Journal of Ornithology
https://doi.org/10.1007/s10336-019-01714-2

ORIGINAL ARTICLE

Mitochondrial phylogeography of the genus *Eremophila* confirms underestimated species diversity in the Palearctic

Fatemeh Ghorbani[1] · Mansour Aliabadian[1,2] · Urban Olsson[3] · Paul F. Donald[4] · Aleem A. Khan[5] · Per Alström[6,7,8]

Received: 26 October 2018 / Revised: 10 August 2019 / Accepted: 15 September 2019

Abstract

Phylogeographic analyses of the genus *Eremophila* (Horned Lark *E. alpestris* and Temminck's Lark *E. bilopha*) were carried out based on the mitochondrial cytochrome *b* and ND2 genes. Four primary lineages with para-/allopatric distributions were identified: (1) a Qinghai–Tibetan–Himalayan lineage; (2) a North African and Middle Eastern lineage; (3) a northwest African and southeast European/southwest Asian lineage; and (4) a Northern Palearctic and North American lineage. The relationships between these four lineages were poorly resolved. They were estimated to have diverged in the late Pliocene to early Pleistocene, although the dates are uncertain due to topological ambiguity and wide confidence intervals. The sublineages were estimated to have diverged around the Middle Pleistocene (c. 0.8–0.2 mya). A strong signal of population growth and range expansion was observed from the Middle Pleistocene, at least in the North Palearctic subclade (A2). Morphometric analysis of the Eurasian taxa revealed a high degree of overlap among taxa, although *E. bilopha* and *E. a. longirostris* stood out from the others. We support a recent suggestion to split *E. alpestris* into multiple species, although we propose four instead of six species, corresponding to the four primary lineages identified in this study: (1) Himalayan Horned Lark *E. longirostris* (by priority and on the premise that the genetically unsampled taxon *longirostris* belongs to this clade); (2) Temminck's Lark *E. bilopha*; (3) Mountain Horned Lark *E. penicillata*; and (4) Common Horned Lark *E. alpestris* (sensu stricto). Our results illustrate the discrepancy between phylogenetic relationships and phenotype in larks.

Keywords Horned Lark · Qinghai–Tibetan Plateau · Species complex · Phylogeny · Taxonomy

Communicated by M. Wink.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10336-019-01714-2) contains supplementary material, which is available to authorized users.

✉ Mansour Aliabadian
aliabadi@um.ac.ir

[1] Department of Biology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran

[2] Research Department of Zoological Innovations, Institute of Applied Zoology, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran

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

[4] Conservation Science Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

[5] Zoology Division, Institute of Pure and Applied Biology, Bahauddin Zakariya University, Multan, Pakistan

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

[7] Swedish Species Information Centre, Swedish University of Agricultural Sciences, PO Box 7007, 75007 Uppsala, Sweden

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

Published online: 27 September 2019

Zusammenfassung

Mitochondriale Phylogeografie der Gattung *Eremophila* bestätigt eine unterschätzte Artenvielfalt in der Paläarktis. Phylogeografische Untersuchungen der Gattung *Eremophila* (Ohrenlerche *E. alpestris* und Hornlerche *E. bilopha*) auf der Basis der Analyse des mitochondiralen Cytochrom b und ND2 Gene ergaben vier primäre Abstammungslinien mit para-/allopatrischen Verbreitungen: (1) eine Qinghai–Tibetische–Himalaya Linie; (2) eine nordafrikanische und nahöstliche Linie; (3) eine nordostafrikanische und südosteuropäische/südwestasiatische Linie; und (4) eine nordpaläarktische und nordamerikanische Linie. Die Verbindungen zwischen diesen vier Abstammungslinien waren bisher nur unzureichend aufgeklärt. Die Aufspaltung wurde im späten Pliozän bis ins frühe Pleistozän geschätzt, obwohl diese Daten, aufgrund von räumlichen Unsicherheiten und weiten Konfidenzintervallen, als unsicher gelten. Für die Unterstämme wurde eine Aufspaltung im mittleren Pleistozän geschätzt (c. 0.8–0.2 Millionen Jahre). Deutliche Hinweise auf ein Populationswachstum und eine Arealausweitung gibt es im mittleren Pleistozän, zumindest im nordpaläarktischen Unterstamm (A2). Morphometrische Analysen der eurasischen Taxa ergaben einen hohen Überlappungsgrad zwischen den Taxa, obwohl *E. bilopha* und *E. a. longirostris* sich von den anderen abheben. Wir unterstützen einen neuen Vorschlag, *E. alpestris* in mehrere Arten aufzuspalten. Wobei wir vier anstatt sechs Arten vorschlagen, entsprechend der vier, in dieser Studie identifizierten primären Abstammungslinien: (1) Himalaya Ohrenlerche *E. longirostris* (vorrangig und unter der Voraussetzung, dass das genetisch bisher nicht beprobte Taxon *longirostris* zu diesem Stamm gehört); (2) Hornlerche *E. bilopha*; (3) Gebirgsohrenlerche *E. penicillata*; und (4) Ohrenlerche *E. alpestris* (sensu stricto). Unsere Ergebnisse verdeutlichen die Diskrepanz zwischen phylogenetischen Beziehungen und dem Phänotyp bei Lerchen.

Introduction

Genetic diversity and demographic structuring of many species have been formed by Quaternary glacial and interglacial climatic oscillations and fueled by consequent ecological changes (Avise 2000; Hewitt 2004; Weir and Schluter 2004; Sandel et al. 2011; Lei et al. 2014). The effects of the late Pliocene and Pleistocene climatic fluctuations on the genetic structure of organisms inhabiting the Holarctic region have been documented for many forest birds (Klicka and Zink 1999; Zink et al. 2002; Weir and Schluter 2004; Saitoh et al. 2010; Zhao et al. 2012). Other studies have focused on open-habitat adapted species (Voelker 1999; Pavlova et al. 2003; Qu et al. 2010; Aliabadian et al. 2012; Li et al. 2016; Liu et al. 2017; Harris et al. 2018; Song et al. 2018), which may differ from forest species in their responses to climate change.

The genus *Eremophila* has a complex taxonomic history due to the high spatial (e.g. de Juana et al. 2004) and temporal (Mason and Unitt 2018) plasticity of its plumage coloration and structure. In his review of the genus (then known as *Otocorys*), Bianchi (1904) recognized eight species, but the tendency since then has been to reduce this number and most sources (e.g. Peters 1960; Cramp 1988; de Juana et al. 2004; Dickinson and Christidis 2014; del Hoyo and Collar 2016; Clements et al. 2018; Gill and Donsker 2018; Shirihai and Svensson 2018) currently recognize two species: Horned Lark *E. alpestris* (Linnaeus, 1758), which is widespread across the Holarctic and also occurs south through Mesoamerica into northern Colombia, in many different open habitats such as grasslands, tundra, farmland, deserts, and alpine habitats and Temminck's Lark *E. bilopha* (Temminck, 1823), which is limited to desert areas of North Africa and the Middle East. Horned Lark breeds on five continents, and is the only lark that has successfully colonized and extended its range into the New World. Among the extensive range of Horned Lark, up to 42 subspecies have been recognized together with a large number of synonyms (Peters 1960; Dickinson 2003; Dickinson and Christidis 2014; del Hoyo and Collar 2016; Gill and Donsker 2018). Temminck's Lark is currently considered a monotypic species (Dickinson and Christidis 2014; del Hoyo and Collar 2016; Gill and Donsker 2018). Horned Lark subspecies have been divided into two groups: one in which the black face mask is separated from the black breast band (in all North American, North Palearctic, and Qinghai–Tibetan–Himalayan subspecies and North African *atlas*, although in *atlas*, there is often a thin band connecting the two; van den Berg 2005), and one in which the black face mask is broadly connected to the black breast band, thereby completely surrounding the pale throat (in the Palearctic subspecies, *penicillata*, *bicornis*, *albigula*, and *balcanica*) (Sharpe 1890; Hellmayr 1929; Whistler 1932; Vaurie 1951; de Juana et al. 2004).

Based on a molecular phylogenetic analysis of the family Alaudidae, the genus *Eremophila* was recently suggested to be a sister to one of the two clades in the traditional genus *Calandrella* (Alström et al. 2013). The same study suggested, albeit with poor support, that the Horned Lark was paraphyletic with respect to Temminck's Lark, and also found deep mitochondrial DNA (mtDNA) divergences among different subspecies of the former, indicating that multiple species should be recognized. Drovetski et al. (2014) undertook a comprehensive phylogeographic analysis of *Eremophila* based on one mitochondrial gene and two nuclear introns. They identified at least seven deeply diverged lineages and corroborated the previous finding that

Temminck's Lark is nested within the Horned Lark complex. They proposed that the Horned Lark should be treated like the five Palearctic (*E. atlas*, *E. penicillata*, *E. brandti*, *E. flava E. elwesi*) and one Nearctic (*E. alpestris* sensu stricto) species in an arrangement that closely mirrored Bianchi's (1904) original proposal.

In this study, we analyzed the phylogeography of the genus *Eremophila* based on mtDNA from throughout the range of the genus, using both new samples and sequences from GenBank. We also estimated divergence times and demography, analyzed biometric data, and finally proposed a revised taxonomy. Compared to earlier studies, we increased the sampling, in particular from Central Asia and the Qinghai–Tibetan–Himalayan region, and have tried to improve the divergence time estimates and demographic history of the genus.

Material and methods

DNA sampling

In total, 40 new DNA samples were obtained; 20 tissue samples, including liver, pectoral muscles, and blood, were collected specifically for this study and deposited at the Department of Zoology, University of Gothenburg, Sweden (DZUG) and Zoology Museum, Ferdowsi University of Mashhad, Iran (ZMFUM), and 20 tissue samples were borrowed from the ornithological collections at the Burke Museum, University of Washington, Seattle, USA [UWBM] and Swedish Museum of Natural History, Stockholm, Sweden [NRM]; Table 1).

Muscle and liver samples were preserved in 20% dimethylsulfoxide (DMSO) or ethanol (75–95%). Blood samples were mixed immediately in a blood storage buffer (0.1 M Tris–HCL, 0.04 M EDTA.Na_2, or 1.0 M NaCl, 0.5% SDS). We also included GenBank sequences in our analyses (from Tieleman et al. 2003; Qu et al. 2010; Alström et al. 2013; Mason et al. 2014; Drovetski et al. 2014; Fig. 1). Additionally, *Calandrella brachydactyla* (GenBank sequences KX379944 and KF735311) and *Alaudala* (formerly *Calandrella*) *raytal* (KF060423 and KJ455343) were used as outgroups.

DNA extraction and amplification

Total genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the recommended protocol for tissue and blood samples. The mitochondrial cytochrome *b* gene (cyt *b*) was amplified and sequenced using the primers and protocol described by Olsson et al. (2005) and the mitochondrial ND2 gene was sequenced following Sorenson et al. (1999). All sequences have been deposited in GenBank (Table 1).

Phylogenetic reconstruction

The sequences were aligned and trimmed using MegAlign 4.03 in the DNAstar package (DNAstar Inc.). Three different datasets were created: one for each gene, including the GenBank sequences plus the new sequences (from 40 tissue samples, 36 ND2 sequences, and 34 cyt *b* sequences were provided; some amplification processes failed), and one for both genes concatenated including some GenBank sequences (Table 1). We used jModelTest (Darriba et al. 2012) and PartitionFinder (Lanfear et al. 2012, 2017) to select the best-fit models of molecular evolution based on the Bayesian Information Criterion (BIC; Schwarz 1978) and the corrected Akaike Information Criterion (AICc), respectively. The best model was GTR + G for cyt *b* and HKY + I + G for ND2, according to jModelTest and HKY + I (cytb_1stpos and 2ndpos), TIM (cytb_3rdpos) and TVM + I (ND2_1stpos), HKY + G (ND2_2ndpos), TRN + G (ND2_3rdpos), according to PartitionFinder. Xml files were generated in BEAUti 2 (Bouckaert et al. 2014). The data were analyzed by Bayesian inference (BI) in BEAST v.2.2.1 (Bouckaert et al. 2014), separately and concatenated, both under a relaxed log-normal clock and a strict clock models. Default priors were used. The analysis with higher posterior probability (a relaxed log-normal clock model) was adopted in the final analyses. In the concatenation analyses, the two loci were treated as separate partitions. We ran the Markov Chain Monte Carlo (MCMC) method for 100 million generations. Convergence to the posterior distributions of the parameter estimates was evaluated by monitoring the effective sample size (ESS > 200) and trace plots in Tracer 1.6 (Rambaut et al. 2014). We used TreeAnnotator v.2.2.1 (Rambaut and Drummond 2015) in the BEAST package to summarize trees with mean height and discarded the first 10% of the trees as burn-in. The tree was displayed in FigTree 1.4.3 (Rambaut 2016).

Divergence times

Divergence times were estimated on the combined dataset (cyt *b* + ND2) in BEAST, with a normally distributed clock prior on cyt *b*, with a mean rate of 0.0105 substitutions/site/lineage/million years (my) and standard deviation 0.001, corresponding to a rate of 2.1% per million years (my) (Weir and Schluter 2008) and allowed ND2 to be estimated. We applied the substitution models GTR + G for cyt *b* and HKY + I + G for ND2, and ran several analyses, with and without outgroup, and using different models (relaxed log-normal and strict clock). The final analyses were run under

Table 1 List of taxa and loci for which original sequence data were produced for this study, with geographic origin, museum reference number, and GenBank accession number

No.	Sample ID	Taxon	Museum ID	Cl.	Locality	Country	Accession number	
							cyt *b*	ND2
1	U5084	*E. a. alpestris*	UWBM 68253	A1	Newfoundland	USA	MN317107	MN365855
2	U5085	*E. a. alpestris*	UWBM 68254	A1	Newfoundland	USA	MN317108	MN365856
3	U5086	*E. a. alpestris*	UWBM 68255	A1	Newfoundland	USA	MN317109	MN365857
4	U5093	*E. a. arcticola*	UWBM 53941	A1	Alaska	USA	MN317116	MN365864
5	U5094	*E. a. arcticola*	UWBM 53943	A1	Alaska	USA	MN317117	MN365865
6	U5095	*E. a. strigata*	UWBM 96010	A1	California	USA	MN317118	MN365866
7	U5097	*E. a. strigata*	UWBM 97721	A1	Nevada	USA	MN317120	MN365868
8	U5098	*E. a. strigata*	UWBM 97722	A1	Nevada	USA	MN317121	MN365869
9	U5096	*E. a. utahensis*	UWBM 99265	A1	Utah	USA	MN317119	MN365867
10	U5099	*E. a. lamprochroma*	UWBM 46848	A1	California	USA	MN317122	MN365870
11	U5091	*E. a. flava*	UWBM 59589	A2	N Ural	Russia	MN317114	MN365862
12	U5092	*E. a. flava*	UWBM 59593	A2	N Ural	Russia	MN317115	MN365863
13	U5223	*E. a. flava*	NRM946610	A2	Yamalia	Russia	MN317123	MN365871
14	U5224	*E. a. flava*	NRM946619	A2	Nenetsie	Russia	MN317124	MN365872
15	U5225	*E. a. flava*	NRM946642	A2	Nenetsie	Russia	MN317125	MN365873
16	U2491	*E. a. brandti*	DZUG	A2		Kazakhstan	NA	MN365851
17	U5087	*E.a. brandti*	UWBM 59834	A2	Choibalsan	Mongolia	MN317110	MN365858
18	U5088	*E.a. brandti*	UWBM 59836	A2	Choibalsan	Mongolia	MN317111	MN365859
19	U5089	*E.a. brandti*	UWBM 57949	A2	Gobi desert	Mongolia	MN317112	MN365860
20	U5090	*E.a. brandti*	UWBM 57960	A2	Gobi desert	Mongolia	MN317113	MN365861
21	U4606	*E. a. brandti*	DZUG	A2	Dundgobi	Mongolia	MN317102	MN365846
22	U4609	*E. a. brandti*	DZUG	A2	Khentii	Mongolia	MN317103	MN365847
23	U4610	*E. a. brandti*	DZUG	A2		Mongolia	MN317104	MN365848
24	U4611	*E. a. brandti*	DZUG	A2	Khentii	Mongolia	MN317105	MN365849
25	U4635	*E. a. brandti*	DZUG	A2	Khentii	Mongolia	MN317106	MN365850
26	1008	*E. a. albigula*	ZMFUM	B	Semnan	NE Iran	MN317099	MN365842
27	2539	*E. a. albigula*	ZMFUM	B	Mashhad	NE Iran	MN317101	MN365844
28	1832	*E. a. albigula*	ZMFUM	B	Mashhad	NE Iran	MN317100	MN365843
29	1204061164	*E. a. penicillata*	ZMFUM	B	Zanjan	NW Iran	MN317097	MN365841
30	1204061134	*E. a. penicillata*	ZMFUM	B	Zanjan	NW Iran	MN317094	NA
31	1204061163	*E. a. penicillata*	ZMFUM	B	Zanjan	NW Iran	MN317096	MN365840
32	1204061166	*E. a. penicillata*	ZMFUM	B	Zanjan	NW Iran	MN317098	NA
33	1204061158	*E. a. penicillata*	ZMFUM	B	Zanjan	NW Iran	MN317095	MN365839
34	U0615	*E. a. atlas*	DZUG	B		Morocco	NA	MN365845
35	U0612	*E. bilopha*	DZUG	C		Morocco	NA	MN365874
36	U4551	*E. a. elwesi*	DZUG	D	Qinghai	China	NA	MN365853
37	U4565	*E. a. elwesi*	DZUG	D	Qinghai	China	NA	MN365852
38	U1555	*E. a. deosaiensis*	DZUG	D		Pakistan	NA	MN365854
39	U1557	*E. a. deosaiensis*	DZUG	D		Pakistan	MN317126	NA
40	U1558	*E. a. deosaiensis*	DZUG	D		Pakistan	MN317127	NA
GenBank	–	*E. a. penicillata*		B		Iran	*KF060442	NA
GenBank	–	*E. a. atlas*		B		Morocco	NA	**KF735322
GenBank	–	*E. a. atlas*		B		Morocco	NA	**KF735321
GenBank	–	*E. bilopha*		C		Morocco	NA	**KF735318
GenBank	–	*E. a. elwesi*		D	Qinghai	China	NA	**KF735315
GenBank	–	*E. a. elwesi*		D	Qinghai	China	NA	**KF735314

The GenBank sequences marked with * were obtained from Alström et al. (2013) and ** from Drovetski et al. (2014). The samples Eral1 and Erbi1 were provided by Pierre–André Crochet and Alban Guillaumet. (UWBM): The Burke Museum, University of Washington, Seattle, USA, (NRM): Swedish Museum of Natural History, Stockholm, Sweden, (DZUG): Department of Zoology, University of Gothenburg, Sweden, (ZMFUM): Zoology Museum, Ferdowsi University of Mashhad, Iran

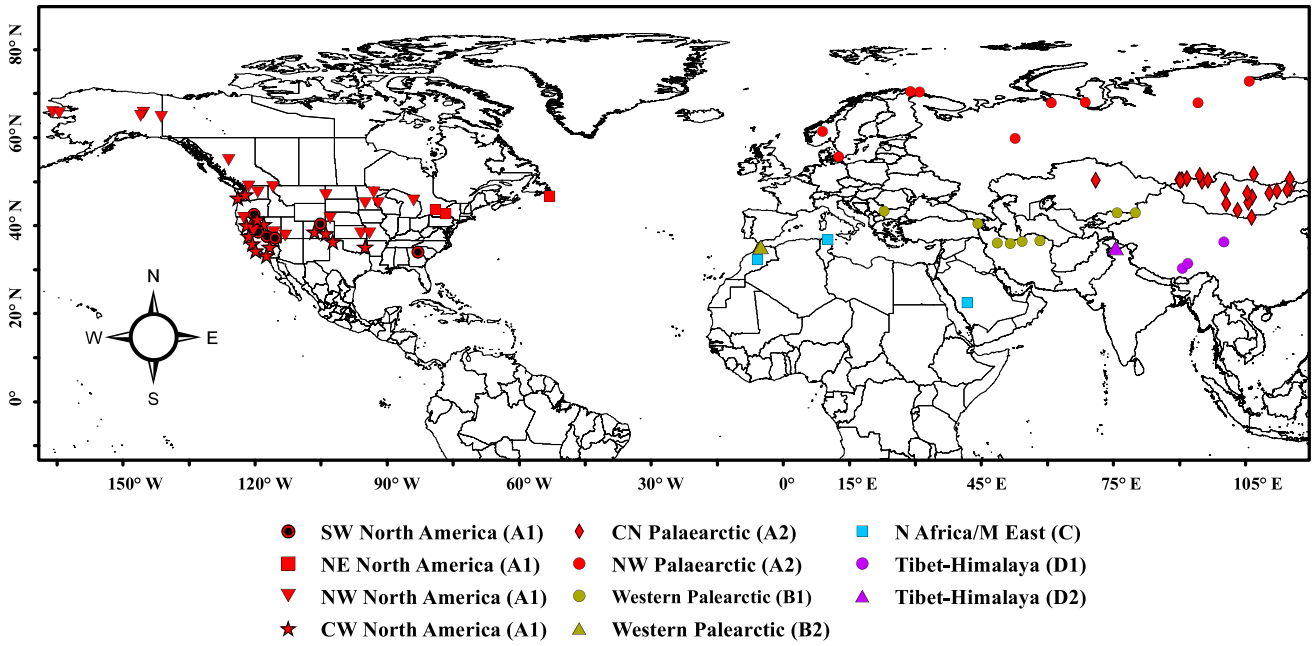

Fig. 1 Genetic sampling, based both on samples used in previous studies (Alström et al. 2013; Drovetski et al. 2014) and obtained specifically for the present study. Colors represent the primary clades defined in Fig. 2 (Red: Clade A; Olive: Clade B; Sky blue: Clade C; Violet: Clade D) (color figure online)

an uncorrelated log-normal relaxed clock model (Drummond et al. 2006) and a birth–death incomplete sampling tree prior. See above for further details.

Demographic analyses

The historical demography was assessed using neutrality tests to determine departure from different null hypotheses. To test the effect of sample size on the demographic results we ran the analyses on four different datasets, including the total number of sequences available for ND2 and cyt *b* (from GenBank and the new sequences obtained for this study), two separate datasets of in total 30 individuals for cyt *b* and ND2, including the same number of individuals from each locality for both genes. The descriptive statistics of nucleotide variation and population differentiation was calculated in DNAsp 6.0 (Rozas et al. 2017) for each lineage. Fu's Fs (Fu 1997), Tajima's D (Tajima 1989), and Ramos-Onsin and Rozas' statistic R2 (Ramos-Onsins and Rozas 2002) were calculated in DNAsp with 1000 simulation replications for each population to test the significance. The time since expansion for each clade were calculated from the equation $t_{exp} = \tau/2u$, using the mismatch calculator by Schenekar and Weiss (2011; tool provided by S. Weiss), where τ (Tau) is a unit of mutational time and u is the cumulative substitution rate per generation across the DNA fragment.

Patterns of historical demography were also inferred from the estimates of effective population size over time using Bayesian skyline in BEAST. Sequence data representing each specific lineage were analyzed separately. We set the substitution rate to 2.1% per million years (Weir and Schluter 2008) for cyt *b*. Since there is a lack of agreement on the substitution rate of ND2, three substitution rates were adopted and compared from different studies: 0.0073 (1.46% sequence divergence; Pereira and Baker 2006), 0.01 (2%; Smith and Klicka 2013), and 0.029 (5.8%; Lerner et al. 2011) substitutions/site/lineage/million years, respectively. The null hypothesis is the constant population size coalescent model. Together, among-site rate heterogeneity across all branches and a log-normal relaxed clock were used for this calculation. Markov chains were run for 500 million generations and sampled every 100,000 generations with the first 25% samples discarded as burn-in. Other parameters were set as default values and results were visualized in Tracer 1.6.

Morphometric analyses

Morphometric data from 228 adult male specimens of *E. alpestris* were obtained from the Natural History Museum, Tring, UK (NHMUK, formerly BMNH). Information on subspecies identification, location, age, and sex of all samples used in morphometric analyses was taken from the corresponding museum labels. Specimens in the NHMUK are grouped by subspecies; among these groupings, the identity of each specimen was checked with reference to collection

date, locality, and the type description; specimens that appeared to have been misclassified were excluded. Wing length (flattened and stretched) and tail length, were measured to the nearest 1 mm with a ruler, and bill length (tip to skull) was measured to the nearest 0.1 mm with a digital caliper. The same measurements from 17 males of *E. bilopha* were provided by Lars Svensson. The data were standardized and the squared cosine was calculated as a measure of the linear correlation between variables (Abdi and Williams 2010). Shapiro tests were used to check whether variables were normally distributed and Wilks' lambda applied for contributing variables in the discriminant function. We examined the assumption of homogeneity of variance with the Bartlett test. We used ANOVA to assess differences in means between taxa. Pairwise significant differences between taxa were determined by post hoc Tukey tests.

Linear discriminant function analysis (LDA) was used to assess subspecies classification. LDA describes intergroup differences by linear combinations of all variables. LDA was fitted by the 'lda' function of the 'MASS' package (Venables and Ripley 2002) in R (R Development Core Team 2018). The contribution of individuals (in percentage) to the principal components was also calculated. A principal component analysis (PCA) of the covariance matrix of log-transformed measurements was used to visualize variation among taxa by plotting multivariate ordination of specimens on the first two principal component (PC) axes (PC1 and PC2).

The morphometric data were also grouped and analyzed based on distribution of different subspecies in the clades derived from molecular analyses and the results were compared. *Eremophila a. montana* ($n = 6$) was not included in this analysis due to its unclear position in the phylogenetic tree (resulting from lack of molecular data), and to a suspicion during specimen examination that two taxa may have been involved. All analyses were performed in R (R Development Core Team 2018).

Results

Molecular analysis

In total, 1038 base pairs (bp) were obtained for ND2 and 999 base pairs for cyt *b* after excluding sites with missing data for some samples. For cyt *b* there were 150 polymorphic sites (69 singleton variable sites and 81 parsimony informative sites), while in ND2 there were 293 polymorphic sites (200 singleton variable sites and 93 parsimony informative sites, Table S1).

Phylogenetic analysis of concatenated dataset

In the BEAST analysis of the combined cyt *b* and ND2 data (46 samples), four well-supported (posterior probability, PP 1.00), deeply diverged primary clades corresponding to different geographical regions/taxa were recovered (Figs. 1, 2): (A) Russia (*E. a. flava*), Kazakhstan and Mongolia (*E. a. brandti*), and North America (*E. a. alpestris*, *E. a. utahensis, E. a. lamprochroma*, *E. a. strigata,* and *E. a. arcticola*); (B) Iran (*E. a. penicillata* and *E. a. albigula*) and Morocco (*E. a. atlas*); (C) Morocco (*E. bilopha*); and (D) Qinghai–Tibetan Plateau (*E. a. elwesi*) and North Pakistan (*E. a. deosaiensis*). The statistical support for the relationships among these four lineages was low, and accordingly *E. bilopha* was nested within *E. alpestris*, but there was no support for the exact position.

Two of these primary clades were further subdivided in accordance with different regions/taxa: (A1) North America (PP 1.00), with further geographical subdivision into four, less well defined, geographical areas; (A2) North Palearctic (PP 1.00), with North Russian and Mongolian samples in separate well-supported (PP 1.00) clades, and with the single Kazakhstan sample in an unsupported position (PP 0.5); (B1) Iran (PP 1.00); and (B2) Morocco (PP 1.00).

Divergence time estimates

In the analysis based on the combined dataset (cyt *b* + ND2), the most recent common ancestor (MRCA) is inferred to have lived in the late Pliocene [3.3 million years ago (mya); highest posterior density (HPD) 2.2–4.7 mya]. The splits among the four primary lineages (A–D) were inferred to have occurred in the late Pliocene, with wide HPDs extending into the Pleistocene (Fig. 2). The splits between the subclades were all dated to the Middle Pleistocene (A1–A2: 0.7 mya; HPD 0.4–1 mya; B1–B2: 0.6 mya; HPD 0.3–0.9 mya), well before the onset of the most recent of the four major Pleistocene glaciations.

Phylogenetic analyses of single loci

The BEAST tree based on ND2 including 310 samples (Fig. S1) is topologically similar to the tree based on the concatenated sequences (Fig. 2) (PP 1.0) except that clade A1 shows well-supported structure and that the position of *E. bilopha* (clade C) as nested within *E. alpestris* was highly supported (PP 1.0). It also shows *E. a. balcanica* belongs to clade B (no cyt *b* sequences available for this taxon). In the BEAST tree based on cyt *b*, including 70 samples (Fig. S2), the sister relationship between the primary clades A and D and between B and C shows low support (PP 0.79 and 0.47, respectively).

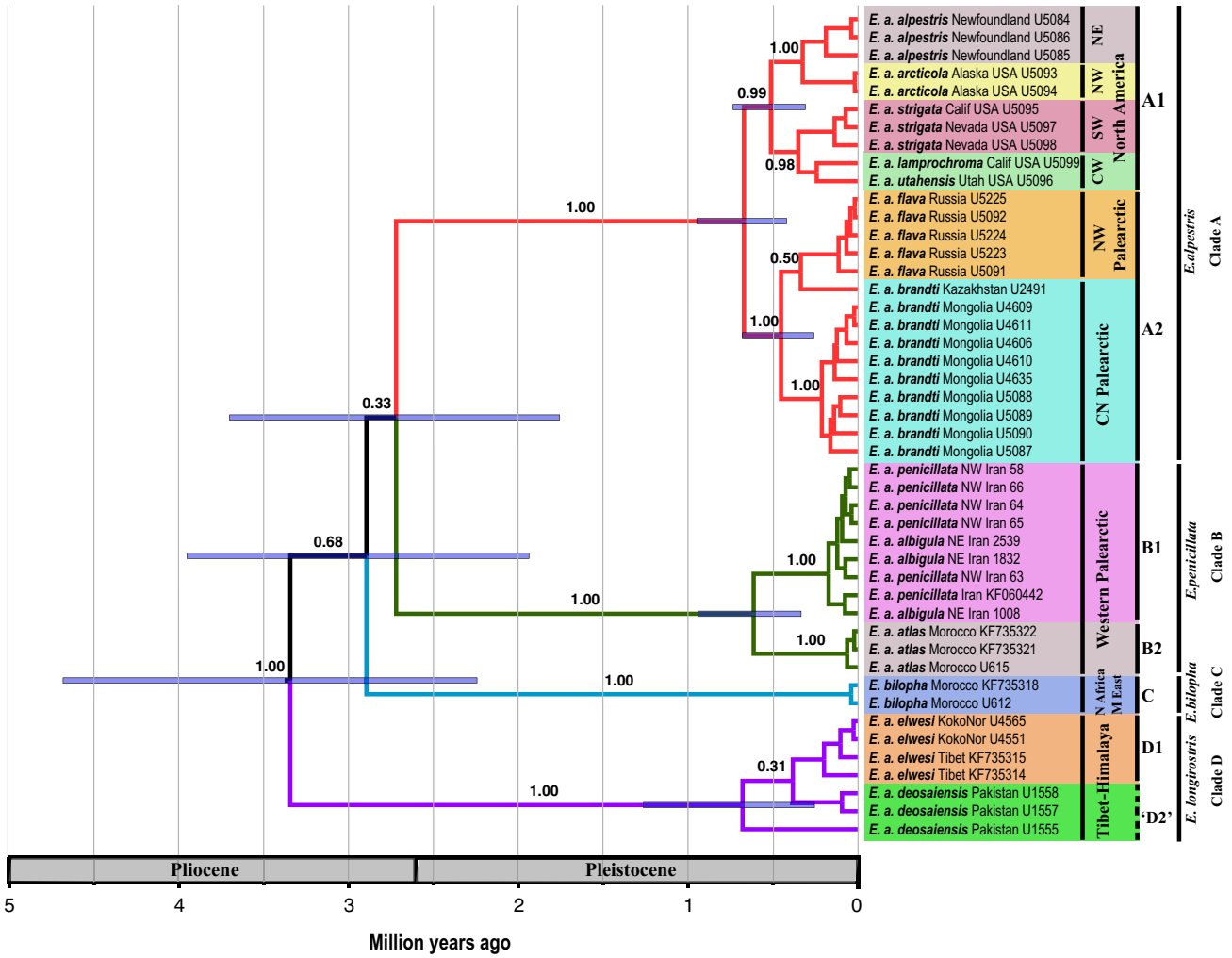

Fig. 2 Chronogram, estimated by Bayesian analysis of concatenated sequences of the mitochondrial cytochrome *b* and ND2 (in total 2037 bp). The values on the branches are posterior probabilities. The labels A1–D2 represent clades discussed in the text. The names on clades A–D refer to the revised species taxonomy proposed here. Outgroups are not shown

Table 2 Genetic diversity based on cyt *b* and ND2 mitochondrial sequences (30 samples) from the same individuals

Clade	*N*	nH	Hd	π	*V*	*S*	*P*	Tajima's *D*	Fu's Fs	R^2	Tau	T_{exp}
Cyt *b*												
A1	10	8	0.96 ± 0.59	0.0078 ± 0.0008	24	12	12	− 0.55	− 0.86	0.13	6.31	NA
A2	14	5	0.93 ± 0.12	0.0026 ± 0.0007	22	16	6	− 1.19	**− 3.71***	**0.08*****	3.64	173.555
B	6	5	0.93 ± 0.12	0.0026 ± 0.0007	7	1	1	− 1.01	− 1.62	**0.15***	2.40	114.400
ND2												
A1	10	8	0.96 ± 0.06	0.0058 ± 0.0007	17	5	12	− 0.02	− 1.52	0.15	4.61	NA
A2	14	10	0.87 ± 0.09	0.0070 ± 0.0014	27	18	9	− 0.81	− 0.66	0.14	2.82	NA
B	6	5	0.93 ± 0.12	0.0063 ± 0.0024	18	18	0	**− 1.59*****	0.11	0.25	1.50	NA

Statistical parameters indicate number of samples (*N*), number of haplotypes (nH), haplotype diversity (Hd ± SD), and nucleotide diversity ($\pi \pm$ SD), Singleton variable sites (*S*), number of variable sites (*V*), parsimony informative sites (*P*), Tajima's *D*, Fu's Fs and R2 (Ramos–Onsin and Rozas' statistic). $P < 0.05$ are in bold

A, *alpestris*; B, *penicillata*; Tau, age of expansion in units of mutational time; T_{exp}, time since onset of population expansion (1000 years ago) (calculated only for populations inferred to deviate from demographic equilibrium)

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Haplotype and demographic analyses

Thirty samples were available for both cyt *b* and ND2 from the same individual (Table 2). No departure from the null hypothesis of demographic equilibrium was indicated for the different populations based on this cyt *b* dataset, except for clade A2 and B. Both Fu's Fs and R2 suggest deviation from equilibrium for clade A2, corroborated by the skyline plot. Since none of the estimators, Tajima's *D*, Fu's Fs, and R2 suggested departure from equilibrium for clade A1 on the cyt *b* dataset, T_{exp} was only calculated for clade A2 and B (Table 2). The estimated time of onset of the inferred expansion for clade A2 was 174 thousand years ago (kya) (Middle Pleistocene), which is consistent with the skyline plot (Fig. 3). The skyline plots based on cyt *b* show a steady increase in population size, leveling out toward the present

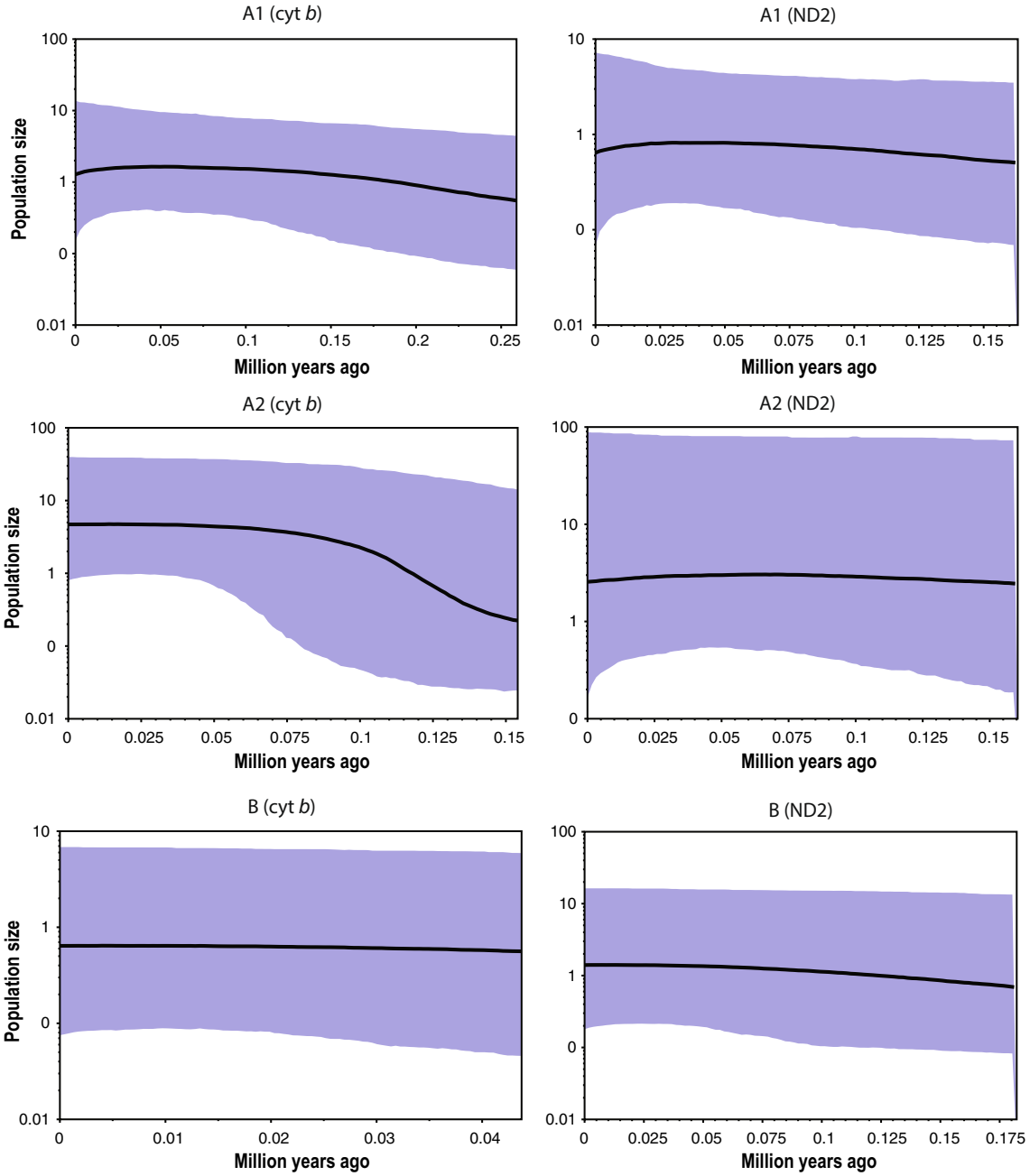

Fig. 3 Skyline plots for the cytochrome *b* and ND2 (30 samples same as in Table 1). The Bayesian skyline plots indicate effective population sizes (scaled by 2.1% divergence rate per million years) for cyt *b* plotted as a function of time in million years. Black lines indicate the median value of effective population size; solid areas denote the 95% highest posterior density. Skyline plots based on cyt *b* for clades A1 and A2 were estimated under the HKY + I model and for the clade B under the HKY model

for clade A1, starting at least 250 kya (Middle Pleistocene), but this is not corroborated by the demographic estimators (Table 2). The skyline plots for clade B suggest nearly constant population sizes, while the expansion time estimated is 114 kya (Fig. 3, Table 2). The skyline plots based on the ND2 dataset (2% sequence divergence) estimated a very slight tendency of increase in population size with some decline at the start of the Holocene for clade A1 (Fig. 3, Table 2). The population sizes of clades A2 and B are inferred to have remained stable (Fig. 3, Table 2). T_{exp} was not calculated for any of the clades in the ND2 dataset since only the weakest indicator, Tajima's *D* indicates any deviation from equilibrium (Table 2). The skyline plots based on the ND2 dataset with time estimates calculated based on 1.46% or for 5.8% sequence divergence per million years are very similar to those based on a 2% divergence, except for clade B, for which a slight more or less steady population expansion is indicated by the skyline plot (data not shown).

By combining data from GenBank and own sequencing, a total of 70 cyt *b* sequences and 310 ND2 sequences were available (Table S1). The skyline plots based on all available sequences of ND2 (2% sequence divergence per million years) indicate a tendency of increase in population size during the Middle Pleistocene for clade A1 (Fig. S3). For clade A2, the results differ from the calculations based on the smaller data set (Fig. 3, Table 2) shows a more evident population increase which is inferred to have started around 125 kya and leveled off around 50 kya (Fig. S3). These results were corroborated by significant departure from the null hypothesis of demographic equilibrium for Tajima's D, Fu's Fs, and R2 for clades A1 and A2 (Table S1). For clade B, a more or less stable population is inferred by the ND2 skyline plot, but with a slight tendency for a decline in effective population size increase coinciding with the start of the Last Glacial Maximum (LGM; 22 kya) (Fig. S3). Contrary to this, the demographic estimators Tajima's D, Fu's Fs, and R2 suggest deviation from population equilibrium for clade B (Table S1). The skyline plot indicates a stable population size for clade C (Fig. S3), corroborated by the demographic estimators (Table S1). For clade D, the skyline plot for the ND2 data indicates a continuous slight decline in effective population size for most of the period, with a steeper decline after the LGM (around 15 kya) (Fig. S3). The demographic estimators Fu's Fs and R2 do not contradict this for the ND2 data (Table S1). However, for the cyt *b* data, based on 32 sequences (clade D), the demographic indicators suggest deviation from equilibrium and a population expansion around 163 kya (Table S1). The skyline plot is consistent with this, but leveling off from around 100 kya, in a similar pattern, as for clade A2 based on the ND2 data set (Fig. S3), but both onset and decline in increase happening somewhat earlier.

Univariate morphometric analyses

ANOVAs of the linear measurements revealed significant differences in wing length ($F_{13,215} = 112.61$; $P < 0.001$), tail length ($F_{13,215} = 58.40$; $P < 0.001$), tail/wing (%) ($F_{13,215} = 13.56$; $P < 0.001$), and bill length ($F_{13,215} = 47.98$; $P < 0.001$), among the taxa (Table 3 and Fig. S4).

Pairwise comparisons between *E. bilopha* and the 13 subspecies of *E. alpestris* revealed that *E. bilopha* differs significantly from all subspecies of *E. alpestris* at least in two characters (wing and tail lengths; $P \leq 0.001$; Table S2

Table 3 Mean and standard deviations of three morphometric data for 228 studied individuals of 14 taxa

Subspecies	*N*	Wing (mean ± SD)	Min	Max	Tail	Min	Max	Bill	Min	Max
E. a. elwesi	19	117.26 ± 2.96	110	121	84.89 ± 4.69	77	93	16.20 ± 0.77	14.5	17.4
E. a. elwesi (*khamensis*)	3	121.67 ± 2.89	120	125	84.67 ± 5.03	80	90	17.27 ± 0.64	16.8	18
E. a. elwesi (*nigrifrons*)	2	115.50 ± 2.12	114	117	82.50 ± 3.54	80	85	16.85 ± 1.34	15.9	17.8
E. a. longirostris	26	125.58 ± 3.01	119	131	87.77 ± 3.46	79	93	19.96 ± 1.25	17.2	21.7
E. a. albigula	20	117.35 ± 2.41	114	123	80.80 ± 4.17	71	88	17.19 ± 0.73	15.8	18.6
E. a. atlas	14	111.86 ± 2.77	106	116	77.93 ± 3.10	73	83	16.81 ± 1.01	15.1	18.5
E. a. bicornis	16	114.00 ± 2.76	109	118	77.69 ± 2.27	73	82	17.72 ± 0.79	16.4	19.1
E. a. penicillata	13	119.77 ± 1.64	117	122	82.54 ± 3.82	75	88	17.81 ± 0.50	16.6	18.5
E. a. brandti	29	110.76 ± 3.51	105	119	78.10 ± 3.27	71	83	15.85 ± 0.87	14	17.5
E. a. flava	44	110.86 ± 1.68	107	115	71.98 ± 3.15	64	78	15.52 ± 0.78	13.7	16.8
E. a. montana	6	115.83 ± 2.93	112	120	81.67 ± 2.80	79	87	16.90 ± 0.33	16.6	17.5
E. a. argalea	16	120.56 ± 2.22	117	126	87.06 ± 2.43	83	91	17.09 ± 0.61	16.2	18.3
E. a. przewalski	4	122.75 ± 3.59	120	128	83.50 ± 3.70	78	86	15.85 ± 0.95	15	16.9
E. bilopha	16	114.00 ± 2.58	109	118	77.69 ± 2.72	73	82	17.73 ± 0.79	16.4	19.1

The values were given in mm

and Fig. S4). Generally, *E. a. longirostris* has the highest mean value in all morphometric characters compared to the other taxa (Table 3). Pairwise comparison revealed no significant differences between *E. a. elwesi*, *E. a. nigrifrons*, *E. a. khamensis*, and *E. a. montana* (Table S2).

Multivariate morphometric analyses

The PCA performed on mean values of linear characters on different subspecies shows PC1 that explains 67.3% of the total variance and reflects differences in tail and wing lengths (Table 4, Fig. 4). PC2 explains 23.5% of the total variance and is interpreted mainly as differences in tail length/wing length (%) and bill length (Table 4, Fig. 4). *Eremophila a. longirostris* is the most divergent taxon, being mostly separated from the other taxa in the main scatter on PC1. *Eremophila bilopha* is separated from all except *E. a. flava* on PC1, but is mostly separated from *E. a. flava* on PC2. The rest of the subspecies of *E. alpestris* form a broadly overlapping set of clusters (Fig. 4). The result from the linear discriminant analysis (LDA; Table 4, Fig. S5) was basically similar to the PCA, although *E. a. longirostris* and *E. bilopha* were better separated from the others on LD1.

The PCA based on the four primary clades provided slightly better separation for *E. bilopha* on PC1, which explained 67.4% of the total variance (Fig. S6). However, there was broad overlap between all clades, especially on PC2.

Table 4 Co-efficients of the two linear discriminant functions (LD1–2) and principal component analysis (PCA 1–2)

Variable	LD1	LD2	PC1	PC2
Wing	− 0.453	1.395	− 0.53	− 0.36
Tail	0.180	− 1.768	− 0.59	0.19
Tail/wing (%)	− 0.295	2.143	− 0.39	0.77
Bill	− 0.473	− 1.067	− 0.46	− 0.49
Eigenvalue	–	–	2.69	0.94
Percentage of explained variance (%)	82.1	9.3	67.3	23.5

Discussion

Phylogenetic inference and divergence times

The identification of four primary, deeply diverged lineages (A–D) is well supported by our molecular data, although

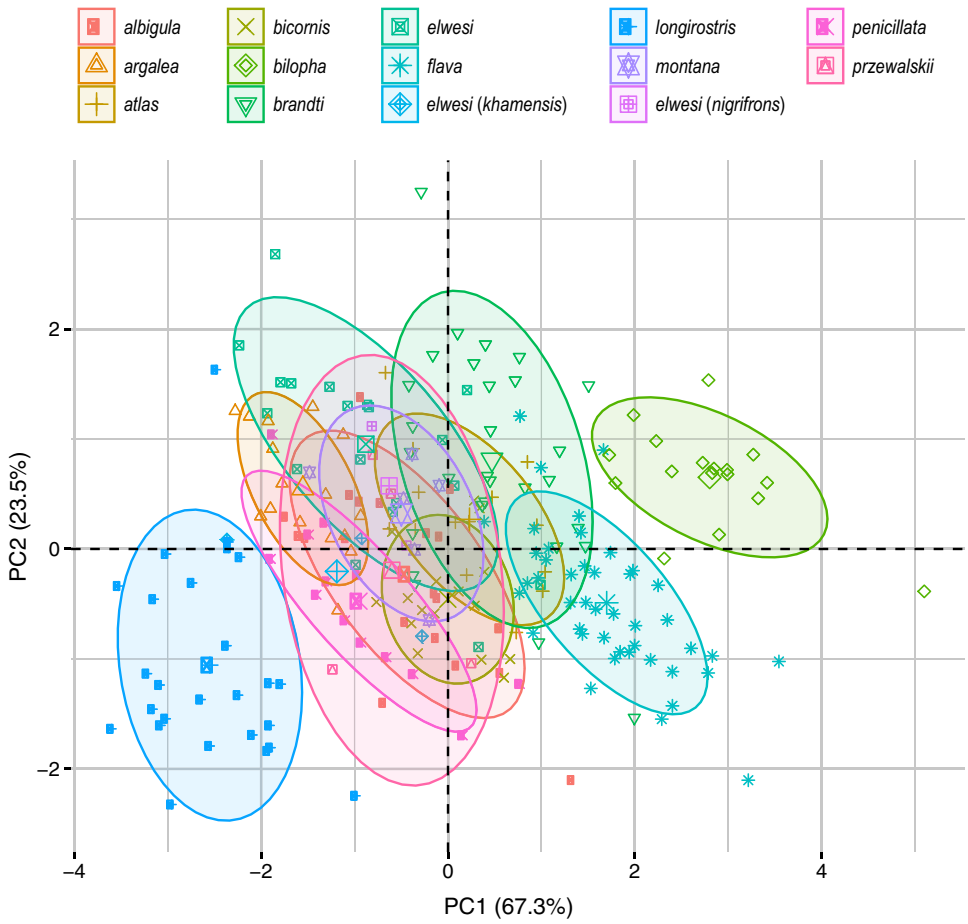

Fig. 4 Scatter-plot of Principal Components Analysis examining differences in four morphometric traits (wing length, tail length, tail/wing [%], and bill length)

the relationships among these lineages are uncertain. The unresolved base of the tree likely stems from conflict between ND2 and cyt *b*, while ND2 inferred the Qinghai–Tibetan–North Pakistan clade (D) as sister to the others with strong support, cyt *b* supported (less strongly) an alternative topology. The separations of North American (A1) vs. North Palearctic (A2), and Iranian (B1) vs. Moroccan (B2) clades are strongly supported. These results are largely congruent with a species tree analysis based on ND2 (partly the same sequences as here) and two nuclear introns (ACO1I9, RHOI1) by Drovetski et al. (2014). Furthermore, in the same analysis, Drovetski et al. (2014) found a sister relationship between *E. a. elwesi* (our clade D1) and *E. bilopha* (our clade C) and between *E. a. alpestris/flava/brandti* (our clade A) and *E. a. penicillata/atlas* (our clade B), although none of these relationships were strongly supported in their analysis (PPs 0.70, 0.88, respectively). Their most strongly supported sister relationship was between *E. a. atlas* and *E. a. penicillata* (PP 0.98), which was also supported in a separate analysis of the nuclear introns, and which is strongly supported in our analyses. The geographical structuring in North America suggested in our analysis of combined ND2 and cyt *b* basically agrees with analyses of ND2 for much larger sample sizes (Drovetski et al. 2014; Mason et al. 2014), however, the larger samples reveal considerable haplotype sharing at least between adjacent areas.

Alström et al. (2013) found that their single *E. a. flava* from Sweden (GenBank KF060442; Swedish Museum of Natural History No. NRM 20046759) is a sister to their single *E. a. atlas*. However, this *E. a. flava* had been accidentally mislabeled and was actually *E. a. penicillata* from Iran, and was accordingly relabeled as *E. a. penicillata* in our analyses.

Our divergence time estimate for the most recent common ancestor (MRCA) of *Eremophila* agrees with that of Alström et al. (2013), based on an analysis of the Alaudidae as a whole. However, the dates are overall considerably older than those of Drovetski et al. (2014) and Mason et al. (2014) (latter only focused on North America, so the only relevant node is the split between Asia and North America). For example, Drovetski et al. (2014) estimated the MRCA of *Eremophila* at 1.4 mya (95% HPD 1.1–1.8 mya), i.e. nearly half of our estimate and with non-overlapping confidence intervals (3.3 mya; 95% HPD 2.2–4.7 mya; Fig. 2). The main reason for the discrepancy between these analyses is that a higher divergence rate gives more recent time estimates. We used the mean molecular clock rate for cyt *b* based on calibrations of more than 70 passerine studies (Weir and Schluter 2008), whereas Drovetski et al. (2014) and Mason et al. (2014) used the mean rate of sequence evolution for ND2 derived from Hawaiian honeycreepers (Drepanidinae) (Lerner et al. 2011). The differing results suggest that one or both of these different divergence rates do not accurately calculate the true clade ages. A parallel case, with similar large differences in dating based on ND2 and cyt *b*, have been found for the avian genus *Prunella* (accentors) (Drovetski et al. 2013; Liu et al. 2017). The relation between the divergence rates of cyt *b* and ND2 within the same datasets is not well understood and the same applies to the possible variation in divergence rates between different groups of birds. We consider our dates more in agreement with the fossil record than those of Drovetski et al. (2014) and Mason et al. (2014). The oldest known fossils of *Eremophila*, of "*Eremophila aff. alpestris*", are from Mongolia/Transbaikal in the late Pliocene, c. 3 mya (MN 16; Zelenkov and Kurochkin 2012). Moreover, there are fossils of *Eremophila prealpestris* from Bulgaria from the early Pleistocene, c. 2.5 mya (MN 17) (Boev 2002, 2012). There are fossils from North America from the mid-Pleistocene, 0.6–0.8 mya (Barnosky 2004), i.e. considerably earlier than the split between the North American and North Palearctic birds estimated by Drovetski et al. (2014) (0.21 mya; 95% HPD 0.13–0.28 mya) and Mason et al. (2014) (c. 0.25 mya, no details given). Although these fossils might represent stem groups rather than the crown group, they nevertheless indicate an older history of the genus than the estimates by Drovetski et al. (2014) and Mason et al. (2014).

The two nuclear markers analyzed by Drovetski et al. (2014) showed strong incomplete lineage sorting and it seems likely that genomic data would be required to more robustly resolve the phylogenetic relationships among the different taxa in this recent radiation.

The demographic results based on cyt *b* suggest an increase in population size for clade A2 starting before or near the beginning of the penultimate glaciation (150–200 kya). The full ND2 data corroborate that both northern clades (A1 and A2) may have experienced population increases since well before the last ice age and the onset of the expansion time (T_{exp}) calculated from Tau is within the confidence interval of the results from the skyline plot. In other words, there are indications that both these clades may have thrived during the penultimate ice age, but not been influenced so much by the most recent one. The cyt *b* data for clade D based on 32 sequences also indicate population expansion about 163 kya, which might suggest that also the Qinghai–Tibetan–Himalayan populations may have been subjected to similar conditions as the northern populations during more or less the same time period. The inferred positive response of the Qinghai–Tibetan–Himalayan and North Palearctic Horned Larks to Pleistocene cooling may be the result of them being able to take advantage of the expansion of steppe-tundra habitats (Yurtsev 1981) that may have been more extensive and less restricted by surrounding features within their ranges, compared to the more southerly populations, as has also been suggested for the genus *Prunella* (Liu et al. 2017). However, the inferred leveling off of the

population increase for clade A2 (Fig. 3) seems to coincide with the increased cooling leading up the LGM. It is not clear that in what way this would have been detrimental to continued population increase, but may have been a consequence of steppe displacement in Siberia and adjacent areas causing shrinking areas of suitable habitat (Sher 1997; Sher et al. 2005). Less severe impact of Pliocene and Pleistocene cooling at lower latitudes (Bobek 1937; Shroder and Bishop 2010) and the buffer role of mountain ranges of the Qinghai–Tibet–Himalayas (Fjeldså et al. 2012; Lei et al. 2015) may have provided more stable conditions for other clades of *Eremophila*, counteracting changes in population sizes.

Biogeography

The topological uncertainties at the base of the tree preclude detailed analysis of the early biogeographical patterns. However, both our tree and the fossil data (see above) strongly suggest that the genus originated in the Palearctic. If the oldest known fossils do indeed represent the ancestral area, this is in agreement with that of another mountain specialist genus, the accentors (*Prunella*), which has been suggested to have originated in the Sino-Himalayan Mountains or these mountains and Central Asia–Mongolia, although considerably earlier than *Eremophila* (Liu et al. 2017; cf. also Drovetski et al. 2013). It seems likely that the four main *Eremophila* lineages originally diverged in different parts of the Palearctic, through habitat fragmentations during the late Pliocene (or possibly early Pleistocene): clade A in the comparatively low-elevation areas in the north/northeast (potentially surviving glacial periods in ice-free areas in the northeast/east, cf. e.g. Saitoh et al. 2010); clade B in mountainous areas in southwest Asia and possibly southern Europe and North Africa (areas with well-established glacial refugia for other taxa; e.g. Frenzel 1973; Hewitt 2000; Tzedakis et al. 2002; Brito 2005; Hughes and Woodward 2017); clade C in lowland areas in North Africa; and clade D within its present range on the Qinghai–Tibetan Plateau (with refugia areas at least along the eastern edge; cf. Qu et al. 2010). During the Pliocene most of the Tarim Basin turned into desert (Zhu et al. 2014). Formation of the Taklamakan desert about 3.6 mya (Zheng et al. 2000, 2002) and expansion of other Central Asian deserts may have created a mosaic of habitats, where some may have acted as geographical barriers to gene flow between northern and southern populations, as well as between populations adapted to either desert-like or tundra-like habitats, respectively. North to south extension of the Pakistan deserts (Kara Kuram, Thar, Cholistan, Thal, and Kharan) and Iranian deserts (Kavir and Lut), which were formed and expanded during the late Pliocene and early Pleistocene (Krinsley 1970; Rajaguru et al. 2014) may have promoted a split between eastern and western populations. This pattern was nearly synchronously observed in magpies *Pica* (Song et al. 2018) and grouse (subfamily Tetraoninae; Drovetski 2003). Clade A later colonized North America, most likely across the Bering Land Bridge during the late Pleistocene. The monophyly of the North American taxa shows that at least the extant taxa originate from a single colonization event. It is possible that *E. a. atlas* (clade B2) and *E. bilopha* (clade C) diverged in different areas in different habitats in North Africa, reflecting their present-day horizontal and vertical parapatry in Morocco (Cramp 1988; Thévenot et al. 2003; Donald 2004), but it is also plausible that *E. a. atlas* colonized its ancestral habitat type in North Africa long after *E. bilopha* had already adapted to and diverged in the desert areas of northern Africa and/or western Asia. The disjunct distribution between *E. a. atlas* and southeast European/southwest Asian *E. alpestris* suggest that the former represents a relict population of a formerly more widespread distribution.

Although the ancestral habitat of the genus is presumably short grass or semi-desert in mountainous or plateau regions, this complex has adapted to an exceptionally wide range of low vegetation and barren habitats such as tundra, alpine grasslands, steppe, stony, and sandy deserts and farmlands from sea-level to c. 5400 m (Cramp 1988; Beason 1995; Alsop 2001; de Juana et al. 2004; Donald 2004). In North America, the colonization of lowland habitats, including desert and farmland, has apparently taken place, rather quickly. This is suggested by both the relatively recent appearance on this continent and by the fact that the earliest fossils from North America are from the Colorado plateau (c. 2900 m a.s.l.) (Barnosky 2004), whereas none of the early fossils are from lowlands (Tommy Tyrberg, in litt.).

Morphological divergence

Plumage divergence within this genus is substantial, which has led to a large number of subspecies being described, although whether plumage tone variation in larks is a useful taxonomic feature is open to question (Donald et al. 2017). This differentiation has apparently taken place fairly rapidly, presumably mainly driven by strong natural selection pressure in the varied habitats colonized by this species. As an example of how rapidly this species can adapt to new environments, Mason and Unitt (2018) demonstrated significant changes in upperpart coloration in a population of *E. alpestris* that took place over just 80 years in response to changes in soil colour following the cultivation of desert soils. However, the plumage variation is not congruent with the genetic differentiation, except for clade B1, in which all taxa show a unique head/breast pattern, which could be

considered a synapomorphy for this clade. The morphometric divergence is limited and our analysis shows that only two taxa (the small *E. bilopha* and the long-billed *E. a. longirostris*) stand out from the rest, at least on the basis of the three variables analyzed.

Taxonomic implications

Drovetski et al. (2014) suggested that the genus *Eremophila* comprises seven independent evolutionary lineages (*alpestris*, *flava*, *brandti*, *penicillata*, *atlas*, *bilopha*, and *elwesi*), and proposed that these should all be treated as separate species. However, they did not specify how these species should be circumscribed, i.e. to which species the unsampled taxa should belong. In our opinion, it is more parsimonious to recognize the four primary, deeply diverged lineages (A–D in Fig. 2), as separate species. Since not all subspecies have been studied genetically, a taxonomic revision at this stage needs to infer relationships based on other data to circumscribe the different species. The plumage data are available for this assessment, whereas the morphometrics proved to be inconclusive. However, it should be stressed that any taxonomic revision based on the available data should be considered tentative, as the mtDNA data published here and in earlier studies (Drovetski et al. 2014; Mason et al. 2014) would ideally be corroborated by nuclear DNA, and as remarked above, genomic data are probably required to resolve the relationships in this recent radiation.

We propose that clade A should be treated as *E. alpestris* (sensu stricto). We agree with Drovetski et al. (2014) that the New World taxa (clade A1) are likely to represent a unique evolutionary unit. However, because of its recent divergence from its sister lineage (clade A2), we prefer to treat these as conspecific. Based on the analyses of Drovetski et al. (2104) and Mason et al. (2014), confirmed here using mainly the same data, the North American taxa show shallow divergences and lack of concordance in geographical, phenotypic, and ecological characters. The New World taxa are in need of a taxonomic revision. Peters (1960) recognized 26 New World subspecies, whereas del Hoyo and Collar (2016) recognized only 13. The taxon *brandti* is paraphyletic in our analyses, although that is not strongly supported. We suggest that this clade should be given the English name Common Horned Lark.

We suggest that clade B should be treated as a single species, *E. penicillata* (by priority), including the subspecies *E. p. penicillata*, *E. p. balcanica*, *E. p. bicornis*, *E. p. albigula*, and *E. p. atlas*. del Hoyo and Collar (2016) also recognized the recently described (Roselaar 1995) taxon *kumerloevei*, which should also be part of this species, but which was disqualified by Kirwan (2006). Although we agree with Drovetski et al. (2014) that the taxon *atlas* (clade B2) represents a separately evolving lineage, supported by its divergent plumage from the other taxa in clade B (de Juana et al. 2004; Shirihai and Svensson 2018), the comparatively recent divergence from its sister lineage (clade B1) makes us favor treatment of these as conspecific. All of the above subspecies except *atlas* have the black face mask broadly connected to the black breast band; in *atlas*, many males have a thin connection between the black on the face and breast (van den Berg 2005). Being a mountain specialist, *E. penicillata* is ecologically segregated from the parapatrically distributed lowland *E. a. brandti*. However, very little is known about the contact zone (if any) between *E. penicillata* and *E. a. brandti* (see below). We propose Mountain Horned Lark as the common name for *E. penicillata*.

We support continued recognition of *E. bilopha* as a distinct species. It is deeply divergent as well as morphologically and ecologically distinct from the parapatrically distributed *E. p. atlas* (Cramp 1988; de Juana et al. 2004). Its juvenile plumage is uniquely different from all other juvenile *Eremophila* that have been studied (de Juana et al. 2004; Shirihai and Svensson 2018). Additionally, *atlas* and *bilopha* are altitudinally wholly segregated in summer, but *atlas* undergoes altitudinal migration that brings some birds into contact with *bilopha* and the two species have been seen together in southern Morocco.

We propose treating the taxa in clade D as a distinct species, *E. longirostris* (Himalayan Horned Lark). Although only two taxa representing this clade have been studied phylogenetically, we tentatively include the subspecies *longirostris*, *deosaiensis* (but see below), *elwesi*, *khamensis*, *przewalskii*, *argalea, teleschowi*, and *nigrifrons* in this species, noting that some of these may be too subtly different to warrant recognition. The oldest available name is *longirostris*, which unfortunately has not been analyzed phylogenetically. In spite of the morphometric distinctness of *longirostris* due to its long bill, we preliminarily include it in this clade, because of its close similarity in plumage and ecology and geographical connectivity to other subspecies from the Qinghai–Tibetan Plateau. Moreover, there is evidence that *longirostris* grades into *argalea*, which some authors consider a synonym of *elwesi* (Whistler 1932; Vaurie 1951). The taxon *nigrifrons* was treated as a synonym of *elwesi* by Bianchi (1904) and Cheng (1987) and there seems to be little evidence to treat *khamensis* as anything other than a synonym of *elwesi.* It is uncertain whether there is any contact between the northernmost taxa in this group, *deosaiensis*, *argalea* and *teleschowi*, and *E. a. brandti* or *E. p. albigula*, and, if there is, what the outcome is.

The taxon *E. alpestris deosai* Meinertzhagen, 1926 was first described from the Deosai Plateau, Pakistan, but was synonymized with *E. a. longirostris* by Vaurie (1951) and Peters (1960). Khan (1999) described a new subspecies, *E. a. deosaiensis*, from the Deosai Plateau. He stated that it has the black mask joining the black breast band, which

is a characteristic of the *penicillata* group and not of the *longirostris* group. However, the wing length of the typical specimen of *deosaiensis* (128 mm; Khan 1999) places it within the range of the *longirostris* group, but outside the *penicillata* group (max 126 mm; Paul Donald, unpublished). Moreover, in photographs of the type of *deosaiensis*, a distinct gap is obvious between the mask and breast band, again suggesting that it is part of the *longirostris* group. We preliminarily use the name *deosaiensis* Khan, 1999 in this paper despite that it is not listed in any checklist or handbook, while acknowledging that it might be better treated as a synonym of *longirostris*.

The form *montana* of the mountains of northern Central Asia (Tien Shan) was proposed by Bianchi (1904) on the basis of its longer, thinner bill than *brandti*. However, it was treated as a synonym of *brandti* by Kozlova (1933), Ludlow and Kinnear (1933), Vaurie (1954), and Peters (1960). The small number of specimens we have examined of this form appears to be appreciably larger than most *brandti*, which is primarily a lowland form; the mean wing length of typical *brandti* in the steppes is around 110 mm, whereas that of "*brandti*" in the Tien Shan (=*montana*) is 118 mm (Cramp 1988). We therefore suggest that *montana* requires further study and that two taxa, one of them perhaps falling in the *longirostris* group, may occur in this area and are separated by altitude. Clearly, much remains to be learned about the complex distribution of different taxa in this genus.

Conclusion

We propose that four separately evolving lineages within the genus *Eremophila* deserve to be recognized at species level: (1) Himalayan Horned Lark *E. longirostris* (comprising *E. l. longirostris*, *E. l. deosaiensis*, *E. l. elwesi*, *E. l. khamensis*, *E. l. przewalskii*, *E. l. argalea*, *E. l. teleschowi*, and *E. l. nigrifrons*) from the Himalayas and Qinghai–Tibetan plateau; (2) Temminck's Lark *E. bilopha* (monotypic), from North Africa to the Middle East; (3) Mountain Horned Lark *E. penicillata* (*E. p. penicillata*, *E. p. atlas*, *E. p. albigula*, *E. p. balcanica,* and *E. p. bicornis*) from northwest Africa and southeast Europe/southwest Asia; and (4) Common Horned Lark *E. alpestris* sensu stricto (*E. a. alpestris* and many other American subspecies, *E. a. flava*, *E. a. brandti*) from the Northern Palearctic and North and northern South America.

Acknowledgements This research was supported by the Faculty of Science, Ferdowsi University of Mashhad (FUM) (Grant No. 3/29628 to FG). Permission to collect samples in Iran was authorized by the Iranian Department of Environment (Permission Number: 93/52337; 2015; 25th January). We are indebted to Lars Svensson for morphometric data of *Eremophila bilopha*; to Takema Saitoh for providing photographs of the type of *Eremophila alpestris deosaiensis*; and to Tommy Tyrberg for information on fossils. P.A. acknowledges support from the Swedish Research Council (Grant No. 2015-04402) and Jornvall Foundation. We also acknowledge all the museums that helped us and provided DNA samples or allowed access to their specimen collections: Pierre-André Crochet and Alban Guillaumet, the Burke Museum, University of Washington, Seattle, USA; Swedish Museum of Natural History, Stockholm, Sweden and The Natural History Museum, Tring, UK. FG is also grateful to Dr. Zeinolabedin Mohammadi and Dr. Asghar Khajeh for their kind cooperation in sampling and to the staff of the University of Gothenburg, Department of Biology and Environmental Sciences for their kind cooperation. The study complied with the laws of the countries in which sampling was undertaken. We also thank five anonymous reviewers for their valuable comments.

References

Abdi H, Williams LJ (2010) Principal component analysis. Inc. WIREs Comp Stat 2:433–459

Aliabadian M, Kaboli M, Forschler MI, Nijman V, Chamani A, Tillier A, Zuccon D (2012) Convergent evolution of morphological and ecological traits in the open-habitat chat complex (Aves, Muscicapidae: Saxicolinae). Mol Phylogenet Evol 65:35–45

Alsop FJ III (2001) Birds of North America, Western region. Smithsonian Handbook. DK Publishing Inc., New York

Alström P, Barnes KN, Olsson U, Barker FK, Bloomer P, Khan AA, Qureshi MA, Guillaumet A, Crochet PA, Ryan PG (2013) Multilocus phylogeny of the avian family Alaudidae (larks) reveals complex morphological evolution, nonmonophyletic genera and hidden species diversity. Mol Phylogenet Evol 69:1043–1056

Avise JC (2000) Phylogeography: the history and formation of species. Harvard University Press, Cambridge

Barnosky AD (2004) Biodiversity response to climatic change in the middle Pleistocene: the Porcupine cave Fauna from Colorado. University of California Press, Berkeley

Beason RC (1995) Horned lark (*Eremophilia alpestris*). In: Poole A, Gill F (eds) The birds of North America No. 195. The academy of natural sciences and The American Ornithologists' Union, Philadelphia and Washington DC, pp 1–21

Bianchi V (1904) Key to the Palaearctic species of larks of the genus *Otocorys*. Ibis 46:370–372

Bobek H (1937) Die rolle der eiszeit in nordwestiran. Z Gletscher 25:130–183

Boev Z (2002) Neogene avifauna of Bulgaria. In: Zhou Z, Zhang F (eds) Proceedings of the 5th symposium of the society of avian palaeontology and evolution, Beijing, 01–04.06.2000. Science Press, Beiging, pp 29–40

Boev Z (2012) Neogene Larks (Aves: Alaudidae (Vigors, 1825)) from Bulgaria. Acta Zool Bulgar 64:295–318

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:e1003537

Brito PH (2005) The influence of Pleistocene glacial refugia on tawny owl genetic diversity and phylogeography in Western Europe. Mol Ecol 14:3077–3094

Cheng TH (1987) A synopsis of the Avifauna of China. Beijing, pp 1–1223

Clements JF, Schulenberg TS, Iliff MJ, Roberson D, Fredericks TA, Sullivan BL, Wood CL (2018) The eBird/Clements checklist of birds of the world: v2018. http://www.birds.cornell.edu/clementschecklist/download/

Cramp S (1988) The Birds of the Western Palearctic, vol V. Oxford University Press, Oxford
Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9:772
de Juana E, Suárez F, Ryan P, Alström P, Donald P (2004) Family Alaudidae (Larks). In: del Hoyo J, Elliott A, Christie DA (eds) Handbook of the Birds of the World, vol 9. Lynx Edicions, Barcelona, pp 496–601
del Hoyo J, Collar NJ (2016) The HBW–BirdLife International illustrated checklist of the birds of the world, 2: passerines. Lynx Edicions, Barcelona
Dickinson E (ed) (2003) The Howard and Moore complete checklist of the birds of the world, 3rd edn. Helm, London
Dickinson EC, Christidis L (2014) The Howard and Moore complete checklist of the birds of the world. Aves Press, Eastbourne
Donald PF (2004) Horned Lark *Eremophila alpestris*. In: del Hoyo J, Elliott A, Christie D (eds) Handbook of the birds of the world, Lynx Edicions, vol 9. Barcelona and BirdLife International, Cambridge, p 863
Donald PF, Alström P, Engelbrecht D (2017) Possible mechanisms of substrate colour-matching in larks (Alaudidae) and their taxonomic implications. Ibis 159:699–702
Drovetski SV (2003) Plio–Pleistocene climatic oscilations, Holarctic biogeography and speciation in an avian subfamily. J Biogeogr 30:1173–1181
Drovetski SV, Semenov G, Drovetskaya SS, Fadeev IV, Red'kin YA, Voelker G (2013) Geographic mode of speciation in a mountain specialist avian family endemic to the Palearctic. Ecol Evol 3:1518–1528
Drovetski SV, Rakovic M, Semenov G, Fadeev IV, Red'kin YA (2014) Limited Phylogeographic signal in sex-linked and autosomal loci despite geographically, ecologically, and phenotypically concordant structure of mtDNA variation in the Holarctic avian genus *Eremophila*. PLoS One 9:1–10
Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. PLoS Biol 4:e88
Fjeldså J, Bowie RCK, Rahbek C (2012) The role of mountain ranges in the diversifcation of birds. Annu Rev Ecol Evol Syst 43:249–265
Frenzel B (1973) Climatic fluctuations of the ice age. The Press of Case Western Reserve University, Cleveland and London
Fu YX (1997) Statistical tests of neutrality of mutations against population growth: hitchhiking and background selection. Genetics 147:915–925
Gill F, Donsker D (2018) IOC world bird list (v 8.1). https://doi.org/10.14344/ioc.ml.8.1
Harris DJ, Taylor SD, White EP (2018) Forecasting biodiversity in breeding birds using best practices. PeerJ 6:e4278
Hellmayr CE (1929) Birds of the James Simpson–Roosevelts Asiatic expedition. Field Mus Nat Hist Zool 17:27–144
Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. Nature 405:907–913
Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. Phil Trans R Soc Lond B 359:183–195
Hughes P, Woodward J (2017) Quaternary glaciation in the mediterranean mountains, 1st edn. Geological Society of London, Bath **(Special Publications)**
Khan AA (1999) A New Horned Lark *Eremophila alpestris deosaiensis*, (subsp. Nov.) from Deosai, Baltistan Pakistan. Pak J Ornith 3:27–33
Kirwan GM (2006) Comments on two subspecies of passerine birds recently described from Turkey, *Eremophila alpestris kumerloevi* and *Pyrrhula pyrrhula paphlagoniae*, with remarks on geographical variation in related forms of bullfinch from the Balkans and Caucasus. Sandgrouse 28:10–21
Klicka J, Zink RM (1999) Pleistocene effects on North American songbird evolution. Proc R Soc Lond B 266:695–700
Kozlova EV (1933) The birds of south-west Transbaikalia, northern Mongolia and central Gobi. Ibis 13(3):59–87
Krinsley DH (1970) A geomorphological and paleoclimatological study of the playas of Iran. USGS Final Scientific Report, Contract, PROCP, pp 700–800
Lanfear R, Calcott B, Ho SY, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695–1701
Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017) PartitionFinder 2: new methods for selecting partitioned models of evolution formolecular and morphological phylogenetic analyses. Mol Biol Evol 34:772–773
Lei FM, Qu YH, Song G (2014) Species diversification and phylogeographical patterns of birds in response to the uplift of the Qinghai–Tibet plateau and quaternary glaciations. Curr Zool 60:149–161
Lei F, Qu Y, Song G, Alström P, Fjeldså J (2015) The potential drivers in forming avian biodiversity hotspots in the East Himalaya Mountains of Southwest China. Integr Zool 10:171–181
Lerner HRL, Meyer M, James HF, Hofreiter M, Fleischer RC (2011) Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Curr Biol 21:1838–1840
Li X, Dong F, Lei F, Alström P, Zhang R, Ödeen A, Fjeldså J, Ericson PGP, Zou F, Yang X (2016) Shaped by uneven Pleistocene climate: Mitochondrial phylogeographic pattern and population history of white wagtail *Motacilla alba* (Aves: Passeriformes). J Avian Biol 47:263–274
Liu B, Alström P, Olsson U, Fjeldså J, Quan Q, Roselaar KCS, Saitoh T, Yao C, Hao Y, Wang W, Qu Y, Lei F (2017) Explosive radiation and spatial expansion across the cold environments of the Old World in an avian family. Ecol Evol 7:6346–6357
Ludlow F, Kinnear NB (1933) A contribution to the Ornithology of Chinese Turkestan. Ibis 13(3):658–694
Mason NA, Unitt P (2018) Rapid phenotypic change in a native bird population following conversion of the Colorado Desert to agriculture. J Avian Biol. https://doi.org/10.1111/jav.01507
Mason NA, Title PO, Cicero C, Burns KJ, Bowie RCK (2014) Genetic variation among western populations of the Horned Lark (*Eremophila alpestris*) indicates recent colonization of the Channel Islands off southern California, mainland-bound dispersal, and postglacial range shifts. Auk 131:162–174
Olsson U, Alström P, Ericson PGP, Sundberg P (2005) Non-monophyletic taxa and cryptic species-evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*, Aves). Mol Phylogenet Evol 36:261–276
Pavlova A, Zink RM, Drovetski SV, Redkin Y, Rohwer S (2003) Phylogeographic patterns in *Motacilla flava* and *Motacilla citreola*: species limits and population history. Auk 120:744–758
Pereira SL, Baker AJ (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Mol Biol Evol 23:1731–1740
Peters JL (1960) Family Alaudidae. In: Mayr E, Greenway JC Jr (eds) Check-list of birds of the world, vol IX. Museum of comparative zoology. Cambridge, Mass, pp 3–80
Qu Y, Lei F, Zhang R (2010) Comparative phylogeography of five avian species: implications for Pleistocene evolutionary history in the Qinghai–Tibetan plateau. Mol Ecol 19:338–351
R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/
Rajaguru SN, Deo SG, Gaillard C (2014) Pleistocene geoarchaeology of Thar Desert. Ann Arid Zone 53:63–76

Rambaut A (2016) FigTree. Version 1.4.3. http://tree.bio.ed.ac.uk/software/figtree
Rambaut A, Drummond AJ (2013) Tracer. v1.6.0. http://tree.bio.ed.ac.uk/
Rambaut A, Drummond AJ (2015) TreeAnnotator. Version 2.2.1. http://beast2.org
Rambaut A, Suchard M, Xie D, Drummond A (2014) Tracer v1.6. http://beast.bio.ed.ac.uk/Tracer
Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. Mol Biol Evol 19:2092–2100
Roselaar CS (1995) Songbirds of Turkey, an atlas of boidiversity of Turkish passerine birds. Haarlem
Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sánchez-Gracia A (2017) DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol 34:3299–3302
Saitoh T, Alström P, Nishiumi I, Shigeta Y, Williams D, Olsson U, Ueda K (2010) Old divergences in a boreal bird supports long-term survival through the ice ages. BMC Evol Biol 10:35. https://doi.org/10.1186/1471-2148-10-35
Sandel B, Arge L, Dalsgaard B, Davies RG, Gaston KJ, Sutherland WJ, Svenning JC (2011) The influence of Late Quaternary climate-change velocity on species endemism. Science 334:660–664
Schenekar T, Weiss S (2011) High rate of calculation errors in mismatch distribution analysis results in numerous false inferences of biological importance. Heredity 107:511–512
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464
Sharpe RB (1890) Catalogue of the passeriformes of perching birds in the collection of the British Museum, Sturnifonnes. Trustees of the Brit Mus (Nat Hist), London
Sher AV (1997) Nature restructuring in the East-Siberian Arctic at the Pleistocenee Holocene boundary and its role in mammal extinction and emerging of modern ecosystems. Earth Cryosphere 1:21–29
Sher AV, Kuzmina SA, Kuznetsova TV, Sulerzhitsky LD (2005) New insights into the Weichselian environment and climate of the east Siberian Arctic, derived from fossil insects, plants, and mammals. Quat Sci Rev 24:533–569
Shirihai H, Svensson L (2018) Handbook of Western Palearctic birds. Hel, London
Shroder JF, Bishop MP (2010) Glaciers of Pakistan. In: Williams RS, Ferrigno JG (eds) Satellite image atlas of glaciers. US Geological Survey Professional Paper, 1386-F-4
Smith BT, Klicka J (2013) Examining the role of effective population size on mitochondrial and multilocus divergence time discordance in a songbird. PLoS One 8:e55161
Song G, Zhang R, Alström P, Irestedt M, Cai T, Qu Y, Ericson PGP, Fjeldså J, Lei F (2018) Complete taxon sampling of the avian genus *Pica* (magpies) reveals ancient relictual populations and synchronous Late-Pleistocene demographic expansion across the Northern Hemisphere. J Avian Biol 49:e01612
Sorenson MD, Ast JC, Dimchev DE, Yuri T, Mindell DP (1999) Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol Phylogenet Evol 12:105–114
Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595
Thévenot M, Vernon R, Bergier P (2003) The birds of Morocco (BOU Checklist Series No. 20). Tring: British Ornithologists' Union
Tieleman BI, Williams JB, Bloomer P (2003) Adaptation of metabolism and evaporative water loss along an aridity gradient. Proc R Soc Lond B 270:207–214
Tzedakis PC, Lawson IT, Frogley MR, Hewitt GM, Preece RC (2002) Buffered tree population changes in a Quaternary refugium: evolutionary implications. Science 297:2044–2047
van den Berg AB (2005) Morphology of Atlas Horned Lark. Dutch Birding 27:256–258
Vaurie C (1951) A study of Asiatic larks. Bull Am Mus Nat Hist 97:435–526
Vaurie C (1954) Systematic notes on Palearctic birds. No. 7. Alaudidae and Motacillidae (Genus *Anthus*). Am Mus Novit 1672:1–13
Venables WN, Ripley BD (2002) Modern applied statistics with S, 4th edn. Springer, New York
Voelker G (1999) Dispersal, vicariance, and clocks: historical biogeography and speciation in a cosmopolitan passerine genus (*Anthus*: Motacillidae). Evol 53:1536–1552
Weir JT, Schluter D (2004) Ice sheets promote speciation in boreal birds. Proc R Soc Lond Biol Sci 271:1881–1887
Weir JT, Schluter D (2008) Calibrating the avian molecular clock. Mol Ecol 17:2321–2328
Whistler H (1932) On some larks of the Kashmir State. Ibis 13:470–479
Yurtsev BA (1981) Relict Steppe complexes in Northeast Asia (problems of reconstruction of cryoxerotic landscapes of Beringia). Nauka, Novosibirsk **(in Russian)**
Zelenkov NV, Kurochkin EN (2012) The first representative Pliocene assemblages of passerine birds in Asia (northern Mongolia and Russian Transbaikalia). Geobios 45:323–334
Zhao N, Dai CY, Wang WJ, Zhang RY, Qu YH, Song G, Chen K, Yang XJ, Zou FS, Lei FM (2012) Pleistocene climate changes shaped the divergence and demography of Asian populations of the great tit *Parus major*: evidence from phylogeographic analysis and ecological niche models. J Avian Biol 43:297–310
Zheng H, Powell CM, An Z et al (2000) Pliocene uplift of the northern Tibetan Plateau. Geology 28:715–718
Zheng H, Chen H, Cao J (2002) Palaeoenvironmental implication of the Plio–Pleistocene loess deposits in southern Tarim Basin. Chin Sci Bull 47:700–704
Zhu B, Yu J, Qin X, Patrick R, Liu Z, Xiong H (2014) Formation and evolution of sand deserts in Xinjiang, Northwest China: I. Provenances of desert sands. J Geogr Sci 24:177
Zink RM, Drovetski SV, Rohwer S (2002) Phylogeographic patterns in the great spotted woodpecker *Dendrocopos major* across Eurasia. J Avian Biol 33:175–178