Phylogenetic relationships of endemic bunting species (Aves, Passeriformes, Emberizidae, *Emberiza koslowi*) from the eastern Qinghai-Tibet Plateau

MARTIN PÄCKERT¹, YUE-HUA SUN², PATRICK STRUTZENBERGER¹, OLGA VALCHUK³, DIETER THOMAS TIETZE⁴ & JOCHEN MARTENS⁵, #

¹ Senckenberg Natural History Collections, Museum of Zoology, Königsbrücker Landstraße 159, 01109 Dresden, Germany — ² Key Laboratory of Animal Ecology and Conservation, Institute of Zoology, Chinese Academy of Science, 100101 Beijing, China — ³ Institute of Biology and Soil Sciences, Russian Academy of Sciences, Vladivostok, 690022 Russia — ⁴ Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Im Neuenheimer Feld 364, 69120 Heidelberg, Germany — ⁵ Institut für Zoologie, Johannes Gutenberg-Universität, 55099 Mainz, Germany — # Results of the Himalaya-Expeditions of J. Martens, No. 281. – For No. 280 see: Genus (Wrocław) 25 (3): 351–355, 2014

Accepted 20.iii.2015. Published online at www.senckenberg.de/vertebrate-zoology on 4.v.2015.

Abstract

In this study we reconstructed the phylogenetic relationships of a narrow-range Tibetan endemic, *Emberiza koslowi*, to its congeners and shed some light on intraspecific lineage separation of further bunting species from Far East Asia and along the eastern margin of the Tibetan Plateau in China. The onset of the Old World bunting radiation was dated to the mid Miocene and gave rise to four major clades: i) one group comprising mainly Western Palearctic species and all high-alpine endemics of the Tibetan Plateau; ii) a clade including *E. lathami*, *E. bruniceps* and *E. melanocephala*; iii) one group comprising mainly Eastern Palearctic species and all insular endemics from Japan and Sakhalin; iv) an exclusively Afrotropic clade that comprised all African species except *E. affinis*, whose phylogenetic relationships were ambiguous and only poorly supported in all reconstructions. The Tibetan bunting, *E. koslowi*, turned out as an early offshoot of the Western Palearctic-Tibetan clade 1 and thus represents an ancient relic lineage that dates back to a mid Miocene colonization event of its ancestors to the alpine plateau habitats. This temporal scenario of an early Miocene origin of alpine Tibetan endemics coincides with recent results for two further species, the Tibetan ground tit, *Pseudopodoces humilis*, and the Tibetan rosefinch, *Carpodacus roborowskii*. The origin of extant intraspecific phylogeographic patterns and splits among sister species in Eastern Asia were dated back to the Pleistocene with earliest lineage splits occurring among taxa from the Japanese Archipelago including Sakhalin and their mainland counterparts. A similarly ancient split separated a southern clade of *E. godlewskii yunnanensis* from S Sichuan and Yunnan from a northern clade including populations from central and northeastern China, Mongolia and S Siberia. Ecological segregation among breeding habitats of southern *E. g. yunnanensis* at lower elevations and those of other conspecifics at high-alpine habitats might have played a key role in the spatial genetic diversification of this species.

Key words

*Emberiza*, phylogeny, Tibetan bunting, endemics, Qinghai-Tibet Plateau, lineage splits, intraspecific differentiation.

Introduction

The Old World buntings of the genus *Emberiza* are widely distributed across the Palearctic, the Middle East, the Himalayas, East Asia and Africa. Throughout the genus’ entire range 42 currently accepted *Emberiza* species (Rose, 2011; 38 species in Byers et al., 1995) occupy a diverse variety of breeding habitats, however most species tend to avoid very densely forested areas and prefer semi-open to open taiga forest habitats, Central Asian
steppes and wetlands (Byers et al., 1995; Rose, 2011). A geographically extensive hotspot of species richness is located in the Eastern Palearctic (Fig. 1). There, many bunting species occupy breeding habitats of open and semi-open taiga forest and forest edges (all information on habitat preferences according to Dementiev & Gladkov, 1954; Panov, 1973; Rose, 2011). Among these E. tristrami is typically bound to closed mixed and pine forests with dense bush undergrowth while E. elegans has a preference of mixed and deciduous forests with a presence of oak stands and E. leucocephalos even settles forest islands in steppe habitats. Other species have a clear preference for wetland forests, marshes and riverbanks with sparse birch and willow stands (E. pusilla, E. aureola, E. rutila). Pallas’ bunting (E. pallasi) even breeds at high latitudes of the subarctic tundra belt in shrub vegetation of dwarf willow and alder stands along river valleys and in subalpine tundra with Rhododendron bushes and grasslands of the Siberian Altai. Twelve to sixteen Emberiza species (wintering species included) have been recorded in local co-occurrence at different local field stations in Far East Russia (Mattes & Alfer, 2010; Mattes & Shokhin, 2010; Heim et al., 2012). At Muraviovo Park in the Middle Amur region the black-faced bunting (E. spodocephala) and Pallas’ bunting (E. pallasi) ranged among the five locally most abundant species with respect to the number of ringed individuals (Heim & Smirenski, 2013). In southern Primorye, at Lazo Reserve highest numbers in the period of migration were observed for E. rustica, E. elegans, E. spodocephala and E. tristrami (Mattes & Alfer, 2010), at Litovka river for the same species and for E. rutila (Valchuk & Yuasa, 2002).

Besides Central and Far Eastern Siberia, further regions of high species richness in Asia are found at the northwestern margin of the Qinghai-Tibet Plateau and in China at the southeastern plateau margin (Fig. 1). There, some species occupy bushy mountain slopes and montane steppe, such as Godlewski’s bunting (E. godlewskii; Fig. 2) that has a patchy breeding distribution at the eastern, the northern and the western plateau margins. Some endemic species are confined to very small breeding ranges at the northern plateau margins, such as the endangered rufous-backed bunting (E. jankowskii) from the grass steppes of Inner Mongolia and W Jilin (Rose, 2011) and at the eastern margins such as the Tibetan bunting (E. koslowi; Fig. 2). The latter is restricted to a narrow breeding distribution around the border between the Chinese provinces Xizang and Qinghai where it occupies alpine shrubs of juniper, rhododendron and cotoneaster on steep slopes and alpine grasslands between 3600 and 4600 m a.s.l. (Schafer, 1938; Olsson, 1995; Thewlis & Martins, 2000; Rose, 2011; Ju & Golok, 2013). The phylogenetic relationships of this Tibetan endemic to other congeners are so far unresolved.

A first near-complete molecular phylogeny of Old World buntings, based on one mitochondrial and one
nuclear marker, was published by Alstrom et al., 2008. Though some among-and within-clade relationships were not resolved in the two-marker phylogeny the authors found evidence of four major Emberiza clades and confirmed six sister species pairs previously identified by traditional morphology-based systematics. Furthermore, they showed that three members of three traditional monotypic emberizid genera, Miliaria, Latoucheornis and Melophus (cf. Byers et al., 1995) were firmly nested within the Emberiza tree (Alstrom et al., 2008). These results confirmed earlier taxonomic recommendations by Sangster et al., (2004) for a transfer of the corn bunting (Miliaria calandra sensu Vos, 1977) into Emberiza based on early molecular studies with rather incomplete taxonomic samplings (Graputto et al., 2001; Lee et al., 2001). In contrast, contrary to molecular evidence from the near-complete bunting phylogeny (Alstrom et al., 2008) the other two monotypic Asian genera (Latoucheornis and Melophus) have been maintained and separated from Emberiza by some authors still to date (Clements et al., 2014). In a recent phylogenetic study on some polytypic species of African brown buntings Olsson et al., (2013) found genetic support for the species status of North African E. sahari as a sister of East African and Middle Eastern E. striolata (these two were already separated by Rose, 2011) and they recommended a further species-level split of E. goslingi from E. tahapisi. Genetic distinctiveness of the latter two species furthermore coincides with differences of territorial songs (Osi- euk, 2011). Traditionally, the insular endemic of Socotra (E. socotrina) was thought to be affiliated with this group (Rose, 2011). This relationship was recently confirmed by Schweizer & Kirwan (2014) who compared sequence data obtained from museum specimens of the Socotra bunting with the African brown bunting data set by Olsson et al. (2013).

Apart from the latter study only little attention has been drawn to intraspecific genetic variation of buntings. Very few phylogeographic studies were dedicated to subspecific variation in the reed bunting, E. schoeniclus (Kvist et al., 2011; Zink et al., 2008). Further evidence of intraspecific lineage divergence among continental and insular populations of E. spodocephala was inferred from RAPD analysis (Dolgova & Valchuk, 2008), from DNA-barcoding (Saitoh et al., 2015) and from multi-locus phylogeographies and accompanying morphometric analyses (Weissensteiner, 2013).

In our study we focused on the phylogenetic relationships of the Tibetan bunting (E. koslowi), a so far virtually unstudied high alpine endemic of the Tibetan plateau, as well as on intraspecific lineage separation in Asian bunting species at the southeastern Chinese plateau margins (such as E. godlewskii; Fig. 2) and in the Himalayas. We added sequence data from taxa not included in previous studies to the Emberiza tree and expanded the dataset to include three mitochondrial genes and two nuclear introns. In addition, we used a time-calibrated multi-locus phylogeny in order to provide an evolutionary time scale of Emberiza bunting evolutionary history.

Material and Methods

DNA analysis

We extracted DNA from 103 blood and tissue samples of 34 Emberiza species plus a few focal subspecies from China and the Himalayas (and two outgroup species, Lapland longspur (Calcarius lapponicus) and snow bunting (Plectrophenax nivalis). For origin of samples and sequences see electronic supplement 1.

We amplified a 706-bp fragment of the mitochondrial cytochrome-b gene for all samples available using the primer combination L14841-Cytb (5’ – AAA AAG CTT CCA TCC AAC ATC TCA GCA TGA TGA AA – 3’, Kocher et al., 1989) and H-15547-Cytb (5’ – AAT AGG
AAG TAT CAT TCG GGT TTG ATG – 3’, Edwards et al., 1991). The PCR protocol was 94°C for 2 min followed by 35 cycles of 92°C for 45 s, 56°C for 1 min and 72°C for 1.5 min with a final extension in 72°C for 5 min. A larger 1079-bp-long fragment was amplified for at least one sample of each taxon investigated with the primer combination O-L14851 (5’ – CCT ACC TAG GAT CAT TCG CCC T – 3’) and O-H16065 (5’ – AGT CTT CAA TCT TTG GCT TAC AAG AC – 3’; Weir & Schluter, 2007). The PCR protocol was 94°C for 10 min followed by 35 cycles of 92°C for 1 min, 53°C for 1 min and 72°C for 2 min with a final extension in 72°C for 10 min.

For a few toe pad samples obtained from whole skins of target taxa we performed DNA extractions using the sbeadex® forensic kit (LGC Genomics). Extraction was performed according to the manufacturer’s instructions except for overnight incubation of tissue with proteinase K (instead of one hour) and only 60 µl elution volume (instead of 100 µl) in order to yield a sufficiently high concentration of DNA extracts. All toe pad samples were analysed in a separate clean lab. There, each step of analysis (sampling, extraction and PCR) was done on separate working benches. In order to avoid cross-contamination working benches were cleaned with DNA-away (Molecular Bio Products, Inc.), after each step both benches and lab rooms were decontaminated with UV-light for at least four hours. Using a set of sequences derived from DNA analysis of fresh samples we designed several primer combinations for amplification of shorter gene fragments from toe pad DNA extracts (110 to 250 bp, depending on the quality and age of template DNA). Primer design was carried out using software OLIGOEXPLORE 1.2 (http://www.genelink.com/tools/gl-oe.asp) and gradient PCRs were performed in order to determine the optimum annealing temperature for each primer pair. PCR primer combinations for amplification of short cytb fragments are provided in electronic supplement 2. For comparison with our own data set we obtained 99 further cytb sequences from GenBank (see electronic supplement 1).

For reconstruction of a multi-locus phylogeny we amplified and sequenced four additional markers for a reduced taxon set with one representative sample per species (or mitochondrial lineage identified within species). We added sequences of two further mitochondrial markers to the data set, 16S rRNA and the barcoding marker cytochrome oxidase (COI) as well as two nuclear introns, myoglobin intron 2 and fibrinogen intron 7. For amplification of short cytb fragments were provided in electronic supplement 2. For comparison with our own data set we obtained 99 further cytb sequences from GenBank (see electronic supplement 1).

We estimated the best-fitting substitution models for each of the five molecular markers using MrModeltest v2 (Nylander, 2004). For both mitochondrial genes (COI and cytb) we estimated separate best-fit models for each codon position separately. Separate models were estimated for the full cytb data set (195 sequences) and for the reduced data set used for multi-locus reconstruction (66 sequences). According to the Akaike Information Criterion (AIC), the best fitting model was GTR+I+Γ for all mitochondrial genes, regardless of which cytb data set was being used, and HKY+Γ was the best fitting model for the two nuclear introns (for model settings refer to Table 1).

We reconstructed a phylogeny based on our full cytb data set using Maximum Likelihood (ML) with RAxML v. 7.2.6 (Stamatakis, 2006; using the GUI python application v. 0.93 by Silvestro & Michalak, 2010) and Bayesian Inference with MrBayes v3.1.2 (Ronquist & Huelsenbeck, 2003). We performed two independent runs: In the first run we treated the entire alignment as a single partition, in the second run we analyzed each codon position as a separate partition. In the latter run we allowed the overall rate to vary between partitions by setting the priors <ratep = variable> and model parameters such as gamma shape, proportion of invariable sites and the rate matrix were unlinked across partitions. Bayesian

Data analysis

with annealing temperatures decreasing from 58°C (5 cycles) and 56°C (5 cycles) to 54°C (30 cycles). For primer combinations, PCR and sequencing protocols for 16S rRNA refer to Spicer & Dupinace (2004), for fib7 refer to Przychtko & Moore (1997) and for protocols for the COI barcoding marker refer to Litmaer et al. (2012).

For hierarchical outgroup rooting we used sequences of representative species of New World Emberizini (Passerellidae sensu Dickinson & Christidis, 2014; Ammodramus humeralis, Junco hyemalis, Passerulus sandwichensis, Zonotrichia albicollis and Z. leucophyrs), five species of longspurs and snow buntings (Calcarius and Plectrophenax; Plectrophenacidae sensu Dickinson & Christidis, 2014) plus a few distantly related taxa (Dendroica virens, Cardinalis cardinalis, Sturnella supercilialis, Fringilla coelebs) all of which had also been used for outgroup rooting by Alström et al. (2008). The final cytb alignment of 657 bp included 195 sequences. Sequences were aligned with MEGA v5 (Tamura et al., 2011) and manually corrected after visual inspection. For comparison with our own COI barcode sequences of target species we used additional sequence data from GenBank (mainly from Kerr et al., 2009 and Saitoh et al., 2015) in order to control for intraspecific mtDNA lineage divergence found in our cytb data set. The multi-locus alignment for all five markers was 3933 bp long (cytb: 1035 bp; 16S rRNA: 838 bp; COI: 693 bp; fib7: 687 bp; myo: 680 bp; Table 1).
analysis was performed using the Metropolis-coupled Markov Chain Monte Carlo algorithm with two parallel runs, each with one cold and three heated chains. The heating parameter λ was set to 0.1. The chains ran for 10,000,000 generations, trees were sampled every 100th generation. The first 3,000 samples were discarded as burn-in and the model parameters and the posterior probabilities were estimated from the remaining samples. The remaining trees were summarized in a 50% majority rule consensus tree.

One sequence of *E. lathami* (from Alström et al., 2008) did not clade with three further sequences as burnin and the model parameters and the posterior probabilities were estimated from the remaining samples. The remaining trees were summarized in a 50% majority rule consensus tree.

For illustration of intraspecific mitochondrial lineage separation we reconstructed haplotype networks for selected taxa (taxon pairs) with TCS v1.21 (Clement, 2007). We assigned multi-copy sequences with TCS v1.21 to the correct taxa and show overwhelming support for the nine-partition scheme over all other schemes tested (Table 2).

In the absence of reliable passerine fossils providing appropriate node age constraints, molecular dating was performed using an empirical substitution rate of 0.0105 substitutions per site per lineage per million years for cytb as evaluated by Weir & Schluter (2008). We applied this rate to the cytb partition and left the rates of all other loci to be estimated relative to the cyt-b partition. The log files were examined with Tracer v1.4.8 (Drummond & Rambaut, 2007) in order to ensure effective sample sizes (ESS; which yielded reasonable values for all parameters after 30,000,000 generations). Trees were summarized with TreeAnnotator v1.4.8 (posterior probability limit = 0.5) using a burn-in value of 9,000 (trees) and the median height annotated to each node.

Node support in a ML framework was obtained by 1,000 bootstrap replicates with RAxML (thorough bootstrap option). In two separate runs, we partitioned the concatenated matrix (5 partitions by gene; 9 partitions by gene and codon see above) and applied the GTR+I+Γ model across partitions. We assigned multi-copy matrices to the correct taxa and show overwhelming support for the nine-partition scheme over all other schemes tested (Table 2).

In order to determine the best-fitting partitioning regime for our data we compared four partitioning schemes using AIC (McGuire et al., 2007) and AICM (Baele et al., 2012). ML likelihood values were obtained with RAxML v8.1.7 using 100 replicates of the new rapid hill-climbing algorithm under the GTR+Γ model. AIC values based on the likelihood of the best tree for each partitioning strategy were calculated with Microsoft Excel. AICM values were obtained with Tracer v1.6 based on BEAST v1.8.1 runs with 11 million generations and trees being sampled at every 1000th state, the first 1000 trees were discarded as burn-in. Substitution models applied to each partition were identical to the ones used in final analyses. A GTR+I+Γ model was applied to each partition of the one and two-partition schemes. The results of AIC and AICM both give the same ranking of the four strategies and show overwhelming support for the nine-partition scheme over all other schemes tested (Table 2).

<table>
<thead>
<tr>
<th>Table 1. Substitution models applied to the different partitions of single-locus and multi-locus alignments.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>bp</strong></td>
</tr>
<tr>
<td><strong>single loc</strong></td>
</tr>
<tr>
<td>657</td>
</tr>
<tr>
<td>GTR+I+Γ</td>
</tr>
<tr>
<td>nA</td>
</tr>
<tr>
<td>nC</td>
</tr>
<tr>
<td>nT</td>
</tr>
<tr>
<td>nG</td>
</tr>
<tr>
<td>α</td>
</tr>
<tr>
<td>l</td>
</tr>
<tr>
<td>T/Tv ratio</td>
</tr>
<tr>
<td>R(a)[A-C]</td>
</tr>
<tr>
<td>R(c)[A-T]</td>
</tr>
<tr>
<td>R(d)[G-C]</td>
</tr>
</tbody>
</table>

For illustration of intraspecific mitochondrial lineage separation we reconstructed haplotype networks for selected taxa (taxon pairs) with TCS v1.21 (Clement, 2007).

We reconstructed a multi-locus phylogeny based on all five markers with BEAST v1.8.1 (Drummond & Rambaut, 2007). We ran BEAST for 30,000,000 generations (trees sampled every 1000 generations) under the uncorrelated lognormal clock model for all loci with the “auto-optimize” option activated and a birth-death process prior (with incomplete sampling assumed) applied to the tree. As for the single-locus analysis, we performed two different runs: i) 5 partitions: by gene only and 5 best-fit substitution models applied to each partition; ii) 9 partitions: by gene and additionally by codon position for the two coding mtDNA markers with the best-fitting-model settings applied to each codon position (for model settings compare Table 1).

In order to determine the best-fitting partitioning regime for our data we compared four partitioning schemes using AIC (McGuire et al., 2007) and AICM (Baele et al., 2012). ML likelihood values were obtained with RAxML v8.1.7 using 100 replicates of the new rapid hill-climbing algorithm under the GTR+Γ model. AIC values based on the likelihood of the best tree for each partitioning strategy were calculated with Microsoft Excel. AICM values were obtained with Tracer v1.6 based on BEAST v1.8.1 runs with 11 million generations and trees being sampled at every 1000th state, the first 1000 trees were discarded as burn-in. Substitution models applied to each partition were identical to the ones used in final analyses. A GTR+I+Γ model was applied to each partition of the one and two-partition schemes. The results of AIC and AICM both give the same ranking of the four strategies and show overwhelming support for the nine-partition scheme over all other schemes tested (Table 2).
Table 2. AIC and AICM values for each partitioning strategy including partitioning of the two coding mitochondrial genes cytochrome-\textit{b} (cyt-\textit{b}) and ND\textit{b} by Codon position. \(K\): number of parameters; \(w\): Aikake weight.

<table>
<thead>
<tr>
<th>Description</th>
<th>lnL(best tree)</th>
<th>(K)</th>
<th>AIC</th>
<th>DAIC</th>
<th>(w)</th>
<th>AICM</th>
<th>DAICM</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 partitions, each locus, each codon position of cyt-\textit{b} and ND\textit{b}</td>
<td>(-28334.36)</td>
<td>1242</td>
<td>59152.72</td>
<td>0</td>
<td>1</td>
<td>57353.90</td>
<td>0</td>
</tr>
<tr>
<td>2 partitions, nuclear loci and mitochondrial genes</td>
<td>(-29706.08)</td>
<td>276</td>
<td>59964.10</td>
<td>811.44</td>
<td>0</td>
<td>59410.32</td>
<td>1874.42</td>
</tr>
<tr>
<td>1 partition, unpartitioned</td>
<td>(-30122.77)</td>
<td>138</td>
<td>60541.53</td>
<td>1388.82</td>
<td>0</td>
<td>59875.85</td>
<td>2340.05</td>
</tr>
<tr>
<td>5 partitions; one partition for each locus</td>
<td>(-29616.12)</td>
<td>690</td>
<td>60612.25</td>
<td>1459.53</td>
<td>0</td>
<td>59917.29</td>
<td>2381.38</td>
</tr>
</tbody>
</table>

Results

Topology of the \textit{Emberiza} tree

The multi-locus phylogeny of \textit{Emberiza} buntings is shown in Fig. 3. Tree topologies inferred from the two independent runs with different partition schemes were widely congruent and major clades were equally well supported with one apparent exception (Fig. 3): The widely congruent and major clades were equally well independent runs with different partition schemes were shown in Fig. 3. Tree topologies inferred from the two \textit{Emberiza} buntings is tree \textit{Emberiza}.

Results

Topology of the \textit{Emberiza} tree

The multi-locus phylogeny of \textit{Emberiza} buntings is shown in Fig. 3. Tree topologies inferred from the two independent runs with different partition schemes were widely congruent and major clades were equally well supported with one apparent exception (Fig. 3): The 5-partition scheme (by gene) reflected a monophyletic African clade (with the exception of \textit{E. affinis}) while the 9-partition scheme (by gene and codon) yielded two non-monophyletic African clades plus \textit{E. affinis} as a separate lineage with unresolved relationships of all three clades (for details see below).

A monophyletic clade of Old World buntings (\textit{Emberiza}) received full Bayesian support in all reconstructions and was sister to an equally well supported clade of New World sparrows (with good support for this sister group relationship; Fig. 3). The Old World buntings were split into four major clades that according to our molecular dating back to a mid Miocene separation (12.3 – 16.6 Ma; Table 3, node \textit{Emberiza}). However, the phylogenetic relationships among these four clades (sister clades I+II and III+IV) were poorly supported in all reconstructions. A strongly supported clade I mostly comprised Western Palearctic species and several endemics from the high alpine elevations of the Tibetan Plateau. The Tibetan bunting, \textit{E. koslowi}, turned out as an early offshoot of that clade and the split age from the crown group was dated to 9.0 – 11.9 Ma (Table 3). The basal split of clade I separated corn buntings, \textit{E. calandra}, as the oldest lineage split (10.0 – 13.4 Ma; Table 3) from all other members of that clade. All nodes of clade I received strong to full Bayesian support except for the relationships among two clades of \textit{E. godlewskii} and its closest relatives \textit{E. cia} and \textit{E. cioides} (see intraspecific patterns, below). Generally, partitioning by gene and codon position yielded slightly older age estimates for the deeper splits, however, estimates for younger nodes (e.g. sister species) were the same for both partitioning schemes (Table 3).

The Western Palearctic-Tibetan Clade I was sister to a rather heterogeneous clade containing only four species (Fig. 3). Eastern Mediterranean \textit{E. melanocephala}, Central Asian \textit{E. bruniceps} and East Asian \textit{E. lathamii} formed a terminal monophyletic group with full support (Fig. 3, clade II). Surprisingly, the only African species not included in the African clade, \textit{E. affinis}, was recovered as sister to clade II, however without support from Bayesian posterior probabilities (Fig. 3). In the maximum likelihood analysis this species was sister to clade III, albeit with equally weak bootstrap support (27%).

Clade III included several Eastern Palearctic species with breeding ranges extending into northern and central China as well as all bunting species endemic to Japan. A basal split separated the sister species pair \textit{E. siemsseni} and \textit{E. elegans} (clade IIIa) from a second clade (both subclades received strong support; Fig. 3). The crown clade (IIIb, Fig. 3) was comprised of two Japanese endemics (\textit{E. sulphurata} and \textit{E. variabilis}) and several Eastern Palearctic species (four of them including populations from Japan). Six of those species occur exclusively in the Eastern and Central Palearctic with three of them reaching into northeastern Scandinavia (\textit{E. aureola}, \textit{E. pusilla}, \textit{E. rustica}; the western range limits of \textit{E. pallasi} reach NE European Russia) and only the red bunting, \textit{E. schoenicius} occupies an extensive range in the Western Palearctic.

Finally, clade IV contained all African species except for \textit{E. affinis}. A basal split separated \textit{E. cabanisi}, \textit{E. flaviventris} and \textit{E. poliopleura} (Fig. 3; clade IVa) from a terminal clade including some Middle East endemics such as \textit{E. striolata} and \textit{E. socotrana} (Fig. 3; clade IVb). Both subclades received full Bayesian support, however their sister group relationship was only poorly supported when the data set was partitioned only by gene but not by codon position. When mtDNA markers were additionally partitioned by codon positions the two subclades of the African clade were not sister to each other and their relationships to the other three main \textit{Emberiza} clades were only poorly resolved (not shown).

Intra- and interspecific lineage separation

The tree based on the full cyt-\textit{b} data set is shown in Fig. 4. The same four major \textit{Emberiza} clades as in the multi-locus analysis were recovered, however, with weaker support for main clades and relationships among the clades were entirely unresolved. Strikingly, no remarkable East-West Palearctic split could be detected in any bunting species. No subclades coinciding with geographic distribution were recovered across trans-Palearctic breeding ranges of focal species and the corresponding haplotype networks did not show any phylogeographic structure ei-
Table 3. Split ages estimates for selected nodes of the Old World bunting tree inferred from two independent runs with BEAST (means [95% highest posterior density intervals] in Ma); the African clade IV was not recovered as a monophyletic unit when the alignment was partitioned by gene and codon (−); * = excluding E. affinis.

<table>
<thead>
<tr>
<th>Node</th>
<th>Regional split</th>
<th>5 partitions by gene only</th>
<th>9 partitions by gene and codon positions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emberiza</td>
<td></td>
<td>12.3 [10.5–14.1]</td>
<td>16.6 [14.8–18.6]</td>
</tr>
<tr>
<td>Clade I</td>
<td></td>
<td>10.0 [8.5–11.6]</td>
<td>13.4 [11.8–15.1]</td>
</tr>
<tr>
<td>Clade II*</td>
<td></td>
<td>6.4 [5.0–8.1]</td>
<td>8.0 [6.2–10.0]</td>
</tr>
<tr>
<td>Clade III</td>
<td></td>
<td>9.5 [8.0–10.9]</td>
<td>12.0 [10.5–13.5]</td>
</tr>
<tr>
<td>Clade IV</td>
<td></td>
<td>10.2 [8.5–11.9]</td>
<td>—</td>
</tr>
<tr>
<td>E. koslowi</td>
<td></td>
<td>9.0 [7.7–10.5]</td>
<td>11.9 [10.5–13.5]</td>
</tr>
<tr>
<td>E. s. personata / E. s. spodocephala</td>
<td>Japan vs. cont. E Asia</td>
<td>1.6 [1.2–2.1]</td>
<td>1.7 [1.2–2.2]</td>
</tr>
<tr>
<td>E. variabilis / E. tristrami</td>
<td>Japan vs. cont. E Asia</td>
<td>2.5 [1.9–3.2]</td>
<td>2.8 [2.0–3.6]</td>
</tr>
<tr>
<td>E. godlewskii South / E. cia</td>
<td>S China vs. W Palearctic</td>
<td>2.2 [1.7–2.7]</td>
<td>2.5 [2.0–3.1]</td>
</tr>
<tr>
<td>E. godlewskii North / E. cioides</td>
<td>N China vs. W Palearctic</td>
<td>2.4 [1.9–2.9]</td>
<td>2.6 [2.0–3.2]</td>
</tr>
<tr>
<td>E. e. elegantula / E. e. elegans</td>
<td>C China vs. E Palearctic</td>
<td>1.0 [0.7–1.3]</td>
<td>1.0 [0.7–1.4]</td>
</tr>
<tr>
<td>E. c. flemingorum / E. c. cia</td>
<td>Himalaya vs. N Palearctic</td>
<td>0.5 [0.3–0.8]</td>
<td>0.5 [0.3–0.8]</td>
</tr>
</tbody>
</table>

ther (see E. rustica as an example; Fig. 4D; also E. schoeniclus, network not shown). Haplotypes of two Himalayan taxa turned out to be distinct from their northwestern or northeastern counterparts: E. fucata aracata and E. cia flemingorum (Fig. 4). Though the branching pattern in E. cia received no support from the partitioned cyt data set, E. cia and E. c. flemingorum were fully supported sister taxa in the multi-locus analysis (and in the cyt tree based on the unpartitioned data set) and were dated to a rather young, late Pleistocene origin (Fig. 3; Table 3). Comparison of barcode sequences separated an Asian clade of E. cia (subspecies E. c. flemingorum and E. c. par) from Western Palearctic E. c. cia (Fig. 5B).

A remarkable degree of intraspecific differentiation was found in another three East Asian species. Far East Russian E. elegans elegans were separated from Chinese E. e. elegantula from Shaanxi and Sichuan by seven substitutions in the haplotype network (Fig. 4F). Like for the Himalayan subspecies mentioned above the split between Russian and Chinese E. elegans was dated to the late Pleistocene (Fig. 3; Table 3). In contrast, Russian E. spodocephala spodocephala were not notably distinct from Chinese E. s. sordida, however one haplotype of the Japanese subspecies E. s. personata was separated from all continental haplotypes by eleven substitutions (Fig. 4E). Comparison of COI barcode sequences confirmed the strict separation of haplotypes from Japan and Sakhalin from continental haplotypes in Far Eastern Russia, Mongolia and China (Fig. 5A). Unlike split ages in the previous examples, the latter split within E. spodocephala was dated to the Pliocene-Pleistocene boundary and roughly coincides with the slightly older split between Far East Russian E. tristrami and Japanese E. variabilis (Fig. 3; Table 3).

The most striking and unexpected intraspecific diversification pattern was found in E. godlewskii. Southwestern Chinese haplotypes from Yunnan, Sichuan and Qinghai were separated from northern and northeastern haplotypes from Gansu, Ningxia and Hebei by a minimum of 30 substitutions (Fig. 4C). The two clades of E. godlewskii did not form a monophyletic clade neither in single-locus nor in multi-locus reconstructions (Figs 3, 4). In the latter, a sister group relationship of southern E. godlewskii with E. cia received full support, while a sister group relationship of northern E. godlewskii with E. cioides received moderate support (Fig. 3). The striking differentiation between the two E. godlewskii clades was also supported by COI barcodes: Along with our samples from Chinese provinces Gansu and Ningxia the northern godlewskii clade included further samples from adjacent Mongolia and Russia (Fig. 5B).

In contrast to the findings outlined above, no mtDNA lineage sorting could be neither found among E. hortulana and E. caesia nor among E. citrinella and E. leucocephalos (Fig. 4A, B).

Discussion

Genus-level systematics

Our five-gene tree topology generally confirmed the four major clades of Old World Emberiza buntings already found by Alström et al. (2008). Likewise, our topology is well in accordance with the recently revised genus-level systematics of Old World buntings by Dickinson & Christidis (2014). They restricted the genus Emberiza to those species contained in our clade I (including Ti- betan E. koslowi), and our clade III fully reflects their genus Schoeniclus. The terminal sister species of our clade II were separated as genus Granatitova from the monotypic genus Melophus by Dickinson & Christidis (2014). This separation is compatible with our topology and seems well justified with respect to the strong morphological distinctiveness of Melophus lathami as the only Old World bunting having a prominent crest in
both males and females. The only striking discordance with the generic classification by Dickinson & Christidis (2014) is the highly unexpected position of *Emberiza affinis* that did not cluster with the other African buntings (*Fringillaria* sensu Dickinson & Christidis, 2014; our clade IV) but was sister to the clade unifying *Melophus* and *Granatitvora* sensu Dickinson & Christidis (2014) in the Bayesian multi-locus tree. However, this placement of *Emberiza affinis* is highly doubtful because it received poor support in all analyses as it was based on evidence from two mitochondrial markers only. In fact, close affinities of *Emberiza affinis* to other African species seem likely with respect to zoogeography but particularly with respect to strong similarities in external morphology with *E. flaviventris*, *E. cabanisi* and *E. poliopleura* (our clade IVa).

These four African species were unified in a separate subgenus by some authors (*Cosmospina Wolters, 1972*; cf. Kobluk, 2007) and likewise separated from other African species of subgenus *Fringillaria*. Within our clade IVa we were able to confirm the supposed sister species relationship between previously unstudied *E. poliopleura* and *E. flaviventris* (Rose, 2011). Notably, the two main African clades (IVA and IVb) were not recovered as sister clades in all multi-locus reconstructions, because their position was highly dependent of sequence data partition. Thus the phylogenetic relationships among the subclades of African *Fringillaria* sensu Dickinson & Chistidis (2014) remain subject of further multilocus or genomic studies.
Fig. 4. Single-locus phylogeny of Old World buntings (Emberiza) based on the mitochondrial cytochrome-\(b\) gene, Bayesian tree (no codon partition applied); node support from posterior probabilities above nodes; node support from RAxML analysis below nodes; * = full support from Bayesian posterior probabilities (1.0) and thorough likelihood bootstrap (100), posterior probabilities < 0.9 not shown; haplotype networks for selected species (pairs) are shown in boxes A–F at the right side of corresponding clades; D) two clades of \(E. rustica\) did not reflect a separation of East (grey) and West (white) Palearctic samples; # = two sequences from Åström et al., (2008) that did not clade with other sequences from the same species (for \(E. lathami\), see material); non-Emberizidae outgroups omitted.
Inter- and intraspecific differentiation of buntings

The few studies dedicated to phylogeographic structure within bunting species found remarkably low differentiation across the wide trans-Palaearctic breeding range of the reed bunting, *E. schoeniclus*. Even though two mtDNA lineages could be identified these reflected only a shallow split and did not coincide with geographic distribution (Zink et al., 2008) – similar to the lack of within-clade differentiation of *E. rustica* (our study). In all cases where significant genetic differentiation among reed bunting populations could be confirmed, this was reflected by distribution patterns of haplotype and allele frequencies, e.g. on the Iberian Peninsula (Kvist et al., 2011). Evidently, the use of mitochondrial genes seems to be problematic for the assessment of intraspecific phylogeographic structure and species identification in buntings. Moreover, in some cases discrimination between sister species is not possible with mtDNA markers. The best-known example is the case of the yellowhammer and the pine bunting (*E. citrinella, E. leucocephalos*) that share a common mitochondrial gene pool but are strongly distinguished by nuclear markers even in regions of local sympatry (Irwin et al., 2009; for a lack of a barcode gap compare Kerr et al., 2009). According to their close phylogenetic relationship, territorial songs of these two species were mutually understood in playback experiments (Tietze et al., 2012). A similar lack of genetic distinctiveness among species was documented for the subarctic genetic differentiation in buntings. Genetic distinctiveness among buntings is required through sampling of further specimens and nuclear markers (E. caesia in this study (although confirmation is required through sampling of further specimens and nuclear markers).

In contrast to these cases of low mtDNA divergence among species, there are examples of striking subspecific genetic differentiation in buntings. Genetic distinctiveness of *E. spodocephala* populations from Japan and Sakhalin (subspecies *personata*) as compared to mainland conspecifics was previously inferred from RADP analysis (Dolgoval & Valchuck, 2008) from and COI barcoding analyses (Kerr et al., 2009; Saitoh et al., 2015). This differentiation pattern was also reflected by a relatively old split age in our multi-locus phylogeny. In fact, *E. s. personata* also exhibits strong distinctiveness of external morphology against continental populations where phenotypical variation follows a cline between continental and Siberian populations. This cline is reflected by a barcode gap analysis and COI barcoding analyses (Kerr et al., 2009, for *E. hortulana* and *E. caesia* in this study (although confirmation is required through sampling of further specimens and nuclear markers).

In contrast to the few studies dedicated to phylogeographic structure within bunting species, species boundaries between buntings can be problematic for the assessment of intraspecific phylogenetic relationships. Thereby following the suggestion by Hartert (1928), Portenko (1960) and Vaurie (1962) later adopted that classification. In contrast, Maurer (1972) separated *E. godlewskii* at the species level from *E. cia* mainly based on variation in body size, feather proportion and plumage coloration. He stated that variation is clinal within each of the two taxa but at the same time there is no evidence of clines or intermediate populations among vicariant populations of the two species in Asia. Strikingly, clinal morphological variation of *E. godlewskii* does not reflect strong character discontinuities, as one would expect from the deep split in our phylogeny. The southernmost forms *yunnanensis* and *khamensis* differ from all other subspecies in having a less pronounced grayish neck-ring (Maurer, 1972). *yunnanensis* is rather short-winged with mean wing-lengths of 81 mm [75–86 mm] in males and 77.5 mm [73–82 mm] in females (Vaurie, 1956). The short wing length distinguishes *yunnanensis* from the neighboring subspecies *omissa* in the North, which has slightly longer wings (Vaurie, 1956: measurements for males between 78.5 and 87 mm). Within the North-South cline along the eastern Tibetan Plateau margin, the subspecies *khamensis* represents an intermediate form that intergrades with the paler nominate *godlewskii* in the North and with the more colourful reddish-brown *yunnanensis* in the South (Vaurie, 1956). Similar to the latter form, its Himalayan counterpart *E. cia flemingorum* was shown to be the most short-winged subspecies of *E. cia* (Martens, 1972). However, in that case similarity of wing proportions may not necessarily indicate common ancestry as suggested in our phylogeny but may rather be due to the fact that migration of *E. c. flemingorum* is probably limited to seasonal elevational movements like in many other Himalayan passerines. The morphological distinctiveness of the Nepalese subspecies *flemingorum* against its western counterparts *stracheyi* and *par* (Martens, 1972) is reflected by the clear split among one haplotype of that form from all western *E. cia* included in our molecular tree (but compare the separation of a Western Palaearctic cluster from a Central Asian/Himalayan cluster in the COI barcode analysis). However, all these results have to be substantiated by more extensive sampling and further integrative research including molecular, morphological and bioacoustic markers before any taxonomic consequences should be drawn.
Fig. 5. Intraspecific genetic variation of Asian buntings based on a 613 bp alignment of COI barcodes; Bayesian inference of phylogeny with MrBayes: A) black-faced bunting, *E. spodocephala*; B) Godlewski’s bunting, *E. godlewskii*; note that the *E. cioides* clade includes two GenBank sequences from samples of false species determination (EU847682 listed as *E. cia*, EU847683 as *E. godlewskii*).
Biogeographic history of Tibetan Plateau species

The high elevations of the central Qinghai-Tibet Plateau have been repeatedly considered a “cradle of evolution” harbouring rather old relic genetic lineages of cold-adapted mammals (DENG et al., 2011; TSENG et al., 2013; WANG et al., 2014) and of passerine birds (WEIGOLD, 2005; LEI et al., 2014; PÄCKERT et al., 2015). According to our dated multi-locus phylogeny the Tibetan bunting, E. koslowi represents one of those rather ancient endemic species of the Tibetan Plateau. This species was already highlighted by VAURIE (1972) in his list of Tibetan endemics and WEIGOLD (2005) listed this species as one of his “first degree endemics” and a character species for the treeless grasslands of the plateau region. Our split age estimates suggest a mid Miocene separation of the Tibetan bunting from its closest relatives and a likewise early colonization of the alpine plateau habitats by this species. This is in accordance with recent evidence that vast parts of the plateau region had already reached present-day altitudes in the Miocene (MULCH & CHAMBERLAIN, 2006; FÄRRE et al., 2014). The hypothesis of a rather ancient Miocene colonization of alpine Tibetan Plateau habitats by ancestors of extant passerine species received recent support from studies of other plateau endemics that have comparably isolated phylogenetic positions and were dated back to similarly ancient split ages, such as the Tibetan rosefinch, Carpodacus roborowiskii (ZUCCON et al., 2012; TIETZE et al., 2013) and the Tibetan ground tit, Pseudopodoces humilis (JOHANSSON et al., 2013; QU et al., 2013).

There is recent evidence from a genomic study that during long lasting separation the latter species strongly adapted to ground-dwelling life at high-elevation environments due to an effect of positive selection on genes related to hypoxia response and skeletal development and expansions in genes involved in energy metabolism (QU et al., 2013). Whether similar genetic adaptations will be found in other regional endemics, such as E. koslowi remains the subject of forthcoming genomic studies.

In the Eastern Palearctic lineage splits among some sister-taxon pairs roughly coincide with the Pliocene-Pleistocene boundary at 2.58 Ma (according to COHEN et al., 2013). However, our dated phylogeny hints to several successive phases of faunal interchange among the Japanese Archipelago and the Far East Asian mainland with the oldest colonization of southern Japan by ancestors of E. sulphurata in the early Pliocene followed by three parallel events of faunal interchange at the Pliocene-Pleistocene boundary (for E. spodocephala compare SAITOH et al., 2015; WEISENSTEINER, 2013). A lack of considerable differentiation among Japanese and continental mtDNA haplotypes of E. cioides and E. yessoensis suggest two further very recent events of interchange due to Holocene range expansion from either an insular or a continental glacial refuge (compare SAITOH et al., 2015). Generally, these examples of Pleistocene lineage divergence between populations from Japan and Sakhalin against their closest relatives from the mainland are paralleled for example in Phylloscopus leaf warblers (P. borealis group: ALSTRÖM et al., 2011; SAITOH et al., 2010; species pair P. tenellipes vs. P. borealoides: MARTENS, 1988; PÄCKERT et al., 2012).

Late Pleistocene lineage diversification among morphologically distinct subspecies along the eastern margin of the Qinghai-Tibet Plateau is evident in E. elegans but not in E. spodocephala occurring in the Eastern Palearctic and China respectively. In contrast, the most striking phylogeographic pattern along the eastern plateau margin is the deep north-south divide in E. godlewskii. Even more unexpectedly this deep divergence does not separate allopatric populations from central/northern China and southern Siberia (that are separated by a large distribution gap) but bisects a continuous Chinese distribution range (see maps in MAUERSBERGER & PORTENKO, 1971; WEIGOLD, 2005; ROSE, 2011). Signatures of Quaternary climate oscillations in the demographic history and thus the extant phylogeographic structure of alpine endemic species from the Tibetan Plateau are not uncommon and were previously documented for a number of species (QU et al., 2009, 2010; YANG et al., 2009; LEI et al., 2014). However, none of these previously investigated examples involved a deep north-south divide comparable to our finding in E. godlewskii. Strikingly, ranges of southern E. godlewskii (north of the Himalayan main range) and E. cia (south of it) come close to each other on the southern macroslope in the Central Himalayas. Even though adjacent populations of their circum-Tibetan ranges are separated by distribution gaps in several places, MAUERSBERGER (1972) already drew a parallel to the few classical examples of “circular overlap”. Previously, MAUERSBERGER & PORTENKO (1971) already cast into doubt whether the assumed distributional gap in southern Tibet was real or due to a lack of field records from this region. In the Himalayas their easternmost record of E. godlewskii originated from Kharta in the eastern Mt Everest area (north of the main range). MARTENS (1972) commented on these easternmost Everest populations that the species had never been recorded West of this location or anywhere throughout Nepal south of the Himalayan main range (not even as a vagrant or migrant) and postulated an ecological component of the observed distribution limits. Throughout its Tibetan range E. godlewskii seems to be restricted to the drier open habitats at high elevations and strictly avoids the monsoon-humid southern flanks of the Himalayas (MARTENS, 1972). This ecological segregation is paralleled on the eastern plateau margins where SCHÄFER (1938) found three different forms of E. godlewskii co-occurring in elevational parapathy in the surroundings of Batang. Generally, E. c. yunnanensis rather occupies subtropical valleys at lower elevations while E. c. khamensis occurs at the high alpine and subalpine shrubs up to 4000 m (with omissa at intermediate elevations in areas of local parapathy where dense forests separate the habitats of the lowland and the alpine subspecies; SCHÄFER, 1938; MAUERSBERGER & PORTENKO, 1971). This kind of elevational segregation among alpine and lowland breeding habitats of northern and southern
E. godlewskii is in fact the only convincing hypothesis that might explain the deep lineage separation among E. g. yunnanensis and its northern conspecifics. Elevational parapatry and ecological segregation of closely related congeners is a common phenomenon in forest-dwelling passerines of the southern and southeastern flanks of the Tibetan Plateau (Johansson et al., 2007; Martens et al., 2011; Packert et al., 2012; Price, 2010; Price et al., 2014). In contrast, a similar phenomenon has rarely been documented for alpine species of the high elevation plateau habitats. Ecology and elevational distribution has been assumed to play a role in lineage separation and speciation processes in the beautiful rosefinch and its allies (C. pulcherrimus, C. waltoni; Tietze et al., 2013; Packert et al., 2015), however, like in the case of E. godlewskii the knowledge on local habitat preferences and the breeding biology of the populations involved is scarce and geographically more extensive sampling is required in order to evaluate range-wide phylogeographic patterns.

Acknowledgements

These results were based on the rich material and field observations from a multitude of nearly annual expeditions of JM to various parts of Asia from the year 1969 until today. This study received substantial funding from Deutsche Forschungsgemeinschaft, PA 1818/3-1 and from Staatsministerium für Forschung und Kunst (SMWK) State of Saxony (AZ-4-7531.50-02-621-08/1). During the years JM received several grants from Deutsche Ornithologen-Gesellschaft (DO-G), Gesellschaft für Tropenornithologie (GTO) and from Feldbausch-Stiftung and Wagner-Stiftung both at Fachbereich Biologie, Johannes Gutenberg-Universität Mainz, Germany. Y-HS received research grants from the National Natural Science Foundation of China, project No. 31272286. DTT was funded by the Deutsche Forschungsgemeinschaft (Ti 679/1-1, Ti 679/2-1). We would like to thank the following colleagues who provided samples for genetic analysis: M. Adams & R. Prŷs-Jones, Bird Collections of the Natural History Museum at Tring, UK; B. Haberl, Mainz; J. Fjeldså, Natural History Museum of Copenhagen, Denmark; S. Frahnert, Museum für Naturkunde Berlin, Germany; H. Grimm & M. Fischer, Naturkundemuseum Erfurt, Germany; P. Kessler, Mainz; J.O. Kriegs, LWL Museum für Naturkunde, Münster, Germany; P. Lymberakis, Natural History Museum of Crete, Irakleio, Greece; N. Rice, Bird Collection at the Academy of Natural Sciences at Drexel University, Philadelphia, USA; R. van den Elzen and T. Töpfer, Forschungsmuseum Alexander Koenig, Bonn, Germany. A part of the 657-bp cytochrome-b data set was generated and analysed by C. Wallmann within the scope of her diploma thesis at Johannes Gutenberg University of Mainz, Germany. We cordially thank A. Rauh for steady assistance and support during lab analyses of toe pad samples. We would furthermore like to thank D. Lu for helping with the lab work in China and for providing sequences for two markers from one Tibetan bunting sample. We are particularly grateful to Miss L. Li for establishing contact with a local Tibetan community in south-eastern Qinghai, especially to Kenpo Tashi Zangpo-Ju. We are very thankful for his hospitality and for giving us guided tours in the habitats of the Tibetan bunting. We would like to dedicate this work to Kenpo Tashi Zangpo and his passionate scholarly efforts for the conservation of endemic wildlife in the region of Nyanpo Yutse.

References


Electronic Supplement File

at http://www.senckenberg.de/vertebrate-zoology ("Contents")

(Table, origin of samples for genetic analysis)

(Table, PCR primer combinations for PCR with toe pad samples from museum specimens).