Resource stability and geographic isolation are associated with genome divergence in western Palearctic crossbills

THOMAS L. PARCHMAN*, PIM EDELAAR†, KATHRYN UCKELE‡, EDUARDO T. MEZQUIDA§, DANIEL ALONSO¶, JOSHUA P. JAHNER*, RON W. SUMMERS** & CRAIG W. BENKMAN††

*Department of Biology, University of Nevada Reno, Reno, NV, USA
†Department of Molecular Biology and Biochemical Engineering, University Pablo de Olavide, Seville, Spain
‡Program in Ecology, Evolution, and Conservation Biology, University of Nevada Reno, Reno, NV, USA
§Department of Ecology, Faculty of Sciences, Autonomous University of Madrid, Madrid, Spain
¶Department of Ornithology, Aranzadi Sciences Society, Donostia-S. Sebastian, Spain
**Royal Society for the Protection of Birds Centre for Conservation Science, North Scotland Regional Office, Inverness, UK
††Department of Zoology and Physiology, University of Wyoming, Laramie, WY, USA

Keywords:
divergent selection; geographic isolation; Loxia curvirostra; Loxia pytyopsittacus; Pinus halepensis; residency; serotiny.

Abstract

While many conifers produce annually variable seed crops, serotinous species (which hold seeds in cones for multiple years) represent unusually stable food resources for seed predators. Such stability is conducive to residency and potentially population divergence of consumers as exemplified by the Cassia crossbill (Loxia xinesciuris) in North America. We used genotyping by sequencing (GBS) to test whether three Mediterranean subspecies of common crossbills (Loxia curvirostra) associated with the serotinous Aleppo pine (Pinus halepensis) were more genetically distinct than European crossbills associated with nonserotinous conifers. We assembled a Cassia crossbill draft genome as a reference for mapping GBS reads and as a first step towards a more contiguous genome assembly. We found clear patterns of genetic divergence for each of the Pinus halepensis-associated subspecies. Geographic isolation, as promoted by resource stability and residency, is associated with genetic divergence of two of these subspecies. However, geographic isolation cannot account for divergence of L. c. hispana. Instead, resource stability likely contributed to divergence by reducing dispersal and increasing resource competition that may limit breeding by immigrants. In contrast, we found no differentiation among common crossbills associated with less stable resources, and only slight differentiation between common crossbills and parrot crossbills (Loxia pytyopsittacus). The substantial morphological divergence between common and parrot crossbills has likely originated or been maintained by selection despite gene flow generated by spatiotemporal resource fluctuation. Our results indicate that phenological as well as morphological characteristics of conifers have influenced crossbill diversification, and suggest a possible link between resource stability and population divergence.

Introduction

Climatic stability over geological time appears to facilitate small range size, population divergence and specialization in biotic interactions (Dynesius & Jansson, 2000; Martin & McKay, 2004; Dalsgaard et al., 2011; Sandel et al., 2011). Similarly, resource stability across seasons and years could be critical to local population
the highest seed kernel intake rates (a direct measure of seed availability) for crossbills during a year exceed the lowest intake rates by only 68% when foraging on the serotinous cones of lodgepole pine (Benkman et al., 2012), but by 200–500% when foraging on nonserotinous conifers (Benkman, 1987). Importantly, more stable resources allow crossbills, such as the Cassia crossbill, to form resident populations and to grow closer to carrying capacity (Benkman et al., 2010, 2012), resulting in stronger and more consistent food limitation (Benkman et al., 2012). Furthermore, residency favours local adaptation in bill traits (Benkman, 2003) that in combination with strong food limitation prevents nonlocally adapted, immigrant ecotypes from persisting (surviving, reproducing), thereby limiting gene flow (Smith & Benkman, 2007; Benkman, 2017).

Within the western Palearctic crossbills, the only serotinous conifer in western Europe regularly fed on by crossbills is Aleppo pine (P. halepensis), which has a Mediterranean distribution (Fig. 1). Although we lack data on feeding intake rates in the western Palearctic, Aleppo pine holds its seeds in cones for several years and consistently produces seed crops every year (Table 1; Tapias et al., 2004) like other serotinous species. Three sub-species of common crossbills are thought to be seden-
tary and associated with P. halepensis in the western Mediterranean: L. c. balearica on the island of Mallorca, L. c. poliogyna in northern Africa and L. c. hispana in south-eastern Spain (Fig. 1). In contrast, the seed crops of other conifers used by crossbills in Europe are spa-
tiotemporally variable (Table 1). This promotes large and long-distance irruptive nomadic movements (New-
ton, 2006) and mass reproduction on bumper crops, both of which can be conducive to gene flow. For example, Europe’s largest billed crossbill (parrot crossbill L. pytopsittacus; see Fig. 1) breeds sympathetically with the much smaller billed L. c. curvirostra when large cone crops of Scots pine (P. sylvestris) or Norway spruce (Picea abies) result in abundant and easily accessible seeds (Summers et al., 2007; see also Nethersole-
Thompson, 1975). (Hereafter, we refer to common crossbill subspecies by their subspecific epithet, including the widespread continental L. c. curvirostra, and use the common name of L. curvirostra to refer to the species as a whole; we refer to parrot crossbills by their specific epithet, pytopsittacus.)

Suggestive of a positive effect of resource stability on genetic divergence, evidence for genetic differentiation between balearica (a P. halepensis specialist) and curviro-
stra from Europe has been reported (Björklund et al., 2013), although little is known about genetic variation in hispana (Edelaar et al., 2012) or poliogyna. Geographic isolation alone should lead to genetic differentiation for balearica and poliogyna (Fig. 1). Balearica is resident on
Mallorca (Massa, 1987), and dispersal by nomadic *curvirostra* from the continent to the island is uncommon (Summers & Jardine, 2005; Björklund et al., 2013). *Poliogyna* is resident in northern Africa (Massa, 1987) and is even more isolated from European populations of *curvirostra* due to expanses of sea and nonsuitable habitat (Griscom, 1937; Cramp & Perrins, 1994). Consequently, *balearica* and *poliogyna* do not provide independent tests for the importance of resource stability to population divergence. However, resource stability has likely played a key role in their ability to form resident, geographically isolated populations and could thereby favour population divergence via drift alone.

*Hispana*, however, is not geographically isolated from *curvirostra* because multiple pine species used by nomadic *curvirostra* occur near and within the mainland range of *P. halepensis* (Fig. 2). During periodically large cone crops of these nonserotinous conifers, nomadic movements of *curvirostra* likely bring them into regular contact with *hispana*. Thus, the relative extent to which *hispana* is genetically differentiated provides a test of the importance of stable seed availability to population divergence, more comparable to the Cassia crossbill in North America. *Hispana* was once described as a subspecies (Hartert, 1904), but because of morphological and geographical similarity with *balearica* and *curvirostra* has not been recognized in subsequent taxonomic treatments (Griscom, 1937; Vaurie, 1956; Cramp & Perrins, 1994; Clement et al., 2010). Nevertheless, crossbills are resident in south-eastern Spain, where *P. halepensis* forests having high frequencies of serotiny are extensive (Fig. 2; Hernández-Serrano et al., 2014) and *hispana* is thought to occur (Cramp & Perrins, 1994). They also have distinct bill morphologies relative to *curvirostra* from northern Spain associated with other conifers (Edelaar et al., 2012), suggesting local adaptation.

In contrast to the evidence for genetic divergence of *balearica* (Björklund et al., 2013), other studies have reported a lack of discernible genetic differentiation among most populations of common crossbills in Europe (Questiau et al., 1999), and a surprising lack of differentiation between two morphologically differentiated species: common crossbills and *pytyopsittacus* (Piertney et al., 2001; Björklund et al., 2013). As in North America, this pattern is likely influenced by ongoing gene flow arising from the nomadic movements that most crossbills exhibit, as well as their recent diversification (e.g. post-Pleistocene, < 11 000 years BP; Björklund et al., 2013).

Fig. 1 Map showing the geographic distributions of common crossbills (*L. curvirostra*; intermediate grey), parrot crossbills (*L. pytyopsittacus*; stippling) and *Pinus halepensis* (dark grey around the Mediterranean), and sampling locations of crossbills (circles with colour corresponding to taxon). Also shown are an illustration of *Pinus halepensis* cones (left side), and head profiles of representative common crossbills from Europe (*Loxia c. curvirostra*), Mallorca (*L. c. balearica*), the Atlas Mountains of North Africa (*L. c. poliogyna*) and parrot crossbill (*L. pytyopsittacus*). Distribution maps of crossbills are from the Atlas of European Breeding Birds (Hagemeijer & Blair, 1997), and the map for *Pinus halepensis* is from EUFORGEN (www.euforgen.org).
Table 1 Seed production characteristics of four species of pine (Pinus) and one spruce (Picea) commonly consumed by crossbills in western Europe and the Mediterranean [from Tapias et al. (2004) for pines; from Young & Young, (1992) for spruce]. Only Pinus halepensis is serotinous. The crossbill taxa in this study that commonly feed on these conifers, and whether they are nomadic or sedentary are provided in the fourth column.

<table>
<thead>
<tr>
<th>Species</th>
<th>Cone persistence in canopy (years)</th>
<th>Interval between large cone crops (years)</th>
<th>Associated crossbills (nomadic/resident)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pinus halepensis</td>
<td>5–20*</td>
<td>1</td>
<td>L. c. balearica, L. c. hispana, L. c. poldogyna (resident)</td>
</tr>
<tr>
<td>Pinus sylvestris</td>
<td>1–3†</td>
<td>2–4</td>
<td>L. c. curvirostra, L. pytyopsittacus (nomadic)</td>
</tr>
<tr>
<td>Pinus nigra</td>
<td>1–3†</td>
<td>2–4</td>
<td>L. c. curvirostra (nomadic)</td>
</tr>
<tr>
<td>Pinus uncinata</td>
<td>1–3†</td>
<td>3–4</td>
<td>L. c. curvirostra (resident)</td>
</tr>
<tr>
<td>Picea abies</td>
<td>1</td>
<td>3–13</td>
<td>L. c. curvirostra (nomadic)</td>
</tr>
</tbody>
</table>

*Within Spain, ~40–80% of the trees produce serotinous cones that hold seeds for several years or even decades (Tapias et al., 2004).
†Nearly, all seeds are shed from cones during the first year (Mezquida et al., 2018; E. T. Mezquida & C. W. Benkman, personal observations).

et al., 2013). However, past studies were based on small sets of molecular markers, and more comprehensive genomic sampling should provide additional insight into patterns of population genetic structure characterizing crossbills in the western Palearctic.

Here, we use a genotyping-by-sequencing (GBS) approach (a variant of dRADseq; Parchman et al., 2012; Peterson et al., 2012) to quantify patterns of phylogenetic and population genetic structure among western Palearctic crossbills. We used Illumina sequencing of variable insert size libraries to assemble a Cassia crossbill draft genome to serve as a reference onto which we mapped GBS reads. We compare genetic differentiation of the three Pinus halepensis-associated subspecies to population samples of curvirostra from regions with less stable seed resources: Scotland where curvirostra feeding on P. sylvestris and various non-native conifers are nomadic (Marquiss & Rae, 2002; Summers et al., 2007) and from the Pyrenees in northern Spain where crossbills feed on Pinus uncinata and P. sylvestris (Senar et al., 1993; Clouet, 2000; Alonso et al., 2016; Fig. 1). Although the annual seed crops in the Pyrenees are sufficiently stable from year to year to support crossbills year-round, all cones open in late winter and spring providing a pulse of easily accessible seeds. During such pulses, food competition is unlikely to prevent multiple forms of crossbills from breeding, unlike when crossbills feed on serotinous species. Consequently, we expect less genetic divergence among the Scottish and Pyrenees populations of curvirostra than for those associated with P. halepensis. In addition, we ask whether more extensive genomic sampling detects evidence for genetic differentiation between curvirostra and pytyopsittacus, as might be expected based on their phenotypic distinctiveness (Summers et al., 2002) and strong assortative mating (Summers et al., 2007). Finally, we have vocal recordings of many of the curvirostra from Scotland and the Pyrenees, which allowed us to assign them to vocal types (Robb, 2000; Summers et al., 2002) and to test for genetic differentiation among them. In contrast to North America, evidence of morphological, ecological and evolutionary distinctiveness of curvirostra vocal types in Europe is more limited (Robb, 2000; Summers et al., 2002; Edelaar, 2008; Edelaar et al., 2008), and analyses of genetic differentiation have yet to be conducted. Our results, based on more than 18 000 single nucleotide polymorphisms (SNPs), provide little to no evidence of genetic divergence among crossbill vocal types, populations and species associated with highly fluctuating seed resources. In contrast, the P. halepensis-associated subspecies were differentiated from one another and all other crossbills, suggesting that resource stability and

Fig. 2 Natural distributions of four species of pines (Pinus spp.) in Spain fed on by crossbills. Pinus halepensis is in light grey, P. nigra in medium grey, P. sylvestris in dark grey and P. uncinata in black. Map is modified from Edelaar et al. (2012), which is based on species maps from the Spanish Forestry Service at http://www.inia.es/genfored.
geographic isolation may interact to influence population divergence.

**Materials and methods**

**Reference genome sequencing and assembly**

We sequenced and assembled a Cassia crossbill (L. s. nesitsiria) draft genome to provide a reference for aligning GBS data, and to provide a resource for future efforts to produce a more contiguous and complete genome assembly. Aligning GBS data to reference genomes can improve analyses by reducing problems associated with paralogous genomic regions and can provide information on the genomic distribution of markers. Genetic divergence between North American red crossbills and European common crossbills is low (Questiau et al., 2007), making this reference suitable for use in this study. We extracted DNA from a male Cassia crossbill collected in the South Hills (Sawtooth National Forest, Twin Falls County, ID) in September 2012. We constructed four distinct sequencing libraries with insert sizes of 500 base pairs (bp), 800 bp, 2 kilobases (kb) and 5 kb and used each to generate 125-bp paired-end reads on an Illumina HiSeq 2000 at the Beijing Genome Institute (BGI). After quality and contaminant filtering, we retained 94 gigabases (Gb) of data (Table S1).

Sequence data were assembled by BGI using an approach utilized for other recently assembled avian genomes (Jarvis et al., 2014; Zhang et al., 2014) using SOAPdenovo v2.04 (http://soap.genomics.org.cn). First, de Bruijn graphs were constructed by splitting corrected-reads with short insert-sizes into kmers before merging kmers, clipping tips, merging bubbles and removing low-coverage links. Contigs with unambiguous de Bruijn graph connections were collected using a kmer length of 23. All usable reads were mapped to contig sequences using the paired-end information to construct scaffolds step by step, from small-insert libraries to large-insert libraries, with at least three read pairs required to form a connection. Gaps within scaffolds were filled by the gapcloser module of SOAPdenovo. This assembly resulted in 64 453 scaffolds, 2,980 of which were greater than 2kb in length (Table S2). The scaffold N50 was 3.1Mb, and the largest scaffold was 12.3Mb long. The total length of the assembly was 1.1Gb, which covered much of the estimated 1.25Gb genome size (Table S2).

**DNA sequencing, assembly and variant calling**

We obtained DNA from 114 common crossbills in five geographic regions, and from six pyttopsittacus in three regions (Table 2; Fig. 1). We assume that the crossbills sampled in Morocco represent poliogyna, those from Mallorca balearica, and those from Málaga in southeastern Spain hispana. Additionally, nine of the curvirostra sampled from Scotland represented two different vocal types (vocal types C and E of Robb, 2000; or 1A and 4E of Summers et al., 2002) recorded at the time of capture, and 10 of the curvirostra sampled in the Pyrenees represented two vocal types. We also sampled three two-barred crossbills (L. leucoptera) from Europe (Table 2) for use as an outgroup in phylogenetic analyses. Blood or feathers were taken from live-caught birds and stored in Queen’s lysis buffer, or on FTA cards. DNA was extracted using Qiagen DNeasy 96 Blood and Tissue Kits (Qiagen, Inc., Valencia, CA, USA) and quantified with a QIAxpert microfluidic analyzer (Qiagen, Inc.).

We constructed reduced-representation libraries using a GBS method that we have previously used with crossbills and other taxa (Gompert et al., 2012; Parchman et al., 2012). We used two restriction enzymes, EcoRI and MspI, to digest genomic DNA, before ligating

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Sample location</th>
<th>Number of individuals</th>
<th>Source*</th>
<th>Expected heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common crossbill</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L. c. balearica</td>
<td>Mallorca, Spain</td>
<td>19</td>
<td>DA</td>
<td>0.212</td>
</tr>
<tr>
<td>L. c. poliogyna</td>
<td>Atlas Mountains, Morocco</td>
<td>23</td>
<td>CWB, ETM</td>
<td>0.214</td>
</tr>
<tr>
<td>L. c. hispana</td>
<td>Málaga, Spain</td>
<td>27</td>
<td>DA</td>
<td>0.235</td>
</tr>
<tr>
<td>L. c. curvirostra</td>
<td>Pyrenees, Spain</td>
<td>26</td>
<td>PE</td>
<td>0.234</td>
</tr>
<tr>
<td>L. c. curvirostra</td>
<td>Scotland</td>
<td>19</td>
<td>RS</td>
<td>0.232</td>
</tr>
<tr>
<td>Parrot crossbill</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loxia pyttopsittacus</td>
<td>Sweden</td>
<td>2</td>
<td>NHMD</td>
<td>0.220</td>
</tr>
<tr>
<td></td>
<td>Norway</td>
<td>2</td>
<td>NHMD</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Russia</td>
<td>2</td>
<td>Ni</td>
<td></td>
</tr>
<tr>
<td>Two-barred crossbill</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loxia leucoptera</td>
<td>Denmark</td>
<td>3</td>
<td>NHMD</td>
<td></td>
</tr>
</tbody>
</table>

*CWB, Craig W. Benkman; DA, Daniel Alonso; ETM, Eduardo T. Mezquida; NHMD, Natural History Museum of Denmark; NHMO, Natural History Museum of Oslo; NI, Natalia Iovchenko and Ladoga Ornithological Station (Leningrad); PE, Pim Edelaar; RS, Ron Summers.

customized adaptors to the digested fragments. We ligated unique adaptors that contained eight- to 10-bp DNA barcode sequences to the EcoRI ends of fragments from each individual, and ligated standard Illumina adaptors to the MseI ends. Barcoded, restriction-ligation products from all individuals were PCR amplified using Illumina-based PCR primers and a high-fidelity proof-reading polymerase (iProof, Bio-Rad, Inc., Hercules, CA, USA) and were pooled. We ran libraries on a 2% agarose gel and excised fragments between 350 and 450 bp in length, which we purified using Qiagen Gel Extraction Kits (Qiagen, Inc.). This step was used to reduce the portion of the genome targeted for sequencing in order to increase sequencing coverage depth. For quality control, the fragment size distribution was quantified with a BioAnalyzer (Agilent, Inc., Santa Clara, CA, USA), and sequencability was verified with qPCR. We sequenced libraries at the National Center for Genome Research (Santa Fe, NM, USA) using one lane of single-end 100-base sequencing on an Illumina HiSeq 2500.

We removed reads potentially representing contaminant DNA (PhiX, E. coli) or portions of Illumina oligos using bowtie 2_dr (Langmead & Salzberg, 2012) and a series of scripts specifically designed for this purpose (https://github.com/ncgr/tapioca). We trimmed barcodes and restriction cut-site associated bases from sequences, and matched barcodes from each read to the correct individual using a Perl script that allows for correction of sequencing errors in the barcode sequences. We executed reference-based assemblies to align all reads from each individual to the Cassia crossbill reference genome using the ALN and SAMSE algorithms in BWA. The resulting alignments were merged with MUSCLE (Edgar, 2004). To compensate for low sequencing depth, we allowed majority rule consensus base calls to be made for sites with coverage depths from three to five. Consensus sequences with more than 10 undetermined or 10 heterozygous sites were discarded, and remaining reads were clustered with an 88% similarity threshold and mapped to the reference genome with BWA. The resulting alignments were merged with MUSCLE (Edgar, 2004). We also used a de novo assembly approach in order to assess the extent to which the P. halepensis-associated common crossbill sub-species form reciprocally monophyletic groups, and to ask whether they represent more recently derived lineages, as might be predicted based on the recent expansion of P. halepensis into the western Mediterranean (Grivet et al., 2009). For these analyses, we used the entire set of samples above, as well as three L. leucoptera as an outgroup. We aligned reads to the reference genome, created multiple alignments and identified polymorphisms using pyRAD 0.7.8 (Eaton, 2014) due to its use of a global alignment clustering algorithm and allowance of indel variation. Parameter values were set to default unless indicated herein. Reads with more than seven low-quality bases were discarded, and remaining reads were clustered with an 88% similarity threshold and mapped to the reference genome with BWA. The resulting alignments were merged with MUSCLE (Edgar, 2004). To compensate for low sequencing depth, we allowed majority rule consensus base calls to be made for sites with coverage depths from three to five. Consensus sequences with more than 10 undetermined or 10 heterozygous sites were discarded, and remaining consensus sequences were clustered with 88% similarity and mapped to the reference genome. After discarding loci with more than 20 SNPs or 10 indels, our final data set included 203 963 SNPs that were represented in more than 80 samples. Alternative data sets with variable numbers of samples per locus and resulting SNPs were also used for phylogenetic inference and produced highly comparable topologies and support.

The nexus file exported by pyRAD was used as input for SVDQUARTETS (Chifman & Kubatko, 2014) as implemented in paup* 4.0a (Swofford, 2002). SVDQUARTETS is a method of phylogenetic estimation under the coalescent model that uses all or a subset of induced quartets to infer phylogenies; it is well-suited to unlinked SNP data and computationally efficient with large data sets (Chifman & Kubatko, 2015). We randomly sampled 11% of the distinct quartets for the 123 samples and used the QFM algorithm (Reaz et al., 2014) for quartet amalgamation. We quantified support using 1000 non-parametric bootstrap replicates, and added quartets until bootstrap support values stabilized. The final tree
was manually rooted with *L. leucoptera* in FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) based on past studies of mtDNA sequence variation (Questiau et al., 1999; Parchman et al., 2007).

**Population genetic analyses**

We tested for population genetic structure within and among populations, species and vocal types using model-free analyses (principal components analysis, PCA) as well as a hierarchical Bayesian clustering method (*ENTROPY*, Gompert et al., 2014) that infers the number of ancestral populations and estimates individual ancestry proportions. We estimated genotype probabilities and ancestry coefficients starting with the genotype likelihoods generated above using *ENTROPY*, which uses a model similar to the correlated allele frequency model of *STRUCTURE* (Pritchard et al., 2000; Falush et al., 2003). *ENTROPY* simultaneously estimates genotype probabilities and ancestry parameters while incorporating uncertainty due to variation in coverage depth across individuals and loci and sequencing error. *ENTROPY* does not use a priori information about the sampled populations of individuals and assumes that individuals consist of loci with ancestry from one of *k* genetic clusters. The model estimates ancestry proportions for each individual based on a specified value of *k* and calculates deviance information criterion (DIC) as a metric for model comparison. We used *ENTROPY* to characterize population structure across all individuals, and to simultaneously estimate genotype probabilities for each locus in each individual.

To speed up the convergence and stabilization of MCMC chains, we initialized individual ancestry estimates using cluster membership probabilities generated from a *k*-means clustering analysis (*KMEANS* package in R) on principal components estimated from the geno-type covariance matrix in a linear discriminant analysis (*MASS* package in R; Jombart et al., 2010). We used this approach to generate starting values of *q* (admixture coefficients) to initialize MCMC in order to facilitate mixing and convergence. This approach did not constrain posterior sampling and was run with no reference to sample identity. We executed four independent MCMC chains for predefined values of *k* = 2–9; 9 corresponded to the number of potential taxa based on the species, subspecies and vocal types represented in our sampling. We ran each chain for 80,000 steps, following 60,000 burn-in steps and thinned by saving every tenth step. We set the remaining *ENTROPY* parameters to their default values. We checked for mixing and convergence by plotting MCMC steps and inspected mixing and convergence by plotting MCMC steps and inspected mixing and convergence. This approach did not constrain posterior sampling and was run with no reference to sample identity. We executed four independent MCMC chains for predefined values of *k* = 2–9; 9 corresponded to the number of potential taxa based on the species, subspecies and vocal types represented in our sampling. We ran each chain for 80,000 steps, following 60,000 burn-in steps and thinned by saving every tenth step. We set the remaining *ENTROPY* parameters to their default values. We checked for mixing and convergence by plotting MCMC steps and inspected mixing and convergence during the burn-in period. Parameter estimates and DIC values were highly similar for independent chains. We set the remaining *ENTROPY* parameters to their default values. We checked for mixing and convergence by plotting MCMC steps and inspected mixing and convergence during the burn-in period. Parameter estimates and DIC values were highly similar for independent chains. We set the remaining *ENTROPY* parameters to their default values. We checked for mixing and convergence by plotting MCMC steps and inspected mixing and convergence during the burn-in period. Parameter estimates and DIC values were highly similar for independent chains.

To further characterize patterns of genotypic variation across individuals, populations and species, we executed PCA on genotype covariance matrices using *PRCOMP* in R (R Core Team, 2013). We first conducted a PCA on all crossbills except the *L. leucoptera* used as an outgroup above. We used permutational multivariate analysis of variance (*PERMANOVA*, Anderson, 2001) implemented in the *vegan* package (Oksanen et al., 2013) of R to test for significant differentiation among populations based on Euclidian distances of the principal components (PCs). As variation in this PCA was dominated by the crossbills associated with *P. halepensis* and potentially obscured additional differentiation between other groups, we executed two additional PCAs. To test for genetic differentiation between *curvirostra* and *pytyopsittacus*, we executed an additional PCA on a subset of samples including *pytyopsittacus* as well as *curvirostra* sampled from Scotland and the Pyrenees. Finally, to examine whether *curvirostra* vocal types were genetically differentiated, we executed a PCA on the subset of *curvirostra* sampled in Scotland and the Pyrenees for which we had vocal type information. For these additional PCAs, we again used *PERMANOVA* to test for differentiation between groups.

To further quantify relative levels of divergence among population samples, we used population allele frequencies to calculate Nei’s genetic distance (Nei’s D; Nei, 1972) among all populations and species. Additionally, we calculated genome-wide and locus-specific (for each of the 18,047 loci) pairwise Hudson’s *F*<sub>ST</sub> (Hudson et al., 1992) based on allele frequencies at all loci for each population. The relationship between geographic and genetic distance for all populations (including *pytyopsittacus* and all *L. curvirostra* subspecies) was evaluated with a multiple regression on distance matrices (MRM; Lichstein, 2007) as implemented in the *ECODIST* package (Goslee & Urban, 2007) in R. Haversine distance was used as the metric of geographic distance between sampling locales, as calculated from the latitude and longitude of sampling locations using the *FOSSIL* package (Vavrek, 2011) in R, and Nei’s D was used as a metric of genetic distance. As a metric of genetic diversity for each population sample, we calculated expected heterozygosity (*H*<sub>E</sub>) from population allele frequencies based on the expectation from Hardy–Weinberg equilibrium.

**Results**

After quality filtering the raw data, parsing barcodes and discarding individuals with read counts below the 0.25 quantile of the read count distribution for all individuals, we retained ~87 million reads across the 114 common crossbills and six *pytyopsittacus*. Reference-based assemblies executed with *bwa* aligned 73 million reads (84% of the total reads) onto the Cassia crossbill reference genome. We identified 63,658 SNPs where 85% of individuals had at least a single read in the final assemblies. After discarding SNPs with minor allele frequencies < 0.03 and sampling a single SNP per 2,000
bases in the reference genome, we retained a final set of 18,047 SNPs for population genetic analyses. This set of genotypes represented ~9.3 million DNA sequences with a mean coverage per individual per locus of 4.9x.

We also evaluated a de novo approach, by assembling a subset of the raw reads into a GBS reference onto which we aligned all reads using BWA. Using alignment, variant calling and filtering steps identical to those used for the reference genome approach above, we identified a very similar number of variants (17,711) with similar average coverage depth (4.7x) (Supporting Information). Population genetic analyses conducted with this alternative set of genotypes were qualitatively indistinguishable to those presented below (Figs S2 and S3). Hence, here we present analyses for variants that were called by aligning reads to the *Cassia crossbill* draft reference genome.

For phylogenetic analyses, we used *pyRAD* to separately genotype 203,963 SNPs across 29,094 contigs in the aligned data representing common crossbills, *pytopsittacus*, and *L. leucoptera* (included as an outgroup). As expected, *L. leucoptera* formed a strongly supported monophyletic outgroup in the SVDSQUARTETS phylogeny. In addition, *balearica* and *poliogyna* each formed strongly supported monophyletic groups (Fig. 3). *Hispana* individuals largely clustered together in the tree (with a few exceptions), but bootstrap support was weak. Nonetheless, patterns in the phylogeny are consistent with the resident *P. halepenis*-associated *curvirostra* subspecies representing more recently derived and independently evolving groups. The six *pytopsittacus* were not monophyletically clustered, but instead were scattered among *curvirostra* from the Pyrenees and Scotland (Fig. 3). Similarly, *curvirostra* (sampled in the Pyrenees and Scotland) and the different vocal types from these populations were scattered throughout the tree (Fig. 3), with no tendency to cluster monophyletically.

Estimates of genetic differentiation based on population allele frequencies revealed low-to-moderate levels of genetic differentiation among populations (Table 3; mean $F_{ST} = 0.055$, range 0.017 – 0.101; mean Nei’s D = 0.018, range 0.004 – 0.037). Nonetheless, population genetic analyses using PCA of genotype covariance matrices and the Bayesian ancestry-based clustering analysis (*ENTROPY*) revealed clear patterns of genetic structure for the three subspecies associated with *P. halepenis*, consistent with phylogenetic analyses. The first two principal components from the PCA of the genotypic covariance matrix explained 36.0% and 16.5% of the total variation, respectively, and separated *balearica* and *poliogyna* into distinct clusters (Fig. 4a). PC3 and PC4 explained smaller proportions of variance, but clearly differentiated *hispana* from *curvirostra* and *pytopsittacus* (Fig. 4b).

Although the *ENTROPY* model with $k = 3$ had the lowest DIC estimate across all five replicate iterations (Fig. S1), models with ancestry proportions of $k = 2$ and 4 also produced patterns of clustering reflecting genetic differentiation of the *P. halepenis*-associated subspecies. The $k = 2$ model assigned all *poliogyna* to one cluster, and all other individuals to another (Fig. 5a). The $k = 3$ model assigned all *poliogyna* to one cluster, all *balearica* to a second cluster and all other individuals to a third (Fig. 5b). The $k = 4$ model further distinguished the ancestry proportions of *hispana* from other continental European crossbills (Fig. 5c). As might be expected given residency, geographic isolation, and presumably smaller population sizes, expected heterozygosity estimates were lower for *poliogyna* and *balearica* than for nomadic *curvirostra* (Table 2). *Hispana* had heterozygosity estimates similar to those of nomadic *curvirostra*, which could be due to a variety of factors, including gene flow from nomadic forms, recent colonization or larger population sizes.

*ENTROPY* analyses and the full PCA including all individuals indicated no evidence of genetic differentiation between *curvirostra* and *pytopsittacus*, among *curvirostra* from Scotland and the Pyrenees, or among vocal types of *curvirostra* (Figs 4a and b, 5). However, the PCA restricted to *curvirostra* and *pytopsittacus* revealed statistically significant (*PERMANOVA: $F_{1, 49} = 1.19, P = 0.046$; Fig. 4c) albeit very minor genetic differentiation between these recognized species (Table 3; average genome-wide $F_{ST} = 0.034$) that was not evident in other analyses. In addition, the distribution of locus-specific $F_{ST}$ estimates revealed few loci with exceptional differentiation. Ninety-four per cent of the locus-specific $F_{ST}$ estimates were less than 0.1, the maximum was only 0.39, and only 152 loci had $F_{ST}$ greater than 0.2. Plots of $F_{ST}$ estimates by their positions in scaffolds of the genome assembly reveal that the weakly differentiated loci (those with $F_{ST} > 0.2$) were evenly distributed across scaffolds rather than being colocalized.

The PCA restricted to nomadic *curvirostra* from Scotland and the Pyrenees revealed no evidence for genetic differentiation among them (*PERMANOVA, $F_{1, 43} = 0.95$, $P = 0.51$; Fig. 4d). In addition, there was no evidence of differentiation among the two vocal types in Scotland (*PERMANOVA, $F_{1, 8} = 0.98$, $P = 0.58$) or between the two vocal types from the Pyrenees (*PERMANOVA, $F_{1, 9} = 0.84$, $P = 0.82$) (Fig. 4d). Across all population samples, there was no relationship between geographic and genetic distances (Fig. 6). This result is consistent with the nomadic movements of most crossbills and a lack of genetic differentiation among nomadic *curvirostra* and *pytopsittacus* despite large distances between sampling locations.

**Discussion**

Our results support the hypothesis that variation in resource stability has, to varying degrees, influenced patterns of population divergence in western Palearctic crossbills. The three subspecies of common crossbills
associated with *P. halepensis*, a serotinous conifer and stable food resource, are the most genetically distinct, a result also found for the Cassia crossbill in North America (Parchman *et al.*, 2016). In contrast, population samples of other common crossbills and of *pytyopsittacus*, as well as the *curvirostra* vocal types, exhibit little to no evidence of genetic differentiation from one another. All of these crossbills occur in regions with less stable food resources, and most of them are nomadic. The absence of a relationship between geographic and genetic distance among population samples (Fig. 6) is presumably the result of a lack of genetic differentiation among geographically distant samples of nomadic *curvirostra* and *pytyopsittacus* as well as elevated divergence of the geographically proximate *P. halepensis*-associated and resident subspecies. Below, we consider how variation in resource stability could have contributed to these contrasting patterns of genetic divergence.

**Multiple consequences of resource stability for reductions in gene flow**

Several factors related to seed crop stability have likely influenced genetic divergence of *P. halepensis*-associated crossbills. First, temporally stable food resources allow residency and population persistence, thereby increasing the opportunity for geographically isolated populations to establish themselves and become locally adapted. Geographically isolated, resident crossbill populations utilizing conifers with strong temporal fluctuation in seed production do not exist, presumably because such populations would rapidly go extinct during years with little or no seed production. Such a pattern also appears evident for hummingbirds where smaller ranges are correlated with increasing climatic stability, which may reflect more stable food resources (Sonne *et al.*, 2016; see also Sandel *et al.*, 2011). Geographic isolation, which is typically critical for genetic divergence in birds (Price, 2008) and other organisms (Coyne & Orr, 2004), has undoubtedly contributed to genetic divergence for *balearica* on Mallorca and *polioygyna* in northern Africa. In both regions, *P. halepensis* is the only conifer commonly used by crossbills (see Critchfield & Little, 1966), and immigration of nomadic *curvirostra* to these regions is uncommon or unlikely (Cramp & Perrins, 1994; Thévenot *et al.*, 2003; Summers & Jardine, 2005). However, the opportunity for immigration by nomadic *curvirostra* into the range of *hispana* is substantial given the widespread occurrence of common crossbills in eastern Spain (Borras & Senar, 2003) and the nearby occurrence of alternative nonserotinous conifers (Fig. 2; Edelaar *et al.*, 2012) commonly

---

**Fig. 3** A quartet-based phylogeny generated with SVDQUARTETS. Bootstrap values > 95 are shown. The three subspecies of common crossbills associated with *Pinus halepensis* are designated by their subspecific epithets (*balearica*, *polioygyna* and *hispana*); monophyly for *balearica* and *polioygyna* was strongly evident with 100% bootstrap support for each. *L. curvirostra* samples from Scotland and the Pyrenees are designated by their location, and the numbers or letters after the location name indicate the vocal type (1, 2, 1A and 4E). Letters after parrot crossbills (*L. pytyopsittacus*) indicate sample location (N: Norway, R: Russia and S: Sweden).
utilized by nomadic curvirostra (Castro et al., 1999; Summers et al., 2007; Benkman & Parchman, 2009, 2013). This could explain the lower levels of genetic divergence of hispana compared to balearica and poliogyna (Figs 3, 4, 9). However, like the sedentary North American Cassia crossbill that is specialized on a serotinous pine and genetically distinct in spite of regular movements into its range by two other ecotypes (Smith & Benkman, 2007; Benkman, 2017), the genetic differentiation of hispana (Fig. 4b) suggests that factors other than geographic isolation, such as those discussed below, can contribute to population divergence in crossbills.

Unlike pairs of nomadic ecotypes, dispersal between nomadic and resident crossbills is asymmetric with nomadic crossbills dispersing into the habitat of resident crossbills but not vice versa. This reduction in dispersal should increase geographic isolation and act to reduce gene flow (Balkau & Feldman, 1973; Price, 2008; Eroukhmanoff et al., 2011). In addition, serotinous conifers cannot escape pre-dispersal seed predators in space and time, and often have evolved elevated physical defences (see Janzen, 1971; Hulme & Benkman, 2002), which can enhance divergent selection between crossbill ecotypes exploiting alternative conifers (e.g. Benkman, 2003; Benkman et al., 2003, 2010). In several cases, including the Cassia crossbill in North America, divergent selection has been enhanced by coevolutionary arms races between resident crossbills and their local conifer (Benkman et al., 2001, 2003, 2010; Parchman & Benkman, 2002; Mezquida & Benkman, 2005). Balearica, poliogyna and hispana all differ in bill structure from curvirostra occurring farther north (Fig. 1; Griscom, 1937; Massa, 1987; Edelaar et al., 2012; C. W. Benkman, P. Edelaar, & E. T. Mezquida, unpubl. data), potentially consistent with local adaptation and with each of these populations experiencing divergent selection from nominate curvirostra (see Benkman, 2003). Such divergent selection should further reduce gene flow (Nosil, 2012; Shafer & Wolf, 2013). Finally, the consistent availability of seed resources associated with cone serotiny allows crossbills to more regularly deplete a greater proportion of the available seeds (i.e. to be in equilibrium with the resource; Benkman et al., 2010, 2012). This prevents immigrant and less-adapted crossbills from coexisting on the same conifers because of greater exploitative competition (Benkman, 2017), which could further reduce gene flow (Smith & Benkman, 2007; Benkman, 2017). In contrast, cones of the non-serotinous pines generally open and begin shedding seeds in winter or spring resulting in a large pulse of accessible seeds and allowing a wide range of crossbill phenotypes to breed sympatrically. This annual pulse of increased seed availability could account for the absence of genetic differentiation of resident crossbills in the Pyrenees, while the high frequency of serotiny in southern Spain (Hernández-Serrano et al., 2014) has likely influenced population genetic divergence of hispana.

In sum, through its potential effects on residency, population persistence, and geographic isolation, as well as enhanced divergent selection, local adaptation, and greater exploitative competition, resource stability likely contributed to population genetic divergence of P. halepensis-associated crossbills. Phylogenetic analyses also indicate that balearica and poliogyna are more recently diverged and independently evolving monophyletic lineages (Fig. 3), consistent with reduced genetic diversity observed in these subspecies (Table 2). As genetic and palaeoecological studies suggest that P. halepensis expanded into the western Mediterranean only 10-14 kya (Pons, 1992; Gómez et al., 2005; Grivet et al., 2009), divergence of these resident populations likely occurred quite recently. Our results along with those for the Cassia crossbill (Parchman et al., 2016), also resident and adapted to a serotinous pine, suggest that the stable resources produced by serotinous conifers may be conducive to population divergence. Such a relationship between resource stability and population divergence is likely to exist for other taxa experiencing resource-based divergent selection as well.

**Phenotypic divergence without genome-wide divergence in the absence of stable resources**

As expected in view of their morphological differentiation and status as distinct species, by employing far more comprehensive genomic sampling than past studies (Piertney et al., 2001; Björklund et al., 2013), we managed to detect some evidence of genetic
differentiation between *curvirostra* and *pytyopsittacus* (Fig. 4c). However, the divergence we detected was very slight, as indicated by a genome-wide $F_{ST}$ of only 0.034 (Table 3) and a lack of distinctiveness in PCA and Bayesian ancestry-based analyses (Figs 4 and 5). Adaptive differences in morphology between *curvirostra* (specialized on Norway spruce, *Picea abies*) and *pytyopsittacus* (specialized on Scots pine) are nevertheless clear (Fig. 1; Lack, 1944; Newton, 1972; Nethersole-Thompson, 1975; Summers *et al.*, 2010). Indeed, the bill of *pytyopsittacus* averages 2.6 mm (about 25%) deeper than that of *curvirostra* (Summers *et al.*, 2002), which is an

---

**Fig. 4** Genotypic variation (based on 18,047 SNPs) among individuals summarized by principal components from PCAs on genotype covariance matrices. Panel a includes the three subspecies of common crossbills (*L. curvirostra*) associated with *P. halepensis*, the two geographic samples of nominate *curvirostra* and parrot crossbills (*L. pytyopsittacus*). Panel b is based on the same PCA as in panel a, but shows PC3–PC4 instead of PC1–PC2, and excludes *balearica* and *poligyna* because they are so distinct from the other individuals in PC1–PC2. Panel c displays PC1–PC2 for a PCA conducted on nominate *curvirostra* and *L. pytyopsittacus*. Panel d displays PC1–PC2 for a PCA conducted on only the nominate *curvirostra*, with different vocal types identified by different colours and symbols. Panel c does not distinguish different vocal types because of their lack of differentiation in panel d.
order of magnitude greater than the 0.2–0.3 mm difference between some ecologically and genetically differentiated North American ecotypes (Groth, 1993; Benkman et al., 2009; Irwin, 2010; Parchman et al., 2016). Thus, the very limited genetic differentiation between these two morphologically and vocally divergent species (Summers et al., 2002) remains surprising. Our results suggest that ongoing gene flow has limited overall genome-wide divergence, while substantial phenotypic divergence between *curvirostra* and *pytyopsittacus* likely originated and is maintained by divergent selection due to specialization on different conifer species. The limited genetic differentiation between these formally recognized crossbill species is similar to a growing number of examples where even genome-level data reveal that phenotypically differentiated bird populations or even species can sometimes exhibit little genome-wide differentiation (Ellegren et al., 2012; Farrington et al., 2014; Poelstra et al., 2014; Lamichhaney et al., 2015).

The distribution of locus-specific $F_{ST}$ estimates suggests that our data detect few if any genomic regions with elevated differentiation between these two species. Such highly differentiated regions likely exist given substantial bill and body size divergence between *curvirostra* and *pytyopsittacus*, but are difficult to detect with the relatively low marker density of our data (one SNP per 72 kb).

We found no evidence of genetic differentiation between vocal types within *curvirostra* (Figs 3 and 4d). This apparent lack of genetic divergence outside of the *P. halepensis*-associated common crossbills contrasts with recent results for some of the sympatric vocal types (ecotypes) in North America (Parchman et al., 2016). Why genetic differentiation should be more limited in Europe than in North America is unclear. Without studies to distinguish whether vocal variation corresponds to morphological variation (e.g. Groth, 1993) and to specialization on alternative conifers (e.g. Benkman, 1993, 2003), we cannot rule out the possibility

---

**Fig. 5** Ancestry proportion estimates ($q$) from the hierarchical Bayesian model implemented in *ENTROPY*. Each vertical bar represents an individual, and colours correspond to the admixture proportions for each of $k$ clusters. Results with $k$ equal to 2, 3 and 4 are shown (those with the best fit to the data, see Fig. S1). Vertical black bars indicate boundaries between subspecies of common crossbills (*L. curvirostra*) associated with *P. halepensis*, geographic samples from nominate *curvirostra* and parrot crossbills (*L. pytyopsittacus*).
that some vocal types represent nonadaptive vocal variants that evolved by mechanisms such as cultural drift during past geographic/demographic isolation (Galis & van Alphen, 2000). Differences in the extent of genetic divergence within European and North American crossbills are unlikely to be the result of differences in when suitable resources for ecological specialization became historically available. Divergence in North America appears to have occurred mostly in the Holocene (Parchman et al., 2016), consistent with estimates of crossbill divergence in Europe (Björklund et al., 2013). Moreover, *P. halepensis* appears to be a recent colonist to the western Mediterranean (post–last glacial maximum; Gómez et al., 2005; Grivet et al., 2009) where the three genetically distinct *P. halepensis*-associated crossbills reside (Fig. 1), suggesting that population genetic divergence in crossbills can occur rapidly given the right conditions.

Conclusions

Our results illustrate clear patterns of genetic structure characterizing the *P. halepensis*-associated crossbill subspecies and suggest that divergence may have been mediated, to varying degrees, by resource stability and geographic isolation. Geographic isolation can account for the genetic distinctiveness of two of these subspecies (*balearica* and *poliogynus*), although the seasonal and annual stability of serotinous seed crops is surely critical for their persistence as isolated populations. The genetic distinctiveness of *hispana* in Spain, however, cannot be accounted for by geographic isolation alone. Instead, a stable food resource has potentially contributed to its divergence by reducing dispersal and because resource competition could limit sympatric breeding, as found in the Cassia crossbill (Smith & Benkman, 2007; Benkman, 2017). Although this requires direct study, along with more thorough sampling of crossbills in Spain to reveal the distribution and phenotypic characteristics of *hispana* compared to other common crossbills. In contrast to the *P. halepensis*-associated *curvirostra* subspecies, genetic differentiation is very slight for a morphologically divergent and formally recognized species (*L. pytyopsittacus*), and differentiation among different vocal types and populations of *curvirostra* sampled from geographically distant regions appears absent. All of these groups are associated with nonserotinous conifers that generally fluctuate greatly in seed availability to crossbills within and usually across years (Table 1; Hagner, 1965). This limited genetic differentiation indicates that divergent selection alone is insufficient to cause population genetic divergence, and is consistent with previous studies indicating that resource fluctuations could impede or even reverse population divergence and ecological speciation (Grant & Grant, 2008; Price, 2008; Nosil et al., 2009). Further studies of the possible connections between resource stability and divergence are warranted, as resource stability can contribute to residency, population persistence, local adaptation, competitive exclusion, range size and geographic isolation, all of which can affect population divergence and speciation.

Acknowledgments

This research was funded by the Robert B. Berry Chair to C.W.B. and UNR start-up funds to T.L.P. G. Zhang provided valuable assistance with the assembly of the Cassia crossbill reference genome. We thank M. E. Prieto for drawing the cones and head profiles of crossbills in Fig. 1, and Natalia Iovchenko (Ladoga Ornithological Station [Leningrad]), and curators of the Natural History Museum of Oslo (Norway) and Natural History Museum of Denmark for tissue samples. A. van den Berg provided location information for capturing crossbills. We thank C. Porter for comments on the manuscript, and J. Garcia for help with DNA extractions. Finally, we thank N. Bailey and three anonymous reviewers for comments that improved the manuscript.

Author contributions

T.L.P. and C.W.B. designed research; C.W.B., P.E., R.S., D.A. and E.T.M. collected field samples for DNA extraction; T.L.P. generated and analysed genetic data; J.P.J. and K.U. contributed to analyses; T.L.P. and C.W.B. wrote paper with contributions from P.E., J.P.J., K.U., E.T.M. and R.S.
Data accessibility
Fastq files containing filtered DNA sequences for each individual, a genotype probability matrix, and the Cassia crossbill preliminary draft genome assembly are available at the Dryad digital repository: https://doi.org/10.5061/dryad.hf492bp.

References


Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Summary characteristics of Illumina sequencing data used for the Loxia sinescicirius draft genome assembly, after filtering.

Table S2 Summary statistics for Loxia sinescicirius draft genome assembly.

Figure S1 Deviance Information Criterion estimates for entropy runs across models for k = 2–9 for five independent iterations of each.
**Figure S2** Genotypic variation (based on 17,711 SNPs) among individuals summarized by principal components from PCAs on genotype covariance matrices for genotype probabilities generated by aligning reads to the *de novo* assembly of GBS reads (rather than alignment to the *L. sinesciuris* reference genome).

**Figure S3** Admixture proportion estimates ($q$) from the hierarchical Bayesian model implemented in ENTROPY based on genotype likelihoods generated by aligning reads to the *de novo* assembly of GBS reads (rather than alignment to the *L. sinesciuris* reference genome).

Data deposited at Dryad: https://doi.org/10.5061/dryad.hf492bp.

Received 6 February 2018; revised 25 July 2018; accepted 11 August 2018