bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Neo-sex chromosomes and demography shape genetic diversity in the Critically Endangered Raso lark

Elisa Dierickx[1,2], Simon Sin[3,5], Pieter van Veelen[6,7], Michael de L. Brooke[1], Yang Liu[4], Scott Edwards[3] and Simon Martin[1,8]

[1] Department of Zoology, University of Cambridge
[2] Fauna & Flora International
[3] Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University
[4] Department of Ecology, Sun Yat-sen University
[5] School of Biological Sciences, University of Hong Kong
[6] Wetsus, European Centre of Excellence for Sustainable Water Technology
[7] Groningen Institute for Evolutionary Life Sciences, University of Groningen
[8] Institute of Evolutionary Biology, University of Edinburgh

KEYWORDS

Skylark genome, genetic diversity, Cape Verde, conservation, Alauda, island endemic

ABSTRACT

Generally small effective population sizes expose island species to inbreeding and loss of genetic variation. The Raso lark has been restricted to a single islet for ~500 years, with a population size of a few hundred. To investigate the factors shaping genetic diversity in the species, we assembled a reference genome for the related Eurasian skylark and then assessed genomic diversity and demographic history using RAD-seq data (26 Raso lark samples and 52 samples from its two most closely related mainland species). Genetic diversity in the Raso lark is lower than in its mainland relatives, but is nonetheless considerably higher than anticipated given its recent population size. We found that suppressed recombination on large neo-sex chromosomes maintains divergent alleles across 13% of the genome in females, leading to a two-fold increase in overall diversity in the population. Moreover, we infer that the population contracted from a much larger size recently enough, relative to the long generation time of the Raso lark, that much of the pre-existing genetic variation persists. Nevertheless, the current small population size is likely to lead to considerable inbreeding. Overall, our findings allow for optimism about the ongoing reintroduction of Raso larks to a nearby island, but also highlight the urgency of this effort.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Island species have suffered 89% of all recorded avian extinctions, despite only representing 20% of all bird species (1-4). Underlying this is island species' vulnerability to alien invasive species (3) and to threats linked to the intrinsic geographical characteristics of islands such as isolation and small distributional area (e.g. 5-8). Stochastic environmental events, habitat destruction or resource depletion can also harm island species more than continental counterparts, since the former may be unable to disperse into alternative habitat. Species with a small effective population size (N_e) are also subject to three types of genetic risk: inbreeding depression through the exposure of deleterious recessive alleles and loss of heterozygote advantage (9-11); accumulation of deleterious alleles due to increased drift ("mutational meltdown") (9,12,13); and loss of potentially adaptive genetic variation limiting future adaptive potential (5,14-16). Previous studies have found that island species, particularly island endemics, often show reduced genetic diversity and increased inbreeding compared to their mainland counterparts, consistent with a reduction in N_e (5,17-18). However, the initial bottleneck associated with island colonisation may be more to blame than long-term reductions in N_e for island populations (18). For most species, poor census data makes it difficult to assess whether island existence *per se* is likely to expose species to the genetic risks above.

The Raso lark *Alauda razae* is endemic to the uninhabited 7km^2 -islet of Raso in Cape Verde (19). Irregular counts since1965 and yearly counts since 2002 indicate a small population fluctuating from 20 breeding pairs to 1558 individuals (20-22) (Table S1). Raso lark sub-fossils indicate a larger past distribution encompassing neighbouring islands: Santa Luzia (35km^2), São Vicente (227km^2) and Santo Antão (779km^2) (19; Figure 1). Raso larks disappeared abruptly from the neighbouring islands following the arrival of humans along with cats, dogs and rodents in 1462 (19). Today, Raso is the last of the larger islets of Cape Verde that remains free of these mammals.

Previous work suggests that a change to genome architecture might buffer Raso larks from genetic diversity loss. The genus *Alauda* has enlarged neo-sex chromosomes which appear to derive from ancestral autosomes (23). Cessation of recombination on neo-sex chromosomes could represent a source of heterozygosity, because females (the heterogametic sex) could retain distinct alleles at homologous loci on their neo-Z and neo-W chromosomes. Indeed, Brooke et al. (24) observed a microsatellite locus with sex-linked segregation and excessive heterozygosity in female Raso larks, and hypothesised that the neo-sex chromosomes could allow the retention of diversity in the face of population contraction. This could impact the genetic diversity, and therefore

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the survival, of Raso larks, particularly if the neo-W and neo-Z represent a large proportion of the genome (as suggested by cytogenetic analysis;25) and if recombination is suppressed across large tracts of these chromosomes, allowing retention of distinct alleles in females. These conditions have hitherto not been directly tested with genome-wide data.

To investigate the impact of the population contraction on genetic diversity in the Raso lark, and the potential role of the neo-sex chromosomes in buffering this impact, we produced a high-quality draft genome assembly for the related Eurasian skylark *Alauda arvensis* and used restriction site associated DNA sequencing (RAD-seq) of 78 individuals from four lark species: Raso lark, Eurasian skylark, Oriental skylark *A. gulgula* (the third currently-recognised *Alauda* species) and crested lark *Galerida cristata*. Our findings reveal effects of both demographic change and neo-sex chromosomes in shaping genetic diversity, and provides insights relevant for Raso lark conservation.

METHODS

Sample collection

Twenty-six Raso lark blood samples were collected on Raso between 2002 and 2014. From colleagues, we also obtained blood and tissue samples for related species: 36 Eurasian skylarks, that we group into three populations, 15 Oriental skylarks from two locations, and 5 crested larks from Saudi Arabia (Figure 1, Table S2). None of the sampled birds was likely to be a migrant based on the sampling dates (Table S2) and/or the migration pattern of the species (26,27). For samples with unknown sex, sex was determined using PCR (28) and/or by examination of heterozygosity on the Z-linked scaffolds (see Results).

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

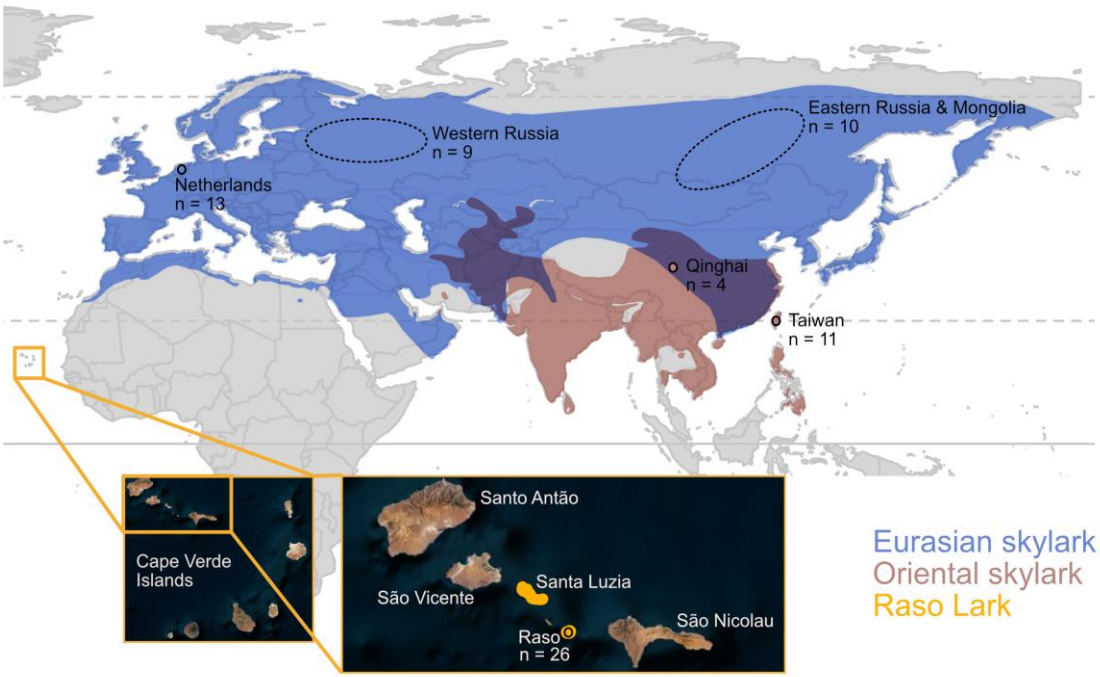

Figure 1. Species ranges, sampling locations and samples sizes

Whole genome sequencing and assembly

A draft reference genome was obtained through the whole-genome sequencing of a male Eurasian skylark sample (individual 0) collected in Mongolia (Table S2). Whole genomic DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen, Venlo, The Netherlands) following the manufacturer's protocol. Two libraries were prepared: a 220bp insert size fragment library using a PrepX ILM 32i DNA Library Kit for an Apollo 324 robot, following the manufacturer's protocol (TaKaRa, Kusatsu, Japan), and a 3 kb mate-pair library using an Illumina Nextera Mate Pair Sample Preparation kit and following the manufacturer's protocol (Illumina, San Diego, CA, USA). Both libraries were sequenced on an Illumina HiSeq 2500 at the Bauer Core,Harvard, producing 125bp paired-end reads.

The Eurasian skylark genome was assembled following the methods outlined in Gnerre et al. (29) (pipeline: https://github.com/simonhmartin/Raso_lark_diversity). Trimmomatic 0.32 (30) was used for adaptor trimming; FastQC (31) to check read quality; and Allpaths-LG (29) to assemble the genome. Assembly summary statistics were calculated with Allpaths-LG.

RAD library preparation

DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen, Venlo, the Netherlands), following the manufacturer's protocol. Single digest RAD-seq libraries for each individual (except individual 0) were prepared according to the protocol of Merrill

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

et al. (32) using the enzyme *PstI*. Each individual was assigned an 8-base pair (bp) inline barcode, and equimolar concentrations of 16 uniquely barcoded individuals were pooled and double-indexed by 16 cycles of high-fidelity PCR using Phusion High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) with Illumina barcodes. The PCR products were pooled in equimolar quantities and sequenced on an Illumina HiSeq 1500 at the Gurdon Institute, University of Cambridge, producing 100bp single-end reads.

Sequence processing and alignment

We used *process_radtags* in Stacks 1.35 (33) without quality filters to sort sequence reads by barcode. We then used Trimmomatic to trim restriction sites (6bp) and remove all trimmed reads shorter than 95bp. *Process_radtags* was then used again to filter for quality. Reads were aligned to the Eurasian skylark genome using Bowtie 2 (34). Reads with multiple significant matches to the reference genome were removed.

Allele frequency spectra and genetic diversity

We used a two-step pipeline in ANGSD (35) to infer allele frequency spectra from the RAD-seq reads mapped to the reference genome, accounting for uncertainty in low-coverage sequencing data. Genotype likelihoods were inferred using ANGSD with likelihood method 2, and only sites with a base alignment quality (baq) ≥1 and SNP quality ≥20 were considered. A mapping quality adjustment of 50 was applied. The *realSFS* tool was then used to infer the allele frequency spectrum with a maximum of 100 iterations, with 20 bootstraps. Nucleotide diversity (π) was computed from the frequency spectrum as the sum of the weighted products of the major and minor allele counts for each allele count category, including the zero category (invariant sites).

Pseudo-chromosomal assembly

To explore how diversity varies across the genome, chromosome positional information is required. We inferred the approximate chromosomal location and orientation of each Eurasian skylark scaffold based on homology with the zebra finch *Taeniopygia guttata* genome (i.e. a pseudo-chromosomal assembly). We used the *nucmer* tool in MUMmer v3.23 (36) to identify regions of strong homology between the two species. Only scaffolds larger than 1 Mb were considered and alignments shorter than 5kb were discarded using the *delta-filter* tool. We used the *mummerplot* tool to visualise all alignments and determine the optimal scaffold order and orientation. Additional manual changes were then made based on visual inspection of the scaffold arrangement. In total, 311 scaffolds, totalling 648 Mb (63% of the genome) were placed on chromosomes.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Proportion of heterozygous sites across the genome

To visualise how the proportion of heterozygous sites varies across the genome in each individual, we called genotypes using SAMtools mpileup version 1.2.1 and BCFtools call version 1.2.1 (37), with default parameters. We considered only genotypes with ≥5x read depth, extracted using BCFtools filter. Thirteen individuals with poor coverage (< 3 million of sites in the genome with ≥5x coverage) were excluded. Individual heterozygosity was computed for 100 kb windows across each scaffold using the Python script popgenWindows.py (github.com/simonhmartin/genomics_general). Windows with fewer than 100 sites (with ≥5x coverage) genotyped across the dataset were excluded.

Demographic inference

We applied two different approaches to investigate historical demographic changes in the Raso lark based on the frequency spectrum (averaged across 20 bootstrap replicates). First we used δaδi (38) to compare four different models of increasing complexity. The first model imposes a constant population size. Since δaδi only optimises the shape of the frequency spectrum, this model has no free parameters. The second model adds a single change in population size at some point in the past (two free parameters: time and relative size of the new population). The third and fourth models each added an additional change (along with two free parameters). Model optimisation was performed using grid sizes of 50, 60 and 70, and repeated 10 to 50 times to confirm optimisation.

In the second approach we used Stairway Plot (39) to infer the optimal population size history given the SFS. We used the “two-epoch” model, with the recommended 67% of sites for training and 200 bootstraps. We tested four different numbers of random breakpoints: 12, 25, 37 and 50.

To convert inferred population sizes and times to numbers of individuals and years, respectively, we used the collared flycatcher *Ficedula albicollis* mutation rate estimate of 4.6×10^{-9} per site per generation (40). We estimated the generation time of the Raso lark, defined as the mean age of the parents of the current cohort at age at first breeding + (1/mean annual mortality) (41). This gave a generation time of 6.5 years.

Relatedness

We estimated relatedness among individuals using two methods suited to low coverage genomic data. The first was NgsRelate (42), which considers genotype

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

likelihoods. These were computed for each individual using the GATK version 3.4 (43,44) HaplotypeCaller tool in GVCF mode with default parameters. NgsRelate was run on a filtered VCF file that included sites covered by at least one read in each individual in the population. The second approach was KGD (45), which is designed for GBS data such as RAD-seq data, and also accounts for low sequencing depth. The input file was generated using ANGSD (35) with the option -dumpcounts 4 to give read counts of each base for each individual at each site. Only sites covered by at least 100 reads across the 26 Raso larks or 50 reads across the 13 Eurasian skylarks from the Netherlands were included. Following Dodds et al. (45), we explored filtering options to find SNP subsets that gave realistic values of self-relatedness (~1). The chosen filter was to use only SNPs with a Hardy-Weinberg disequilibrium value between 0 and 0.1.

RESULTS

Genome-wide diversity

We generated a high-quality draft genome assembly for the Eurasian skylark, totalling 1.06 Gb with a scaffold N50 length of 1.44 Mb (71.5 Kb for contigs). This was based on 154,342,128 reads in the 220bp library and 263,949,984 reads in the 3 kb library.

RAD-seq reads for 78 individuals were mapped to the reference genome, yielding a high density of RAD loci sequenced to low depth (min = 1.5x, max = 6.9x, mean = 2.8x) (Table S2). Using a threshold of at least 100 reads across the dataset to designate a shared RAD locus, this gives 85 million genotyped sites across the dataset, of which 6 million are SNPs.

Average nucleotide diversity (π) across the 26 Raso larks based on inferred allele frequency spectra is 0.0018. This is 18.6% of that in Eurasian skylark from the Netherlands (0.0097), and nearly half of that in the Oriental skylark from Taiwan (0.0044) (Figure 2). Assuming an equilibrium population with $\Theta = 4N_e\mu$ and a per-generation mutation rate equivalent to that of the collared flycatcher, this translates to an effective population size (N_e) of ~100,000 in the Raso lark, compared to ~500,000 in the Eurasian skylark. Therefore, genetic diversity in the Raso lark, despite being five-fold lower than that in the Eurasian skylark, is consistent with a population size far larger than its current census population size of ~1000 individuals. Our subsequent analyses therefore investigated two factors that could explain the unexpectedly high diversity in Raso larks: (1) large neo-sex chromosomes (23) could elevate diversity in females (24) and (2) the population contraction to its current size could be too recent to have eliminated genetic diversity.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

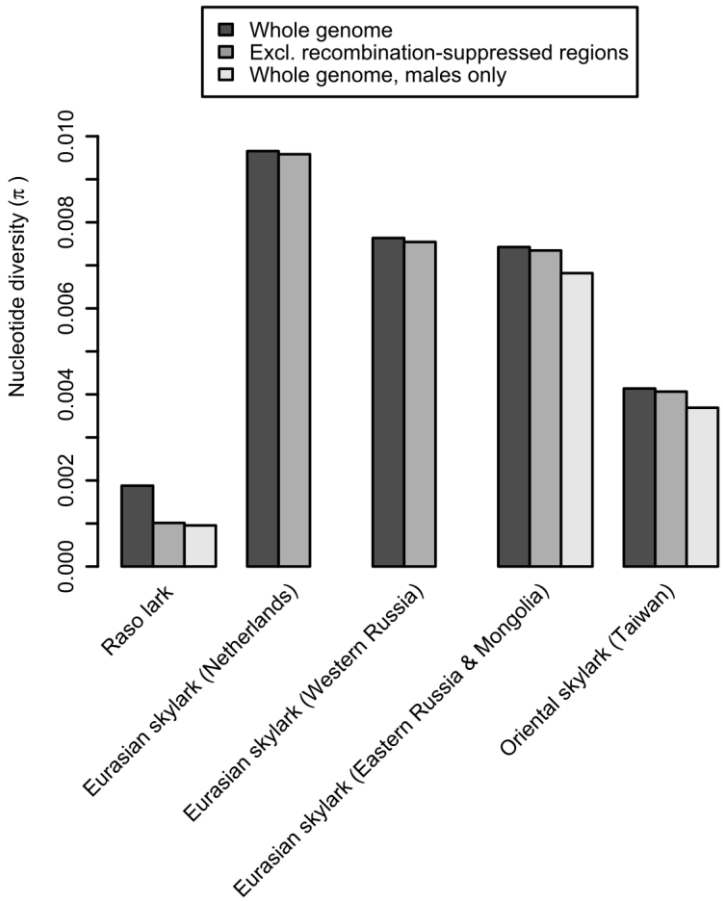

Figure 2. Diversity is reduced in the Raso lark. Nucleotide diversity (π) computed from inferred allele frequency spectra (i) using all genome scaffolds across all individuals in each population, (ii) excluding scaffolds identified as showing suppressed recombination in females (see Figure S1), or (iii) using all scaffolds, but only males. Values are plotted for populations with ≥5 representative individuals. Detailed values are given in Table S3.

Elevated diversity across neo-sex chromosomes

Genetic diversity in Raso larks may be maintained on the large neo-sex chromosomes (24). This would occur if recombination was suppressed between the neo-Z and neo-W, allowing the co-existence of divergent alleles in females. Estimated heterozygosity in each individual based on the RAD-seq reads mapped to the reference genome, partitioned in 100 kb windowsis strongly elevated in female Raso larks across large portions of chromosomes 3, 4a and 5 (Figure 3). By contrast, males show a consistently low proportion of heterozygous sites across the whole genome (Figure 3). Previously, only part of chromosome 4a had been identified as sex-linked in several members of the Sylvioidea clade, including the Eurasian skylark (23). Our findings suggest that large portions of chromosomes 3 and 5 also form part of the neo-sex chromosomes. The elevated heterozygosity in females indicates that recombination is

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

suppressed between parts of the *neo-W* and *neo-Z* . It is important to note that the chromosome map used here - based on the zebra finch genome - does not reflect the true karyotype for larks. For example, it has already been shown that only part of chromosome 4a (called 4a-1) has become sex-linked. The other fragment (4a-2) remains autosomal and therefore its lack of elevated heterozygosity is expected. Without a linkage map, we cannot determine whether similar fragmentation of chromosomes 3 and 5 has occurred, but given the broad extent of recombination suppression, it is likely that these two chromosomes are entirely sex-linked.

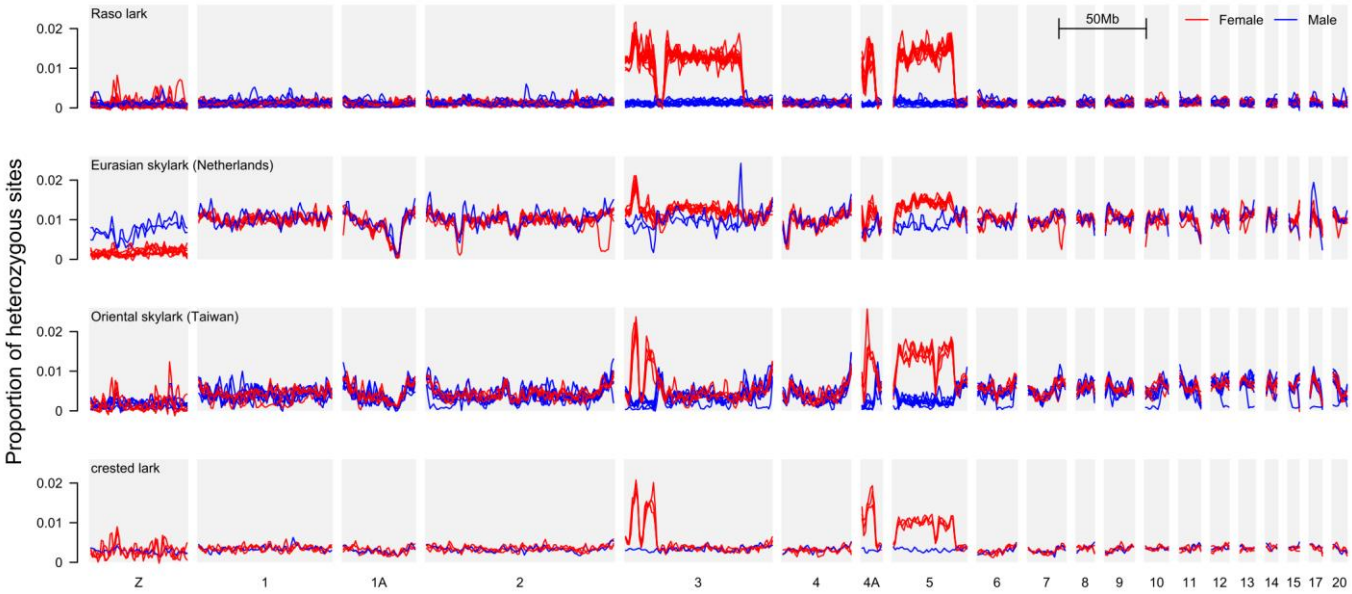

Figure 3. Individual heterozygosity reveals neo-sex chromosomes. The proportion of heterozygous sites per 100 kb window in each individual, plotted across each chromosome (based on homology with the zebra finch), with locally-weighted smoothing (loess, span = 5 Mb). Females and males are indicated by red and blue lines, respectively. One population from each species is shown (see Figure S2 for all populations). Three Eurasian skylark samples from the Netherlands (samples 6, 4 and 28) are excluded from this plot on the basis that they had excessive heterozygosity or homozygosity resembling sequencing errors.

Comparison with the other species gives insights into the progression of recombination suppression through time. A difference between male and female heterozygosity exists across the same three chromosomes in Eurasian skylarks from the Netherlands, but it is far less pronounced (Figure 3). This is due to the much higher genome-wide background heterozygosity in this population, whereas heterozygosity in the regions of suppressed recombination is similar between Eurasian skylark and the Raso lark. The same three chromosomes also show enhanced heterozygosity in female Oriental skylark and crested lark, but over smaller portions of chromosomes 3 and 5. A similar analysis in a recent study of two additional outgroup species, the bearded reedling

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Panurus biarmicus and the horned lark *Eremophila alpestris*, which also carry neo-sex chromosomes, revealed an even more reduced pattern of recombination suppression, that excludes chromosome 5 (46). This allows us to partially reconstruct the timing of recombination suppression, which has occurred in a stepwise manner in "strata," following fusions of chromosomes 4a-1 and later 3 and 5, to the W (Figure 4). The initial suppression occurred in two narrow strata on chromosomes 3 and 4a-1 (46). This increased marginally on chromosome 3 in the common ancestor of the larks, and was followed by the formation of a large new stratum across most of chromosome 5 in the common ancestor of the crested lark and the *Alauda* larks. Further recombination suppression in several parts of chromosome 3 then probably occurred in the ancestor of *Alauda*. However, this was likely polymorphic, because it is lacking in the Oriental skylark, and also in the eastern populations of the Eurasian skylark, but it is present in the western populations of the Eurasian skylark as well as in the Raso lark (Figure S2, Figure 4).

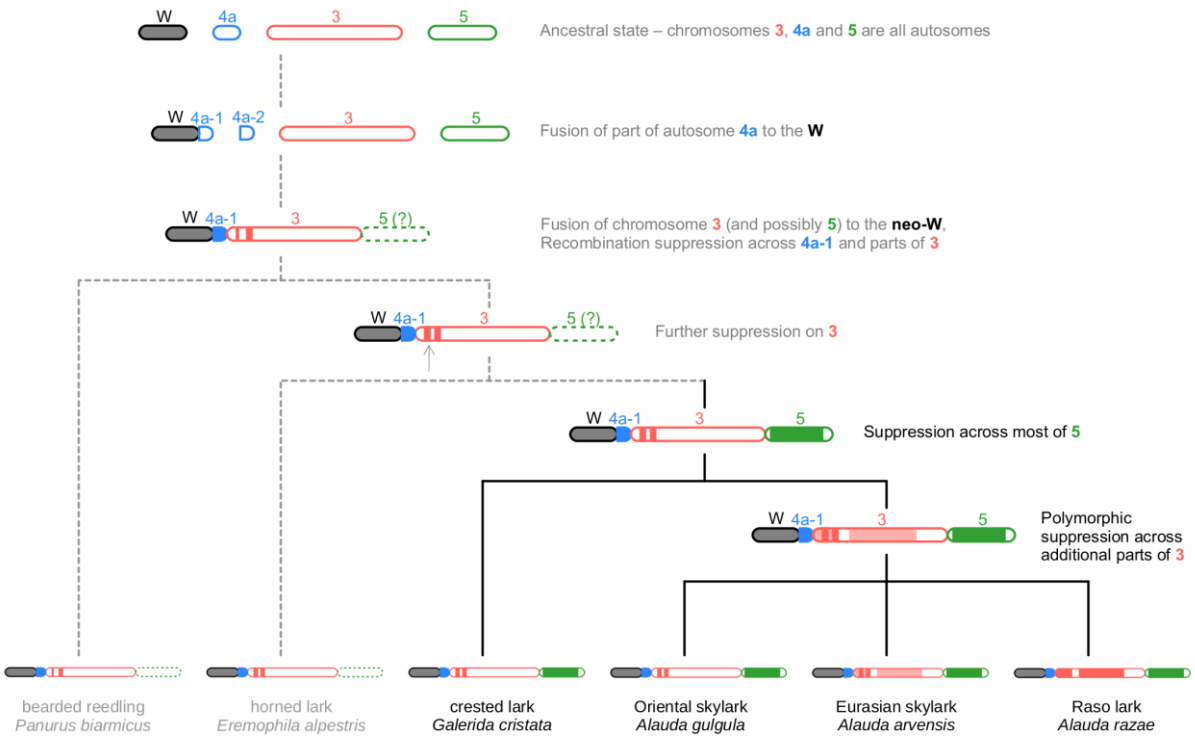

Figure 4. A model for the evolution of neo-sex chromosomes in larks. The most parsimonious ancestral states are indicated based on linkage mapping (23) and patterns of recombination suppression in present-day species (Figure 3 and see 46). Inferences based on other studies (23,50) are shown in grey, and uncertain fusions are shown by dashed lines. Note that the order and orientation of the fused chromosomes 3 and 5 is unknown, so the representation here is one of several possibilities. Regions of suppressed recombination (Figure 3) are indicated by filled boxes. Light shading indicates polymorphic recombination suppression (i.e. in some populations but not others; Figure S2)

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Raso lark diversity remains higher than expected when neo-sex chromosomes are excluded

When the regions showing recombination suppression in females (Figure S1) are excluded, π in the Raso lark is nearly halved (Figure 2, Table S3). A similar reduction is seen when females are simply excluded. Recombination suppression across the neo-sex chromosomes therefore does indeed contribute to the maintenance of genetic variation in the Raso lark. This suppression influences only 12.6% of the genome, but contributes to almost half of the observed diversity in this species. Less pronounced reductions in diversity are seen in the Eurasian skylark and Oriental skylark populations, as expected given their higher genome-wide background diversity.

Nevertheless, even after accounting for the neo-sex-chromosomes, π in the Raso lark is ~0.001, about 10% of that in Eurasian skylarks from the Netherlands. This is still far greater than expected given the difference in current population sizes between the two species. The discrepancy between census population size and diversity implies that our second hypothesis may also be valid: there has been insufficient time since the population contraction for significant loss of genetic diversity in Raso larks.

Genomic signatures of ancient and recent bottlenecks

To investigate more closely whether a recent population collapse has left a detectable signature on Raso lark genetic diversity, we examined the allele frequency spectrum, which carries information about historical population size changes (47,48). Surprisingly, the frequency spectrum (computed after excluding scaffolds showing evidence for suppressed recombination; see Figure S1) is skewed towards an excess of rare variants (Figure S3), which is also captured in the negative Tajima's D value of -0.7. An excess of rare variants is typical of population expansion rather than contraction and is therefore not consistent with our understanding of the recent history of the Raso lark. This skew was consistent across 20 bootstrap replicates and 26 'drop-one-out' replicates, in which a single individual was excluded in each case (Figure S3), implying that it is not a sampling artefact.

We therefore used two related approaches to explore the historical demography of this species based on the frequency spectrum. First, we used δaδi (38) to compare the fit of simple models allowing zero, one, two or three changes in population size in the past. The simplest model imposes a constant population size, and shows a very poor fit to the data, as expected (Figure S4). A model that allows a single change in population size in the past shows a far better fit and a greatly improved composite likelihood (Figure S4). The inferred model involves an ancient population expansion from ~50,000

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

to ~450,000 individuals ~60,000 years ago (Figure S4). Note that the inferred values for population size and timing of these events should not be interpreted as exact historical estimates, as these depend on our imperfect knowledge of generation time and mutation rate. Instead, the broad pattern of population size change over time is most relevant here, as this determines the shape of the allele frequency spectrum. This one-change model is able to recreate the excess of singleton variants in Raso larks, but still shows a fairly poor fit for to the frequency of other rare variants. The model allowing two changes in population size again shows a major increase in likelihood and better fit to the frequency spectrum (Figure S4). Note that the inferred demographic model does not include a recent contraction. Instead, it consists of an ancient bottleneck down to ~6,000 individuals followed by an expansion ~100,000 years ago to ~300,000 individuals. These findings show that the distribution of genetic variation in the Raso lark can be explained very accurately by a few ancient demographic changes, without inclusion of a recent population contraction.

Given the recent census estimates of ~1000 individuals for the Raso lark, we attempted to fit a final model that allows for a recent population contraction. As δaδi failed to optimise a model with a third change in population size (i.e. two additional free parameters), we instead used a fixed population contraction, and performed a manual search to approximate this parameter estimate by re-running the optimization across a range of final N_e values from 0.1% to 1% of the ancestral size. The resulting best fit model once again shows an improved fit to the data over models without a recent contraction, although it results in minimal detectable difference to the expected frequency spectrum (Figure S4). The model again involves an ancestral bottleneck that ended over 100,000 years ago, but with a larger population of over 3 million individuals since then, followed by a recent contraction down to about 350 individuals that occurred just 50 years (around 8 generations) ago (note again that parameter estimates are not to be to be interpreted as exact). Because all events in δaδi are timed relative to the estimated ancestral N_e, we are unable to fit a model in which the contraction is fixed precisely at a given date, such as the predicted 85 generations ago for human settlement of Cape Verde. Nonetheless, we can conclude from these models that the distribution of genetic variation in the Raso lark is consistent with a recent and dramatic population contraction, but one that is so recent as to have had very little impact on the overall distribution of genetic diversity in this species.

Our second approach used Stairway Plot (41) to estimate the optimal population size history given the frequency spectrum. The inferred history is remarkably similar in its general structure to the 3-change model inferred using δaδi, with a strong ancestral

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

bottleneck around 100,000 years ago followed by expansion to a large population of nearly a million individuals, with a sharp recent contraction (Figure 5). The contraction is most pronounced in the most recent past, it is inferred to have initiated further in the past, ~10,000 years ago. Given the δaδi results showing, that any signal of the most recent population contraction in the frequency spectrum is fairly weak, it is likely that the exact timing of this event would be difficult to infer. Nevertheless, both approaches indicate that the Raso lark population was probably very large up until fairly recently, thus agreeing with our second hypothesis that the population still retains much genetic variation that pre-dates its recent contraction.

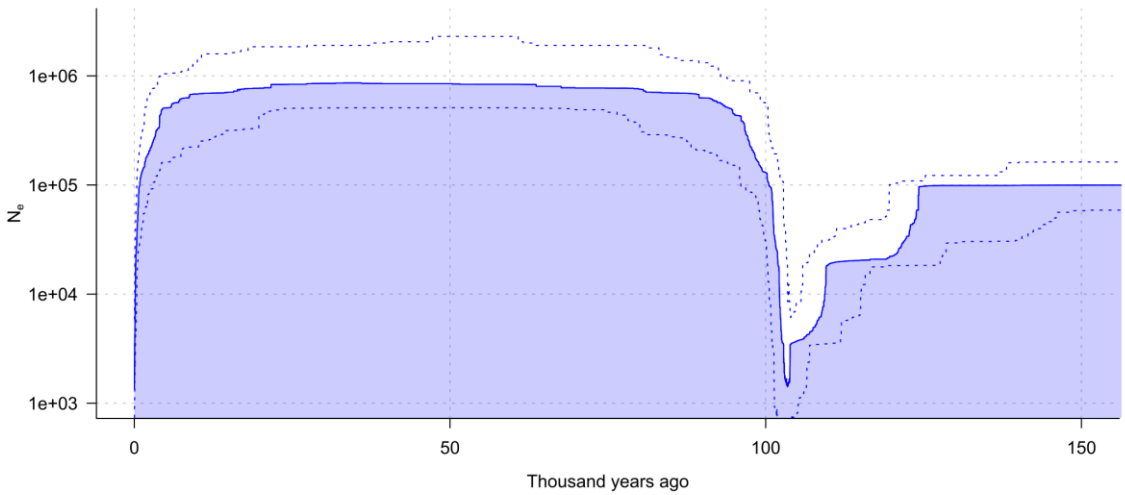

Figure 5. Demographic model fitting suggests an ancient bottleneck. The inferred N_e (log scale) over time (with the present on the left), based on inference using Stairwayplot (41). Dashed lines indicate the 95% confidence interval.

Relatedness

Even if it has had little effect on overall diversity, a recent population collapse may be detectable by an increase in the number of related individuals in the population. Using two different methods, we find that most of the 26 Raso lark samples show little or no detectable relatedness, but that three pairs of individuals show levels of relatedness of ~0.5, indicating either sibling or parent-offspring relationships, while several other pairs showed non-zero relatedness of up to 0.2 (Figure S5). One of the pairs of close relatives consists of individuals sampled three years apart, in 2011 and 2014, making this unlikely to be a sampling artefact (Figure S5). Unfortunately, due to a labelling error of some Raso lark DNA samples, the dates of sampling for the other two related pairs are unknown. The equivalent analyses performed for the 13 Eurasian skylarks from the Netherlands found no consistent evidence of high relatedness (Figure S5). Given our relatively small sample sizes, we cannot draw strong conclusions from these results,

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

but they may indicate that inbreeding is beginning to become a threat to the Raso lark population.

DISCUSSION

Genetic markers have long been used to investigate the genetic risks facing species thought to have small N_e, such as island endemics (5,17,18). Genomic approaches now allow us to address these questions at far greater resolution, revealing how different parts of the genome have been affected, and inferring past demography. We find that the island-endemic Raso lark has reduced genetic diversity compared to its widespread *Alauda* relatives, but that this difference is smaller than expected. The Eurasian skylark is comparable in diversity to other highly diverse birds such as the zebra finch ($\pi \approx 0.01$; 50), and diversity in the Raso lark is only five-fold lower than this. This is inconsistent with the large difference in census population sizes: there are an estimated one million breeding pairs of Eurasian skylarks in the United Kingdom alone (51), and ~40-1500 Raso larks on Raso. Our findings indicate that the unexpectedly high diversity in Raso larks can be explained by the recency of the population contraction from a much larger ancestral size, with an added buffering effect provided by enlarged neo-sex chromosomes that retain excess heterozygosity in females.

Previous work (23) suggested that neo-W and neo-Z chromosomes arose at the base of the *Sylvioidea* 42.2 million years ago through fusion of part of chromosome 4a to both the W and Z. Our results indicate that two other large chromosomes, 3 and 5, have become fused to the W in the genera *Alauda* and *Galerida*. Another recent study (46) supports these observations and further indicates that at least chromosome 3 is also sex-linked in two additional outgroup species (horned lark, bearded reedling). Without a linkage map we cannot determine whether homologs of chromosomes 3 and 5 have also fused to the Z to form a neo-Z (as is known to be the case for chromosome 4a-1; 23). The observation by Bulatova (25) that both W and Z chromosomes are enlarged in larks supports this.

Despite the deep age of the neo-sex chromosomes, our results indicate fairly strong homology between the neo-W and neo-Z in the regions of recombination suppression, with around 1-1.5% divergence. This value may be somewhat underestimated, as the most divergent parts would be excluded due to poor read mapping (46). Nevertheless, even a divergence of 2% would translate to ~2.2 million generations to coalescence, much less than the age of the neo-sex chromosomes. Importantly, this coalescence time reflects not the age of the fusions but rather the age and extent of recombination

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

suppression. Recombination and gene conversion may continue to occur at low levels even between divergent parts of the chromosomes. Moreover, large tracts of all three chromosomes show low levels of heterozygosity consistent with ongoing recombination. Comparisons among species indicate that recombination suppression has progressed further in Raso lark than in Oriental skylark and crested lark. Furthermore, we find evidence for variation in the extent of recombination in Eurasian skylarks, implying that populations can exist for a long time with variable levels of recombination suppression in the sex chromosomes.

It is tempting to speculate that the large extent of recombination suppression seen in Raso larks has been favoured by selection following the loss of genetic diversity. However, the trend in Eurasian skylarks does not support this reasoning, as the greatest extent of suppression is seen in the more diverse western populations. A perhaps more likely explanation, is that recombination suppression has been favoured by sexually antagonistic selection (46, 52). In any case, the resulting maintenance of high heterozygosity across roughly 13% of the genome, which accounts for about half of the genetic diversity in Raso larks overall, may have fitness benefits for females, but these would be difficult to distinguish from other sex differences.

When we exclude the sex chromosomes, levels of diversity in Raso larks are still far higher than their tiny population size would predict. Such a mismatch between differences in genetic diversity and census population size is a well-documented phenomenon (53,54). Research in this field is typically aimed at explaining a lower-than-expected diversity in larger populations, either through a mismatch between census and effective population sizes (55), or through increased action of selection affecting linked sites (56,57). Higher-than-expected diversity in smaller populations requires a different explanation, such as gene flow from other populations, increased mutation rate, or recent contraction from larger ancestral size (18). In the case of the Raso lark, gene flow is unlikely, given that the other two *Alauda* species are largely restricted to Eurasia. We can also confidently rule out an increase in mutation rate in the Raso lark compared to its *Alauda* relatives, because the level of divergence between the neo-Z and neo-W is similar in all three species. A recent reduction from a larger size is therefore the most likely explanation. Assuming this coincided with the settlement of Cape Verde in 1462, the generation time of Raso larks means that this would correspond to just 85 generations ago. Under a simple model of loss of 1/2N times the ancestral variation through drift each generation, with a current N_e of 1000, this translates to a retention of over 95% of the pre-existing genetic diversity (Figure S6). Even in the more extreme case of N_e = 100, > 60% of the diversity is retained.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

While it may be unsurprising that the recent contraction has failed to eliminate genetic diversity, we were surprised by the skew towards rare variants in the allele frequency spectrum, indicating that the genetic make-up of this population is largely shaped by a major population expansion that occurred deeper in the past. Our estimate of just over 100,000 years ago for this expansion following a strong bottleneck may coincide with their colonisation of Cape Verde. Summing the areas of the surrounding islands, Santa Luzia, São Vicente and Santo Antão (19), we can predict a minimum ancestral range of 1,048km^2, which is ~150 times the area of Raso. Extrapolating from the current population of ≈ 1000 predicts an ancestral population size of 150,000. Our modelling estimates a much larger N_e of nearly a million prior to the recent collapse. It is possible that Raso larks were more abundant due to higher population density in the past, or that their colonisation of Cape Verde from a larger mainland population occurred more recently.

Despite the considerable genetic diversity in Raso larks relative to their population size, continued existence at this size will inevitably increase their genetic risks. We found three pairs of closely related Raso larks out of 26 sampled. While we cannot rule out a chance effect, finding related individuals is to be expected given that the population dropped to just 57 individuals in 2004, around two generations ago. With sufficient sample sizes, individual-level relatedness may provide more sensitive detection of recent population collapse than population-level diversity, which can take multiple generations to change appreciably.

Another likely and serious risk for this species is the loss of future adaptive potential (58). Cape Verde is particularly vulnerable to environmental changes such as climate change (59)). Adaptive potential is also relevant for the ongoing reintroduction of Raso larks to Santa Luzia (the first birds were translocated in April 2018). While this translocation project is crucial for the conservation of the species, it could constitute another bottleneck on Santa Luzia if the founder population is too small. The success of previous reintroduction programmes for vertebrates has been highly variable, with inbreeding in small founder populations and the resulting vulnerability to disease often being cited as a cause of failure (60,61). The strong skew towards an excess of rare variants in Raso larks, as a result of their historical bottleneck and subsequent expansion, may make the challenge greater than usual, as by definition most mutations segregating in the population are present in just one or a few individuals and are thus more likely to be missed in a small sample set.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In conclusion, our findings give cause for both optimism and concern for the Raso lark. The appreciable genetic diversity retained means that there is hope for the ongoing reintroduction to other islands, but loss of diversity and inbreeding depression seem inevitable if the population persists at its current size.

ACKNOWLEDGEMENTS

Nick Horrocks and Per Alström shared crested lark and Eurasian skylark samples. Allison Shultz provided feedback over the course of this project. Marco van der Velde and Sarah Barker advised on genetic sexing and on RAD library preparation. Chris Jiggins let us use his laboratory. Thanks to the field assistants on Raso: Mark Bolton, Ewan Campbell, Simon Davies, Mike Finnie, Tom Flower, Lee Gregory, Sabine Hille, Mark Mainwaring, Jason Moss and Justin Welbergen.

FUNDING

We were supported by the Sir Peter Scott Studentship and the Rouse Ball Eddington Fund of Trinity College, Cambridge (to EGD), the VOCATIO Award (to EGD), the William Bateson Fellowship of St John's College, Cambridge (to SHM), Julian Francis, RSPB, CEPF, and BirdLife International's Preventing Extinctions Initiative.

REFERENCES

1. Johnson, TH, and AJ Stattersfield. 1990. "A Global Review of Island Endemic Birds." Ibis 132 (2): 167–80.
2. Manne, LL, TM Brooks, and SL Pimm. 1999. "Relative Risk of Extinction of Passerine Birds on Continents and Islands." Nature 399 (May): 258–61.
3. Butchart, SHM, AJ Stattersfield, and TM Brooks. 2006. "Going or Gone: Defining 'Possibly Extinct' Species to Give a Truer Picture of Recent Extinctions." Bulletin of the British Ornithological Club 126 (A): 7–24.
4. BirdLife International. 2011. "We Have Lost over 150 Bird Species since 1500." http://datazone.birdlife.org/sowb/casestudy/we-have-lost-over-150-bird-species-since-1500.
5. Frankham, R. 1998. "Inbreeding and Extinction: Island Populations." Conservation Biology 12 (3): 665–75.
6. Parmesan, C. 2006. "Ecological and Evolutionary Responses to Recent Climate Change." Annual Review of Ecology Evolution and Systematics 37 (May): 637–69.
7. Bellard, C, C Bertelsmeier, P Leadley, W Thuiller, and F Courchamp. 2012. "Impacts of Climate Change on the Future of Biodiversity." Ecology Letters 15 (4): 365–77.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

8. Ferreira, MT, P Cardoso, PAV Borges, R Gabriel, EB de Azevedo, F Reis, MB Araújo, and RB Elias. 2016. "Effects of Climate Change on the Distribution of Indigenous Species in Oceanic Islands (Azores)." Climate Change 138: 603–15.
9. O'Grady, JJ, BW Brook, DH Reed, JD Ballou, DW Tonkyn, and R Frankham. 2006. "Realistic Levels of Inbreeding Depression Strongly Affect Extinction Risk in Wild Populations." Biological Conservation 133 (1): 42–51.
10. Edmands, S. 2007. "Between a Rock and a Hard Place: Evaluating the Relative Risks of Inbreeding and Outbreeding for Conservation and Management." Molecular Ecology 16: 463–75.
11. Charlesworth, D, and JH Willis. 2009. "The Genetics of Inbreeding Depression." Nature Reviews Genetics 10: 783-96.
12. Lynch, M, J Conery, and R Burger. 1995. "Mutational Meltdowns in Sexual Populations." Evolution 49 (6): 1067–80.
13. Reed, DH. 2005. "Relationship between Population Size and Fitness." Conservation Biology 19 (2): 563–68.
14. Tollington, S, CG Jones, A Greenwood, V Tatayah, C Raisin, T Burke, DA Dawson, and JJ Groombridge. 2013. "Long-Term, Fine-Scale Temporal Patterns of Genetic Diversity in the Restored Mauritius Parakeet Reveal Genetic Impacts of Management and Associated Demographic Effects on Reintroduction Programmes." Biological Conservation 161: 28–38.
15. Nair, P. 2014. "Conservation Genomics." Proceedings of the National Academy of Sciences of the United States of America 111: 569.
16. Romiguier, J, P Gayral, M Ballenghien, A Bernard, V Cahais, A Chenuil, Y Chiari, et al. 2014. "Comparative Population Genomics in Animals Uncovers the Determinants of Genetic Diversity." Nature 515 (7526): 261–63.
17. Frankham, R. 1997. "Do Island Populations Have Less Genetic Variation than Mainland Populations?" Heredity 78 (3): 311–27.
18. James, JE, R Lanfear, and A Eyre-Walker. 2016. "Molecular Evolutionary Consequences of Island Colonization." Genome Biology and Evolution 8 (6): 1876–88.
19. Mateo, JA, LF Lópex Jurado, and P Geniez. 2009. "Historical Distribution of Razo Lark *Alauda razae* in the Cape Verde Archipelago." Alauda 77 (4): 309–12.
20. Ratcliffe, N, LR Monteiro, and CJ Hazevoet. 1999. "Status of the Raso lark *Alauda razae* with Notes on Threats and Foraging Behaviour." Bird Conservation International 9: 43–46.
21. Donald, P, and M de L Brooke. 2006. "An Unlikely Survivor: The Peculiar Natural History of the Raso lark." British Birds 99: 420–30.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22. Brooke, M de L. 2018. "Extreme Rainfall-Related Clutch Size Variation in the Critically Endangered Raso lark *Alauda Razae* of the Arid Cape Verde Islands." Bird Conservation International. doi:10.1017/S0959270918000163.
23. Pala, I, S Naurin, M Stervander, D Hasselquist, S Bensch, and B Hansson. 2012. "Evidence of a Neo-Sex Chromosome in Birds." Heredity 108 (3): 264–72.
24. Brooke, M de L, JA Welbergen, MC Mainwaring, M Van Der Velde, AMF Harts, J Komdeur, and W Amos. 2010. "Widespread Translocation from Autosomes to Sex Chromosomes Preserves Genetic Variability in an Endangered Lark." Journal of Molecular Evolution 70: 242–46.
25. Bulatova, NS. 1973. "Unusually Large Sex Chromosomes in Some Larks (Aves: Alaudidae)." Mammalian Chromosomes Newsletter 14: 150–51.
26. BirdLife International. 2016. "Species Factsheet." http://www.birdlife.org/datazone/species- factsheet.php?id=3957.
27. IUCN. 2017. "IUCN Red List of Threatened Species. Version 2016.3." Accessed January 18. http://www.redlist.org/.
28. Griffiths, R, MC Double, K Orr, and RJG Dawson. 1998. "A DNA Test to Sex Most Birds." Molecular Ecology 7 (8): 1071–75.
29. Gnerre, S, I Maccallum, D Przybylski, FJ Ribeiro, JN Burton, BJ Walker, T Sharpe, et al. 2011. "High-Quality Draft Assemblies of Mammalian Genomes from Massively Parallel Sequence Data." Proceedings of the National Academy of Sciences of the United States of America 108 (4): 1513–18.
30. Bolger, AM, M Lohse, and B Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." Bioinformatics 30 (15): 2114–20.
31. Andrews, S. 2010. "FastQC: A Quality Control Tool for High Throughput Sequence Data." http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/. doi:citeulike-article-id:11583827.
32. Merrill, RM, P Rastas, SH Martin, MC Melo, S Barker, J Davey, WO McMillan, and CD Jiggins. 2019. "Genetic Dissection of Assortative Mating Behavior." PLOS Biology 17 (2): e2005902.
33. Catchen, J, PA Hohenlohe, S Bassham, A Amores, and WA Cresko. 2013. "Stacks: An Analysis Tool Set for Population Genomics." Molecular Ecology 22 (11): 3124–40.
34. Langmead, B, and SL Salzberg. 2012. "Fast Gapped-Read Alignment with Bowtie 2." Nature Methods 9 (4): 357–59.
35. Korneliussen, TS, A Albrechtsen, and R Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." BMC Bioinformatics 15 (1). doi:10.1186/s12859-014-0356-4.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

36. Kurtz, S, S Salzberg, M Shumway, A Delcher, C Antonescu, A Phillippy, and M Smoot. 2004. “Versatile and Open Software for Comparing Large Genomes.” Genome Biology 5 (2): R12.
37. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. “The Sequence Alignment/Map format and SAMtools.” Bioinformatics 25, 2078–9.
38. Gutenkunst, RN, RD Hernandez, SH Williamson, and CD Bustamante. 2009. “Inferring the Joint Demographic History of Multiple Populations from Multidimensional SNP Frequency Data.” PLoS Genetics 5 (10). doi:10.1371/journal.pgen.1000695.
39. Liu, X, and YX Fu. 2015. “Exploring Population Size Changes Using SNP Frequency Spectra.” Nature Genetics 47 (5): 555–59.
40. Smeds, L, A Qvarnström, and H Ellegren. 2016. “Direct Estimate of the Rate of Germline Mutation in a Bird.” Genome Research 26 (9): 1211–18.
41. Dierickx, EG, R Robinson and M de L Brooke. 2019. “Survival of a Long-Lived Single Island Endemic, the Raso lark *Alauda razae*, in Relation to Fluctuating Population and Rainfall.” Scientific Reports. In review.
42. Korneliussen, T, and I Moltke. 2015. “NgsRelate: A Software Tool for Estimating Pairwise Relatedness from next-Generation Sequencing Data.” Bioinformatics 31 (24): 4009–11.
43. Van der Auwera, GA, M Carneiro, C Hartl, R Poplin, G del Angel, A Levy-Moonshine, T Jordan, et al. 2013. “From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline.” Current Protocols in Bioinformatics 43: 11.10.1-11.10.33.
44. DePristo, MA, E Banks, RE Poplin, KV Garimella, JR Maguire, C Hartl, AA Philippakis, et al. 2011. “A Framework for Variation Discovery and Genotyping Using.” Nature Genetics 43 (5): 491–98.
45. Dodds, KG, JC McEwan, R Brauning, RM Anderson, TC Stijn, T Kristjánsson, and SM Clarke. 2015. “Construction of Relatedness Matrices Using Genotyping-by-Sequencing Data.” BMC Genomics 16 (1): 1–15.
46. Sigeman, H, S Ponnikas, P Chauhan, EG Dierickx, M de L Brooke, and B Hansson. 2019. “Repeated Evolution During a Recent Avian Sex Chromosome Expansion.” doi: https://doi.org/10.1101/585059.
47. Marth, GT, E Czabarka, J Murvai, and ST Sherry. 2004. “The Allele Frequency Spectrum in Genome-Wide Human Variation Three Large World Populations.” Genetics 372 (January): 351–72.

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

48. Polanski, A, and M Kimmel. 2003. “New Explicit Expressions for Relative Frequencies of Single-Nucleotide Polymorphisms with Application to Statistical Inference on Population Growth.” Genetics 165 (1): 427–36.
49. Hedrick, PW, and RC Lacy. 2015. "Measuring Relatedness Between Inbred Individuals." Journal of Heredity 106 (1):20-5.
50. Balakrishnan, CN, and SV Edwards. 2009. “Nucleotide Variation, Linkage Disequilibrium and Founder-Facilitated Speciation in Wild Populations of the Zebra Finch (*Taeniopygia guttata*).” Genetics 181 (2): 645–60.
51. Donald, PF. 2004. The Skylark. London: T & AD Poyser.
52. Wright, AE, R Dean, F Zimmer, and JE Mank. 2016. “How to Make a Sex Chromosome.” Nature Communications 7 (May): 12087.
53. Lewontin, RC. 1974. The Genetic Basis of Evolutionary Change. New York: Columbia University Press.
54. Leffler, EM, K Bullaughey, D. Matute, WK Meyer, L Ségurel, A Venkat, P Andolfatto, and M Przeworski. 2012. “Revisiting an Old Riddle: What Determines Genetic Diversity Levels within Species?” PLoS Biology 10 (9). doi:10.1371/journal.pbio.1001388.
55. Avise, JC, J Arnold, RM Ball, E Bermingham, T Lamb, JE Neigel, CA Reeb, and NC Saunders. 1987. “Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics.” Annual Review of Ecology and Systematics 18 (1): 489–522.
56. Cutter, AD, and BA Payseur. 2013. “Genomic Signatures of Selection at Linked Sites: Unifying the Disparity among Species.” Nature Reviews Genetics. doi:10.1038/nrg3425.
57. Corbett-Detig, RB, DL Hartl, and TB Sackton. 2015. “Natural Selection Constrains Neutral Diversity across A Wide Range of Species.” PLoS Biology 13 (4). doi:10.1371/journal.pbio.1002112.
58. de Villemereuil, P, A Rutschmann, KD Lee, JG Ewen, P Brekke, and AW Santure. 2019. “Little Adaptive Potential in a Threatened Passerine Bird.” Current Biology 29 (5): 889–94.

59. Ministry of Environment Housing and Territory Planning of Cape Verde. 2011. “Climate Change Vulnerability Assessment of Cape Verde: Summary for Policy Makers.” Praia.
60. Frankham, R, JD Ballou, and DA Briscoe. 2002. Introduction to Conservation Genetics. Cambridge: Cambridge University Press.
61. Brekke, P, PM Bennett, J Wang, N Pettorelli, and JG Ewen. 2010. “Sensitive Males: Inbreeding Depression in an Endangered Bird.” Proceedings of the

bioRxiv preprint doi: https://doi.org/10.1101/617563; this version posted April 24, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Royal Society B: Biological Sciences 277 (1700): 3677–84.
doi:10.1098/rspb.2010.1144.