ARTICLES
https://doi.org/10.1038/s41559-018-0606-3
nature ecology & evolution

Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates

Hernán E. Morales[1,2*], Alexandra Pavlova[1], Nevil Amos[1,3], Richard Major[4], Andrzej Kilian[5], Chris Greening[1] and Paul Sunnucks[1]

Metabolic processes in eukaryotic cells depend on interactions between mitochondrial and nuclear gene products (mitonuclear interactions). These interactions could have a direct role in population divergence. Here, we study mitonuclear co-evolution in a widespread bird that experienced population divergence followed by bidirectional mitochondrial introgression into different nuclear backgrounds. Using >60,000 single nucleotide polymorphisms, we quantify patterns of nuclear genetic differentiation between populations that occupy areas with different climates and harbour deeply divergent mitochondrial lineages despite ongoing nuclear gene flow. We find that strong genetic differentiation and sequence divergence in a region of ~15.4 megabases on chromosome 1A mirror the geographic pattern of mitochondrial DNA divergence. This result is seen in two different transects representing populations with different nuclear backgrounds. The chromosome 1A region is enriched for genes performing mitochondrial functions (N-mt genes). Molecular signatures of selective sweeps in this region alongside those in the mitochondrial genome suggest a history of adaptive mitonuclear co-introgression. Moreover, evidence for large linkage disequilibrium blocks in this genomic region suggests that low recombination could facilitate functional interactions between co-evolved nuclear alleles. Our results are consistent with mitonuclear co-evolution as an important mechanism for population divergence and local adaptation.

Genomic studies of the early stages of population divergence enhance our understanding of the genetic basis of local adaptation, reproductive isolation and speciation[1–3]. Genomic differentiation between closely related populations is often heterogeneous, with low differentiation across most of the genome accompanied by high differentiation at some regions[4]. Regions of elevated genetic differentiation are commonly referred to as genomic islands of differentiation[5,6]. In some cases, genomic islands of differentiation can form under divergent natural selection, which limits gene flow at barrier loci[6] (genes involved in local adaptation and/or reproductive isolation) and their linked neighbours[7]. In others, they are a by-product of past directional or background selection on linked loci[8,9]. Investigating patterns of genomic differentiation can enhance understanding of the genomic basis of local adaptation and the evolution of barrier loci during population divergence[10–12]. Few studies have directly investigated the co-evolution of nuclear and mitochondrial genomes during the early stages of divergence in natural populations[13–17] and none have found regions of genomic differentiation enriched for nuclear genes with mitochondrial functions. Genes encoded by mitochondrial DNA (mtDNA) and nuclear-encoded genes with mitochondrial functions (N-mt genes) are nevertheless prime candidates for driving population divergence because they maintain essential functions of energetics, metabolism and gene regulation[18,19]. Most significantly, mtDNA and N-mt genes co-encode and regulate oxidative phosphorylation (OXPHOS) complexes, which serve as the primary source of available chemical energy in the cell[20]. Accordingly, genetic variation in mtDNA can have strong fitness effects, often expressed through mitonuclear interactions[21,22], which can be environment dependent[23].

Mitonuclear interactions occur despite mitochondrial and nuclear genomes having different modes of inheritance, recombination and mutation rates, implying that their co-evolution is enforced by natural selection[24,25]. Two types of selective pressures can act on mitonuclear interactions during population divergence. First, selection to maintain metabolic functionality drives co-evolution of mtDNA and N-mt genes: in response to rapid accumulation of slightly deleterious mutations and/or adaptive variation in mtDNA, N-mt genes evolve compensatory or complementary changes[26,27]. Thus, mitonuclear co-evolution can result in populations having different co-adapted sets of mtDNA and N-mt alleles. Second, environmental variation can drive selection for locally adapted metabolic phenotypes, modulated by mitonuclear interactions[28,29]. For example, variation in mitochondrial coupling between substrate oxidation and chemical energy production can be adaptive in environments that differ in thermal profiles and nutrient availability[30,31]. Thus, environmental variation can act as an extrinsic barrier to gene flow through diverging populations having different locally adapted sets of mtDNA and N-mt alleles. Moreover, intrinsic co-evolution and extrinsic adaptation to thermal environments could result in poor fitness of hybrid mitonuclear genotypes due to incompatibilities when two populations with different sets of co-adapted mitonuclear alleles admix (for example, after secondary contact)[28,32]. Existing or novel mechanisms to reduce recombination between

[1]School of Biological Sciences, Monash University, Clayton, Victoria, Australia. [2]Department of Marine Sciences, University of Gothenburg, Gothenburg, Sweden. [3]Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, Heidelberg, Victoria, Australia. [4]Australian Museum Research Institute, Australian Museum, Sydney, New South Wales, Australia. [5]Diversity Arrays Technology, University of Canberra, Canberra, Australian Capital Territory, Australia. *e-mail: hern.moral@gmail.com

co-adapted N-mt alleles (for example, chromosomal rearrangements) might help to maintain the integrity of mitonuclear combinations[33–35]. Thus, mitonuclear genetic incompatibilities and/or recombination-inhibiting genomic architectures may act as intrinsic barriers to gene flow.

The endemic eastern Australian songbird *Eopsaltria australis* (eastern yellow robin (EYR)) provides a model in which to study mitonuclear interactions and co-evolution during divergence with gene flow because it has two diverged populations that appear to have experienced extensive mitochondrial introgression from each other[36,37] (Supplementary Fig. 1). Coalescent modelling indicates that the main axis of north versus south nuclear DNA differentiation (mean genome-wide $G_{ST} = 0.084$) originated when northern and southern populations split approximately 2 million years ago[36], accompanied by divergence in mtDNA lineages (mitolineages; now having 6.8% difference in DNA sequence)[36]. Subsequently, two episodes of mitochondrial introgression rearranged the geographic distribution of mitochondrial divergence to coast versus inland: the northern mitolineage (mito-A) introgressed southwards through the relatively arid and climatically variable inland range ~270 thousand years ago (ka), and the southern mitolineage (mito-B) introgressed northwards through the more temperate and climatically stable coastal range ~90 ka[36]. This history approximates a 'natural experiment', in which a pattern of mitochondrial divergence between lineages occupying contrasting climates occurs in two different (that is, north and south) nuclear backgrounds (Supplementary Fig. 1). Following southward mito-A introgression, the southern population began to diverge into an inland (more arid) mito-A-bearing sub-population and a coastal (more temperate) mito-B-bearing sub-population ~60 ka. These two sub-populations are now connected by male-mediated gene flow[36]. Signatures of positive selection on six amino acids of mitochondrially encoded OXPHOS genes support non-neutral mtDNA divergence between the mitolineages[38,39]. Moreover, signatures of selective sweeps in the mitogenome, and significant correlations of mitolineage distributions with climatic variation after controlling for the effects of geography and distance, suggest that one or both mitochondrial introgression events may have been adaptive[37,38].

We hypothesize that adaptive inland–coastal mitolineage differentiation should be accompanied by corresponding differentiation in regions of the nuclear genome linked to mitochondrial function, reflecting selection to maintain optimal mitonuclear interactions. Specifically, if the initial north–south population divergence generated N-mt genes co-adapted to their regional mitochondrial type, mitochondrial introgression should have been accompanied by introgression of co-evolved alleles of N-mt genes (that is, mitonuclear co-introgression[17,40]). Here we examine these predictions using genomic data collected along two geographic transects intersecting EYR's inland–coastal mitochondrial divergence in both northern and southern nuclear backgrounds.

Results

Narrow mitolineage contact zones at regions of environmental divergence occur in both nuclear backgrounds. A total of 407 EYR individuals sampled across the species' range were categorized for mitochondrial lineage membership by sequencing the mitochondrial *ND2* gene (Fig. 1a and Supplementary Table 1). Based on earlier nuclear genetic analyses[36], sampling efforts were focused on two transects ~700 km apart—one in the northern and one in the southern nuclear background. In each transect, we identified a contact zone (where both mitolineages are present) that is narrow (~20–40 km) relative to the EYR dispersal distance (2–25 km per generation[41]). Each contact zone occurs in a region of climatic transition (Fig. 1a). The correlation between mitolineage distribution and climatic variation is stronger in the southern than the northern transect (Fig. 1b).

Strong genetic differentiation of outlier loci suggests a history of mitochondrial–nuclear co-evolution and co-introgression. We obtained 60,444 single nucleotide polymorphisms (SNPs) by performing complexity-reducing representative sequencing of the genome using Diversity Arrays Technology (DArT) sequencing (DArTseq)[42] for 164 individuals (Supplementary Table 2). We assigned each of the 153 individual birds sampled at the transects intersecting the mitochondrial and climatic divergence (Fig. 1) to one of four groups for analysis ('analysis groups') according to their nuclear background (that is, north or south transect) and mitolineage membership (mito-A or mito-B). Genome-wide genetic differentiation between mito-A- and mito-B-bearing analysis groups was modest in both transects (mean F_{ST} in north = 0.02 and south = 0.05; Supplementary Fig. 2). We used three approaches—top 1% F_{ST} values, BayeScEnv and PCAdapt—to identify strongly differentiated loci (outlier loci) in each transect (see Methods). These methods each use different assumptions and take into account potential confounding factors (for example, genetic drift) to distinguish true outliers from false positives[43]. Each approach identified 227–323 outliers in each transect, with appreciable commonality in the loci identified across methods and transects (Table 1, Fig. 2 and Supplementary Figs. 2–5). A principal component analysis (PCA) of genome-wide variation (that is, all SNPs) for all 164 samples revealed 2 main axes of EYR genetic differentiation: north–south (PC1, explaining 4.7% of variation) and inland–coastal (PC2, explaining 3.4% of the variation) (Fig. 1c and Supplementary Fig. 6). In contrast, genetic variation along the first two principal component axes of outlier loci was exclusively structured in the inland–coastal direction (PC1: 45.2%; PC2: 3.2%) with no north–south signal (Fig. 1d). These results indicate that divergent mtDNA lineages and nuclear outlier loci have parallel or coincident inland–coastal geographic distributions in north and south transects, and suggests that some nuclear outlier loci co-introgressed with mtDNA (Supplementary Fig. 1). In support of mitonuclear co-introgression, when northern and southern analysis groups bearing the same mitolineage were compared, the allele frequencies of outliers were more strongly correlated than those of random non-outliers (Supplementary Fig. 7). The earlier date for the proposed north-to-south inland mitochondrial introgression (~270 ka)[36], as well as stronger climatic differences in the southern region, could explain the overall higher coast–inland neutral nuclear differentiation in the south (Fig. 1c and Supplementary Fig. 2).

Highly differentiated loci are concentrated in two genomic clusters of differentiation. We mapped DArT tags to the reference genome of the zebra finch *Taeniopygia guttata*[44], obtaining unique, high-quality hits for 35,030 tags. Most SNPs that were strongly differentiated between mito-A- and mito-B-bearing analysis groups in each transect clustered into only a few genomic regions (Fig. 2a and Supplementary Figs. 3–5). A hidden Markov model (HMM) analysis (Methods[45]) revealed the presence of statistically supported genomic clusters of differentiation between mito-A- and mito-B-bearing analysis groups in each transect: one on chromosome 1A (the chromosome 1A cluster; ~15.4 megabases (Mb) long) and one on chromosome Z (~0.75 Mb long) (Fig. 2).

The genomic cluster of differentiation on chromosome 1A has an over-representation of nuclear-encoded genes with mitochondrial functions. We tested whether the two genomic clusters of differentiation were significantly enriched for N-mt genes. The chromosome 1A cluster had a significant excess of N-mt genes: 32 genes compared with a genome-wide average of 12.2 ($P < 0.001$; Supplementary Fig. 9 and Supplementary Tables 3 and 5). One of the nuclear-encoded regulatory genes located in this cluster, *YARS2* (mitochondrial tyrosyl-transfer RNA (tRNA) synthetase), interacts with the mitochondrially encoded tRNA-tyrosine. We have

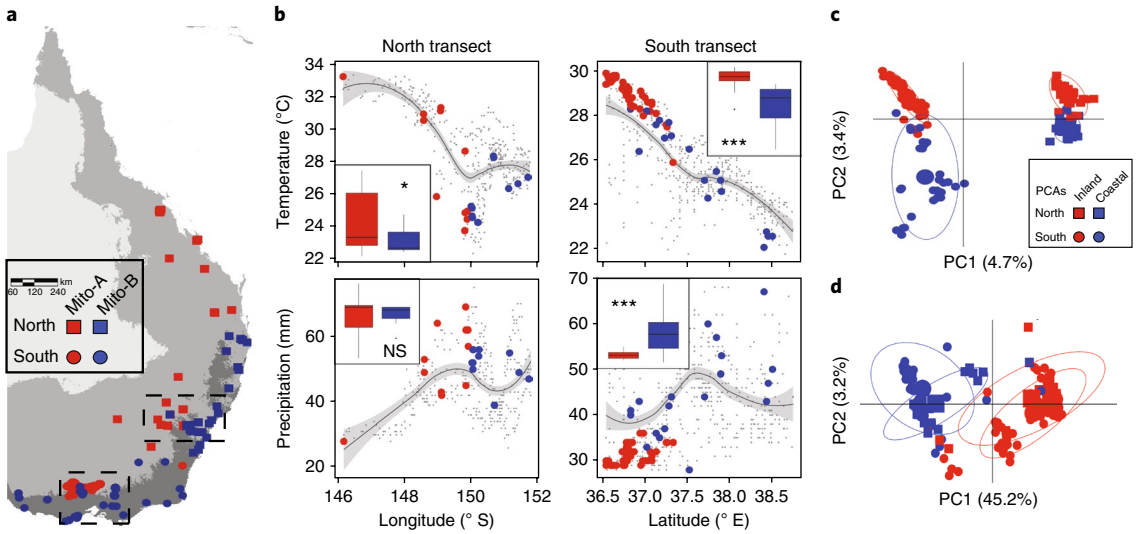

Fig. 1 | Geographic and climatic distribution of mitochondrial and nuclear genetic variation across the EYR range. a, Geographic distribution of 6.8% divergent mitolineages—inland mito-A (red; $n=262$) and coastal mito-B (blue; $n=145$)—in different nuclear backgrounds (north (squares) and south (circles)). The same symbols apply to the remaining panels. Sampling was focused along two transects (black dashed rectangles), one in each nuclear background. Shades of grey represent the maximum temperature of the warmest month (BIO5): dark grey: <28 °C; medium grey: 28–33 °C; light grey: >33 °C. **b**, Climatic distribution of mitolineages along each transect ($n=50$ in the northern transect and $n=103$ in the southern transect). The *y* axes show climatic variables; that is, the maximum temperature of the warmest month (BIO5) and the precipitation of the driest month (BIO14). Grey points represent the climatic space occupied by the EYR, obtained from presence records of the Atlas of Living Australia. Trends and 95% confidence intervals are shown by a locally weighted (LOESS) smoothing (grey line) and shading, respectively. Large coloured dots represent our sampling. Insets: box plots show the climatic divergence between mitolineages. Horizontal line, median; upper and lower hinge of the box, 75% and 25% quantiles, respectively; error bars, largest and smallest values that are, respectively greater and less than the upper and lower hinge (1.5× interquartile range). The significance of the mitolineage–climate correlation was obtained using a binomial generalized linear model (***$P<0.001$; *$P<0.05$; NS, not significant). **c**, First two axes of a PCA of nuclear genome-wide variation (60,444 SNPs) for all EYR samples ($n=164$); two major axes of genetic differentiation corresponded to north–south direction (PC1) and inland–coastal direction (PC2). **d**, PCA plot for 565 SNP loci that are outliers between groups of individuals that harbour different mitolineages, in one or both transects. Outlier loci show strong genetic differentiation in the inland–coastal direction along the two first PCA axes, but notably do not contain any signal of the north–south genome-wide divergence as in **c**. In **c** and **d**, ellipses encapsulate 80% of the data spread, and symbol size reflects the number of overlapping individual data points.

Table 1 | Number of outlier loci identified in northern and southern transects using three different methods

Method	North	South	In common (transects)
F_{ST}[a]	315	285	210
BayeScEnv[b]	227	303	191
PCAdapt[c]	323	235	171
In common (methods)	185	203	

A total of 565 different outliers were identified by one or more methods in one or both transects. Only loci mapped to the reference zebra finch genome are included. For a summary of unmapped loci, see Supplementary Fig. 8. [a]Top 1% quantile of per-marker F_{ST} between mito-A- and mito-B-bearing analysis groups in each transect within 40 km of the mitolineage contact zone centre. [b]Outliers of population differentiation ($Q\leq0.05$) with mitolineage membership used as a binomial variable. [c]Outliers of population differentiation ($Q\leq0.001$) across the first latent factor, which captures inland–coastal differentiation.

shown previously that the tRNA-tyrosine arm has a single fixed substitution in mito-A (compared with the reconstructed ancestral state)[38]. Incompatibilities between these genes have been implicated in mitochondrial dysfunction in human populations and *Drosophila* hybrids[23,46,47]. The chromosome 1A cluster also includes four N-mt genes that encode structural subunits or assembly factors of OXPHOS complexes (Supplementary Fig. 9 and Supplementary Tables 3 and 5); namely, three supernumerary subunits of OXPHOS complex I (*NDUFA6*, *NDUFA12* and *NDUFB2*) and an assembly chaperone of ATP synthase (*FMC1*)[48–52]. We used a recently published three-dimensional protein structure of OXPHOS complex I[52] (Supplementary Fig. 10) to infer the location of the three identified nuclear subunits and the three mitochondrially encoded OXPHOS subunits previously inferred to be under positive selection in EYR (ND4, ND4L and ND5; Supplementary Table 5)[38,39]. We infer that NDUFB2 interacts directly with ND5. While the positions of *NDUFA6* and *NDUFA12* suggest that they do not interact directly with the mitochondrially encoded subunits, both subunits occur at regions critical for energy transduction[48–52]. Accordingly, their substitutions may modulate the function of the complex via long-range conformational changes dependent on mitochondrially encoded subunits[49,51] (Supplementary Fig. 10). In contrast, the cluster on the Z chromosome contained only one N-mt gene (*SLC25A46*) and no OXPHOS genes.

The chromosome 1A genomic cluster of differentiation is characterized by signatures of selective sweeps and low recombination. We compared genome-wide linkage disequilibrium decay with that of chromosome 1A. Genome-wide linkage disequilibrium decayed rapidly and reached average linkage disequilibrium levels at a genetic distance of ~7.8 kilobases (kb) between markers (Supplementary Fig. 11 and Supplementary Table 6). Chromosome 1A had the highest values of linkage disequilibrium of all chromosomes and a substantially slower average linkage disequilibrium decay (~140 kb; Supplementary Table 6 and Supplementary Figs. 11 and 12). The chromosome 1A cluster is also characterized by low genetic diversity in both transects: observed heterozygosity was lower within the cluster than outside it (Fig. 2b). High—albeit

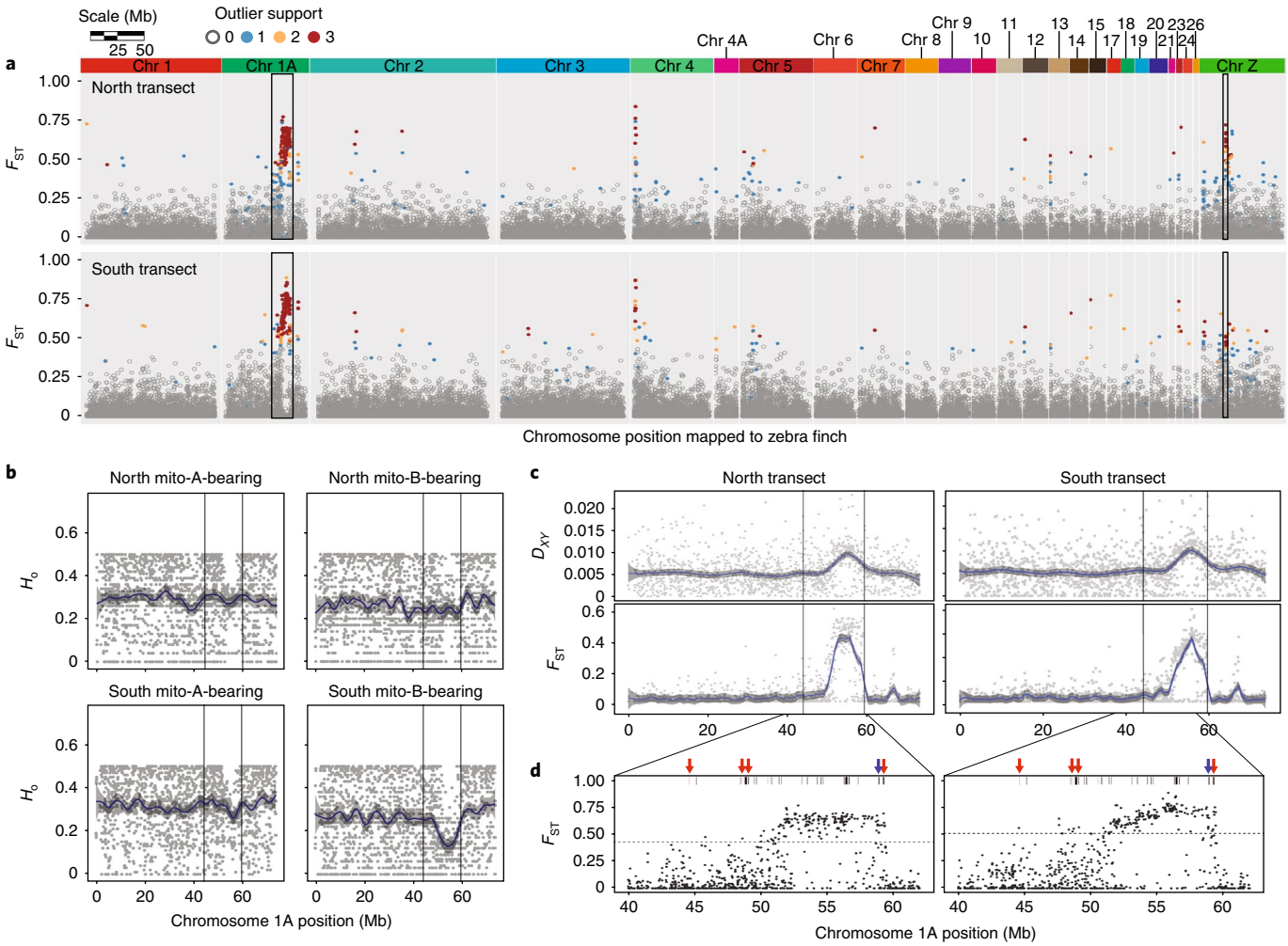

Fig. 2 | Heterogeneous genomic differentiation between mito-A- and mito-B-bearing analysis groups, and properties of the chromosome 1A cluster of differentiation. a, Manhattan plots. The *y* axis shows the F_{ST} value for each SNP calculated from individuals sampled within 40 km of the contact zone. The *x* axis shows the genomic position of each SNP relative to the zebra finch reference genome. Data points are blue, yellow or red if the outliers were detected by one, two or three methods, respectively. The outlier methods used were: top 1% F_{ST}, BayeScEnv with mitolineages as covariates ($Q \leq 0.05$) and PCAdapt ($Q \leq 0.001$). Black rectangles represent genomic clusters of differentiation on chromosome 1A and chromosome Z identified using a hidden Markov model (HMM) ($Q \leq 0.05$). Chr, chromosome. **b**, Observed heterozygosity (H_o) was lower within the chromosome 1A cluster, as indicated by the vertical black lines (mean $H_o = 0.28$) than outside it (mean $H_o = 0.32$) for mito-A- and mito-B-bearing analysis groups in each transect. Grey dots represent per-SNP values. Blue lines with grey shading show generalized additive model smoothing with 95% confidence intervals. **c**, Average sequence divergence (D_{XY}) and genetic differentiation (F_{ST}) between mito-A- and mito-B-bearing analysis groups in each transect. Lines show generalized additive model smoothing with 95% confidence intervals. Vertical black lines represent the chromosome 1A cluster of differentiation. **d**, F_{ST} values within the chromosome 1A cluster of differentiation for each transect. Horizontal lines represent the upper 99% quantile. The relative positions of genes with functional annotations to mitochondrial activity (N-mt genes; GO term: 0005739) are indicated by black bars. Protein-coding OXPHOS genes (GO term: 0006119) are indicated by red arrows (from left to right: *NDUFA12*, *FMC1*, *NDUFA6* and *NDUFB2*). The mitochondrial tyrosyl-tRNA synthetase (*YARS*) is indicated by a blue arrow.

variable—linkage disequilibrium and low genetic diversity at the chromosome 1A cluster suggest a selective sweep[53] for this region, paralleling the pattern of selective sweep previously observed in mtDNA[38]. Absolute sequence divergence (D_{XY}) was elevated at the chromosome 1A cluster region in both transects (Fig. 2c); this mirrors the elevated relative divergence seen in F_{ST}, thus indicating that high F_{ST} is driven by differences in gene flow and not simply by differences in levels of within-species polymorphism[9]. Next, we investigated whether the chromosome 1A cluster contained large blocks of linkage disequilibrium, as expected under low recombination for this genomic region. In the presumed source populations of the introgression events (north mito-A-bearing and south mito-B-bearing analysis groups[36]), no large linkage disequilibrium blocks were observed (Fig. 3). In contrast, the contact zones and populations away from the contact zones that experienced past introgression (that is, north mito-B-bearing and south mito-A-bearing analysis groups) showed large linkage disequilibrium blocks within the chromosome 1A cluster (Fig. 3). This suggests that both classes of chromosome 1A cluster haplotypes ('inland' and 'coastal') still segregate largely intact in these localities. Together, these linkage disequilibrium data point to low recombination within the chromosome 1A cluster despite ongoing gene flow.

Away from the contact zones, individual genetic assignment based on nuclear outliers is concordant with mitolineages. We used STRUCTURE version 2.3.4[54] to estimate the probability of assignment (Q) of each individual to nuclear genetic clusters based on each of three subsets of data: putatively neutral loci (6,947

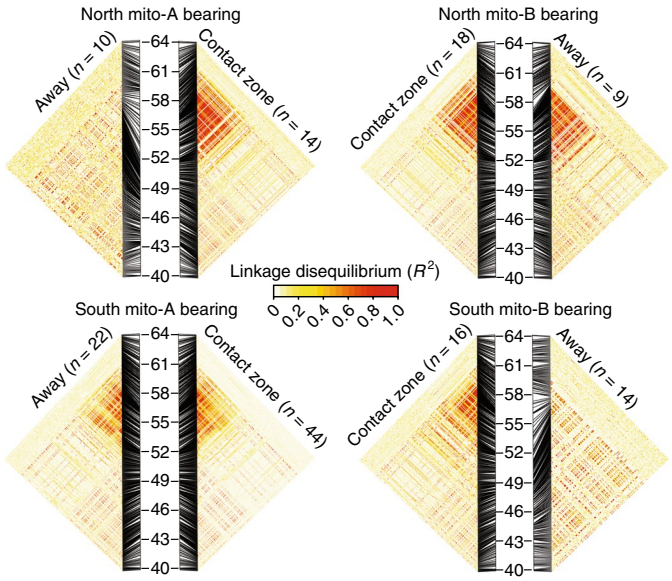

Fig. 3 | Linkage disequilibrium blocks within the chromosome 1A cluster of differentiation for individuals in the contact zone and away from the contact zone. Each triangular plot shows a heat map of pairwise linkage disequilibrium (R^2) between all SNP markers. SNPs are arranged vertically according to their relative position within chromosome 1A in a region containing the genomic cluster of differentiation (44.2–59.6 Mb). The numbers between heat maps show the genomic coordinates (in Mb) on chromosome 1A of the zebra finch reference genome. Thin black lines show the position of each SNP relative to the reference.

genome-wide non-outlier loci and not linked to outlier loci); autosomal outliers except those in the chromosome 1A cluster (227 outliers); and outliers restricted to the chromosome 1A cluster (236 outliers). Consistent with earlier findings[36], analysis of neutral loci showed that individuals fell into two major clusters—north or south (Fig. 4a), with further subdivision of the southern cluster into two admixed subclusters (inland and coastal). In contrast, both outlier analyses identified just two clusters—inland and coastal (red and blue, respectively, in Fig. 4a and Supplementary Figs. 13 and 14), with both clusters present in the north and south transects. Given that the chromosome 1A cluster is enriched for N-mt genes and potentially under selection for mitonuclear interactions, cluster membership for chromosome 1A cluster outliers is expected to match mitolineage membership. Based on the results of the chromosome 1A cluster outlier analysis, we classified each individual as being inland ($Q \geq 0.9$ in the inland cluster), coastal ($Q \geq 0.9$ in the coastal cluster) or admixed ($0.9 < Q > 0.1$) according to the outliers, and assessed whether each inland and coastal individual possessed nuclear outliers that matched their mitolineage (inland/mito-A or coastal/mito-B) or mismatched it (inland/mito-B or coastal/mito-A). Most individuals (>66%) were classified as matched in both transects. Mismatched (<13%) and admixed (<28%) individuals were consistently fewer and tended to occur near the contact zone between mitolineages in each transect (Fig. 4c,d).

Discussion

Patterns of genome-wide genetic differentiation between groups of individuals that harbour two deeply divergent mitolineages, occupy contrasting climates and undergo gene flow provide evidence of a genomic cluster of differentiation implicated in mitonuclear co-evolution in a widespread songbird. Genomic differentiation is concentrated in a ~15.4-Mb genomic region on chromosome 1A that contains an over-representation of N-mt genes. This is consistent with a model in which fast-evolving mitochondrial variation generates selection for compensatory evolution of N-mt genes. Co-evolving mitochondrial and nuclear genes appear to have undergone mitonuclear co-introgression in response to Pleistocene climate changes[36] (Supplementary Fig. 1). Many loci within the detected cluster of differentiation on chromosome 1A have low genetic diversity and high linkage disequilibrium, as is typically seen under selective sweeps, paralleling the genomic signatures previously inferred in EYR mitogenomes[38]. The inferred patterns of selective sweeps at the cluster on chromosome 1A and localized suppressed recombination are consistent with co-introgression of mitonuclear interacting alleles under ecologically divergent selection[7,9,55]. Moreover, the elevated sequence divergence (D_{XY}) and relative differentiation (F_{ST}) within the chromosome 1A cluster suggest that divergent selection acting at putative barrier loci (for example, N-mt genes) counteracts the otherwise homogenizing effects of gene flow[7,9]. In contrast, genome-wide variation in recombination rate or background selection in a region of low recombination (for example, the incidental island model[56]) are unlikely to explain our results. However, we note that future whole-genome analyses could reveal smaller genomic clusters of differentiation throughout the genome. Overall, our findings are consistent with climate-driven mitochondrial divergence accompanied by non-neutral divergence of nuclear genes with mitochondrial function, enforced by selection to maintain optimal mitonuclear interactions.

Natural selection might play major roles in EYR's mitonuclear divergence in different ways. Inland–coastal climatic gradients could act as an extrinsic barrier to gene flow[28,29,57]. Despite the capability to travel several kilometres per day[41], EYR individuals with inland genotypes in the chromosome 1A outlier region are never found deep into coastal habitats and vice versa, and admixed individuals are rare away from the intermediate habitats of the contact zone (Fig. 4c,d). This effect is apparently less pronounced in the southern transect because the inland range of the species stops not far north of the contact zone (Figs. 1a and 4d). Intrinsic barriers to gene flow would also be in operation if the fitness of hybrids is lowered by genetic incompatibilities[32,33]. The low but measurable proportion of chromosome 1A outlier genotypes that are admixed or mismatched compared with their respective mitolineage indicates that intrinsic barriers to gene flow are not absolute, and that some first-generation offspring of inland–coastal crosses are sufficiently viable and fecund to produce backcrosses, in both directions (Fig. 4a,b). Empirical tests are required to determine whether the various classes of inland–coastal crosses have lower fitness (work in progress), but it is possible that compatibility of the maternal set of N-mt alleles with mtDNA is sufficient to moderate the loss of fitness in first-generation offspring of inland–coastal crosses and subsequent maternal backcrosses[28]. Similarly, the fitness of individuals bearing a mitolineage mismatching their chromosome 1A cluster alleles might depend on whether they have matching alleles at autosomal outliers elsewhere in the genome (Fig. 4a). It is also possible that selection on mitonuclear combinations might be additive rather than epistatic (that is, selection might independently target mitochondrial and nuclear variation).

Genomic architecture characterized by low recombination can facilitate functional interactions between co-evolved nuclear alleles, and maintain the integrity of divergent haplotypes, which under selection could more readily remain associated with their respective mitolineages, even in the presence of gene flow[33,58]. Low recombination can arise through a variety of mechanisms. Phylogenetically conserved recombination cold-spots and chromosomal rearrangements (for example, inversions) can suppress recombination and maintain barrier loci in strong linkage[34,35,59]. Consistently, the cluster overlaps with a region of low recombination shared by zebra finches and flycatchers, including the presumed position of the zebra finch centromere (Supplementary Fig. 15)[44,59,60]. Clusters of genetic differentiation have been identified at a similar genomic location in

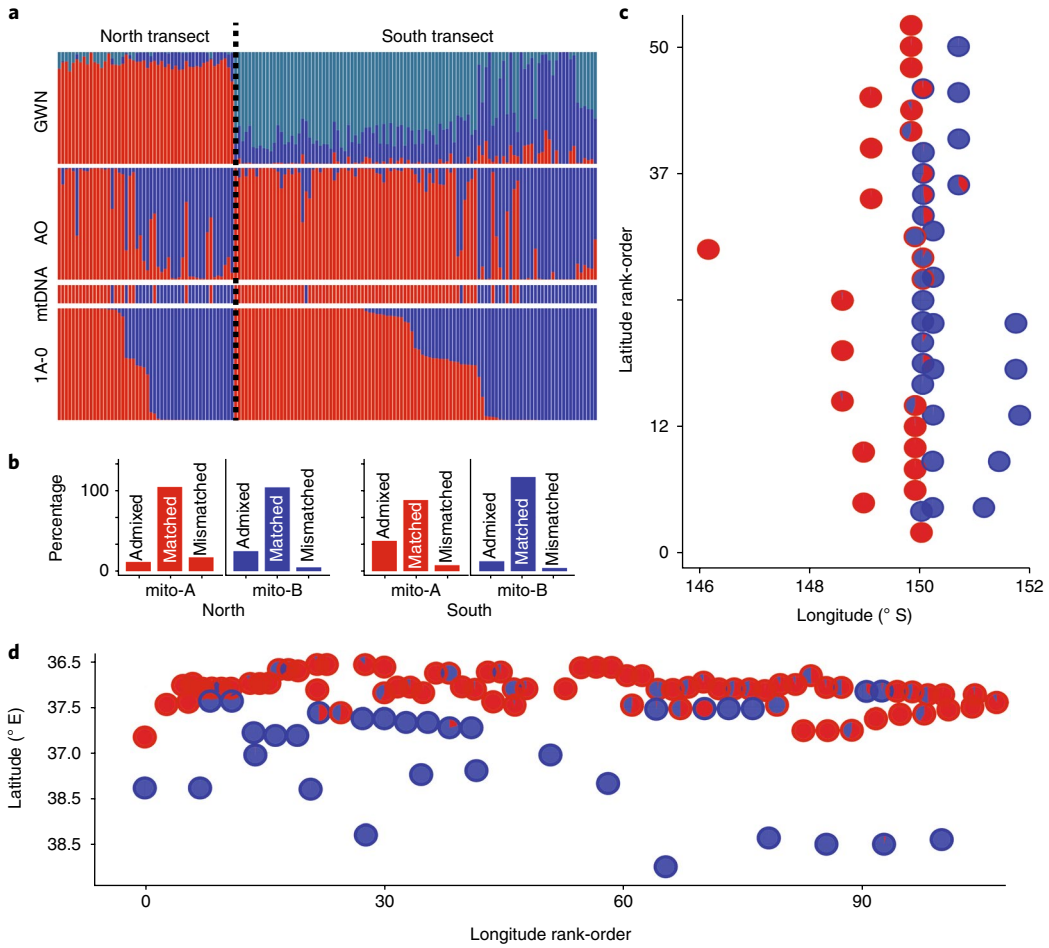

Fig. 4 | Individual admixture at genome-wide non-outliers (GWN), autosomal outliers not in the chromosome 1A cluster of differentiation (AO) and chromosome 1A cluster outliers (1A-O), mitochondrial–nuclear matching of 1A-O, and their geographic distributions. a, Probability of assignment (*Q*; *y* axes; scale 0–1) of each individual to nuclear genetic clusters (nuclear DNA bar plots) for GWN (top; red, northern cluster; light blue, south-inland; dark blue, south-coastal; $n = 6{,}947$ putatively neutral loci; $K = 3$), AO (middle; red, inland; blue, coastal; $n = 227$ autosomal outliers; $K = 2$) and 1A-O (bottom; red, inland; blue, coastal; $n = 236$ chromosome 1A cluster outliers; $K = 2$). The corresponding mitolineage membership for each individual is shown as mtDNA bar plots (third panel from the top; red, mito-A; blue, mito-B). Individuals within each transect are displayed, left to right, in the order of decreasing membership of the inland 1A-O cluster. **b**, Percentage of individuals in different categories of correspondence between nuclear ancestry of inland and coastal clusters inferred from 1A-O, and their mitolineage membership of mito-A and mito-B: matched ($Q \geq 0.9$ in the appropriate cluster), mitonuclear–mitolineage mismatched ($Q < 0.1$) and admixed ($0.9 < Q > 0.1$). **c**,**d**, Geographic distribution of 1A-O clusters for north (**c**) and south (**d**) transects. To avoid complete overlap of points, rank-order represents latitude in the north and longitude in the south. Pies show *Q* values for two 1A-O clusters (red, inland; blue, coastal) for each individual. The outlines of each pie chart show individual mitolineage membership (red, inland; blue, coastal). Plots analogous to panels **b**–**d** for GWN and AO loci can be found in Supplementary Fig. 13.

at least two songbird taxa harbouring deep mitochondrial divergences: greenish warblers *Phylloscopus trochiloides*[55] and *Ficedula* flycatchers[61]. Together, this suggests that chromosome 1A N-mt genes could lie within an ancient recombination cold-spot in songbirds[59,60]. In contrast, chromosomal rearrangements in EYR could explain the observed pattern of large linkage disequilibrium blocks in populations that experienced introgression and at contact zones if polymorphic gene arrangements are maintained under selection promoting recombination suppression. While songbird genomes are generally characterized by strong genome-wide synteny, intrachromosomal inversions can be common[60] and tend to occur more often in lineages with overlapping ranges[62]. Whatever the cause of low recombination and strong linkage disequilibrium blocks in the chromosome 1A cluster, recombination suppression must be incomplete, given the variation in F_{ST}, D_{XY} and linkage disequilibrium across the cluster (Figs. 2 and 3) and the existence of individuals backcrossed for subsets of chromosome 1A cluster outliers (Fig. 4). It is unclear at this point how many genes are targeted by selection in this region of suppressed recombination, and it is possible (especially under the chromosomal inversion model) that a single N-mt gene has a major effect on mitonuclear compatibility, while others are hitchhiking.

It is noteworthy that the Z chromosome had significantly higher levels of genetic differentiation between mito-A- and mito-B-bearing analysis groups in each transect compared with all autosomes except chromosomes 1A and 23 (pairwise Wilcoxon test, *P* value for all significant tests: ≤0.007). Faster divergence of sex chromosomes compared with autosomes is expected due to their lower effective population size and expression of recessive deleterious alleles in the heterogametic sex[63,64]. In EYR, stronger genomic incompatibilities in hybrid heterogametic (ZW) females than homogametic males (ZZ) (that is, Haldane's rule[65]) could explain reduced female-mediated Z-linked and mitochondrial gene flow compared with male-mediated nuclear gene flow[66], despite females being the more

dispersing sex[41,67]. Alternatively, the Z chromosome could contain barrier loci directly related to lineage recognition (for example, via plumage colour or song) and/or mitonuclear metabolism[63,68]. While north–south is the main axis of colour differentiation in EYR, we have previously shown that, at regional spatial scales, inland and coastal individuals differed slightly but significantly in plumage colouration[69].

In summary, our study contributes to the mounting evidence that mitonuclear co-evolution is an important mechanism of population divergence[13–17,28,29,32]. A region of chromosome 1A appears to be particularly influential in EYR inland–coastal divergence and has notable patterns of genetic differentiation in other songbirds. Collecting further data on fitness proxies related to mitonuclear functioning and reproductive interactions between the lineages will enable us to answer important questions about genomic landscapes of differentiation, genomic mechanisms maintaining mitonuclear interactions despite gene flow, relationships between mitochondrial metabolism and environment, and the fitness consequences of mitonuclear co-adaptation and mitonuclear genetic incompatibilities[58].

Methods

Samples and mitolineage identification. We determined the EYR mitolineage for 407 individuals ($n_{\text{mito-A}} = 262$; $n_{\text{mito-B}} = 145$) using ND2 sequences (Supplementary Table 1). A DNeasy Blood and Tissue Kit (Qiagen) was used to extract DNA. PCRs were performed following a previously described protocol[37] and sequenced commercially (Macrogen).

Climatic variables. Values for two Bioclimatic variables—maximum temperature of the warmest month (BIO5) and minimum precipitation of the driest month (BIO14)—were extracted for each sampling location and for 2,350 presence records obtained from the Atlas of Living Australia (http://www.ala.org.au) using the R package raster[70–72]. Presence records were used to represent the entire climatic space of EYR in each transect, and their trend for each climatic variable was represented by fitting a locally weighted (LOESS) smoothing against a geographic variable (longitude in the north transect and latitude in the south transect) with the function geom_smooth from the R package ggplot2. We quantified the correlation between mitolineage membership and climatic variation using a binomial generalized linear model in R.

Sequencing, genotyping and mapping. We genotyped samples using the reduced-representation approach implemented in DArTseq (Diversity Arrays Technology). Samples were digested using a combination of *Pst*I and *Nsp*I enzymes following ref. [42]. Reduced-representation libraries were prepared for Illumina sequencing, including a varying length barcode region (including 63 technical replicates). Libraries were sequenced with 77 single-read sequencing cycles on Illumina HiSeq 2000. Sequencing reads were processed using DArT P/L's proprietary analytical pipelines (see Supplementary Materials for details).

A total of 68,258 DArT tags (60 base pairs each after trimming barcodes) containing 97,070 SNP markers were obtained. We retained 60,444 SNPs that had a call rate of at least 80%. We approximated the genomic position of each SNP by mapping DArT tags to the reference genome of the zebra finch *T. guttata*, taeGut3.2.4 (ref. [44]) using BLASTn version 2.3.0 (ref. [73]). Only unique high-quality hits were considered, ensuring high confidence in the mapping. Due to the limitation of not having a reference genome for EYR, we were able to map only 55% of the DArT tags using this method, and a few hundred outlier loci were not mapped (Supplementary Fig. 8). Mapped markers were distributed over all autosomes and the Z chromosome except the microchromosomes 16, LG2 and LG5. There was strong correlation between chromosome size and the number of mapped tags, indicating that tags mapped uniformly across the zebra finch genome (Pearson's $r = 0.99$; Supplementary Fig. 16). We were unable to assess the coverage of the female-specific W chromosome as it was absent from the zebra finch reference genome.

We explored potential inconsistencies in mapping EYR DArT tags to the zebra finch reference genome[44] by comparing it with mapping to the collared flycatcher reference genome[61]. DArT tags mapped to similar genomic positions in both genomes for most chromosomes, including the region of the chromosome 1A genomic cluster of differentiation (Supplementary Fig. 14). This result was expected given the extreme level of gene order conservation between these two passerine birds; their gene order is even similar to that of the chicken, from which they differ karyotypically in two chromosomes—1A and 4A (which resulted from chromosomal fission and fusion, respectively[74]). However, the inferred mapping locations of DArT tags on the Z sex chromosome of the two reference genomes were substantially different (Supplementary Fig. 14), indicating potential inconsistencies with our mapping in this chromosome. The zebra finch genome is our preferred reference given its more advanced level of annotation.

Genetic differentiation between mito-A- and mito-B-bearing analysis groups in each transect and identification of outliers. We summarized the genetic structure of all samples with PCA using the dudi.pca function in Adegenet 2.0.1 (ref. [75]). We performed PCA analyses for all 60,444 SNPs and 565 outliers discovered by one or more methods in one or both transects. For each transect, we estimated genetic differentiation between 'analysis groups' defined by the birds' mitolineage, and identified outlier loci (that is, loci of extreme differentiation compared with the average) using three different methods:

(1) F_{ST} -outlier detection at fine spatial scales.We measured per-SNP genetic differentiation (Weir and Cockerham's F_{ST}) between mito-A- and mito-B-bearing analysis groups in each transect using the diffCalc function of the R package DiveRsity and considered the upper 1% values as outlier loci[76,77]. We limited the samples to individuals within 40 km from the mid-point of the contact zone (14 north mito-A-bearing versus 22 north mito-B-bearing, and 20 south mito-A-bearing versus 12 south mito-B-bearing). We assumed that within an 80-km-wide region the effect of genetic drift due to geographic distance is negligible, given that EYR disperses 2–25 km per generation.

(2) Correlations between nuclear loci and mitolineage membership with BayScEnv.We used BayeScEnv[78], which detects loci that depart from neutral expectations (outliers) based on their F_{ST} values and associations of allele frequencies with environmental variables, while correcting for the confounding effects of population history. SNPs are assigned to either a neutral model or one of two non-neutral models: the environmental correlation model and the locus-specific model. We used all the samples contained in each transect for this analysis. Individuals were grouped into sample sets for analysis according to their geographic location and mitolineage (Supplementary Fig. 17). We used the mitolineage membership of each sample set as a binomial environmental variable (mito-A coded 1 and mito-B coded −1). We also conducted BayeScEnv analyses using climatic variables that correlate to the mitochondrial divergence (Fig. 1b) as continuous environmental variables: maximum temperature of the warmest month (BIO5) and precipitation of the driest month (BIO14). These analyses yield similar results to the analysis with mitolineage membership as a binomial variable, highlighting the strong mtDNA–climate correlation (Supplementary Fig. 4). For all BayeScEnv analyses, we used 20 pilot runs of 4,000 iterations with a burn-in of 80,000 iterations and samples taken every 10 steps, as well as a false discovery rate significance threshold of 5%. Convergence of every run was confirmed using the R package coda[79]. We assigned equal prior probability to both non-neutral models.

(3) PCA. We used the programme PCAdapt[80] to identify SNPs contributing the most to genetic differentiation across each transect, without assuming any kind of prior grouping. We used all the samples contained in each transect for this analysis. PCAdapt uses a hierarchical Bayesian model to determine population structure with latent factors (K, analogous to PCA axes) and identify outlier loci that contribute disproportionately to explaining each of the K factors. Among the important differences compared with other methods are that PCAdapt does not rely on F_{ST} estimates, does not require classification of individuals into populations, performs well across a range of demographic models and it is agnostic to environmental information (or mitolineage membership in our case). An initial inspection of 76 K factors revealed that while the first 7 Ks explain most of the genetic variation, only $K1$ differentiated well between mito-A- and mito-B-bearing analysis groups in both transects (Supplementary Fig. 18). Accordingly, we performed the analysis with $K = 2$ and extracted outliers along $K1$ with Qvalues lower than 0.01.

Allelic frequency correlations. To test whether alleles from outlier loci were segregating in the same direction in both divergent nuclear genomic backgrounds, we performed a correlation analysis of allelic frequencies within mito-A- or mito-B-bearing analysis groups between the northern and southern transects: north mito-A versus south mito-A and north mito-B versus south mito-B. We tested two types of outlier loci: autosomal outliers except those in the chromosome 1A cluster ($n = 227$) and outliers in the chromosome 1A cluster ($n = 236$). We compared outlier loci correlations with correlations drawn from non-outlier loci across the rest of the genome. To correct for the effect of linkage as much as possible, we obtained a random distribution of correlation values for each group. For outlier loci, we used 100 iterations of a random selection of 100 markers. For the genome-wide neutral loci, we randomly selected 100 markers without replacement until all the markers were used (~350 iterations). We tested whether distributions of allelic correlations of each outlier loci set were greater than those for non-outlier loci with a t-test in R.

Identification of genomic clusters of differentiation. HMMs are useful to identify genomic regions that contain contiguous SNPs of high differentiation without having to rely on methods that require defining arbitrary sliding window sizes[45]. HMMs assume that genetic differentiation changes across the genome between hidden states, assigns each SNP to a given state level and identifies state transitions. We defined three hidden states of genetic differentiation—low, intermediate and high—and considered clusters

of differentiation as regions of contiguous SNPs belonging to the high-differentiation state. As input for the HMM analysis, we used the cumulative distribution function of *Q* values from the outlier analysis of BayeScEnv, with mitolineage membership as a covariate. We corrected for multiple testing with a false discovery rate significance threshold of 1%, and used a reduced dataset of 27,912 SNPs with a call rate higher than 90% and a minor allele frequency >10% to avoid rare variants[12]. To overcome low marker density for some chromosomes, we modelled hidden state changes across the entire genome. This decision did not bias our results because we did not identify any significant state transition between chromosomes. The analysis was performed with the R package HiddenMarkov[81], adapting a script written by D. Marques (https://github.com/marqueda/HMM-detection-of-genomic-islands).

Functional significance of candidates for mitochondrial–nuclear interactions. We counted zebra finch genes with functional annotations for mitochondrial activity (that is, N-mt genes; Gene Ontology (GO) term: 0005739) extracted from Ensembl (accessed November 2015)[82], and a subset of those encoding supernumerary subunits and assembly factors for OXPHOS complexes (that is, OXPHOS genes; GO term: 0006119) extracted from the Kyoto Encyclopaedia of Genes and Genomes (accessed June 2015)[83]. We performed the counts within each of the genomic clusters of differentiation and equal-sized regions across the whole genome (that is, random distribution). For each cluster of differentiation, we computed the probability that the observed count was significantly higher than the random distribution with the t.test function in R.

We mapped the protein products of three nuclear OXPHOS complex I genes located within the chromosome 1A genomic cluster of differentiation and three EYR mitochondrial OXPHOS complex I genes previously suggested to be under divergent selection[38,39] onto the 4.2-Å-resolution published structure of the homologous cryo-EM bovine mitochondrial complex I[52] in the molecular graphics programme UCSF Chimera[84].

Genetic diversity, linkage disequilibrium and D_{XY}. We calculated observed heterozygosity as a proxy for per-locus genetic diversity using the basicStats function of the R package DiveRsity[76]. We estimated per-chromosome linkage disequilibrium (R^2) with PLINK[85] using a reduced dataset of 27,912 SNPs with a call rate higher than 90% and a minor allele frequency >10% to avoid rare variants. We calculated pairwise linkage disequilibrium per chromosome and then examined the rate of linkage disequilibrium decay as a function of physical distance, using data from all chromosomes combined and independently for two chromosomes with genomic clusters of differentiation, 1A and Z, following Hill and Weir[86] as implemented in ref. [87]. We then focused on the genomic cluster of differentiation on chromosome 1A (see Results) to calculate pairwise linkage disequilibrium between all markers with the R package genetics[88] and visualized it as a triangular matrix with the R package LDheatmap[89]. For each mito-A- and mito-B-bearing analysis group in each transect, we analysed two subsets of individuals: one including individuals from the contact zone between mito-A- and mito-B-bearing analysis groups, and one including individuals away from the contact zone.

D_{XY} was estimated by first concatenating all chromosome 1A DArT tags (monomorphic positions included) and then calculating average Nei's D_{XY}[9] for non-overlapping 100-kb windows with python scripts developed by S. Martin[90] (https://github.com/simonhmartin/genomics_general/blob/master/popgenWindows.py). Relatively large windows were needed to capture several tags per window, although this is probably not a problem given the slow linkage disequilibrium decay in chromosome 1A (see Results). Trials with shorter (50 kb) and longer windows (200 kb and 500 kb) gave qualitatively similar results.

Admixture analyses. We estimated the probability of assignment (Q) of 153 individuals from two transects ($n=50$ northern; $n=103$ southern) to different nuclear genetic clusters using the admixture model with correlated allele frequencies implemented in STRUCTURE version 2.3.4[54]. Three analyses were run on all individuals. The first was run on putatively neutral loci assuming $K_{max}=3$ in accordance with previous analyses of genome-wide neutral variation[36,69]. It involved a dataset of 6,947 genome-wide non-outlier loci. First, all outliers (defined as those identified by any of the three methods above) and loci within 100,000 bases of any outliers were removed from the list of all mapped loci, and the remaining putatively neutral loci were randomly sampled every 100,000 bases along the entire genome. The second and third analyses were run on outlier loci assuming $K_{max}=2$. The second dataset included 227 autosomal outliers excluding those in the chromosome 1A cluster of differentiation. The third dataset comprised 236 outliers from the chromosome 1A cluster of differentiation. To account for unequal representation of individuals in each of the source populations, population-specific ancestry priors were used[91]. A total of 20 replicates of 200,000 burn-in iterations were followed by 1,000,000 iterations run for each K. Runs were summarized using the web server CLUMPAK[92].

Although strong linkage among outliers violated the STRUCTURE assumption of no linkage disequilibrium among loci, qualitatively similar results regarding genetic structure were obtained using assumption-free PCA (Supplementary Fig. 19).

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Code availability. R custom scripts have been deposited in GitHub at https://github.com/hmoral/EYR_DArT. These include: (1) script for HMM analyses; (2) script for allelic frequency correlations; and (3) script for N-mt gene enrichment test analyses.

Data availability. All datasets were deposited in Figshare at https://dx.doi.org/10.6084/m9.figshare.5072143. These include: (1) a table with information for every individual, including geo-climatic data, ND2 sequences and DArT genotypes; (2) an unfiltered SNP dataset with 97,070 SNP markers; and (3) BLAST results for reference genomes.

Received: 22 June 2017; Accepted: 13 June 2018;
Published online: 9 July 2018

References

1. Seehausen, O. et al. Genomics and the origin of species. *Nat. Rev. Genet.* **15**, 176–192 (2014).
2. Harrison, R. G. & Larson, E. L. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Mol. Ecol.* **25**, 2454–2466 (2016).
3. Payseur, B. A. & Rieseberg, L. H. A genomic perspective on hybridization and speciation. *Mol. Ecol.* **25**, 2337–2360 (2016).
4. Wu, C. I. The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001).
5. Wolf, J. B. & Ellegren, H. Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* **18**, 87–100 (2017).
6. Ravinet, M. et al. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.* **30**, 1450–1477 (2017).
7. Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402 (2009).
8. Noor, M. A. & Bennett, S. M. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**, 439–444 (2009).
9. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
10. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
11. Soria-Carrasco, V. et al. Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738–742 (2014).
12. Marques, D. A. et al. Genomics of rapid incipient speciation in sympatric threespine stickleback. *PLoS Genet.* **12**, e1005887 (2016).
13. Gagnaire, P.-A., Normandeau, E. & Bernatchez, L. Comparative genomics reveals adaptive protein evolution and a possible cytonuclear incompatibility between European and American eels. *Mol. Ecol. Evol.* **29**, 2909–2919 (2012).
14. Bar-Yaacov, D. et al. Mitochondrial involvement in vertebrate speciation? The case of mito-nuclear genetic divergence in chameleons. *Genome Biol. Evol.* **7**, 3322–3336 (2015).
15. Sambatti, J., Ortiz-Barrientos, D., Baack, E. J. & Rieseberg, L. H. Ecological selection maintains cytonuclear incompatibilities in hybridizing sunflowers. *Ecol. Lett.* **11**, 1082–1091 (2008).
16. Baris, T. Z. et al. Evolved genetic and phenotypic differences due to mitochondrial–nuclear interactions. *PLoS Genet.* **13**, e1006517 (2017).
17. Boratyński, Z., Ketola, T., Koskela, E., & Mappes, T. The sex specific genetic variation of energetics in bank voles, consequences of introgression?. *Evol. Biol.* **43**, 37–47 (2016).
18. Allen, J. F. The function of genomes in bioenergetic organelles. *Phil. Trans. R. Soc. B* **358**, 19–38 (2003).
19. Horan, M. P., Gemmell, N. J. & Wolff, J. N. From evolutionary bystander to master manipulator: the emerging roles for the mitochondrial genome as a modulator of nuclear gene expression. *Eur. J. Human Genet.* **21**, 1335–1337 (2013).
20. Bar-Yaacov, D., Blumberg, A. & Mishmar, D. Mitochondrial–nuclear co-evolution and its effects on OXPHOS activity and regulation. *Biochim. Biophys. Acta* **1819**, 1107–1111 (2012).
21. Ballard, J. W. O. & Pichaud, N. Mitochondrial DNA: more than an evolutionary bystander. *Funct. Ecol.* **28**, 218–231 (2014).
22. Wolff, J. N., Ladoukakis, E. D., Enríquez, J. A. & Dowling, D. K. Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Phil. Trans. R. Soc. B* **369**, 20130443 (2014).
23. Hoekstra, L. A., Siddiq, M. A. & Montooth, K. L. Pleiotropic effects of a mitochondrial–nuclear incompatibility depend upon the accelerating effect of temperature in *Drosophila*. *Genetics* **195**, 1129–1139 (2013).
24. Rand, D. M., Haney, R. A. & Fry, A. J. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol. Evol.* **19**, 645–653 (2004).

25. Dowling, D. K., Friberg, U. & Lindell, J. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends Ecol. Evol.* **23**, 546–554 (2008).
26. Osada, N. & Akashi, H. Mitochondrial–nuclear interactions and accelerated compensatory evolution: evidence from the primate cytochrome c oxidase complex. *Mol. Ecol. Evol.* **29**, 337–346 (2012).
27. Sloan, D. B., Havird, J. C. & Sharbrough, J. The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol. Ecol.* **26**, 2212–2236 (2017).
28. Burton, R. S., Pereira, R. J. & Barreto, F. S. Cytonuclear genomic interactions and hybrid breakdown. *Annu. Rev. Ecol. Evol. Syst.* **44**, 281–302 (2013).
29. Hill, G. E. Mitonuclear ecology. *Mol. Ecol. Evol.* **32**, 1917–1927 (2015).
30. Das, J. The role of mitochondrial respiration in physiological and evolutionary adaptation. *BioEssays* **28**, 890–901 (2006).
31. Stier, A. et al. Mitochondrial uncoupling as a regulator of life-history trajectories in birds: an experimental study in the zebra finch. *J. Exp. Biol.* **217**, 3579–3589 (2014).
32. Burton, R. S. & Barreto, F. S. A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Mol. Ecol.* **21**, 4942–4957 (2012).
33. Lindtke, D. & Buerkle, C. A. The genetic architecture of hybrid incompatibilities and their effect on barriers to introgression in secondary contact. *Evolution* **69**, 1987–2004 (2015).
34. Yeaman, S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl Acad. Sci. USA* **110**, E1743–E1751 (2013).
35. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
36. Morales, H. E., Sunnucks, P., Joseph, L. & Pavlova, A. Perpendicular axes of differentiation generated by mitochondrial introgression. *Mol. Ecol.* **26**, 3241–3255 (2017).
37. Pavlova, A. et al. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. *Evolution* **67**, 3412–3428 (2013).
38. Morales, H. E., Pavlova, A., Joseph, L. & Sunnucks, P. Positive and purifying selection in mitochondrial genomes of a bird with mitonuclear discordance. *Mol. Ecol.* **24**, 2820–2837 (2015).
39. Lamb, A. et al. Climate-driven mitochondrial selection: a test in Australian songbirds. *Mol. Ecol.* **27**, 898–918 (2018).
40. Beck, E. A., Thompson, A. C., Sharbrough, J., Brud, E. & Llopart, A. Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome *c* oxidase: a potential case of cytonuclear cointrogression. *Evolution* **69**, 1973–1986 (2015).
41. Debus, S. & Ford, H. Responses of eastern yellow robins *Eopsaltria australis* to translocation into vegetation remnants in a fragmented landscape. *Pac. Conserv. Biol.* **18**, 194–202 (2012).
42. Kilian, A. et al. Diversity arrays technology: a generic genome profiling technology on open platforms. *Methods Mol. Biol.* **888**, 67–89 (2012).
43. Hoban, S. et al. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am. Nat.* **188**, 379–397 (2016).
44. Warren, W. C. et al. The genome of a songbird. *Nature* **464**, 757–762 (2010).
45. Hofer, T., Foll, M. & Excoffier, L. Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genom.* **13**, 107 (2012).
46. Riley, L. G. et al. Mutation of the mitochondrial tyrosyl-tRNA synthetase gene, *YARS2*, causes myopathy, lactic acidosis, and sideroblastic anemia—MLASA syndrome. *Am. J. Human Genet.* **87**, 52–59 (2010).
47. Meiklejohn, C. D. et al. An incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in *Drosophila*. *PLoS Genet.* **9**, e1003238 (2013).
48. Yip, C. Y., Harbour, M. E., Jayawardena, K., Fearnley, I. M. & Sazanov, L. A. Evolution of respiratory complex I: "supernumerary" subunits are present in the alpha-proteobacterial enzyme. *J. Biol. Chem.* **286**, 5023–5033 (2011).
49. Angerer, H. et al. The LYR protein subunit NB4M/NDUFA6 of mitochondrial complex I anchors an acyl carrier protein and is essential for catalytic activity. *Proc. Natl Acad. Sci. USA* **111**, 5207–5212 (2014).
50. Fiedorczuk, K. et al. Atomic structure of the entire mammalian mitochondrial complex I. *Nature* **538**, 406–410 (2016).
51. Ostergaard, E. et al. Respiratory chain complex I deficiency due to *NDUFA12* mutations as a new cause of Leigh syndrome. *J. Med. Genet.* **48**, 737–740 (2011).
52. Zhu, J., Vinothkumar, K. R. & Hirst, J. Structure of mammalian respiratory complex I. *Nature* **536**, 354–358 (2016).
53. Kim, Y. & Nielsen, R. Linkage disequilibrium as a signature of selective sweeps. *Genetics* **167**, 1513–1524 (2004).
54. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
55. Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H. & Owens, G. L. Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol. Ecol.* **25**, 4488–4507 (2016).
56. Turner, T. L. & Hahn, M. W. Genomic islands of speciation or genomic islands and speciation? *Mol. Ecol.* **19**, 848–850 (2010).
57. Qvarnström, A., Ålund, M., McFarlane, S. E. & Sirkiä, P. M. Climate adaptation and speciation: particular focus on reproductive barriers in *Ficedula* flycatchers. *Evol. Appl.* **9**, 119–134 (2016).
58. Sunnucks, P., Morales, H. E., Lamb, A. M., Pavlova, A. & Greening, C. Integrative approaches for studying mitochondrial and nuclear genome co-evolution in oxidative phosphorylation. *Front. Genet.* **8**, 25 (2017).
59. Singhal, S. et al. Stable recombination hotspots in birds. *Science* **350**, 928–932 (2015).
60. Kawakami, T. et al. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Mol. Ecol.* **26**, 4158–4172 (2017).
61. Ellegren, H. et al. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**, 756–760 (2012).
62. Hooper, D. M. & Price, T. D. Chromosomal inversion differences correlate with range overlap in passerine birds. *Nat. Ecol. Evol.* **1**, 1526–1534 (2017).
63. Qvarnström, A. & Bailey, R. I. Speciation through evolution of sex-linked genes. *Heredity* **102**, 4–15 (2009).
64. Mank, J. E., Nam, K. & Ellegren, H. Faster-Z evolution is predominantly due to genetic drift. *Mol. Ecol. Evol.* **27**, 661–670 (2010).
65. Haldane, J. B. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12**, 101–109 (1922).
66. Beekman, M., Dowling, D. K. & Aanen, D. K. The costs of being male: are there sex-specific effects of uniparental mitochondrial inheritance? *Phil. Trans. R. Soc. B* **369**, 20130440 (2014).
67. Harrisson, K. A. et al. Fine-scale effects of habitat loss and fragmentation despite large-scale gene flow for some regionally declining woodland bird species. *Landsc. Ecol.* **27**, 813–827 (2012).
68. Hill, G. E.., & Johnson, J. D.. The mitonuclear compatibility hypothesis of sexual selection. *Proc. R. Soc. B* **280**, 20131314 (2013).
69. Morales, H. E. et al. Neutral and selective drivers of colour evolution in a widespread Australian passerine. *J. Biogeogr.* **44**, 522–536 (2017).
70. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
71. Hijmans R. J. *raster: Geographic Data Analysis and Modeling* R package version 2.3-12 (2014); http://www.rspatial.org/
72. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2014); http://www.R-project.org/
73. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 1 (2009).
74. Kawakami, T. et al. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* **23**, 4035–4058 (2014).
75. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
76. Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W. & Prodöhl, P. A. diveRsity: an R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods Ecol. Evol.* **4**, 782–788 (2013).
77. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population-structure. *Evolution* **38**, 1358–1370 (1984).
78. Villemereuil, P. & Gaggiotti, O. E. A new F_{ST}-based method to uncover local adaptation using environmental variables. *Methods Ecol. Evol.* **6**, 1248–1258 (2015).
79. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R News* **6**, 7–11 (2006).
80. Duforet-Frebourg, N., Luu, K., Laval, G., Bazin, E. & Blum, M. G. Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 Genomes data. *Mol. Ecol. Evol.* **33**, 1082–1093 (2016).
81. Harte, D. *Package 'HiddenMarkov': Hidden Markov Models* R package version 1.8-4 (2015); https://cran.r-project.org/web/packages/HiddenMarkov/index.html
82. Cunningham, F. et al. Ensembl 2015. *Nucleic Acids Res.* **43**, D662–D669 (2015).
83. Kanehisa, M.., & Goto, S.. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
84. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
85. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Human Genet.* **81**, 559–575 (2007).
86. Hill, W. & Weir, B. Variances and covariances of squared linkage disequilibria in finite populations. *Theor. Popul. Biol.* **33**, 54–78 (1988).
87. Marroni, F. et al. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (*CAD4*) gene. *Tree Genet. Genomes* **7**, 1011–1023 (2011).

88. Warnes, G. & Leisch, F. *Genetics: Population Genetics* R package version 1.1-5 (2005); https://cran.r-project.org/web/packages/genetics/index.html
89. Shin, J.-H., Blay, S., McNeney, B. & Graham, J. LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* **16**, 1–10 (2006).
90. Richards, E. J. & Martin, C. H. Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialist pupfishes. *PLoS Genet.* **13**, e1006919 (2017).
91. Wang, J. The computer program structure for assigning individuals to populations: easy to use but easier to misuse. *Mol. Ecol. Resour.* **17**, 981–990 (2017).
92. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191 (2015).

Acknowledgements

H.E.M. was supported by the Holsworth Wildlife Research Endowment (2012001942) and Stuart Leslie Bird Research Award from BirdLife Australia, PhD scholarships from Monash University and the Department of Public Education of the Mexican Government, and a Monash Postgraduate Publication Award. C.G. was supported by an ARC DECRA Fellowship (DE170100310). Other funding came from Monash internal sources. Genomic analyses were undertaken at the Monash High-Performance Computing facility and on the Albiorix computer cluster at the Department of Marine Sciences, University of Gothenburg. Field samples were collected under scientific research permits issued by the Victorian Department of Environment and Primary Industries (numbers 10007165, 10005919 and 10005514), New South Wales Office of Environment and Heritage (SL100886), in accordance with Animal Ethics approvals AM13-05, BSCI_2012_20 and BSCI_2007_07, using bands issued by the Australian Bird and Bat Banding Scheme. We are grateful to L. Joseph, R. Palmer, H. Sitters and C. Connelly for providing genetic samples. A. Gonçalves da Silva, D. Marques, S. Martin and V. Soria-Carrasco provided valuable inputs regarding data analysis, L. Joseph provided input on EYR evolution, and J. Wolf provided input on functional properties of the mitonuclear candidates. We thank S. Edwards, M. Webster, L. Kvistad and S. Falk for comments on earlier versions of the manuscript.

Author contributions

H.E.M., A.P. and P.S. conceived the project. H.E.M., A.P., N.A. and R.M. obtained the field samples. H.E.M. obtained the genetic data and performed the analyses. C.G. performed the protein structural analyses. H.E.M. wrote the paper with the help of A.P., C.G. and P.S. All co-authors read and approved the final version.

Competing interests

A.K. is employed by the commercial service provider that produced genome marker data for the paper.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0606-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.E.M.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Corresponding author(s): Hernan Morales (hern.moral@gmail.com

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Pipelines for DArT sequencing post-proccessing are proprietary: http://www.diversityarrays.com/dart-application-dartseq
Data analysis	All the programs we used are listed in the methods sections and custom scripts here: https://github.com/hmoral/EYR_DArT

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All datasets were deposited in Figshare doi: 10.6084/m9.figshare.5072143. (1) a table with information for every individual including geo-climatic data, ND2 sequences and DArT genotypes; (2) unfiltered SNP dataset with 97,070 SNP markers; and (3) BLAST results for reference genomes.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Evolutionary genomics study of 164 individual birds for which SNP genotypes were obtained
Research sample	Endemic Australian songbird, Eastern Yellow Robin, Eopsaltria australis
Sampling strategy	Samples were collected along geographical transects that intersecting the divergence process we wanted to measure. Two groups in each transect = four "populations". Collecting wild bird genetic samples is challenging due to ethics restrictions, lack of suitable habitat bird-territoriality and capture logistics. A first trip was made to identify sites with presence of target birds. Successive trips included miss-netting on each site, targeting at least four birds in each site. Sites were commonly proximate to each other (within 5-10 km). Considering that each bird pair has a territory of 1 ha this sampling density was enough to capture population-wide genetic diversity. After each trip we compared the sites collected against a habitat map (https://nvim.delwp.vic.gov.au/) and randomly target sites with different habitat types in following trips. We aimed to collect at least 20 individuals per population for the > 60,000 SNPs that we genotypified. These numbers are within the sample sizes and loci required to have full power in population genomics according to Morin et al. 2008; Mol. Ecol. Res; Robinson et al. 2014; BMC Evol Biol and Kalinowski 2005; Heredity
Data collection	DNA was extracted by HM, ND2 gene amplified by PCR by HM. DNA was shipped to diversity arrays (http://www.diversityarrays.com). Diversity arrays performed genotyping and provided the SNP table
Timing and spatial scale	Samples included
Data exclusions	Only samples that render significantly less SNPs than others were excluded. These samples were considered as failed during sequencing protocol.
Reproducibility	Alternative approaches were used to uncover outlier loci. All data and code has been deposited in public repositories
Randomization	Samples were randomly sampled in the field. Genetic markers were nearly-randomly sampled across the genome
Blinding	Not relevant

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Samples were taken from wild caught birds under a wide range of conditions
Location	Eastern Australia
Access and import/export	All animal ethics permits were obtained and we followed their guidelines: Field samples were collected under scientific research permits issued by the Victorian Department of Environment and Primary Industries (numbers 10007165, 10005919 and 10005514), New South Wales Office of Environment and Heritage (SL100886), in accordance with Animal Ethics approval AM13-05, using bands issued by the Australian Bird and Bat Banding Scheme.
Disturbance	Sampling is minimally invasive by capturing birds and taking a small blood sample, no animals were harmed

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	Blood samples and DNA extractions are stored at the School of Biological Sciences, Monash University and can be made available upon request