ARTICLES
https://doi.org/10.1038/s41559-017-0364-7
nature ecology & evolution

Mosaic genome evolution in a recent and rapid avian radiation

Katherine Faust Stryjewski* and Michael D. Sorenson*

Recent genomic analyses of evolutionary radiations suggest that ancestral or standing genetic variation may facilitate rapid diversification, particularly in cases involving convergence in ecological traits. Likewise, lateral transfer of alleles via hybridization may also facilitate adaptive convergence, but little is known about the role of ancestral variation in examples of explosive diversification that primarily involve the evolution of species recognition traits. Here, we show that genomic regions distinguishing sympatric species in an extraordinary radiation of small finches called munias (genus *Lonchura*) have phylogenetic histories that are discordant with each other, with the overall pattern of autosomal differentiation among species, and with sex-linked and mitochondrial components of the genome. Genome-wide data for 11 species sampled in Australia and Papua New Guinea indicate substantial autosomal introgression between sympatric species, but also identify a limited number of divergent autosomal regions, several of which overlap known colour genes (*ASIP*, *EDN3*, *IGSF11*, *KITLG*, *MC1R* and *SOX10*). Phylogenetic analysis of these outlier regions shows that different munia species have acquired unique combinations of alleles across a relatively small set of phenotypically relevant genes. Our results demonstrate that the recombination of ancestral genetic variation across multiple loci may be an important mechanism for generating phenotypic novelty and diversity.

Natural selection and recombination combine to produce heterogeneous patterns of genomic divergence between nascent and recently evolved species[1,2]. Particularly when there is an ongoing exchange of genes via hybridization, divergence may be limited to a tiny fraction of the genome underlying phenotypic differences and species identity[3–5]. An increasing number of whole-genome comparisons have identified genes and genomic regions with elevated divergence between recently evolved species[5–9], but pairwise comparisons reveal little about the origins of genetic variants in these regions. Extending genomic analyses to examples of explosive diversification[10] provides an opportunity to assess the role of ancestral genetic variation in speciation and phenotypic diversification[11–14]. While recent work has established that ancestral variation may promote adaptive radiation[11,15], particularly in cases involving the evolution of convergent phenotypes in multiple lineages[11–14], less is known about the role of ancestral variants in cases of rapid diversification that primarily involve the evolution of phenotypes involved in sexual selection and species recognition.

The genus *Lonchura* (family Estrildidae) includes one of the most extraordinary examples of recent and rapid radiation in birds (Supplementary Table 1). Diversifying from a common ancestor over the past ~0.5 million years or less (Supplementary Fig. 1), 13 munia species in Australia, New Guinea and nearby islands have evolved discrete and unambiguous differences in plumage colour and pattern (Fig. 1a); in contrast, evidence of ecological divergence is limited, although populations of three species (*Lonchura grandis*, *Lonchura forbesi* and *Lonchura melaena*) have larger bills. The clade is also notable for the coexistence of several pairs or trios of species in broad sympatry (Fig. 1a), fulfilling a key criterion for biological species that more typically requires millions of years to achieve[16]. Among birds, perhaps only the southern capuchinos (genus *Sporophila*) are comparable in all these respects[17,18].

Here, we examine patterns of genomic divergence in munias using robust restriction-site-associated DNA sequencing (RAD-seq) and complete mitochondrial genome datasets along with low-coverage whole-genome sequencing of individuals from 18 populations representing 11 of the 13 species in the *Lonchura* radiation. Our analyses indicate extensive autosomal introgression between sympatric species, but also identify divergent portions of the genome, variably including mitochondrial DNA (mtDNA), a portion of the Z chromosome that likely represents a major inversion, and a relatively small set of mostly narrow autosomal outlier regions, including several with known colour genes. Phylogenetic analysis of these regions indicates that each is characterized by a limited number of divergent allelic lineages, the sharing of which among species is often discordant both with genome-wide relationships and among outlier loci, generating a mosaic pattern of genetic similarity and divergence across species and loci. Our results suggest that differential selection on ancestral genetic variation and lateral transfer of alleles via introgression have contributed to the phenotypic diversification of the *Lonchura* munias by generating unique combinations of alleles across a relatively small set of phenotypically relevant genes.

Results

Analyses of mitochondrial genomes (Fig. 1b) and genome-wide RAD-seq data (Fig. 2) confirm recent and rapid speciation in munias, but also reveal incongruent relationships that suggest extensive autosomal admixture between sympatric species. With the exception of two *L. grandis* populations, which are relatively divergent from each other and all other sampled species, STRUCTURE analyses based on autosomal RAD-seq data group populations first by geographic region and begin to discriminate sympatric species only at higher values of k (number of populations), although generally with a potential signal of admixture (Fig. 2a and Supplementary Fig. 2). Indeed, most sampled populations are genetically most similar to a sympatric population of another munia species (Fig. 2), with pairwise divergence varying from essentially zero ($\Phi_{ST}=0.007$ for *Lonchura castaneothorax* versus *Lonchuraflaviprymna* in Western Australia, where Φ_{ST}, an analogue of F_{ST}, is a DNA sequence-based measure of population differentiation resulting from analysis of molecular variance) to moderate (maximum sympatric $\Phi_{ST}=0.24$ for *L. castaneothorax* versus *L. grandis* in Milne Bay) (Supplementary Figs. 3 and 4

Department of Biology, Boston University, Boston, MA 02215, USA. *e-mail: stryjewski@fas.harvard.edu; msoren@bu.edu

Table 1 | Genomic outlier regions

Chr	Start	End	Length	Number of SNPs with $F_{ST} > 0.8$ in four focal comparisons				Gene(s)
				WA	C	MD	TF	
1	93,368,000	93,394,000	26,000	0	0	19*	6*	*IGSF11**
1	97,906,000	97,924,000	18,000	0	32*	13*	9*	*POLD3*
1A	42,210,000	42,392,000	182,000	0	305*	1	3	*TMTC3, KITLG**
1A	50,782,000	50,810,000	28,000	0	15*	0	0	*SOX10**
1A	72,376,000	72,616,000	240,000	23*	7*	4*	20*	*(DDX11), CCDC91, (PTHLH)*
2	26,566,000	26,598,000	32,000	16*	0	1	8*	*SLC25A13*
2	78,752,000	80,860,000	2,108,000	7	119*	141*	1	*CDH18, BASP1, MYO10, FAM134B, ZNF622, MARCH11, FBXL7, ANKH, OTULIN*
3	34,918,000	35,720,000	802,000	0	0	1	0	*CRIM1, FEZ2, VIT, STRN*
5	32,380,000	32,420,000	40,000	0	6*	25*	0	*(STXBP6, NOVA1)*
6	30,418,000	30,536,000	118,000	18*	73*	38*	35*	*GRK4*
11	10,212,000	10,324,000	130,000	0	1	1	62*	*(NKD1, GSE1)*
11	11,642,000	11,652,000	10,000	1*	3*	3*	0	*TCF25, MC1R*, DBNDD1*
18	1,992,000	2,024,000	32,000	0	1	8*	4	*SLC16A3, CSNK1D*
18	3,812,000	4,020,000	208,000	0	0	0	0	*COX10, HS3ST3*
20	1,820,000	1,928,000	108,000	26*	127*	28*	85*	*AHCY, ASIP**
20	11,986,000	12,038,000	52,000	0	24*	2	19*	*EDN3*, ZNF831*
24	3,866,000	4,008,000	142,000	0	1	16*	34*	*ST14, NFRKB, PRDM10, APLP2*
26	3,058,000	3,114,000	56,000	0	19*	4*	2	*KIAA1324, SARS, CELSR2, PPIL1*
27	3,384,000	3,396,000	12,000	0	5*	8*	0	*MRC2, TLK2*
28	580,000	650,000	70,000	0	2	0	37*	*SLC1A6, CERS4, MARCH2, RAB11B, ANGPTL4, KANK3, RPS28, ADMP*
Z	28,000,000	48,000,000	20,000,000	390	42,940*	6,779*	1,995*	>100 genes, including *SLC45A2** and *FST**

Regions of elevated divergence (F_{ST}) identified in one or more of four focal comparisons of sympatric species pairs and/or in the sliding-window phylogenetic analysis. The column 'Chr' shows the chromosome number in the zebra finch reference genome, 'Start' and 'End' mark the approximate extent of the region with elevated F_{ST} and/or Bs values and the region from which polymorphisms were extracted for PCA and phylogenetic analyses (see Supplementary Figs. 7–26 and 28). The number of SNPs with $F_{ST} > 0.8$ in each genomic region is shown for each of four pairwise comparisons (Western Australia (WA), Central Province, Papua New Guinea (C), Madang Province, Papua New Guinea (MD) and Trans Fly, Papua New Guinea (TF)); regions ultimately scored as divergent between species (fixed or nearly fixed for alternative alleles) are indicated by an asterisk. Note that the number of SNPs depends in part on both the size of the genomic region and average sequencing depth and coverage, which were somewhat lower in the WA and MD comparisons. Annotated vertebrate genes overlapping each genomic region are listed. Genes just outside the analysed region are listed in parentheses. Asterisks indicate genes associated with colouration in other vertebrates (*IGFS11*[73], *KITLG*[74–77], *SOX10*[27,74,78–80], *MC1R*[74,81–86], *ASIP*[5,74,87–92], *EDN3*[74,93–95], *SLC45A2*[8,27,96,97] and *FST*[5]). We also note four additional genes associated with bill size (*CRIM1*[98,99]), bone development (*PTHLH*[100] and *MRC2*[101]) and physiological adaptation (*POLD3*[102,103]).

and Supplementary Table 2). In contrast, genetically similar sympatric species are in some cases entirely distinct in their mtDNA (for example, *Lonchura nevermanni* and *Lonchura stygia* in the Trans Fly and *L. castaneothorax* and *Lonchura spectabilis* in Madang; Figs. 1b and 2a)—a result consistent with low fitness in hybrid females, in accordance with Haldane's rule[19], impeding the introgression of maternally inherited mtDNA. Similarly, sympatric populations in mainland New Guinea (Central, Milne Bay and Madang) are highly differentiated at RAD-seq loci mapping to an ~20 megabase pair (Mbp) region in the centre of the Z chromosome, resulting in a pattern of genetic structure that is incongruent with both mtDNA and the overall pattern for autosomal loci (Fig. 2a).

Autosomal admixture between sympatric species is further indicated by direct evidence of hybridization, analysis of rare single-nucleotide polymorphisms (SNPs) and estimates of co-ancestry based on recent coalescent events. Hybrid munias are infrequently encountered in the field, but are reported in the ornithological literature[20,21], and our sample includes two individuals of recent mixed ancestry. Genetic data confirm that a putative *L. castaneothorax* × *L. grandis* hybrid male captured in Central Province, Papua New Guinea is an F1 hybrid (Fig. 2 and Supplementary Figs. 2 and 4). Of course, introgression occurs only if hybrids survive and reproduce; another individual from Madang was identified as *L. grandis* based on its phenotype, but clearly had some *L. castaneothorax* ancestry, consistent with a backcross to *L. grandis* (Fig. 2 and Supplementary Figs. 2 and 4).

Interestingly, hybridization between *L. flaviprymna* and *L. castaneothorax* in Western Australia—the least divergent species pair in our study ($\Phi_{ST} = 0.007$)—was described as common in the 1950s, involving 10% of breeding pairs[20]. In 2010, however, we detected no obvious hybrids among over 100 birds captured in the same region and principal component analysis (PCA) of RAD-seq polymorphisms separates these populations into discrete clusters with no evidence of intermediate individuals (Supplementary Fig. 4). Individuals of sympatric *L. stygia* and *L. nevermanni* also cluster into discrete populations based on autosomal data, but show some evidence of Z chromosome admixture (Fig. 2a), implying that introgression occurred at some point in the past. Thus, hybridization may be episodic, perhaps occurring at a relatively high rate following secondary contact and then declining in frequency via reinforcement of reproductive isolation[22].

The distribution of rare SNPs among populations provides one test for recent introgression in sympatry. Shared ancestral polymorphism among recently evolved species is expected, but the rarest alleles, which are of more recent origin on average[23], should be restricted to a single population or species in the absence of recent gene flow[24]. In the munias, we find that rare alleles are more often shared between sympatric populations than between allopatric

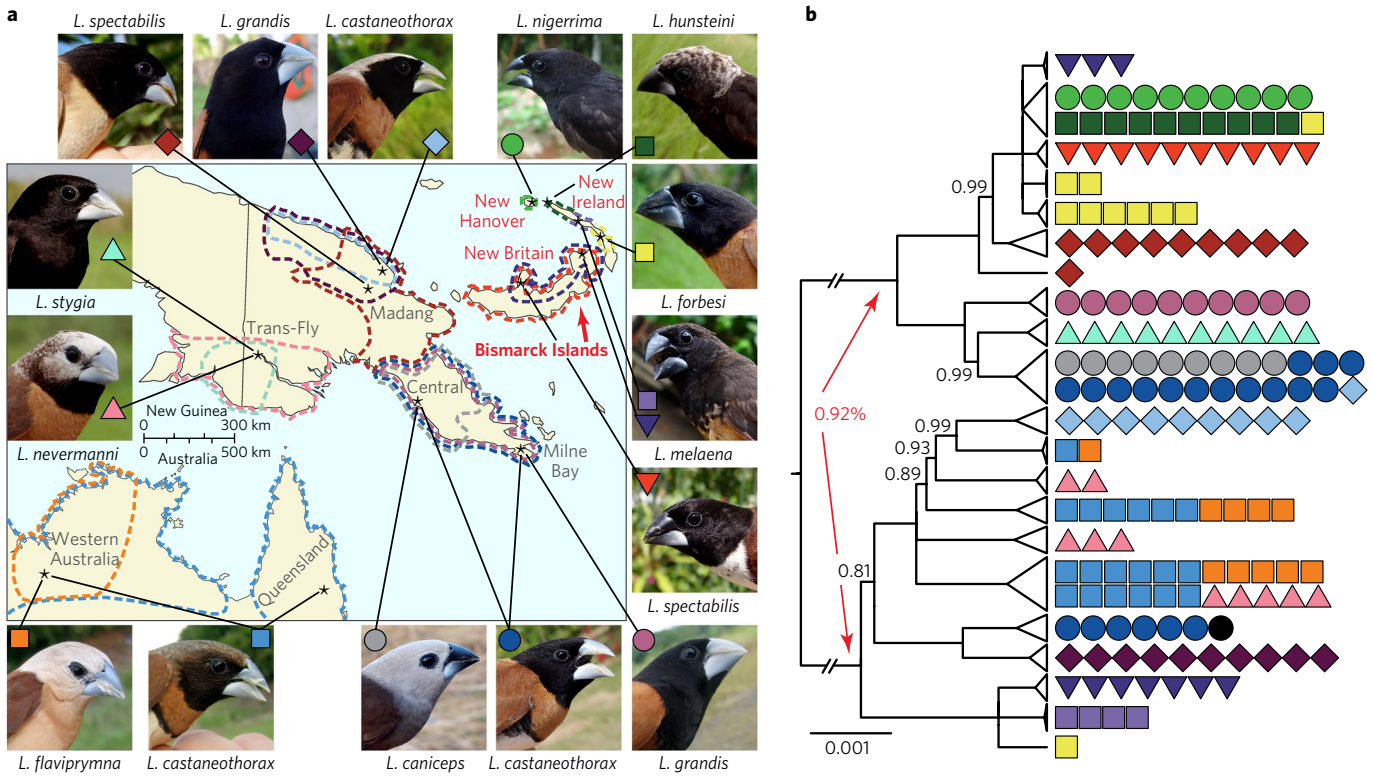

Fig. 1 | Geographic distributions and mtDNA phylogeny. a, Approximate distributions of the 11 *Lonchura* species sampled in this study, with collection localities indicated by asterisks. Note that Australia is drawn at 60% scale relative to New Guinea. Genetic and phenotypic divergence was minimal between *L. castaneothorax* populations in Central and Milne Bay Provinces in Papua New Guinea, and also between two *L. castaneothorax* populations in Australia, so these pairs of populations share the same symbols and are combined in some analyses. **b**, Phylogeny based on complete mtDNA sequences for 173 birds representing 11 species and 18 populations. Closely related haplotypes are collapsed to simplify the figure. Individual birds are indicated by colour-coded symbols as in **a**. One *L. castaneothorax* × *L. grandis* hybrid (black circle) from Central Province, Papua New Guinea, is included. Posterior probabilities are indicated for clades with values of <1. The average sequence divergence between the two main clades is only 0.92%.

populations, including conspecific populations in different regions (Fig. 2c and Supplementary Fig. 5). This pattern includes the two *L. grandis* populations, which are relatively divergent from other species, but are nonetheless more likely to share rare SNPs with sympatric populations of other species than with each other (Fig. 2c and Supplementary Fig. 5).

ThefineRADstructure[25] software package—a conceptually similar approach that measures recent co-ancestry by focussing on the most recent coalescent events among the haplotypes at each locus—produced similar results while also revealing interesting signals of mixed ancestry (Fig. 3). For example, the two populations of *L. grandis* represent the only exception to the general pattern of highest co-ancestry between sympatric populations, but *L. grandis* in Madang has higher co-ancestry with sympatric populations of *L. castaneothorax* and *L. spectabilis* than with any other 'non-*grandis*' populations. Likewise, *L. castaneothorax* and *L. spectabilis* in Madang have higher recent co-ancestry with sympatric *L. grandis* than with allopatric *L. grandis*. As is more typical of other sympatric comparisons, *L. castaneothorax* and *L. spectabilis* in Madang have their highest recent co-ancestry with each other, but *L. castaneothorax* in Madang also has relatively high co-ancestry with the remaining *L. castaneothorax* populations in other regions, whereas *L. spectabilis* has relatively high co-ancestry with allopatric *Lonchura caniceps*. These results indicate substantial admixture between sympatric populations, but also suggest that genome-wide patterns of genetic similarity and divergence retain some information about older historical relationships among species and populations. More broadly, all the above results suggest that hybridization has had a homogenizing effect on the genomes of sympatric munias, but has occurred at a frequency insufficient to impede or reverse divergence at the loci responsible for phenotypic differences between species.

To identify genomic regions associated with phenotypic diversification and examine the phylogenetic history of these regions, we generated low-depth (~1.6× per individual) whole-genome sequencing data for 176 birds, representing all 18 sampled populations (Fig. 1), plus one outgroup sample (*Lonchura leucosticta*). After alignment to the zebra finch *Taeniopygia guttata* reference genome[26], we calculated F_{ST} (fixation index) between sympatric species for 10- and 100-kilobase pair (kbp) sliding windows. Focusing on four pairwise comparisons of sympatric species with minimal genome-wide divergence (Fig. 2b), we analysed 20 regions of elevated divergence scattered across the autosomal genome (Supplementary Fig. 6). Most of these regions were relatively narrow (for example, 10^4–10^5 bp) and many were identified in more than one pairwise comparison, often with peaks of maximal divergence in nearly identical genomic locations, which in most cases fall within or overlap one or more annotated genes (Supplementary Figs. 7–26). Notably, these include six regions encompassing genes involved in melanogenesis or melanocyte development (*ASIP*, *EDN3*, *IGSF11*, *KITLG*, *MC1R* and *SOX10*), all of which have been associated with colour phenotypes in other animals, as well as genes involved in bone development (*CRIM1*, *MRC2* and *PTHLH*) and cold adaptation (*POLD3*) (Table 1).

A sliding-window phylogenetic analysis indicates that species/population-level relationships within these outlier regions are often discordant with the genome-wide pattern (Fig. 4 and Supplementary Fig. 6). For example, *L. caniceps* from Central Province, Papua New Guinea is divergent from sympatric *L. castaneothorax* at the

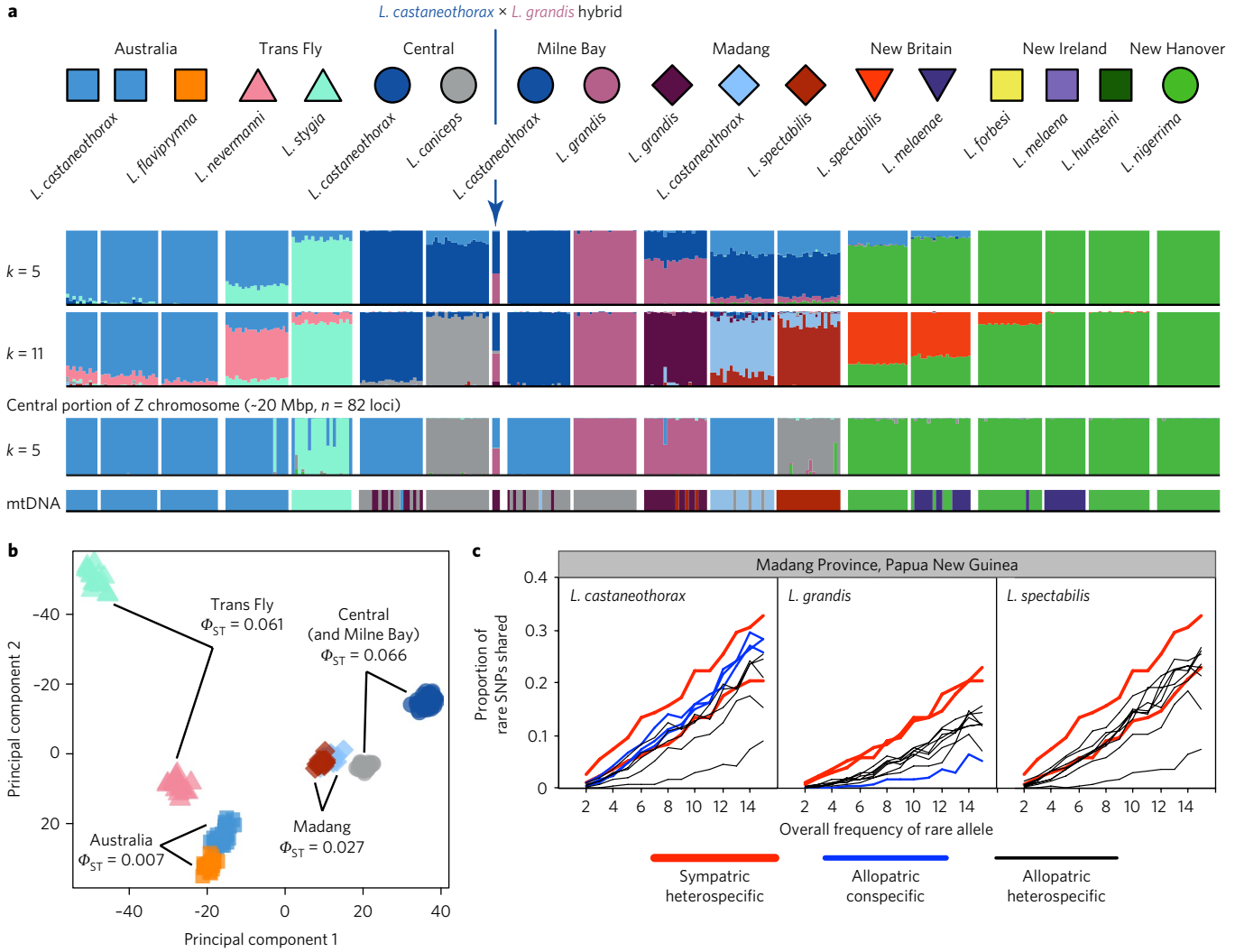

Fig. 2 | Results of ddRAD-seq analyses. a, STRUCTURE results for $k=5$ and $k=11$ populations based on 6,759 autosomal loci, and for $k=5$ populations based on 82 loci mapping to the central portion of the Z chromosome (positions 28 M to 48 M in the zebra finch reference genome[26]). For comparison, the mtDNA plot was constructed manually based on the results from Fig. 1b augmented by NADH dehydrogenase subunit 2 (ND2) gene sequences for additional samples. The total sample size was 336 individuals, including 18–20 for most populations. The bar representing an *L. castaneothorax* × *L. grandis* F1 hybrid is doubled in width to improve visibility. **b**, PCA of genetic diversity at autosomal loci for a set of closely related species/populations in Australia and mainland Papua New Guinea (analysis excludes two relatively divergent populations of *L. grandis* and species/populations in the Bismarck Islands). Pairwise Φ_{ST} values for sympatric populations are indicated. **c**, Sharing of rare SNPs among three sympatric species in Madang Province compared with sharing with allopatric populations in Australia and elsewhere in mainland Papua New Guinea. Note that this analysis includes only one population of *L. spectabilis*, so there is no 'allopatric conspecific' comparison for this population. See Supplementary Figs. 2, 4 and 5 for additional STRUCTURE, PCA and rare SNP results.

KITLG gene, but shares its divergent allelic lineage and largely black bill with species/populations in the Bismarck Islands (Fig. 4c and Supplementary Fig. 9), suggesting that variation in or near *KITLG* underlies this derived phenotype. A different pattern is observed at *ASIP*—another well-known colour gene—at which *L. caniceps*, *L. nevermanni* and *L. flaviprymna* are highly divergent from their respective sympatric congeners, but share similar haplotypes with each other and with the Bismarck Islands populations (Fig. 4c). In addition, *ASIP* distinguishes the five populations of *L. castaneothorax* from all others—a pattern observed in only two other autosomal regions (Supplementary Figs. 11 and 18), one of which is *MC1R* (another colour gene).

As in the above examples, the remaining outlier loci have only two to four distinct allelic lineages per locus (Fig. 5). This includes a few loci at which a novel allele appears to have swept to fixation in a single species (for example, Supplementary Fig. 22). Thus, some outlier loci may have three or four functionally distinct allelic lineages, but the number of divergent lineages per locus is always smaller than the number of species in our analysis. Moreover, different subsets of populations share similar alleles at different loci, such that each species has a unique combination of alleles across the set of outlier loci (Fig. 5b). We also note that outgroup sequences from *L. leucosticta* are consistently divergent from all ingroup alleles at all outlier loci (Supplementary Fig. 27), suggesting that functionally relevant variants originated within the focal clade and were not acquired from more distantly related *Lonchura* species.

Other than the potential link between *KITLG* and bill colour (see above), we did not detect obvious associations between individual loci and specific phenotypic traits, but this is not necessarily expected given the likelihood of epistatic and combinatorial effects among loci[27]. Likewise, epistasis and simple Mendelian dominance[5] may allow for polymorphism at a subset of loci in each species,

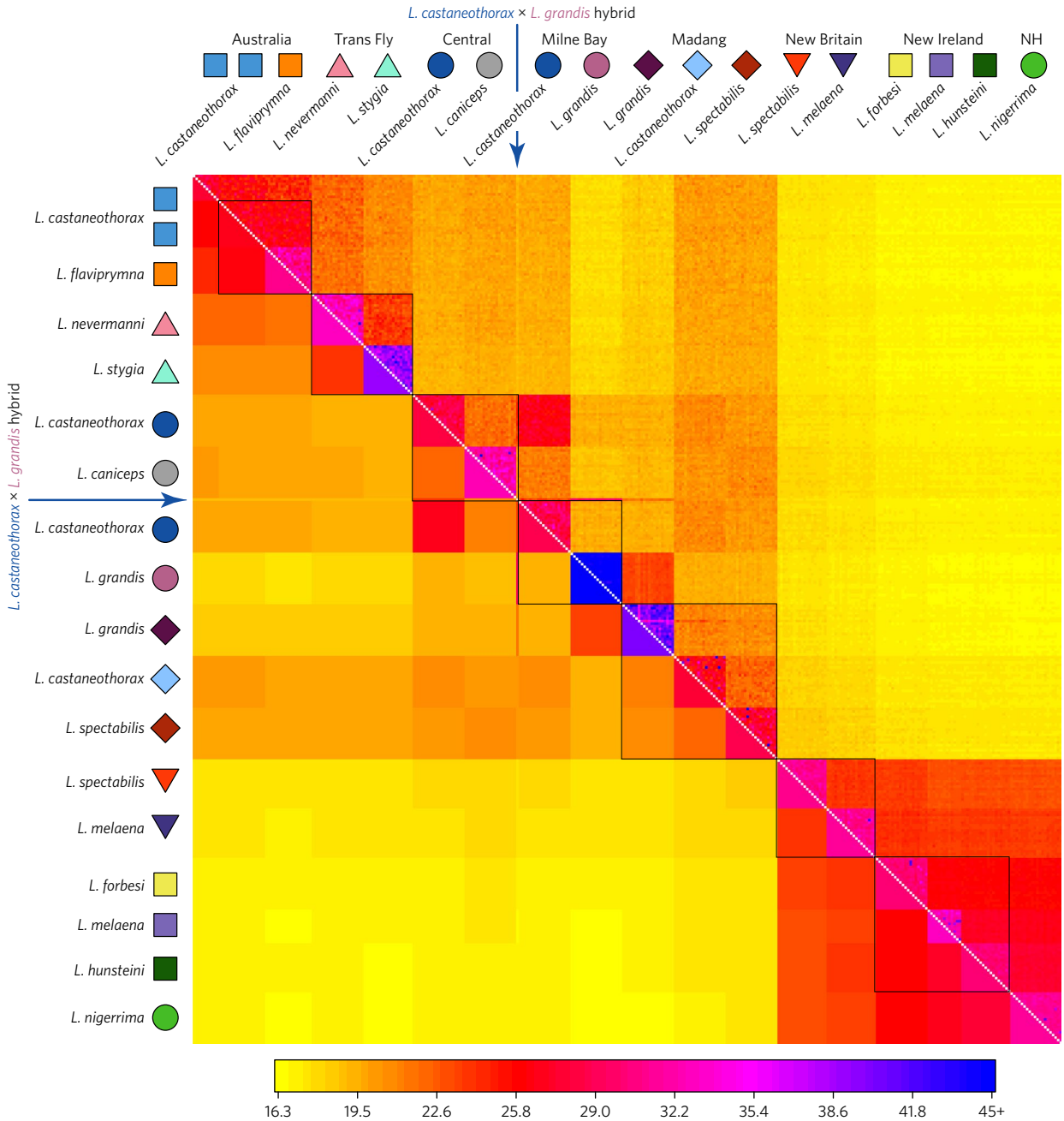

Fig. 3 | Co-ancestry matrix from fineRADstructure. Co-ancestry coefficients between individual munia samples are plotted above the diagonal; average coefficients within and among populations are plotted below. Sympatric populations are enclosed in black boxes, but note that populations of *L. castaneothorax* from Central and Milne Bay Provinces represent two samples from one geographically contiguous population and are thus effectively sympatric. Note also that fineRADstructure achieves greater resolution of subtle population structure than STRUCTURE, with individuals from the six New Britain and New Ireland populations discernible as discrete clusters.

potentially facilitating the lateral transfer of alleles among populations. As a simple example, a major locus producing the nearly all-black plumage of *L. stygia* could make variation at other colour and patterning loci effectively neutral, allowing ongoing segregation of alleles that would have phenotypic effects in other genetic backgrounds.

Confirming RAD-seq results, whole-genome sequencing shows that a large portion of the Z chromosome conflicts with genome-wide autosomal relationships (Fig. 4d and Supplementary Fig. 28). Uniformly high divergence across this region (Fig. 4e) and evidence of divergent Z chromosomes segregating within *L. stygia* (Supplementary Fig. 29) suggest that an ~20-Mbp inversion suppresses recombination in this region.

Discussion

Following the approach of other recent studies[4–8,13,18,28–30], we identified genes and genomic regions likely to be involved in the phenotypic divergence of closely related species by scanning the genome for regions of elevated divergence as measured by F_{ST}. These 'islands of divergence' are expected whenever closely related populations experience divergent selection, particularly if ongoing gene flow impedes genome-wide divergence[3–5]. Several recent

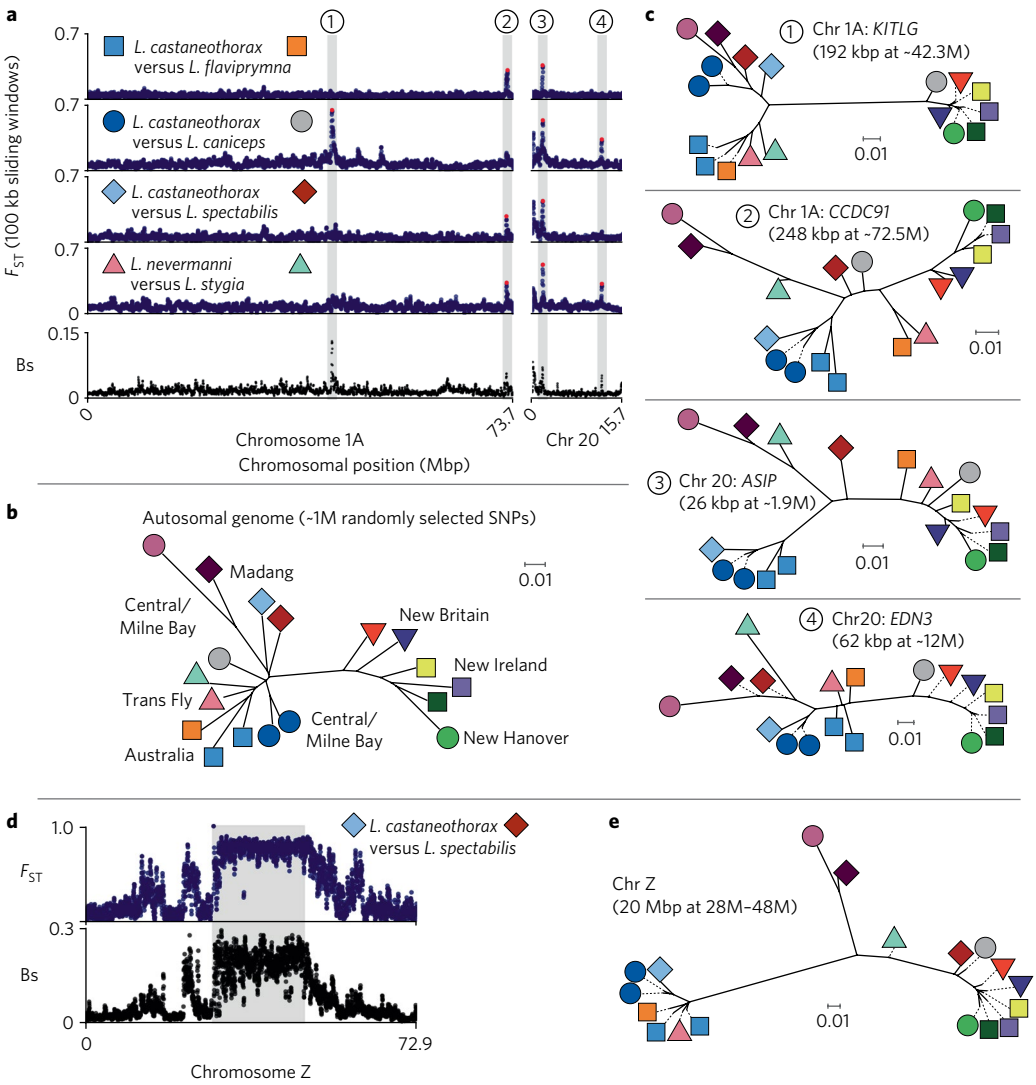

Fig. 4 | Illustration of phylogenetic heterogeneity among genomic outlier regions. a, Sliding-window plots of divergence (F_{ST}) between four pairs of sympatric populations (top four plots) and phylogenetic branch score (Bs), measuring discordance between local and genome-wide trees for all 18 populations (bottom graph). A single F_{ST} or Bs value was calculated for each 100-kbp window, sliding by 20 kbp. Results are shown for selected chromosomes (see Supplementary Fig. 6 for the entire genome). **b**, Population-level neighbour-joining tree based on pairwise net divergence (d_a) values across ~1 million randomly selected autosomal SNPs. **c**, Population-level neighbour-joining trees for the four regions of elevated divergence highlighted in **a**. Three of these four regions overlap genes associated with colouration in other animals (*KITLG*, *ASIP* and *EDN3*). See Supplementary Figs. 7–26 and 28 for additional detail on these and other genomic regions with elevated divergence. **d**, Z chromosome sliding-window plots, including F_{ST} for one pairwise comparison and Bs for all populations, comparing local 100-kbp trees with the overall autosomal genome tree. **e**, Population-level tree for the 20-Mbp region highlighted in grey in **d**. Chr, chromosome.

papers have pointed out that similar patterns can be generated by selective sweeps occurring after the cessation of gene flow[31], which may involve genes unrelated to speciation and reproductive isolation, or by processes unrelated to adaptive divergence such as background or linked selection in regions of reduced recombination such as centromeres[30,32–35].

Several considerations suggest that these alternative processes do not account for the outlier regions we identified. First, the species in our analysis diversified recently, allowing little time for the relatively slow process of divergence via background selection. Second, most divergence peaks we identified are relatively narrow and are not associated with centromeres or other regions of reduced genetic diversity. Finally, and most importantly, the mosaic patterns of divergence (or lack thereof) among *Lonchura* species across multiple outlier loci are inconsistent with a history of recurrent divergence in independent lineages. For example, the largest autosomal outlier region we identified overlaps the chromosome 2 centromere, where reduced recombination has likely accentuated divergence. Nevertheless, there are only four divergent lineages in this region among 11 *Lonchura* species (Supplementary Fig. 13), one of which appears to be a recombinant version of the other three (Supplementary Fig. 30). In some cases, sympatric species pairs share closely related haplotypes across the chromosome 2 centromere (for example, *L. castaneothorax* and *L. flaviprymna, and L. stygia* and *L. nevermanni*), despite being divergent at other outlier loci (Fig. 5b); conversely, species with divergent chromosome 2 haplotypes share similar alleles at other outlier loci and show little evidence of divergence at other centromeres (Supplementary Fig. 6 and Supplementary Figs. 31 and 32).

The lack of Z chromosome and/or mtDNA admixture between several pairs of sympatric munias is potentially consistent with the idea that inter-locus genetic incompatibilities, including cyto-nuclear

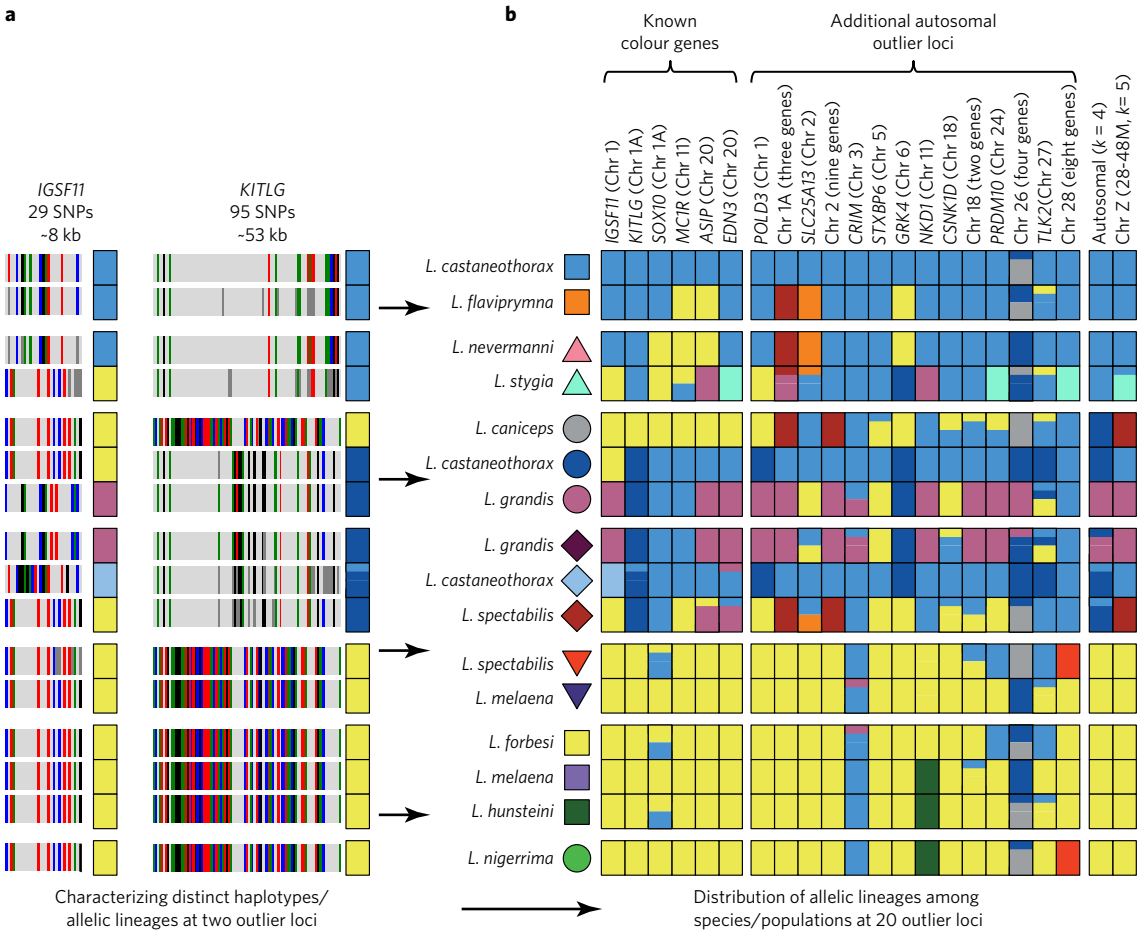

Fig. 5 | Mosaic distribution of alleles at outlier loci. a, Two examples of characterizing distinct haplotypes/allelic lineages at outlier loci. The left column for each gene plots consensus nucleotides that differ from the zebra finch reference genome for each population (A = green; C = blue; G = black; T = red; dark grey = polymorphic, common allele frequency <0.7). The right column for each gene categorizes and colour codes each population based on its most common haplotype. The plots include biallelic SNPs with large frequency differences ($\Delta \geq 0.8$) between groups of populations identified by PCA and population-level trees for each region (Fig. 4 and Supplementary Figs. 7–26 and 28). Note that there are a limited number of distinct alleles/haplotypes at each locus and that the pattern of shared alleles among populations is strikingly different at these two loci. **b**, Comparable results for 20 autosomal outlier regions. Six known colour genes are shown on the left and, for comparison, the last two columns provide a summary of genetic similarity and divergence for the autosomal genome overall and for a putative ~20-Mbp inversion in the centre of the Z chromosome. Populations that are polymorphic at a given locus (most common haplotype at less than ~80% frequency) are coded accordingly. Chr, chromosome.

incompatibilities, contribute to speciation by reducing hybrid fitness and generating divergent sexual selection on plumage patterns and mate preferences[36]. For example, *L. castaneothorax* and *L. spectabilis* in Madang Province are minimally divergent across the autosomal genome ($\Phi_{ST} = 0.027$), but remain distinct in mtDNA and are fixed for alternative states of the Z chromosome inversion. Other comparisons, however, are inconsistent with a specific role for the Z chromosome inversion in cyto-nuclear incompatibility because there is no consistent association with mtDNA (Supplementary Fig. 2). For example, *L. castaneothorax* and *L. caniceps* in Central Province are also fixed for alternative states of the Z chromosome inversion, but share closely related mtDNA haplotypes. Conversely, *L. stygia* is polymorphic for the inversion, but remains distinct from sympatric *L. nevemanni* in mtDNA. At least two additional genes associated with plumage colour in birds (*FST* (ref. [5]) and *SLC45A2* (ref. [18,27])) map to the central portion of the Z chromosome, raising the simple alternative that divergence in this region is maintained as it is at any other locus with alternative alleles controlling a phenotypic trait. Nonetheless, the lack of mitochondrial introgression between some sympatric munias remains intriguing and potentially consistent with Haldane's rule[19], although other explanations have been suggested[37].

The mosaic pattern of genomic divergence in munias, in which the phylogenetic histories of putatively functional loci are largely decoupled from overall patterns of autosomal divergence, probably reflects a complex history of speciation that included episodes of dispersal, allopatric divergence, secondary contact, genome-wide introgression and reinforcement. While the full history of the *Lonchura* radiation will be difficult to reconstruct, some biogeographic patterns are evident in our results. For example, *L. nevermanni* in the Trans Fly region of southern New Guinea is genetically similar to *L. flaviprymna* in Western Australia, both genome wide and at most outlier loci (Fig. 5b), suggesting allopatric divergence from a recent common ancestor. In contrast, sympatric *L. stygia* is divergent from *L. nevermanni* at numerous outlier loci, including several at which *L. stygia* shares alleles with mainland Papua New Guinea populations of *L. grandis* and/or *L. castaneothorax*. Thus, *L. nevermanni* and *L. stygia* probably originated from different ancestral source populations that colonized the Trans Fly region independently. Similarly, allopatric populations of *L. spectabilis*

and *L. caniceps* in mainland Papua New Guinea share similar alleles at several outlier loci, including the Z chromosome inversion and a broad region of divergence overlapping the chromosome 2 centromere, and are more likely to share closely related alleles with species/populations in the Bismarck Islands than are mainland populations of *L. castaneothorax*. Thus, colonization of the Bismarck Islands may have occurred before the expansion of *L. castaneothorax* populations across mainland Papua New Guinea.

The precise history of allelic variants at individual outlier loci is also difficult to reconstruct, but differential selection on retained ancestral polymorphisms and/or lateral transfer of adaptive alleles via introgression must have been involved in generating the mosaic patterns we observed. These processes comprise two forms of 'collateral evolution'[38], defined as the parallel evolution of ancestral genetic variants in independent lineages and recognized as an important mechanism for convergent evolution[14,39,40]. In Darwin's finches, for example, ancestral alleles at two loci are associated with changes in bill morphology across multiple species[13,28]. Likewise in the munias, ancestral alleles may underlie convergent components of each species' unique phenotype, but we suggest that collateral evolution also contributed to phenotypic diversification by generating new combinations of alleles across a relatively small set of potentially interacting colour genes and other functionally relevant loci. The role of ancestral variation and collateral evolution in producing phenotypic novelty and diversity may be under-appreciated.

Methods

Sample collection. Samples from 17 munia populations representing 11 of the 13 species comprising the focal clade were collected in Australia and Papua New Guinea from 2010 to 2012 (Supplementary Table 3). We focused on regions in which two or more species occur in sympatry, including Western Australia and seven Papua New Guinea provinces—Milne Bay, Madang, Central, East New Britain, West New Britain, New Ireland and Western (Fig. 1). Pectoral muscle from birds prepared as museum specimens (up to 20 per population) was stored in dimethyl sulfoxide buffer[41]. Blood samples from birds that were captured and released (up to 30 additional per population) were preserved on Whatman FTA cards. Finally, ten tissue samples of *L. castaneothorax* from Queensland, Australia were obtained from the Australian National Wildlife Collection at the Commonwealth Scientific and Industrial Research Organisation.

Double-digest RAD-seq (ddRAD-seq). Genomic DNA was extracted from muscle tissue or blood samples using a DNeasy Kit (Qiagen) with 4 µl RNAse added to the lysate following incubation. RAD-seq was completed for 336 samples using the double-digest method described in ref. [42], with the following modifications: the concentration of ligated fragments was quantified for each sample by quantitative polymerase chain reaction, after which sets of 12 samples were pooled in equimolar concentration, thereby reducing the time and cost associated with the rest of the protocol. After size-selection, the pooled libraries were amplified by polymerase chain reaction for 22 cycles using Phusion High-Fidelity DNA Polymerase and quantified again using quantitative polymerase chain reaction. Multiple 12 sample pools were then combined in equimolar amounts for sequencing on an Illumina HiSeq 2500 system using two-lane flow cells in RAPID mode; 151-base pair (bp) single-end reads were obtained for all samples. After filtering low-quality singletons, the number of sequence reads per sample averaged ~1.4 million, ranging from ~683,000 to 3.96 million.

The sequence data were processed using a combination of custom Python scripts and publicly available software as described in ref. [42]; current versions of the code are available at https://github.com/BU-RAD-seq. Analyses were based on a set of 6,759 autosomal and 284 Z-linked loci, each of which had a median per-sample sequencing depth of at least 20, missing data for less than 5% of individuals and 'flagged' genotypes in less than 2% of individuals. Flagged genotypes included putative heterozygotes in which one allele accounted for <29% of reads and/or in which 'extra' reads were inconsistent with two primary alleles (see ref. [42] for more details). A large number of 'flagged' genotypes at a putative locus is indicative of sequences from paralogous loci being incorrectly clustered together. Z-linked loci were identified as those with a roughly 2:1 ratio of average sequencing depth in males versus females (Supplementary Fig. 33). In subsequent analyses, flagged genotypes were scored as missing and one allele was scored as missing for low-depth genotypes. Each unique indel, regardless of length, was scored as a single 0/1 polymorphism. Finally, we used Geneious (version 8.0.3; https://www.geneious.com) to examine and manually adjust the alignments of loci with five or more unique indels, two or more correlated SNPs in the last five bases (often indicative of misalignment at the end of the locus) and/or four or more perfectly correlated SNPs at any position within the locus (potentially due either to misalignment or not-yet-detected sequences from paralogous loci). End-of-locus adjustments were guided by comparison with the zebra finch *Taeniopygia gutatta* genome[26] where possible. Eight loci were discarded due to ambiguous alignment of repetitive elements.

Overall, less than 1% of the ddRAD-seq data matrix was represented by missing data (0.28%), low-depth genotypes (< 5 sequence reads; 0.35%) or flagged genotypes (0.16%), whereas the median sequencing depth was 98 reads per sample per locus, generally allowing genotypes to be scored unambiguously. The robust genotypic (and haplotypic) information provided by ddRAD-seq was ideal for characterizing genome-wide patterns of population structure, particularly for analyses based on rare SNPs and recent coalescent events (see below).

Genome-wide patterns of genetic diversity and divergence among the 18 munia populations were assessed using analysis of molecular variance (AMOVA)[43], STRUCTURE[44], fineRADstructure[25] and PCA[45]. Nucleotide diversity[46] and pairwise Φ_{ST} values[43] among populations were calculated with a custom Python script. For STRUCTURE and PCA, we included multiple SNPs and/or indels per locus; analyses based on alleles/haplotypes or a randomly selected SNP per locus generated similar results. For PCA, we coded biallelic polymorphisms and analysed the data in R (ref. [47]) following the approach of ref. [45]. We found empirically that excluding rare SNPs resulted in a larger number of principal component axes (for example, PC6 and above) capturing variation among, rather than within, populations. Thus, we excluded rare SNPs with a global frequency < 1% (that is, 6 or fewer copies of the rare allele among the 336 individuals = 672 alleles). Missing genotypes at autosomal loci (which comprised < 1% of the data matrix) were assigned a score equal to two times the respective population allele frequency for that SNP or indel. For Z-linked loci, at which males have two alleles and females one, male genotypes were coded as 0, 0.5 or 1 (heterozygote = 0.5), whereas females were coded as 0 or 1.

We analysed the same sets of biallelic polymorphisms in STRUCTURE. Nine replicate runs were completed for each value of k up to 13, with 20,000 steps in the Markov chain following a burn-in of 10,000 steps. We used the admixture model with allele frequencies correlated among populations. The Evanno method[48], as implemented in STRUCTURE HARVESTER[49], inferred just two distinct populations based on the autosomal data, but additional structure among populations was clearly evident and interpretable for values up to $k = 11$ (Supplementary Fig. 2). The same approach was used to analyse a set of 153 biallelic polymorphisms from 82 RAD-seq loci mapping to a 20-Mbp region in the centre of the Z chromosome. For females, one allele was scored as missing at each Z-linked locus. The Evanno method indicated three distinct populations for this Z-linked dataset, but values up to $k = 7$ produced interpretable results (Supplementary Fig. 2).

As a test of the relative levels of gene flow between sympatric populations of different species and allopatric populations of the same species, we calculated the proportion of rare autosomal SNPs shared among 11 populations of 7 species sampled in Australia and mainland Papua New Guinea (Fig. 2 and Supplementary Fig. 5). We tallied the number of autosomal SNPs with frequencies from 2 (0.46%) to 15 (3.5%) in this sample of 216 individuals ($n = 432$ alleles) and then calculated the proportion of SNPs in each frequency category that was shared between each pair of populations. We incorporated a small correction for differences in sample size ($n = 18$ for two populations, $n = 19$ for one population and $n = 20$ for eight populations) because the probability of detecting a rare allele increases with the number of individuals sampled.

As an additional and conceptually similar approach to examining patterns of recent shared ancestry, we analysed the autosomal RAD-seq data using fineRADstructure, which exploits the haplotype linkage information within each locus to derive a co-ancestry matrix based on the most recent coalescent events (that is, the sharing of identical or nearest-neighbour haplotypes among individuals)[25]. While the species in our analysis are all closely related, the median autosomal RAD-seq locus was 147 bp with 9 SNPs and 11 unique alleles/haplotypes, the relationships among which should provide substantially more information about recent ancestry than is available in individual SNPs.

Whole-genome sequencing. Low-depth, whole-genome sequencing data were generated for 177 samples, including representatives of 18 munia populations (Supplementary Table 4), plus a putative *L. grandis* × *L. castaneothorax* hybrid and one sample of a more distantly related congener, *L. leucosticta*, as an outgroup species (obtained from the Western Australia Museum). In most cases, we collected whole-genome sequencing data for 10 individuals per population, except for *L. castaneothorax* from Queensland, Australia ($n = 9$), *L. caniceps* from Central Province, Papua New Guinea ($n = 9$) and *L. melaena* from New Ireland Province, Papua New Guinea ($n = 7$) (Supplementary Table 4).

Genomic DNA were prepared using an Illumina Nextera kit following the manufacturer's protocol with the minor modifications noted below. We measured the DNA concentration using a NanoDrop instrument (Thermo Fisher Scientific) and used 25–45 ng of input genomic DNA rather than the recommended 50 ng because we found that the larger amount was not sufficiently fragmented in the 'tagmentation' step. We also prepared our own solid phase reversible immobilization beads[50] rather than using Agencourt AMPure XP beads. We assessed the distribution of fragment sizes and library concentration using

a High Sensitivity DNA Analysis kit on an Agilent 2100 Bioanalyzer and then pooled equimolar amounts of indexed libraries for 20–39 individuals per sequencing run.

The results from our first two sequencing runs revealed an excess of short fragments, making some paired-end reads redundant and reducing the overall yield of useful data. For subsequent runs, we ran pooled libraries on a Pippin Prep (Sage Science), selecting fragments between 400 and 1,400 bp. Fragment libraries were sequenced on either an Illumina HiSeq 2000 with 100-bp paired-end reads (*L. castaneothorax* and *L. flaviprymna* from Western Australia) or on an Illumina HiSeq 2500 with 150-bp paired-end reads (all other populations).

Reads were assigned to individual samples based on the Nextera dual indexing system using a custom script that allowed single-base errors in the 8-bp index read as long as all other possible indices could be unambiguously excluded. Adapter sequences were trimmed using CutAdapt[51], allowing up to a 15% mismatch, a minimum adapter length of 12 bp and a minimum remaining fragment length of 20 bp. Paired reads that overlapped were combined using PEAR (ref. [52]) with a minimum overlap of 30 bp; this step trimmed any remaining adapter sequences (that is, those < 12 bp). Finally, low-quality bases at the ends of non-overlapping reads were trimmed using a custom script. We aligned our data to the zebra finch *Taeniopygia guttata* reference genome[26] using the 'very-sensitive-local' option in Bowtie2 (ref. [53]). A high proportion of reads aligned to the genome (median per sample = 95.3%) and a large fraction of the genome was covered for each population (range = 92.7–96.0%), whereas genomic coverage for individual samples varied in relation to the sequencing depth (Supplementary Table 4). These results indicate that the zebra finch genome is an effective reference genome for munias. The average aligned sequencing depth per sample was ~1.6× for the autosomal genome (Supplementary Table 4). Subsequent data processing was completed using Picard, Samtools[54] and the Genome Analysis Toolkit[55–57]. Following indel realignment, SNPs were called using the Genome Analysis Toolkit's UnifiedGenotyper tool, excluding variants with a Phred confidence score of < 30. We used VCFtools[58] to further filter the SNPs, excluding variants with a mapping quality of < 30.

All subsequent analyses were based on the site-by-site genotypes inferred by the Genome Analysis Toolkit. We used the most likely genotype at each site for each sample, but for the calculation of population-level allele frequencies counted only one allele when an inferred homozygous genotype was based on a single read. Following filtering, the Genome Analysis Toolkit identified a total of 58.9 million SNPs, 26.2 million of which were fixed differences between the zebra finch and the focal clade. Another 3.4 million SNPs were unique changes in the outgroup species *L. leucosticta*. Thus, most analyses were based on ~28.4 million SNPs that were biallelic within the ingroup and had a total read depth across all 176 ingroup samples of ≤ 600 to avoid repetitive elements in the genome (this threshold was determined empirically from the distribution of the total sequencing depth per site; for the Z chromosome, we set the threshold at ≤ 400 reads for 116 males). This set of SNPs had a median sequencing depth of 308 across all samples—an average of ~1.7 reads per SNP per individual or ~17 reads per SNP per population.

We used two methods to identify genomic outlier regions. First, we completed sliding-window analyses of F_{ST} to identify regions of elevated divergence in four pairwise comparisons of genetically similar, sympatric populations: (1) *L. castaneothorax* and *L. flaviprymna* in Western Australia; (2) *L. castaneothorax* and *L. caniceps* in Central Province, Papua New Guinea; (3) *L. castaneothorax* and *L. spectabilis* in Madang Province, Papua New Guinea; and (4) *L. nevermanni* and *L. stygia* in Western Province, Papua New Guinea (Figs. 1 and 2). We used a custom Python script to calculate site-by-site F_{ST} values using equation 10 in ref. [59], restricting the analysis to SNPs with sequence data for at least five individuals in each focal population (or four individuals per population for the *L. castaneothorax* versus *L. flaviprymna* comparison due to the lower overall sequencing depth for these populations). To obtain a composite estimate of F_{ST} for each window, we calculated a ratio of averages rather than an average ratio[59–61]—an approach that yields an unbiased and increasingly accurate estimate of F_{ST} as the number of SNPs included in a sliding window increases, even with a small sample of individuals per population[61]. For comparison, we also completed sliding-window calculations of F_{ST} for the four focal comparisons using ANGSD[62], which incorporates genotype likelihood information in the calculation of population genetic parameters. The results from ANGSD produced essentially identical inferences about the genomic locations of divergence peaks, but the range of calculated F_{ST} values in each pairwise comparison was strongly correlated with the average sequencing depth per population, suggesting an undesirable bias.

Second, we included all 18 sampled populations in a sliding-window 'phylogenetic outlier' analysis. We calculated absolute divergence (d_{XY}) between all pairs of populations and used the resulting distance matrix to generate a neighbour-joining tree for each 100-kilobase (kb) region (sliding by 20 kb). We then compared each 100-kb tree to a genome-wide tree (based on ~1 million randomly sampled autosomal SNPs) using a tree-to-tree distance metric (the phylogenetic branch score; Bs)[63] that incorporates differences in both topology and relative branch lengths. All trees were re-scaled to a total length of 1 before comparisons were made and we excluded terminal branches to put greater emphasis on differences in topology, although this step had little effect on the regions identified as outliers. Likewise, similar results were obtained using trees based on matrices of net divergence (d_a)[46] values, with or without terminal branches included. We completed additional pairwise comparisons of selected populations to investigate phylogenetic outliers that did not correspond to regions of elevated divergence already identified in the four pairwise comparisons above.

As a metric of pairwise divergence, we found the sliding-window results for absolute divergence (d_{XY}) to provide little additional insight for the recently evolved species in our study. Differences in d_{XY} may help distinguish genomic regions affected by post-speciation selective sweeps from those diverging in the face of ongoing gene flow, but many generations (for example, ~0.5 N_e) are required for d_{XY} to become significantly elevated relative to the genomic background in the latter scenario[31]. Distinguishing between these alternatives—and other possible scenarios[9]—is not critical for our key conclusions, and we present other analyses supporting the inference of autosomal introgression between sympatric species. Empirically, nucleotide diversity (π) across the genome was highly correlated between sympatric munias (Supplementary Fig. 34), and d_{XY} was in turn highly correlated with average π, except in regions of elevated relative divergence (F_{ST}) where d_{XY} was also elevated relative to the average nucleotide diversity (Supplementary Fig. 35). Regions of elevated F_{ST} were often associated with greater differences in π (Supplementary Fig. 34), but were generally well within the typical range of d_{XY} values (Supplementary Fig. 36).

Allelic diversity and relationships within outlier regions were further characterized using PCA and phylogenetic analysis. Start and end points for each region (Table 1) were set to encompass the extent of clearly elevated F_{ST} values in one or more pairwise population comparisons (see Supplementary Figs. 7–26 and 28). For PCA, SNPs were coded as described above for the RAD-seq data, excluding sites with a global minor allele frequency of < 1%, as well as sites with data for < 81 of 176 samples (< 53 of 116 samples for the Z chromosome). Population-level neighbour-joining trees for each outlier region were based on the matrix of pairwise net divergence (d_a) values among all 18 populations. Trees including the outgroup taxon *L. leucosticta* were generated in the same manner, but were based on a larger set of SNPs (that is, including those representing fixed differences between the outgroup and focal clade). We do not report detailed results for regions of elevated F_{ST} associated with the centromeres of telocentric chromosomes such as chromosomes 12, 14, 17, 19 and 20, which show moderately elevated divergence between *L. castaneothorax* and *L. spectabilis* from Madang Province, for example (Supplementary Fig. 6).

Genes within and adjacent to outlier regions were identified using the zebra finch genome annotation in the Ensembl genome browser[64]. We plotted the approximate locations of centromeres for each chromosome based on published information[26,65].

Mitochondrial genomes. We assembled complete mitochondrial genome sequences for 174 of the 177 individuals included in the whole-genome sequencing analysis. We used the zebra finch mitochondrial genome to guide the assembly of a draft mtDNA for a single *L. flaviprymna* individual and then iteratively refined the consensus sequence by re-aligning the original whole-genome sequencing data. Once completed, the mtDNA sequence for this sample was used as the reference for the alignment and assembly of the remaining munia samples. Excluding three DNA samples extracted from blood, the mtDNA sequencing depth ranged from ~45× to over 5,000× per sample (median = 896×). For the phylogenetic analysis, we excluded three small regions, including two variable-length regions comprising mononucleotide C repeats (12 and 18 alignment positions, respectively), and 5 bp of the control region in which two positions appeared to be heteroplasmic in several samples. The resulting alignment was 16,803 bp. We generated a mitochondrial phylogeny using BEAST[66] (version 1.8.2), implementing an HKY + I + G model with an uncorrelated lognormal relaxed clock and a constant population size coalescent prior; the Markov chain was run for 25 million generations and sampled every 5,000 generations, with 10% excluded as burn-in. Convergence was examined in Tracer version 1.4 (ref. [67]) and a consensus tree was generated using TreeAnnotator[66].

To provide an approximate timescale for the diversification of the focal clade, we assembled mtDNA sequence data for additional estrildid finches and relevant outgroups and completed a second BEAST analysis based on 10,890 alignment positions from the 12 light-strand-encoded protein-coding genes. Given the greater level of divergence among taxa in this analysis, we partitioned the dataset into codon positions and estimated base frequencies and GTR + I + G substitution parameters separately for each partition, but linked the topology and branch lengths among partitions. The tree was calibrated using divergence time estimates from ref. [68], including the divergence of: (1) estrildid and parasitic finches (14.32 million years; 95% highest posterior density range = 11.04–17.92 million years); and (2) a clade of Old World finches (comprising parasitic finches, estrildids, ploceids and accentors) and their sister clade (comprising Passeridae, Motacillidae and the 'nine-primaried oscines') (20.03 million years; 95% highest posterior density range = 16.69–23.49 million years). These estimates are in turn based on a calibration point of 57.3 million years for the divergence of oscines and suboscines (excluding Acanthisittidae)[69], which is older than estimated in other recent studies[70–72], potentially biasing our result towards a higher estimate.

Data availability. All sequencing data are available under NCBI BioProject accession number PRJNA397589. Voucher specimens have been deposited in the Museum of Comparative Zoology at Harvard University.

Received: 3 November 2016; Accepted: 29 September 2017;
Published online: 30 October 2017

References

1. Via, S. & West, J. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* **17**, 4334–4335 (2008).
2. Nosil, P., Funk, D. J. & Ortiz-Barrientos, D. Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402 (2009).
3. Wu, C. I. The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001).
4. Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**, 1572–1578 (2005).
5. Toews, D. P. L. et al. Plumage genes and little else distinguish the genomes of hybridizing warblers. *Curr. Biol.* **26**, 2313–2318 (2016).
6. Ellegren, H. et al. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **461**, 756–760 (2012).
7. Poelstra, J. W. et al. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**, 1410–1414 (2014).
8. Malinsky, M. et al. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**, 1493–1498 (2015).
9. 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).
10. Givnish, T. J. Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytol.* **207**, 297–303 (2015).
11. Brawand, D. et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (2014).
12. Jones, F. C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
13. Lamichhaney, S. et al. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
14. The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
15. Berner, D. & Salzberger, W. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet.* **31**, 491–499 (2015).
16. Weir, J. T. & Price, T. D. Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *Am. Nat.* **177**, 462–469 (2011).
17. Campagna, L. et al. Rapid phenotypic evolution during incipient speciation in a continental avian radiation. *Proc. R. Soc. B* **279**, 1847–1856 (2012).
18. Campagna, L. et al. Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Sci. Adv.* **3**, e1602404 (2017).
19. Haldane, J. B. S. Sex ratio and unisexual sterility in hybrid animals. *J. Genet.* **12**, 101–109 (1922).
20. Immelmann, K. Besiedlungsgeschichte und Bastardierung von *Lonchura castaneothorax* und *Lonchura flaviprymna* in Nordaustralien. *J. Ornithol.* **103**, 352–357 (1962).
21. Restall, R. *Munias and Mannikins* (Pica Press, Crowborough, 1996).
22. Servedio, M. R. & Noor, M. A. F. The role of reinforcement in speciation: theory and data. *Annu. Rev. Ecol. Syst.* **34**, 339–364 (2003).
23. Kimura, M. & Ohta, T. The age of a neutral mutant persisting in a finite population. *Genetics* **75**, 199–212 (1973).
24. Slatkin, M. Rare alleles as indicators of gene flow. *Evolution* **39**, 53–65 (1985).
25. Malinsky, M., Trucchi E., Lawson, D. & Falush D. RADpainter and fineRADstructure: population inference from RADseq data. Preprint at https://www.biorxiv.org/content/early/2016/06/7/057711 (2016).
26. Warren, W. C. et al. The genome of a songbird. *Nature* **464**, 757–762 (2010).
27. Domyan, E. T. et al. Epistatic and combinatorial effects of pigmentary gene mutations in the domestic pigeon. *Curr. Biol.* **24**, 459–464 (2014).
28. Lamichhaney, S. et al. A beak size locus in Darwin's finches facilitated character displacement during a drought. *Science* **352**, 470–474 (2016).
29. Nadeau, N. J. et al. Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Phil. Trans. R. Soc. B* **367**, 343–353 (2012).
30. Vijay, N. et al. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Commun.* **7**, 13195 (2016).
31. 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).
32. Burri, R. et al. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* **25**, 1656–1665 (2915).
33. Cutter, A. D. & Payseur, B. A. Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat. Rev. Genet.* **14**, 262–274 (2013).
34. Vijay, N. et al. Genome-wide signatures of genetic variation within and between populations—a comparative perspective. Preprint at https://www.biorxiv.org/content/early/2017/01/31/104604 (2017).
35. Wolf, J. B. W. & Ellegren, H. Making sense of genomic islands of differentiation in light of speciation. *Nat. Rev. Genet.* **18**, 87–100 (2017).
36. Hill, G. E. & Johnson, J. D. The mitonuclear compatibility hypothesis of sexual selection. *Proc. R. Soc. B* **280**, 20131314 (2013).
37. Petit, R. J. & Excoffier, L. Gene flow and species delimitation. *Trends Ecol. Evol.* **24**, 386–393 (2009).
38. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
39. Colosimo, P. F. et al. Widespread parallel evolution in sticklebacks by repeated fixation of *Ectodysplasin* alleles. *Science* **307**, 1928–1933 (2005).
40. Norris, L. C. et al. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. *Proc. Natl Acad. Sci. USA* **112**, 815–820 (2015).
41. Seutin, G. B., White, N. & Boag, P. T. Preservation of avian blood and tissue samples for DNA analyses. *Can. J. Zool.* **69**, 82–90 (1991).
42. DaCosta, J. M. & Sorenson, M. D. Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS ONE* **9**, e106713 (2014).
43. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
44. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
45. Novembre, J. & Stephens, M. Interpreting principal components analyses of spatial population genetic variation. *Nat. Genet.* **40**, 646–649 (2008).
46. Nei, M. & Li, W.-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci. USA* **76**, 5269–5273 (1979).
47. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, 2008).
48. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* **14**, 2611–2620 (2005).
49. Earl, D. A. & von Holdt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361 (2012).
50. Roland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).
51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
52. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
53. Langmeade, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
54. Li, H. et al. The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
55. DePristo, M. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genetics* **43**, 491–498 (2011).
56. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
57. Van der Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).
58. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
59. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
60. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
61. Willing, E.-M., Dreyer, C. & van Oosterhout, C. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* **7**, e42649 (2012).
62. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).

63. Kuhner, M. K. & Felsenstein, J. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* **11**, 459–468 (1994).
64. Cunningham, F. et al. Ensembl 2015. *Nucleic Acids Res.* **43**, D662–D669 (2015).
65. Knief, U. & Forstmeier, W. Mapping centromeres of microchromosomes in the zebra finch (*Taeniopygia guttata*) using half-tetrad analysis. *Chromosoma* **125**, 757–768 (2016).
66. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012).
67. Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. *Tracer* v1.6 (2014); http://tree.bio.ed.ac.uk/software/tracer/
68. Gibb, G. C. et al. New Zealand passerines help clarify the diversification of major songbird lineages during the Oligocene. *Genome Biol. Evol.* **7**, 2983–2995 (2015).
69. Pratt, R. C. et al. Toward resolving deep neoaves phylogeny: data, signal enhancement, and priors. *Mol. Biol. Evol.* **26**, 313–326 (2009).
70. Claramunt, S. & Cracraft, J. A new time tree reveals Earth history's imprint on the evolution of modern birds. *Sci. Adv.* **1**, e1501005 (2015).
71. Jarvis, E. D. et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
72. Prum, R. O. et al. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573 (2015).
73. Eom, D. S. et al. Melanophore migration and survival during zebrafish adult pigment stripe development requires the immunoglobulin superfamily adhesion molecule Igsf11. *PLoS Genet.* **8**, e1002899 (2012).
74. Fariello, M.-I. et al. Selection signatures in worldwide sheep populations. *PLoS ONE* **9**, e103813 (2014).
75. Guenther, C. A., Tasic, B., Luo, L., Bedell, M. A. & Kingsley, D. M. A molecular basis for classic blond hair color in Europeans. *Nat. Genet.* **46**, 748–752 (2014).
76. Miller, C. T. et al. *Cis*-regulatory changes in Kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. *Cell* **131**, 1179–1189 (2007).
77. Sulem, P. et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443–1452 (2007).
78. Dutton, K. A. et al. Zebrafish colourless encodes *sox10* and specifies non-ectomesenchymal neural crest fates. *Development* **128**, 4113–4125 (2001).
79. Gunnarson, U. et al. The dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of *SOX10*. *Pigment Cell Melanoma Res.* **24**, 268–274 (2011).
80. Hubbard, J. K., Uy, J. A. C., Hauber, M. E., Hoekstra, H. E. & Safran, R. J. Vertebrate pigmentation: from underlying genes to adaptive function. *Trends Genet.* **26**, 231–239 (2010).
81. Baião, P. C., Schreiber, E. A. & Parker, P. G. The genetic basis of plumage polymorphism in red-footed boobies (*Sula sula*): a melanocortin-1 receptor (MC1R) analysis. *J. Hered.* **98**, 287–292 (2007).
82. Mundy, N. I. et al. Conserved genetic basis of a quantitative plumage trait involved in mate choice. *Science* **303**, 1870–1873 (2004).
83. Nachman, M. W., Hoekstra, H. E. & D'Agostino, S. L. The genetic basis of adaptive melanism in pocket mice. *Proc. Natl Acad. Sci. USA* **100**, 5268–5273 (2003).
84. Nadeau, N. J., Minvielle, F. & Mundy, N. I. Association of a Glu92Lys substitution in *MC1R* with extended brown in Japanese quail (*Coturnix japonica*). *Anim. Genet.* **37**, 287–289 (2006).
85. Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R. E. & Mundy, N. I. The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. *Curr. Biol.* **11**, 550–557 (2001).
86. Uy, J. A., Moyle, R. G., Filardi, C. E. & Cheviron, Z. A. Difference in plumage color used in species recognition between incipient species is linked to a single amino acid substitution in the melanocortin-1 receptor. *Am. Nat.* **174**, 244–254 (2009).
87. Hiragaki, T. et al. Recessive black is allelic to the yellow plumage locus in Japanese quail and associated with a frameshift deletion in the *ASIP* gene. *Genetics* **178**, 771–775 (2008).
88. Nadeau, N. J. et al. Characterization of Japanese quail yellow as a genomic deletion upstream of avian homolog of the mammalian *ASIP* (agouti) gene. *Genetics* **178**, 777–786 (2008).
89. Oribe, E. et al. Conserved distal promoter of the agouti signaling protein (*ASIP*) gene controls sexual dichromatism in chickens. *Gen. Comp. Endocrinol.* **177**, 231–237 (2012).
90. Berryere, T. G., Kerns, J. A., Barsh, G. S. & Schmutz, S. M. Association of an agouti allele with fawn or sable coat color in domestic dogs. *Mamm. Genome* **16**, 262–272 (2005).
91. Bonilla, C. et al. The 8818G allele of the agouti signalling protein (*ASIP*) gene is ancestral and associated with darker skin color in African Americans. *Hum. Genet.* **116**, 402–406 (2005).
92. Manceau, M., Domingues, V. S., Mallarino, R. & Hoekstra, H. E. The developmental role of agouti in color pattern evolution. *Science* **331**, 1062–1065 (2011).
93. Dorshorst, B. et al. A complex genomic rearrangement involving the endothelin-3 locus causes dermal hyperpigmentation in the chicken. *PLoS Genetics* **7**, e1002412 (2011).
94. Shinomiya, A. et al. Gene duplication of endothelin 3 is closely correlated with the hyperpigmentation of the internal organs (fibromelanosis) in silky chickens. *Genetics* **190**, 627–638 (2012).
95. Kaelin, C. B. et al. Specifying and sustaining pigmentation patterns in domestic and wild cats. *Science* **337**, 1536–1541 (2012).
96. Xu, X. et al. The genetic basis of white tigers. *Curr. Biol.* **23**, 1031–1035 (2013).
97. Gunnarsson, U. et al. Mutations in *SLC45A2* cause plumage color variation in chicken and Japanese quail. *Genetics* **175**, 867–877 (2007).
98. Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R. & Tabin, C. J. *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465 (2004).
99. Wilkinson, L. et al. CRIM1 regulates the rate of processing and delivery of bone morphogenetic proteins to the cell surface. *J. Biol. Chem.* **278**, 34181–34188 (2003).
100. Miao, D. et al. Parathyroid hormone-related peptide is required for increased trabecular bone volume in parathyroid hormone-null mice. *Endocrinology* **145**, 3554–3562 (2004).
101. Fasquelle, C. et al. Balancing selection of a frame-shift mutation in the *MRC2* gene accounts for the outbreak of the crooked tail syndrome in Belgian blue cattle. *PLoS Genetics* **5**, e1000666 (2009).
102. Cardona, A. et al. Genome-wide analysis of cold adaptation in indigenous Siberian populations. *PLoS ONE* **9**, e98076 (2014).
103. Raj, S. M., Pagani, L., Gallego Romero, I., Kivisikd, T. & Amos, W. A general linear model-based approach for inferring selection to climate. *BMC Genet.* **14**, 87 (2013).

Acknowledgements

This research was supported by the National Geographic Society (8933-11), the National Science Foundation (DEB 1210810 and DEB 1446085) and by small grants from the American Ornithologists' Union, American Museum of Natural History, American Society of Naturalists, Society for the Study of Evolution, Society of Systematic Biologists, Systematics Association and Linnean Society. The Departments of Environment and Conservation in Papua New Guinea and Western Australia provided permits for the fieldwork. We thank the Western Australian Museum and Commonwealth Scientific and Industrial Research Organisation for tissue loans. J. Robins at the National Research Institute facilitated our applications for permits and visas in Papua New Guinea. This work was conducted under Boston University's Institutional Animal Care and Use Committee protocol number 10-011. We thank J. Berv, S. Billy, C. Kieswetter, J. Lewis, R. McKay, P. Saguba, T. Stryjewski and many others for assistance with fieldwork, and C. Balakrishnan, J. DaCosta, D. Irwin and C. Schneider for comments on the paper.

Author contributions

M.D.S. conceived the study. K.F.S., with limited assistance from M.D.S., completed the fieldwork, collected and prepared the specimens and collected all of the genomic data. The authors worked together on analysing the data and writing and approving the paper.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-017-0364-7.

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

Correspondence and requests for materials should be addressed to K.F.S. or M.D.S.

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

nature research

Corresponding author(s): Katherine Faust Stryjewski
Michael D. Sorenson

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

We sampled a total of 435 birds from 17 populations representing 11 of the 13 species in the radiation. We attempted to acquire RAD-seq data for 20 birds per population and whole-genome sequencing (WGS) data for 10 birds per population, determining that 40 and 20 chromosomes per population, respectively, would be sufficient to call genotypes and obtain an accurate sampling of population genetic variation.

2. Data exclusions

Describe any data exclusions.

For the RAD-seq data, loci with a median per sample sequencing depth of less than 20, missing data for more than 5% of individuals, or flagged genotypes in more than 2% of individuals were omitted (see Methods for more specifics on flagged genotypes). For WGS data, only loci that aligned with the Zebra Finch genome were included in the analyses. These were filtered to exclude variants with a Phred confidence score and/or mapping quality of less than 30. SNPs that were identified as fixed differences between the Zebra finch and/or the outgroup and the focal clade were not included in most analyses. See the Methods section for more specifics.

3. Replication

Describe whether the experimental findings were reliably reproduced.

N/A

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

N/A

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☒	☐ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

RAD-seq data were processed using custom Python scripts available at https://github.com/BU-RADseq. These make use of three freely available software packages: BLAST, USEARCH v5, MUSCLE v3.8.31. We used Geneious v8.0.3 to examine alignments. Analyses were performed in STRUCTURE, fineRADstructure, and R. For whole-genome sequence data, adapters were trimmed using CutAdapt and overlapping paired reads were combined using PEAR. Sequences were aligned to the Zebra Finch genome using Bowtie2. Sequence data was processed using Picard, Samtools, the Genome Analysis Toolkit, and VCFtools. Genes were identified using the Ensembl genome browser. Sliding window analyses and phylogenetic analyses of individual loci were completed using custom scripts. PCA was performed in R. A mitochondrial phylogeny was constructed using BEAST v1.8.2, Tracer v1.4, and TreeAnnotator.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Bird tissues and specimens collected for this study have been accessioned into the Harvard Museum of Comparative Zoology collection and made available for loan or study.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

N/A

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used. N/A

b. Describe the method of cell line authentication used. N/A

c. Report whether the cell lines were tested for mycoplasma contamination. N/A

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. N/A

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

A total of 435 birds were included in this study. Of these, 10 came from tissue loans from the Australian National Wildlife Collection, CSIRO. The remaining 425 birds were sampled during fieldwork in Australia and Papua New Guinea. In most cases, tissue samples were taken and an accompanying study skin was prepared (up to 20 birds per population). Blood samples also were collected for up to 30 birds per population. This fieldwork was conducted with the necessary permits from local government and samples were imported into the US with permission from US Fish & Wildlife and the USDA. Fieldwork was conducted under Boston University IACUC protocol number 10-011.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants. N/A