

# A supergene underlies linked variation in color and morphology in a Holarctic songbird

Erik R. Funk <sup>1</sup>✉, Nicholas A. Mason<sup>2</sup>, Snæbjörn Pálsson <sup>3</sup>, Tomáš Albrecht <sup>4,5</sup>, Jeff A. Johnson <sup>6</sup> & Scott A. Taylor <sup>1</sup>

The genetic architecture of a phenotype can have considerable effects on the evolution of a trait or species. Characterizing genetic architecture provides insight into the complexity of a given phenotype and, potentially, the role of the phenotype in evolutionary processes like speciation. We use genome sequences to investigate the genetic basis of phenotypic variation in redpoll finches (*Acanthis* spp.). We demonstrate that variation in redpoll phenotype is broadly controlled by a ~55-Mb chromosomal inversion. Within this inversion, we find multiple candidate genes related to melanogenesis, carotenoid coloration, and bill shape, suggesting the inversion acts as a supergene controlling multiple linked traits. A latitudinal gradient in ecotype distribution suggests supergene driven variation in color and bill morphology are likely under environmental selection, maintaining supergene haplotypes as a balanced polymorphism. Our results provide a mechanism for the maintenance of ecotype variation in redpolls despite a genome largely homogenized by gene flow.

<sup>1</sup>Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA. <sup>2</sup>Museum of Natural Science and Department of Biological Science, Louisiana State University, Baton Rouge, LA 70803, USA. <sup>3</sup>Department of Life and Environmental Sciences, University of Iceland, Askja, Sturlugata 7, 101 Reykjavik, Iceland. <sup>4</sup>Department of Zoology, Charles University, Vinicna 7, CZ-12844 Prague, Czech Republic. <sup>5</sup>Institute of Vertebrate Biology, Czech Academy of Sciences, Kvetna 8, CZ-60365 Brno, Czech Republic. <sup>6</sup>Wolf Creek Operating Foundation, Wolf, WY 82844, USA.  
✉email: [erikfunk@gmail.com](mailto:erikfunk@gmail.com)

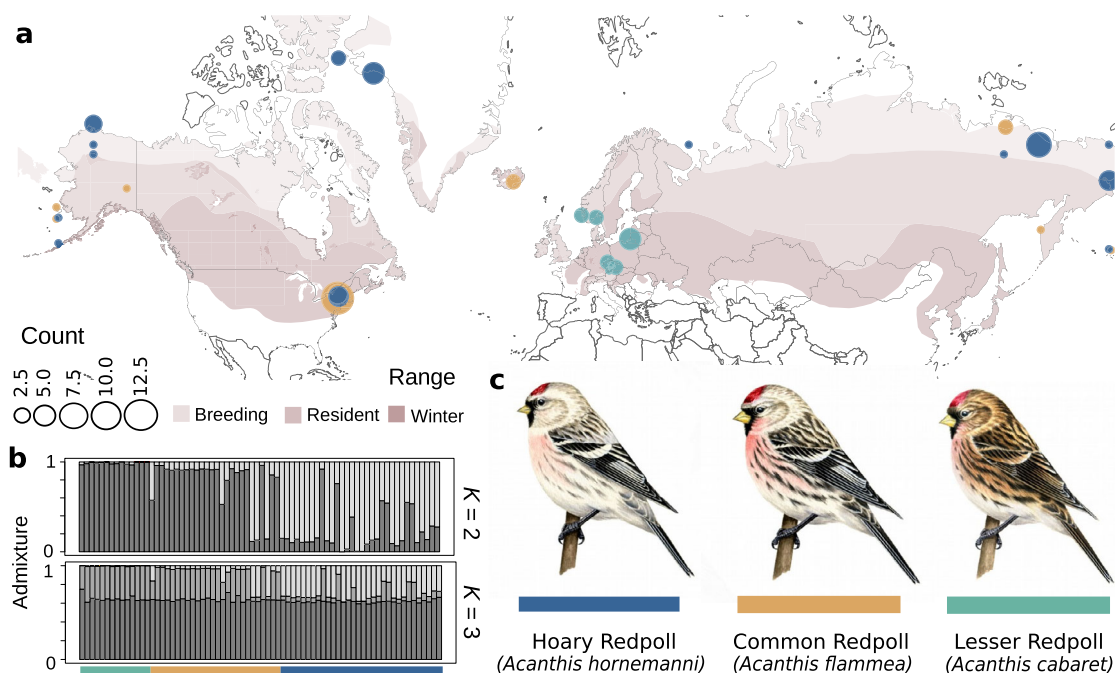
Identifying the genetic basis of divergent traits has become a major goal in biology. Studies demonstrating associations between genotype and phenotype have provided evidence for key evolutionary processes, such as adaptive introgression<sup>1–3</sup> and speciation<sup>4</sup>. Understanding the genetic architecture of traits (e.g., single gene, polygenic, supergene) may provide insight into the complexity of the phenotype, and its role in various evolutionary processes<sup>5</sup>. Speciose radiations, such as the Capuchino seedeaters (*Sporophila* spp.), demonstrate that traits under independent modular genetic architecture might readily generate novel phenotypes through genetic recombination, with the potential to lead to increased species diversity<sup>6–8</sup>. Alternatively, traits may become tightly linked by large structural variants, such as chromosomal inversions<sup>9,10</sup>. When chromosomal inversions link multiple genetic elements that control a suite of traits (e.g., phenotypic, behavioral, etc.) they are commonly referred to as supergenes<sup>11</sup>. Because supergenes can maintain complex phenotypic differences within a population, supergenes may have unique evolutionary implications. For example, different evolutionary outcomes may depend on if (and how) incompatibilities arise between supergene haplotypes. Mutations that arise along one inversion haplotype might be locally advantageous in an isolated population, but incompatible with the alternative inversion haplotype, resulting in a classic model of lineage divergence and speciation through Bateson–Dobzhansky–Muller incompatibilities<sup>12,13</sup>. However, the immediate effects of an inversion (e.g., altering gene expression at inversion breakpoints) or subsequent point mutations within an inversion may result in an inversion genotype being lethal or highly deleterious to fitness, despite providing a fitness benefit to heterozygous individuals<sup>9,10,14</sup>. This latter model can generate a stable polymorphism that promotes intraspecific variation (such as different ecotypes) but not lineage divergence and subsequent speciation. Importantly, even in the absence of lethal genotypes, inversions may persist as stable polymorphisms if they affect traits involved in local adaptation across a heterogeneous environment<sup>15</sup>.

Redpoll finches (*Acanthis* spp.) are high-latitude Holarctic songbirds that have a long history of taxonomic controversy due to low levels of genetic divergence and overlapping geographic distributions despite variation in body size, bill morphology, and plumage coloration (Fig. 1a; refs. 16,17). Although these species co-occur in much of their ranges, environmental niche models demonstrate slight latitudinal differences between common and hoary redpolls during the breeding months<sup>18</sup>. Specifically, lighter-plumaged individuals with shorter and narrower bills (i.e., hoary redpolls) are more common at higher latitudes while lesser redpolls are restricted to western Europe and southern Scandinavia. This variation in redpoll morphology across latitude suggests phenotype may be playing a role in local adaptation; however, it is unclear the extent to which adaptation is contributing to reproductive isolation and divergence. While redpolls are commonly defined as three species, including the common redpoll (*Acanthis flammea*), the hoary redpoll (*A. hornemanni*), and the lesser redpoll (*A. cabaret*), we refer to the different redpoll groups described above as ecotypes based on phenotype and breeding distribution, given the lack of taxonomic consensus.

Here, we identify regions of the genome underlying phenotypic variation, including numerous genes inside a 55-Mb chromosomal inversion that links multiple phenotypic traits as a supergene. Genomic divergence appears constrained to these regions while the genomic background appears homogenized by ongoing gene flow. The associations identified here provide insight into the role that traits and their genetic architecture play in maintenance of phenotypic variation despite widespread gene flow.

## Results and discussion

To evaluate population structure in redpolls, we sequenced genomes of 73 individuals from the three described redpoll ecotypes (Supplementary Data File 1). Our results from whole genomes



**Fig. 1** The redpoll finches. **a** Seasonal distribution of redpolls shown in purple, with lightest indicating breeding range, and darkest indicating wintering range. Colored dots indicate sample number and locations with blue corresponding to hoary redpoll, gold corresponding to common redpoll, and green corresponding to lesser redpoll. **b** conStruct plots for a  $K$  of 2 and 3 showing that genetic clustering poorly aligns with taxonomy. **c** Illustrations of the three redpoll ecotypes demonstrating differences in size, bill, and plumage coloration (Illustrations by Liz Clayton Fuller). Distribution was generated using shape files provided by BirdLife International.

confirm findings from a previous study using a reduced-representation approach (ddRAD-seq)<sup>18</sup>: redpolls lack population genetic structure by either geography or ecotype boundaries (Fig. 1b), with spatially explicit clustering analyses supporting  $K = 2$ , and failing to group all individuals according to their species classification. In addition, principal component analysis (PCA) (Supplementary Fig. 1) of 25 million single nucleotide polymorphisms (SNPs) further reveal that PC1 explains only 3.14% of total genomic variation across all ecotypes and the majority of their global distribution. However, both PCA and population assignment analyses nonetheless indicate some degree of genetic clustering (Fig. 1b, Supplementary Fig. 1). PC1 visually separated samples into three clusters, with a left-most cluster containing both lesser and common redpolls, a right-most cluster containing almost entirely hoary redpolls, and a central cluster containing a mix of both common and hoary redpolls. However, many localities were recovered in all three groups, suggesting no influence of geography on genetic structure (Supplementary Fig. 2). Because neither geography nor ecotype were perfectly assigned to clusters, we were interested in identifying the genomic regions responsible for generating these clusters, and in investigating their potential evolutionary impacts.

**Genetics of redpoll divergence.** To identify divergent regions of the genome in redpolls, we aligned sequences to a brown-capped rosy-finch (*Leucosticte australis*) reference genome and searched for local peaks of differentiation between PCA clusters by calculating  $F_{ST}$  and  $d_{XY}$  in 25-kb windows across all chromosomes including all ecotypes. These scans identified a highly differentiated region across 55 Mb of chromosome 1 (Fig. 2a, c, Supplementary Fig. 3). Rerunning PCA and population assignment analyses after the removal of this chromosome either eliminated, or reduced, the variation explained (Supplementary Figs. 1 and 4), demonstrating the strong contribution of this region to total genetic differentiation in redpolls. Further, conducting a PCA using only chromosome 1 qualitatively produced much stronger definition in the three clusters originally identified (Fig. 2e). Within-group heterozygosity of the middle cluster for the highly differentiated region (0.626) was roughly double that of the outside clusters (0.388 and 0.378 for left and right clusters, respectively), suggesting that the PCA groups represent three possible inversion genotypes. We hereafter refer to these putative genotypes as AA, AB, and BB in left to right order across PC1. We do not distinguish between the ancestral and inverted haplotypes, and use the term inversion to refer to the inversion region rather than the inverted haplotype.

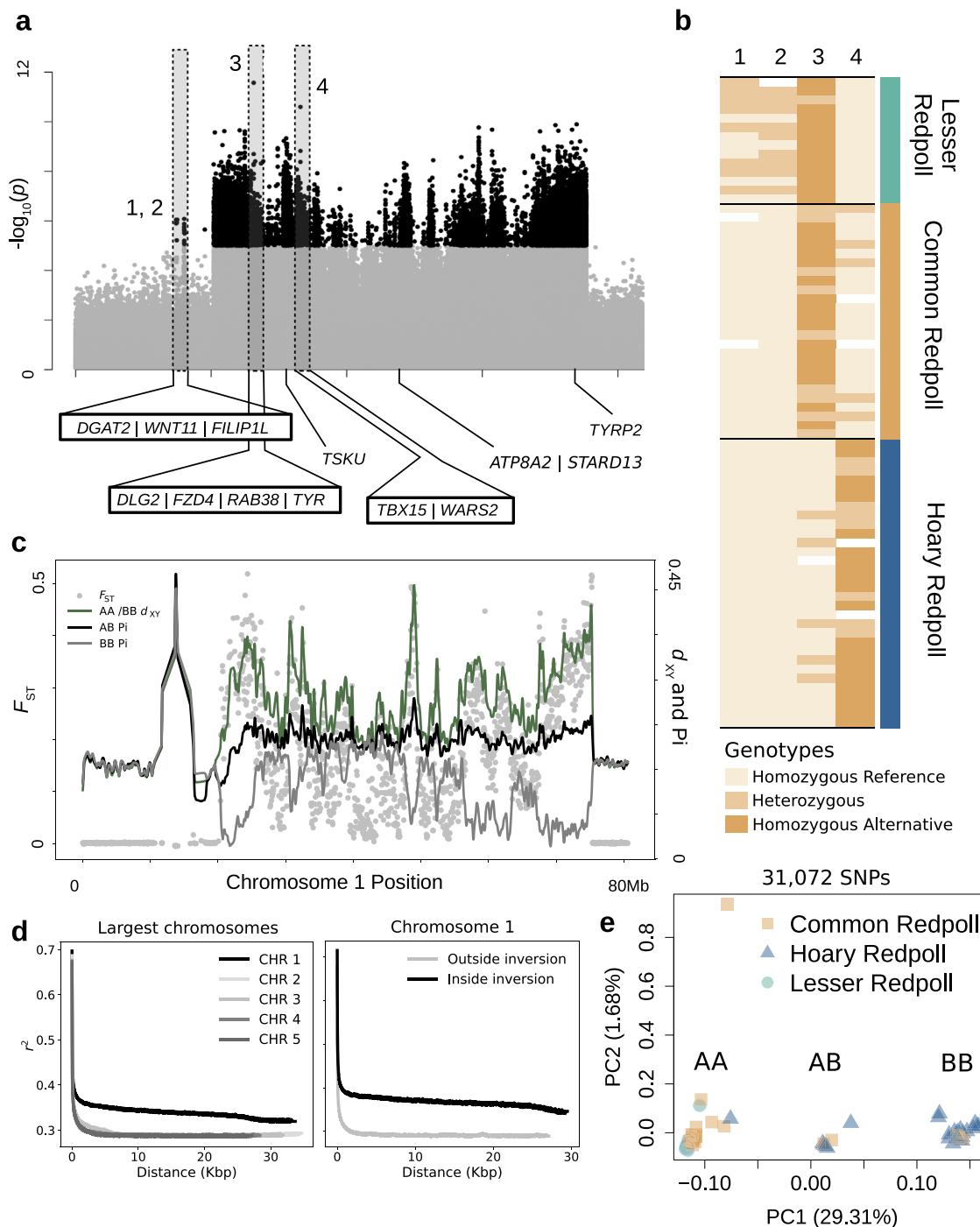
Broadly, the pattern of divergence recovered here is consistent with a large pericentric chromosomal inversion<sup>10,19,20</sup>, including abrupt changes in  $F_{ST}$  corresponding to the inversion breakpoints, and a central spike at the centromere (Fig. 2a, c). Reduced recombination within an inversion is expected to produce patterns of elevated linkage disequilibrium (LD), along with a decrease in nucleotide diversity along the inversion in homozygotes. These patterns are both confirmed here, including a within-cluster decrease in homozygote (AA and BB) nucleotide diversity ( $\pi$ ), and elevated LD within the inversion when both compared to regions outside the inversion and along other chromosomes (Fig. 2c, d). To further characterize the inversion, we selected one individual each from the AA and BB genotype groups to resequence using Oxford Nanopore Technologies MinION long-read sequencing. Structural variant calling with SVIM v1.4.2<sup>21</sup> identified an inversion extending from 18.9 to 75 Mb along chromosome 1; however, overall number of reads supporting the variant call was low due to the size of the inversion and low yield from the MinION runs.

Because redpolls overlap extensively in distribution, species identification is made primarily on the basis of a suite of morphological characters, including plumage coloration (extent of brown and red pigments), bill size and shape, and body size. Transitioning from the AA, to AB, to BB genotype also broadly mirrors a transition in phenotype from dark to light plumage coloration, where the AA genotype is associated with dark plumage, BB is associated with light plumage, and AB is intermediate. Mason and Taylor<sup>18</sup> paired phenotypic measurements of plumage and bill morphology with gene expression data to reveal a strong, linear correlation between gene expression and morphology (see ref. <sup>18</sup>, Fig. 3a). Superimposing inversion genotype on this relationship for the same individuals reveals that inversion homozygotes form the extremes of these categories, while the single heterozygote forms an intermediate (Fig. 3a). Although sample size in this comparison is small, it provides strong independent evidence that the chromosome 1 inversion plays a large role in redpoll morphology, and that phenotypic variation may be additive with respect to inversion haplotype copy number.

**Genetic associations and candidate loci.** In total, we identified 498 annotated genes within the chromosome 1 inversion region. While all genes within the inversion are likely to be linked through the suppression of recombination, and thus could be contributing equally to phenotype, we nonetheless attempted to narrow down candidate gene regions in order to infer which biological processes and pathways were potentially influenced by the inversion and identify associated regions elsewhere in the genome. To do so, we applied two approaches: (1) by compiling a list of genes that fell within the highest  $F_{ST}$  peaks, and (2) by identifying SNPs significantly associated with species classification using a genome-wide efficient mixed model analysis (GEMMA)<sup>22</sup>. While species identity does not perfectly correlate with redpoll phenotype because they exhibit continuous phenotypic variation, the fact that species classification relies almost entirely on morphology makes it a reasonable proxy for total phenotypic variation. Finally, we annotated missense mutations within the identified genes based on a variant's location with respect to open reading frames using SNPeff v4.3<sup>23</sup>. Our results suggest that the vast majority of SNPs associated with phenotypic variation in redpolls are within or close to the inversion: 99% of 20,443 SNPs significantly associated with redpoll phenotype were located on chromosome 1, with only 167 located elsewhere in the genome (Supplementary Fig. 3). To evaluate the reliability of these SNPs in predicting phenotype, we used a Bayesian sparse linear mixed model in a leave-one-out cross validation framework. Predicted phenotypes suggest that allelic variation of the identified SNPs explain a significant proportion of the observed phenotypic variation ( $R^2 = 0.79$ ; Supplementary Fig. 5).

We filtered annotations for genes that either contained, or were adjacent to, significant SNPs as identified by GEMMA or  $F_{ST}$  outlier analysis, resulting in 322 genes across 7 chromosomes (Supplementary Data File 2). Within this gene set, the gene ontology category of biological regulation was overrepresented. While this category is broad and difficult to interpret meaningfully, we note that a number of genes on chromosomes 1 and 2 identified by our analysis had annotations that either relate to coloration or bird bill development or have been implicated in coloration or bird bill development in previous studies (Table 1).

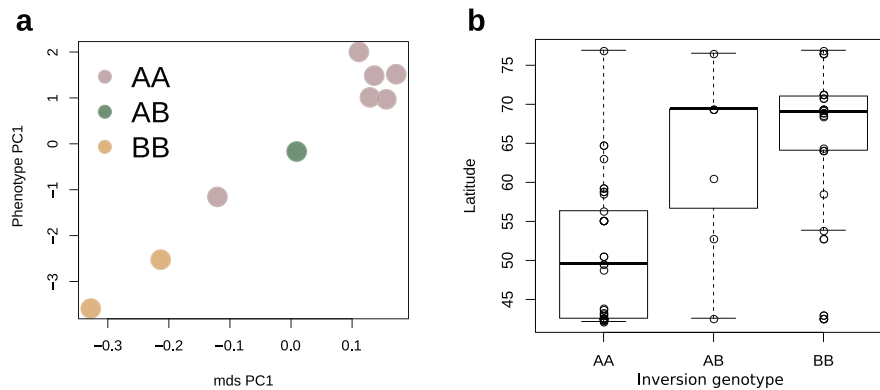
Within the chromosome 1 inversion region, some of the most differentiated and significantly associated regions include key genes relating to melanin synthesis: *TYR*, *TYRP2*, *FZD4*, *TSKU*, *FSTL1*<sup>24–29</sup>. Both *TYR* and *TYRP2* produce melanogenic enzymes that directly synthesize melanin. In addition, *FZD4* produces a G



**Fig. 2 Chromosome 1 supergene.** **a** Chromosome 1 SNPs significantly associated with phenotype (black dots) using mixed model analysis in GEMMA with an alpha of  $1 \times 10^{-5}$  to correct for multiple comparisons. Numbers indicate the two most significant SNP associations outside (1,2) and inside (3,4) the inversion region and correspond to genotypes in **(b)** for those specific SNPs. Lightest yellow cells indicate individuals homozygous for the reference allele, darkest indicate homozygous for the alternative allele, intermediate indicate heterozygotes, and white are missing data. **c**  $Pi$  (black and gray lines for AA and AB individuals, respectively) and  $d_{XY}$  (green line) in 50-kb windows for the first 80 Mb of chromosome 1. **d** LD calculated as  $r^2$  comparing chromosome 1 (black line) to 4 of the other largest chromosomes (shades of gray) and LD within the supergene (black) to the rest of chromosome 1 (gray). **e** PCA of chromosome 1 with a 40% minor allele frequency cutoff showing three main clusters along PC1. Taxa colors correspond to Fig. 1. All sequences were aligned to a brown-capped rosy-finch (*Leucosticte australis*) reference genome. Source data are provided as Source data file.

protein-coupled receptor in the Wntless-type signaling pathway, which acts as one of the main pathways affecting the regulation of MITF<sup>29,30</sup>. Previous studies of gene expression in redpolls<sup>18</sup> also reported differential gene expression of *FZD3*, suggesting that Frizzled family receptors may play a significant role in further modulating melanogenesis in this group.

Redpoll phenotype also varies in the amount of red feather coloration resulting from carotenoid pigmentation. Carotenoid pigments are unique in animals in that they cannot be synthesized endogenously and must instead be taken in through their diet before they can be deposited in feathers. Previous studies of genes involved in carotenoid pigmentation in birds



**Fig. 3 Correlations between inversion genotype, morphology, and latitude.** **a** Phenotype PC1 and gene expression data (as mds PC1) from Mason and Taylor (2015) colored by inversion genotype, with extreme phenotypes produced by homozygotes shown in gold and pink, and an intermediate phenotype produced by a heterozygote shown in green. **b** Latitude of sampling site for each individual grouped by inversion genotype demonstrating B haplotypes increase in frequency with latitude (AA  $n = 37$ , AB  $n = 7$ , BB  $n = 28$ ). Box hinges represent the first and third quartiles, with centers representing medians. Whiskers represent maximum and minimum values except for BB, where outliers are values exceeding 1.5 times the interquartile range. Source data are provided as Source data file.

**Table 1 Candidate genes for plumage color and bill morphology.**

Locus name	Location	Synonymous mutations	Missense mutations	Description
FILIP1L	CHR1	3	0	Wnt pathway inhibitor
FSTL1	CHR1 <sup>a</sup>	1	0	Bone morphogenic protein inhibitor
TSK	CHR1 <sup>a</sup>	12	2	Bone morphogenic protein inhibitor
FZD4	CHR1 <sup>a</sup>	0	0	Wnt pathway inhibitor
TYR	CHR1 <sup>a</sup>	17	1	Catalyzes the initial step leading to melanin production
ATP8A2	CHR1 <sup>a</sup>	14	8	ATPase phospholipid transporting
STARD13	CHR1 <sup>a</sup>	27	3	StAR-related lipid transfer domain
TYRP2	CHR1 <sup>a</sup>	24	7	Regulates eumelanin and pheomelanin
IL1R1	CHR1	13	13	SFRP4 enhancer and differentially expressed in redpolls
PKS15/1	CHR2	67	37	Possible homology with locus affecting color accumulation in Budgerigars
SFRP4	CHR2	1	1	Wnt and bone morphogenesis modulator

<sup>a</sup>Located within the supergene.

highlight the role of two scavenger receptor genes (*SCARB1*<sup>31</sup>, *SCARF2*<sup>32</sup>). The proteins produced by these genes likely function in the recognition of the lipoproteins that transport the hydrophobic carotenoid pigments. We identified two genes (*ATP8A2*, *STARD13*) within the inversion region that may also be related to carotenoid pigmentation through their involvement in lipid transport. Specifically, *STARD13* produces a stAR-related lipid transfer protein, which as a protein family, are involved in intracellular lipid transport, metabolism, and cell signaling events<sup>33</sup>. While further validation studies are required to understand the role of these genes in carotenoid variation, their functions appear to be in line with other recently reported genes associated with carotenoid pigmentation.

Two additional genes within the chromosome 1 inversion region that could be affecting phenotype are well-characterized: *TSKU* and *FSTL1* are known antagonists of bone morphogenic protein (BMP) signaling<sup>24,26</sup>. However, the effects of BMP inhibition may influence phenotype in at least two disparate ways: through the regulation of melanogenesis, or by contributing to differences in bill morphology. BMPs are regulators with important roles in epidermal homeostasis and hair follicle growth and pigmentation<sup>34</sup>. Specifically, *BMP4* and *BMP6* products have both been demonstrated as inhibiting or stimulating melanogenesis, respectively<sup>34</sup>. However, other studies have also implicated

*BMP4* in the development of bird bill morphology<sup>35,36</sup>. For example, studies of *BMP4* in Darwin's Finches find strong correlations of *BMP4* expression with both bill depth and width<sup>35</sup>, two traits known to vary in redpolls<sup>18,37</sup>. Similar to Frizzled, *TSKU* was also shown to be differentially expressed in redpolls<sup>18</sup>. We therefore emphasize the observed differences in *TSKU* and *FSTL1* documented here could influence biologically important phenotypic variation in redpoll coloration, bill morphology, or both. Given the implication of *BMP4* in multiple pathways affecting different phenotypes, there could be pleiotropic effects resulting from one or more loci altering BMP signaling. Taken together, these candidate loci provide evidence that multiple aspects of redpoll phenotype are likely affected by a single genomic region maintaining associated SNPs from numerous genes in tight physical linkage.

While nearly all associated SNPs with gene annotations were within the chromosome 1 inversion region, three additional genes containing, or neighboring, associated SNPs may also have important phenotypic effects. Two of these—*FILIP1L* (chromosome 1 but outside of the inversion region), and *SFRP4* (chromosome 2)—act as regulators in the WNT pathway, suggesting they likely play roles in further modulating melanogenesis<sup>38,39</sup>. Similar to *TSKU* and *FSTL1*, *SFRP4* has also been demonstrated to regulate BMP, further emphasizing the

possibility of singular or joint effects on plumage coloration and bill morphology.

A third locus outside of the inversion region near an associated SNP on chromosome 2 includes a polyketide synthase (PKS). While this gene was annotated based on similarity to *Mycobacterium PKS15/1*, its function in birds has yet to be fully validated. However, its synteny with *RAB18*, and *YME1L1*, suggests homology with a PKS described in budgerigars (*Melopsittacus undulatus*)<sup>40</sup>. Functional validation through yeast-based expression demonstrated that the budgerigar PKS plays a critical role in the accumulation of red/yellow, parrot-specific pigments known as psittacofulvins. The association of PKS with redpoll phenotype indicates that it might play a similar role for organisms that contain carotenoids instead of psittacofulvins. While this requires further investigation, PKSs have been demonstrated elsewhere as important in animal pigment biosynthesis<sup>41</sup>.

**Evolutionary consequences.** Broadly, redpoll phenotype appears to function as a balanced polymorphism resulting from a 55-Mb inversion that affects plumage coloration and bill morphology. Genetic associations that include loci outside of the inversion region suggest that phenotype is likely modulated further by several independent gene regions to generate the varied forms seen across all redpoll ecotypes. Examination of genotypes at SNPs associated with redpoll morphology (Fig. 2b) suggest that the inversion region primarily separates the hoary redpoll from both the common and lesser redpolls, while the additional associations with other genomic regions separate the lesser redpoll from both the hoary and common redpolls. These results demonstrate that the chromosome 1 inversion contains multiple, linked genetic elements that together affect a suite of phenotypic traits in redpolls, providing evidence that redpoll phenotype is broadly controlled by a supergene genetic architecture<sup>11</sup>. As lesser redpolls form the darkest and smallest end of the redpoll phenotype distribution, the associated SNPs located outside the inversion may also be additive with respect to overall phenotype. Given the range restriction and more extreme phenotype of the lesser redpoll, there is less opportunity for disassortative mating, and it is unclear how the derived SNPs outside of the inversion that further modulate phenotype interact with the B inversion haplotype. A previous study of an avian supergene in white-throated sparrows (*Zonotrichia albicollis*)<sup>10</sup> demonstrated that one of the supergene haplotypes in sparrows had likely introgressed from a closely related species. However, topology weighting across windows of the redpoll supergene favored a topology that included a sister relationship between the two haplotypes, with a combined average weight of 54% among the three topologies that included this sister relationship (Supplementary Fig. 6), providing evidence that the redpoll supergene likely evolved within the redpoll lineage<sup>42</sup>.

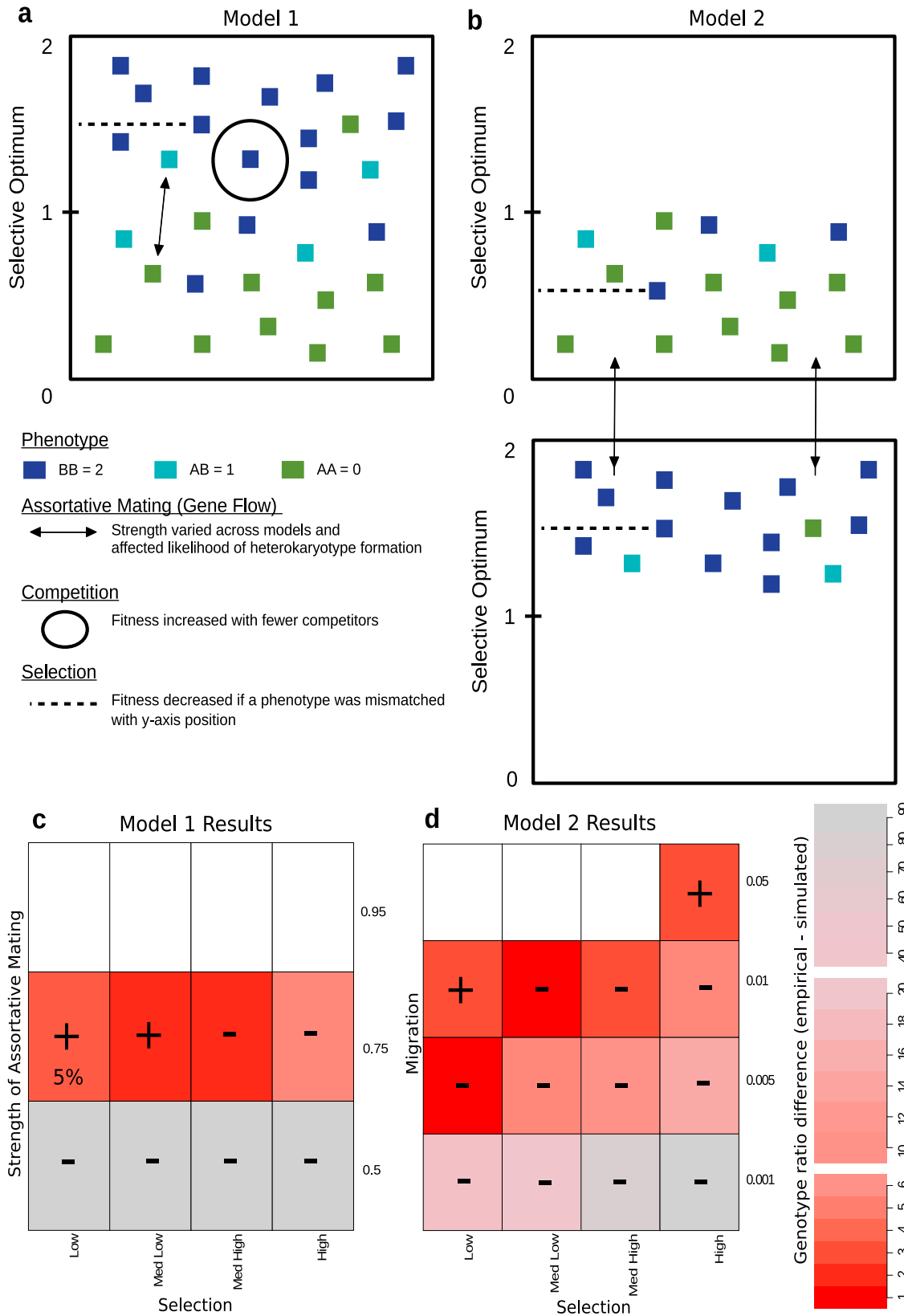
Considerable theoretical attention has recently been given to the evolution and degradation of supergenes<sup>43,44</sup>. A primary consequence of supergene-bearing inversions is increased mutational load<sup>12,44</sup> due to the difficulty of purging deleterious mutations in the absence (or severe reduction) of recombination. This simple scenario could result in a balanced polymorphism stemming from associative overdominance, where inversion heterozygotes perform best because heterozygosity masks some of the deleterious mutations<sup>12</sup>. However, redpolls heterozygous for the inversion appear to occur in fewer numbers than homozygotes (7/73 samples in this study), suggesting an alternative mechanism may be responsible for maintaining the polymorphism. Given the presence of all three inversion genotypes in redpolls, no combination of the supergene appears to be lethal—a finding in contrast to other recently described

supergenes of similar size<sup>9,45</sup>. Because there is no lethal inversion genotype, recombination likely occurs regularly in homozygotes (and possibly at low levels in heterozygotes), potentially allowing for some purging of deleterious mutations. This could have a considerable influence on the maintenance of the variation, and the evolutionary consequences of this supergene.

Understanding the effects of the redpoll supergene, and the forces responsible for its maintenance, is difficult. In the absence of selection (imposed by the environment, or through mate choice), the supergene would function as a single locus with one of the haplotypes eventually becoming fixed or lost due to drift<sup>12</sup>. Even with some selection, high levels of migration (between inversion genotypes) would swamp out any loci contributing to local adaptation. The persistence of the redpoll supergene is therefore likely dependent on both selection and migration. One scenario that is supported by field data is that the supergene remains balanced through assortative mating. Redpolls often mate assortatively<sup>46</sup>, but, intermediates and mixed pairs have also been observed from multiple localities<sup>37,47</sup>. Thus, the strictness of assortative mating may vary depending on the locality or may relax during irruptive population years<sup>47</sup>. Relaxation of mate choice and mixed pairings would produce the intermediate number of inversion heterozygotes seen in our data, and ultimately maintain the supergene as a stable polymorphism. However, this scenario alone does not provide an explanation for the maintenance of latitudinal differences between ecotypes. Furthermore, no hybrid zone has ever been documented in redpolls, which would be expected under a strict assortative mating scenario. While regions of hybridization have been suggested in places like Iceland, where high color variation exists<sup>48,49</sup>, previous genetic studies have not recovered support for this hypothesis<sup>50</sup>.

The phenotypes produced by the supergene are likely subject to environmentally mediated selection: notably, the more northerly distributed redpoll ecotype demonstrates features associated with high-latitude adaptation in other bird species (e.g., whiter color, smaller bill)<sup>51,52</sup>. Despite including some individuals sampled during the non-breeding season ( $n = 27$ ), we are able to detect differences in latitude by inversion genotype group, with B haplotypes significantly more common at higher latitudes (Fig. 3b). This pattern holds when examining only breeding season birds. While it is plausible that an alternative locus is affecting ecotypic distribution, the overall low levels of background genetic variation reflect ongoing gene flow within this system. This pattern could instead reflect incomplete lineage sorting and recent divergence times, however, tests for introgression using an ABBA-BABA framework detected a significant signal of gene flow among redpoll taxa ( $D = 0.0027$ ,  $p = 0.0003$ ). Gene flow among ecotypes would be expected to disrupt linkage between any latitude-associated loci and phenotype through recombination unless those loci were tightly linked as in an inversion. In light of the link between the redpoll supergene and phenotype and differences in breeding distribution between ecotypes<sup>18</sup>, the supergene may impart local adaptation to the environment. However, given the detection of inversion heterozygotes and the presence of gene flow, the inversion likely does not influence reproductive isolation. Thus, redpolls appear to function as a single species harboring ecotypic variation, rather than as three distinct species.

To explore the evolutionary conditions under which the observed pattern of the inversion polymorphism can remain balanced, we used the program SLiM<sup>53</sup> to simulate data under two spatial models of evolution informed by the aspects described above (Fig. 4a, b). Both models simulated 100-kb chromosomes, including a 50-kb inversion that contributed to phenotype, in diploid individuals<sup>54</sup>. The first model included one population with spatially varying selection along the y-axis to approximate



differences in fitness for a particular inversion genotype by latitude. In addition, we included assortative mating as determined by an adjustable parameter, and spatial competition such that individuals surrounded by fewer individuals in space received an increase in fitness. The second model also included spatially

varying selection but considered two ecotypes as two populations with gene flow. We then varied the strength of selection and the strength of assortative mating or migration parameters and quantified (1) whether or not a simulation resulted in a stable polymorphism, and (2) the inversion genotype ratios that were

**Fig. 4 Simulation models and results.** **a, b** 1000 diploid individuals were randomly placed in (x,y) coordinate space across one population in model 1 and two populations in model 2, respectively. Phenotype was additive, determined by inversion genotype, with blue squares representing BB individuals, green squares represent AA individuals, and cyan squares represent AB individuals. The strength of selection was determined for all individuals by a combination of the difference between phenotype and position along the y-axis (dotted line), and by a varied selection parameter. The strength of assortative mating or gene flow (solid arrow) was included in model 1 and model 2, respectively, and was varied across iterations. Model 1 also included a fitness adjustment for individuals based on the number of other individuals close by (solid circle). **c, d** Results from model 1 and model 2, respectively, show the similarity between the genotype ratios produced by simulation and genotype ratio recovered in our sampling of redpolls, where red represents the greatest similarity, gray represents the least similarity, and white represents simulations that failed to produce a stable polymorphism. Scale bar numbers represent the difference between the simulated and empirical genotype ratios with sign indicating either greater or fewer simulated heterozygotes. Numbers inside a cell indicate that only a portion of the 50 simulation iterations produced a stable polymorphism.

produced. We compared these ratios to the inversion genotype ratio in redpolls captured by our sampling. These simulations revealed that the strength of assortative mating, or amount of migration, played a larger role in the balancing of the inversion polymorphism than selection did at the levels tested (Fig. 4c, d; Supplementary Table 1). Regardless of the strength of selection, weak assortative mating or high migration invariably led to the loss of an inversion haplotype due to drift. Strong assortative mating or low levels of migration did maintain both haplotypes but failed to produce inversion heterozygotes. Further, all levels of selection produced spatial stratification of inversion genotypes along the selection gradient.

While these models are relatively simple and only represent two possibilities, they provide a starting point for further exploration of the complex dynamics that affect the maintenance of supergenes. For example, in redpolls, these simulations suggest that even very weak selection can produce the spatial variation seen among ecotypes, and that some relaxation of assortative mating is likely occurring, as has been proposed for populations in Canada and Alaska (USA)<sup>37</sup> and Norway<sup>47</sup>.

As whole-genome sequences proliferate, an emerging body of literature is providing empirical evidence of intraspecific variation maintained through inversion polymorphism<sup>9,10,14,15,55</sup>. The maintenance of redpoll ecotypes via an inversion across an environmental gradient places redpolls within this growing number of species. In some cases, such as monkeyflowers (*Mimulus guttatus*)<sup>14</sup>, inversion polymorphisms may confer sex-specific effects, and can be maintained within a population through a balance of positive and negative fitness interactions. In other cases, though not exclusive to sex-specific effects, inversions may affect phenotypes related to local adaptation, and species distributed across a heterogeneous environment may retain an inversion polymorphism through spatially varying selection, as suggested here for redpolls. This has recently been demonstrated in seaweed flies (*Coelopa frigida*)<sup>15</sup> and deer mice (*Peromyscus maniculatus*)<sup>55</sup>. In addition, studies of *Drosophila* have reported clinal variation in multiple survival traits controlled by an inversion as a result of spatially varying selection across latitude<sup>56</sup>. These findings in *Drosophila* highlight the need for further investigation into the selection pressures and fitness effects of the inversion we report in redpolls.

We provide evidence from whole-genome sequence data of loci associated with redpoll plumage coloration and bill morphology contained within a ~55-Mb inversion supergene. While some authorities classify redpolls as three separate species (e.g., ref. 57), we find no evidence of genome-wide population genetic structure consistent with current taxonomy. Instead, we provide evidence that the suite of morphological traits used to describe redpoll species differences are linked within the identified supergene. The presence of all possible inversion genotypes suggests there are no lethal supergene combinations and indicate that while these traits are likely involved in local adaptation, they are not involved in reproductive isolation. Though breeding distributions vary latitudinally, even minor levels of contemporary gene flow within

broad areas of sympatry likely maintain these traits as stable polymorphisms. Manipulations involving common garden experiments, or aviary crosses will help elucidate the strength of selection and may reveal additional unknown genetic interactions with the supergene that are affecting the evolution of redpolls.

With the explosive growth in the number of sequenced genomes and increasing sophistication of analytical tools, detecting structural variants or complex genetic architectures is likely to become common. The large size and high gene content of these classes of variants may in some cases translate to large evolutionary effects. While key theoretical work continues to emerge, the further exploration of empirical patterns between phenotype and supergenes will provide useful insight into the evolutionary effects of similar genetic architectures for a wide range of organisms.

## Methods

**Sampling.** We sampled 73 individuals from across the three redpoll species currently recognized by many authorities (e.g., ref. 57), including common redpoll ( $n = 26$ ), hoary redpoll ( $n = 33$ ), and lesser redpoll ( $n = 14$ , Fig. 1, Supplementary Data File 1) (Fig. 1 range map generated using ref. 58). We extracted genomic DNA using a salt extraction protocol. Samples were first lysed using a homogenizing solution (0.4 M NaCl, 10 mM Tris-HCl pH 8.0, and 2 mM EDTA pH 8.0, Proteinase K), and a 20% SDS solution. We added 2  $\mu$ l of glycoblue dye to aid in the identification of the DNA pellet, and precipitated DNA using a 6 M NaCl solution and 100% EtOH. DNA pellets were resuspended in 100  $\mu$ l of TE buffer (10 mM Tris, 1 mM EDTA at pH 8–9). We prepared genomic libraries using the Nextera XT kit with half-reaction volumes. We pooled all 73 individuals, and sequenced whole genomes using two lanes of an S4 flow cell on an Illumina Novaseq (Illumina Inc., CA, USA). The collection and handling of all samples were done with approval and in accordance with the ethical guidelines set out by the University of Colorado Boulder Institutional Animal Care and Use Committee, the University of Iceland, Reykjavik Institution of Life and Environmental Science, the Czech Academy of Sciences Institute of Vertebrate Biology, the Greenland Home Rule Government, Danish Polar Center, and the Cornell University Institutional Animal Care and Use Committee.

Raw reads were trimmed using TrimmomaticPE<sup>59</sup> and aligned to a brown-capped rosy-finch (*Leucosticte australis*) reference genome using the BWA mem algorithm with default settings<sup>60</sup>. We called variants using bcftools mpileup<sup>61</sup>, and filtered to keep only single nucleotide polymorphisms (SNPs) with a quality score higher than 80. We dropped one individual (R97) due to its sibling status with another individual. We removed all potentially paralogous loci by filtering SNPs with a depth lower than 2x and higher than 12x coverage, removed all SNPs with a minor allele frequency lower than 5%, and generated multiple datasets based on allowed missing data using VCFtools v 0.1.16<sup>62</sup>. We ran subsequent analyses on datasets allowing for no missing data (100p), and 25% missing data (75p) with concordant results between both datasets. We present results from the 75p dataset, unless noted otherwise. Scripts used in our bioinformatic pipeline are publicly available on Github ([https://github.com/erikfunk/whole\\_genome\\_bioinformatics](https://github.com/erikfunk/whole_genome_bioinformatics))<sup>63</sup>.

In addition, we selected two individuals to generate long-read whole genomes resequenced using Oxford Nanopore Technologies MinION sequencing platform. These individuals (R29, R47) were selected as representatives of two distinct inversion genotypes within redpolls. Library preparation and sequencing were carried out at Colorado State University using the SQK-LSK-109 Ligation Sequencing Kit (Oxford Nanopore Technologies, OX, UK), across 5 flow cells. Long reads were mapped using NGMLR v0.2.7<sup>64</sup> with default settings. Resulting bam alignment files were used to detect structural variants with the program SVIM v1.4.2<sup>21</sup>. We adjusted the maximum detectable variant size to 60 Mb using the `-max_sv_size` argument.

**Clustering analyses.** To visualize population genomic structure, we ran principal component analyses using the R package SNPrelate v1.19.3<sup>65</sup>. To further assess



clustering of individuals by species designation, we ran the program conStruct v.1.0.4<sup>66</sup>. We assessed models for all number of populations from  $K=1$  to  $K=6$  using both spatial and non-spatial models. Model performance was evaluated using a combination of cross validation and layer contribution scores. We selected the best value of  $K$  as the model that exhibited the largest gains in predictive accuracy, while maintaining observable layer contributions.

**Diversity and linkage statistics.** We assessed heterogeneity in divergence across the genomic landscape using windowed calculations of diversity statistics for variant sites, including  $P_i$ ,  $F_{ST}$ , and  $d_{XY}$ . All calculations were made using the python script popgenWindows.py ([https://github.com/simonhmartin/genomics\\_general.com](https://github.com/simonhmartin/genomics_general.com)). We conducted these genome scans as pairwise comparisons between named species, and distinct clusters of individuals identified in PCAs. After identification of the chromosome 1 inversion, we calculated heterozygosity of the inversion for each of the three PCA clusters using R package adegenet v.2.1.3<sup>67</sup>.

To evaluate linkage across the chromosome 1 inversion, we calculated linkage disequilibrium (LD) as  $r^2$  using plink v1.9<sup>68</sup>. LD was calculated between pairwise SNPs across each chromosome for the entire genome. We also calculated LD separately along sections of chromosome 1 that were outside, and inside the inversion. LD for each region was averaged by combining all calculations of  $r^2$  for a given distance between SNPs, resulting in a plot of LD decay.

**Association and annotations.** We used two approaches to identify regions of the genome associated with PCA clusters and phenotypic differences:  $F_{ST}$  peaks, and a genome-wide mixed model analysis using GEMMA<sup>22</sup>. While continuous variation exists in redpolls, species classifications are based on morphological characters, including plumage coloration, and bill size and shape. We used species classification as a summary of phenotype to test for genetic associations in GEMMA and ran analyses using sex as a covariate, and without using any covariates. No differences were recovered between these two analyses and we report the results with sex as a covariate here. We included a matrix of relatedness, generated using the -gk 1 command within GEMMA. Positions of significant SNPs identified by GEMMA, and peak regions of  $F_{ST}$  were used to compile lists of genes that either encompassed or neighbored these SNPs (within 100-kb) from the annotated brown-capped rosy-finch reference genome. We numbered our chromosomes based on gene content and synteny with the zebra finch (*Taeniopygia guttata*). Significance for associated SNPs were drawn using a false discovery rate of  $p < 1e-5$ . To evaluate the degree to which the associated SNPs can be used to explain phenotype, we used GEMMA to performed a leave-one-out cross validation using a Bayesian sparse linear mixed model (BSLMM) and the -predict 1 argument. To generate a predicted phenotype, we systematically dropped the phenotype for a single individual, rerunning the model each time. This resulted in a separate model, and a predicted phenotype for each individual. The variance in predicted phenotype was quantified using a regression against observed phenotype.

To evaluate how the genes identified by each approach may be playing a role in the generation of divergent phenotypes, we examined known gene ontologies using Panther v.16<sup>69</sup>. We tested for overrepresented GO categories and examined functional classification across three GO databases (molecular function, cellular component, and biological process), as well as signaling and metabolic pathways. In addition, we compared our gene lists to the over- and under-expressed gene lists produced by Mason and Taylor<sup>18</sup>. Finally, we categorized each redpoll variant by comparing its chromosomal position to the gene model coordinates in our reference genome using SNPeff v4.3<sup>23</sup>. Based on its position with respect to gene model reading frames, variants were classified as either intergenic, upstream/downstream, intronic, synonymous, or missense mutations.

**Supergene origin and maintenance.** To explore the possibility that one of the inversion haplotypes introgressed into redpolls from a closely related species, we used phylogenetic methods to test if the two inversion haplotypes were (1) sister, indicating an origin within the redpoll lineage, or (2) not sister, indicating an origin outside of the redpoll lineage followed by introgression. We generated phylogenies using 50, 100, and 200 SNP windows along the chromosome 1 inversion using Twisst<sup>42</sup>. Results were congruent among the three different window sizes so we only present results from the 100 SNP window analyses. Trees were generated using 4 redpolls, 2 from each of the inversion homozygote groups, and 9 additional individuals from across 5 taxa at varying degrees of divergence from redpolls. We selected these taxa based on the most recently published phylogeny for the family Fringillidae<sup>70</sup>. These additional individuals included two species grouped as crossbills (*Loxia leucoptera* and *Loxia curvirostra*), two species grouped as rosy-finches (*Leucosticte atrata* and *Leucosticte tephrocotis*), and the tree was rooted using a *Fringilla coelebs* genome from the NCBI Sequence Read Archive (SRR11537170).

All sequences were aligned using the same brown-capped rosy-finch (*Leucosticte australis*) reference genome and the same bioinformatics pipeline described above for redpolls. We converted the resulting VCF file into a .geno file using the parseVCF.py python script from Simon Martin (<https://github.com/simonhmartin>) and generated rooted phylogenies using PhyML<sup>71</sup>. PhyML was run

using the Twisst script phymL\_sliding\_windows.py with default settings. Topology weights were calculated and visualized in R using the plot\_twisst.R script.

The evolutionary processes that act on a population in order to maintain a balanced polymorphism are likely complex and numerous. As a first step in understanding the maintenance of the supergene polymorphism in redpolls, we simulated 100-Kb chromosomes with a 50-kb inversion in 1000 diploid individuals using the program SLiM<sup>53</sup> under two different models of evolution. Individual phenotype was determined by the number of copies of the inversion an individual possessed. Each model included two spatial dimensions, with selection varying along the y-axis. The selective optimum was determined by an individual's spatial position and implemented as a fitness adjustment based on the difference between an individual's position and their phenotype. Model 1 simulated a single population with assortative mating and spatial competition, while model 2 simulated two populations (one corresponding to each homozygote group), allowing for migration each generation. Each simulation was allowed to run for 10,000 generations and was iterated 50 times. Each simulation also varied in one of two parameters, including selection, strength of assortative mating (model 1), or amount of migration (model 2). We generated a custom output that tallied the total number of iterations that resulted in a balanced polymorphism (i.e., both haplotypes still present) and extracted the average count of each inversion genotype, along with its spatial position. Eidos code for the models we used can be found on github at [https://github.com/erikfunk/redpoll\\_slim\\_models](https://github.com/erikfunk/redpoll_slim_models).

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

## Data availability

The raw sequence data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) database under accession code PRJNA753137. Genomic datasets are available at the Dryad Digital Repository [<https://doi.org/10.5061/dryad.q83bk3jjm>]. Sequence data for *Fringilla coelebs* is available from the NCBI Sequence Read Archive (SRR11537170). The Panther Gene Ontology database is available online at [<http://www.pantherdb.org/>]. Source data are provided with this paper.

## Code availability

Previously published code used in this study is available on github at [https://github.com/erikfunk/whole\\_genome\\_bioinformatics](https://github.com/erikfunk/whole_genome_bioinformatics) [<https://doi.org/10.5281/zenodo.5542029>], [https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general), and <https://github.com/simonhmartin/twisst>. Eidos code written for the SLiM models used in this study is available at [https://github.com/erikfunk/redpoll\\_slim\\_models](https://github.com/erikfunk/redpoll_slim_models) [<https://doi.org/10.5281/zenodo.5542015>].

Received: 30 April 2021; Accepted: 8 November 2021;

Published online: 25 November 2021

## References

1. The Heliconius Genome Consortium. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
2. Jones, M. R. et al. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**, 1355–1358 (2018).
3. Oziolor, E. M. et al. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* **364**, 455–457 (2019).
4. Westram, A. M. et al. Clines on the seashore: the genomic architecture underlying rapid divergence in the face of gene flow. *Evol. Lett.* **2**, 297–309 (2018).
5. Hansen, T. E. The evolution of genetic architecture. *Annu. Rev. Ecol. Evol. Syst.* **37**, 123–157 (2006).
6. Campagna, L. et al. Repeated divergent selection on pigmentation genes in a rapid finch radiation driven by sexual selection. *Sci. Adv.* **3**, e1602404 (2017).
7. Turbek, S. P. et al. Rapid speciation via the evolution of pre-mating isolation in the Iberá Seedeater. *Science* **371**, eabc0256 (2021).
8. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends Ecol. Evol.* **34**, 531–544 (2019).
9. Küpper, C. et al. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).
10. Tuttle, E. M. et al. Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350 (2016).
11. Thompson, M. J. & Jiggins, C. D. Supergenes and their role in evolution. *Hereditas* **113**, 1–8 (2014).
12. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
13. Faria, R., Johannesson, K., Butlin, R. K. & Westram, A. M. Evolving inversions. *Trends Ecol. Evol.* **34**, 239–248 (2019).

14. Lee, Y. W., Fishman, L., Kelly, J. K. & Willis, J. H. A segregating inversion generates fitness variation in yellow monkeyflower (*Mimulus guttatus*). *Genetics* **202**, 1473–1484 (2016).
15. Mérot, C. et al. Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msab143> (2021).
16. Clement, P. In *Handbook of the Birds of the World* (eds del Hoyo, J., Elliott, A. & Christie, D. A.) 564–565 (Lynx Edicions, 2010).
17. Clement, P. In *Handbook of the Birds of the World* (eds del Hoyo, J., Elliott, A. & Christie, D. A.) 565–566 (Lynx Edicions, 2010).
18. Mason, N. A. & Taylor, S. A. Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Mol. Ecol.* **24**, 3009–3025 (2015).
19. Pearse, D. E. et al. Sex-dependent dominance maintains migration supergene in rainbow trout. *Nat. Ecol. Evol.* **3**, 1–12 (2019).
20. Kim, K. W. et al. A sex-linked supergene controls sperm morphology and swimming speed in a songbird. *Nat. Ecol. Evol.* **1**, 1168–1176 (2017).
21. Heller, D. & Vingron, M. SVIM: Structural variant identification using mapped long reads. *Bioinformatics* **35**, 2907–2915 (2019).
22. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
23. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly. (Austin)* **6**, 80–92 (2012).
24. Geng, Y. et al. Follistatin-like 1 (Fstl1) is a bone morphogenetic protein (BMP) 4 signaling antagonist in controlling mouse lung development. *Proc. Natl Acad. Sci. USA* **108**, 7058–7063 (2011).
25. Yamada, T. et al. Melanocyte stem cells express receptors for canonical Wnt-signaling pathway on their surface. *Biochem. Biophys. Res. Commun.* **396**, 837–842 (2010).
26. Ohta, K. et al. Tsukushi functions as an organizer inducer by inhibition of BMP activity in cooperation with chordin. *Dev. Cell* **7**, 347–358 (2004).
27. Körner, A. & Pawelek, J. Mammalian tyrosinase catalyzes three reactions in the biosynthesis of melanin. *Science* **217**, 1163–1165 (1982).
28. Bertolotto, C. et al. Different *cis*-acting elements are involved in the regulation of TRP1 and TRP2 promoter activities by cyclic AMP: pivotal role of M boxes (GTCATGTGCT) and of microphthalmia. *Mol. Cell. Biol.* **18**, 694–702 (1998).
29. Serre, C., Busuttill, V. & Botto, J. M. Intrinsic and extrinsic regulation of human skin melanogenesis and pigmentation. *Int. J. Cosmet. Sci.* **40**, 328–347 (2018).
30. Schulte, G. The class frizzled receptors. *Pharmacol. Rev.* **62**, 632–667 (2010).
31. Toomey, M. B. et al. High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proc. Natl Acad. Sci. USA* **114**, 5219–5224 (2017).
32. Brelford, A., Toews, D. P. L. & Irwin, D. E. Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proc. R. Soc. B* **284**, 20171106 (2017).
33. Soccio, R. E. & Breslow, J. L. STAR-related lipid transfer (START) proteins: mediators of intracellular lipid metabolism. *J. Biol. Chem.* **278**, 22183–22186 (2003).
34. Singh, S. K., Abbas, W. A. & Tobin, D. J. Bone morphogenetic proteins differentially regulate pigmentation in human skin cells. *J. Cell Sci.* **125**, 4306–4319 (2012).
35. 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).
36. Knief, U. et al. QTL and quantitative genetic analysis of beak morphology reveals patterns of standing genetic variation in an Estrildid finch. *Mol. Ecol.* **21**, 3704–3717 (2012).
37. Troy, D. M. A phenetic analysis of the redpolls *Carduelis flammea flammea* and *C. hornemanni exilipes*. *Auk* **102**, 82–96 (1985).
38. Kwon, M. et al. Filamin A interacting protein 1-like inhibits WNT signaling and MMP expression to suppress cancer cell invasion and metastasis. *Int. J. Cancer* **135**, 48–60 (2014).
39. Berndt, T. et al. Secreted frizzled-related protein 4 is a potent tumor-derived phosphatidic acid. *J. Clin. Invest.* **112**, 785–794 (2003).
40. Cooke, T. F. et al. Genetic mapping and biochemical basis of yellow feather pigmentation in budgerigars. *Cell* **171**, 427–439 (2017).
41. Calestani, C. & Wessel, G. M. *These Colors Don't Run: Regulation of Pigment Biosynthesis in Echinoderms*. (Springer International Publishing, 2018).
42. Martin, S. H. & Bellegghem, Van S. M. Exploring evolutionary relationships across the genome using topology weighting. *Genetics* **206**, 429–438 (2017).
43. Gutiérrez-valencia, J., Hughes, W., Berdan, E. L. & Slotte, T. The genomic architecture and evolutionary fates of supergenes. *Genome Biol. Evol.* **13**, evab057 (2021).
44. Berdan, E. L., Blanckaert, A., Butlin, R. K. & Bank, C. Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genet.* **17**, e1009411 (2021).
45. Wang, J. et al. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).
46. Lifjeld, J. T. & Bjerke, B. A. Evidence for assortative pairing by the cabaret and flammea subspecies of the common redpoll *Carduelis flammea* in SE Norway. *Fauna Nor. Ser. C. Cinclus* **19**, 1–8 (1996).
47. Harris, M. P., Norman, F. I. & Mccoll, H. S. A mixed population of redpolls in northern Norway. *Br. Birds* **58**, 288–294 (1965).
48. Amouret, J., Hallgrímsson, G. T., Kolbeinsson, Y. & Pálsson, S. Morphological differentiation of Icelandic Redpolls, *Acanthis flammea islandica*. *Bird. Study* **63**, 37–45 (2016).
49. Herremans, M. Taxonomy and evolution in redpolls *Carduelis flammea-hornemanni*; a multivariate study of their biometry. *Ardea* **78**, 441–458 (1990).
50. Amouret, J., Steinauer, K., Hallgrímsson, G. T. & Pálsson, S. Evolutionary status of icelandic redpolls *Carduelis flammea islandica* (Aves, Passeriformes, Fringillidae). *J. Ornithol.* **156**, 1035–1048 (2015).
51. Zink, R. M. & Remsen, J. V. Evolutionary processes and patterns of geographic variation in birds. *Curr. Ornithol.* **4**, 1–69 (1986).
52. Symonds, M. R. E. & Tattersall, G. J. Geographical variation in bill size across bird species provides evidence for Allen's rule. *Am. Nat.* **176**, 188–197 (2010).
53. Haller, B. C. & Messer, P. W. SLiM 3: forward genetic simulations beyond the Wright-Fisher model. *Mol. Biol. Evol.* **36**, 632–637 (2019).
54. Funk, E. R. et al. A supergene underlies linked variation in color and morphology in a Holarctic songbird. Redpoll slim models. <https://doi.org/10.5281/zenodo.5542015> (2021).
55. Hager, E. R. et al. A chromosomal inversion drives evolution of multiple adaptive traits in deer mice. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.01.21.427490> (2021).
56. Durmaz, E., Benson, C., Kapun, M., Schmidt, P. & Flatt, T. An inversion supergene in *Drosophila* underpins latitudinal clines in survival traits. *J. Evol. Biol.* **31**, 1354–1364 (2018).
57. Clements, J. F. et al. The eBird/Clements checklist of birds of the world: v2019. <http://www.birds.cornell.edu/clementschecklist/download> (2019).
58. BirdLife International and NatureServe. Bird species distribution maps of the world (BirdLife International, Cambridge, UK and NatureServe, Arlington, USA, 2015).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics* **30**, 2114–2120 (2014).
60. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2009).
61. Wyszoker, A. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
62. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
63. Funk, E. R. *GitHub* (2021).
64. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. *Nat. Methods* **15**, 461–468 (2018).
65. Zheng, X. et al. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
66. Bradburd, G. S., Coop, G. M. & Ralph, P. L. Inferring continuous and discrete population genetic structure across space. *Genetics* **210**, 33–52 (2018).
67. Jombart, T. ADEGENET: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
68. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 1–16 (2015).
69. Mi, H. et al. PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, 394–403 (2021).
70. Zuccon, D., Prys-Jones, R., Rasmussen, P. C. & Ericson, P. G. P. The phylogenetic relationships and generic limits of finches (Fringillidae). *Mol. Phylogenet. Evol.* **62**, 581–596 (2012).
71. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

## Acknowledgements

We would like to thank David Toews, Erica Larson, Erik Enbody, Ethan Linck, and members of the Taylor Lab for comments on previous versions of this manuscript, and Aaron Westmoreland for comments on our SLiM models. We would like to thank the Natural History Museum in Oslo, the Yale Peabody Museum, the University of Alaska Museum, the Cornell University Museum of Vertebrates, and the University of Washington Burke Museum for providing samples for this project, along with Craig Benkman for providing the crossbill genome sequences used in our topology weighting analyses. We would also like to thank Kurt Burnham and the High Arctic Institute for field support in obtaining samples in Greenland, with permits provided by the Greenland Home Rule Government. T.A. would like to thank the Czech Science Foundation (projects 15-11782S and 19-22538S) for funding and the Bird Ringing Centre at the National Museum in Prague (namely Jaroslav Cepak and Petr Klvna) for their support during the sampling of redpolls. We would like to thank the Society of Systematic

Biologists, American Ornithological Society, and the Denver Field Ornithologists for providing funding for this project.

### Author contributions

S.A.T., E.R.F., and N.A.M. conceived the study. E.R.F. performed data analysis, generated figures, and wrote the manuscript with assistance from S.A.T. N.A.M., J.J., T.A., and S.P. provided critical samples and manuscript revisions.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27173-z>.

**Correspondence** and requests for materials should be addressed to Erik R. Funk.

**Peer review information** *Nature Communications* thanks Emily Moore, Rusty Gosner, and the other anonymous reviewer(s) for their contribution to the peer review this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021