

SCIENTIFIC REPORTS



OPEN

Genomic regions of speciation and adaptation among three species of grouse

Radoslav Kozma, Patrik Rödin-Mörch & Jacob Höglund 

Understanding the molecular basis of adaptation is one of the central goals in evolutionary biology and when investigated across sister species it can provide detailed insight into the mechanisms of speciation. Here, we sequence the genomes of 34 individuals from three closely related grouse species in order to uncover the genomic architecture of speciation and the genes involved in adaptation. We identify 6 regions, containing 7 genes that show lineage specific signs of differential selection across the species. These genes are involved in a variety of cell processes ranging from stress response to neural, gut, olfactory and limb development. Genome wide neutrality test statistics reveal a strong signal of population expansion acting across the genomes. Additionally, we uncover a 3.5 Mb region on chromosome 20 that shows considerably lower levels of differentiation across the three grouse lineages, indicating possible action of uniform selection in this region.

The identification of genes involved in species' adaptation to their environment is one of the main aims of evolutionary biology¹. By uncovering the genetic basis of adaptation we improve our understanding of the mechanisms responsible for adaptation and speciation and by which phenotypic diversity arises. Until now, this was easier to achieve in domesticated animals where the strong artificial selection and constant, human-monitored environment quickly drove favoured alleles to fixation (e.g. chicken² and horses³). This approach has proved fruitful in forming the foundations for the study of speciation. But in order to understand the way natural selection shapes individuals and species, study of wild populations has to be undertaken. This allows the examination of the architecture of adaptive traits under natural processes, free from intentional human-driven selection^{4,5}.

One tool that enables the identification of genomic regions involved in speciation is the F_{ST} outlier method. By comparing the amount of differentiation at a particular locus to the overall levels of differentiation across the genome, the method indicates whether selection is acting upon any particular region. Higher levels of differentiation signal the presence of positive/directional selection, while lower levels of differentiation signal the presence of balancing/purifying selection. For example, the technique has been used to identify the gene controlling beak morphology in Darwin's finches⁶, plumage colour in crows⁷ and adaptation to hypoxia in wolves⁸.

In this study, the aim is to utilise outlier methods to analyse the genomic architecture of species specific differences by exploring the genes involved in adaptation in three grouse; the willow grouse (*Lagopus lagopus lagopus*), the red grouse (*Lagopus lagopus scoticus*) and the rock ptarmigan (*Lagopus muta*). All belong to subfamily Tetraoninae and the *Lagopus* common ancestor diverged from the lineage leading to black grouse (*Tetrao tetrix*) and relatives approximately 3 million years ago (Mya)^{9,10}. Consequently, the rock ptarmigan speciated from the *Lagopus lagopus* lineage around 2-1 Mya^{9,10}. The red grouse is recognised as a subspecies of the willow grouse¹¹, however it is endemic to the moorlands of Great Britain and Ireland and has thus been separated from continental populations at least since the time the British Isles separated from mainland Europe (~6000 years ago^{12,13}). No record of any current gene flow between red grouse and willow grouse exists, with even the more proximal Irish and British red grouse showing no gene flow and substantial genetic differentiation¹⁴. Both the rock ptarmigan and the willow grouse have a circumpolar distribution, however, the rock ptarmigan occurs on Greenland, Iceland and Svalbard while the willow grouse does not^{11,15}. Furthermore, the rock ptarmigan is the most cold adapted of the three, whereby it is a sedentary species that breeds across open arctic and subarctic habitats¹⁶. The willow grouse inhabits open subalpine habitat, boreal forests and moorland¹¹. Where the distribution of the two species overlaps, the rock ptarmigan occurs at higher altitudes^{17,18}. The three taxa also exhibit differences

Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, Uppsala, SE-75236, Sweden. Correspondence and requests for materials should be addressed to J.H. (email: jacob.hoglund@ebc.uu.se)

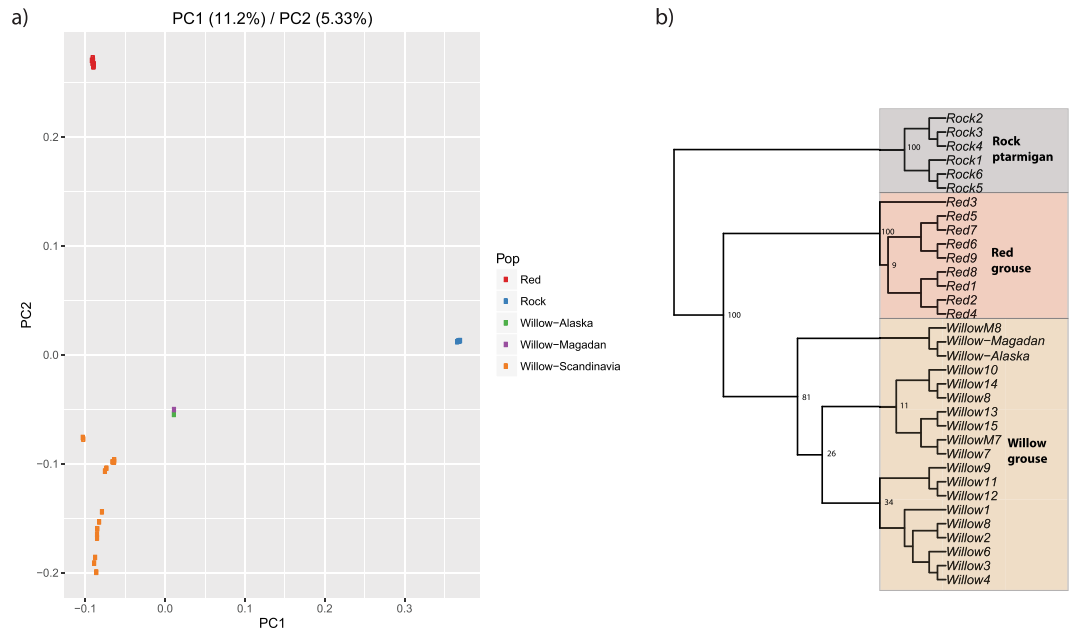


Figure 1. (a) Principal component analysis (PCA) and (b) Maximum likelihood inferred cladogram of the 34 grouse individuals used in the study (see Table S1 for sampling location of each individual). Bootstrap support values are given at each branch.

in plumage colouration; the rock ptarmigan and willow grouse have a brown summer plumage (full in females and partial in males) which moults into an all-white winter plumage, while both sexes of the red grouse forgo the white winter plumage instead remaining brown all year round¹¹. The varying levels of arduous habitat and differences in plumage as well as population history thus provide a compelling system for the study of molecular adaptation and speciation.

Results and Discussion

Population structure and phylogeny. In total we identified 8 307 719 variable sites across the three taxa and a PCA based on these sites reveals a clear species clustering. PC1 separates the rock ptarmigan from the other two taxa (explaining 11.2% of the total variation) while PC2 further separates the willow and red grouse (explaining 5.33% of the total variation) (Fig. 1a). In the PCA, the willow grouse shows the most overall intraspecific variation, where the Scandinavian individuals show relatively large scatter while the Siberian and Alaskan individuals are more differentiated still (Fig. 1a). The red grouse and rock ptarmigan individuals show little intraspecific variation and cluster tightly together, but these differences in comparison to willow grouse most certainly is due to the relatively smaller and localised sample sizes for the latter two species. The species tree based on a random 59 574 autosomal SNP subset is in line with previously published phylogenies of the species complex^{11,13,19}, whereby the red grouse is the sister taxon to the willow grouse (Fig. 1b). Intraspecific divergence largely followed expectations based on geographic origin.

Genomic regions associated with differentiation. Overall levels of differentiation follow the prediction set by the phylogeny, with the lowest differentiation being between the willow grouse and red grouse (mean $F_{ST} = 0.08 \pm 0.04$), followed by willow grouse and rock ptarmigan (mean $F_{ST} = 0.17 \pm 0.07$) and the largest differentiation being between the red grouse and rock ptarmigan (mean $F_{ST} = 0.21 \pm 0.08$, Supplementary Fig. 4).

Taxon specific test of neutrality revealed strong demographic effects acting across the whole genome in all three taxa. Tajima's D was highly negative (genome-wide mean \pm SD: willow grouse = -2.12 ± 0.13 , red grouse = -2.31 ± 0.14 , rock ptarmigan = -1.88 ± 0.09) while Fay and Wu's H was positive (genome-wide mean \pm SD: willow grouse = 0.32 ± 0.01 , red grouse = 0.28 ± 0.01 , rock ptarmigan = 0.37 ± 0.02 , Supplementary Figs S5–S11), indicating strong signatures of population expansion in all three taxa.

The ZF_{ST} outlier test detected 2 non-overlapping regions for each taxon that were identified in more than one comparison (Table 1, Fig. 2). For the willow grouse these are located on chromosome 2 (position ~95 Mb) and 26; for the red grouse they are on chromosome 2 (position ~81 Mb) and Z (position ~61 Mb); for the rock ptarmigan they are on chromosome 20 and Z (position ~43 Mb). These regions include a number of genes as deduced from the alignment to the chicken genome - willow grouse: *CDH7* and *FOXP4*; red grouse: *SUN3* and *EDIL3*; rock ptarmigan: *ROMO1*, *CPNE1* and *GADD45A* (Fig. 3). Both *CDH7* and *FOXP4* are important developmental genes, with *FOXP4* being involved in neural, gut and pulmonary development and *CDH7* being integral in neural, optic, branchial and olfactory development, whose mutations in humans gives rise to the multiple-malformation 'CHARGE' syndrome^{20,21}. *CPNE1* is part of the calcium-dependent membrane-binding protein family and plays a role in the neuronal progenitor cell differentiation²². *ROMO1* and *GADD45A* on the other hand are both involved in cellular stress response following reactive oxygen species (ROS) accumulation and environmental stress, respectively^{23,24}. *EDIL3* (also known as *DEL-1*) is a gene involved in the cardiovascular system that has been

Species	Chromosome	Window (Mb)	Gene	Function
Willow grouse	2	95.610–95.625	<i>CDH7</i>	Cell adhesion protein involved in neural, optic, branchial and olfactory development
Willow grouse	26	4.800–4.815	<i>FOXP4</i>	Transcription factor involved in neural, gut and pulmonary development
Red grouse	2	81.855–81.870	<i>SUN3</i>	Protein anchor involved in cytoskeleton anchoring and spermatogenesis
Red grouse	Z	61.710–61.725	<i>EDIL3</i>	Integrin ligand involved in angiogenesis and vessel wall development
Rock ptarmigan	20	1.095–1.110	<i>ROMO1</i>	Involved in DNA damage and replication senescence by reactive oxygen species (ROS) build up
			<i>CPNE1</i>	Involved in calcium dependent neuronal cell differentiation
Rock ptarmigan	Z	43.650–43.665	<i>GADD45A</i>	Involved in cellular response to environmental stresses by activation of the p38/JNK pathway

Table 1. Summary of the outlier windows detected across the three F_{ST} comparisons.

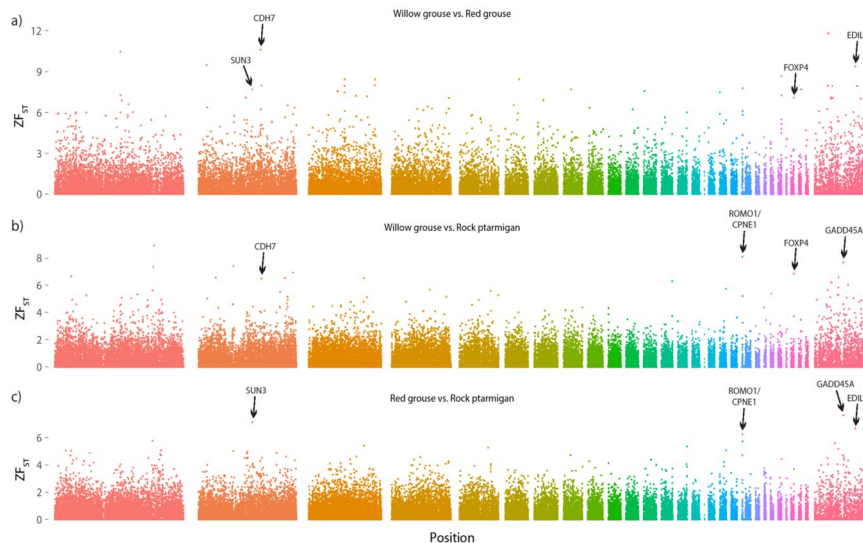


Figure 2. Genome-wide F_{ST} outlier test, showing the three pairwise comparisons of the three study taxa; (a) willow grouse vs. red grouse, (b) willow grouse vs. rock ptarmigan and c) red grouse vs. rock ptarmigan. The y-axis shows the ZF_{ST} score, where a 15 kb window with a score ≥ 6 is deemed an outlier. Each colour represents a different chromosome, with the autosomes arranged 1–28 (left to right) and chromosome Z located on the far right. The genes lying within same outlier window detected in two of the three comparisons are shown.

shown to play a key function in angiogenesis in mice^{25,26}. Lastly, *SUN3* is a gene involved in cytoskeletal anchoring and has been shown to play a role in mammal spermatogenesis²⁷.

FOXP4, which was found to be under positive selection in the willow grouse, is an interesting candidate gene for speciation. It is part of the forkhead box (FOX) group of transcription factors²⁸, another member of which (*FOXP2*) has been shown to be highly involved in neural development, specifically important for learning of bird song and human speech^{29,30}. Moreover, the product of the *FOXP4* gene has been shown to function in similar fashion to *FOXP2*³¹. This is interesting, because unlike in passerine birds, the technicality and repertoire of willow grouse calls are comparatively nominal. This raises two potential issues. Firstly, it is perhaps possible that the call repertoire of the willow grouse hides more intricacies than once thought and would thus select for higher call learning capability. Or, it may be the case that *FOXP4* has a more ubiquitous role in bird calls and may be selected for even in species without intricate song repertoires. Further complicating the picture, studies on the mouse have shown that the *FOXP4* transcription factor is also expressed in the developing pulmonary and gut tissue³², indicating the potential that this gene may also be the target of diet induced selection.

Even less is known about the specific function in birds of the remaining positively selected genes. *CDH7* appears to have a broad effect on early development, being involved in neural, optic, brachial and olfactory ontogeny^{20,33}. Further studies in the chicken also reveal the gene's important function in limb development, by regulation of limb bud mesenchymal cell motility and migration³⁴. Due to such broad scope of possible traits in which *CDH7* can have an effect in the willow grouse it requires further research and insight to be able to specify more precisely what are the ramification of positive selection acting on this gene.

No experimental evidence exists for the function of *ROMO1* and *GADD45A* in birds; therefore it is hard to specify exactly why these genes are selected for in the rock ptarmigan. However, out of the three taxa studied

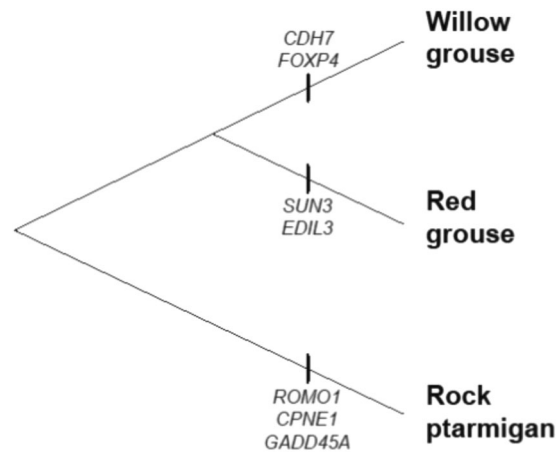


Figure 3. Genes under divergent selection in each of the three grouse lineages, as revealed by the F_{ST} outlier test.

here, the rock ptarmigan lives in the most extreme environment, so genes that would help cope with the effects of increased stress are likely to be more important for this species than for the other studied grouse. Likewise, no experimental evidence of *CPNE1*, *SUN3* or *EDIL3* in birds exists. Therefore, more will have to be known before we can hypothesise why any of the genes identified in this study might be positively selected in their respective grouse taxon. Furthermore, in all these regions of speciation, the identification of individual SNPs will aid in determining the overall target and effect of selection.

One methodological aspect of the F_{ST} outlier analysis is that it judges a window to be an outlier by looking at the number of standard deviations it is away from the overall mean. And since the range of F_{ST} values is bound between 0 and 1, the larger the overall mean, the harder it is for a window to be considered an outlier. As such, much fewer outliers were identified in the comparisons involving the rock ptarmigan, because the overall level of differentiation is higher (see mean F_{ST} values in Results and Supplementary Fig. S4). This was a conservative approach to eliminate capturing false positives. By lowering the threshold (e.g. consider anything with a $ZF_{ST} > 5$ to be an outlier) might reveal more windows that are consistently identified as outliers among the different comparisons. However, different methods for identifying loci under selection consistently identify different loci even when applied to the same data^{35,36}. To avoid false positives we felt a conservative cut-off was appropriate. This is also why we have chosen the same cut-off for all species contrasts.

Region of low differentiation. Apart from finding regions of high differentiation, the F_{ST} outlier analysis also revealed a ~3.5 Mb region of low differentiation on chromosome 20, in which the F_{ST} values are considerably smaller when compared to the rest of the chromosome (Fig. 4). This pattern is seen in all three pairwise comparisons, suggesting the suppression of differentiation is acting across all three taxa. Blast alignment to the chicken genome revealed that coding sequences of 30 genes are present within the region (Supplementary Table S2). Interestingly, the first gene to overlap the start of this region (5' end) is *ASIP*, also known as *Agouti*, which has a well-established role in melanogenesis^{36–38} but has also been shown to have a regulatory role in lipid metabolism in adipocytes³⁹.

There are multiple scenarios that could account for this pattern. Recombination rates might be lower in this region, either, due to natural variation in recombination rate or suppressed due to genomic architecture like inversions. However, inversions lead to higher not lower signals of differentiation⁴⁰. Uniform selection might be acting in the region, whereby the same alleles of all genes in this block are selected for in all three taxa. This would, in effect, produce the same long haplotype block in the three grouse. Alternatively, it could be that selection is acting on any of the genes within this region and the remainder of the block gets fixed through hitchhiking, but this should lead to reduced variation at linked sites.

The aforementioned strong demographic signal could, however, confound this inference. The significantly negative Tajima's *D* and positive Fay and Wu's *H* would indicate an overarching effect of population expansion and not selection shaping the genomic landscape of this region. However dips in Fay and Wu's *H* statistic, while never reaching negative values, do signal that selection may also be acting.

From a functional perspective, the region contains the *Agouti* gene, which is situated immediately at its 5' start. In mice studies, this gene has been shown to directly affect the coat colouration by binding to the *MC1R* cell surface protein thereby causing an increase in the production of pheomelanin (yellow/red pigment) and down-regulating the production of eumelanin (brown/black pigment)^{37,41}. In quail it has also been shown that the *yellow* phenotype is caused by a mutation >90 kb upstream of the *Agouti* gene consequently changing its promoter⁴². In the three grouse species studied here, pigmentation plays an integral role in their ecology. Both the willow grouse and rock ptarmigan moult their brown summer plumage into a pure white plumage prior to winter in order to match the predominantly snowy environment. The red grouse on the other hand, living in the generally snow-free British Isles, forgoes this winter moult and instead retains its brown plumage year-round⁴³. Because the genomic scan did not show any differentiation in this region between the red grouse and the other two species, this hints to the involvement of *Agouti* in the strongly conserved brown summer plumage, rather than the

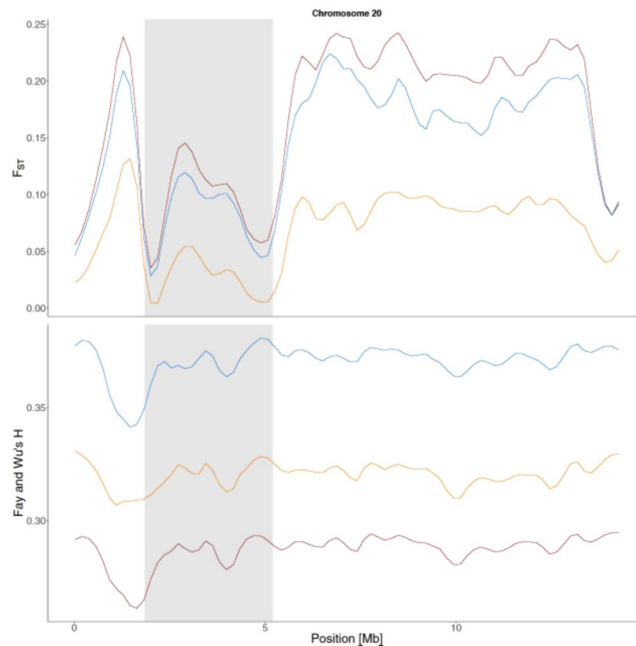


Figure 4. Chromosome 20, with the ~3.5-Mb long region of low differentiation highlighted in grey. Top panel shows the variation in F_{ST} (blue: willow grouse vs. rock ptarmigan, red: red grouse vs. rock ptarmigan, orange: willow grouse vs. red grouse), bottom panel shows the variation in Fay and Wu's H (blue: rock ptarmigan, red: red grouse, orange: willow grouse).

differentiated white-winter plumage. Further possible regulatory changes to *Agouti* or other genes involved in the melanogenesis pathway thus might be responsible for the change in winter plumage colour in the willow grouse and rock ptarmigan.

Complicating the issue further, the gene has also been found to be involved in regulation of lipid metabolism in mice and humans by antagonistically binding to another melanocortin receptor (*MC2R*)³⁹. This raises the other possibility that the regulation of *Agouti* could be involved in the adaptation to the colder, more open habitats and sedentary lifestyle seen in all three grouse. However, more knowledge of the gene's function in avian systems will have to be gathered before this issue can be fully resolved.

Strong demographic signal. By performing tests of neutrality, we have also been able to show strong signals of expanding population across the whole genome of each of the studied taxa. This was revealed by a significantly negative Tajima's D and a positive Fay and Wu's H . This result goes in line with demographic studies on these species^{44,45}, which show that the Scandinavian willow grouse (also containing the eventual red grouse lineage) and Greenland rock ptarmigan underwent a bottleneck during the Last Glacial Maximum (LGM) and only upon the deglaciation of northern Europe and Greenland did population sizes start to increase. Rock ptarmigan consistently showed higher genetic diversity based on θ than the other two taxa (Figs S5–S11).

A striking result in our data is the consistency of diversity patterns over chromosomal regions, seen in the supplementary figures (Figs S5–S11). The plots look almost identical for the three species, especially for the comparisons among red and willow grouse. Intuitively differential selection should cause more differences. A possible cause for this pattern could be similar strong post/inter-glacial demographic effects^{15,45}. A not mutually exclusive explanation may be consistent uniform purifying selection among these ecologically similar species which share a long common evolutionary history.

Conclusions

Two avenues should be followed in order to understand the dynamics of grouse speciation better. The first is to explore the extent of recombination within the genomes. By studying the changes in recombination rate along chromosomes, further insight into the nature of selection acting upon particular regions can be gained^{46,47}. If positive selection is indeed responsible for the outlier loci presented in this study, one would expect the suppression of recombination within the same and neighbouring outlier windows to follow.

The second necessary step is to further investigate the function of the genes discovered under selection in this study. Five out of the 7 genes found under divergent selection here have no experimental evidence of their function in avian systems, which makes the interpretation of the effect of selection acting on them difficult. To ultimately link gene function and adaptive phenotype, selection experiments should be carried out in order to test the consequences of different alleles⁴⁴. Possible validation of these candidate genes can be achieved by a combination of selection experiments with resequencing and gene expression studies. Other, albeit less straightforward ways to validate function would be applying functional assays such as knock-out/knock-down in the chicken at least, if not the grouse themselves.

Regardless this study provides an important genome-wide insight: grouse genomes vary among genomic regions as result of speciation and adaptation. Some parts are more divergent than expected by drift while other regions are more similar than expected by background differentiation. Regions which show signatures of positive and divergent selection vary among taxon comparisons while the region showing signs of uniform selection displays the same pattern across species contrasts.

Methods

Sampling, DNA extraction, sequencing and filtering. In total 34 individuals were sampled, comprising of 19 willow grouse (17 from Scandinavia, 1 from Magadan, Eastern Russia and 1 from Paxson, Alaska, USA), 9 red grouse from Yorkshire Dales National Park (Northern England) and 6 rock ptarmigan from south-western Greenland (Supplement Table S1). DNA extraction was performed using the Qiagen DNeasy Blood & Tissue Kit[®] following the manufacturer's instructions (Qiagen) and DNA quality of each individual was checked on an agarose gel and subsequently measured using a Qubit[®] Fluorometer. After library preparation with the Illumina TruSeq protocol, the samples were sequenced at the SNP&SEQ technology platform of Uppsala University using an Illumina HiSeq. 2500 to generate 125 bp paired end reads with a target insert size of 350 bp. Quality trimming was performed using Trimmomatic v0.36⁴⁸, following a 4 step procedure: (i) removing Illumina TruSeq adaptors, (ii) removing leading and trailing bases with quality score <5, (iii) scanning the read with a 4 base-pair sliding window and cutting when the average quality per base dropped below 15 and (iv) removing reads that were <50 bp after trimming. Overall read quality was checked using FastQC v0.11.4 (available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>).

Mapping and Analysis. All properly paired reads that passed quality control were then mapped to the closely related black grouse genome⁴⁹ using the BWA-MEM alignment algorithm⁵⁰ with default setting. Duplicate reads were marked with Picard (<http://broadinstitute.github.io/picard/>) and local realignment around indels was performed with the GATK IndelRealigner tool^{51,52} producing the final filtered bam alignment files. The resultant mean coverage across all individuals was 28x (range: 23–38x, willow grouse mean: 29x, red grouse and rock ptarmigan mean: 28x) (Supplement Table S1, Supplement Figs S1–S3). All subsequent population genetic parameter estimation was based on genotype likelihoods obtained from these bam-alignment files.

We used the program ANGSD v0.902⁵³ to calculate the unfolded allele frequency likelihoods for each taxon separately, using the black grouse reference genome to polarise the polymorphisms. The SAMtools genotyping model was implemented (-GL 1) and only sites that had mapping quality >50, base quality >30, Minor Allele Frequency (MAF) >0.05, minimum depth of $[1/3 * \text{mean taxon coverage} * \text{number of individuals}]$ and maximum depth of $[2 * \text{mean taxon coverage} * \text{number of individuals}]$ were considered. From the allele frequency likelihoods we then calculated the 2 dimensional site frequency spectrum (2D-SFS) for each taxon pair, which was then used to estimate the F_{ST} across non-overlapping 15 kb windows. The F_{ST} scores were subsequently Z-transformed and any region with a ZF_{ST} score above 6 was deemed an outlier⁶ and was subsequently aligned to the chicken genome (galGal4) using BLAST⁵⁴ to identify gene content. The same outliers that were identified in both comparisons for each taxon (e.g. willow grouse vs. red grouse and willow grouse vs. rock ptarmigan) were thus assumed to be taxon specific and of higher confidence than outliers identified in one comparison but not the other.

From the allele frequency likelihoods we also estimated the unfolded marginal SFS for each taxon separately from which diversity and neutrality-test statistics (pairwise theta $[\pi]$, Tajima's D , Fay and Wu's H^{55}) were calculated across non-overlapping 15 kb windows. Principal component analysis (PCA) was performed using ngsTools⁵⁶, whereby all individuals were genotyped together and subsequently a covariance matrix was created based on all variable sites.

To reconstruct the phylogeny of the taxa we first genotyped all individuals at a random subset of 500 k autosomal SNPs using ANGSD's SAMtools genotyping model. In order to remove redundancy due to linkage disequilibrium (LD), we filtered the dataset using SNPhylo⁵⁷ applying the LD threshold of 0.4 according to the authors recommendations for genomes with relatively high levels of LD, leaving 59 574 informative SNPs. Maximum likelihood phylogeny tree was subsequently inferred using the rapid bootstrapping algorithm with the "GTRGAMMA" substitution model and bootstrapping criterion set to "autoMRE" employed in RaxML v8.2.4⁵⁸.

All methods were carried out in accordance with relevant guidelines and regulations and all experimental protocols followed Uppsala University guidelines. None of the protocols involved live animals and birds were hunted under the guides and regulations of each respective country where a bird was shot⁵⁹.

Data Availability

Raw data is archived at the sequence read archive (SRA) under bioproject accession number: PRJNA512999.

References

1. Stapley, J. *et al.* Adaptation genomics: the next generation. *Trends Ecol. Evol.* **25**, 705–712, <https://doi.org/10.1016/j.tree.2010.09.002> (2010).
2. Rubin, C.-J. *et al.* Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–591, <https://doi.org/10.1038/nature08832> (2010).
3. Andersson, L. S. *et al.* Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice Locomotion in mammals relies on a central pattern-generating circuitry of spinal interneurons established during development that coordinates limb movement. *Nature* **488**, 642–646, <https://doi.org/10.1038/nature11399> (2012).
4. Sousa, V. & Hey, J. Understanding the origin of species with genome-scale data: modelling gene flow. *Nat. Rev. Genet.* **14**, 404–14, <https://doi.org/10.1038/nrg3446> (2013).
5. Kardos, M. *et al.* Whole genome resequencing uncovers molecular signatures of natural and sexual selection in wild bighorn sheep. *Mol. Ecol.* **24**, 5616–5632, <https://doi.org/10.1111/mec.13415> (2015).

6. Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375, <https://doi.org/10.1038/nature14181> (2015).
7. Poelstra, J. W. *et al.* The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**, 1410–1414, <https://doi.org/10.1126/science.1253226> (2014).
8. Zhang, W. *et al.* Hypoxia adaptations in the grey wolf (*Canis lupus chanco*) from Qinghai-Tibet Plateau. *PLoS Genet.* **10**, e1004466, <https://doi.org/10.1371/journal.pgen.1004466> (2014).
9. Drovetski, S. Plio-Pleistocene climatic oscillations, Holarctic biogeography and speciation in an avian subfamily. *J. Biogeogr.* **30**, 1173–1181, <https://doi.org/10.1046/j.1365-2699.2003.00920.x> (2003).
10. Gutierrez, R. J., Barrowclough, G. F. & Groth, J. G. A classification of the grouse (Aves: Tetroninae) based on mitochondrial DNA sequences. *Wildlife Biol.* **6**, 205–212 (2000).
11. Höglund, J., Wang, B., Axelsson, T. & Quintela, M. Phylogeography of willow grouse (*Lagopus lagopus*) in the Arctic: taxonomic discordance as inferred from molecular data. *Biol. J. Linn. Soc.* **110**, 77–90, <https://doi.org/10.1111/bij.12109> (2013).
12. Ward, I., Larcombe, P. & Lillie, M. The dating of Doggerland – post-glacial geochronology of the southern North Sea. *Environ. Archaeol.* **11**, 207–218, <https://doi.org/10.1179/174963106x123214> (2006).
13. Quintela, M., Berlin, S., Wang, B. & Höglund, J. Genetic diversity and differentiation among *Lagopus lagopus* populations in Scandinavia and Scotland: evolutionary significant units confirmed by SNP markers. *Mol. Ecol.* **19**, 2380–2393, <https://doi.org/10.1111/j.1365-294X.2010.04648.x> (2010).
14. McMahon, B. J., Johansson, M. P., Piertney, S. B., Buckley, K. & Höglund, J. Genetic variation among endangered Irish red grouse (*Lagopus lagopus hibernicus*) populations: implications for conservation and management. *Conserv. Genet.* **13**, 639–647, <https://doi.org/10.1007/s10592-011-0314-x> (2012).
15. Kozma, R., Melsted, P., Magnússon, K. P. & Höglund, J. Looking into the past - the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Mol. Ecol.* **25**, 570–580, <https://doi.org/10.1111/mec.13496> (2016).
16. Watson, A., Moss, R. & Rae, S. Population dynamics of Scottish rock ptarmigan cycles. *Ecology* **79**, 1174–1192, [https://doi.org/10.1890/0012-9658\(1998\)079\[1174:PDOSRP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(1998)079[1174:PDOSRP]2.0.CO;2) (1998).
17. Storch I. *Grouse Status Survey and Conservation Action Plan 2006–2010*. (Gland, 2006).
18. Storch, I. Conservation Status and Threats to Grouse Worldwide: An Overview. *Wildlife Biol.* **6**, 195–204 (2000).
19. Lucchini, V., Höglund, J., Klaus, S., Swenson, J. & Randi, E. Historical biogeography and a mitochondrial DNA phylogeny of grouse and ptarmigan. *Mol. Phylogenet. Evol.* **20**, 149–162, <https://doi.org/10.1006/mpev.2001.0943> (2001).
20. Aramaki, M. *et al.* Embryonic expression profile of chicken CHD7, the ortholog of the causative gene for CHARGE syndrome. *Birth Defects Res. Part A Clin. Mol. Teratol.* **79**, 50–57, <https://doi.org/10.1002/bdra.20330> (2007).
21. Teufel, A., Wong, E. A., Mukhopadhyay, M., Malik, N. & Westphal, H. FoxP4, a novel forkhead transcription factor. *Biochim. Biophys. Acta* **1627**, 147–152, [https://doi.org/10.1016/S0167-4781\(03\)00074-5](https://doi.org/10.1016/S0167-4781(03)00074-5) (2003).
22. Park, N. *et al.* Copine1 enhances neuronal differentiation of the hippocampal progenitor HiB5 cells. *Mol. Cells* **34**, 549–554, <https://doi.org/10.1007/s10059-012-0235-7> (2012).
23. Chung, Y. M. *et al.* Replicative Senescence Induced by Romo1-derived Reactive Oxygen Species. *J. Biol. Chem.* **283**, 33763–33771, <https://doi.org/10.1074/jbc.M805334200> (2008).
24. Takekawa, M. *et al.* A Family of Stress-Inducible GADD45-like Proteins Mediate Activation of the Stress-Responsive MTK1/MEKK4 MAPKKK. *Cell* **95**, 521–530, [https://doi.org/10.1016/S0092-8674\(00\)81619-0](https://doi.org/10.1016/S0092-8674(00)81619-0) (1998).
25. Hidai, C. *et al.* Cloning and characterization of developmental endothelial locus-1: An embryonic endothelial cell protein that binds the alpha vbeta 3 integrin receptor. *Genes Dev.* **12**, 21–33, <https://doi.org/10.1101/gad.12.1.21> (1998).
26. Fan, Y. *et al.* Del-1 gene transfer induces cerebral angiogenesis in mice. *Brain Res.* **1219**, 1–7, <https://doi.org/10.1016/j.brainres.2008.05.003> (2008).
27. Göb, E., Schmitt, J., Benavente, R. & Alsheimer, M. Mammalian Sperm Head Formation Involves Different Polarization of Two Novel LINC Complexes. *PLoS One* **5**, e12072, <https://doi.org/10.1371/journal.pone.0012072> (2010).
28. Rouso, D. L. *et al.* Foxp-Mediated Suppression of N-Cadherin Regulates Neuroepithelial Character and Progenitor Maintenance in the CNS. *Neuron* **74**, 314–330, <https://doi.org/10.1016/j.neuron.2012.02.024> (2012).
29. Miller, J. E. *et al.* Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *J. Neurophysiol.* **100**, 2015–2025, <https://doi.org/10.1152/jn.90415.2008> (2008).
30. Scharff, C. & Haesler, S. An evolutionary perspective on FoxP2: strictly for the birds? *Curr Opin Neurobiol.* **15**, 694–703, <https://doi.org/10.1016/j.conb.2005.10.004> (2005).
31. Mendoza, E. *et al.* Differential coexpression of FoxP1, FoxP2, and FoxP4 in the Zebra Finch (*Taeniopygia guttata*) song system. *J. Comp. Neurol.* **523**, 1318–1340, <https://doi.org/10.1002/cne.23731> (2015).
32. Lu, M. M., Li, S., Yang, H. & Morrisey, E. E. FoxP4: A novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Gene Expr Patterns* **2**, 223–228, [https://doi.org/10.1016/S1567-133X\(02\)00058-3](https://doi.org/10.1016/S1567-133X(02)00058-3) (2002).
33. Lalani, S. R. *et al.* Spectrum of CHD7 mutations in 110 individuals with CHARGE syndrome and genotype-phenotype correlation. *Am. J. Hum. Genet.* **78**, 303–314, <https://doi.org/10.1086/500273> (2006).
34. Kim, D., Kang, S.-S. & Jin, E.-J. Alterations in the temporal expression and function of cadherin-7 inhibit cell migration and condensation during chondrogenesis of chick limb mesenchymal cells in vitro. *J. Cell Physiol.* **221**, 161–170, <https://doi.org/10.1002/jcp.21840> (2009).
35. Vitti, J. J. *et al.* Detecting natural selection in genomic data. *Ann. Rev. Genet.* **47**, 97–120, <https://doi.org/10.1146/annurev-genet-111212-133526> (2013).
36. Duforet-Frebourg, N. *et al.* Detecting genomic signatures of natural selection with principal component analysis: application to the 1000 genomes data. *Mol. Biol. Evol.* **33**, 1082–1093, <https://doi.org/10.1093/molbev/msv334> (2015).
37. Hoekstra, H. E. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* **97**, 222–234, <https://doi.org/10.1038/sj.hdy.6800861> (2006).
38. Steiner, C. C., Weber, J. N. & Hoekstra, H. E. Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* **5**, e219, <https://doi.org/10.1371/journal.pbio.0050219> (2007).
39. Manceau, M., Domingues, V. S., Mallarino, R. & Hoekstra, H. E. The Developmental Role of Agouti in Color Pattern Evolution. *Science* **331**, 1062–1065, <https://doi.org/10.1126/science.1200684> (2011).
40. Xue, B., Moustaid-N, N., Wilkison, W. O. & Zemel, M. B. The agouti gene product inhibits lipolysis in human adipocytes via a Ca²⁺-dependent mechanism. *FASEB J.* **12**, 1391–1396 (1998).
41. Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics* **48**, 84–88, <https://doi.org/10.1038/ng.3430> (2016).
42. Graham, A., Wakamatsu, K., Hunt, G., Ito, S. & Thody, A. J. Agouti protein inhibits the production of eumelanin and pheomelanin in the presence and absence of alpha-melanocyte stimulating hormone. *Pigment Cell Res* **10**, 298–303 (1997).
43. Nadeau, N. J. *et al.* Characterization of Japanese quail yellow as a genomic deletion upstream of the mammalian ASIP (agouti) gene. *Genetics* **178**, 777–786, <https://doi.org/10.1534/genetics.107.077073> (2008).
44. Skoglund, P. & Höglund, J. Sequence polymorphism in candidate genes for differences in winter plumage between Scottish and Scandinavian Willow Grouse (*Lagopus lagopus*). *PLoS One* **5**, e10334, <https://doi.org/10.1371/journal.pone.0010334> (2010).

45. Pardo-Diaz, C., Salazar, C. & Jiggins, C. D. Towards the identification of the loci of adaptive evolution. *Methods Ecol. Evol.* **6**, 445–464, <https://doi.org/10.1111/2041-210X.12324> (2015).
46. Kozma, R., Lillie, M., Benito, B. M., Svenning, J.-C. & Höglund, J. Past and potential future population dynamics of three grouse species using ecological and whole genome coalescent modeling. *Ecol Evol.*, on-line early, <https://doi.org/10.1002/ece3.4163> (2018).
47. 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, <https://doi.org/10.1101/gr.196485.115> (2015).
48. Nosil, P. & Feder, J. L. Genomic divergence during speciation: causes and consequences. *Philos. Trans. R. Soc. B.* **367**, 332–342, <https://doi.org/10.1098/rstb.2011.0263> (2012).
49. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
50. Wang, B., Ekblom, R., Bunikis, I., Siitari, H. & Höglund, J. Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution. *BMC Genomics* **15**, 180, <https://doi.org/10.1186/1471-2164-15-180> (2014).
51. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv Prepr arXiv13033997. <http://arxiv.org/abs/1303.3997> (2013).
52. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303, <https://doi.org/10.1101/gr.107524.110> (2010).
53. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498, <https://doi.org/10.1038/ng.806> (2011).
54. Korneliusen, T., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356, <https://doi.org/10.1186/s12859-014-0356-4> (2014).
55. Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A Greedy Algorithm for Aligning DNA Sequences. *J. Comput. Biol.* **7**, 203–214, <https://doi.org/10.1089/10665270050081478> (2000).
56. Fay, J. C. & Wu, C. I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405–1413 (2000).
57. Fumagalli, M., Vieira, F. G., Linderth, T. & Nielsen, R. NgsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* **30**, 1486–1487, <https://doi.org/10.1093/bioinformatics/btu041> (2014).
58. Lee, T.-H., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 162, <https://doi.org/10.1186/1471-2164-15-162> (2014).
59. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, <https://doi.org/10.1093/bioinformatics/btu033> (2014).

Acknowledgements

The authors wish to thank Giulia Tuveri for helpful comments on the manuscript and David Newborn from the Game and Wildlife Conservation Trust (www.gwct.org.uk) for collecting the red grouse samples. Funding for the project was received from the Research Council of Sweden (VR), grant to J.H. and the Zoologiska Stiftelsen, grant to R.K. The computations were performed on the resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

Author Contributions

The paper was drafted by R.K., P.R.M. and J.H. R.K. performed the analyses and wrote the first draft of the paper. The paper was finalised after contributions of all authors.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-36880-5>.

Competing Interests: The authors declare no competing interests.

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) 2019