Supplementary information for:

Predicted deleterious mutations reveal the genomic mechanisms underlying fitness variation in a lekking bird

Rebecca S. Chen, Carl D. Soulsbury, Kosmas Hench, Kees van Oers, Joseph I. Hoffman

Correspondence: Rebecca S. Chen, rebecca.chen@uni-bielefeld.de, Department of Evolutionary Population Genetics, Faculty of Biology, Bielefeld University, Bielefeld, Germany

Supplementary results and discussion

Population and relatedness structure

To contextualise our mutation load analyses, we first characterised the genetic structure of the black grouse population and tested for the presence of close kin both within and among the lekking sites (Supplementary Fig. 1a). Genetic differentiation was overall rather weak, with no pairwise comparisons of lekking sites yielding F_{ST} values differing significantly from zero (Supplementary Table 9). However, there was a hint of population structure along a west-east axis, with KUM and NYR being the most differentiated lekking sites (Supplementary Fig. 1b). These results are consistent with previous studies revealing weak to moderate population structure among black grouse leks[1–3] and reflect the interplay between female dispersal and male site fidelity[1,4,5]. Accordingly, no significant differences were found among the lekking sites in individual mutation loads (Supplementary Table 8).

To characterise the relatedness structure of our dataset, we calculated R0, R1 and KING-robust kinship values for all pairwise combinations of individuals and visualised the results by plotting R1 against KING-robust kinship (Supplementary Fig. 1c). Pairs of individuals were assigned to specific relatedness categories by computing Z scores and comparing these with the inference criteria of Manichaikul et al.[6]. We identified a total of 466 pairs of close kin among 17,949 pairwise comparisons (2.6%) comprising 27 full siblings (5.8%), 60 parent-offspring pairs (12.9%), 176 second-degree relatives (37.8%) and 203 third-degree relatives (43.6%). The proportion of relatives sampled at the same lekking site was highest for full siblings and lowest for third-degree relatives (Supplementary Fig. 1d). The presence of close male kin on black grouse leks has been described previously[7] and arises due to strong male philopatry[1,4,5]. While the risk of inbreeding is reduced by female-biased dispersal[8], consanguineous matings may still occur as older females have a greater chance of mating with philopatric male kin[9].

Supplementary methods

Reference genome

For this study, we scaffolded a black grouse reference genome assembled by the 10K Bird Project (B10K)[10,11]. B10K constructed a single-tube long fragment read co-barcoded (stLFR) library which was subsequently 100bp paired-end sequenced on a DNBseq platform at the Beijing Genomics Institute (BGI). Raw reads with missing data (>10%), low quality (>40% bases with Phred score <= 10) or below expected insert sizes were excluded and PCR duplicates were filtered out using SOAPfilter2 v2.2[12]. GapCloser v1.12[13] was used to close gaps within scaffolds. The resulting B10K assembly had a total length of 1,002,957,384 bp and consisted of 26,930 scaffolds with an N50 of 5,658,217 bp (for details, see Supplementary Table 10).

To improve contiguity, we scaffolded the B10K assembly using HiRise together with Cantata Bio. They used 200 ml of blood to prepare an Omni-C library for genome scaffolding with HiRise. First, chromatin in the nucleus was fixed with formaldehyde[14]. DNase I was then used to extract and digest the chromatin, after which the ends were repaired and ligated to a biotinylated bridge adapter. The DNA was subsequently purified and unligated fragments were discarded. A library was then generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Using streptavidin beads, biotin-containing fragments were isolated, and the sequences were duplicated in a polymerase chain reaction (PCR). Prior to deep sequencing, the B10K assembly was used to quality check the OmniC library. Finally, the sequencing was performed using an Illumina HiSeqX platform with a target coverage of 30X and the resulting reads were quality filtered (MQ > 50) and used to scaffold both pseudo-haplotypes with Cantata Bio's HiRise software[14]. The contiguity of the resulting reference genome was considerably improved, with the scaffold N50 increasing over ten-fold to 69,550,540 (Supplementary Table 10). Lastly, we identified the scaffold corresponding to the Z chromosome by aligning the chicken Z chromosome (NCBI RefSeq assembly GCF_016699485.2) to the black grouse reference genome using BLAST v2.12.0[15]. We identified a long scaffold that showed 83% identity to the chicken Z chromosome, which we assigned as the black grouse Z chromosome.

Genome annotation

To annotate the scaffolded reference genome, we generated transcriptomic data from an 11–14 day old black grouse embryo obtained from a captive breeding facility in the Netherlands. Total RNA was extracted from a mixture of randomly selected embryo sections using the Qiagen RNeasy Plus Kit. The extract was treated with DNase and the RNA was cleaned using AMPure beads and the QIAGEN FastSelect HMR RNA depletion kit. The library was prepared by Genewiz Multiomics & Synthesis Solutions using the NEBNext Ultra II RNA Library Prep Kit and sequenced on an Illumina NovaSeq6000 platform with 150 bp paired-end reads.

The genome annotation was performed by Cantata Bio using the resulting RNA sequencing data together with publicly available RNA sequencing data (NCBI SRA Bioproject SRP006680). Repeat families in the genome were identified *de novo* and classified using RepeatModeler v2.0.1[16]. The output from RepeatModeler was then used to identify and mask repeated segments in the genome annotation file using RepeatMasker v4.1.0[17]. Two initial *ab initio* gene models were trained with the coding sequences of the chicken (*Gallus gallus*), Japanese quail (*Coturnix japonica*), rock ptarmigan (*Lagopus muta*) and wild turkey (*Meleagris gallopavo*) using AUGUSTUS v2.5.5[18] and SNAP v2006-07-28[19] respectively. The AUGUSTUS predictions were optimised in six rounds. RNA-Seq reads were aligned to the reference genome using STAR v2.7[20] and intron hints were generated with the bam2hints tools within AUGUSTUS. Subsequently, we used MAKER[21], SNAP and AUGUSTUS (with intron-exon boundary hints provided from the RNA-Seq data) to predict genes in the repeat-masked reference genome. To help guide the gene prediction process, Swiss-Prot peptide sequences from the UniProt database were downloaded and used in combination with the protein sequences from the avian species described above to generate peptide evidence in the Maker pipeline. The final gene sets comprised only genes that were predicted by both SNAP and AUGUSTUS. The quality of the predictions was assessed using AED scores generated for predicted genes by MAKER. Genes were further characterised for their putative function by performing a BLAST[15] search of the peptide sequences against the UniProt database. Transfer RNAs were predicted using tRNAscan-SE v2.05[22].

Population structure and relatedness

We used principal component analysis (PCA) to test for population genetic structure. For this analysis, PLINK v1.90[23] was used to produce a stringently filtered SNP dataset from which the following were excluded (i) strongly linked SNPs, where linkage disequilibrium (LD) was computed in window sizes of 50 base pairs with shifts of 5 SNPs with a variance inflation factor threshold of 2 (--indep 50 5 2); (ii) SNPs deviating significantly from Hardy-Weinberg equilibrium (HWE) with an alpha level of 0.001 (--hwe 0.001); and (iii) SNPs with a minor allele frequency (MAF) below 0.01 (--maf 0.01). The PCA was implemented using PLINK (--pca) and the results were visualized with the R tidyverse package set v1.3.1[24], including dplyr v1.1.4 and ggplot2 v3.5.1.

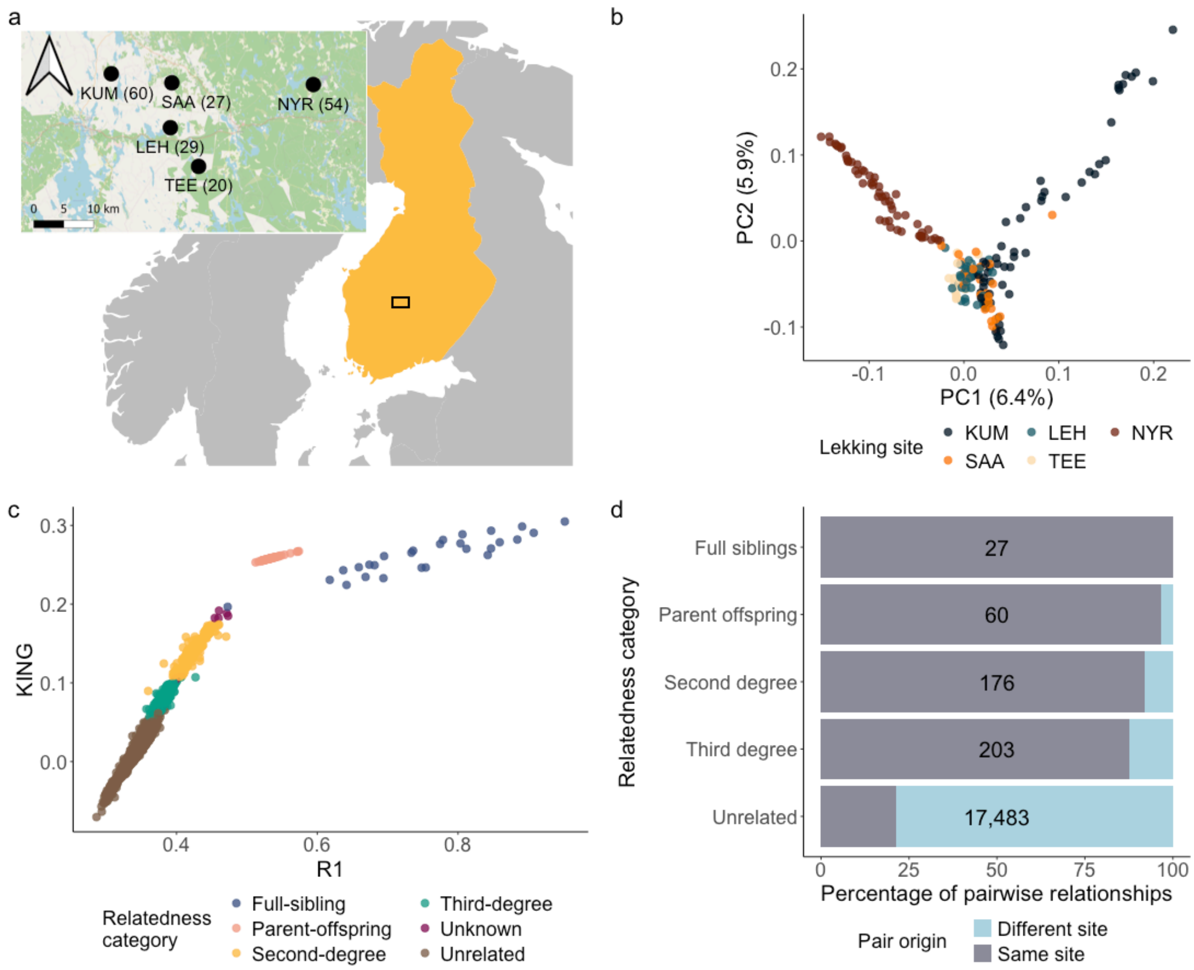
Next, we quantified genetic differentiation by calculating F_{ST} values between each pair of lekking sites based on the stringently filtered dataset described above. F_{ST} values were computed per SNP using VCFtools v0.1.17[25] and then averaged across loci. To test whether the mean F_{ST} values were significantly different from zero, we computed 1,000 bootstrap replicates by resampling the pairwise F_{ST} values across loci with replacement using the boot package v.1.3.28[26]. Finally, we calculated the absolute differences between the resampled mean F_{ST} values and the mean F_{ST} of the original data, summed these values and divided them by the total number of bootstrap replicates to obtain the *p*-value, which represents the probability of observing a mean F_{ST} different from zero.

We used the same stringently filtered dataset to infer patterns of pairwise genomic relatedness among individuals following the workflow of Humble *et al.*[27]. NgsRelate v2[28] was used to compute three relatedness indices: KING-robust kinship, R0 and R1[29]. The KING-robust kinship estimate can be used to distinguish between different levels of relatedness when allele frequencies are unknown and is robust to population structure[6]. The R0 and R1 statistics specify whether zero or one copies of an allele are shared, respectively. Different categories of relatedness are associated with non-overlapping ranges of R1, R0 and KING-robust kinship values.

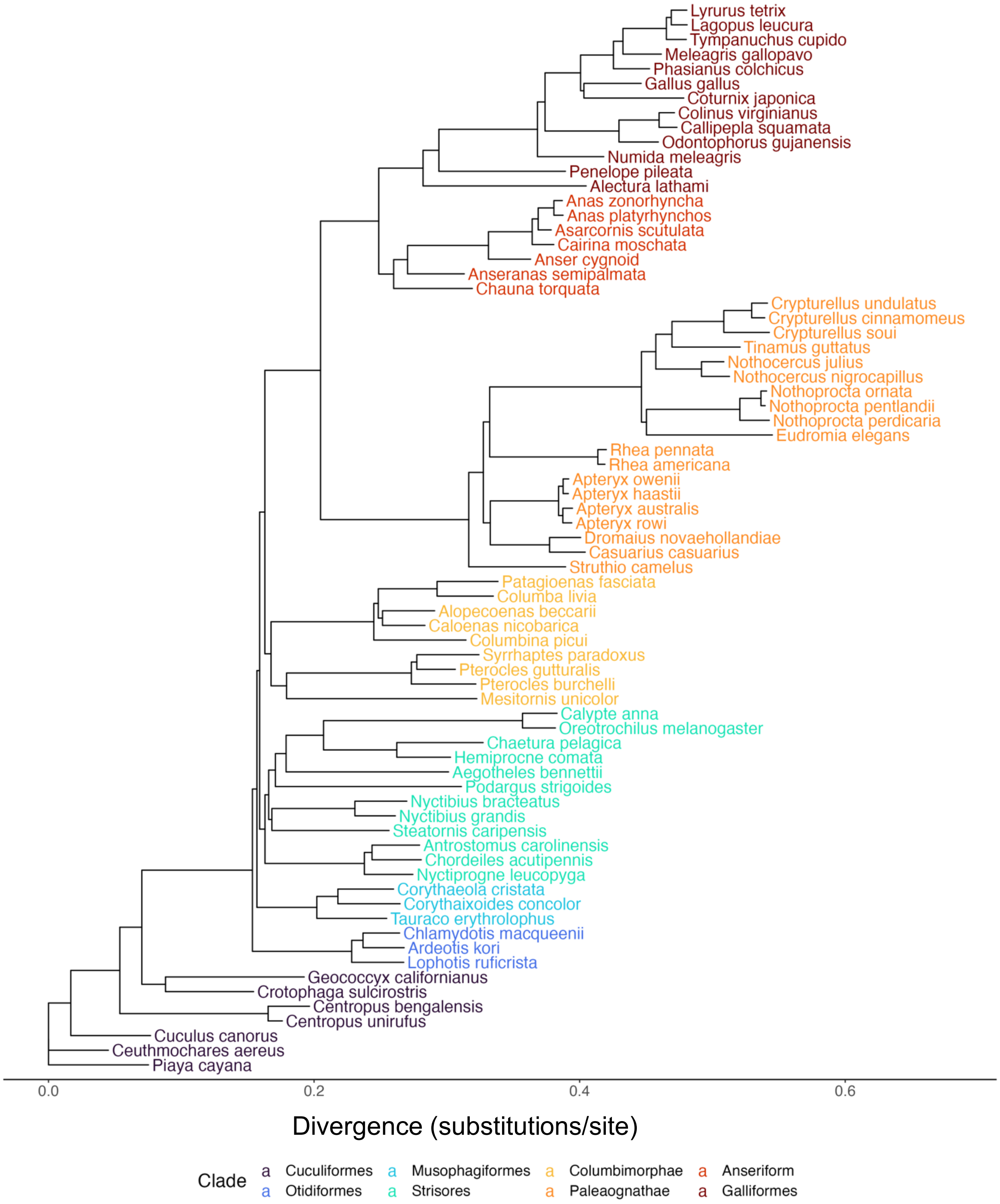
To allocate pairs of individuals to specific relationship categories, we calculated the relatedness coefficients $\hat{p}$, Z0, Z1 and Z2 using the -genome function in PLINK[23]. $\hat{p}$ is the overall proportion of the genome that is identical by descent between any pair of individuals. Z0, Z1 and Z2 are coefficients that estimate the proportion of the genome for which zero, one or two alleles of a pair of individuals are identical by descent, respectively. We used the method described in Manichaikul *et al.*[6] to assign relationship categories to each pair of individuals based on theoretical thresholds of the various relatedness coefficients. Pairs of individuals were assigned to one of five relatedness categories: parent-offspring, full siblings, second-degree relatives (e.g. half-siblings and grandparents-grandchildren), third-degree relatedness (e.g. cousins) or unrelated individuals. To allow minor deviations from the theoretical expectations for certain relationships, we classified pairs of individuals as falling within a given relatedness category if they were within 0.01 of the respective inference threshold, following Waples *et al.*[29]. Any pair that did not fall within the theoretical ranges of any category including this additional margin were classified as "unknown".

Supplementary figures

Supplementary Fig 1. Genetic and relatedness structure of the black grouse study population in central Finland. (a) Geographical locations of the lekking sites, with circle sizes being proportional to the number of sampled lekking males (total $n = 190$) as shown in the legend. Abbreviations: KUM = Kummunsuo, NYR = Nyrölä, SAA = Saarisuo, LEH = Lehtosuo, TEE = Teerisuo; (b) Results of the principal component analysis, with the lekking sites colour coded as shown in the legend; (c) R1 coefficients plotted against KING-robust kinship coefficients for all individual pairwise comparisons. The colours of the points indicate relationship categories inferred by comparing PLINK Z scores with the inference criteria derived from Manichaikul *et al.*[6] as shown in the legend; (d) A breakdown of the relatedness structure of the population divided into comparisons within and among lekking sites.



Supplementary Fig 2. Phylogenetic tree used for the calculation of GERP scores. Shown is an unrooted phylogenetic tree consisting of 74 avian species that was used for calculating GERP scores. Different avian clades are colour-coded as shown in the legend.



Supplementary tables

Supplementary Table 1. Output of Bayesian GLMMs of lifetime mating success (LMS). Shown are point estimates, credible intervals (95% and 80%) and R^2 values of the standardised beta coefficients of the genomic predictors on LMS. Predictors included the total GERP and SnpEff loads, the total GERP load based on mutations with GERP scores between 3 and 4, and the homozygous and heterozygous GERP and SnpEff loads.

Predictor	**Median**	**95% CI (lower, upper)**	**80% CI (lower, upper)**	**Conditional R^2 [95%CI]**	**Marginal R^2 [95%CI]**
Total GERP load	-0.21	-0.27, -0.14	-0.25, -0.17	0.07 [0.04, 0.11]	0.02 [$3.64e^{-5}$, 0.06]
Total SnpEff load	-0.11	-0.18, -0.04	-0.16, -0.06	0.05 [0.03, 0.08])	0.01 [$3.69e^{-5}$, 0.03]
Total GERP load based on GERP scores 3-4	-0.12	-0.20, -0.05	-0.17, -0.08	0.06 [0.03, 0.09]	0.01 [$6.81e^{-5}$, 0.03]
Homozygous GERP load	-0.57	-0.76, -0.39	-0.70, -0.45	0.07 [0.04, 0.11]	0.02 [$1.68e^{-4}$, 0.06]
Heterozygous GERP load	-0.60	-0.78, -0.41	-0.72, -0.48		
Homozygous SnpEff load	-0.04	-0.17, -0.01	-0.15, -0.04	0.05 [-.03, 0.08]	0.01 [$2.87e^{-4}$, 0.03]
Heterozygous SnpEff load	-0.09	-0.24, -0.06	-0.21, 0.09		

Supplementary Table 2. Output of Bayesian GLMMs of lifetime mating success (LMS) broken down by gene region for GERP. Shown are point estimates, credible intervals (95% and 80%) and R^2 values of the standardised beta coefficients of the total GERP load calculated per genomic location.

Region	**Median**	**95% CI (lower, upper)**	**80% CI (lower, upper)**	**Conditional R^2 [95%CI]**	**Marginal R^2 [95%CI]**
Promoter	-0.18	-0.26, -0.09	-0.23, -0.13	0.07 [0.04, 0.11]	0.01 [4.18e^{-5}, 0.04]
TSS	-0.27	-0.35, -0.20	-0.32, -0.22	0.07 [0.04, 0.11]	0.03 [1.00e^{-3}, 0.07]
Intron	-0.29	-0.37, -0.21	-0.35, -0.23	0.06 [0.04, 0.09]	0.03 [2.0e^{-3}, 0.06]
Exon	0.23	0.15, 0.31	0.18, 0.29	0.07 [0.04, 0.11]	0.02 [3.00e^{-4}, 0.05]

Supplementary Table 3. Output of Bayesian GLMMs of lifetime mating success (LMS) broken down by gene region for SnpEff. Shown are point estimates, credible intervals (95% and 80%) and R^2 values of the standardised beta coefficients of total SnpEff load calculated per genomic location.

Region	**Median**	**95% CI (lower, upper)**	**80% CI (lower, upper)**	**Conditional R^2 [95%CI]**	**Marginal R^2 [95%CI]**
Promoter	-0.26	-0.34, -0.18	-0.31, -0.20	0.06 [0.04, 0.10]	0.02 [$1.00e^{-3}$, 0.06]
TSS	-0.04	-0.12, 0.04	-0.09, 0.01	0.05 [0.03, 0.09]	0.05 [$1.56e^{-5}$, 0.02]
Intron	-0.08	-0.15, 0.01	-0.13, -0.02	0.05 [0.03, 0.08]	0.01 [$8{,}93e^{-5}$, 0.03]
Exon	-0.06	-0.20, -0.03	-0.17, -0.06	0.05 [0.03, 0.08]	0.01 [6.60^{-6}, 0.03]

Supplementary Table 4. Output of Bayesian GLMMs of annual mating success (AMS). Shown are point estimates, credible intervals (95% and 80%) and R^2 of the standardised beta coefficients of the six sexual traits, total load, and age category.

Approach	**Predictor**	**Median**	**95% CI (lower, upper)**	**80% CI (lower, upper)**	**Conditional R^2 [95%CI]**	**Marginal R^2 [95%CI]**
GERP	Lyre size	0.31	-0.07, 0.69	0.07, 0.57	0.55 [0.40, 0.69]	0.11 [0.02, 0.26]
	Eye comb size	0.13	-0.21, 0.48	-0.10, 0.35		
	Blue chroma	0.15	-0.05, 0.36	0.01, 0.29		
	Attendance	1.32	0.71, 2.01	0.91, 1.74		
	Fighting rate	-0.06	-0.33, 0.21	-0.24, 0.11		
	Centrality	-0.59	-0.93, -0.23	-0.81, -0.36		
	Total load	-0.12	-0.38, 0.11	-0.28, -0.03		
	Age category - yearling	-0.17	-1.09, 0.74	-0.78, 0.42		
SnpEff	Lyre size	0.33	-0.08, 0.72	0.07, 0.58	0.55 [0.41, 0.68]	0.10 [0.02, 0.25]
	Eye comb size	0.14	-0.21, 0.49	-0.09, 0.37		
	Blue chroma	0.15	-0.06, 0.36	0.02, 0.29		
	Attendance	1.32	0.72, 1.94	0.92, 1.73		
	Fighting rate	-0.06	-0.32, 0.20	-0.23, 0.11		
	Centrality	-0.57	-0.94, -0.22	-0.82, -0.34		
	Total load	-0.11	-0.38, 0.16	-0.27, 0.06		
	Age category - yearling	-0.15	-1.04, 0.79	-0.73, 0.45		

Supplementary Table 5. Output of Bayesian GLMMs of the six sexual traits. Shown are point estimates, credible intervals (95% and 80%) and R^2 values of the standardised beta coefficients of total GERP load and age category. One model was constructed for each of the sexual traits (see Methods for details).

Response	Predictor	Median	95% CI (lower, upper)	80% CI (lower, upper)	Conditional R^2 [95%CI]	Marginal R^2 [95%CI]
Lyre size	Total GERP load	-0.03	-0.10, 0.04	-0.07, 0.02	0.88 [0.86, 0.89]	0.74 [0.72, 0.75]
	Age - yearling	-1.77	-1.86, -1.68	-1.82, -1.72		
Eye comb size	Total GERP load	$-3.3e^{-3}$	-0.09, 0.09	-0.06, 0.05	0.57 [0.49, 0.63]	0.40 [0.34, 0.45]
	Age - yearling	-1.31	-1.47, -1.16	-1.41, -1.21		
Blue chroma	Total GERP load	-0.03	-0.13, 0.08	-0.10, 0.04	0.42 [0.32, 0.50]	0.10 [0.06, 0.15]
	Age - yearling	-0.67	-0.85, -0.49	-0.79, -0.55		
Attendance	Total GERP load	-0.10	-0.19, -0.01	-0.16, -0.05	0.42 [0.33, 0.50]	0.33 [0.27, 0.39]
	Age - yearling	-1.15	-1.31, -0.99	-1.26, -1.05		
Fighting rate	Total GERP load	0.01	-0.09, 0.12	-0.06, 0.08	0.18 [0.09, 0.30]	0.07 [0.02, 0.12]
	Age - yearling	-0.60	-0.83, -0.37	-0.75, -0.44		
Centrality	Total GERP load	0.04	-0.09, 0.16	-0.04, 0.12	0.44 [0.32, 0.55]	0.08 [0.03, 0.13]
	Age - yearling	0.62	0.40, 0.84	0.48, 0.76		

Supplementary Table 6. Output of Bayesian GLMMs of the six sexual traits. Shown are point estimates, credible intervals (95% and 80%) and R^2 values of the standardised beta coefficients of total SnpEff load and age category. One model was constructed for each of the sexual traits (see Methods for details).

Response	**Predictor**	**Median**	**95% CI (lower, upper)**	**80% CI (lower, upper)**	**Conditional R^2 [95%CI]**	**Marginal R^2 [95%CI]**
Lyre size	Total SnpEff load	0.04	-0.04, 0.10	-0.01, 0.08	0.88 [0.86, 0.89]	0.73 [0.72, 0.75]
	Age - yearling	-1.77	-1.84, -1.69	-1.82, -1.72		
Eye comb size	Total SnpEff load	-0.01	-0.10, 0.07	-0.07, 0.04	0.57 [0.50, 0.63]	0.40 [0.35, 0.45]
	Age - yearling	-1.31	-1.46, -1.15	-1.41, -1.21		
Blue chroma	Total SnpEff load	-0.07	-0.17, 0.04	-0.13, $9.80e^{-4}$	0.42 [0.33, 0.50]	0.11 [0.06, 0.15]
	Age - yearling	-0.67	-0.84, -0.49	-0.79, -0.55		
Attendance	Total SnpEff load	-0.03	-0.11, 0.06	-0.08, 0.03	0.43 [0.34, 0.51]	0.32 [0.26, 0.38]
	Age - yearling	-0.17	-1.32, -0.99	-1.27, -1.06		
Fighting rate	Total SnpEff load	-0.02	-0.12, 0.08	-0.09, 0.04	0.18 [0.09, 0.29]	0.07 [0.02, 0.12]
	Age - yearling	-0.58	-0.82, -0.36	-0.74, -0.43		
Centrality	Total SnpEff load	-0.01	-0.14, 0.11	-0.09, 0.07	0.45 [0.32, 0.55]	0.08 [0.03, 0.12]
	Age - yearling	0.62	0.40, 0.84	0.48, 0.76		

Supplementary Table 7. Output of Bayesian GLMMs for the direct and indirect effects of total load on AMS. Shown are point estimates and credible intervals (95% and 80%) of the standardised beta coefficients of the direct and indirect effects of total load on AMS through the six sexual traits calculated using the Point Method (see Methods for details). The direct effect of total GERP load on AMS was estimated by correcting for all the sexual traits.

Approach	**Effect**	**Mediator**	**Median**	**95% CI (lower, upper)**	**80% CI (lower, upper)**
GERP	Direct	-	-0.13	-0.36, 0.11	-0.29, 0.03
	Indirect	Lyre size	-0.01	-0.04, 0.01	-0.03, 0.01
		Eye comb size	0.00	-0.02, 0.02	-0.01, 0.01
		Blue chroma	0.00	-0.03, 0.01	-0.02, 0.01
		Attendance	-0.13	-0.28, -0.01	-0.22, -0.05
		Fighting rate	0.00	-0.02, 0.02	-0.01, 0.01
		Centrality	-0.02	-0.11, 0.05	-0.07, 0.02
SnpEff	Direct	-	-0.11	-0.38, 0.16	-0.27, 0.06
	Indirect	Lyre size	0.01	-0.01, 0.05	0.00, 0.03
		Eye comb size	0.00	-0.03, 0.02	-0.01, 0.01
		Blue chroma	-0.01	-0.04, 0.01	-0.03, 0.00
		Attendance	-0.03	-0.16, 0.09	-0.11, 0.04
		Fighting rate	0.00	-0.01, 0.02	-0.01, 0.01
		Centrality	0.01	-0.06, 0.09	-0.04, 0.06

Supplementary Table 8. Output of frequentist linear models of mutation load. Shown are beta estimates, standard errors, *t*-values and *p*-values. Lekking site was used as a predictor, where Kummunsuo was the reference lek.

Response	**Lekking site**	**Beta estimate**	**Standard error**	***t*-value**	***p*-value**
Total GERP load	Lehtusuo	-8.57 e-5	1.17 e-4	-0.73	0.47
	Nyrölä	-5.35 e-5	9.71 e-5	-0.55	0.58
	Saarisuo	-5.70 e-5	1.20 e-4	-0.48	0.64
	Teerisuo	4.85 e-5	1.34 e-4	0.36	0.72
Total SnpEff load	Lehtusuo	-1.09 e-4	6.66 e-4	-0.16	0.87
	Nyrölä	-1.79 e-4	5.52 e-4	-0.32	0.74
	Saarisuo	-9.17 e-4	6.82 e-4	-1.34	0.18
	Teerisuo	-5.43 e-4	7.60 e-4	-0.72	0.48

Supplementary Table 9. Population differentiation among lekking sites. Shown above the diagonal are mean F_{ST} values and corresponding *p*-values are shown below the diagonal for each pairwise comparison.

Lekking site	**Kummunsoa**	**Lehtusuo**	**Nyrölä**	**Saarisuo**	**Teerisuo**
Kummunsoa	-	0.012	0.017	0.014	0.016
Lehtusuo	0.512	-	0.014	0.009	0.009
Nyrölä	0.527	0.457	-	0.017	0.017
Saarisuo	0.500	0.524	0.501	-	0.013
Teerisuo	0.488	0.506	0.536	0.527	-

Supplementary Table 10. Summary statistics of the B10k and the scaffolded black grouse reference genome.

Statistic	B10k genome	Scaffolded genome
Total length (bp)	1,002,957,384	1,003,452,484
N50	5,658,217	69,550,540
L50	49	5
N90	500,392	12,704,504
L90	267	18
Number of scaffolds	26,930	21,979
Number of scaffolds > 1kbp	10,613	5,662
Largest scaffold (bp)	32,946,576	189,864,486
Number of gaps	10,955	15,906

References

1. Höglund, J., Alatalo, R. V., Lundberg, A., RintamÎki, P. T. & Lindell, J. Microsatellite markers reveal the potential for kin selection on black grouse leks. *Proc. R. Soc. Lond. B Biol. Sci.* **266**, 813–816 (1999).
2. Chen, R. S. *et al.* Effects of hunting on genetic diversity, inbreeding and dispersal in Finnish black grouse (*Lyrurus tetrix*). *Evol. Appl.* **16**, 625–637 (2023).
3. Corrales, C. & Höglund, J. Maintenance of gene flow by female-biased dispersal of Black Grouse Tetrao tetrix in northern Sweden. *J. Ornithol.* **153**, 1127–1139 (2012).
4. Caizergues, A. & Ellison, L. N. Natal dispersal and its consequences in Black Grouse Tetrao tetrix: Natal dispersal and its consequences in Black Grouse. *Ibis* **144**, 478–487 (2002).
5. Warren, P. K. & Baines, D. Dispersal, survival and causes of mortality in black grouse Tetrao tetrix in northern England. *Wildl. Biol.* **8**, 91–97 (2002).
6. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
7. Lebigre, C., Alatalo, R. V., Forss, H. E. & Siitari, H. Low levels of relatedness on black grouse leks despite male philopatry. *Mol. Ecol.* **17**, 4512–4521 (2008).
8. Lebigre, C., Alatalo, R. V. & Siitari, H. Female-biased dispersal alone can reduce the occurrence of inbreeding in black grouse (*Tetrao tetrix*). *Mol. Ecol.* **19**, 1929–1939 (2010).
9. Soulsbury, C. D., Alatalo, R. V., Lebigre, C., Rokka, K. & Siitari, H. Age-dependent inbreeding risk and offspring fitness costs in female black grouse. *Biol. Lett.* **7**, 853–855 (2011).
10. Zhang, G. Bird sequencing project takes off. *Nature* **522**, 34–34 (2015).

11. Feng, S., Stiller, J. & Deng. Dense sampling of bird diversity increases power of comparative genomics. *Nature* **587**, 252–257 (2020).
12. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 18 (2012).
13. Xu, M. *et al.* TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* **9**, giaa094 (2020).
14. Putnam, N. H. *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).
15. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
16. Smit, A., Hubley, R. & Green, P. RepeatModeler Open. (2008).
17. Smit, A., Hubley, R. & Green, P. RepeatMasker Open. (2013).
18. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**, 637–644 (2008).
19. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
20. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
21. Cantarel, B. L. *et al.* MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
22. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).
23. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

24. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
25. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
26. Canty, A. & Ripley, B. D. boot: Bootstrap R (S-Plus) functions. (2021).
27. Humble, E., Paijmans, A. J., Forcada, J. & Hoffman, J. I. An 85K SNP Array Uncovers Inbreeding and Cryptic Relatedness in an Antarctic Fur Seal Breeding Colony. *G3 GenesGenomesGenetics* **10**, 2787–2799 (2020).
28. Korneliussen, T. S. & Moltke, I. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. *Bioinformatics* btv509 (2015) doi:10.1093/bioinformatics/btv509.
29. Waples, R. K., Albrechtsen, A. & Moltke, I. Allele frequency-free inference of close familial relationships from genotypes or low-depth sequencing data. *Mol. Ecol.* **28**, 35–48 (2019).