***Progress Report*:**

Analysing the genetic diversity of the capercaillie population in the Cairngorms National Park

DATE: 31/05/2021

Prepared for:

Carolyn Robertson
Cairngorms Capercaillie Project
Cairngorms National Park Authority
14 The Square
Grantown on Spey
PH26 3HG
E: carolynrobertson@cairngorms.co.uk

Prepared by:

Dr Alex Ball
RZSS WildGenes Laboratory
Royal Zoological Society of Scotland
134 Corstorphine Road
Edinburgh
EH12 6TS, UK
E: aball@rzss.org.uk

CONTENTS

EXECUTIVE SUMMARY

This report outlines the progress made regarding the genetic work performed by the Royal Zoological Society of Scotland's WildGenes laboratory under contract number CCP_003, awarded by the Cairngorms National Park Authority. The RZSS WildGenes laboratory has been contracted for a period of 16 months (Dec 2020-March 2021) to analyse DNA extracted from feather samples collected by the Cairngorms Capercaillie Project. This constitutes a 6-month progress report summarising the achievements up to 31st May 2021:

1. DNA extraction was attempted from 5 tissue samples provided by CCP.
2. 34 blood/tissue samples were obtained from 7 European regions:
    i. France (n=12)
    ii. Norway (n=2)
    iii. Sweden (n=5)
    iv. Poland (n=4)
    v. Germany (n=5)
    vi. Austria (n=4)
    vii. Scotland (n=2)
3. Mitochondrial sequences have been produced for 350 feather samples.
4. A genomic library was produced containing DNA from 85 samples.
5. All 85 samples have produced genomic sequences suitable for use in SNP filtering.
6. Four museums have been contacted to facilitate access to historic Scottish capercaillie specimens.
7. Contractual actions 1-3 have been completed and progress started towards actions 8 and 9.

GLOSSARY

CCP – Cairngorms Capercaillie Project

ddRAD – double-digest Restriction-site Associated DNA, a technique used to target a reduced region of the genome that is useful when working with species with limited previously produced genetic resources.

d-loop – a hyper-variable domain of the mitochondrial genome that is often used to investigate genetic diversity across populations.

Genomic library – a mixture of genomic DNA that has originated from multiple individuals that is often indexed with unique barcodes for bioinformatic separation after sequencing.

Haplotype – a collection of mutations that are inherited together within the same genome and are used to identify individuals with more recent shared ancestry.

Hybrid capture – A target enrichment technique whereby genetic probes are designed to target known variable genomic regions to enrich degraded samples by binding to (aka capturing) these regions.

Mitochondria – organelles within cells that contain a small section of an organism's genome. They are present in multiple copies within each cell.

mtDNA – refers to mitochondrial DNA; the roughly 16,000 base pair genome found within mitochondria.

nDNA – refers to nuclear DNA; the majority of an organism's genome found within each cell nucleus.

SNP – a Single Nucleotide Polymorphism is a single variation between otherwise identical genetic sequences.

INTRODUCTION

The Cairngorms Capercaillie Project, funded by the National Lottery Heritage Fund, is focused on saving the capercaillie from extinction in the UK. In the 1970s there were thought to be around 20,000 capercaillie in Scotland, but recent estimates suggest less than 1000 birds remain. The Cairngorms National Park is home to approximately 90% of these remaining birds. Action in the National Park is therefore critical to prevent extinction in the UK and build a long-term future for the species. Saving a species on the brink of extinction and planning for the future at the same time will always be complex; there is no single solution. The project's work therefore involves five essential actions delivered across the Cairngorms National Park.

1. Helping communities to create and deliver their own community-led actions for capercaillie.
2. Raising awareness of the plight of capercaillie and how people can help.
3. Researching the genetic diversity of capercaillie in the National Park to help inform action.
4. Improving and creating more habitat for capercaillie.
5. Strengthening current capercaillie monitoring to enable more informed decisions.

This report is focused on the aspect related to understanding the genetic diversity of the capercaillie in the National Park. In 2019 over 1,200 feathers were collected by the Cairngorms National Park. The Royal Zoological Society of Scotland's (RZSS) WildGenes laboratory processed these samples to identify those with DNA suitable for further genetic analysis.

The RZSS WildGenes laboratory is now proceeding to evaluate the genetic diversity of the capercaillie population in the Cairngorms National Park and learn more from the collected material. The specific services identified in the tender were as follows:

1. Prepare a genomic library using a subset of 45 high-quality samples from those collected.
2. Extract DNA from 4 additional tissue samples collected for inclusion in the genomic library. The genomic library would then include 49 high-quality samples.
3. Sequence and prepare a library report.
4. Deposit the sequence data in a public repository in order to create a library of reference data and provide a public resource for others to use.
5. Analyse 619 low-quality samples (DNA extracted from feathers), potentially using a targeted enrichment approach.
6. Provide genetic individual-level identification for each feather sample.
7. Report on the DNA analysis, including recommendations to inform the project's work.
8. If possible, source reference material from capercaillie populations outside the UK and draw comparisons in the final report to aid a broader understanding of the Cairngorms capercaillie population.
9. Reference, as far as possible, previous DNA research involving the Scottish capercaillie population to aid a historical understanding of the Cairngorms capercaillie population.
10. Work with the project to help develop and deliver a programme of communication activities that will enable a wider range of people to access and learn about the research.
11. Present the research results and recommendations at a workshop hosted by the project for all organisations and individuals interested in developing an action plan in response.

We have achieved actions 1-2, and this report delivers on action 3. We have also begun working towards actions 8 and 9. Below I summarise the progress against each of the aforementioned actions.

PROGRESS

Action 1 - Prepare a genomic library

The RZSS WildGenes laboratory have prepared a genomic library via a double-digest Restriction site-associated DNA (ddRAD) technique consisting of DNA from 85 capercaillie samples. The DNA was obtained from the following samples i) 53 samples extracted from the feathers collected by the CCP, ii) Four samples extracted from tissue samples provided by the CCP (see Action 2) and iii) 28 samples extracted from blood/tissue provided from 7 European locations (see Action 8). The samples were selected based on DNA quality as determined by gel electrophoresis and fluorometry quantification (Qubit® 2.0) and are summarised in Table 1.

Table 1. *Summary of the samples included in the genomic library.*

Sample ID	Qubit (ng/ul)	Sample type	Origin
CAP0003	7	Tissue	Scotland
CAP0189	7	Tissue	Scotland
CAP0287	3.64	Feather	Scotland - CCP
CAP0305	2.53	Feather	Scotland - CCP
CAP0309	3.55	Feather	Scotland - CCP
CAP0310	1.72	Feather	Scotland - CCP
CAP0319	3.51	Feather	Scotland - CCP
CAP0320	1.27	Feather	Scotland - CCP
CAP0334	3.89	Feather	Scotland - CCP
CAP0337	3.12	Feather	Scotland - CCP
CAP0338	2.83	Feather	Scotland - CCP
CAP0339	7.27	Feather	Scotland - CCP
CAP0376	4.73	Feather	Scotland - CCP
CAP0443	0.88	Feather	Scotland - CCP
CAP0470	1.77	Feather	Scotland - CCP
CAP0481	6.68	Feather	Scotland - CCP
CAP0488	2.30	Feather	Scotland - CCP
CAP0514	1.34	Feather	Scotland - CCP
CAP0522	2.49	Feather	Scotland - CCP
CAP0570	4.20	Feather	Scotland - CCP
CAP0576	4.24	Feather	Scotland - CCP
CAP0580	1.01	Feather	Scotland - CCP
CAP0589	6.96	Feather	Scotland - CCP
CAP0639	7.53	Feather	Scotland - CCP
CAP0669	4.03	Feather	Scotland - CCP
CAP0715.3	7	Feather	Scotland - CCP
CAP0754	0.56	Feather	Scotland - CCP
CAP0776	2.15	Feather	Scotland - CCP
CAP0777	7.44	Feather	Scotland - CCP
CAP0838	1.53	Feather	Scotland - CCP
CAP0845	7	Feather	Scotland - CCP
CAP0848	1.63	Feather	Scotland - CCP
CAP0849	3.97	Feather	Scotland - CCP
CAP0873	1.54	Feather	Scotland - CCP
CAP0899	1.41	Feather	Scotland - CCP
CAP0916	3.38	Feather	Scotland - CCP
CAP0942	0.54	Feather	Scotland - CCP
CAP1034	0.60	Feather	Scotland - CCP

CAP1036	1.91	Feather	Scotland - CCP
CAP1063	2.10	Feather	Scotland - CCP
CAP1084	0.43	Feather	Scotland - CCP
CAP1101	1.33	Feather	Scotland - CCP
CAP1117	2.68	Feather	Scotland - CCP
CAP1119	0.57	Feather	Scotland - CCP
CAP1120	0.60	Feather	Scotland - CCP
CAP1126	0.25	Feather	Scotland - CCP
CAP1132	1.07	Feather	Scotland - CCP
CAP1133	7	Feather	Scotland - CCP
CAP1134	7	Feather	Scotland - CCP
CAP1135	3.46	Feather	Scotland - CCP
CAP1139	0.76	Feather	Scotland - CCP
CAP1171	2.22	Feather	Scotland - CCP
CAP1182	8.48	Feather	Scotland - CCP
CAP1184	3.41	Feather	Scotland - CCP
CAP1186	1.57	Feather	Scotland - CCP
CAP1195	5.20	Tissue	Scotland - CCP
CAP1196	7	Tissue	Scotland - CCP
CAP1197	2.39	Tissue	Scotland - CCP
CAP1198	7	Tissue	Scotland - CCP
CAP1201	7	Tissue	France
CAP1202	7	Tissue	France
CAP1203	7	Tissue	France
CAP1204	7	Tissue	France
CAP1205	7	Tissue	France
CAP1207	7	Tissue	France
CAP1209	7	Tissue	France
CAP1210	7	Tissue	France
CAP1211	7	Tissue	France
CAP1213	7	Blood	Germany
CAP1214	7	Blood	Germany
CAP1216	7	Tissue	Austria
CAP1217	7	Tissue	Austria
CAP1218	7	Tissue	Austria
CAP1220	7	Tissue	Germany
CAP1222	7	Tissue	Norway
CAP1223	2.84	Tissue	Norway
CAP1224	7	Blood	Sweden
CAP1225	7	Blood	Sweden
CAP1226	7	Blood	Sweden
CAP1227	4.93	Blood	Sweden
CAP1228	7	Blood	Sweden
CAP1229	7	Blood	Poland
CAP1230	7	Blood	Poland
CAP1231	7	Blood	Poland
CAP1232	7	Blood	Poland

Action 2 – Extract DNA from 4 additional tissue samples provided by CCP

The RZSS WildGenes laboratory received 5 tissue samples from the CCP, collected from carcasses found within the Cairngorms National Park. DNA was extracted from each, quality checked and run through a species ID test. One sample was identified as a Common pheasant (*Phasianus colchicus*) and four were identified as capercaillie. The DNA quality was highly variable, with only one (CAP1198) exhibiting high

quality and quantity of DNA (Figure 1). However, the presence of some high molecular weight DNA in each of the extractions meant that all four were included in the genomic library (see Action 1).

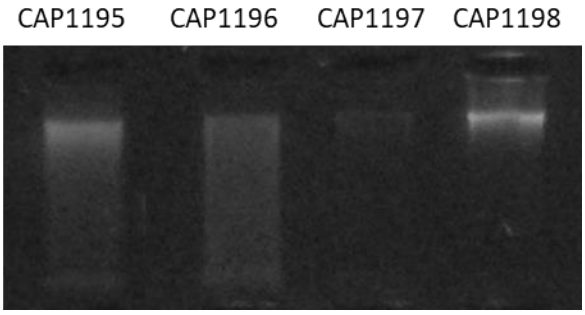

Figure 1. *Gel electrophoresis image of the four tissue samples identified as capercaillie showing highly variable DNA quality. CAP1198 is the only sample with a high quantity of extracted DNA*.

Action 3 – Sequence and prepare a library report

This progress report constitutes delivery towards this action. Three sequencing approaches are being used in this project i) mitochondrial sequencing, ii) ddRAD genomic library preparation followed by illumina next-generation sequencing and iii) Hybrid capture followed by illumina next-generation sequencing. Progress has been made on the first two approaches; however the third approach relies on completion of approach 2.

i) Mitochondrial sequencing

To date, sequences of the d-loop region of the mitochondrial genome have been sequenced and quality checked for 350 of the feather samples collected by the CCP. They currently reveal 5 haplotypes within the sample-set. The mitochondrial d-loop region has also been sequenced for all 85 samples included on the ddRAD genomic library. These haplotypes have been compared with those previously published by Segelbacher and Piertney (2007), revealing that within our dataset there are 9 of the 21 haplotypes previously identified from across European capercaillie populations. There is also an additional novel haplotype within our dataset. The samples included in our genomic library therefore include mitochondrial haplotype representation from across a substantial proportion of the known European diversity (Figure 2).

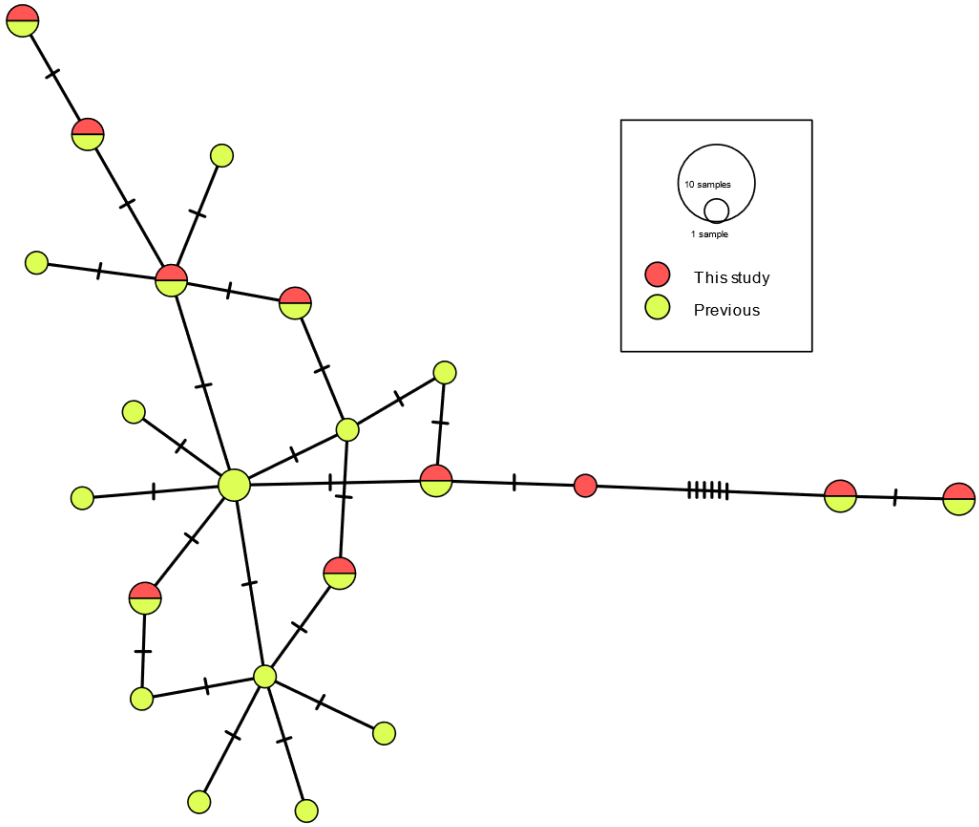

Figure 2. *Median joining haplotype network for a 252bp mitochondrial d-loop sequence alignment. Each circle represents a unique maternal lineage found in Europe. The colours denote lineages identified in a previous genetic study (Yellow - Segelbacher & Piertney 2007) and those identified in samples being used within this study (red). The cross hairs on the lines between the circles denote the number of mutational differences between each haplotype.*

ii) ddRAD genomic library

As outlined under action 1, a genomic library has been prepared using DNA extracted from 85 capercaillie samples originating from 7 regions in Europe, including DNA from blood/tissue samples and the highest quality feather samples. The library has been sequenced on an Illumina Hi-seq platform and the sequence data demultiplexed before use in a single-nucleotide polymorphism (SNP) selection analysis. All 85 samples have each produced >250,000 sequence reads (the criteria for progressing to the next analysis stage). Appendix 1 contains our quality-check report that is produced after de-multiplexing. The library contains 10 replicates (suffixed a & b) and a negative control. As the library contained samples from a

range of sample types, DNA quantity was standardised during library preparation to improve the success of the more degraded samples. The number of reads produced is a rough indication of sample success and as can be seen in Figure 3, although variable, a substantial proportion of the feathers have produced over 5 million sequence reads. The number of sample reads associated with each sample can be found in Appendix 2.

iii) Hybrid capture

SNP selection for the hybrid capture probes is now underway. The development of this approach will be used to sequence the more degraded feather samples to support action 5.

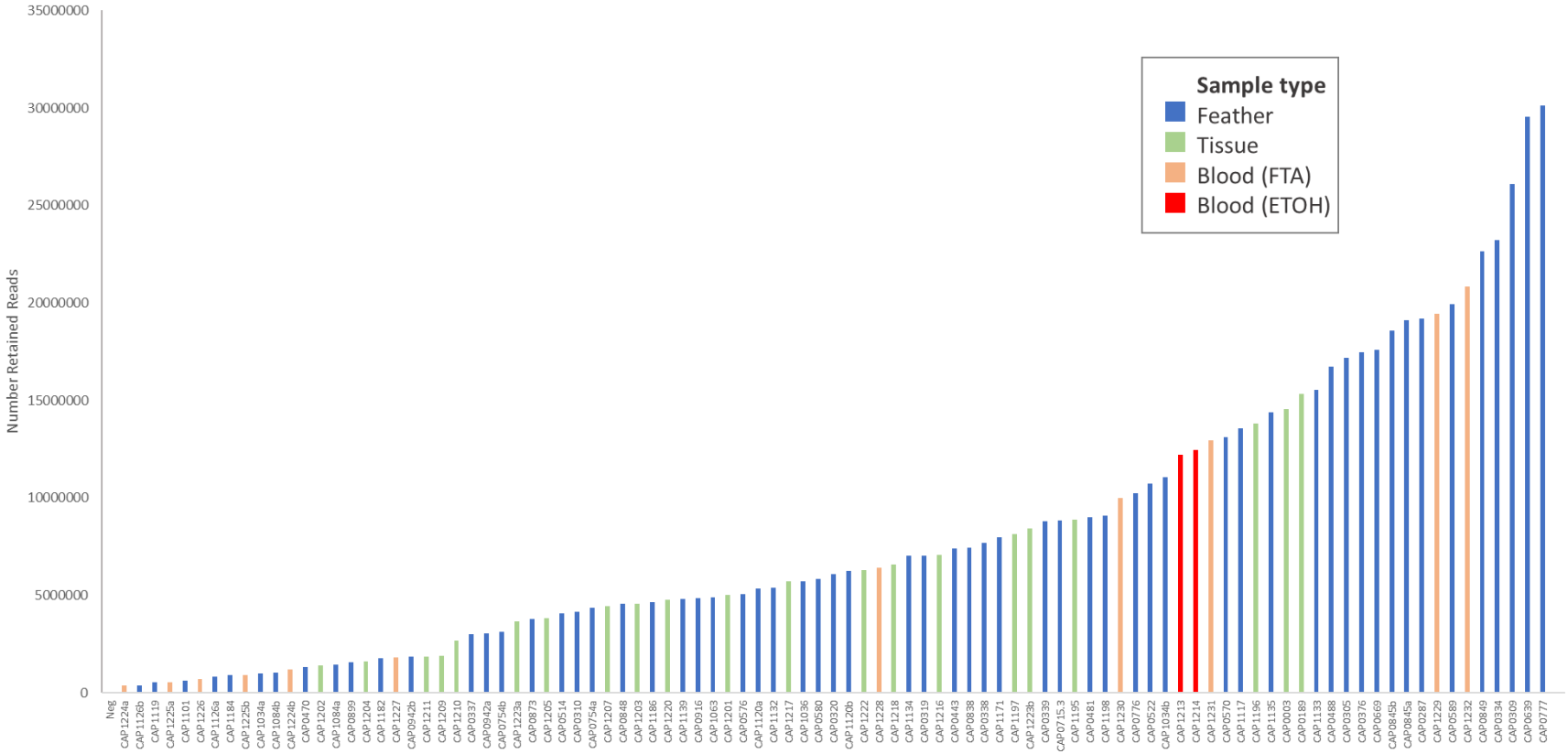

Figure 3. *The number of sequence reads produced for each sample included in the ddRAD genomic library after sequencing on an Illumina Hi-seq platform. The quality threshold of >250,000 reads was reached for every sample.*

Action 8 – Source material from capercaillie populations outside of the UK

Tissue/blood samples have been sourced from wild capercaillie in 7 regions across Europe: Scotland, Norway, Sweden, Poland, Austria, Germany and France (Table 2). DNA extraction was attempted from all 33 samples which were then tested for species identity (Verma & Singh 2002), quantified via a NanoDrop™ ND-1000, genetically sexed and quality checked using gel electrophoresis. DNA was extracted from all samples, however two samples were identified as golden eagle (CAP1219 & CAP1221) and four samples failed the ID test (Table 2). The remaining 28 samples were included in the ddRAD genomic library alongside the DNA from 57 CCP feathers. As outlined in action 3, all have successfully passed the first quality check and will progress to the SNP selection stage. The large number of populations sampled means that the genomic markers identified in the next analysis stage will not only be applicable to studying the Scottish capercaillie population but will be relevant for comparisons with other capercaillie populations across Europe.

Table 2. *Summary of the additional tissue/blood samples sourced for inclusion in the marker development to increase the applicability of the markers to multiple populations across Europe.*

Sample ID	Sample type	DNA conc (ng/ul)	Sex	Species ID	DNA quality	Origin	Used in ddRAD
CAP0003	tissue	7	Male	Capercaillie	Good	Scotland	Yes
CAP0189	tissue	116	Male	Capercaillie	Medium	Scotland	Yes
CAP1201	tissue	199.7	Male	Capercaillie	Medium	France	Yes
CAP1202	tissue	1356	Male	Capercaillie	Medium	France	Yes
CAP1203	tissue	439.4	Male	Capercaillie	Medium	France	Yes
CAP1204	tissue	2401.9	Male	Capercaillie	Medium	France	Yes
CAP1205	tissue	799.8	Male	Capercaillie	Medium	France	Yes
CAP1206	tissue	1228.9	Male	fail	Medium	France	No
CAP1207	tissue	226.5	Male	Capercaillie	Medium	France	Yes
CAP1208	tissue	2304.9	Male	fail	Medium	France	No
CAP1209	tissue	1139.3	Male	Capercaillie	Medium	France	Yes
CAP1210	tissue	1515.8	Male	Capercaillie	Medium	France	Yes
CAP1211	tissue	2394.5	Male	Capercaillie	Medium	France	Yes
CAP1212	tissue	431.6	Male	fail	Poor	France	No
CAP1213	Blood (ETOH)	24.5	Male	Capercaillie	Good	Germany	Yes
CAP1214	Blood (ETOH)	55.2	Male	Capercaillie	Good	Germany	Yes
CAP1215	Blood (ETOH)	107.9	Female	fail	Good	Germany	No
CAP1216	tissue	127.7	Male	Capercaillie	Medium	Austria	Yes
CAP1217	tissue	49.7	Male	Capercaillie	Medium	Austria	Yes

CAP1218	tissue	44.9	Male	Capercaillie	Good	Austria	Yes
CAP1219	tissue	56.3	Male	Golden eagle	Poor	Germany	No
CAP1220	tissue	59.4	Male	Capercaillie	Good	Austria	Yes
CAP1221	tissue	40.5	Male	Golden eagle	Poor	Germany	No
CAP1222	tissue	43.3	Female	Capercaillie	Good	Norway	Yes
CAP1223	tissue	33.6	Male	Capercaillie	Poor	Norway	Yes
CAP1224	Blood (FTA)	41.2	Male	Capercaillie	Poor	Sweden	Yes
CAP1225	Blood (FTA)	62.9	Female	Capercaillie	Poor	Sweden	Yes
CAP1226	Blood (FTA)	19.4	Male	Capercaillie	Good	Sweden	Yes
CAP1227	Blood (FTA)	8.4	Male	Capercaillie	Medium	Sweden	Yes
CAP1228	Blood (FTA)	40	Male	Capercaillie	Medium	Sweden	Yes
CAP1229	Blood (FTA)	27.8	Male	Capercaillie	Good	Poland	Yes
CAP1230	Blood (FTA)	27.7	Male	Capercaillie	Good	Poland	Yes
CAP1231	Blood (FTA)	18.7	Male	Capercaillie	Good	Poland	Yes
CAP1232	Blood (FTA)	23.4	Female	Capercaillie	Good	Poland	Yes

Action 9 – Historical understanding of the Cairngorms capercaillie population

We have approached four museum collections that contain capercaillie specimens obtained from populations in Scotland and which include two of the most extension zoology collections within the UK (National Museums Scotland in Edinburgh & the Natural History Museum in London). The specimen collection dates range between 1798 and 1997, potentially providing insight into the population during the last 200 years. However, any DNA remaining within them will be highly degraded. The plan is to use the targeted enrichment approach we are developing for the feathers to also increase the amount of genetic data we can obtain from these museum samples.

SUMMARY

Currently the genetic lab work is on course. Mitochondrial sequencing of the DNA extracted from the feather samples has progressed substantially in order to generate a dataset that is comparable to those previously published. Thirty-nine additional tissue/blood samples have been obtained and evaluated for inclusion in the genomic marker development. The samples now contain representatives from wild populations in seven countries and this will be invaluable for designing a genomic toolkit that is not only applicable to capercaillie populations in Scotland but also throughout mainland Europe. Eighty-five samples are currently being used to generate the toolkit required to work with the degraded feather samples. The first stage has been the production of the ddRAD genomic library which has now been sequenced via next generation sequencing. The next step is to use these sequences to select genomic markers for the design of genetic probes that will allow us to enrich for targeted regions in the genomes of the degraded samples. We have also approached four museum collections that contain capercaillie specimens, and the hope is these will provide us with an historical perspective to the capercaillie population in Scotland.

REFERENCES

Segelbacher, G. & Piertney, S. 2007. "Phylogeography of the European capercaillie (*Tetrao urogallus*) and its implications for conservation" *Journal of Ornithology* 148: 269–274. https://doi.org/10.1007/s10336-007-0153-1http://doi.wiley.com/10.1046/j.1471-8286.2002.00180.x.

Verma, S. K., and L. Singh. 2002. "Novel Universal Primers Establish Identity of an Enormous Number of Animal Species for Forensic Application." *Molecular Ecology Notes* 3(1): 28–31. http://doi.wiley.com/10.1046/j.1471-8286.2003.00340.x

APPENDICES

Appendix 1. *Quality check report produced after demultiplexing of the Illumina sequencing data.*

This can be found in the document entitled: **Appendix_1_QC_post_demultiplex_RZSS51.pdf**

Appendix 2. *Details of the number of sequence reads for each sample in the ddRAD library.*

The results can be found in this document: **Appendix_2_sequence_reads_ddRAD_samples.xlsx**