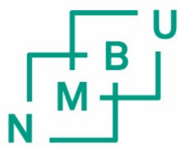



Norges miljø- og
biovitenskapelige
universitet

Master's Thesis 2017 60 ECTS
Department of Animal and Aquacultural Sciences

# Genetic diversity in five chicken lines from the Norwegian live poultry gene bank

Cathrine Brekke
Animal Science

**Forord**

Denne oppgaven ble skrevet våren 2017 og er avsluttende arbeid på en 5-årig mastergrad i husdyrvitenskap ved institutt for husdyr- og akvakulturvitenskap ved Norges miljø- og biovitenskapelige universitet. Det er en oppgave innen avl og genetikk, som er det fagområdet jeg har valgt å fordype meg i. Jeg har likt godt å fordype meg i én problemstilling og har lært mye det siste året, både faglig og teknisk.

Jeg vil gjerne takke hovedveilederen min Theo Meuwissen for god veiledning og hjelp med koder og programvarer. Takk for lærerike diskusjoner om metoder og resultater. I tillegg vil jeg takke Peer Berg for god faglig veiledning og for at jeg fikk jobbe med ett av deres prosjekter ved NordGen. Takk for nyttige tilbakemeldinger på oppgaven. En stor takk går også til Linn Groeneveld som har sittet mange timer og hjulpet meg med datamaterialet og alltid vært tilgjengelig for spørsmål og veiledning av skrivearbeidet. Tusen takk!

Til slutt vil jeg gjerne takke Martine Cornelia Franck for grafisk hjelp og Nikolai Aksnes for viktig hjelp med programmering.

Ås, mai 2017

Cathrine Brekke

# Table of Contents

**Forord** 1

**Sammendrag** 3

**Abstract** 4

**1. Introduction** 5

**2. Material and methods** 7

2.1 Genetic lines 7

2.2 DNA extraction and Genotyping 9

2.3 Comparable dataset 11

2.4 Genomic relationship matrix G 13

2.5 Contributions to genetic diversity 14

**3. Results** 16

3.1 Genomic relationships 16

3.2 Contributions to genetic diversity 21

**4. Discussion** 23

4.1 Data reliability 23

4.2 Genomic relationships 24

4.3 Contributions to genetic diversity 26

4.4 perspectives 28

**5. Conclusion** 30

**References** 31

**Appendix A** 33

## Sammendrag

Siden 1995 har det ikke vært aktivt avlsarbeid på verpehøner i Norge. De siste aktive linjene er bevart på den levende genbanken for fjørfe på Hvam sammen med andre bevaringsverdige raser. Målet med dette studiet var å kartlegge genetisk diversitet innen og mellom fem av de norske linjene I tillegg til å kvantifisere deres eventuelle genetiske unikhet i et internasjonalt perspektiv. De fem linjene i studiet består av fire tidligere aktive verpehønslinjer og etterkommeren av den tidligere landhøna, Jærhøns. De fem linjene ble genotypet med SNP-chipen Affymetrix® Axiom® Chicken Genotyping Array. Genomisk slektskap innen og mellom de fem linjene ble estimert basert på 468 973 SNP markører. Observert heterosygositet lå mellom 0.32 for Jærhøns og 0.40 for NorBrid 8, den eneste brune egg leggeren i studiet. De norske linjene ble sammenlignet med 70 andre raser av ulik internasjonal opprinnelse. For å kvantifisere bevaringsverdi ble de norske linjene rangert etter relative bidrag til genetisk diversitet. Estimerte tap av genetisk diversitet dersom en norsk linje gikk tapt lå mellom 0.17 og 2.84 % relativt til det fulle settet med 75 linjer.
Begge analysene indikerer høyere grad av genetisk diversitet i NorBrid 8, som verper brune egg. Lavere genetisk diversitet i linjer som verper hvite egg enn i linjer som verper brune egg er i samsvar med andre studier.

## Abstract

Since 1995, there has been no active breeding of commercial egg layers in Norway. The last active breeding lines, as well as other breeds of conservation value are conserved in the Norwegian live poultry gene bank. The aim of this study was to evaluate genetic diversity within and between five of the Norwegian lines as well as the genetic uniqueness in an international context. Included in the study is four of the last active breeding lines as well as the national Norwegian landrace, Jærhøns. The five lines were genotyped with the Affymetrix® Axiom® Chicken Genotyping Array. Genomic relationships within and between the five lines were estimated from 468 973 SNP markers. The observed heterozygosity ranged from 0.32 in Jærhøns to 0.40 in NorBrid 8, the only brown egg layer in the study. Comparative analyses were carried out with 70 other populations of different origins to assess conservation value in an international context. The Norwegian lines were ranked according to their relative contributions to genetic diversity. Losses in genetic diversity ranged from 0.17 to 2.84 % when one of the Norwegian lines were lost relative to the entire set of 75 lines. Both analyses indicate higher genetic diversity in the brown egglayer. Lower genetic diversity estimates for white egg layers than for brown egg layers is in agreement with other studies.

# 1. Introduction

The first signs of domestication of chicken are nearly 8000 years old. It is commonly known that the Jungle fowl (Gallus gallus gallus), originating from Southeast Asia, is the ancestor of all domesticated chicken breeds (Fumihito et al, 1996). During the last century, man has created a wide variety of breeds for different purposes. The isolation of populations and development of breed specific standards of phenotypes has lead to decreased genetic diversity within breeds. However, the extent of different breeds of chicken and phenotypic variation still existing, suggests that a substantial genetic variation may still be present between breeds. Since the 1990's, there has been a drastic reduction in companies supplying the global poultry market with genetic material. Today, more than 70% of the worlds commercial egg production is based on genetics from only 2 large companies producing specialized, highly productive lines (Gura et al, 2007). Hence, the evaluation of existing genetic diversity, and establishment of conservation priorities in chicken is of great importance.

In the 1960s there were 23 different breeding stations in Norway and a total of 26 different lines of egg layers. From 1969 to 1973 a national project was carried out to compare the lines for a number of production traits. In the following years, several breeding stations were shut down and only the most productive lines were continued. In 1973, the Norwegian live poultry gene bank was established. The initial purpose of the gene bank was to have a copy of the active Norwegian breeding lines as a security in case of disease or accidents at the breeding stations. The gene bank was also going to conserve the national breed and descendant of the Norwegian landrace, Jærhøns. In 1974 the gene bank was moved to Hvam Agricultural College and it has since been there. In 1994 the European Economic Area Agreement opened for import of grandparental livestock for egg production. The active Norwegian egg laying lines were soon outcompeted by highly productive lines from international breeding companies, and the Norwegian Poultry Breeding Association (NPBA) ended their breeding work in 1995. Today, the aim of the gene bank is to conserve the last active Norwegian breeding lines as well as other breeds of conservation value (Genressursutvalget for husdyr, 2002).

Poultry breeds can be of cultural and historical conservation value. They may also be a genetic resource for future poultry breeding and food production if a breed or population is found to be genetically distinguishable from other domesticated breeds. Quantifying the conservation value of the genetic material in the gene bank is of importance in order to evaluate the conservation priorities and management of genetic diversity in the gene bank.

The first high-density genotyping array developed for high-resolution genetic analysis of chicken has recently become publically available. The genotyping Affymetrix ® Axiom® array for chicken includes more than 580 000 single nucleotide polymorphisms (SNPs) segregating in a wide variety of breeds and populations (Kranis et al, 2013). One of the advantages of using SNPs to analyze genetic diversity is that they can be found in any region throughout the genome including introns, exons, regulatory sites etc., as opposed to molecular markers like microsatellites and minisatellites that are assumed to be selectively neutral (Fernández and Bennewitz, 2017). The possibility to genotype high densities of SNPs across the chicken genome gives a broad genetic coverage and may be the best available representation of the true genetic diversity. It is also of interest to use a genotyping platform that enables comparison with recent, ongoing or future analysis of other populations internationally. This will give a more complete picture of the genetic diversity and potential uniqueness of the Norwegian lines.

Included in this study are five genetic lines from the Norwegian gene bank; the national breed, Jærhøns, as well as 4 lines that were previously bred for commercial egg production in Norway. Comparative analyses are carried out with 70 populations of different origins. The aim of this thesis was to evaluate genetic diversity, measured as observed heterozygosity, within and between five conserved genetic lines of laying hens from the live poultry gene bank at Hvam as well as quantifying conservation value in terms of relative contributions to genetic diversity. The results will aid in the discussion of the conservation value of the five lines in a national perspective, as well as the future of the gene bank. The results may also be valuable in the discussion of chicken genetic diversity in an International perspective.

# 2. Material and methods

## 2.1 Genetic lines

From the live poultry gene bank at Hvam Agricultural College in Norway the following five genetic lines of Norwegian egg layers are included in this study. Jærhøns (Figure 2.1) is the national Norwegian chicken breed and represents the original Norwegian landrace. Roko 1 is the oldest, closed breeding line in Norway and the only line left of Rokohøns. NorBrid 1 and NorBrid 4 were the maternal and paternal line, respectively, in the crossing NorBrid 41 which was the most commonly used commercial egg layer in Norway until 1994. NorBrid 8 was the paternal line in the most common brown egg layer hybrid, NorBrid 78, and is the only brown egg layer in the study. Number of samples genotyped from each line can be found in Table 2.1. Throughout this report chicken breeds and populations will be referred to as genetic lines. There is a further description of the lines in Table 2.2.

**Table 2.1** Number of male and female samples that were genotyped from each of the five genetic lines.

| Line | # Individuals | # Males | # Females |
|---|---|---|---|
| Jærhøns | 19 | 5 | 14 |
| Rokohøns | 20 | 5 | 15 |
| NorBrid 1 | 20 | 5 | 15 |
| NorBrid 4 | 19 | 5 | 14 |
| NorBrid 8 | 18 | 5 | 13 |
| Total | 96 | 25 | 71 |

**Figure 2.1** Jærhøns. Photo: Anna Rehnberg / © Norsk genressurssenter / NIBIO

The description of the five genetic lines of laying hens in this study (Table 2.2) must be considered to be current at the time when they were taken into the gene bank. The lines have since been bred as closed lines and there has been no systematic documentation of traits. The selection has been minimal, only leaving out dysfunctional and unfertile animals and concentrating on maintaining typical breed signatures and the phenotypic variation present.

**Table 2.2** Description of the five genetic lines included in the study. Information is collected from a report on conservation of poultry genetic recourses in Norway (Genressursutvalget for husdyr, 2002).

| Line | Description |
|---|---|
| Jærhøns | Jærhøns is the only breed left of the Norwegian landrace. The Norwegian landrace was almost extinct when the work to save the breed was initiated at Jæren in 1916 and the breed was named Jærhøns. At that time, the breed encompassed a large variety in color, but there was a demand for color standardization and work was put into creating a breed with a stable inheritance of color. Jærhøns is a white egg layer that was systematically bred for production up until 1973, when it was no longer commercially competitive. In 1974 the breeding population arrived to the newly started gene bank at Hvam, and has since been kept here in a pure line. Jærhøns is known to have a low bodyweight with relatively large eggs and good shell quality. The breed is auto-sexing with different color in down in day old male and female chicken. Jærhøns appears in a dark and lighter shade of brown with a striped pattern, the light variety is in minority. |
| Roko 1 | Roko 1 is the last active breeding line of Rokohøns. It is the oldest existing purebred line in Norway. Roko 1 is an active line of White Leghorn, originating in Italy. Roko 1 was the maternal line in the white egg layer hybrid Rokohøns 21. The line has a relatively low bodyweight and high productivity and is therefore characterized by high feed efficiency. The line will be referred to as Rokohøns in this study. |
| NorBrid 1 | NorBrid 1 is a line of White Leghorn, derived from the 1977 generation of the line Roko 4. NorBrid 1 was used as the maternal line in the white egg laying hybrid NorBrid 41, which was the hybrid that covered 80 % of the national egg market for many years, until Norway opened for import of grandparental livestock for egg production in 1994. NorBrid 1 hens are calm and productive egg layers and the eggs have a good shell quality. |
| NorBrid 4 | NorBrid 4 is a line descending from White Leghorn derived from the genetic line Melsomlinjen in 1972. NorBrid 4 is known as a productive line with a calm temperament and good heterosis effect in crossbreeding. It was the paternal line in the white egg laying hybrid NorBrid 41. |
| NorBrid 8 | NorBrid 8 is a brown egg layer, with brown coat color, descending from Red Rhode Island (Brenøe & Kolstad 2001). It was imported from Hissex in Sweden in 1981 and bred in a pure line in Norway since 1982. This line was the paternal line in the crossing NorBrid 87 which was the main brown egg layer on the Norwegian market until Norway opened for import of grandparental livestock for egg production in 1994. |

## 2.2 DNA extraction and Genotyping

Blood samples were collected from animals from all the lines at the gene bank in 2011 and DNA was extracted in 2012. Blood samples (approximately 2 ml) were obtained from the wing vein with a 21ga, 1.5 in needle. Blood was stored at -20°C in sodium heparin tubes. One drop of thawed blood was transferred to FTA Elute Micro Card (GE Healthcare, piscataway, USA) and allowed to dry. DNA was extracted from the dried blood spot on the FTA card following the manufacturer's instructions. The quality and concentration of the DNA was tested at Biobank Hamar, and 96 samples were selected for further analysis. The samples were normalized to 10 ng/μL and 50 μL.

The genotyping was carried out at Centre for Integrative Genetics (CIGENE) with the Affymetrix® Axiom® Chicken Genotyping Array. The high density SNP Array consists of 580,961 SNPs evenly distributed along the chicken genome and segregating in a wide variety of chicken breeds and populations (Kranis et al. 2013). The array will be referred to as the 600K SNP array. Two genotyping quality control steps were carried out by CIGENE; dish quality control (DQC) and call rate quality control (QC-CR) (Affymetrix 2017, p. 15). One Individual from NorBrid 8 failed DQC and was removed from the dataset. One individual initially failed the QC-CR. Because of a relatively low number of individuals in the study and high number of SNPs it was decided to include the individual and remove the 1955 SNPs that caused the sample to fail QC-CR from the whole dataset (see Table 2.3).

The following filtering was carried out using PLINK 1.9 (Chang et al. 2015). Some SNPs have duplicate probe sets reading the same SNP, as they were originally mapped to different locations (Affymetrix 2017 p. 64). For each SNP with duplicate probe sets, 1 random probe set was removed. Within lines it was filtered for SNPs and samples with missing call rates above 10%. A Hardy Weinberg equilibrium (HWE) exact test was also performed within lines, using the mid p-value option in PLINK (Graffelman & Moreno, 2013). SNPs deviating significantly from HWE ($p < 0.001$) were excluded.

The whole dataset was filtered again for missing SNP and sample call rates above 10%. As each line represent more than 10% of the whole dataset, SNPs excluded within lines are consequently excluded from the whole dataset when filtering for missing SNP call rates above 10 %. SNPs with minor allele frequencies lower than 0.01 were removed from the dataset. An overview of the SNP filtering can be found in table 2.3. The final dataset consisted of 468 973 SNP markers and 95 samples, from 17 to 20 individuals per line. This dataset will be referred to as the Norwegian dataset. An overview of the samples from each line included in the analysis can be found in table 2.4.

**Table 2.3** Genotyping quality controls performed in PLINK1.9. MCR = Missing call rates

| Filter | # SNPs left | # SNPs lost | % Lost |
|---|---|---|---|
| Initial dataset | 580 961 | 0 | 0 |
| QC-CR | 579 006 | 1 955 | 0.34 |
| Duplicate probe sets | 578 962 | 44 | 0.01 |
| Deviation from HWE $p < 0.001$ and SNP MCR >10% | 539 499 | 39 463 | 6.82 |
| Sample MCR >10% | 539 499 | 0 | 0 |
| SNPs with MAF < 0.01 | 468 973 | 70 526 | 13.07 |

**Table 2.4** Number of samples in the final Norwegian dataset

| Line | # Individuals | # Males | # Females |
|---|---|---|---|
| Jærhøns | 19 | 5 | 14 |
| Rokohøns | 20 | 5 | 15 |
| NorBrid 1 | 20 | 5 | 15 |
| NorBrid 4 | 19 | 5 | 14 |
| NorBrid 8 | 17 | 5 | 12 |
| Total | 95 | 25 | 70 |

### 2.3 Comparable dataset

A set of comparable data was provided from an ongoing study led by Steffen Weigend at the Institute of Farm Animal Genetics, Friedrich-Loeffler-Institut Germany as part of the Synbreed project. A report of the first results from this ongoing research has been published (Weigend et al. 2014). The dataset provided for comparison with the Norwegian dataset consist of two individuals, one male and one female, from 70 different populations of chicken. The genotyping was carried out with the same 600 K SNP array. Genetic lines in this dataset range from fancy breeds to traditional breeds with Asian and European origin, as well as two wild types, *Gallus gallus gallus* and *Gallus gallus spadiceus*. An overview of the populations in the dataset and their origin is provided in Table 2.6. The two datasets were merged and the same SNPs that were excluded in the filtering process of the Norwegian dataset were excluded. Filtering for SNPs and samples with missing call rates > 10 % was carried out. The final comparable dataset consist of 468 918 and 140 individuals from 70 lines and will be referred to as the Synbreed dataset. The final merged dataset consist of 468 918 SNPs and 235 individuals from 75 genetic lines and will be referred to as the international dataset. An overview of the final three datasets used in the analysis can be found in Table 2.5

**Table 2.5** An overview of the contents of the three datasets used in the analysis

| Dataset | # Lines | #Individuals | # Males | # Females | # SNPs |
|---|---|---|---|---|---|
| Norwegian | 5 | 95 | 25 | 70 | 468 973 |
| Synbreed | 70 | 140 | 70 | 70 | 468 918 |
| International | 75 | 235 | 95 | 140 | 468 918 |

**Table 2.6.** An overview of the genetic lines in the Synbreed dataset (Weigend et al. 2014).

| Origin | Type | Lines |
|---|---|---|
| **Asia** | Long tailed breeds | *PHxx, SAsch, YOwr* |
| | Game type and related breeds | *ASrb, IKxx, Maxx, OFrbx, SHsch* |
| | Asian type breeds | *BHrg, BHwsch, COsch, DLIa, MRschk, NHbr, NHL68, ORge, PRgp, ROro, SNwsch, TOgh, WYsschs, WYw* |
| | Crested breeds | *SEsch, Sew* |
| | Bantam breeds | *CHgesch, CHschw, KSgw, OHgh, OHsh, ZCsch, ZCw* |
| **Europe** | Intermediate type breeds | *ARsch, ARwi, DOxx, VWco, VWcoE* |
| | Mediterranian type breeds | *ITrh, ITsch, KAsch, LER11, LEw, Misch* |
| | Northwest-European breeds | *HAsI, KRsch, KRw, LAco, OMsschg, RHrh, RHsch, THsch, WTs* |
| | Crested breeds | *APsscht, HOxx, PAxx* |
| | Bantam breeds | *ABwa, BAsch, DZgh, FZgpo, FZsch, GBxx, SBgschs, SBsschs* |
| **Wild** | *Gallus gallus gallus* | *GGg* |
| | *Gallus gallus spadiceus* | *GGsc* |

**Line abbreviations**

Bwa - Barbue d'Anvers quail
AKxx - Carlise Old English Game any colour
APsscht - Appenzeller Pointed Hood silver spangled
ARsch - Rumpless Araucana black
ARwi - Rumpless Araucana black breasted red
ASrb - Aseel red mottled
BAsch - Rosecomb Bantam black
BHrg - Brahma gold
BHwsch - Brahma light
BKschg - Bergische Crower
BLxx - Brakel silver
BSsch - Berg-Schlotter black
CHgesch - Japanese Bantam black tailed buff
CHschw - Japanese Bantam black mottled
COsch - Cochin black
DLIa - German Faverolles salmon
DOxx – Dorking any colour
DSgp - German Grey Chickens cuckoo
DZgh - German Bantam gold partridge
FRgew - Frisian Fowl chamois penciled
FZgpo - Booted Bantam millefleur
FZsch - Booted Bantam black
GBxx - Barbue du Grubbe any color
HAsl - Hamburgh silver spangled
HOxx - Poland White Crested black
IKxx - Indian Game dark
ITrh - Leghorn brown
ITsch - Leghorn black
KAsch - Castilians black
KRsch - Creeper black
KRw - Creeper white
KSgw - Ko Shamo black-red
LAco - Lakenvelder black and white
LER11- White Leghorn line R11
LEw - White Leghorn
Maxx - Malay black red
Misch – Minorca black

MRschk - Marans copper black
NHbr - New Hampshire red
NHL68 - New Hampshire line 68
OFrbx - Orloff red spangled
OHgh - Ohiki red duckwing
OHsh - Ohiki silver duckwing
OMsschg - East Friesian Gulls silver pencilled
ORge - Orpington buff
PAxx - Poland any color
PHxx - Phoenix golden or golden duckwing
PRgp - Plymouth Rocks barred
RHrh - Rhinelander Chicken brown
RHsch - Rhinelander Chicken black
ROro - Rhode Island Red red
SAsch - Sumatra black
SBgschs - Sebright Bantam golden
SBsschs - Sebright Bantam silver
SEsch - Silkies black
SEw - Silkies white
SHsch - Shamo black
SNwsch - Sundheimer light
THsch - Thuringian Bearded Chicken black
TOgh - Toutenkou black breasted red
VWco - V orwerk buff columbian
VWcoE - Vorwerk conservation program
WTs - Westphalian Chicken silver
WYsschs - Wyandotte silver laced
WYw - Wyandotte white
YOwr - Y okohama red saddled white
ZCsch - Pekin Bantam black
ZCw - Pekin Bantam white

Ggg - Gallus gallus gallus
Gsc – Gallus gallus spadiceus

### 2.4 Genomic relationship matrix G

Genomic relationships were calculated with the software Gghat (version 3, Meuwissen 2015), based on the following formula presented by Vanraden (2008).

Let **M** be the matrix with marker alleles for each individual at each locus, with dimension n (number of individuals) by m (number of loci). **P** is the allele frequencies at each loci *i,* expressed as the difference from 0.5 and multiplied by 2, i.e column *i* of P is $2(p_i - 0.5)$. $\boldsymbol{Z} = \boldsymbol{M} - \boldsymbol{P}$. Subtracting **P** from **M** is done to set the mean value of allele effects to zero. The following gives the genomic relationship matrix **G**:

$$\mathbf{G} = \frac{\boldsymbol{ZZ'}}{2\sum \mathrm{p_i}(1-\mathrm{p_i})} \quad (1)$$

The diagonal of **G** contains the genomic inbreeding coefficients plus 1, i.e. the genomic inbreeding coefficient F for individual j is measured as:

$$F_j = \mathrm{G_{jj}} - 1. \quad (2)$$

Because there are no available allele frequencies from a base population in the study, allele frequencies were set to 0.5. Note that setting the allele frequencies to 0.5 gives $2(p_i - 0.5) = 0$, ultimatley giving the following formula for calculation of **G**:

$$\mathbf{G} = \frac{\boldsymbol{MM'}}{2\sum \mathrm{p_i}(1-\mathrm{p_i})} \quad (3)$$

Observed heterozygosity within or across lines is measured as:

$$H_o = 1 - F_{mean} \quad (4)$$

The heat maps presenting the genomic relationship matrices are created in R studio Version 1.0.136 (R Core Team, 2016) with the Heatmaply package (Galili, 2016). The heatmaply function calculates a distance matrix from the G matrix following Euclidean distance. The dendrogram clustering follows complete linkage agglomerative hierarchical clustering. The branches are rotated to find the optimal ordering of rows and columns so that the sum of distances between each adjacent sample is minimized (Galili 2015).

**2.5 Contributions to genetic diversity**

The core set method, as suggested by Eding et al. (2002) was followed to measure the contribution of a genetic line to genetic diversity. Let G be the n × n matrix containing within and between line genomic relationships for n lines in set S. Also consider an n-dimensional vector c that contains the relative contributions of each line to the core set, where the elements of c sum to one. The average relationship in the set S, given c, can be calculated as follows:

$$f(S) = c'Gc \tag{5}$$

The core set is constructed so that no lines get negative contributions. The line with the most negative contribution is removed from the core set and the contribution is set to zero. This is repeated until all contributions are equal or greater than zero. The vector c is restricted so that the elements sum to one and the average relationship in the core set is minimized. The vector is estimated as:

$$c_{min} = \frac{G^{-1}1_n}{1'_n G^{-1} 1_n}. \tag{6}$$

The minimum relationship in the core set, $f(S)_{min}$, can than be obtained from

$$c'_{min} G c_{min} = \frac{1}{1'_n G^{-1} 1_n} \tag{7}$$

As the genetic variance within set S is proportional to $(1 - f(S)_{min})$, the genetic diversity Div(S) is defined as:

$$Div(S) = (1 - f(S)_{min}). \tag{8}$$

The following was done to measure the relative loss of genetic diversity if one line is lost. If *l* is a single line in set S, and $S_{-l}$ is the set of lines where *l* is excluded, then the diversity in the set when line *l* is lost is:

$$Div(S_{-l}) = (1 - f(S_{-l})_{min}). \tag{9}$$

Relative genetic diversity lost if this line is lost from set S is then:

$$\% \, lost = \frac{Div(S) - Div(S_{-l})}{Div(S)} \cdot 100 \tag{10}$$

Unless else is stated, calculations are carried out with ad hoc scripts in R studio Version 1.0.136 (R Core Team, 2016).

# 3. Results

## 3.1 Genomic relationships

Figure 3.1 is a heat map presenting the 5 x 5 **G** matrix containing average genomic relationships within and between the lines in the Norwegian dataset. A lighter shade reflects higher relationship estimates. The ordering of the lines is decided by the clustering in the dendrogram. The dendrogram in Figure 3.1 shows two major clusters; NorBrid 8, which is the only brown egg layer in the Norwegian dataset, forms one cluster and the white egg layers NorBrid 1, NorBrid 4, Rokohøns and Jærhøns forms another cluster. The column and row representing the genomic relationships between NorBrid 8 and the other lines is clearly darker than the rest, reflecting lower relationship estimates. The between line genomic relationship estimates range from 0.19 between NorBrid 8 and NorBrid 1, to 0.76 between NorBrid 4 and Rokohøns.

Figure 3.2 is a heat map presenting the 95 x 95 **G** matrix containing the genomic relationships between and within individuals in the Norwegian dataset. The diagonal is the estimated genomic inbreeding coefficient + 1 for each individual. The Inbreeding coefficients range from 0.28 to 0.72. Mean genomic inbreeding and observed heterozygosity in the Norwegian lines are presented in Table 3.1.

**Table 3.1** Means and SD of the genomic inbreeding coefficient (F) and observed heterozygosity ($H_o$) for the genetic lines in the Norwegian dataset.

| Genetic line | Nr of samples | Mean (F) | SD | $H_o$ |
|---|---|---|---|---|
| Jærhøns | 19 | 0.68 | 0.02 | 0.32 |
| Rokohøns | 20 | 0.61 | 0.04 | 0.39 |
| NorBrid 1 | 20 | 0.61 | 0.02 | 0.39 |
| NorBrid 4 | 19 | 0.63 | 0.09 | 0.37 |
| NorBrid 8 | 17 | 0.60 | 0.03 | 0.40 |
| Total | 95 | 0.63 | 0.05 | 0.37 |

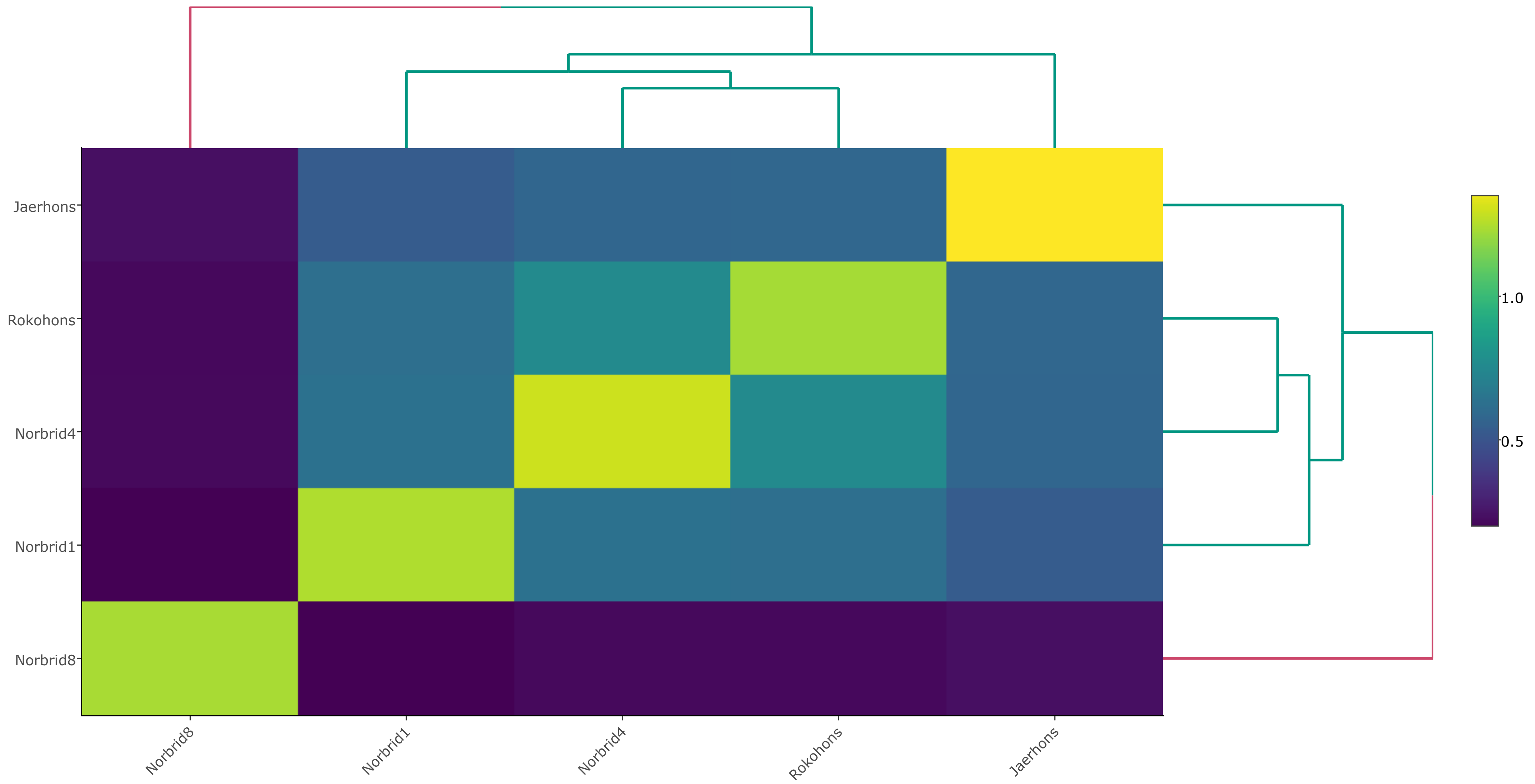



**Figure 3.1.** Graphic presentation of the matrix containing genomic relationships within and between lines in the Norwegian dataset. Color is dependent on the genomic relationship estimate, where a lighter color towards yellow reflects a higher relationship estimate.

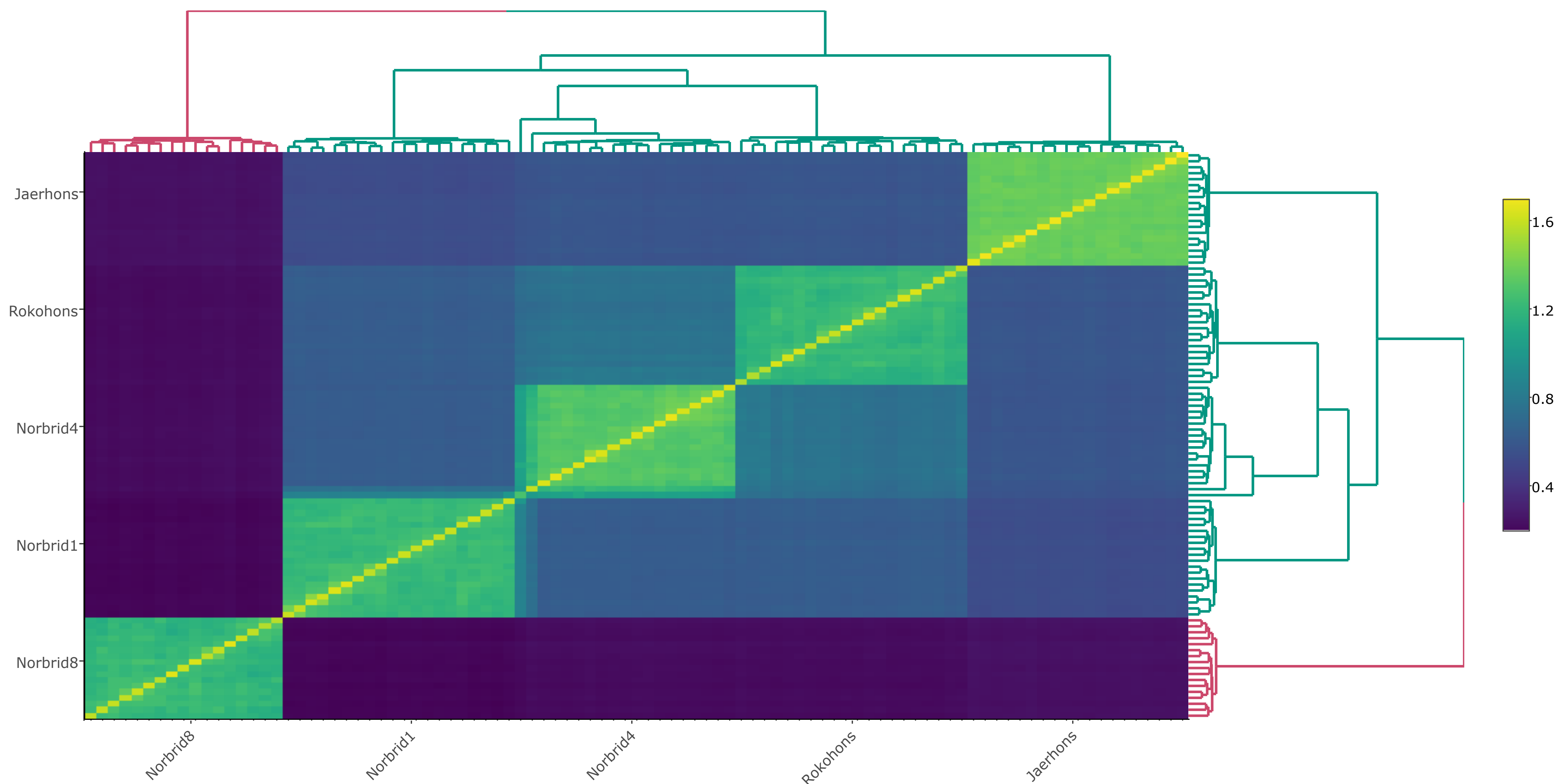



**Figure 3.2.** Graphic presentation of the matrix containing genomic relationships between individuals in the Norwegian dataset. The diagonal is the Inbreeding coefficient +1. Color is dependent on the genomic relationship estimate, where a lighter color towards yellow reflects a higher relationship estimate.

Figure 3.3 is a graphic presentation of the 75 x 75 G matrix containing average genomic relationships between and within lines in the international dataset. The matrix is symmetric and the diagonal represents the average genomic relationship between the individuals within the lines. The diagonal elements do not include the within individual genomic relationships. The ordering of the lines in the heat map follows the clustering in the dendrogram. The dendrogram shows two main clusters. The first main cluster includes only populations with Asian origin, including the two wild types *Gallus gallus gallus* and *Gallus gallus spadiceus*. The distinction of this cluster is also visible in the heat map with a darker square in the area representing the relationships between the lines in this cluster and the other lines. Within the first main cluster there are three clusters. The first (pink) cluster consists of only the NorBrid 8 line from the Norwegian dataset. NorBrid 8 is a line of the breed Red Rhode Island, which is of Asian origin. The second cluster (orange) consists of lines of Asian origin including the two wild types. The third cluster (olive green) consists of three lines of Asian origin, including two Asian bantam breeds. The Second main cluster consists of mostly European origin breeds, although in the first (green) cluster there are lines of both Asian and European origins. The next two clusters (turquoise and blue) consist of only European bantam lines. In the purple cluster all the lines are of European origin. Jærhøns is included in this cluster. The last cluster (violet) is also a purely European origin cluster and consists of 4 White Leghorn lines; NorBrid 1, NorBrid 4 and Rokohøns from the Norwegian dataset and the White Leghorn line R11 from the Synbreed dataset. The average genomic relationship estimates between lines in the international dataset ranges from 0.19 to 1.18. The lowest estimate is the genomic relationship between the two Norwegian lines NorBrid 8 and NorBrid 1. The highest estimate is between the two European bantam lines Sebright golden and Sebright silver from the Synbreed dataset. The row and column representing the relationship estimates between NorBrid 8 and the other lines is relatively dark across the heat map, reflecting lower relationship estimates. The average genomic relationship estimates within lines ranges from 0.69 to 1.62. The highest and lowest within line genomic relationship estimate is for the Sebright bantam silver and the Marans copper black respectively, both lines from the Synbreed dataset.

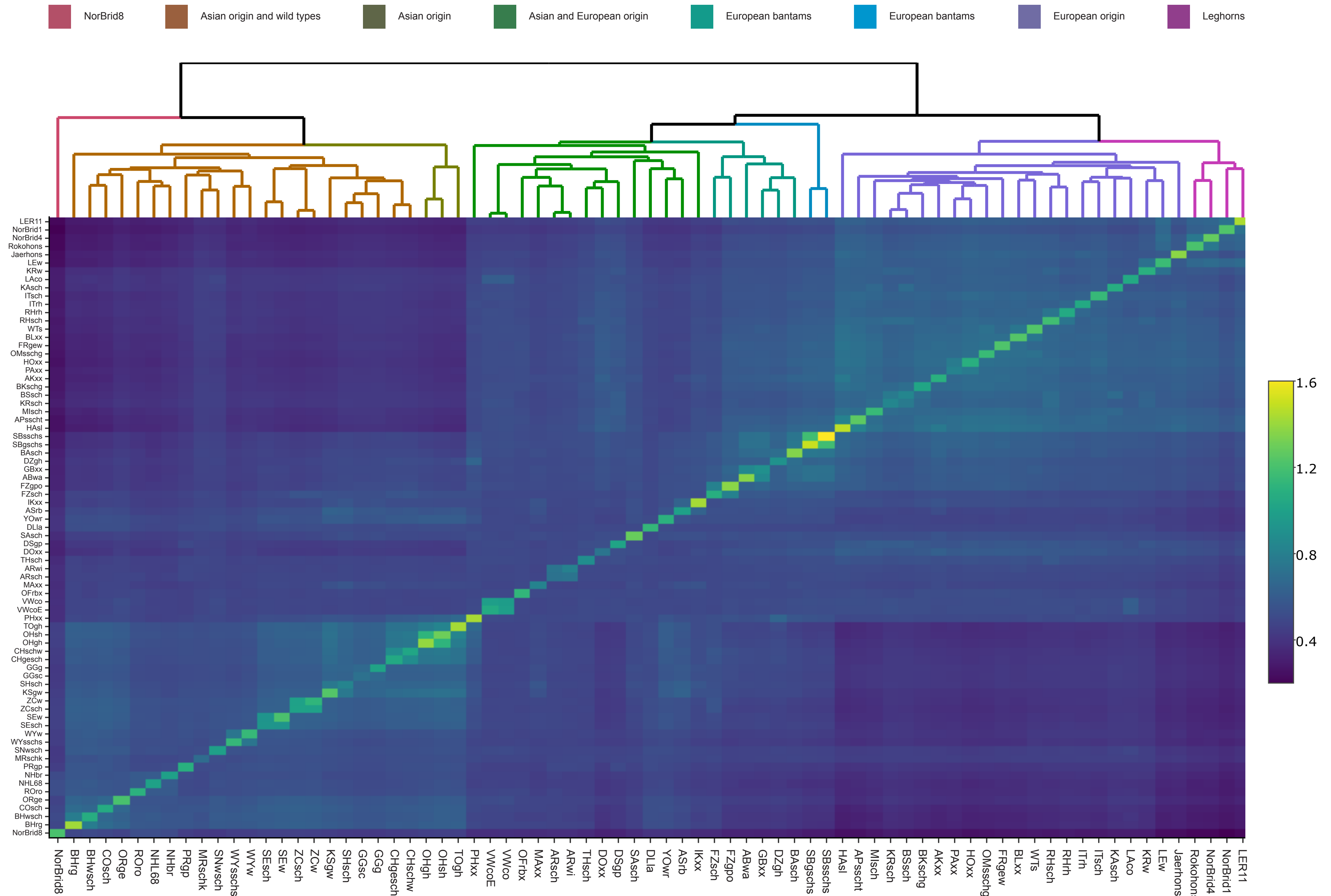



**Figure 3.3.** Graphic presentation of the matrix containing genomic relationships within and between lines in the International dataset. Color is dependent on the genomic relationship estimate, where a lighter color towards yellow reflects a higher relationship estimate.

## 3.2 Contributions to genetic diversity

The core set method was applied to the Norwegian dataset to measure relative loss of genetic diversity if a line is lost in a national perspective. This was done by comparing the diversity of the core set when all 5 lines are retained to the diversity of the core set constructed from the Norwegian set minus one line. The results can be found in Table 3.2. Genetic diversity lost is shown in percentage. The lines are ranked according to percentage genetic diversity lost. NorBrid 8 receives the highest conservation priority. The percentage diversity lost when NorBrid 8 is lost is considerably higher than the other four lines. NorBrid 4 receives the lowest priority for conservation of genetic diversity and NorBrid 1 receives priority 2, the highest priority of the white egg layers.

**Table 3.2.** Relative loss of genetic diversity when a line is lost from the Norwegian dataset and priority ranking for conservation of genetic diversity. $f(S)_{min}$ is the minimum relationship (8) and Div(S) is the genetic diversity calculated as 1- $f(S)_{min}$ (9). Losses are calculated relative to genetic diversity in the Full set.

| **Set(S)** | $\boldsymbol{f(S)_{min}}$ | **Div(S)** | **% Lost** | **Priority** |
|---|---|---|---|---|
| Norwegian | 0.5774 | 0.4226 | - | - |
| **Norwegian - 1** | | | | |
| Jærhøns lost | 0.6054 | 0.3946 | 6.63 | 3 |
| Rokohøns lost | 0.5936 | 0.4064 | 3.84 | 4 |
| NorBrid 1 lost | 0.6121 | 0.3879 | 8.22 | 2 |
| NorBrid 4 lost | 0.5895 | 0.4105 | 2.88 | 5 |
| NorBrid 8 lost | 0.7829 | 0.2171 | 48.64 | 1 |

The core set method was also applied to the International dataset to measure relative loss of genetic diversity in an international perspective if a Norwegian line is lost. The percentage lost is the difference between the diversity in the core set constructed from the international dataset, and the diversity in the core set when one Norwegian line is removed. NorBrid 8 comes out with the highest percentage diversity lost. The priority ranking for NorBrid 1, NorBrid 4 and NorBrid 8 is the same as in the ranking for national conservation value. The priority ranking between Jærhøns and Rokohøns is shifted in an international perspective, Jærhøns receives priority 4 and Rokohøns priority 3. The results can be found in Table 3.3.

**Table 3.3.** Relative loss of genetic diversity when a Norwegian line is lost from the International dataset and priority for conservation of maximum genetic diversity. $f(S)_{min}$ is the minimum relationship (8) and Div(S) is the genetic diversity (9). Losses are calculated relative to genetic diversity in the Full set.

| Set(S) | $f(S)_{min}$ | Div(S) | % lost | Priority |
|---|---|---|---|---|
| International | 0.4701 | 0.5299 | - | - |
| **International -1** | | | | |
| Jærhøns lost | 0.4713 | 0.5287 | 0.22 | 4 |
| Rokohøns lost | 0.4713 | 0.5287 | 0.23 | 3 |
| NorBrid 1 lost | 0.4751 | 0.5249 | 0.94 | 2 |
| NorBrid 4 lost | 0.4710 | 0.5290 | 0.17 | 5 |
| NorBrid 8 lost | 0.4851 | 0.5149 | 2.84 | 1 |

A third calculation was carried out using the core set method. The diversity in the core set constructed from only the lines in the Synbreed dataset was compared to the diversity in the core set if one Norwegian line was added. This was done in order to measure the gain in genetic diversity from each Norwegian line in an international context in isolation from the other Norwegian lines. In this calculation the relationships between the Norwegian lines will not affect the priority ranking. NorBrid 8 also receives the highest ranking in this context. The ranking between NorBrid 4 and Jærhøns is shifted so that Jærhøns receives the lowest priority for conservation of genetic diversity. These results are presented in table 3.4.

**Table 3.4**. Relative genetic diversity gained when a Norwegian line is added to the Synbreed dataset and priority for conservation of maximum genetic diversity. $f(S)_{min}$ is the minimum relationship (8) and Div(S) is the genetic diversity calculated as $1 - f(S)_{min}$ (9). Losses are calculated relative to genetic diversity in the Full set.

| Set(S) | $f(S)_{min}$ | Div(S) | % Gained | Priority |
|---|---|---|---|---|
| International | 0.4701 | 0.5299 | - | - |
| Synbreed | 0.4954 | 0.5046 | - | - |
| **Synbreed+1** | | | | |
| Jærhøns | 0.4941 | 0.5059 | 0.26 | 5 |
| Rokohøns | 0.4924 | 0.5076 | 0.59 | 3 |
| NorBrid 1 | 0.4890 | 0.5110 | 1.27 | 2 |
| NorBrid 4 | 0.4927 | 0.5073 | 0.54 | 4 |
| NorBrid 8 | 0.4810 | 0.5190 | 2.85 | 1 |

# 4. Discussion

The first section of the discussion focuses on the results of the analysis; Data reliability, the genomic relationship estimates and the measures of relative contributions to genetic diversity. In the second section perspectives and the possible applications of the results are discussed.

### 4.1 Data reliability

There is one individual from the NorBrid 4 line that stands out in the Norwegian dataset with the lowest inbreeding coefficient and a higher relationship with the NorBrid 1 line than the other NorBrid 4 individuals (Figure 3.2). This sample is the one that first failed a quality control in the genotyping process (see chapter 2.2). Possibly there was something wrong with the blood sample or genotyping process of this sample. There could also be biological reasons behind these results. For example, crossbreeding due to misplacement in the handling of the animals at the gene bank could have taken place.

As there are only two samples from each line in the comparable dataset, inbreeding coefficients may not be a good representation of the true inbreeding in this line. The inbreeding coefficients of these individuals are therefore not presented or discussed. In any analysis including these lines, the diagonal is not included when calculating average relationships within lines as it would count for 50% of the within line genomic relationships (Figure 3.3, Table 3.2 and Table 3.3).

It was not the aim to specifically analyze single lines in the Synbreed dataset, but rather to compare the Norwegian lines to this dataset as a whole to be able to analyze the Norwegian lines in an international context. Even though there is a higher uncertainty related to these lines than the Norwegian lines, one can expect the average relationships to be a good representation of the relationships between these lines. The dendrogram in Figure 3.3 indicates this, as the lines cluster well according to their origin. This is also an indication that the method used to cluster the genetic lines works well on genomic relationships.

## 4.2 Genomic relationships

When calculating genomic relationships, reference allele frequencies have to be set for each of the SNPs. Optimally; allele frequencies from a common ancestor are available. When this is not the case, one option is to use mean allele frequencies for the SNPs between the lines in the dataset. This option assumes that the base population is an average of the populations studied. White egg layers dominate the dataset in this study, and three out of five lines are lines of the White Leghorn breed. It would therefore be wrong to assume that these lines have drifted equally far from a common ancestor. The lines dominating the dataset would look less inbred as they would have drifted less from the mean allele frequencies. Another option is to use 0.5 as the reference allele frequencies for all the SNPs, which is what was chosen in this study. In this case the analysis is not biased by the distribution of lines in the study, and it provides a fair comparison between the lines. However, assuming that all allele frequencies have drifted from 0.5 may result in the lines looking more inbred as the allele frequencies were likely far from 0.5 in the common ancestor.

There is a clear distinction between the brown egg layer NorBrid 8 and the four White egg layers in the two figures presenting the genomic relationships between the Norwegian lines (Figure 3.1 and 3.2), this is to be expected as the line is the only brown egg layer and the only one of Asian origin. NorBrid 1, NorBrid 4 and Rokohøns all originated from the White leghorn breed from Livorno, Italy, which explains why they are relatively closely related. The Norwegian landrace, Jærhøns is slightly less related to the other white egg layers. Though they are all of European origin, their common origin is probably further back in time. The average inbreeding coefficient F in the Norwegian lines range from 0.60 to 0.68. This is relatively high if one considers that 5 generations of full sib mating would lead to an expected inbreeding coefficient F of 0.59. However, this may be affected by the reference allele frequencies being set to 0.5. Also, to decide if the level of inbreeding is sustainable in these lines one would have to look at the development of F over several generations. A study of the relationship status based on pedigree data has indicated that the conservation of existing genetic diversity has been effective in the Norwegian lines (Groeneveld et al, 2015).

Jærhøns shows the highest average inbreeding of the lines in the study, 0.68. The relatively high inbreeding in Jærhøns may be explained by the previous Norwegian landrace being almost extinct in the beginning of the 1900s.

Conversely, the diversity (Ho) in the Norwegian lines ranges from 0.32 to 0.40. This is similar to what was found in a microsatellite study on 5 local Swedish breeds, where observed heterozygosity ranged from 0.23 to 0.41 (Abebe et al, 2015). The brown egg layer NorBrid 8 shows the highest diversity estimate. Lower genetic diversity estimates in white egg layers than brown egg layers is in agreement with other studies (Hillel et al 2003, Weigend et al 2014, Groeneveld et al 2010). The observed heterozygosity in the Norwegian white egg layers range from 0.32 to 0.39. Observed heterozygosity in the brown egg layer was 0.40. This is higher than what was found in commercial white and brown egg layers genotyped with the same 600K SNP array, which was 0.15 and 0.23, respectively (Weigend et al 2014).

The lowest genomic relationship estimate between two lines in the international dataset is between two of the Norwegian lines, NorBrid 8 and NorBrid 1. And in the dendrogram resulting from the average genomic relationships between lines in the international dataset, the white egg layers NorBrid1, NorBrid4, Rokohøns and Jærhøns and the brown egg layer NorBrid 8 ends up in opposite ends of the dendrogram. This distinction between brown and white egg layers is in agreement with other studies that have clustered chicken breeds in neighbor-joining trees. Commercial white and brown egg layers form clear distinct clusters in opposite ends of the center, and other populations like fancy breeds and wild types are closer to the center (Weigend et al 2014, Rosenberg et al 2001). Even though Brown and white egg layers are bred with similar breeding goals and have similar production qualities, they are clearly genetically distinct.

### 4.3 Contributions to genetic diversity

In a national perspective, 48.64 % genetic diversity is lost if Norbrid 8 is lost. This value is much higher than for the other lines, which range from 2.88 to 8.22 %. The results may have looked different if there were more brown egg layers in the study. NorBrid 1, NorBrid 4 and Rokohøns are closely related. Therefore, losing one of these lines does not lead to much loss in diversity when the other two lines are still retained. One could expect Jærhøns to receive the highest priority out of the white egg layers, as it has the lowest relationship estimates with the other white egg layer lines. The lower diversity within Jærhøns may explain why it does not get the highest ranking out of the white egg layers. One could argue that the brown and white egg layers should be separated when making a priority between lines. It would also be natural that Jærhøns receive a higher priority due to the cultural value of the breed. One could use this analysis to prioritize between the three White Leghorn lines. In that case, NorBrid 1 would receive the highest priority and NorBrid 4 the lowest. If one was forced to conserve fewer lines, one could also argue that the three Leghorn lines could be merged into one line rather than only conserving NorBrid 1. This would conserve more genetic diversity than if two lines were lost.

In an international perspective, there are smaller differences in relative contribution to genetic diversity between the lines. Average diversity lost ranges from 0.22 to 2.84 % (Table 3.2). In the international dataset there are other White Leghorn lines. There are also other brown egg layers including one Red Rhode Island line. This explains why the diversity lost when NorBrid 8 is lost is much lower in an international context. NorBrid 8 still receives the highest priority for conservation of genetic diversity. This is also true when the lines are made independent of one another by looking at genetic diversity gained when one line is added to the Synbreed dataset (Table 3.3). The three White Leghorn lines appear more genetically unique in this context, but not by much. It is noteworthy, however, that NorBrid 8 still receives the highest priority. Jærhøns gets the lowest priority in this context. If one looks at Figure 3.3, these findings are in agreement with the dendrogram. NorBrid 8 forms its very own cluster, suggesting that relatively little genetic material is shared with any other line. Jærhøns is clustering with a larger group of lines with European origin. NorBrid 1, NorBrid 4 and Rokohøns form a cluster with only one other White Leghorn line. This suggests that

there will be more lines sharing genetic material with Jærhøns in the Synbreed dataset than with the three White Leghorns. As the number of lines is higher in the international dataset, the within line diversity will affect the results less than in the same analysis on the Norwegian dataset. The reason why Jærhøns receives a low priority is more likely because there are more lines of European origin sharing relatively recent ancestors with Jærhøns, than to the lower diversity within the line.

When the core set method was applied to the international dataset, the relative contributions to genetic diversity was calculated for all the lines in the international dataset. The results for the lines in the Synbreed dataset were not presented, as the aim was to make a prioritization between the Norwegian lines. It is worth mentioning, however, that the five Norwegian lines received five out of the six highest priorities for conservation of genetic diversity. The estimated relative diversity lost if a line is lost is lower for all the other lines in the dataset except one line, the Malay Black Red. These results can be found in appendix A. As there are only two individuals per line, there is a higher uncertainty related to the results for the Synbreed lines. However, as the analysis is based on average relationships, it should not affect the results directly that there is a difference in number of individuals. It would be natural to expect some lines to appear more unique than the Norwegian lines and some less unique. When looking at the heat map of the genomic relationships in Figure 3.3 it does suggest that the Norwegian lines are somewhat unique in the context of this dataset. NorBrid 8 forms a cluster all on its own. NorBrid 1, NorBrid 4 and Rokohøns form a cluster with only one other leghorn line. Jærhøns is in a bigger cluster, but in a separate cluster within that cluster. Most of the lines in the Synbreed dataset cluster with many other lines, it is natural that if you loose one from the dataset, there will not be much genetic diversity lost as there will be other lines closely related that encompasses some of the same genetic material. The comparable dataset used in this study consist mainly of fancy breeds or small, local populations. It does not include any commercial lines. Although including a wide variety of breeds from various origins, it may not reflect the true global poultry genetic diversity. It would have been of interest to include active commercial lines of white and brown egg layers as well as broilers. Also, an addition of more Nordic populations that may share more recent common ancestors with the Norwegian lines would give a more complete picture of the true genetic uniqueness of these five lines.

### 4.4 perspectives

Even though the international dataset consists of many type of breeds with a large variety in size, color and patterns, the genomic relationship between all lines is quite high (Table 3.2). A variety in color and other visible phenotypes is often considered a sign of diversity, but this may not always be a good indicator of the genomic diversity. A recent study on genetic diversity in Nordic type goats found that breeds with the lowest diversity estimates had a large variety in coat color and color pattern and discuss the importance of separating conservation of phenotypic diversity and molecular diversity (Lenstra et al., 2016). Previously, conservation of diversity was based on phenotype and pedigree information (Woolliams & Oldenbroek 2007). Poultry landraces are often more colorful and patterned than the commercial chicken breeds. Caution has to be taken when assuming that these landraces are more genetically diverse. Conversely, if there is not a direct link between a variety in color and genomic diversity, this also suggests that maintaining genomic diversity may not ensure retention of diversity in visual phenotypes, if this is desired.

This study looks at mean genetic diversity across the genome. It may also be of interest to study diversity in specific areas of the genome and compare this to commercial breeds. A study on the Finnish landrace found a substantial diversity in a cluster of genes involved in immune responses (Fulton et al. 2017). One could also look at regions around QTLs associated with traits that have undergone strong selection. For example the major QTL in a study by Kerje et al (2003) found to explain a large part of the difference in body weight and egg weight between the Red Jungle Fowl and the White Leghorn breed. However, one could also argue that conservation decisions should in fact be based on diversity across the genome, as one of the arguments for conserving genetic diversity often is that we don't know what traits may be important or desirable in the future.

The priority between lines for conservation in this study is purely based on genetic diversity. When making decisions on which lines to prioritize for conservation, there are other parameters that may be equally important. John Ruane (2000) suggested the following key criteria to include in decision making when prioritizing between breeds for conservation on a national level; degree of endangerment, presence of traits of current economic value, presence of traits of current scientific value, agro ecological value in a special landscape, cultural-historical value as well as genetic uniqueness. The results in this study give a first insight on these lines genetic uniqueness in a national as well as international context. Jærhøns has a historical value, and although not showing an overall genetic uniqueness, may have traits or alleles that are unique to the breed.

## 5. Conclusion

This study gives first insights into the genomic relationships within and between the Norwegian lines. The three White Leghorn lines are highly related. The Norwegian landrace shows the highest level of inbreeding and is relatively closely related to the other white egg layers. The brown egg layer, NorBrid 8, shows the highest level of observed heterozygosity. There is a clear genetic distinction between the white and brown egg layers in the study.

NorBrid 8 receives the highest ranking for conservation of genetic diversity both in a national and international perspective. Losses in genetic diversity ranged from 0.17 to 2.84 % when one of the Norwegian lines were lost relative to the entire set of 75 lines. Although these losses in genetic diversity are not large, the estimates are higher than for all but one of the lines in comparison in this study.

In further analysis of the relative genetic uniqueness of the Norwegian lines, it would be of interest to include active commercial lines of brown and white egg layers.

## References

Abebe A. S., Mikko S., Johansson A. M. (2015) Genetic Diversitet of Five Local Swedish Chicken Breeds Detected by Microsatellite Markers. PLoS ONE 10(4): e0120580. doi:10.1371/journal. pone.0120580

Affymetrix (2017). Affymetrix analysis guide: Axiom® genotyping solution data analysis guide. https://biobank.ctsu.ox.ac.uk/crystal/docs/axiom_geno_analguide.pdf

Brenoe, U. T. & Kolstad, K. (2001) Food Efficiency in Pure Strains and Hybrids of Laying Hens Analysed in a Diallel Cross: III. Direct, Heterosis and Maternal Effects on Food Consumption and Adjusted Food Consumption, Acta Agriculturae Scandinavica, Section A — Animal Science, 51:2, 129-133, DOI: 10.1080/090647001750193468

Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4.

Eding H., Crooijmans R.P.M.A., Groenen M.A.M., Meuwissen T.H.E., Assessing the contribution of breeds to genetic diversity in conservation schemes, Genet. Sel. Evol. 34 (2002) 613–634.

Fulton, J. E., Berres, M. E., Kantanen, J. & Honkatukia, M. 2017. In press. MHC-*B* Variability in the Finnish Landrace Chicken Conservation Program. Poultry Science. (accepted 19.3.2017)

Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., Ohno, S. (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(13), 6792–6795.

Galili, T. 2016. Heatmaply: Interactive Heat Maps Using 'plotly'. R package version 0.6.0. https://CRAN.R-project.org/package=heatmaply

Galili, T. 2015; Dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* 2015; 31 (22): 3718-3720. doi: 10.1093/bioinformatics/btv428

Genressursutvalget for husdyr. (2002). Bevaring a genetiske ressurser på fjørfe I Norge. Norway.

Gura, S. 2007: *Livestock Genetics Companies. Concentration and proprietary strategies of an emerging power in the global food economy*. League for Pastoral Peoples and Endogenous Livestock Development, Ober-Ramstadt, Germany

Graffelman J, Moreno V. 2013. The mid p-value in exact tests for Hardy-Weinberg equilibrium. Statistical Applications in Genetics and Molecular Biology, 12.

Groeneveld L.F., Müller, U., Groeneveld, E., Sæther,N. & Berg, P. 2015. Inbreeding in native Norwegian poultry breeds, with partially unknown maternal pedigree. 66th Annual Meeting of the European Federation of Animal Science. 31.08 – 04.09, 2015. Warsaw, Poland.

Hillel J., Groenen, M.A.M.,Tixier-Boichard, M., Korol, A. B., David, L., Kirzhner V., Burke T., Barre-Dirie A., Crooijmans R.P.M.A., Elo K., Feldman M.W., Freidlin P.J., Maki-Tanila A., Oortwijn M., Thomson P., Vignal A., Wimmers K., Weigend S., Biodiversity of 52 chicken populations assessed by microsatel- lite typing of DNA pools, Genet. Sel. Evol. 35 (2003) 533–557.

Kerje, S., Carlborg, Ö., Jacobsson, L., Schütz, K., Hartmann, C., Jensen, P., & Andersson, L., 2003. The twofold difference in adult size between the red junglefowl and White leghorn chickens is largely explained by a limited number of QTLs. International Society for Animal Genetics, Animal Genetics, 34

Kranis, A., Gheyas, A.A., Boschiero, C., Turner, F., Yu, L., Smith, S., Talbot, R., Pirani, A., Brew, F., Kaiser, P., Hocking, P.M., Fife, M., Salomon, N., Fulton, J., Strom, T.M., Haberer, G., Weigend, S., Preisingen, R., Gholami, M., Qanbari, S., Simianer, H., Watson, K.A., Wooliams, J.A. and Burt, D.W. (2013) Development of a high density 600K SNP genotyping array for chicken, *BMC Genomics* 14: 59.Published online Jan 28, 2013. doi: 10.1186/1471-2164-14-59

Meuwissen, T.H.E. (2015) Gghat 3 [computer software] Retrieved from http://wiki.nmbu.org/index.php/Gghat_version_3

Oldenbroek, K. 2017. Genomic management of animal genetic diversity. Wageningen Academic Publishers.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rosenberg. N. A, Burke, T., Elo, K., Feldman, M. W., Freidlin, P. J., Groenen, M. A. M., Hillel, J., Mäki-Tanila, A., Tixier-Boichard, M., Vignal, A., Wimmers, K., & Weigend, S. 2001. Empirical Evaluation of Genetic Clustering Methods Using Multilocus Genotypes From 20 Chicken Breeds. GENETICS *October 1, 2001 vol. 159 no. 2 699-713*

Szekeres, B., Schönherz, A. A., Nielsen, V. H., Guldbrandtsen, B. 2016. Gamle danske husdyrracers genomer. DCA report nr. 082, August 2016. Aarhus universitet, Denmark

VanRaden, P.M. (2008). Efficient methods to compute genomic predictions. Journal of Dairy Science 91: 4414-4423.

Weigend S, Janßen-Tapken U, Erbe M, Baulain U, Weigend A, Sölkner J, Simianer H (2014) Genome-wide analyses of genetic diversity and phylogenetic relationships in the Synbreed Chicken Diversity Panel. Proceedings XIV$^{th}$ European Poultry Conference, Stavanger, Norway 23.-27. June, S. 164-176

Wigginton JE, Cutler DJ, Abecasis GR (2005) A note on exact tests of Hardy-Weinberg equilibrium. American Journal of Human Genetics, 76.

Wooliams, J.A., Oldenbroek, K. (2017). Chapter 1. Genetic diversity issues in animal populations in the genomic era. In Oldenbroek, K. (Eds.), Genomic management of animal genetic diversity (pp. 13 – 45). Wageningen Academic Publishers, Wageningen, the Netherlands.

Fernández, J., Bennewitz, J. (2017) Chapter 2. Defining genetic diversity based on genomic tools. In Oldenbroek, K. (Eds.), Genomic management of animal genetic diversity (pp. 49 – 72). Wageningen Academic Publishers, Wageningen, the Netherlands.

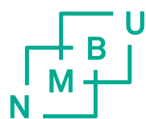
# Appendix A

**Table A.1** Relative loss of genetic diversity when a line is lost from the International dataset. $f(S)_{min}$ is the minimum relationship (8) and Div(S) is the genetic diversity (9). Losses are calculated relative to genetic diversity in the full international set.

| Set (S) | f(S) min | Div(S) | % Lost |
|---|---|---|---|
| International | 0,470105 | 0,529895 | |
| **International - 1** | | | |
| VWcoE | 0,470108 | 0,529892 | 0,000555 |
| ARsch | 0,470187 | 0,529814 | 0,015355 |
| AKxx | 0,470106 | 0,529894 | 0,000233 |
| BAsch | 0,470105 | 0,529895 | 0,000000 |
| BHrg | 0,470173 | 0,529827 | 0,012847 |
| BHwsch | 0,470298 | 0,529702 | 0,036313 |
| BLxx | 0,470106 | 0,529895 | 0,000000 |
| CHgesch | 0,470106 | 0,529894 | 0,000000 |
| CHschw | 0,470112 | 0,529888 | 0,001255 |
| COsch | 0,470244 | 0,529756 | 0,026151 |
| DLla | 0,470111 | 0,529889 | 0,001041 |
| DSgp | 0,470164 | 0,529836 | 0,011168 |
| DZgh | 0,470109 | 0,529891 | 0,000772 |
| DOxx | 0,470473 | 0,529527 | 0,069445 |
| FZgpo | 0,470105 | 0,529895 | 0,000000 |
| FZsch | 0,470165 | 0,529835 | 0,011334 |
| FRgew | 0,470105 | 0,529895 | 0,000000 |
| HOxx | 0,470106 | 0,529894 | 0,000115 |
| ITrh | 0,470114 | 0,529886 | 0,001685 |
| ITsch | 0,470106 | 0,529894 | 0,000174 |
| KAsch | 0,470143 | 0,529857 | 0,007108 |
| KSgw | 0,470109 | 0,529891 | 0,000663 |
| KRw | 0,470106 | 0,529894 | 0,000000 |
| LAco | 0,470181 | 0,529819 | 0,014353 |
| LEw | 0,470346 | 0,529654 | 0,045378 |
| MAxx | 0,471099 | 0,528901 | 0,187634 |
| MRschk | 0,470399 | 0,529601 | 0,055377 |
| MIsch | 0,470105 | 0,529895 | 0,000000 |
| NHbr | 0,470623 | 0,529377 | 0,097698 |
| OHgh | 0,470112 | 0,529888 | 0,001344 |
| ORge | 0,470123 | 0,529877 | 0,003321 |
| OMsschg | 0,470113 | 0,529887 | 0,001450 |
| PAxx | 0,470132 | 0,529868 | 0,005077 |
| PHxx | 0,470105 | 0,529895 | 0,000000 |
| PRgp | 0,470667 | 0,529333 | 0,106071 |
| RHrh | 0,470128 | 0,529872 | 0,004353 |
| RHsch | 0,470105 | 0,529895 | 0,000000 |
| SBgschs | 0,470105 | 0,529895 | 0,000000 |
| SBsschs | 0,470107 | 0,529893 | 0,000263 |

**Table A.1** continued

| Set (S) | f(S)min | Div(S) | %Lost |
|---|---:|---:|---:|
| SEw | 0,470118 | 0,529882 | 0,002399 |
| SAsch | 0,470107 | 0,529893 | 0,000270 |
| SNwsch | 0,470113 | 0,529888 | 0,001377 |
| VWco | 0,470105 | 0,529895 | 0,000000 |
| WTs | 0,470105 | 0,529895 | 0,000000 |
| WYw | 0,470111 | 0,529889 | 0,001099 |
| YOwr | 0,470105 | 0,529895 | 0,000000 |
| ZCsch | 0,470188 | 0,529812 | 0,015666 |
| ZCw | 0,470105 | 0,529895 | 0,000000 |
| TOgh | 0,470113 | 0,529887 | 0,001505 |
| OFrbx | 0,470145 | 0,529855 | 0,007455 |
| ROro | 0,470611 | 0,529389 | 0,095513 |
| SHsch | 0,470105 | 0,529895 | 0,000000 |
| ARwi | 0,470105 | 0,529895 | 0,000000 |
| ASrb | 0,470107 | 0,529894 | 0,000240 |
| WYsschs | 0,470622 | 0,529378 | 0,097528 |
| IKxx | 0,470115 | 0,529886 | 0,001755 |
| OHsh | 0,470105 | 0,529895 | 0,000000 |
| ABwa | 0,470106 | 0,529894 | 0,000139 |
| APsscht | 0,470105 | 0,529895 | 0,000000 |
| KRsch | 0,470105 | 0,529895 | 0,000000 |
| GBxx | 0,470105 | 0,529895 | 0,000000 |
| BKschg | 0,470112 | 0,529888 | 0,001270 |
| HAsl | 0,470105 | 0,529895 | 0,000000 |
| THsch | 0,470118 | 0,529882 | 0,002425 |
| BSsch | 0,470105 | 0,529895 | 0,000000 |
| GGsc | 0,470406 | 0,529595 | 0,056669 |
| GGg | 0,470113 | 0,529888 | 0,001376 |
| LER11 | 0,470819 | 0,529181 | 0,134686 |
| NHL68 | 0,470807 | 0,529193 | 0,132436 |
| NorBrid8 | 0,485133 | 0,514867 | 2,835938 |
| Jaerhons | 0,471282 | 0,528718 | 0,222147 |
| NorBrid4 | 0,471000 | 0,529000 | 0,168950 |
| Rokohons | 0,471320 | 0,528680 | 0,229286 |
| NorBrid1 | 0,475112 | 0,524888 | 0,944887 |




**Norges miljø- og biovitenskapelig universitet**
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway