Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science

Single nucleotide polymorphisms in northeast European wolves *(Canis lupus)*

Federico Javier Fröstl

Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics,
385
Uppsala 2012

Master's Thesis, 30 HEC

Erasmus Mundus Programme – European Master in Animal Breeding and Genetics

Swedish University of Agricultural Sciences
Faculty of Veterinary Medicine and Animal Science
Department of Animal Breeding and Genetics

Single nucleotide polymorphisms in northeast European wolves *(Canis lupus)*

Federico Javier Fröstl

Supervisors:
Tomas Bergström, SLU, Department of Animal Breeding and Genetics
Göran Andersson, SLU, Department of Animal Breeding and Genetics
Mikael Åkesson, SLU, Department of Ecology
Patrick Waldmann, BOKU, Austria
Examiner:
Sofia Mikko, SLU, Department of Animal Breeding and Genetics

Credits: 30 HEC
Course title: Degree project in Animal Science
Course code: EX0556
Programme: Erasmus Mundus Programme - European Master in Animal Breeding and Genetics
Level: Advanced, A2E

Place of publication: Uppsala
Year of publication: 2012
Name of series: Examensarbete / Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, 385
On-line publication: http://epsilon.slu.se

Key words: PLINK, runs of homozygosity, heterozygosity, allele contribution, *Canis lupus*

Single nucleotide polymorphisms in northeast European wolves (*Canis lupus*)

Federico Javier Fröstl
830920P914

THESIS ANIMAL BREEDING AND GENETICS (COURSE CODE)
June 2012

Department of Animal Breeding and Genetics

SUPERVISORS
Tomas Bergström
Göran Andersson
Mikael Åkesson

ABSTRACT

In the 1960s, the gray wolf was extinct as a breeding population in Scandinavia. However, in the 1980s a pair was established and started to reproduce in central Sweden. A third wolf, a male, immigrated to Sweden in 1991 contributing to the genetic variation in Scandinavia. Currently the Scandinavian wolf population consists of more than 280 individuals. The population is thus a typical example of a population that has undergone a severe bottleneck. To define the degree of genetic variation in the Scandinavian wolf population thought regions of homozygosity (ROH) and to evaluate possible contribution of immigrant individuals to increase genetic variation on the Swedish population, twenty-three wolves have been genotyped using the 170k canine-specific single nucleotide polymorphism (SNP) array. SNP data was analyzed with PLINK [1] and R [2] software's. In the estimation of ROH, the individual coverage of Swedish wolves was 4.1% larger compared with immigrant wolves. In the individual heterozygosity estimation, 1.6 % more heterozygous SNPs were found in the Swedish population. From the genetic contribution of immigrants, an increase of 17% of heterozygous SNPs was found among fixed (non-variable) SNPs after the addition of 7 simulated immigrants to the Swedish population. Because of the lack of genetic variation in the Swedish population, new allelic variation could be added through the addition of new immigrants. Nevertheless, the small general heterozygosity difference found among immigrants and the Swedish population suggests that even immigrants wolves lack of substantial genotypic variation.

Keywords: PLINK, runs of homozygosity, heterozygosity, allele contribution, *Canis lupus*

ABBREVIATIONS

SNP: Single Nucleotide Polymorphism
DNA: Deoxyribonucleic acid
Kb: Kilobases
bp: base pair
mtDNA: Mitochondrial Deoxyribonucleic Acid
CNV: Copy Number Variation
GWAS: Genome-Wide Association Studies
MAF: Minor Allele Frequency
IBS: Identical by State
IBD: Identical by Descent
KBAVG: Kilo Base Average
ROH: Runs of Homozygosity
SD: Standard Deviation

CONTENTS

1 - Background 4
2 - Literature Review 5
2.1 Genomic similarity between dogs and wolves. 5
2.2 170 K canine-specific SNP Array 6
2.3 Software 6
3 Materials and methods 6
3.1 The studied wolves 6
3.2 Summary statistics and filtering 8
3.3 Runs of homozygosity 8
3.4 Individual degree of heterozygosity 8
3.5 “Rare” alleles contribution 9
4 Results 9
4.1 Summary statistics and filtering 9
4.2 Runs of homozygosity and comparison between immigrant and Swedish population 10
4.3 Individual heterozygosity 10
4.4 “Rare” allele contribution 11
4.4.1 Immigrant allele contribution 11
4.4.2 Sequence of immigration events 12
5 Discussion 13
5.1 Summary statistics and filtering 13
5.2 Runs of homozygosity 13
5.3 Individual degree of heterozygosity 14
5.4 “Rare” allele contribution 14
6 Conclusion 17
Acknowledgments 17
7 References 18
8 Appendix 19

1 - BACKGROUND

The Scandinavian wolf (*Canis lupus*) population was declared extinct in the 1960s due to extensive hunting [3]. The combination of maternal (mtDNA), paternal (Y chromosome microsatellite), and biparentally inherited genetic markers (autosomal and X chromosome microsatellites) showed that the population had been founded by one male and one female from Finland and Russia [4] emigrated in 1983 [4]. In 1991 one male joined the population and two additional male immigrants successfully bred in Scandinavia in 2008 to 2010. Few immigration events and several generations of low effective population size (Ne < 50) have resulted in average inbreeding levels above 0.25, corresponding to the offspring of full sibs with unrelated parents [4]. As a result, severe inbreeding depression has been observed [4] that may increase the risk of extinction of the population [5]. Even a single immigrant can genetically improve populations that have experienced severe inbreeding. The male that immigrated in 1991 alleviated the inbreeding depression temporarily, causing a quite rapid expansion of the population [6].

Genetic variability often is quantified by gene diversity, by the number of distinct alleles per locus or by the percentage of loci that are polymorphic [7]. Endangered species have small and/or declining populations, so inbreeding and loss of genetic variability are unavoidable in them. Loss of genetic variability reduces the ability of populations to evolve to cope with environmental change, increasing extinction risk [8]. Small populations are prone to loss genetic variability expressed as nucleotide diversity by genetic drift. Moreover this variation carried by each individual may also be reduced in small populations as a consequence of the breeding between closely related individuals, which in turn may lead to inbreeding depression, *i.e.* reduced individual fitness due to inbreeding [9]. Inbreeding depression often results in decreased fertility and increased risk of developing disease. Inbreeding has therefore become a key objective for conservation genetics to monitor genetic variation [10].

Population bottlenecks are defined as temporary but significant reductions of population size [11]. The effects can vary depending on both the size to which the population is reduced and the duration of the bottleneck (*i.e.*, the number of generations). Loss of genetic variation in small endangered species seriously threatens their abilities to persist and recover [12]. Several recent introductions into populations with low fitness appear to have shown genetic restoration of fitness to levels similar to that before the effects of genetic drift [9].

Until 1991, when the third immigrant male wolf was established in the Swedish wolf population, there was only one reproducing pack of wolves, resulting in strong inbreeding and loss of heterozygosity [4]. Following the establishment of this new male wolf, the population heterozygosity increased as did both the number of wolves and breeding packs, suggesting the importance of this immigrant to the successful expansion of the species in Scandinavia [6]. Nevertheless, inbreeding levels continued to grow after some time, correlated with low reproductive success and reduced litter sizes from more related parents, suggesting that the population is still at risk to experiencing inbreeding depression in the absence of additional genetic input from new immigrants [4].

The availability of high density SNPs arrays provides the opportunity to scan the genome for runs of homozygosity (ROH) (13). A ROH is essentially a continuous segment of DNA sequence without heterozygosity in the diploid state (13) and can be used as a potential way to study inbreeding (14).

The main hypothesis and objectives of this study are:

- There is a difference in lengths of ROH between Swedish and immigrant wolves. It is believed that Swedish individuals will present larger regions of the genome with higher similarity degree due to lower recombination rates compared with less related individuals. We believe the Swedish wolves show higher number of larger ROHs due to their higher degree of homozygosity. The objective will be to compare and identify the ROH and their lengths between Swedish and immigrants wolves and therefore to estimate the degree of genetic variation (expressed as homozygosity degree) of both groups.
- The Swedish individuals will have a different degree of heterozygosity compared with immigrant wolves. A mating between closer related individual's increases the frequency of homozygous combinations of deleterious recessive alleles due to the increased chance of offspring inheriting alleles identical by decent from both heterozygous parents [15]. It is predicted that the Swedish wolves will have lower genetic variability expressed as a lower amount of heterozygous SNPs compared with an immigrant group. To address this hypothesis, the average amount of individual heterozygous SNPs of the Swedish wolves will be defined and compared with immigrant individuals.
- Immigration of immigrant wolfs and subsequent breeding with Swedish wolves could increase heterozygosity within the Swedish wolf population. It is predicted that the breeding with individuals of an immigrant population could potentially improve the heterozygosity degree of the Swedish individuals through the addition of heterozygous SNPs. The objective will be to evaluate the possible contribution and increase to genetic variation (heterozygosity) from potential immigrants to Swedish individuals. The result will help to analyze how the mating of dissimilar genetic individuals could vary the heterozygosity levels of the studied Swedish population.

The expected outcome of this study will be a thorough characterization of the degree of genetic variation in the current Swedish wolf population. Consequently, it could help to inform the authorities concerning conservation biology and management of the Swedish wolf population.

2 - LITERATURE REVIEW

2.1 Genomic similarity between dogs and wolves

The gray wolf [17] *Canis lupus* is the largest extant member of the dog family of mammals, the *Canidae*. The species has become extinct in most of Western Europe, in Central America and North America [23]. Genetic studies reaffirm that the gray wolf is the ancestor of the domestic dog; sequences from both dogs and wolves supported the hypothesis that wolves were the ancestors of dogs [17].

Phylogenetic analyses of mitochondrial DNA (mtDNA) and archaeological evidence from south-west Asia indicates that the domestic dog *Canis lupus familiaris* was originated from the domestication of the wild gray wolf approximately 5400 and 16300 before present [6] [18] [19]. The canine larger genetic variation in East Asia compared with other regions and the pattern of phylogeographic variation, suggest an East Asian origin of the domestic dog [20]. From and evolutionary point of view, this divergence is very recent, and the genome sequence of dogs and wolves is almost identical [18], *i.e.* 78 chromosomes: 38 pairs of autosomes and two sex chromosomes [21].

The genetic distance for mitochondrial DNA in dogs and Eurasian wolves confirmed that wolves are the exclusive ancestral species of dogs [22]. Gray wolves and dogs are most closely related (4% & 21% sequence divergence in nuclear exon and intron sequences, respectively), followed by a close affiliation with coyotes, golden jackals and ethiopian wolves [22]. The reason for this conclusion is that the mayor diversity of mtDNA was found here, framed within the comparison of all the surveyed dog populations[18].

2.2 170 K canine-specific SNP Array

The canineHD Genotyping BeadChip contains more than 170,000 markers placed on the CanFam2.0 reference sequence. The CanineHD BeadChip enables the interrogation of genetic variation in any domestic dog breed presenting an average of 70 markers per megabase (Mb), data that provides a SNP density sample **for robust within-breed association and copy number variation (CNV) studies**. The proprietary technology allows unconstrained locus selection and a high-throughput format that provides a practical solution for whole-genome studies about the domestic dog [23].
Due to the genomic similarity between dogs and wolves, the majority of SNPs in the Canine HD array are also polymorphic in gray wolf and ethiopian wolves, what evidence the potential usage of the array in other canids [26].

2.3 Software

PLINK is a free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner. The focus of PLINK is mainly on analysis of genotype/phenotype data [1]. With PLINK, large data sets comprising hundreds of thousands of markers genotyped for thousands of individuals can be rapidly analyzed in their entirety. PLINK provides tools to make analyses computationally efficient, supporting novel approaches to whole-genome data analysis that take advantage of whole-genome coverage [1]. PLINK was used for quality control of data, filtering and running of homozygosity calculation.
R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical (linear and nonlinear modeling, classical statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible [2]. A script to calculate allele individual's heterozygosity contribution degree was developed using GenABEL-package (an R library developed to facilitate Genome-Wide Association (GWA) analysis of binary and quantitative traits [24]).

3 MATERIALS AND METHODS

3.1 The studied wolves

A summary of all wolves used in this study can be seen in Table 1. The samples were taken from sedated or dead animals, collected during the Skandulv project (http://skandulv.nina.no/) and they were provided by Mikael Åkesson, who extracted the DNA samples from blood, tissue and faeces.

Table 1.The sampled wolves included in this study along with gender, type of sample, year of death and comments. Two main groups can be distinguished: Swedish individuals (Nyskoga couple offspring and Gillhov couple offspring) and immigrants

Identity	Sample number	Gender	Type of sample	Year of death	Comment
I1	5097-77	M	Tissue	1977	Immigrant, no surviving offspring
I2	5007-79	F	Tissue	1979	Immigrant, no surviving offspring
I3	D-05-18	M	Tissue	2005	Immigrant, non-reproducing
I4	LU07-236	M	Blood	2008	Immigrant, non-reproducing
I5	5126-86	F	Tissue	1986	Immigrant, Alpha-female Nyskoga 1
I6	LU09-069	M	Blood	Alive 2011	Immigrant, Alpha-male Galven
I7	GR10-077	M	Tissue	Alive 2010	Immigrant, Alpgha-male Kynna 2
I8	GR11-047	F	Blood	Alive 2011	Immigrant 2010
I9	215	M	Tissue	2003	Immigrant
I10	SEP0011739	F	Feaces	Alive 2011	Immigrant 2010
OG1	LU08-115	M	Tissue	1992	Offspring Gillhov couple
OG2	1307/04	F	Tissue	2004	Offspring Gillhov couple
OG3	5015-93	M	Tissue	1993	Offspring Gillhov couple
OG4	5002-96	M	Tissue	1996	Offspring Gillhov couple
OG5	9803	F	Blood	?	Offspring Gillhov couple
OG6	18703	M	Tissue	1992	Offspring Gillhov couple
OG7	5016-99	M	Tissue	1999	Offspring Gillhov couple
OG8	LU-08-114	M	Tissue	1992	Offspring Gillhov couple
ON1	5131-84	M	Tissue	1984	Offspring Nyskoga 1 couple
ON2	5131-86	M	Tissue	1986	Offspring Nyskoga 1 couple
ON3	5216-86	M	Tissue	1986	Offspring Nyskoga 1 couple
ON4	16-8905	F	Tissue	?	Offspring Nyskoga 2 couple
S1	SFT13087	M	Tissue	1997	Shot in Finland
S2	SF230	M	Tissue	1998	Shot in Finland

The pedigree of the individuals used in this study is shown in Figure 1. In total, 24 wolf samples were used. Those samples where taken between 1977 and 2011 from wolves of Swedish and Finnish origin, and the selection of the samples obeyed to the purpose of capturing the total population variation. This sampling was possible because only a few individuals (three wolves and later two more immigrants) were the founding fathers the current Swedish wolf population.

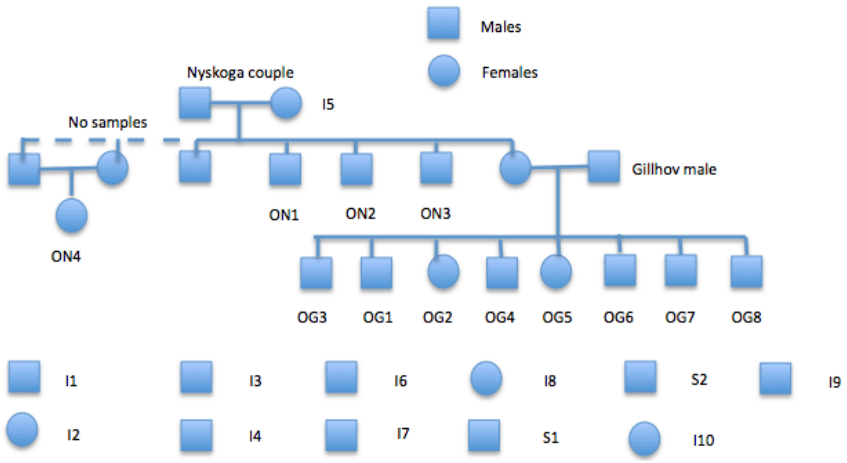

Figure 1. Pedigree of the wolves used for this study. As it can be observed, 3 main groups are distinguished: Gillhov couple offspring, Nyskoga couple offspring and immigrants. Note that some individuals have not been sampled and are thus not accompanied with identity.

3.2 Quality control and filtering

A quality control and filtering of the data was done using PLINK.
First, a quality control analysis was performed to define missing genotypes (which exposes missingness by individual and by SNP) and minor allele frequency (frequency at which the less common allele occurs in a given population for each SNP [25]). For missingness and minor allele frequency calculation (MAF), a threshold of 25% was used.
Secondly, a filter of individuals with too much missing genotype data (with more than 25% missing genotype) was done, followed by an exclusion of SNPs on the basis of minor allele frequency (MAF) set at 25% as well.
A detail of the command used in PLINK for quality control and filtering can be observed in the appendix (list of commands 1 in the appendix).

3.3 Runs of homozygosity

A ROH is a long continuous stretch of DNA sequence without heterozygosity in the diploid state that can help to estimate the degree of genetic variation of the individual [13].
The ROHs were calculated using PLINK for two groups of individuals: Swedish and immigrant wolves. The algorithm takes a window of a number of adjacent SNPs and slides them across the genome, at each window position it determinates whether this window is "homozygous" or not, and then -for each SNP- it calculates the proportion of "homozygous" windows that overlap that position. Finally, call segments based on overlapping event. The homozygous segment criteria were (max. thresholds):

- Length (kb): 1000 (definition of the sliding window)
- Number of adjacent SNPs (N): 100
- Density (kb/SNP): 50
- Largest gap (kb): 1000

The command used for the analysis of both groups is detailed in the appendix (list of commands 2).
After all ROHs for each individual were obtained, an "individual coverage" was calculated, what reveals, how much of the individual genome was covered by ROH. In order to do this, every individuals ROHs sizes (in kb) was compared with the total kb size of the dog genome (2,294,902 kb). The formula used was: Individual coverage = homozygous regions kb sum/ dog genome size.

3.4 Individual degree of heterozygosity

The objective of the section was to analyze the number of heterozygous SNPs of each individual to define the heterozygosity average degree of the Swedish and immigrant individuals.
Under that purpose, a script was developed in R by Mats Persson (as shown with detail at the appendix, list of commands 3). The script considered and compared the 2 alleles of each non-missing SNPs: if a difference appeared, the SNP was considered as heterozygous. Finally, the

script accounted a final number corresponding to those SNPs which alleles differed (heterozygous SNP) and SNPs which alleles were equal (homozygous SNP).
After the heterozygous SNPs were counted for each individual, an average was calculated.

3.5 “Rare” alleles contribution

The main objective of this section was the analysis of the potential contribution of rare new alleles from immigrants to the Swedish wolf group. It is easily inferable that the contribution of SNPs with new different alleles will generate a higher genotypic variation among the Swedish individuals as a primarily outcome of an increased number of new heterozygous SNPs.
For the analysis of the “rare allele contribution” another script was developed in R by Mats Persson (as described in the appendix, list of commands 4). The script allowed to calculate the number of added SNPs with different new alleles of each one of the seven immigrants subjected to analysis.
The next step was to simulate a sequence of immigration events trough a random “input” of immigrant individuals at the Swedish population (simulating breeding) in order to analyze the variable SNPs addition for fixed locus positions. Once an immigrant was compared with the Swedish population in the search of input of new variable SNPs, the individual was thereafter considered as part of the Swedish population (reference population) before the next immigrant was added (by doing that, the new genomic variation was taken into account as part of the reference population). The additions were repeated five times at different random input orders. Finally, the sequences of heterozygous SNPs addition were compared and analyzed.

4 RESULTS

4.1 Summary statistics and filtering

Table A in the appendix shows the result of the missingness set to 25% for each individual. A 9% proportion of missing SNPs was found (range 16-63.0%) among the non-filtered individuals. Furthermore, an average of 17.5% missingness was found among immigrants, 4.6% among offspring of Gillhov couple and 2.4% among offspring of Nyskoga couple. Separating the data between tissue and blood samples, the first group presented an average missingness of 6.5% and, the second group, of 6.3%. Moreover, three individuals with more than 25% missing data were excluded: sample I7 with a missingness of 32% (tissue sample), sample I9 with a value of 34% (tissue sample) and sample I10 (feaces sample) with 63% missingness. The three samples expressed an average missingness of 43%.
The SNPs of the remaining individuals were filtered with a 25% missing data threshold, which excluded 6027 SNPs failed missingness test (GENO>0.25) reducing the amount of markers (after filtering) from 173662 to 168383.

4.2 Runs of homozygosity and comparison between immigrants and Swedish population

PLINK was used to calculate the ROH of each individual of the Swedish and immigrant groups. Based on Table B of the appendix, figure 2 helps to distinguish the differences between both groups.

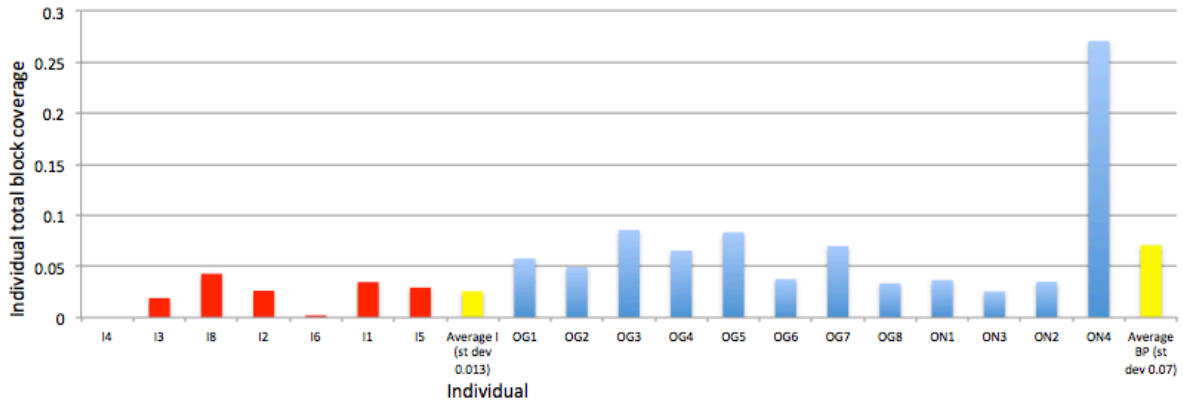

Figure 2. ROH coverage of immigrant individuals (red bars), Swedish individuals (blue bars) and the average of observations for each group (yellow bars) with standard deviations (SD).

Figure 3 shows an average total ROH coverage of 3% for immigrants (with a SD of 1.3%) and 7.1% for the Swedish individuals (with a SD of 1.3% for the Gillhov group, and 1.6% for the Nyskoga group excluding ON4), this means that the Swedish group total ROH average coverage was bigger compared with the immigrants, with the exception of individual ON4 (offspring Nyskoga 2 couple), which shows evident higher homozygosity compared with both groups. This higher homozygosity was confirmed from pedigree reconstruction analysis, which exposed probable high level of inbreeding as well (offspring from full sibs born of Nyskoga 1).
With the exclusion of individual ON4 (the only known inbreed individual, with very high coverage reaching values over 25%) the Swedish population total ROH average coverage decreased from 7,1% to 5,2%, still a higher value compared with the immigrants (3%).
A within-family difference was found between the higher coverage of offspring of Gillhov couple (6%) and the lower coverage of the offspring of Nyskoga couple (9.2% including ON4 and 3.2% excluding ON4).
There was no data from individual I4 (immigrant, non-reproducing), probably related to the amount of missing data.

4.3 Individual heterozygosity

The percentage of heterozygous SNPs was calculated for each individual using an R script developed by Mats Persson.
Figure 3 shows the percentages of heterozygosity for both Swedish and immigrant wolves, calculated form the number of SNPs that worked for each individual.

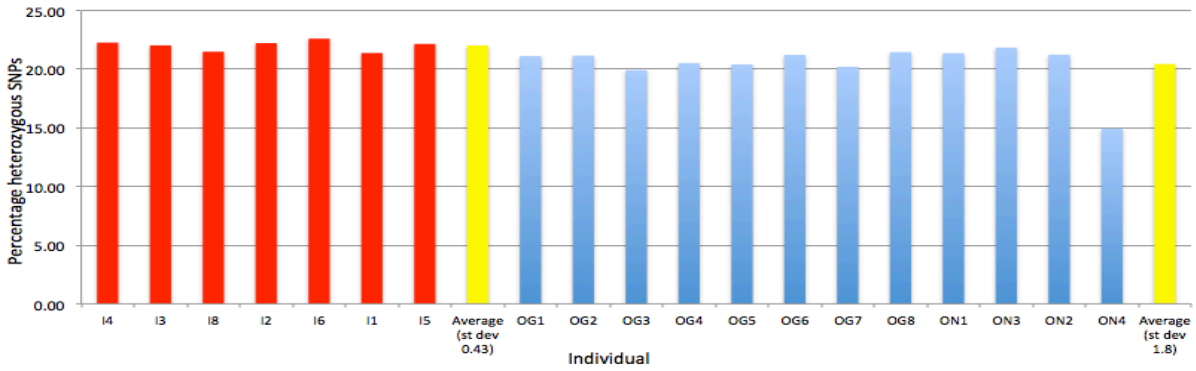

Figure 3. Percentage of heterozygous SNPs (SNPs with different alleles) of immigrant (red bars) and Swedish (blue bars) individuals. Yellow bars represent an average of the values and SD for each group.

A higher percentage of heterozygosity was observed among the immigrants compared with the Swedish group. The difference was bigger compared with individual ON4 (a deeply inbreed wolf).
On average, immigrants showed 22% heterozygous SNPs (with SD of 0.43) whereas Swedish individuals only 20.4% (with a SD of 1.8), which means: immigrants have around 1.6% more heterozygous SNPs.
In case that the ON4 individual was omitted from the Swedish group, the average heterozygosity percentage increased from 20.4% to 20.9% (with a SD of 0.59) which still is a lower value compared with the one of the immigrant group.

4.4 "Rare" allele contribution

4.4.1 Immigrant allele contribution

Using the Swedish wolves as reference population, the heterozygous SNPs contribution of each immigrant was calculated using R. Table 2 shows the new allele contribution and the number of missing SNPs for fixed SNPs to the reference population for each one of the immigrants.

Table 2. "New" SNP contribution to the reference population (offspring of founding wolves) per immigrant with averages.

Identity	Number SNPs with new allele	Number missing SNPs
I4	10244	13609
I2	8983	436
I6	9449	15242
I1	7448	450
I3	8666	2212
I8	9085	1468
Average	8979	5570

The immigrants showed an average of 8979 heterozygous SNPs (with a SD of 924 SNPs). It can be observed a high contribution of individual I4 (10244 SNPs) followed by I6 (9449 SNPs), I8 (9085 SNPs), I2 (8983 SNPs), I3 (8666 SNPs), I1 (7448 SNPs) and finally I5 (108 SNPs). Both

individuals I4 (immigrant non-reproducing) and I6 (immigrant alpha-male Galven) present the higher values of missing SNPs data compared with the rest of the individuals.

4.4.2 Sequence of immigration events

To calculate the added number of new heterozygous SNPs from all immigrants to a reference population in a sequence of immigration, an R script was used, where five different immigration events were simulated with a random input of the six immigrants on the reference population.

Table 3. Number of "new" heterozygous SNPs added to the reference population for 5 random events. The six inputs represent the six immigrants "added" progressively to the reference population.

Random event	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6
Event 1	10248	16479	19895	22937	25538	**28222**
Event 2	8670	15239	18886	22495	24514	**27198**
Event 3	8987	15546	18962	22259	25422	**27330**
Event 4	9453	14943	19636	23627	26785	**28905**
Event 5	7452	13888	18064	21613	24163	**26201**
Average	8962	15219	19088.6	22586.2	25284.4	**27571**

As Table 3 shows, some events add a higher amount of final SNPs compared with other ones, in this case for example; event 4 inputs the higher amount of final SNPs.
The script exposed that of 168.383 SNPs, 88.228 SNPs were fixed, which means, nearly the half (52.39%) showed no allele variation among individuals of the Swedish population. Nevertheless, with the addition of immigrants, an average of 27.553 heterozygous SNPs were added (with a standard deviation of 1034 SNPs) among the positions where no allele variation was originally found, arising therefore a supposed decrease of 31.3% on the amount of fixed SNP positions. Consequently, after the addition of all immigrants, the number of heterozygous SNPs in the Swedish group changed from 80363 to 107916, what corresponds to an increase of approximately 17% in the number of heterozygous SNPs.

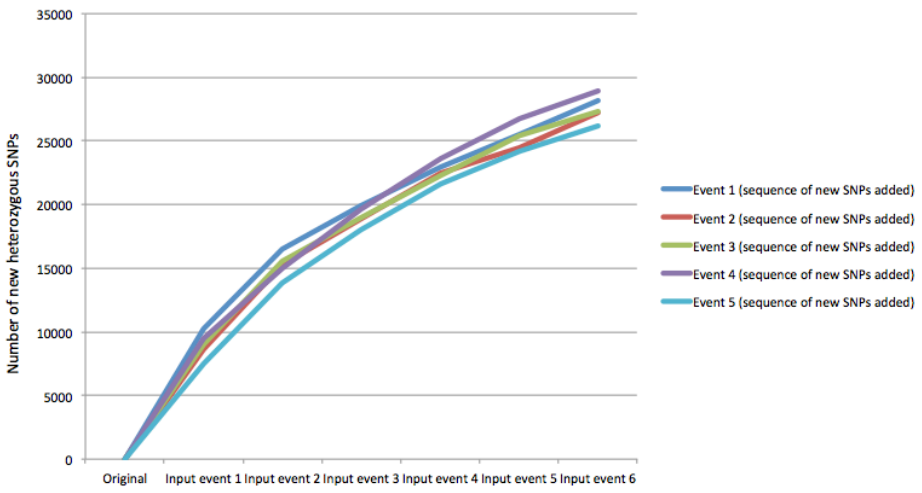

Figure 4. Sequence of heterozygous SNPs addition.

The curve in Figure 4 shows the response to the addition of new heterozygous SNPs with the addition of immigrants into the reference population. The amount of new heterozygous SNPs added (compared with the previous generation) decrease every time a new immigrant joins the reference population.

5 DISCUSSION

In this study, the ROH of a Swedish and immigrant group were compared. The amount of heterozygosity of both groups was calculated and a script was created to analyze the potential increase of heterozygous SNPs when immigrant individuals are bred with Swedish wolves.

5.1 Summary statistics and filtering

We perform a quality control of the SNP data based missingness values. Furthermore, we filter the data based on MAF values and missing genotype rate.
The SNP calling results were as follows: average missingness of 9% was found among all samples and the amount of SNPs left after the filtering was 96.9%. The results demonstrated that the genome-wide canine SNP Array showed a high performance (taking into account the average tagging capability for dog breeds founded by previous studies [26]). The missingness is probably not related with the use of a canine SNP array with wolf samples. Past investigations have demonstrated that the majority of SNPs (>65%) in the Illumina array were also found to be polymorphic in the Gray wolf and Ethiopian wolf samples, indicating the potential usage of such arrays in other canids [26]. The amount of missing data could be due to a poor sample quality, probably related to the DNA origin or extraction method where problems are often encountered in terms of relatively low DNA yields and/or recovering DNA free of inhibitory substances [27].
It seems there is no difference between the missingness between blood (6.3%) and tissue (6.5%) samples. Nevertheless, of the three excluded individuals, two came from tissue sample and one from fecal origin (S9, with missingness of 63%). The use of feces as an alternative source of DNA is becoming increasingly popular; however DNA quality and quantity often are compromised when using fecal material as an alternative source of DNA [28].

5.2 Runs of homozygosity

We calculate the ROH coverage using PLINK software over a group of Swedish and immigrant wolves and then proceeded to compare both group results.
It was observed that the ROH coverage of the immigrant group (3%) was in general smaller than the Swedish wolves average coverage (7,1%), which confirms a larger extent of homozygosity among the Swedish wolf population.
The Swedish wolf population experienced a severe bottleneck numerous generations ago. Following data, we observed a higher percentage of short segments (1Mb) of ROH in the Swedish population comparing with an immigrant population. This pattern was also observed in previous studies on cattle and humans studies which demonstrated that the presence of a high amount of

ROH segments of 1MB or more could help to indicate the presence of old inbreeding that cannot be traced using pedigree data [13] [14].
The Gillhov offspring group compared the Nyskoga offspring group bigger average ROH coverage's (6% vs. 3.2% excluding ON4) suggest a higher genetic distance between the Nyskoga parents compared with the Gillhov parents, which means Gillhov parents are more genetic related than Nyskoga parents. Full sibs show a high amount of genome region with complete identity, meaning that the genome variation among full siblings is lower compared with non-related individuals [29] where more recombination events are more likely to happen.
There were differences observed between the standard deviation values between immigrant and Swedish wolves as well as between (and within) Swedish families. The differences in standard deviations between groups could be related with the differences in sample sizes. To avoid bias, it is recommended to increase the number of samples.

5.3 Individual degree of heterozygosity

We created an R script to estimate the amount of heterozygous SNPs per individual, calculated a heterozygosity SNP percentage and, afterwards, compared it between the Swedish and immigrant group.
On average, Swedish wolves had around a 1.64% more heterozygous SNPs comparing with the Swedish population, which confirms the lower degree of genetic variation among this population.
From the results of the analysis the initial hypothesis was confirmed: a higher degree of heterozygosity was observed among immigrant wolves compared with Swedish wolves (with a substantial bigger difference compared with individual ON4). As it is known, the mating between closer and related individuals increases the chance of homozygosity at each locus, what is associated with an increase in the level of homozygosity [30].
The lower degree of heterozygosity of the Swedish population could be related with the effect of inbreeding. Studies made by Rumbal *et.al.* [31], report the effect of inbreeding in heterozygosity and reproductive success due to the mating of close related individuals in *D. Melanogaster*.

5.4 "Rare" allele contribution

We developed a script in R to calculate the heterozygous allele contribution of each immigrant to a Swedish reference population. In addition, we perform an immigration simulation to a reference population to calculate the total amount of new heterozygous SNPs that the immigrants could add to the Swedish group.
After all immigrants were added to the reference population, it was observed an average increase of 17% of heterozygous SNPs among fixed SNPs. This value helps to understand the initial assumption: the use of immigrant wolves with a higher genotypic variation shows how the mate of genetically dissimilar individuals can increase heterozygosity of the Swedish wolves.
Each immigrant showed a similar contribution of heterozygous SNPs to the Swedish group, but a deeper analysis on the number of missing SNPs per individual revealed that individual I4 (immigrant non-reproducing) and I6 (immigrant alpha-male Galven) presented an evidently higher amount of missing SNPs (although they input the higher amounts of heterozygous SNPs). This amount of missingness could lead to aunderestimation of the real number of potential heterozygous SNPs that they could contribute to the Swedish group. Individuals I4 and I6 could

be taken as an interesting example to analyze the consequences of mating dissimilar individuals if the objective would be to evaluate the increase of amount of heterozygous SNPs on fixed SNPs in the Swedish population. However, there is an uncertainty on the level of trust for these two samples, which could have a high rate of missing SNPs due to contamination or poor quality of the samples [32]. It is interesting to remark that I4 and I6 come from blood samples, suggesting a possible incidence on the type of sample in the missingness rate. Nevertheless, no significant difference between missingness of blood (6,3%) and tissue (6,5%) samples were found, which suggest that the origin of the sample is not related with such missingness. A further analysis must take place to evaluate possible scenarios on influence of final amount of missing SNPs, probably related to DNA quality [32].

The heterozygous SNPs calculation proposed in this study can help to select individuals based on heterozygous SNPs "contribution", an interesting approach for genetic rescue, in which the introduction of unrelated individuals into an inbreed population results in the reduction of detrimental genetic effects and reveals itself as a potential important management tool in the quest to mitigate adverse effects over small populations [6]. The increase of heterozygosity related with increase on population viability was already reported in several species, as the lesser kestrel (*Falco naumanni*) [33] or the panther (*Puma concolor coryi*) [34] for example. Nevertheless, a special emphasis and caution should be considered when new genetic-distanced immigrants are introduced into small populations: the genetic contribution may not be always positive. The carrying of a recessive disease could have a negative impact on the original population if such immigrant is successfully reproducing, with the consequence of an increase on the frequency of the disadvantageous allele. Low genetic variability cannot always have a negative impact on the fitness of small populations, recent studies performed by Ellegreen *et al.* showed that the severe bottlenecked Swedish beaver population (*Castor fiber*) presents low levels of genetic variation and, notwithstanding, the population reached to an effective population size (approx. 1880), thanks to the preservation of the natural habitat and the plastic adaptation in the cultivated landscape [35]. The moose (*Alces alces*) can be taken as another interesting example (as shown by Hundermark *et al.)*. Despite they historically suffer serial population bottlenecks; moose have exhibited notable ability to adapt to a changing environment, indicating that limited neutral genetic variation may not necessarily indicate limited adaptive genetic variation [36]. Studies made by Vissner P. *et.al*. in Chillingham cattle, also demonstrates the viability of the isolated breed even though the presence of an almost total genetic uniformity [37].

The five immigration events showed a dynamic representation of the input of heterozygous SNPs for fixed SNPs to the Swedish group. As expected, as each immigrant is added to the reference population, the number of heterozygous SNPs increased. Nevertheless, when more heterozygous SNPs were "added" from a new immigrant, the amount of fixed SNPs in the reference population decreased but the SNP "improvement" was lower if compared with the previous addition (Fig. 4). The curve exposed the tendency of "new" SNPs addition for fixed positions in the reference population, following a type II functional response of heterozygous SNPs addition -characterized by a decelerating response rate when new variation was added [38]-. Variations are in this case the outcome of different new combinations of already existing genetic information of the species, that do not add any new characteristic to the genetic information. The saturation is produced because this variation occurs in the limits of genetic information, *i.e.* the gene pool of the population. Variations will only produce changes that remain within the boundaries of the genetic information of the species [39].

A generic form of the type II curve is

$$f(R) = \frac{aR}{1 + ahR}$$

Where in this case f(R) represents the number of SNPs, R the number of new immigrants, and “a” and “h” are parameters of this type of functional response curve.

To predict the partial saturation of new heterozygous SNPs added, and to obtain the estimates of the parameters of the curve (“a” and “h”) with nonlinear least squares, the observed values of Table 3 (number of immigrants and the SNPs count) and the statistical software Stata 9.1 were used. The underlying representative curve that fits all variation of observations was obtained with estimates of a=10989.14 and h=0.0000213, both statistically significant at 1% level. With these parameters it was possible to predict the potential SNPs input of any number of immigrant wolves to the Swedish reference population. As it is showed in Table C in the appendix, the first immigrant introduced 8905 new SNPs, while the 10^{th} would introduce 1059, and the 200^{th} only 5 SNPs. This model shows strong evidence of genetic saturation related to heterozygosity improvement after the addition of an important number of potential new immigrants. In an ideal population, genotype frequencies predicted by the Hardy-Weinberg equilibrium estimate that heterozygotes are most frequent when the frequency of the two alleles is equal to 0.5 [26]. As the Swedish wolf population is a finite population, isolated and founded only by three individuals, the heterozygosity that could be added from new immigrants would decrease as the population growth, thus the probability to input new rare alleles will decrease with time.

6 CONCLUSION

This study has identified differences on the ROH between Swedish and immigrant individuals. These regions were 2.2% bigger among the Swedish wolf population compared with immigrants, what confirms a lower degree of genetic variation of the Swedish population. In conclusion, the higher amount of short ROH among Swedish individuals provide a good indication of individual inbreeding and the presence of a past population bottleneck.

The study demonstrates that the Swedish wolf population has a lower heterozygosity degree compared with immigrant individuals. The observations highlights a difference of 1.6% more heterozygous SNPs among immigrants. Interestingly, the immigrant population showed a low heterozygosity difference compared with the Swedish population, pointing out probable problems on genetic diversity in that population as well.

The immigrant individuals used in this study could potentially improve the heterozygosis percentage of the Swedish founding wolves' offspring in a 17%. This is important to analyze the effect of mating more genetic dissimilar individuals in order to increase the allele heterozygosity, helping therefore to decrease the negative effects of homozygosity in small populations.

Based on the present findings, we believe that this study is essential to demonstrate how the mating of more dissimilar wolves can improve Swedish wolf population heterozygosity degree, a fundamental rate vital as management tool for conservation genetics. Nevertheless, the management strategies for the Swedish wolf population must consider all reasonable criterions to improve and to maintain an appropriate fitness, analyzing for the process the advantages and disadvantages of the introduction of new individuals.

ACKNOWLEDGMENTS

I would like to thank my supervisors Göran Andersson and Tomas Bergström for the constant support, patience and guidance trough all the project. I would like also to thank specially Mikael Åkesson for providing me invaluable feedback.

7 REFERENCES

1- Plink. *Plink. Whole genome association analysis toolset.* 2009 10/10/2009 [cited 2012 23/2]; v1.07:Available from: http://pngu.mgh.harvard.edu/~purcell/plink/contact.shtml.
2- *"The R Project for Statistical Computing."* [cited 2012 5/4] Available from: http://www.r-project.org/.
3- Wabakken P, S.H., Liberg O, Bjärvall *The recovery, distribution, and population dynamics of wolves on the Scandinavian peninsula, 1978-98.* Canadian Journal of Zoology, 2001. **79**: p. 710-725.
4- Liberg O, A.H., Pedersen HC et al. , Severe inbreeding depression in a wild wolf (Canis lupus) population. . Biology Letters 2005. 1: p. 17-20.
5- Newman D., P., D. , *Increased probability of extinction due to decreased genetic effective population size: experimental populations of Clarkia pulchella.* . Evolution, 1997. 51: p. 354-362.
6- Vila C., S.A.K., Flagstad O., Seddon J., Bjornerfeldt S., Kojola I., Casulli A., Sand H., Wabakken P., Ellegren H., *Rescue of a severely bottlenecked wolf (Canis lupus) population by a single immigrant*. Proc R Soc B, 2003. 270(1510): p. 91-97.
7- Lacy R. *Importance of genetic variation to the viability of mammalian populations*. Journal of mammalogy, 1997. 78(2):320-335
8- Frankham R. *Genetics and conservation biology*. Academie des sciences. 2003.
9- Keller LF, W.D. and Waller, D.M., *Inbreeding effects in wild populations*. Trends Ecol Evol 2002. 17: p. 230-241
10- Answers. *Genetic variation*. 2012 [cited 17/6]; Available from: http://wiki.answers.com/Q/What_is_genetic_variation
11- Peery M. Z., et.al. *Reliability of genetic bottleneck tests for detecting recent population declines*. Molecular ecology. 2012.
12- Allendorf, F.W., *Genetic drift and the loss of alleles versus heterozygosity*. Zoo Biology, 1986. 5: p. 181-190.
13- Hamzic, E., *Level of Inbreeding Derived from Runs of Homozygosity: A comparison of Austrian and Norwegian Cattle Breeds*. June 2012.
14- Ferencakovic M. Et.al., *Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle.* Home. 2011. Vol 76, No4.
15- Heterozygosity, Fitness and Inbreeding Depression in Natural Populations. 2006. [cited 2/7]; Available from: http://www.studymode.com/essays/Heterozygosity-Fitness-Inbreeding-Depression-Natural-Populations-99810.html
16- Kohts, A.E., *The Variation of Colour in the Common Wolf and its Hybrids with Domestic Dog*. Journal of zoology, 1948. 117(4): p. 784-790.
17- Pulliainen E. 1964. *Studies on the wolves (Canis Lupus L.) in Finland*. Annales Zoologici Fennici, 2, 215–259.
18- Pang J.F., K.C., Zou X.J., Zhang A.B., Luo L.Y., Angleby H., Ardalan A., Ekstrom C., Skollermo A., Lundeberg J., Matsumura S., Leitner T., Zhang Y.P., Savolainen P, mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. Mol. Biol. Evol., 2009. 26: p. 2849-2864.
19- Slovanien P. et.al, *Comprehensive study of mtDNA among Southwest Asian dogs contradicts independent domestication of wolf, but implies dog-wolf hybridization*. Ecol Evol. 2011. 373-385.
20- Savolainen P, Zhang YP, Luo J, Lundeberg J, Leitner T., *Genetic evidence for an East Asian origin of domestic dogs.* Science, 2002. 1610-3.
21- Switonski, M., Reimann, N., Bosma, A.A., Long, S., Bartnitzke, S., Pienkowska, A., Moreno-Milan, M.M., Fischer, P, *Report on the progress of standardization of the G-banded canine (Canis familiaris) karyotype. Committee for the Standardized Karyotype of the Dog (Canis familiaris).* Chromosom. , 1996. Res. 4: p. 306-309.
22- Kerstin Lindblad-Toh, C.M.W., Tarjei S. Mikkelsen, et al, *Genome sequence, comparative analysis and haplotype structure of the domestic dog*. Nature, 2005. 438: p. 803-819.
23- Illumina. *CanineHD BeadChip*. 2010 [cited 2010 23/2]; Available from: http://www.illumina.com/documents//products/datasheets/datasheet_caninehd.pdf.
24- Yurii Aulchenki, *GenABEL tutorial*. 2011.
25- Single Nucleotide Polymorphism. *DNA baser. SNP definition 6* [cited 21/7]; Available from: http://www.dnabaser.com/articles/SNP/SNP-single-nucleotide-polymorphism.html
26- Hamilton M. B. *Population Genetics.* Wiley-Blackwell. 2009.
27- Yu Z, Morrison M. *Improved extraction of PCR-quality community DNA from digesta and fecal samples.* Biotechniques, 2004, 36: p. 808-12.
28- Fike J., et. al. *Genotyping error rates associated with alternative sources of DNA for the north American river otter*. IUCN Otter Spec. 2004. 21 A.
29- Information on genes. *Half-siblings vs. full-siblings visualized.* 2011. [cited 18/6]; Available from: http://ongenes.blogspot.se/2011/02/half-siblings-vs-full-siblings.html
30- Christenten K. *Population Genetics*. Division of Animal Genetics.
31- Heber S. Et.al., *A test of the genetic Rescue Techniche using bottlenecked donor populations of Drosophila Melanogaster.* Plos One. 2012.
32- Adams JR, Vucetich LM, Hedrick PW, Peterson RO, Vucetich JA. *Genomic sweep and potential genetic rescue during limiting environmental conditions in an isolated wolf population.* Proc Biol Sci, 2011. 278: 3336-44.
33- Ortego J. Et.al., *Inreease of heterozygosity in a growing population of lesser kestrels.* Biol. Lett. 2007 vol. 3 no. 5 585-588

34- Pimm S. L. Et.al., *The Genetic rescue of the Florida panther*. Animal Conservation. 2005.
35- Nolet B., Rosell F., Comeback of the beaver *Castor fiber*: An overwiev of old and new conservation problems.
36- Hundermark K., Bowyer T. Genetics, evolution and Phylogeography of Moose. 2004. Alces. 40: 103-122.
37- Visscher P.M. et. Al., *A viable herd of genetically uniform cattle.* Nature. 2001. 409, 303.
38- Prasanna B.M. Indian Agricultural Research Institute. Introduction to population Genetics. 2007. Pag 3.
39- Pulliainen E. 1964. Studies on the wolves (Canis Lupus L.) in Finland. Annales Zoologici Fennici, 2, 215–259.

8 APPENDIX

List of commands 1. PLINK commands used for quality control and filtering of data

Allele frequency command:

./plink --ped wolf.ped --map wolf.map --allow-no-sex --noweb --dog –freq

Gender check command:
./plink --ped wolf.ped --map wolf.map --allow-no-sex --noweb --dog --check-sex
Filtering based on missing rate per individual:
./plink --ped wolf.ped --map wolf.map --allow-no-sex --noweb --dog --mind 0.25
Filtering based on Allele frequency:
./plink --ped wolf.ped --map wolf.map --allow-no-sex --noweb --dog --mind 0.25 --maf 0.00
Filtering based on missing rate per SNP:
./plink --ped wolf.ped --map wolf.map --allow-no-sex --noweb --dog --mind 0.25 --maf 0.00 --geno 0.25.

List of commands 2. PLINK commands to calculate runs of homozygosity
For immigrants
./plink --ped wolf.ped --map wolf.map --allow-no-sex --noweb --dog --mind 0.25 --maf 0.06 --geno 0.25 --homozyg --keep Immigrants.txt
For breeding population
./plink --ped wolf.ped --map wolf.map --allow-no-sex --noweb --dog --mind 0.25 --maf 0.04 --geno 0.25 --homozyg --keep breeding-population.txt

List of commands 3. R script for individual degree of heterozygosity. The script worked as follows:
Take the 2 alleles of each non-missing SNPs and compared them. If a difference appeared, the SNPs is considered as heterozygous.

```
candidate_reshaped <- data.frame(dummy = rep(NA, dim(data_ed)[1]/2))
candidate_reshaped[,1:2] <- NA
names(candidate_reshaped) <- rep(candidate, each = 2)
candidate_reshaped[,1] <- data_ed[odd_row, candidate]
candidate_reshaped[,2] <- data_ed[odd_row+1, candidate]
candidate_reshaped[,"mismatch1"] <- 0
candidate_reshaped[,"mismatch2"] <- 0
candidate_reshaped[candidate_reshaped[,1] != data_reshaped[,"major"],"mismatch1"] <- 1
candidate_reshaped[candidate_reshaped[,2] != data_reshaped[,"major"],"mismatch2"] <- 1
candidate_reshaped[,"mismatch_sum"] <- candidate_reshaped[,"mismatch1"] +
candidate_reshaped[,"mismatch2"]
sum(candidate_reshaped[,1] != candidate_reshaped[,2])
sum(candidate_reshaped[,1] == candidate_reshaped[,2] & candidate_reshaped[,1] != 0)
```

List of commands 4. R script for rare allele contribution. The script worked as follows:

a) Rearrange SNP data to columns (one column per individual).

b) Input of the individuals of the reference set (offspring of founding population) and candidate (immigrant to compare).

c) Fixed SNPs analysis: For the analysis of the contribution of new rare alleles only common fixed SNPs positions (alleles read by pairs) were considered for the analysis, this means, SNPs that presented same alleles for all the individuals of the reference set (which gave an idea of the degree of homozygosity).

d) Candidate comparison calculations: Only fixed SNPs positions of the reference population was used for comparison with the immigrant individual.

e) Results: The script allowed to calculate the number of total fixed SNPs of the reference population, the number of input SNPs of the immigrant with new different alleles for fixed SNPs in the reference population, the number of input SNPs with 1 new different allele of the immigrant for fixed SNPs in the reference population, the number of input SNPs with

2 new alleles of the immigrant for fixed SNPs in the reference population and the number of missing SNPs of the immigrant (candidate).

```
#Imput data
ref_set <- c("")
candidate <- ""
# Data reading block
data <- read.delim(sep = " ", header = F, file="plink.ped", stringsAsFactors = F)
data_ed <- as.data.frame(t(data[,7:dim(data)[2]]), stringsAsFactors =F)
names(data_ed) <- data[,2]
#Reference set calculations
data_reshaped <- data.frame(dummy = rep(NA, dim(data_ed)[1]/2))
data_reshaped[,1:(length(ref_set)*2)] <- NA
names(data_reshaped) <- rep(ref_set, each = 2)
odd_row <- seq(from=1, to = dim(data_ed)[1], by = 2)
odd_col <- seq(from=1, to = dim(data_reshaped)[2], by = 2)
data_reshaped[,odd_col] <- data_ed[odd_row, ref_set]
data_reshaped[,odd_col+1] <- data_ed[odd_row+1, ref_set]
data_reshaped[,"fixed"] <- rowSums(data_reshaped[,] == data_reshaped[,1] | data_reshaped[,] == "0")
== length(ref_set)*2
data_reshaped[,"A"] <- rowSums(data_reshaped[,1:(length(ref_set)*2)] == "A")
data_reshaped[,"T"] <- rowSums(data_reshaped[,1:(length(ref_set)*2)] == "T")
data_reshaped[,"C"] <- rowSums(data_reshaped[,1:(length(ref_set)*2)] == "C")
data_reshaped[,"G"] <- rowSums(data_reshaped[,1:(length(ref_set)*2)] == "G")
data_reshaped[,"0"] <- rowSums(data_reshaped[,1:(length(ref_set)*2)] == "0")
data_reshaped[, "major"] <- "0"
data_reshaped[which(data_reshaped[,"A"] > length(ref_set)), "major"] <- "A"
data_reshaped[which(data_reshaped[,"T"] > length(ref_set)), "major"] <- "T"
data_reshaped[which(data_reshaped[,"C"] > length(ref_set)), "major"] <- "C"
data_reshaped[which(data_reshaped[,"G"] > length(ref_set)), "major"] <- "G"
#Candidate comparison calculations
candidate_reshaped <- data.frame(dummy = rep(NA, dim(data_ed)[1]/2))
candidate_reshaped[,1:2] <- NA
names(candidate_reshaped) <- rep(candidate, each = 2)
candidate_reshaped[,1] <- data_ed[odd_row, candidate]
candidate_reshaped[,2] <- data_ed[odd_row+1, candidate]
candidate_reshaped[,"mismatch1"] <- 0
candidate_reshaped[,"mismatch2"] <- 0
candidate_reshaped[candidate_reshaped[,1] != data_reshaped[,"major"],"mismatch1"] <- 1
candidate_reshaped[candidate_reshaped[,2] != data_reshaped[,"major"],"mismatch2"] <- 1
candidate_reshaped[,"mismatch_sum"] <- candidate_reshaped[,"mismatch1"] +
candidate_reshaped[,"mismatch2"]
candidate_sum_1 <- sum(candidate_reshaped[,"mismatch_sum"] > 0 & data_reshaped[,"fixed"] &
data_reshaped[, "major"] != "0")
candidate_sum_2 <- sum(candidate_reshaped[,"mismatch_sum"] == 1 & data_reshaped[,"fixed"] &
data_reshaped[, "major"] != "0")
candidate_sum_3 <- sum(candidate_reshaped[,"mismatch_sum"] == 2 & data_reshaped[,"fixed"] &
data_reshaped[, "major"] != "0")
candidate_sum_4 <- sum(data_reshaped[,"fixed"] & data_reshaped[, "major"] != "0")


candidate_sum_6 <- sum(candidate_reshaped[,1]== "0" & candidate_reshaped[,2]== "0")
#Results
#Number of fixed SNPs in the reference population: candidate_sum_4
#Number of SNPs with new different allele: candidate_sum_1
#Number of SNPs with 1 different new allele: candidate_sum_2
#Number of SNPs with 2 different new alleles: candidate_sum_3
#Number of SNPs that didnt work candidate_sum_6
```

Table A. Missingness by individual (`proportion of missing SNPs) of the total individuals analyzed calculated by PLINK.

Identity	Number of missing SNPs	Proportion of missing SNPs
I1	2986	0.02
OG1	12533	0.07
OG2	12393	0.07
I2	2910	0.02
OG3	3514	0.02
I3	5657	0.03
ON1	3893	0.02
OG4	3136	0.02
I4	16236	0.09
ON3	3205	0.02
S1	2707	0.01
I6	18377	0.1
ON2	3440	0.02
S2	5833	0.03
I7	56406	0.32
ON3	8513	0.05
OG5	5917	0.03
I10	109898	0.63
OG6	4046	0.02
OG7	3090	0.02
I8	4794	0.03
OG8	20679	0.12
I9	65620	0.34
ON4	2918	0.01
Average	15779.21	0.09

Table B. Individual total ROH coverage of immigrants and Swedish wolves calculated with PLINK defined as the proportion of the genome of the individual located in a region of homozygosity.

Identity	Individual total region of homozygosity coverage
I3	0.02
I8	0.04
I2	0.03
I6	0.01
I1	0.03

I5	0.03
Average	0.03

Identity	Individual total region of homozygosity coverage
OG1	0.06
OG2	0.05
OG3	0.08
OG4	0.06
OG5	0.08
OG6	0.04
OG7	0.07
OG8	0.04
ON1	0.07
ON3	0.02
ON2	0.03
ON4	0.27
Average	0.071

Table C. Total predicted and incremental SNPs to the reference population

Number of immigrants	Total Predicted SNPs	Incremental SNPs
0	0	0
1	8905	8905
2	14970	6065
3	19367	4397
4	22702	3334
5	25317	2615
6	27422	2106
9	31836	4414
10	32895	1059
	(...)	
20	38685	355
	(...)	
100	45025	19
	(...)	
200	45966	5