DEVELOPMENT OF A SNP MARKERS PANEL FOR PARENTAGE TESTING AND INDIVIDUAL IDENTIFICATION IN THE NORWEGIAN WHITE SHEEP (NWS) POPULATION

TESFAYE KEBEDE BELAY

Norwegian University of Life Sciences

Department here of animal aquacultural sciences

Master Thesis 2013 - 30 Credits

NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Department of Animal and Aquaculture Sciences

DEVELOPMENT OF A SNP MARKERS PANEL FOR PARENTAGE TESTING AND INDIVIDUAL IDENTIFICATION IN THE NORWEGIAN WHITE SHEEP (NWS) POPULATION

TESFAYE KEBEDE BELAY

EUROPEAN MASTER IN ANIMAL BREEDING AND GENETICS

THESIS ANIMAL BREEDING AND GENETICS (M30-IHA)

May 2013

SUPERVISOR: Prof. Dag Inge Våge

Centre for integrative genetics (CIGENE)
Department of Animal and Aquaculture Sciences (IHA)

Norwegian University of Life Sciences (UMB)

Table of contents

List of Tables

Table

List of Figures

Figure

List of Tables in Appendix 1

Table

List of Tables in Appendix 2

Table

Abbreviations

AI	Artificial insemination
CIGENE	Centre for integrative genetics
HWE	Hardy Weinberg equilibrium
IBD	Identity by descent
IBS	Identity by state
ISGC	International Sheep Genomics Consortium
LD	Linkage disequilibrium
MAF	Minor allele frequencies
NIBS	Normalized identity by state
NWS	Norwegian White Sheep
OAR	Ovis aries chromosomes
PE	Power of parentage exclusion
PI	Probability of identity
SNP	Single nucleotide polymorphism

Acknowledgements

First of all, I wish to thank the Almighty God for helping me to successfully complete this work.

I want to convey my sincere appreciation and gratitude to my supervisor Prof. Dag Inge Våge for giving me constructive pieces of advice and guidance from the beginning up to the end of this research work. I greatly indebted to him since without his encouragement, suggestion, insight, guidance and professional expertise, the completion of this work would have not been possible.

I would like to thank Erasmus Mundus Scholarship for financial support.

I am also thankful to the Norwegian University of Life Sciences, Department of Animal and Aquaculture Sciences for allowing me to study here and to use available student facilities during the study period.

My sincere appreciation goes to my friend Solomon Antwi Boison for his unreserved help on how to use Plink software as well as for his encouragement and advice throughout the course of my study. Thank you Solomon you are a real friend!

I am thankful to laboratory technicians in CIGENE, Kristina Vagonyte-Hallan and Kristil Kindem Sundsaasen, for their demonstration on how to design primers and perform SNP genotyping.

I would like to thank the Norwegian Sheep and Goat Breeders Association for the data used in this study.

I would also like to thank those people in IHA who supported my study in different aspects.

Last but not least, my special gratitude goes to my love, *Almaz Mengistu*, who has always been with me with her love spiritually. Thank you Almi, you have carried the entire burden in managing our children along with your office and field work. You have been my pillar behind this achievement. I am also greatly indebted to my lovely son *Firaol* and daughter *Koket* for their love and patience.

Summary

Correct parentage information is essential for the success of livestock breeding programs. However, pedigree errors could occur due to human made errors or unintentional misbreeding due to e.g. a broken fence or similar circumstances. Such errors have large impact on the efficiency of a genetic evaluation of breeding program. DNA based parentage information is therefore crucial in this regard. Based on data from 378 rams genotyped by the ovine Illumina 50K SNP chip, we have generated a highly informative SNP markers panel consisting of 68 markers distributed across 24 autosomes. These markers have 0.48 to 0.5 minor allele frequencies and have been located at ≥20Mbp apart from each other when residing on the same chromosome. Theoretical powers for identity and parentage exclusion of these markers were calculated based on allele frequency data. Practical power of the panel in paternity exclusion or assignment was verified using real data from NWS and Spael breeds. The 68 highly informative markers were also tested for their technical feasibility. Combined probability of identity for the 68 markers was estimated as $1.09x10^{-29}$, which increased to $9.81x10^{-15}$ when a subset of 33 markers with excellent technical performances (≥95 call rate) was used. The combined powers of parentage exclusion (PE) was calculated, considering the situations where both parents are genotyped but only one parent is evaluated for exclusion (PE1) or both parents are evaluated for exclusion (PE2) and when only parent is genotyped and evaluated for exclusion (PE3). The power was estimated to be 0.999999 (PE1), 1.00 (PE2) and 0.999886 (PE3). The corresponding powers when 40 markers with very good technical performances (≥90 call rate) were used were 0.999753, 0.999998 and 0.995206, respectively. Practically, the panel revealed 4.9% and 13.3% incompatible genotypes for at least one marker of the parent-offspring pairs studied in the NWS and Spael breeds, respectively. Moreover, the panel with 68 markers achieved an overall paternity assignment rate of 97.3%. These results show that the developed SNP marker panel has sufficient power for either paternity exclusion or assignment and for individual identification. Therefore, the results of this study can immediately be used in the parentage testing practice to provide parentage verification and is expected to contribute to the quality control in the NWS breeding system.

1. Introduction

The sheep industry in Norway has both economic and ecological importance. They are reared for lamb meat (major product) and wool production (Vatn, 2009). Their contribution to the gross domestic product is very low, but they have large importance for maintenance of the rural population and for preservation of landscape (Vatn, 2009). During summer, about 80% of all sheep are kept on common grazing in forest and mountains, whereas during rest of the year, they kept mainly in door on roughage with concentrate supplements (Eikje *et al*. 2008; Vatn, 2009). The sheep breeding scheme in Norway is based on progeny tested ram lambs in ram circles (Eikje *et al*.2008). The ram lambs are selected based on pedigree and own performances. The selected ram lambs (test rams) are moved to different flocks within a ram circle to ensure that their offspring for progeny testing are born in different environmental conditions. The test rams mate all ewes in heat except those elite ewes, which are mated by proven rams. The ram lambs are also progeny tested by use of artificial insemination (AI), which become more common in Norwegian sheep breeding (Eikje *et al*.2008). This technique is used more frequently in the NWS, which is the largest breeding population found in Norway and currently, around quarter of breeding ewes of the NWS are expected to be bred by AI.

Correct parentage information is essential for the success of livestock breeding programs. Genetic evaluation programs and managing inbreeding and undesirable genetic conditions rely completely on correct parentage information. If the basic assumption of correct parentage is hampered, the consequences will be biased estimates of heritabilities (Van Vleck, 1970a; Geldermann *et al*.1986; Visscher *et al*. 2002; Parlato and Van Vleck, 2012), breeding values (Van Vleck, 1970b; Geldermann *et al*. 1986; Long *et al*. 1990; Israel and Weller,2000; Banos *et al*. 2001; Baron *et al*. 2002), inbreeding rate (Banos *et al*. 2001; Visscher *et al*. 2002), and estimates of genetic progress (Van Vleck, 1970b; Geldermann *et al*.1986; Long *et al*. 1990; Israel and Weller, 2000; Visscher *et al*. 2002; Sander *et al*. 2006). This in turn resulted in reduced genetic gain and wrong decisions with regard to inbreeding and undesirable genetic conditions. Pedigree error can also result in a loss of power to detect linkage in linkage analysis of genetic diseases and quantitative traits (Epstein *et al*. 2000). The false evidence for linkage (i.e. reduced or inappropriately increased evidence for linkage) can detracts reproducibility of linkage outcomes and lead to misleading conclusions (Epstein *et al*. 2000; Cherny *et al*. 2001). Moreover, the pedigree error might raise the question of trust in pedigree certificates (Leroy *et al*. 2011).

Pedigree errors in animal breeding could happen due to several factors. It is a common problem in extensive breeding systems where multiple sires natural mating is practiced (Souza *et al.*2012) and where recording system is poorly established. Parentage misidentification may also occur in controlled system due to human errors in mothering up, recording and in artificial insemination process (Heaton *et al*. 2002; Weller *et al*. 2004; Souza *et al*. 2012), and also due to errors made by animal itself like jumping fences. In the Norwegian sheep breeding, the extensive management during summer and increasing trend in the use of AI could possibly result in pedigree error. In cattle breeding, measures such as a good recording and verification systems, keeping single ID throughout animal life and some control measures taken by AI companies are expected to minimize pedigree errors (Visscher *et al*. 2002). On other hand, DNA testing to identify the correct parentage is the best solution to overcome harmful effects of pedigree error on genetic evaluation (Parlato and Van Vleck, 2012; Souza *et al*. 2012).

DNA-based parentage testing is useful to: 1) reduce pedigree errors, improve genetic selection and speed up genetic progress, 2) enable multiple sires mating and identify the most productive sires, 3) reduce labour involved in mothering up and 4) identify untagged animals. It is based on detection and analysis of genetically inherited markers. Any diploid individuals have two copies of each chromosome that are made up of DNA, sections of which can be detected and used as markers. All animals inherit two copies of each chromosome: one copy from dam and one from sire. Therefore, if a marker is present in progeny but absent in both nominated parents, the progeny must be excluded as the offspring of that mating; or one of the nominated parents (sire or dam) is excluded from parentage when his or her genotypes is not compatible with the offspring (http://www.beefcrc.com/publications/fact-sheets.html). This probability of excluding an alleged parent depends on the marker type, the number of alleles, and the allele frequencies in the population to be used for parentage testing (Gomez-Raya *et al.* 2008**)**.

Different types of DNA markers have been used in parentage testing and individual animal identification. Microsatellite markers have commonly been used in parentage testing and determinations of identity because they are highly polymorphic, have high information content, and show a genome wide coverage and are easy to detect (Baron *et al*, 2002; Souza *et al*. 2012). However, microsatellites have high mutation rate that could cause misclassification in parentage testing (Tishkoff *et al*. 2003). In recent years, single nucleotide polymorphisms

(SNPs) have got attention as alternative markers. This is because SNPs are genetically stable (have lower mutation rate), are abundant in the genome, and are amenable to high-throughput automated analysis (Heaton *et al.* 2002; Tishkoff *et al*. 2003; Werner *et al*. 2004; Hara *et al.* 2010). The lower information content of SNPs compared to the highly polymorphic microsatellites are considered as one disadvantage, but this can be compensated for by using larger numbers of SNP markers (Werner *et al.* 2004).

Parentage testing and individual identification that use SNPs have been established in cattle populations in U.S (Heaton *et al.* 2002), European (Werner *et al.* 2004) and Japan (Hara *et al.* 2010). However, the SNP system is not in place or rarely established for other livestock species, including sheep. Moreover, the system developed for one species/breed might not work effectively for other species/breeds, indicating the need to establish the system for specific populations. Based on data from 378 AI rams, we have developed an efficient SNP markers panel from large set of SNP array (50k) in NWS population and the potential utility/power of the panel in parentage exclusion or allocation and individual identification was estimated.

2. Materials and methods

2.1. Animals and their relationships

A total of 378 AI rams of the NWS have been genotyped by the Illumina 50K SNP chip. These data were used in the development of SNP marker panel. To get insight about degree of relationships among individuals in the dataset, pairwise identity by state (IBS) similarity matrix and identity by descent (IBD) was calculated using Plink software v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/ and Purcell *et al*. 2007). Individuals may share alleles that are IBS not only because they are closely related but also due to chance. As a result, the IBS values need to be scaled or normalized to provide better inference to genetic relationships among individuals. We scaled each IBS value based on the smallest IBS estimate as: $NIBS_i=(IBS_i-IBS_l)/(1-IBS_l)$, where $NIBS_i$ is normalized IBS at the i^{th} pair of individuals when i=1,2,3, ...N possible number of pairs of individuals, IBS_i is the i^{th} pairwise IBS value in the matrix and IBS_l is the lowest IBS value in the matrix. This assumes that individual pairs with the lowest IBS value are distantly related or not related, but they shared alleles that are IBS simply due to chance.

2.2. Genotypes and quality control

The Illumina 50k SNP Array (Illumina, San Diego, CA, USA) contained 54246 SNPs. Total numbers of SNPs and other SNP information per chromosome is presented in Table 2 and minor allele frequency (MAF) distribution for SNPs on the array is given in Figure 2. Of the 54246 SNPs, 378 of them were not mapped to any of the *Ovis aries* chromosomes (OAR). In addition, 1452 of the SNPs were located on sex-chromosomes; 1027 SNPs were monomorphic, 4787 SNPs had no genotype call and 2391 SNPs failed Hardy-Weinberg equilibrium (HWE) exact test ($P<0.05$) (Table 2). All SNPs in these categories were discarded from the original set of 54246 SNPs, resulting in set of 44,490 SNPs that were used for further screening in the process of informative markers selection.

2.3. Informative SNP markers selection

Selection of informative markers begun by eliminating markers that failed to pass the usual SNPs quality control parameters (genotype call rate, HWE test, monomorphic). Those markers that passed the preliminary screening were subjected to two other stringent criteria. These criteria were a MAF value in the range of 0.48 to 0.5 and $\geq$ 20 Mbp physical distances between markers located on the same chromosome. SNPs were ranked according to their MAF and those markers that met the threshold value set for MAF were selected (Table 2) and sorted by their chromosome and map position to select those distributed across the genome. The second criterion was implemented on markers that met the first criterion and that reside on the same chromosome. However, four markers that located at $\leq$20Mbp (7 to 18Mbp) distance from each other or from other markers were, by mistake, included in to the panel during primer design.

2.4. Primer design and SNP genotyping

Markers that met both criteria were tested to evaluate their technical performances. Genomic DNA sequences of 120 bases flanking the 50K SNP chip marker were found from the incomplete sheep genome reference assembly build 1. Both amplification and extension primers were designed using the MassARRAY® Designer software from Sequenom. The primers were designed in 2x40 multiplexes PCR reactions. Six functional markers were included in both multiplexed reactions and these markers were not used in paternity testing but added into the reactions for routine diagnostic purposes. Primers sequences used, amplification length and hybridization temperature is given in Appendix 1 Table 3. The iPLEX® Gold system was used for genotyping and analyzes were performed on the

Sequenom MassARRAY Workstation 4.0. DNA samples from 109 Spael rams and 59 NWS rams were used for testing the selected SNP markers. The 59 NWS AI rams were not represented among those genotyped by the Illumina 50k SNP array.

2.5. Statistical analyses

2.5.1. Estimation of allele frequency and heterozygosity

Minor allele frequencies and HWE exact test were estimated with Plink v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/ and Purcell *et al*. 2007). Genotype frequencies and observed heterozygosity (O(H)) were calculated by direct count method using the number of individuals in each genotypic class, while an unbiased estimate of expected heterozygosity (E(H)) was estimated using the formula of Nei et al. (1978): $2P_{ij}(1-P_{ij})*2N_j)/(2N_j-1)$, where P_{ij} is the frequency of the i^{th} allele at the j^{th} locus and N_j is number of individuals surveyed at the j^{th} locus. The Polymorphic information content (PIC) for each marker was calculated as

$$1-\sum_{i=1}^{n} p_i^2-\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} 2p_i^2\,p_j^2$$

where pi is population frequency of the i^{th} allele and n is number of alleles per marker (Botstein *et al*. 1980).

2.5.2. Calculation of power of exclusion and probability of identity

Exclusion power is the probability of excluding a random individual from the population as a potential parent of an animal based on genotype of one or both parents and offspring. The exclusion powers when genotypes for two parents and one offspring are known, but aimed to exclude *a parent* (PE_1) or *both parents* (PE_2) and parentage exclusion power when genotypes available only for one parent and one offspring (PE_3) were calculated using the formula of Jamieson & Taylor (1997):

$$PE1 = 1-2\sum_{i=1}^{n} p_i^2+\sum_{i=1}^{n} p_i^3+2\sum_{i=1}^{n} p_i^4-3\sum_{i=1}^{n} p_i^5-2(\sum_{i=1}^{n} p_i^2)2+3\sum_{i=1}^{n} p_i^2\sum_{i=1}^{n} p_i^3$$

$$PE2 = 1+4\sum_{i=1}^{n} p_i^4-4\sum_{i=1}^{n} p_i^5-3\sum_{i=1}^{n} p_i^6-8(\sum_{i=1}^{n} p_i^2)2+8(\sum_{i=1}^{n} p_i^2)(\sum_{i=1}^{n} p_i^3)+2(\sum_{i=1}^{n} p_i^3)2$$

$$PE3 = 1 - 4\sum_{i=1}^{n} p_i^2 + 2(\sum_{i=1}^{n} p_i^2)2 + 4\sum_{i=1}^{n} p_i^3 - 3\sum_{i=1}^{n} p_i^4$$

, where p_i is allele frequency of the i^{th} allele and n is number of alleles at a locus.

Combined power of exclusion over all independent markers studied might be needed to achieve acceptable power of exclusion. It measures the capacity of the system to detect a false accusation of parentages and was computed (Jamieson & Taylor 1997) as: $1 - [\prod_{j=1}^{k}(1 - \text{PEj})]$, where k is number of loci and PE_j (j=1, 2 - - - k) is power of exclusion for individual markers as calculated from either of the above equations.

Beside the estimated allele frequencies, MAF of 0.1, 0.2, 0.25, 0.3, 0.4 and 0.5 were hypothetically generated for the selected markers. Then, combined powers of exclusion (PE1, PE2 and PE3) were calculated for each categories of the generated MAF and the powers were plotted against the MAF (Figure 6). It was done to see effect of MAF on power, as well as to determine the smallest value of MAF at which reasonable power of parentage exclusion is obtained with this panel.

Power of the panels in uniquely identify individuals (probability of identity, PI) in the population was also estimated based on genotype frequencies. Probability of identity is defined as the estimated probability that two individuals selected at random from a population would possess identical multi-locus genotypes. This probability for a marker is equal to summation of the square of each genotypic frequency (Heaton *et al.* 2002; Hara *et al.* 2010). Combined PI, which is the multiple product of each individual marker probability, was computed as:

$$PI = \prod_{i=1}^{n}(pp_i^2 + 2pq_i^2 + qq_i^2).$$

2.6. Verification of power of the panel

Practical exclusion power of the selected SNP markers was tested using genotypes from 88 fathers (rams) and 185 offspring. These 185 parent-offspring pairs, or 273 individuals, were selected from the 378 AI rams used for the panel development. The numbers of offspring per ram varied from 1 to 16. For the remaining 105 individuals, parent genotypes were not available, and were therefore not included in the verification analysis. In addition to the data from NWS, the verification activity was also conducted with genotype data from Spael sheep

breed (30 parent-offspring pairs). This data consisted of 21 half-sib families with totally 30 offspring and the number of offspring per family (ram) varied from 1 to 5. For the Spael breed, only technically feasible markers were used to detect mismatches (Table 5). Number and proportion of mismatches between each hypothesized parent-offspring pairs were calculated using software found at the Galaxy platform developed at CIGENE. This program checks for errors in Mendelian inheritance by using parent and offspring genotypes and pedigree information. Manual inspection was also used to confirm the mismatches reported by the software. Exclusion was considered whenever the genotype of the father was incompatible with the genotype of the progeny for at least one of the markers.

Moreover, practical power of the panel was further evaluated from paternity assignment perspective. The paternity assignment analysis was conducted using genotypes of the 185 offspring that were used in the exclusion analysis. In this case, two groups of candidate fathers were used (88 and 227 candidate fathers). The 88 males are possible fathers of the 185 offspring as indicated in the pedigree file. In seeking for complete assignment, 139 extra males on which we don't have information about whether they are suspected fathers or not were added into the 88 candidate fathers file, resulting in 227 candidate fathers. It was done assuming that increase in number of males in the candidate fathers file will increase probability for a true father to be included into the file. In this analysis, sib ship was also attempted to be inferred in addition to paternity. The assignment was computed by a program called COLONY v.2 (Jones and Wang, 2009).

3. Results

3.1. Genetic relationships between animals

The degree of relatedness among individuals in the dataset was evaluated based on identity by state and/or identity-by-descent (IBD). The normalized identity by state (NIBS), which is expected to represent the true genetic relationships, was plotted against number of pairs of individuals (Figure 1). It was found that there were 71253 possible pairs of individuals for which genetic relatedness was determined. Pairwise IBS similarity was in the range of 0.6816 to 0.8487 (data not shown), which were normalized with reference to the minimum IBS (0.6816) and resulted in the range of 0 to 0.52 (Table 1). In addition to correcting for alleles shared between two individuals by chance alone, the normalization increased the range of IBS similarity. This made the graphical presentation and interpretation of the NIBS results much easier than that of IBS. The NIBS was categorized into three classes based on the observed nature of its value as shown in Figure 1. These categories represented different type of genetic relationships that exist between individuals in the dataset. Values for each measure of IBD indicated in Table 1 were significantly differing among the classes of NIBS established. Types of genetic relationships given in Table 1 were based on pedigree information.

Table 1: Normalized identity by state and average measures of identity by descent (IBD) and relationship type for all possible pairs of individuals in the dataset

NIBS	IBD0	IBD1	IBD2	PIBD	RT	n
0-0.20	0.9672	0.0244	0.0084	0.0206	-	70023
0.2-0.36	0.5492	0.4225	0.0283	0.2396	HS/GG	1042
0.4-0.52	0.0130	0.9524	0.0346	0.5108	PO/FS	188

NIBS: normalized identity by state; IBD0, IBD1, and IBD2: proportion of allele shared identical by descent in which 0, 1 or 2 alleles are inherited from a recent common ancestor, respectively; PIBD: proportion of alleles that are IBD; RT: relationship types (PO: parent-offspring; FS: full-sib; HS: half-sib; GG: grandfather-grandsons) based on pedigree information; n: number of pairs of individuals in each category of NIBS.

Individual pairs (n=70023) with less than 0.2 NIBS were distantly related. On average each pair in this category shared only about 2% (PIBD) of their alleles that are IBD throughout the genome. In other word, they shared 0 alleles in common at most of their loci in which the IBD0 was about 97% (Table 1). Other pairs of individual (n= 1042) with NIBS values

between 0.2 and 0.4 have shared about 24% of their alleles that are IBD and the proportion of 0 allele share was about 55%. Animals in this category shared reasonable proportions of alleles that were inherited from recent common ancestor. As a result, they are expected to be genetically related, in which half-sib or grandfather-grandson relationships were indicated for most of those pairs in the pedigree file. It was also observed that a group of pairs of individuals (n= 188) relatively have large NIBS values (≥0.4) and this group was clearly separated from the other pairs (Figure 1). The average proportion of IBD alleles for a pair of individuals in this group was around 51% whereas the proportion of IBD0 was only 1.3% (Table 1). This suggested that those pair (s) of individuals with NIBS ≥0.4 were closely related. Based on pedigree information, most of the 188 pairs of individuals have either parent-offspring or full-sib relationships. On average, individual pairs in every category of NIBS have shared more 1 allele than two alleles.

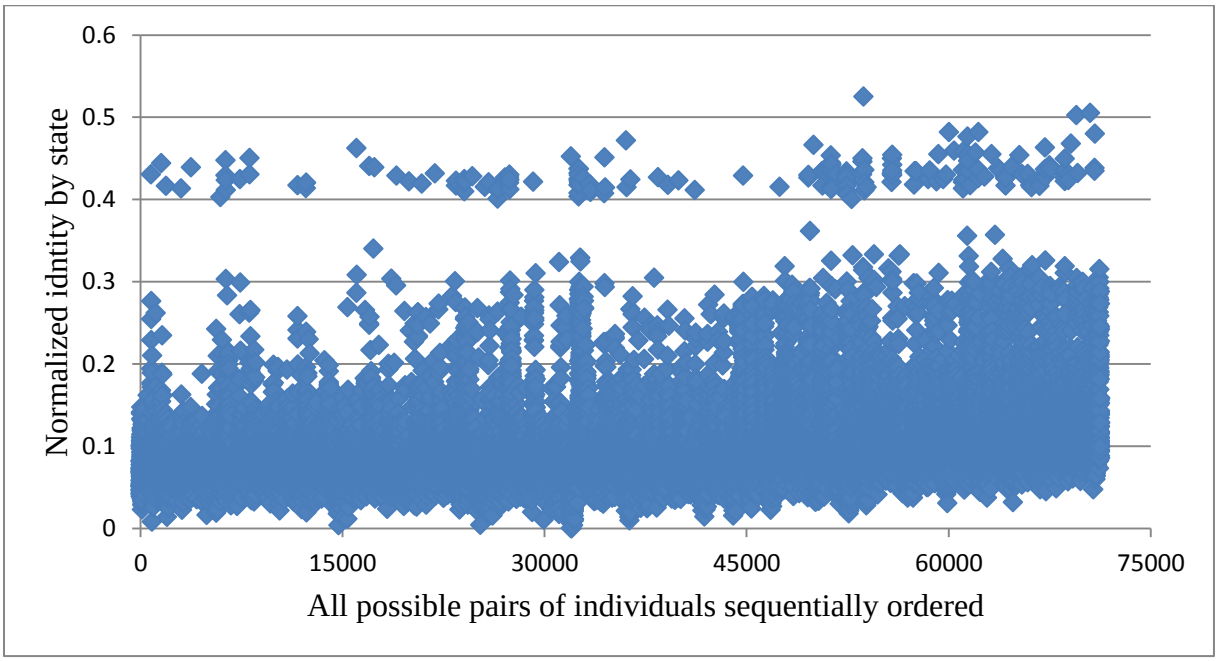

Figure 1: Genetic relationships between possible pairs of individuals in the dataset based on normalized identity by state, which is corrected for relationships due to chance alone

3.2. Highly informative SNP markers and their heterozygosity estimates

Numbers of highly informative markers per chromosome are presented in Table 2. Of 44490 SNPs that passed the preliminary quality control parameters, 2443 of them met the threshold value set for MAF (0.48 to 0.5). Among these, 115 candidate markers that met both criteria were selected from the 50k SNP Array (Table 2). Finally, 68 highly informative markers that

were possible to combine in 2x40 multiplex reactions were selected. The remaining 12 markers, out of the 2x40 reactions, were functional markers that have been included in to the reaction for routine diagnostic purpose. MAF distributions for the entire SNP-set on the chip and for the 68 SNPs are given in Figure 2. The panel is biased towards high MAF markers (Figure 2) to ensure their informativeness and hence the power of the panel. List of the 68 SNP markers, their chromosomal positions, allele and genotype frequencies are given in Appendix 1 Table 1. Of those candidate markers (115) that failed to be included into the primer design, some of them were discarded due to failure in primer design and some were excluded due to the forced inclusion of the six functional (diagnostic) markers into each multiplex. List of amplification and extension primers, amplicon length and hybridization temperatures are given in Appendix 1 Table 3.

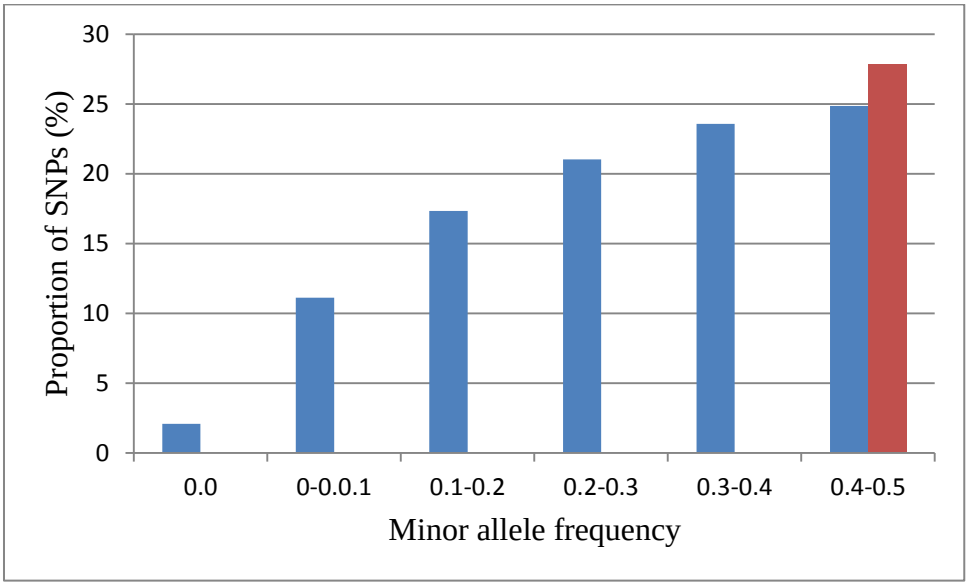

Figure 2: Minor allele frequency distributions for all 49459 SNP with non-zero call rate (blue bars) and for the 68 SNPs in the parentage panel (red bar)

Table 2: SNP information and number selected per chromosome and selection criteria

OAR	SNP	Interval (bp)	Monomorph	Nogenocall	HWE	Selection criteria		Selected SNPs
						MAF (.48-.5)	≥20Mb distance	
0*	378	-	8	44	41	-	-	-
1	5930	50536	96	568	294	272	13	8
2	5474	48073	87	444	249	236	8	4
3	5008	48468	98	421	193	244	11	8
4	2680	47447	58	211	98	118	4	2
5	2365	49386	41	186	93	115	5	4
6	2593	49773	37	220	109	131	5	3
7	2252	48277	51	205	95	103	5	1
8	2057	47604	32	169	101	92	4	3
9	2141	47098	30	198	93	120	4	4
10	1851	50873	38	166	91	84	5	3
11	1180	56712	29	93	56	65	4	0
12	1723	49973	39	146	72	82	4	2
13	1696	52426	34	143	69	84	4	4
14	1174	58610	32	109	66	52	3	1
15	1694	53063	41	147	73	79	4	2
16	1580	48873	30	164	82	65	3	1
17	1420	55280	31	141	69	55	4	2
18	1413	50938	27	111	59	66	3	2
19	1248	51960	29	109	47	59	3	1
20	1148	48426	24	121	54	42	3	2
21	898	61469	17	104	46	49	3	2
22	1097	50199	22	96	54	54	3	2
23	1128	58821	18	110	64	58	3	3
24	741	59833	18	62	27	35	2	0
25	1001	48042	11	88	56	51	3	3
26	924	53977	12	102	40	32	2	1
x	1451	88233	37	109	-	-	-	-
y	1	-	0	0	-	-	-	-
Total	54246		1027	4787	2391	2443	115	68

**SNPs not assigned to any of Ovis aries chromosomes (OAR); total number of SNPs, average interval between SNPs, number of monomorphic SNPs, number of SNPs with no genotyping call (nogenocall), number of SNPs that failed Hardy-Weinberg equilibrium (HWE) test ($p<0.05$), number of SNPs with minor allele frequency (MAF) between 0.48 to 0.5 and number of SNPs both with MAF 0.48 to 0.5 and located at ≥20Mbp distance apart from each other on the same chromosome.*

Average heterozygosities and polymorphic information content estimate for the 68 markers are given in Table 3. Average O(H), E(H) and PIC estimates were 0.4982, 0.5006 and 0.375,

respectively (Table 3). O(H) ranged from a low of 0.4392 for OAR15_6615347.1 to the highest value of 0.5661 for OAR19_1987551.1 (Table 3, Appendix 1 Table 1), while E(H) ranged from 0.5001(OAR2_25624172.1) to 0.5007(OAR1_122906056.1) (Table 3, Appendix 1 Table 2). The width of range for observed heterozygosity was larger than for expected heterozygosity (Table 3). The overall estimate of average O(H) was slightly lower than E(H), indicating deficiency in heterozygosity, but the difference was statistically insignificant. PIC ranged from 0.3747 (OAR2_25624172.1) to 0.375(OAR1_122906056.1) (Appendix 1 Table 2). The PIC estimated was relatively high since the maximum PIC value for SNP is 0.5.

Table 3: Mean (±SE) observed and expected heterozygosity and polymorphic information content for the 68 markers

Parameters	Mean	Standard error (SE)	Minimum	Maximum
O(H)	0.4982	0.0033	0.4392	0.5661
E(H)	0.5006	1.3×10^{-05}	0.5001	0.5007
PIC	0.375	6.6×10^{-06}	0.3747	0.375

O(H) is observed heterozygosity; (E(H) is expected heterozygosity; PIC is polymorphic information content;

3.3 Technical performance of the 68 markers

Two 40-plexes genotyping reaction were run over 168 DNA samples following iPLEX Gold genotyping protocols used in CIGENE. The technical performance of the SNPs was evaluated mainly in terms of total genotyping efficiency (i.e. call rate: percent of genotypes with calls out of the total number of possible calls). In both breeds, 14.71% of the SNPs (n=10) were not performing at all i.e. they had zero call rate while 11.76% (n=8) of them performed poorly with call rate between 62.5 to 84%. The latter group of SNPs showed (very) low signal/intensity in which sometimes it was difficult to see the peaks in the spectrum. Among the remaining SNPs that were considered as good performing (≥85 call rate), SNP OAR10_57586299.1 had unusual genotypes in which it possess both homozygous genotypes (AA and GG), but no heterozygous genotype (AG). This SNP had 98% call rate.

In a check for concordance in polymorphism between the two genotyping methods (Illumina and Sequenom), one SNP (OAR15_50080570.1: with call rate of 92%) was found to be

monomorphic in Spael breed that were genotyped by Sequenom. But, it was polymorphic in NWS with either of the genotyping methods. All other SNPs were found polymorphic in both breeds and the same kinds of polymorphisms were detected by both methods at respective loci. Relatively, the SNPs were less polymorphic in the Spael in which the MAF ranged from 0.0972 to 0.4954 (excluding the monomorphic locus) and more than 82% of the SNPs in this breed had >0.2 MAF (Figure 3). Based on the 59 samples in NWS, MAF ranged from 0.2679 to 0.5 (data not shown since MAF from large sample was available for NWS).

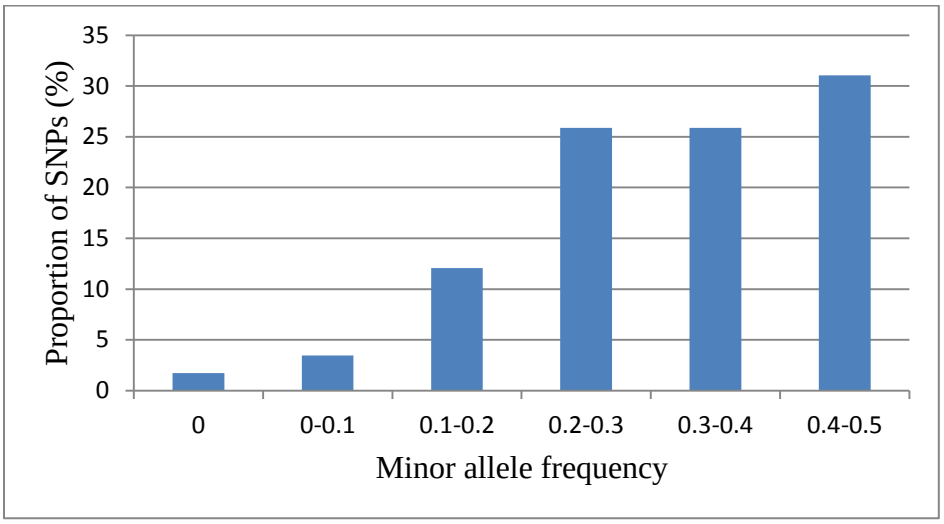

Figure 3: Minor allele frequency distribution for the 58 SNPs with non-zero call rate in Spael breed. One marker with zero MAF (monomorphic) was also included in the distribution.

On the other hand, SNPs analysed in Spael had better call rate than in NWS (Figure 4). For example, about 30% of the SNPs in NWS were found to have a call rate <85%, while only 10% of the SNPs in Spael had a call rate <85%. Similarly, individuals in Spael had better genotype call rate than in NWS (Figure 5). For example, about 14% of the individuals in NWS had no genotype in 22 to 46% of the loci, while only 2.75% of the individuals in Spael had missing genotype in 20 to 28% of the genotyped SNPs.

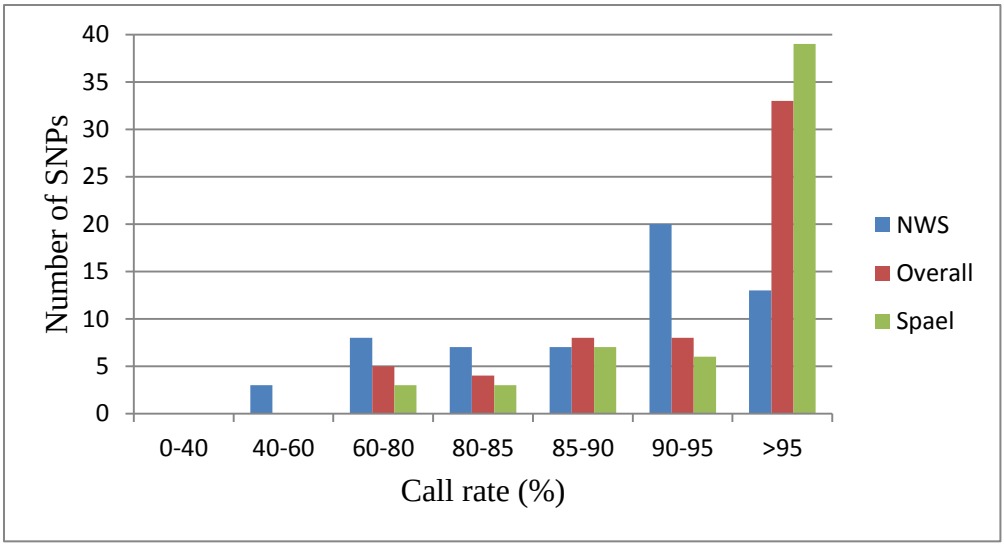

Figure 4: Distribution of genotype call rate for the technically tested SNP markers in NWS and Spael sheep breeds and in the overall dataset. Markers that totally failed the technical test (without any call) were not included in the distribution.

In addition to the call rate and check for concordance, error in Mendelian inheritance patterns was checked to further evaluate technical performance of the 68 SNP markers. This check was only done for Spael sheep using genotypes from the 30 parent-offspring pairs, because no families were found in the NWS animals genotyped by the 2x40 multiplexes (Sequenom). When the genotype of a single SNP failed to match in at least 2 parent-offspring pairs, this SNP was considered to be unreliable provided that no mismatch (s) occurred at other locus (loci). Mismatch at a single locus between paternal and offspring genotypes was observed in 50% of the parent-offspring pairs studied. This SNP was the one with unusual genotypes (as indicated above), confirming that this SNP is truly unreliable. Another bad performing SNP (OAR6_96871879.1: with call rate 96%) that failed to show Mendelian inheritance was also found. Genotypes of this SNP were found to be mismatched in 20% of parent-offspring pairs studied. Mismatches at these two loci were not considered in the pedigree error rate calculation. In addition to these two loci, the monomorphic locus found in Spael and those with no call rate were excluded from the subsequent calculation of technical power of parentage exclusion. Overall, 30.9% of the 68 markers genotyped gave no result (14.7%) or unsatisfactory results (16.2%).

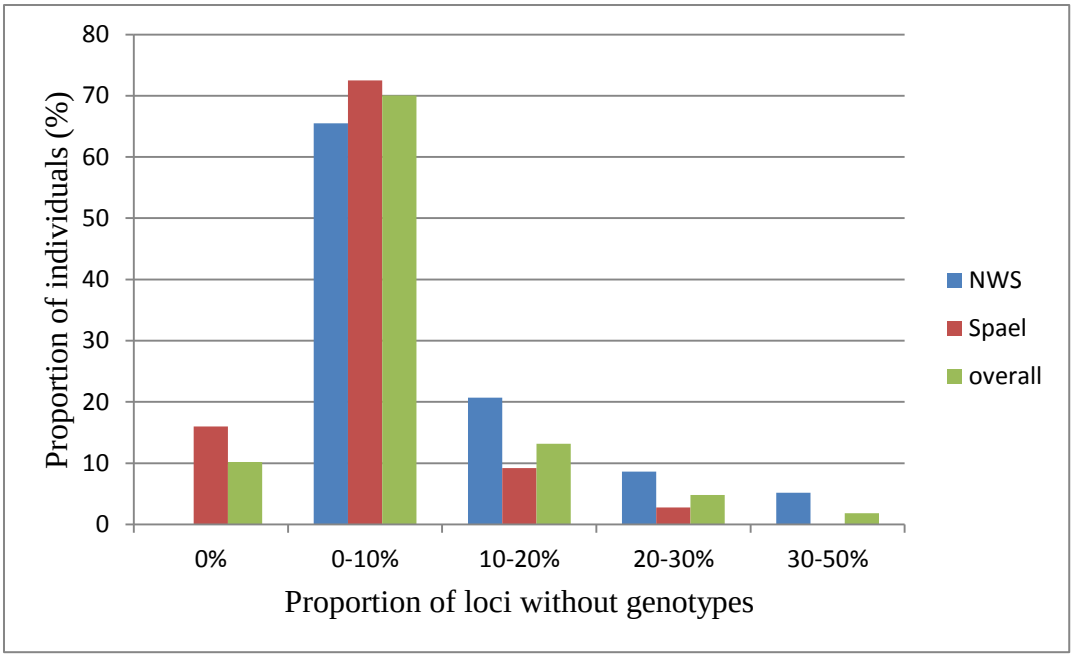

Figure 5: Distribution of loci without genotypes as a function of individuals in the two breeds and in the overall dataset. Majority (>65%) of individuals in the overall dataset, as well as in each breed had no genotypes in less than 10% of the loci. More missing genotypes were observed in NWS than in the Spael breed. One outlier individual in NWS with ≥95% missing genotype was excluded.

3.4. Power of parentage exclusion

The utility of the combined set of 68 markers was evaluated by estimating the theoretical power in parentage testing and individual identification. The power was estimated based on allele and genotype frequencies tabulated in Appendix Table 1. Based on allele frequencies, combined/joint power of estimate for PE_1, PE_2 and PE_3 was 0.999999, 1 and 0.999886, respectively (Table 4). This indicated that the panel is theoretically powerful enough to exclude ≥99.99% of falsely accused parent(s) from parentage.

Theoretical power of the panel in parentage exclusion at different hypothetically set allele frequencies were also estimated and are shown in Figure 6. As expected, the power was increased with increase in MAF and attained the maximum power at frequency of 0.5. The panel showed reasonable power for parentage exclusion (>99%) with MAF values as low as 0.25 (Figure 6). This can be taken as an indication for robustness of the panel. This analysis assumed constant number of markers (68) i.e. all markers will technically be feasible. However, such assumption might not hold always since some markers fail and hence are

excluded from power calculation. Under this condition, reasonable power could not be found with MAF values as low as 0.25.

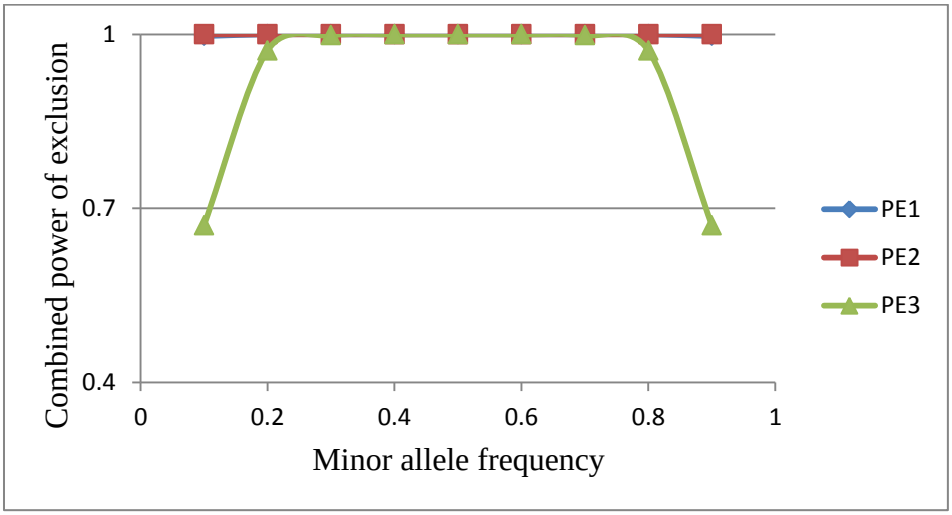

Figure 6: Effect of minor allele frequencies on combined power of parentage exclusion. When only one parent genotyped (PE3), the power of parentage exclusion was 99%, with MAF-value ≥0.25. With both parents genotyped (PE1 and PE2), power were very high even at MAF of 0.1.

Power of parentage exclusion was also calculated after the 68 markers were technically tested. As indicated earlier, some of these markers failed to perform (14.71%) and some performed poorly (16.17%) while majority of them (69.12%) passed the test with good results. As a result, powers at different number of markers were estimated (Table 4). As expected, the powers increased with increasing number of markers (Table 4, Figure 7). The rate of increment was highest for PE2 followed by PE1 and least for PE3. This trend was clearly shown in Figure 7, where the power curve for PE2 was the steepest and above the other two curves. At relatively large number of markers (≥48), PE2 attained the maximum power possible (100%). PE2 even achieved high power (99.93%) at as low as 22 best performing markers. This power might be sufficient for excluding both wrongly alleged father and mother from parentage. However, at least 33 markers are needed to achieve a power of ≥99.99% in PE2 (Table 4). Similar to PE2, PE1 had similar power at large number of markers and it achieved reasonable power (99.89%) with the 33 markers that performed well in Sequenom array with ≥95% call rate (Table 4). In this case, if higher power (≥99.99%) is required to exclude paternity from a given mother-offspring pairs (PE1), at least 46 SNP markers are needed. In the third scenario (PE3), 99.52% power was achieved with 40 technically well

performing 40 markers. Unlike the other two powers, PE3 failed to achieve ≥99.99% and the maximum power it achieved was 99.94% (Table 4, Figure 7). However, practically, it was efficient in excluding wrongly assigned paternity both in NWS and Spael breeds (Table 5).

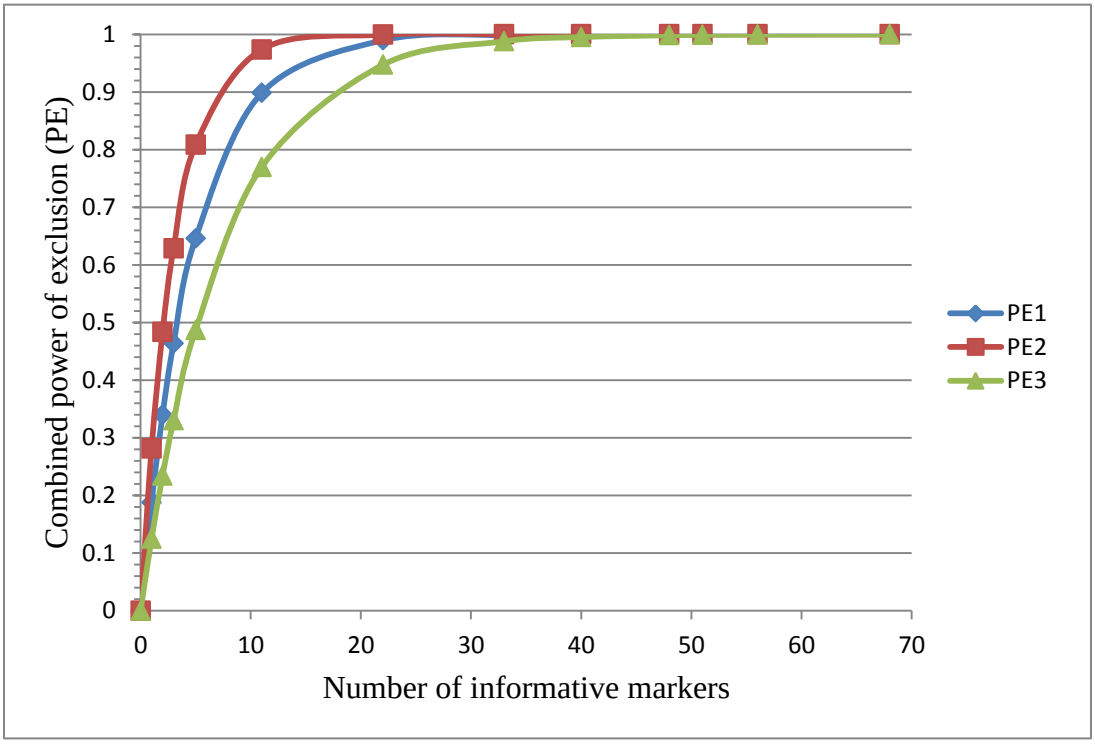

Figure 7: Effect of number of markers on power of parentage exclusion measured by three possible parameters (PE1, PE2 and PE3).

3.5. Probability of identity

The utility of the 68 SNP markers were further evaluated to estimate their power in individual identification, which was estimated based on genotype frequencies tabulated in Appendix 1 Table 1. The estimated probability that two individuals drawn at random from NWS population would possess identical genotypes at the 68 loci was 1.09×10^{-29} (Table 4). Theoretically the panel is sufficient to uniquely identify every individual in NWS population.

Similar to power of parentage exclusion, probability of identity was reduced with decrease in the number of markers (Table 4). With the best performing (≥95% call rate) SNPs, 22 to 33 markers resulted in reasonably low probability of identity, which might be suffice to unique identity every individuals a population.

Table 4: Combined powers of exclusion and probability of identity for a panel with different numbers of markers, which were set mainly based on call rate*

Panel	PE1	PE2	PE3	PI	Call rate (%)
68	0.999999	1.000000	0.999886	1.09×10^{-29}	≥0
56	0.999991	1.000000	0.999434	1.44×10^{-24}	≥63
51	0.999975	1.000000	0.998896	2.06×10^{-22}	≥80
48	0.999953	1.000000	0.998353	3.85×10^{-21}	≥85
40	0.999753	0.999998	**0.995206**	1.06×10^{-17}	≥90
33	**0.998943**	0.999981	0.987798	**9.81×10^{-15}**	≥95
22	0.989622	**0.999301**	0.947009	4.33x10-10	≥98

**Some markers with better call rate were excluded from power calculation due to reasons related to monomorphism, unusual genotypes and failure in Mendelian inheritance. PE_1 and PE_2 is combined power of parentage exclusion knowing genotype of both parents and offspring, but intending to exclude one (PE_1) or both parents (PE_2); PE_3 is combined power of parentage exclusion knowing one parent (either sire or dam) and offspring; PI is probability of identity.*

3.6 Verification of power of the panel

Practical power of the panel with 68 markers was tested in terms of paternity exclusion and paternity assignment. The exclusion test was conducted for the scenario when only one paternal and offspring genotypes are known (PE_3), using genotypes of 88 fathers and 185 offspring. The SNP marker panel revealed incompatible genotypes for at least two markers in 9 of the 185 parent-offspring pairs studied (4.9%). On average, the genotype incompatibility occurred at 6 different loci per parent-offspring pair, ranging from 2 to 14 loci (Table 5). Since mismatch at least in one marker was considered, the nine sires could be excluded from paternity. List of the 185 parent-offspring pairs, number and proportion of incompatible loci per the paternal-offspring pair are presented in Appendix 2 Table 1.

Power verification was also done with markers that performed well in the Sequenom array. With such markers, both the 185 parent-offspring pairs from NWS and 30 parent-offspring pairs from Spael sheep breeds were used to further evaluate the practical power of the panel. In the Spael breed, the panels showed four mismatches out of the 30 pairs (13.3%) at least at one locus (Table 5). These mismatches did not include the two loci that were regarded as

unreliable SNPs according to the Mendelian error check. The numbers of incompatible loci per parent-offspring pair were in a range of 1 to 8 (Table 5). In the NWS, the same numbers of mismatches (nine) that were detected by the 68 markers were also detected with 40 to 56 technically feasible markers. In both breeds, the numbers of incompatible loci per parent-offspring pairs were decreased with decrease in number of markers (Table 5). This results show that the panel has sufficient power for identifying wrong parentage even with 40 markers.

Table 5: Number of incompatible loci per parent-offspring (PO) pair for a panel with different number of markers

PO pair ID	Number of incompatible loci detected in a panel with different markers:			
	68	56	48	40
For NWS:				
1	2	1	1	1
2	6	4	4	3
3	5	5	2	1
4	8	7	7	7
5	6	4	4	3
6	5	5	3	3
7	4	2	1	1
8	14	12	10	10
9	7	6	6	4
For Spael:				
10	-	1	1	1
11	-	4	4	4
12	-	8	8	6
13	-	8	7	6

The power verification has also been evaluated from the paternity assignment perspective, where 88 and 227 candidate fathers were used with the 185 offspring. The assignment was done only for NWS breed. When 88 rams included in the candidate fathers file, each offspring that were not excluded during exclusion analysis was correctly assigned to their biological

father with 100% probability and rate of assignment (Appendix 2 Table 2). Correctness of the assignment was confirmed based on pedigree information and exclusion analysis results. Since there were nine mismatches among the 185 parent-offspring pairs evaluated (Appendix Table 2 Table 1), at least 176 offspring (100% rate of assignment) were expected to be assigned to their true father. However, 180 offspring (n=180: Figure 8; Table 6) were assigned to their biological father. These results showed that four out of the nine offspring that have been associated with wrong fathers were now assigned to their true fathers (Appendix 2 Tables 2 and 3). Overall, the panel correctly assigned 97.3% of the offspring to their biological fathers. For the remaining 5 animals (2.7%) no assignment was found, probably due to lack of the true fathers in the dataset.

With large number of candidate fathers (227), paternity assignment rate was estimated to 82.7% (n=153). In this case, 3 offspring (1.6%) were wrongly assigned, resulting in an overall rate of 81.1% (n=150) assignment compared to 97.3% (n=180) rate of assignment in the other analysis with 88 candidate fathers (Table 6). In this case, one offspring among those with incorrect paternity was assigned to its most likely father. This increased the number of paternity assigned offspring from four to five. List of the mismatched pairs with their assigned fathers (if any) are given in Appendix 2 Table 3.

Table 6: Summary of practical power of the panel (68 markers) in paternity exclusion and assignment verification

Method	#fathers	#offspring	#off. with no sire assigned	#off. with wrongly assigned sire	#off. with true sire assigned
Exclusion	88	185	9	0	176 (4.9%)*
$Assignment_1$	88	185	5	0	180 (97.3%) $
$Assignment_2$	227	185	32	3	150 (81.1%) $

*value in bracket is rate of mismatches or pedigree error rate; $ paternity assignment rate; number of offspring (#off.)

Similarly, most of the inferred full- and half-sib relationships were in accordance with the relationships observed in pedigree file. However, larger numbers of sib ship pairs (half-or – full sib) were detected in the assignment than that observed in pedigree. In fact, most of those sib ships that were found by the assignment but were not found in pedigree had from very low to medium probability of assignment, indicating that they were not true sibs.

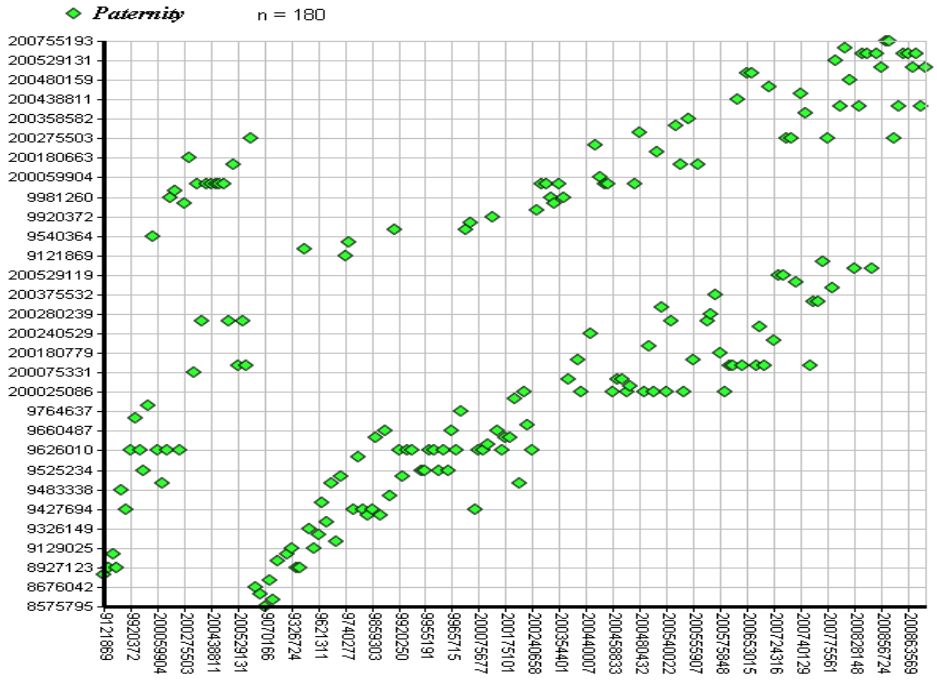

Figure 8: Paternity assignment plot in which the numbers in y-and x-axis represents father and offspring ID, respectively. The plot was condensed to fit to the size of window as a result ID of each parent and offspring was not displayed. The plot was drawn by the program COLONY v.2 (Jones and Wang, 2009). Each star symbol in the plot indicated parent-offspring pair, which is called ‘paternity’. The number of stars on horizontal line in the plot indicated number of offspring per father.

4. Discussion

4.1 Genetic relationships

Genetic relationships among individuals in the population can be determined based on either pedigree or marker information. In the pedigree approach, founder animals are assumed to be unrelated to each other. Also, there is a possible occurrence of recording errors in the pedigree file. Because of these, marker based inference to genetic relationships seems to be an alternative approach. Since related individuals share more alleles than unrelated ones, relationships can be estimated from marker genotypes in every pair of individuals, including the founders. However, determination of relationship types only based on marker information is not an easy exercise. For example in this study, we were able to classify the possible number of pairs of individual into more similar groups (Table 1) based on proportion of NIBS and IBD estimates, but it was not possible to determine the types of genetic relationship. With sophisticated analysis (additional parameter estimated and use of additional software), Stevens *et al*. (2011) were able to infer the type of genetic relationship in population data based on combined IBS and IBD information. Hence, for simple analysis that aimed to get overview of genetic relationships like in this study, either maker based or combination of both approaches could be sufficient.

Individual pairs with half- and full-sib, parent-offspring and grandfather-grandson relationships constitute more than 72% of the 378 AI rams used in the panel development, indicating that the majority of the AI rams used in this study were highly related. All these relationships might put in question whether the AI rams are representative for the whole NWS population. High genetic relationships between individuals are expected to have huge impact on allele frequency. Hence, it is advisable to adjust allele frequency for the degree of relatedness observed among individuals in such kind of studies. However, allele frequency was not adjusted for the observed high level of relatedness in this study. We consider that alleles that are common among the AI rams will over time be common in the population due to extensive use of artificial insemination. Also, an attempt was made to evaluate robustness of the panel under hypothetically set values of MAF (Figure 6). Finally, power of the panel in paternity assignment and sib ship inference was tested (see below).

4.2 Highly informative marker selection criteria and their impact on power of exclusion

Parentage testing and individual identification are essential for efficient management of animal populations and for assuring food identity (Werner *et al*. 2004; Hara *et al*. 2010). This

can be achieved by developing and implementing an efficient DNA markers panel. Efficiency of the panel depends on informativeness of the markers, which is primarily depends on MAF and independency of the other loci constituting the panel. In parentage analysis information over many loci are combined to get reasonable power of exclusion. The loci could be in linkage disequilibrium (LD) so that the alleles at different loci could not assort independently. Such phenomenon decreases expected probability of exclusion due to reduced amount of genetic variation for determining parentage (Jones and Ardren, 2003). Extent of LD in the NWS population is unknown. Compared to cattle, most studies indicated persistence of LD over short chromosomal distance in different sheep populations although it varied across populations. For example, Meadows *et al*. (2008) reported short range LD (0-5cM) in five Australian sheep populations and García-Gámez *et al*.(2012) also reported persistence of LD over much more limited distances in Spanish sheep population than reported in dairy cattle. Assuming that the extent of LD in NWS population is similar to other ovine populations, the condition of ≥20 Mbp physical distance between the selected markers is most likely sufficient to avoid a negative effect on the power of parentage exclusion and probability of identity. The threshold value we used for physical distance (≥20Mbp) in this study is similar to that used in cattle for parentage testing based on SNP panel (Heaton *et al*. 2002; Werner *et al*. 2004; Hara *et al*., 2010).

However, recently emerging SNP panel for parentage testing and/or individual identity in ovine populations has put little/no emphasis on physical distance between markers residing on the same chromosomes. The International Sheep Genomics Consortium (ISGC), for example, released 89 SNP markers that are technically robust for parentage analysis in a wide variety of sheep breeds (Kijas *et al*. 2012). The SNPs in the ISGC panel that reside on the same chromosome are located very close to each other (≥0.1Mbp). Similarly, Paiva *et al.* (2011) used >3 Mbp distance between SNP marker in a panel developed for parentage and traceability testing in Brazilian sheep breeds. Why they paid low attention to physical distance between markers are not clear, but it might be related to lower extent of LD in sheep compared to that in cattle.

Power of parentage exclusion is calculated based on population allele frequencies, which is preferred to be estimated from unrelated or distantly related individuals. In this study, highly informative SNP markers with high MAF (0.48 to 0.5) in the AI rams (Appendix Table 1) were used, unlike most other studies where MAF ≥0.2 has been used. As a result, the

theoretical power of the panel, for parentage exclusion and individual identification, was higher than previously reported panels with either SNP or microsatellite markers. However, the allele frequencies in the current study were estimated from closely related individuals in the population (Figure 1, Table 1). This might have a major negative impact on the power of the panel. Existences of either full-or half-siblings in the pool of candidate parents are known to be the worst condition in parentage analysis (Jones and Ardren, 2003). The currently used population was not free from such problems because, as mentioned, the dataset contained closely related individuals in which majority of them have parent-offspring or half-and full-sib relationships (Table 1). Therefore, if the allele frequencies in our study are adjusted for the high level of relatedness, MAF of the 68 markers will obviously drop below the values currently used and also the power of the panel will be reduced.

To investigate this, the effect of minor allele frequencies on the power of the panel was evaluated (Figure 6). As expected, the power increased with increasing MAF, up to 0.5. At the extreme case where MAF is 0.1, the power in PE_3 was drastically dropped to 67%, but power reduction in PE_1 and PE_2 was extremely low (almost zero). As the MAF increased, PE_3 approached asymptotically to one (Figure 6) and when MAF is $\geq$0.25, PE_3 power reached 99% probability of exclusion, given false parentage. In other word, reduction in MAF, for example due to correction for the degree of relatedness, almost to half of the estimated value (0.48 to 0.5) was found to have little effect on power of parentage exclusion. This indicates that the panel developed in this study is relatively robust in accounting for possible effects of the suboptimal conditions (related AI rams) for estimating the "population allele frequency". Unlike PE_3 power, PE_1 and PE_2 power was hardly affected by the lowest MAF and their power was more or less similar for any values of MAF used (Figure 6); indicating that less informative markers can be used in parentage analysis when genotypes of both suspected parents are known.

In addition to informativeness, numbers of marker have also influenced power of the panel (Figure 7, Table 4). Number of markers needed to achieve reasonable power depends on the scenario under which the power is calculated. In the situation where both parents are genotyped, smaller numbers of markers (22 to 33 markers) than where only one parent is genotyped (40 markers) are required to achieve $\geq$99.5% power (Table 4). Therefore, markers with excellent technical performance ($\geq$95% call rate) can be used for parentage exclusion when both parents are genotyped and for individual identification when identity is lost.

Overall, around 69% of the 68 markers were technically performed well (≥85 call rate). Call rate per SNP and per individuals was higher in the Spael than in the NWS (Figures 4 and 5). This result indicates the existence of differences in DNA quality used for genotyping between the two breeds, since similar technical procedures were used on samples from both breeds. In fact, difference in DNA quality could also be experienced between samples within a breed and could contribute to differences in technical performance among SNPs. Problems related to DNA quality can be evident from lack of genotypes at many loci for an individual or group of individuals. For example, about 14% of the individuals in NWS had no genotype in 22 to 46% of the loci, while only 2.75% of the individuals in Spael had missing genotype in 20 to 28% of the genotyped SNPs (Figure 5). On other hand, the poor performance observed for some SNPs could also be due to problems related to technical procedures, as there were SNPs that performed poorly in both breeds. Moreover, poor technical performance might also be due to presence of unknown polymorphisms at the primers binding sites (Heaton *et al.* 2002) and such phenomenon might lead SNPs to fail to or poorly hybridize to their target sequences.

4.3 Pedigree error rate

The 4.9% mismatch rate found among the AI rams in this study was relatively small and comparable to pedigree errors reported for four sheep breeds in France (1 to 10%, Leroy *et al*. 2011) and for New Zealand sheep flocks (0.5 to 9.4%, Crawford *et al*. 1993). However, the error rate found in Spael breed was higher than that in NWS and in other sheep breeds reported in literature. In several studies of cattle (Geldermann *et al*. 1986; Banos *et al.* 2001; Baron *et al.* 2002; Visscher *et al*. 2002), between 4 to 36% mismatch rates were found based on microsatellite markers. Mismatch rates in Spael sheep was 2.7 times larger than that estimated in NWS population. These results show the need for more accurate ways of animal identification and parentage information. Several studies in cattle with simulation and real data showed that mismatch rate up to 5% have minimal effect on the estimation of genetic values, but the effect increased with increasing error rates (Van Vleck, 1970a). Hence, the pedigree error rate observed in this study (4.9%) expected to have minimal effect on the genetic evaluation of NWS population. Moreover, it is interesting to note that the Norwegian Sheep and Goat Breeders Association have already recognized existences of pedigree errors in the NWS population data used in this study. Such observation confirmed accuracy/power of the panel in excluding falsely alleged paternity in the NWS population.

Error sources are discussed in literature, such as mislabelling of semen straws by AI companies, incorrect identification of semen straws by AI technicians, errors in mothering up and recording, inadvertent misbreeding and use of multiple breeding males (Baron *et al*. 2002; Visscher *et al*. 2002). Some of these factors are expected to be controlled through good recording and verification systems, keeping single ID throughout animal life and through some quality control measures taken by AI companies (Visscher *et al*. 2002). Moreover, DNA based parentage identification was recommended as best solution to overcome such problems.

In this study, exclusion of paternity was based at least on one marker mismatch. This is similar to paternity studies in humans and other livestock that based on the SNP system. However, in human paternity studies with microsatellites, at least 2-3 markers mismatch are required for exclusion. This is because microsatellites have high mutation rate compared to SNP markers (Tishkoff *et al*. 2003).

4.4 Parentage exclusion and assignment

The power of parentage exclusion estimates the probability that the markers will exclude a non-parent from paternity or maternity or both. This parameter was estimated for three possible scenarios, i.e. when two parents are genotyped, and the intention is to exclude one parent (PE1) or both parents (PE2) and when only one parent is genotyped (PE3). Overall, the powers were low for individual marker analysed (Appendix Table 2), despite the fact that each marker had high MAF and relatively high PIC (0.375, where the maximum PIC is 0.5 for SNP). The low power for each marker is clearly due to the bi-allelic nature of SNP. Relatively, the power in the third scenario (PE3) was lower than in the others (Appendix Table 2), which is due to lack of genotype information from maternal side. The powers were almost the same for every marker within a scenario, but strictly spoken, OAR2_25624172.1 was the least powerful marker whereas OAR6_96871879.1 was the most powerful one in all scenarios (Appendix Table 2). After technical testing, however, this most power marker was excluded from the panel due to high rate of Mendelian error observed at this locus in 20% of the Spael families studied.

Combined powers of parentage exclusion measure the power of the panel in detecting falsely accused parentages and were estimated under two conditions: with the 68 theoretically feasible markers and with different number of technically feasible markers. After technical testing, markers were categorized in to different groups (e.g. in to 56, 48 and 40 markers),

based on their call rate (Table 4). With the theoretical panel (68 markers), when only one parent was known (PE3), the power was 0.999886. The panel with 40 well performing markers (≥90% call rate) efficiently excluded wrongly assigned father from paternity in both breeds (PE3=0.995206). Considering the level of technical performance observed and the possible failure in detecting mismatches with lower marker numbers, a panel with at least 48 markers should be used in real life testing. This result shows that the panel with lower numbers of markers (e.g. 40/48 markers), will still provide adequate power for paternity exclusion (PE3), but can be improved with the addition of new markers. These numbers can be increased by improving the technical performance of the poorly performing SNPs or by recruiting new markers that meet the selection criteria used in this study.

The combined estimate of power of exclusion is a generally useful parameter, but it assumes random mating and no relatedness between the alleged parent tested and the true parent (Souza *et al*. 2012). When related rams (half-sibs or even full-sibs) are used for breeding, which is a common practice in NWS breeding, false paternity might not be revealed when half-sib or full-sib of the alleged ram is the true father. In such situations, highly powerful markers panel is needed to precisely determine paternity, particularly when only one parent (ram) is genotyped (PE3), which is the most common situation in many breeding schemes. Our results with real data showed that the panel with ≥40 markers precisely excluded paternity (PE3) even when the candidate parents were a collection of half- or full-sib individuals. This result indicated that the panel has high capability in discriminating two closely related individuals. This can be evident form the estimated probability of identity that raged from $1.06x10^{-17}$ to 1.09×10^{-29} for ≥40 markers (Table 4), indicating almost zero chance of observing two individuals with identical genotypes at least at 40 loci.

In parentage analysis, exclusion is based on Mendelian rules of inheritance where incompatibilities between parents and offspring are used to reject particular parent-offspring hypotheses (Jones and Ardren, 2003). However, genotyping errors, null alleles and mutations can contribute to false exclusions. Hence, parentage assignment approach that is based on likelihood scores derived from parent and offspring genotypes can be an alternative under such condition. This is because the likelihood-based allocation methods usually allow for some degree of transmission errors due to genotyping error or other art facts (Jones and Ardren, 2003).

Paternity assignment analysis was conducted with two groups of candidate fathers (88 and 227 rams) and 185 offspring. In the assignment analysis with 88 candidate fathers, 97.3% rate of assignment was obtained. In the analysis with 227 candidate fathers rather different results were observed: 1) paternity assignment rate decreased by 14.6% 2) wrong assignments were observed and 3) time of analysis more than doubled (2 vs 4 days). This second analysis was actually a difficult situation, where numbers of candidate fathers were much larger than number of offspring tested. The result suggests that an increment in number of candidate fathers to be tested will not necessarily improve the rate of assignment. Rather the opposite could happen if large numbers of highly related individuals are included as candidate fathers. This is because likelihood scores calculated from genotypes of closely related animals may be very similar. As a result, half- or full-sibs of a father could be assigned as the true father. Also, the program may not be sensitive enough in discriminating those likelihood scores when assigning parentage, leading to lack of assignments. However, given the power of this panel, it is less probable to assign half-or full-sibs of a father as the true father. From inspection of the pedigree file, none of the true father and assigned father in the wrong assignments has close relationships. This suggests that it is not necessarily the increase in relationships with increasing number of candidate fathers that resulted in false assignments, but rather the much larger number of candidate fathers itself because the program may not be designed for such unusual condition.

In either paternity exclusion or assignment approach as well as under hypothetically set MAF, the panel showed consistent performance. Therefore, it can be concluded that this panel is robust and can efficiently exclude or assign parentage in the NWS or Spael population. Given their informativiness, these SNP markers can also be used as initial set for other sheep populations and our approach can also be used for developing a universal SNP panel to be used across sheep populations in the world. Moreover, these SNP markers may also provide some information in detecting population stratification in sheep genetic association studies because they are highly informative and believed to be in linkage equilibrium.

5. Conclusions

We have generated a SNP panel for paternity testing and individual identification in NWS consisting of 68 highly informative markers. After technical testing using the Sequenom iPLEX Gold System, 48 markers appeared to perform technically well. The panel has a very high theoretical power to uniquely identify every individual in the population. Similarly, it has high power of parentage exclusion under all the three possible scenarios considered. Practically, when genotypes from only one parent and offspring are known, the panel accurately excluded wrongly assigned fathers from paternity and also correctly assigned paternity to their biological progenies. Therefore, the results of this study can immediately be used in the parentage testing practice to provide parentage verification and thereby contribute to improved management of NWS breeding system.

6. References

Banos G., Wiggans G.R., Powell R.L. (2001) Impact of paternity errors in cow identification on genetic evaluations and international comparisons. *Journal of Dairy Science* **84**:2523–2529.

Baron E.E., Martinez M.L., Verneque R.S., Coutinho L.L. (2002) Parentage testing and effect of misidentification on the estimation of breeding value in Gir cattle. *Genetics of Molecular Biology* **25**, 389-394.

Botstein D., White R.L., Skolnick M.H., Davies R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314–31.

Cherny S.S., Abecasis G.R., Cookson W.O.C., Sham P.C., Cardon L.R. (2001). The effect of genotyping and pedigree errors on linkage analysis: Analysis of three Asthma genome scans. Genetic Epidemiology 21(Suppl 1), S117-S122.

Crawford A.M., Tate M.L., McEwan J.C., Kumaramannickavel G., McEwan K.M. et al. (1993). How reliable are sheep Pedigree? Proceeding of the New Zealand Society of Animal Production 53: 363-370

Eikje L.S., A°dnøy T., Klemetsda G. (2008) The Norwegian sheep breeding scheme: description, genetic and phenotypic change. *Animal* **2(2)**, 167–176.

Epstein M.P., Duren W.L., Boehnke M. (2000) Improved Inference of Relationship for Pairs of Individuals. American Journal of Human Genetics 67, 1219-1231.

García-Gámez E., Sahana G., Gutiérrez-Gil B., Arranz J.J. (2012) Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. *BMC Genetics* 13:43 doi:10.1186/1471-2156-13-43

Geldermann H., Pieper U., Weber W.E. (1986). Effect of misidentification on the estimation of breeding value and heritability in cattle. *Journal of Animal Science* **63**:1759–1768.

Gomez-Raya L., Priest K., Rauw W.M., Okomo-Adhiambo M., Thain D.,et al. (2008) The value of DNA paternity identification in beef cattle: Examples from Nevada's free-range ranches. *Journal of Animal Science* **86**:17-24.

Hara K., Watabe H., Sasazaki S., Mukai F., Mannen H. (2010) Development of SNP markers for individual identification and parentage test in a Japanese Black cattle population. *Animal Science Journal* **81,** 152–157.

Heaton M.P., Harhay G.P., Bennett G.L., Stone R.T., Grosse W.M., et al. (2002) Selection and use of SNP markers for animal identification and paternity analysis in U.S. beef cattle. *Mammalian Genome* **13**, 272–281.

http://www.beefcrc.com/publications/fact-sheets.html. Utilizing DNA for Parentage Verification Fact sheet: CRC FOR BEEF GENETIC TECHNOLOGIES.

Israel C., Weller J.I. (2000) Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *Journal of Dairy Science* **83**:181–187.

Jamieson A, Taylor S.C. (1997) Comparisons of three probability formulae for parentage exclusion. *Animal Genetics* **28,** 397–400.

Jones A.G., Ardren W.R. (2003) Methods of parentage analysis in natural populations. *Molecular Ecology* **12**, 2511–2523.

Jones O., Wang, J. (2009) COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10**: 551–555.

Kijas J., McEwan J., Clarke S., Henry H., Maddox J. et al. (2012) Development of a SNP panel for parentage assignment in Sheep. International Sheep Genomics Consortium (http://sheephapmap.org/news/PAG_2012_ParentagePoster.pdf).

Leroy G., Danchin-Burge C., Palhiere I., Baumung R., Fritz S. et al. (2011) An ABC estimate of pedigree error rate: application in dog, sheep and cattle breeds. Animal Genetics 43, 309–314.

Long T.E., Johnson R.K., Keele J.W. (1990) Effects of errors in pedigree on three methods of estimating breeding value for litter size, backfat and average daily gain in swine. *Journal of Animal Science,* **68(12)**, 4069-4078.

Meadows J.R.S., Chan E.K.F., Kijas J.W. (2008). Linkage disequilibrium compared between five populations of domestic sheep. *BMC Genetics* 9:61 doi:10.1186/1471-2156-9-61.

Nei M. (1978) Estimation of average heterozygosity and genetic distance from asmall number of individuals. *Genetics* **89**, 583-590.

Paiva S.R., Silva M.V.G.B., Andrade L.G., Guedes E., McManus C. et al. (2011) Development of a low density SNP panel for parentage and traceability testing in Brazilian sheep breeds for optimization of breeding and conservation programs. Plant & animal genomes xix conference (p551).

Parlato E., Van Vleck L.D. (2012) Effect of parentage misidentification on estimates of genetic parameters for milk yield in the Mediterranean Italian buffalo population. *Journal of Dairy Science* **95** :4059–4064.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., et al. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.

Sanders K., Bennewitz J. Kalm E. (2006) Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. *Journal of Dairy Science* **89**, 315-321.

Souza C.A., Paiva S.R., McManus C.M., Azevedo H.C., Mariante A.S., et al. (2012) Genetic diversity and assessment of 23 microsatellite markers for parentage testing of Santa Inês hair sheep in Brazil. *Genetics and Molecular Research* **11 (2),** 1217-1229.

Stevens E.L., Heckenberg G., Roberson E.D.O., Baugher J.D., Downey T.J., et al. (2011) Inference of Relationships in Population Data Using Identity-by-Descent and Identity-by-State. PLoS Genet 7(9): e1002287. doi:10.1371/journal.pgen.1002287.

Tishkoff S.A., Verrelli B.C. (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu. Rev. Genomics. *Human Genetics* **4**, 293–340.

Van Vleck L.D. (1970a) Misidentification in estimating the paternalsib correlation. *Journal of Dairy Science* **53**,1469–1474.

Van Vleck L.D. (1970b) Misidentification and sire evaluation. *Journal of Dairy Science* **53**,1697–1702.

Vatn S. (2009)The sheep industry in the Nordic countries. Journal of Small Ruminant Research 86, 80-83.

Visscher P.M., Woolliams J.A., Smith D., Williams J.L. (2002) Estimation of Pedigree Errors in the UK Dairy Population using Microsatellite Markers and the Impact on Selection. *Journal of Dairy Science* **85**:2368–2375.

Werner F.A.O., Durstewitz G., Habermann F.A., Thaller G., Kra¨mer W., et al. (2004) Detection and characterization of SNPs useful for identity control and parentage testing in major European dairy breeds. *Animal Genetics* **35**, 44–49.

7. Appendices

Appendix 1: List of the 68 highly informative SNP marker and their characteristics

Table 1: Position and allele and genotype frequencies for the 68 SNP markers

SNP-ID	OAR	Position (bp)	Alleles (A1/A2)	Allele frequencies		Genotype frequencies		
				A1	A2	A1A1	A1A2	A2A2
s73618.1	01	4845086	A/G	0.5	0.5	0.2513	0.4974	0.2513
s03901.1	01	53672087	A/G	0.4987	0.5013	0.2619	0.4735	0.2646
OAR1_122906056.1	01	122906056	A/G	0.5	0.5	0.2567	0.4866	0.2567
OAR1_194765342.1	01	194765342	G/A	0.496	0.504	0.2646	0.4630	0.2725
OAR1_221021345.1	01	221021345	A/G	0.5	0.5	0.2414	0.5172	0.2414
OAR1_241150187.1	01	241150187	A/G	0.5	0.5	0.2487	0.5026	0.2487
OAR1_261538157.1*	01	261538157	G/A	0.4987	0.5013	0.2381	0.5212	0.2407
OAR1_281552674.1#	01	281552674	A/G	0.5	0.5	0.2566	0.4868	0.2566
OAR2_25624172.1*	02	25624172	A/G	0.4828	0.5172	0.2354	0.4947	0.2698
OAR2_143746835.1	02	143746835	A/G	0.5	0.5	0.2302	0.5397	0.2302
OAR2_182832798.1	02	182832798	A/G	0.5	0.5	0.2513	0.4974	0.2513
OAR2_253397498.1	02	253397498	A/G	0.4987	0.5013	0.2460	0.5053	0.2487
OAR3_26768007.1#	03	26768007	G/A	0.4974	0.5026	0.2593	0.4762	0.2646
OAR3_47567775.1	03	47567775	G/A	0.4841	0.5159	0.2354	0.4974	0.2672
OAR3_66101339.1	03	66101339	A/G	0.4987	0.5013	0.2460	0.5053	0.2487
OAR3_127775703.1	03	127775703	A/G	0.5	0.5	0.2354	0.5291	0.2354
OAR3_149319006.1#	03	149319006	A/G	0.4947	0.5053	0.2354	0.5185	0.2460
OAR3_156574444.1*	03	156574444	G/A	0.4974	0.5026	0.2381	0.5185	0.2434
OAR3_191006426.1*	03	191006426	A/G	0.5	0.5	0.2513	0.4974	0.2513
OAR3_239598117.1	03	239598117	A/G	0.5	0.5	0.2672	0.4656	0.2672
OAR4_21234684.1#	04	21234684	A/G	0.4987	0.5013	0.2487	0.5000	0.2513
OAR4_118010742.1	04	118010742	G/A	0.5	0.5	0.2460	0.5079	0.2460
s15703.1	05	1328422	G/A	0.4974	0.5026	0.2487	0.4974	0.2540
OAR5_26720623.1	05	26720623	C/A	0.4987	0.5013	0.2381	0.5212	0.2407
OAR5_76877234.1	05	76877234	A/G	0.5	0.5	0.2751	0.4497	0.2751
OAR5_105151936.1*	05	105151936	G/A	0.4974	0.5026	0.2593	0.4762	0.2646
OAR6_58614245.1	06	58614245	A/G	0.5	0.5	0.2513	0.4974	0.2513
OAR6_96871879.1	06	96871879	A/C	0.5	0.5	0.2249	0.5503	0.2249
OAR6_106006705.1	06	106006705	A/G	0.4907	0.5093	0.2540	0.4735	0.2725
OAR7_32487594.1	07	32487594	A/G	0.4987	0.5013	0.2513	0.4947	0.2540
OAR8_1203191.1	08	1203191	G/A	0.4987	0.5013	0.2593	0.4788	0.2619
s31851.1	08	25926605	G/A	0.4987	0.5013	0.2381	0.5212	0.2407
OAR8_95188605.1*	08	95188605	A/C	0.4974	0.5026	0.2566	0.4815	0.2619
OAR9_17222799.1*	09	17222799	G/A	0.4987	0.5013	0.2434	0.5106	0.2460
OAR9_55140044.1#	09	55140044	G/A	0.5	0.5	0.2646	0.4709	0.2646
OAR9_77545332_X.1#	09	77545333	G/A	0.4921	0.5079	0.2672	0.4497	0.2831

OAR9_99601991.1	09	99601991	A/G	0.4987	0.5013	0.2460	0.5053	0.2487
OAR10_3635864.1	10	3635864	C/A	0.4934	0.5066	0.2302	0.5265	0.2434
OAR10_57586299.1$	10	57586299	A/G	0.5	0.5	0.2302	0.5397	0.2302
s12970.1	10	88770435	A/G	0.4987	0.5013	0.2593	0.4788	0.2619
OAR12_33175150.1	12	33175150	A/C	0.4973	0.5027	0.2361	0.5225	0.2414
s49565.1	12	61344433	A/G	0.5	0.5	0.2698	0.4603	0.2698
OAR13_8268238.1#	13	8268238	A/G	0.5	0.5	0.2540	0.4921	0.2540
s24404.1#	13	31289320	G/A	0.5	0.5	0.2381	0.5238	0.2381
OAR13_56607666.1	13	56607666	A/G	0.5	0.5	0.2487	0.5026	0.2487
OAR13_87920961.1	13	87920961	C/G	0.496	0.504	0.2487	0.4947	0.2566
OAR14_7280304_X.1	14	7280305	A/C	0.496	0.504	0.2328	0.5265	0.2407
OAR15_6615347.1	15	6615347	C/A	0.4947	0.5053	0.2751	0.4392	0.2857
OAR15_50080570.1@	15	50080570	G/A	0.4907	0.5093	0.2434	0.4947	0.2619
OAR16_49974943.1	16	49974943	G/A	0.4987	0.5013	0.2599	0.4775	0.2626
s32327.1	17	3372466	G/A	0.5	0.5	0.2540	0.4921	0.2540
OAR17_33673034.1	17	33673034	A/G	0.496	0.504	0.2407	0.5106	0.2487
OAR18_1464804.1	18	1464804	G/A	0.4987	0.5013	0.2354	0.5265	0.2381
s46618.1	18	30382827	A/C	0.4947	0.5053	0.2619	0.4656	0.2725
OAR19_1987551.1	19	1987551	A/G	0.4974	0.5026	0.2143	0.5661	0.2196
OAR20_4451114.1	20	4451114	A/G	0.4987	0.5013	0.2169	0.5635	0.2196
OAR20_24583511.1	20	24583511	A/G	0.4854	0.5146	0.2434	0.4841	0.2725
OAR21_1349161.1	21	1349161	G/A	0.4947	0.5053	0.2420	0.5080	0.2500
OAR21_47788299.1*	21	47788299	G/A	0.4987	0.5013	0.2460	0.5053	0.2487
OAR22_1023592.1	22	1023592	A/G	0.5	0.5	0.2804	0.4392	0.2804
s60529.1	22	24732824	A/G	0.5	0.5	0.2698	0.4603	0.2698
OAR23_13626111.1#	23	13626111	A/G	0.5	0.5	0.2487	0.5026	0.2487
s35982.1	23	30551131	G/A	0.4987	0.5013	0.2751	0.4471	0.2778
s26911.1	23	61422821	A/G	0.5	0.5	0.2619	0.4762	0.2619
s63658.1	25	1119928	G/A	0.4921	0.5079	0.2354	0.5132	0.2513
OAR25_23096948.1#	25	23096948	A/G	0.4987	0.5013	0.2434	0.5106	0.2460
DU459122_429.1	25	44278877	C/A	0.5	0.5	0.2460	0.5079	0.2460
OAR26_32558263.1	26	32558263	A/G	0.5	0.5	0.2302	0.5397	0.2302

*#:SNPs with no signal i.e. with zero call rate; @:SNP with high call rate (>90) but monomorphic in Spael breed; $:SNP with unusual genotypes i.e. it has both types of homozygous genotypes (AA and GG), but no heterozygous genotype (AG); **: SNPs with poor signal or with 62 to 84% call rate in 2x40 genotyping by Sequenom.*

Table 2: Expected heterozygosity (E (H)), polymorphic information contents (PIC), power of exclusion and probability of identity (PI) for the 68 markers

SNP-ID	E(H)	PIC	Power of exclusions			PI
			PE1	PE2	PE3	
OAR1_122906056.1	0.50067	0.37500	0.18750	0.28125	0.12500	0.36858
OAR1_194765342.1	0.50063	0.37498	0.18749	0.28124	0.12498	0.35857
OAR1_221021345.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38407
OAR1_241150187.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37633
OAR1_261538157.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38626
OAR1_281552674.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.36865
s03901.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.36283
s73618.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37369
OAR2_143746835.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.39720
OAR2_182832798.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37369
OAR2_253397498.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37769
OAR2_25624172.1	0.50007	0.37470	0.18735	0.28107	0.12470	0.37299
OAR3_127775703.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.39082
OAR3_149319006.1	0.50061	0.37497	0.18749	0.28123	0.12497	0.38483
OAR3_156574444.1	0.50065	0.37499	0.18750	0.28125	0.12499	0.38479
OAR3_191006426.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37369
OAR3_239598117.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.35958
OAR3_26768007.1	0.50065	0.37499	0.18750	0.28125	0.12499	0.36396
OAR3_47567775.1	0.50016	0.37475	0.18737	0.28109	0.12475	0.37419
OAR3_66101339.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37769
OAR4_118010742.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37906
OAR4_21234684.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37500
OAR5_105151936.1	0.50065	0.37499	0.18750	0.28125	0.12499	0.36396
OAR5_26720623.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38626
OAR5_76877234.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.35366
s15703.1	0.50065	0.37499	0.18750	0.28125	0.12499	0.37370
OAR6_106006705.1	0.50049	0.37491	0.18746	0.28120	0.12491	0.36299
OAR6_58614245.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37369
OAR6_96871879.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.40392
OAR7_32487594.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37240
OAR8_1203191.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.36509
OAR8_95188605.1	0.50065	0.37499	0.18750	0.28125	0.12499	0.36627
s31851.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38626
OAR9_17222799.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38046
OAR9_55140044.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.36172
OAR9_77545332_X.1	0.50054	0.37494	0.18747	0.28121	0.12494	0.35378
OAR9_99601991.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37769
OAR10_3635864.1	0.50058	0.37496	0.18748	0.28122	0.12496	0.38936
OAR10_57586299.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.39720
s12970.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.36509

OAR12_33175150.1	0.50065	0.37499	0.18750	0.28125	0.12499	0.38705
s49565.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.35752
OAR13_56607666.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37633
OAR13_8268238.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37113
OAR13_87920961.1	0.50063	0.37498	0.18749	0.28124	0.12498	0.37243
s24404.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38776
OAR14_7280304_X.1	0.50063	0.37498	0.18749	0.28124	0.12498	0.38931
OAR15_50080570.1	0.50049	0.37491	0.18746	0.28120	0.12491	0.37257
OAR15_6615347.1	0.50061	0.37497	0.18749	0.28123	0.12497	0.35019
OAR16_49974943.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.36449
OAR17_33673034.1	0.50063	0.37498	0.18749	0.28124	0.12498	0.38049
s32327.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37113
OAR18_1464804.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38928
s46618.1	0.50061	0.37497	0.18749	0.28123	0.12497	0.35963
OAR19_1987551.1	0.50065	0.37499	0.18750	0.28125	0.12499	0.41464
OAR20_24583511.1	0.50024	0.37479	0.18739	0.28112	0.12479	0.36786
OAR20_4451114.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.41280
OAR21_1349161.1	0.50061	0.37497	0.18749	0.28123	0.12497	0.37912
OAR21_47788299.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37769
OAR22_1023592.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.35013
s60529.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.35752
OAR23_13626111.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37633
s26911.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.36395
s35982.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.35275
DU459122_429.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.37906
OAR25_23096948.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.38046
s63658.1	0.50054	0.37494	0.18747	0.28121	0.12494	0.38200
OAR26_32558263.1	0.50066	0.37500	0.18750	0.28125	0.12500	0.39720
Average/combined	0.50062	0.37498	0.999999	1	0.999886	1.09E-29

Table 3: Amplification and extension primers for the 68 markers

SNP_ID	Amplification primers (forward & backward) and product length (Len)			Hybridization temperature (Tm) & extension primer	
	Backward	Forward	Len	Tm	Primer sequences
DU459122_429_1	ACGTTGGATGTCTGTATGTCCAAAGGAGTC	ACGTTGGATGTTCAAACAGGGATGCACCAG	105	46.1	ggagGGAGTCTATTTGACTGTCT
OAR1_122906056_1	ACGTTGGATGAGCCTATATGTGTTTGTAG	ACGTTGGATGTAAAAGTAGGTACATCTGGG	81	48.4	cTATGTGTTTGTAGTAACAAAGGT
OAR1_194765342_1	ACGTTGGATGGCTCTGTTTTGAGGTATGCG	ACGTTGGATGGGATGCAGATGTAGACTACG	100	46.3	ggGGTATGCGGAAACCTAA
OAR1_221021345_1	ACGTTGGATGCTCAGAAGTACACCATCTGC	ACGTTGGATGCACTGTAAGAGTGATAACAGG	100	47.6	tCTCATCAACTAGATATCAGAACA
OAR1_241150187_1	ACGTTGGATGCAGAATGTTATTGCACCACAC	ACGTTGGATGGAAGAAAGTATCTGGACAAG	99	47.2	gGATTTCAGTAATTGCTTCTTCA
OAR1_261538157_1	ACGTTGGATGCGAAATGTACTGGTTTCTCC	ACGTTGGATGATGACCCTCCTTGAAACCTG	100	50.3	cctcAATGTACTGGTTTCTCCTCTTAAA
OAR1_281552674_1	ACGTTGGATGAGGACCCTTCTGAGTGTGAG	ACGTTGGATGACAGACAGAGAACAGGGTGC	113	46.8	gaccGGTCATCTCTCTGTGTTC
OAR10_3635864_1	ACGTTGGATGATGTGAACCCTGTTCTCACC	ACGTTGGATGAGGAACACAAAGGCACAGAC	99	53.1	acgttAACCCTGTTCTCACCATTCAC
OAR10_57586299_1	ACGTTGGATGATCAACTAAAGGGGACTCTC	ACGTTGGATGGACCAGTGGATAAAACAGGG	88	50.2	agAAAGGGGACTCTCGAGTTA
OAR12_33175150_1	ACGTTGGATGATACCCTGATGTCTCTGTGG	ACGTTGGATGGAGTGCATTCAATAGGTGGG	102	52.7	ggggaGTGGGGCAGGCCAAT
OAR13_56607666_1	ACGTTGGATGACTTCTTAACATCGGGGCAG	ACGTTGGATGCCTCAGCCAGATTCTCAATG	110	48.4	caAGCAGATGAGCCACC
OAR13_8268238_1	ACGTTGGATGTGTTTCTGTAGGCAACACGC	ACGTTGGATGCTGTCAGAATGAATTCAGGG	114	46.4	ggttTACTACTATTCGTCAAAGTACA
OAR13_87920961_1	ACGTTGGATGGTGGAGAACCAAGGACTTTC	ACGTTGGATGGTTCTTATCTGCTACAAACC	96	50	AAGGACTTTCTCAGCTGTG
OAR14_7280304_X	ACGTTGGATGATCCTGAGGTGTTAACAGCG	ACGTTGGATGCTGAGTTCCGGAAGATATAC	103	50.5	GCGAAAAATTTACTTCTCAAGTTTAATC
OAR15_50080570_1	ACGTTGGATGCCTGTTTAACAGCAAAGGAC	ACGTTGGATGACTCATGTTAGTCTTTGTCG	108	49.5	acagaAAAGGACTTAAGCTAACCGT
OAR15_6615347_1	ACGTTGGATGCTAGTTGTGACTTCATTTTG	ACGTTGGATGTGTCATTTCCACAAGGACCC	113	45.8	TACTTATTTAAAGAGTATGGCAA
OAR16_49974943_1	ACGTTGGATGGTCTTCCATCGTTCTGACTC	ACGTTGGATGTAATCTCCATCCTTCCTTGC	104	46.1	TTTACCATAACACAATCATAAGA
OAR17_33673034_1	ACGTTGGATGTCATGACTCTATCTGCTGGG	ACGTTGGATGAAGGAATGACAGCTGTTGAG	105	48.8	TGCCACCATCAAATCCTA
OAR18_1464804_1	ACGTTGGATGGAGCAGAACGGTGTGATTTC	ACGTTGGATGTAGTTCCCCATTCATCAGGC	100	46.3	TCCTTAGAAGCAAGCTG
OAR19_1987551_1	ACGTTGGATGTGCTGCTCTCTGTCCTAAAG	ACGTTGGATGATGAGGGAAGTCCTCTAATC	102	47	CTCTGTCCTAAAGTTTACTGA
OAR2_143746835_1	ACGTTGGATGTGCAGGTTACAGGGAATGAC	ACGTTGGATGCTAGCGGGTTTATTTCCCAG	94	54.8	ggtgACAGGGAATGACTCAGAAGCC
OAR2_182832798_1	ACGTTGGATGTGAATGGGTTGTCCACTTCC	ACGTTGGATGCCATATTGACTCACTCCAGG	94	45.5	taccACCGAATCCTTTTGGT
OAR2_253397498_1	ACGTTGGATGTTACCTTCCATCAGCACCAG	ACGTTGGATGATCTGCACCTTCCTGTTCTC	106	46.2	cccCAGCACCAGGATCAC
OAR2_25624172_1	ACGTTGGATGTGCACACCAGTCAGAAAGTC	ACGTTGGATGTTTCTGTCTTGAGTGAGGGC	113	58.2	ggatAGCACACCAAGATGCACACTGTC
OAR20_24583511_1	ACGTTGGATGTACTAGAAAACCGCCAGCAG	ACGTTGGATGGAGAGAGGGACTATTTTCCG	110	57.8	cccGGCTCCATGCTAAGAGCATTCACTT
OAR20_4451114_1	ACGTTGGATGGGAGGGTTTAGACTCTTCAC	ACGTTGGATGACTTTTTCCTGCATGGGCTC	101	48.1	gagtTCTCTGTGTAGAGCTCATAATA
OAR21_1349161_1	ACGTTGGATGGGATTCCTCTTCTGACTTCC	ACGTTGGATGGTGAAGTCTTACCAGGGTTG	89	49.6	aaTTCCTGAATTTTAACTGTAGCTTC
OAR21_47788299_1	ACGTTGGATGTCTCAGACACAGACACACAC	ACGTTGGATGTCAAAAATCCTGACCCTGGC	99	48.8	gacgCAACTATCTGATGTGCTGTG
OAR22_1023592_1	ACGTTGGATGTCAAATTTAGAGGAGGCAGG	ACGTTGGATGTTTTCCCCACCAGCATTCAC	90	51.2	ggtcGAGGAGGCAGGAAATTTAGATG
OAR23_13626111_1	ACGTTGGATGGAGGGAAAGAGGATGATAGC	ACGTTGGATGTGGGTTGCCATTTCCTTTCC	99	48	tgatcGCTTGGCAAGTTAAGAATTG
OAR25_23096948_1	ACGTTGGATGGGGTTGAGCCCTTATTGTAA	ACGTTGGATGCATCTAGGTCCACATAGATT	119	48.7	ccAACCATTGTAATTGCACTAGAA
OAR26_32558263_1	ACGTTGGATGCTTCCACCTCTGTGCCTTTA	ACGTTGGATGTTAGAAGGCTGGGAAGAAGG	90	49.4	gtcccCCTCTGTGCCTTTATTTGTG
OAR3_127775703_1	ACGTTGGATGTCCTTTCATCTTGAGCATTC	ACGTTGGATGAGCAATAAGGATTGTCTCCG	101	45.1	CATCTTGAGCATTCTAAAATC
OAR3_149319006_1	ACGTTGGATGCAAGATACTGTACCTTCAGC	ACGTTGGATGGTTCTGTTTGCAGAACAGAG	95	47	TCTTCTACTTGGTGTTATGTT
OAR3_156574444_1	ACGTTGGATGGTCCAGTGGTGTTAATCAGG	ACGTTGGATGTGGCATGCGAACTCTGACC	102	52.2	cgCAGGAAAGGGCAGGTAGTTAT
OAR3_191006426_1	ACGTTGGATGCCCAAATTGGAAACCAGGTG	ACGTTGGATGTTGTTCAAGACAAAGAAGG	91	48.9	ggcgTAGCCAGCTGGAAAAGTAT
OAR3_239598117_1	ACGTTGGATGGGAACTATTTCTCTCCTCCG	ACGTTGGATGCCTGTTCATCTCCACCCACA	84	50.8	gaagcGTGTCGTGAAAGCCCG
OAR3_26768007_1	ACGTTGGATGGTCCTATCTGTTCTTTCAGC	ACGTTGGATGTGTGGTTTTGGCCCATATCC	94	47.8	ggttTTCTTTCAGCTAGGGGAT
OAR3_47567775_1	ACGTTGGATGGATACCAGGATTCCACCTTC	ACGTTGGATGTTTCAGCTCTTCCACTCTCC	103	51.7	aacgCCACCTTCCCTGGGTAATAA
OAR3_66101339_1	ACGTTGGATGTGCAGTTTTACCCTCGTGTG	ACGTTGGATGCACTGTTTCTGGTTTCAGCC	105	46.3	tCTCGTGTGGCTTTGT
OAR4_118010742_1	ACGTTGGATGTACATGGGAGAAGAGAAGTC	ACGTTGGATGTAGTCAAGCAGCCACTAAAC	96	47.6	cGAATGATCTGACGGTGTC
OAR4_21234684_1	ACGTTGGATGCATTAAAATGGTTCATTGAGG	ACGTTGGATGAAACCTACTCAGGTTTCCCC	91	47	GTTCATTGAGGAAAACTTAGG
OAR5_105151936_1	ACGTTGGATGCCGAAATGCCAAGAACACTG	ACGTTGGATGGTATGAAGCTTAAGGATGTG	109	50	ggagtCCAAGAACACTGGTATGTCA
OAR5_26720623_1	ACGTTGGATGGAGATGAGTACTGCACAGAG	ACGTTGGATGCTGACCGTCTTTGCAAAGTG	110	45.3	ctCTATAGTATAAGCCTTTCTTTCTA
OAR5_76877234_1	ACGTTGGATGTCTTGTGCTGCACATTGCTC	ACGTTGGATGCTCTTGGTAAATAAAGGGA	110	46.3	GCACATTGCTCTATTAATGTA
OAR6_106006705_1	ACGTTGGATGGCATATAGAGCTCCCCAAAG	ACGTTGGATGCACGTAAAATCCATGGCTAC	99	51.2	CTCCCCAAAGTTCAATAATGATCT
OAR6_58614245_1	ACGTTGGATGCCTATTGATCATGGGACAGC	ACGTTGGATGAACAGCAACAAAAATCTGCC	85	47.2	GGACAGCATAGTAAAATATATGAC
OAR6_96871879_1	ACGTTGGATGGAAGTGAAGGGTAAGCAGTC	ACGTTGGATGAGGCTCTCTTTGCAAAGCTG	99	50.9	aaccAGCAGTCTCCTGGGC
OAR7_32487594_1	ACGTTGGATGTCAAAGGATACTTCTGAGAG	ACGTTGGATGAAGTCAGTGGTCTGACTTGG	98	48.6	CTGAGAGTAAAGAAAGTTCAGG
OAR8_1203191_1	ACGTTGGATGTATTCCTCACCTTAGCACCC	ACGTTGGATGTCAGCTCCTGACAGAGTTTG	112	47.6	CTTTTCTGTATATCTGCATCAATAA
OAR8_95188605_1	ACGTTGGATGCCATAGTCACAGTATGGCTC	ACGTTGGATGCCTTGTTACTTCATCCTGGC	101	52.5	ggttaACAGTATGGCTCAGCTAAAGAC
OAR9_17222799_1	ACGTTGGATGAGCTCACGCTTCTTACAGAG	ACGTTGGATGAACCCATCAGTGGCATCTTG	107	53	cttaAACATCCTTGGTATTAGCTTGTGT
OAR9_55140044_1	ACGTTGGATGGATGAGATCAGTGATAGAGG	ACGTTGGATGTCCACTGTGAAACCTTTAAC	99	46.7	tatcTGGGTAATGATAACATACCATA
OAR9_77545332_X	ACGTTGGATGCTCAATTTATGCTTTACTC	ACGTTGGATGCCACTGAATTGTATACCAAGG	98	46.7	TCAATTTATGCTTTACTCAAAATTAA
OAR9_99601991_1	ACGTTGGATGACTGTGGCAGAGCAAAAGTG	ACGTTGGATGACAACACACCCAAAACAC	112	50.2	aaacGAGCAAAAGTGGTTTTAAGGTT
s03901_1	ACGTTGGATGCCCGATAGCAGGAAGTCCG	ACGTTGGATGGAGGGCTCATAGACAATGTG	106	52.1	tccCCGCACCCGCGATAT
s12970_1	ACGTTGGATGTAACAGTCACTTGAGGCCAG	ACGTTGGATGCAGTAAGAACTAGACAGGAG	83	47.1	gtCCAGTCACCTGGGAT
s15703_1	ACGTTGGATGGTTCCAAAGAGCTTTGCCAG	ACGTTGGATGGGGTGGTGATCTAGAAACTG	83	45	ccaCAGGGCCTTTCTAGG
s24404_1	ACGTTGGATGCAGCATCACGTGTCATCAAG	ACGTTGGATGCTTAAATGGGATGGTGTGGC	98	45.7	ggtgAGATGCTGAGATGGAATAT
s26911_1	ACGTTGGATGAGTAAGTAGGCGCTGTTAGG	ACGTTGGATGCTGCCTAGACGAAGAGTAAG	114	50.2	aaGGGGACGAGGGTAACT
s31851_1	ACGTTGGATGAGTAATTAGCAGGCCAAGGG	ACGTTGGATGCCAAAACAGACATGTGGTGC	99	55.9	GGCCAAGGGGTTCAAGGAG
s32327_1	ACGTTGGATGTCTTCAATGAGCCAAGGAGC	ACGTTGGATGTGAGCATCTTAGCTTGTGCC	104	49.1	GGTTAAAGAAGTCCCCGTA
s35982_1	ACGTTGGATGCAGATGGGTTAATCACGTGC	ACGTTGGATGAGCCAGAAGGGCCAACAAG	99	54	ttccCTGTCCAGGACCTGCCT
s46618_1	ACGTTGGATGTGGCCTTTTCAGGTTCATGC	ACGTTGGATGTCCTCACTCTTACTGACTAC	102	46.1	CAGGTTCATGCTTGTTAC
s49565_1	ACGTTGGATGCACTCAAAGCCTCAAGAGAG	ACGTTGGATGATACGCAGCTGAGTGGCCAG	109	50.9	CTCAAGAGAGCTCCCAGT
s60529_1	ACGTTGGATGAACACTGAGTCAGAGGTGAG	ACGTTGGATGAATCCTCCTACCCTCTTCTG	112	49	gGCAGAAGTGCTGGGT
s63658_1	ACGTTGGATGTTCAGTCTGATCCTCTTCGG	ACGTTGGATGGATCTGAGAAGCCAGATCGG	115	52.1	ggaaAGGAGGGCGACTCTGTA
s73618_1	ACGTTGGATGCTCTGCTGCTCAGTTCTCCA	ACGTTGGATGCTGGGAGGTATGTGATCAAG	96	52.1	GTTCTCCAGCCCCCG

Appendix 2: Paternity exclusion and assignment out puts

Table 1: Number and proportion of incompatible loci between parental and offspring genotypes with the 68 markers in NWS

Individual ID	Father ID	Discords	Discord%	Individual ID	Father ID	Discords	Discord%
8967030	8676042	0	0.0	9859303	9427694	0	0.0
9067747	8670078	0	0.0	9859558	9655410	0	0.0
9070166	8575795	0	0.0	9862104	9427584	0	0.0
9113940	8765624	0	0.0	9862107	9660487	0	0.0
9121257	8665564	0	0.0	9867852	9483136	0	0.0
9121869	8923781	0	0.0	9880057	9680408	0	0.0
9170058	8970084	0	0.0	9920250	9626010	0	0.0
9170130	8575795	2	2.94	9920372	9626010	0	0.0
9225657	8927123	0	0.0	9920427	9521134	0	0.0
9260589	9061241	0	0.0	9920655	9626010	0	0.0
9260609	9061241	0	0.0	9929297	9626010	0	0.0
9326724	9129025	0	0.0	9940537	9540149	5	7.35
9413485	8927123	0	0.0	9950434	9525234	0	0.0
9424040	8927123	0	0.0	9955039	9757403	0	0.0
9427619	9225657	0	0.0	9955191	9525234	0	0.0
9525367	9326149	0	0.0	9955503	9626010	0	0.0
9537732	9129025	0	0.0	9955536	9626010	0	0.0
9540364	8927123	0	0.0	9955616	9525234	0	0.0
9621311	9323103	0	0.0	9955890	9626010	0	0.0
9670164	9470115	0	0.0	9956030	9525234	0	0.0
9680408	9483338	0	0.0	9965715	9660487	0	0.0
9713398	9413040	0	0.0	9970207	9626010	0	0.0
9713505	9513199	0	0.0	9975279	9626010	0	0.0
9713670	9313625	0	0.0	9981260	9525234	0	0.0
9723427	9521134	0	0.0	200025044	9828543	0	0.0
9727192	9427694	0	0.0	200025590	9764637	0	0.0
9740277	9121869	0	0.0	200040506	9680408	0	0.0
9758336	9260589	0	0.0	200040787	9540364	0	0.0
9812772	9427694	0	0.0	200059904	9626010	0	0.0
9813194	9613555	0	0.0	200061019	9727192	0	0.0
9858402	9427694	0	0.0	200061098	9427694	0	0.0
9859266	9427694	7	10.29	200061130	9513199	0	0.0
200061587	9626010	0	0.0	200428176	200180253	0	0.0
200075677	9626010	0	0.0	200433272	200025086	0	0.0
200075680	9626010	0	0.0	200438811	200040787	0	0.0
200113693	9626025	0	0.0	200440007	200025686	8	11.76
200122511	9920372	0	0.0	200440353	200240529	0	0.0
200137594	9660487	0	0.0	200440772	200241045	0	0.0
200156175	9626010	0	0.0	200441336	200059904	0	0.0

200175101	9655410	0	0.0	200450581	200040787	0	0.0
200175105	9655410	0	0.0	200458810	200040787	0	0.0
200175276	9975555	0	0.0	200458811	200040787	0	0.0
200180663	9981260	0	0.0	200458833	200025086	0	0.0
200180736	9513199	0	0.0	200458977	200075328	0	0.0
200225030	200025086	0	0.0	200459117	200075328	0	0.0
200225101	200025044	0	0.0	200460546	200025086	0	0.0
200240495	9680526	0	0.0	200470217	200040787	0	0.0
200240658	9626010	0	0.0	200475334	200025686	0	0.0
200241045	9626010	0	0.0	200480064	200040787	0	0.0
200256375	9955039	0	0.0	200480159	200040787	0	0.0
200275503	9970207	0	0.0	200480410	200280236	0	0.0
200280193	200180663	0	0.0	200480432	200280193	0	0.0
200320103	200040787	0	0.0	200480509	200025086	0	0.0
200320308	200040787	0	0.0	200513424	200213379	0	0.0
200340400	9981260	0	0.0	200522071	200061587	0	0.0
200341279	200075331	0	0.0	200522557	200025086	0	0.0
200353873	200040787	7	10.29	200529095	200225101	0	0.0
200354401	200040787	0	0.0	200529131	200126549	0	0.0
200358582	200040787	0	0.0	200529412	200327021	0	0.0
200364329	9981260	0	0.0	200540022	200025086	0	0.0
200380115	200280236	0	0.0	200540789	200280236	0	0.0
200421174	200075328	0	0.0	200541333	200341279	0	0.0
200421240	200040787	0	0.0	200541366	200061587	0	0.0
200421779	200061130	6	8.82	200555843	200025086	0	0.0
200555868	200358582	0	0.0	200765139	200327474	0	0.0
200555907	200180253	0	0.0	200770052	200570047	0	0.0
200555911	200280236	0	0.0	200775561	200275503	0	0.0
200555937	200061587	0	0.0	200780235	200460513	0	0.0
200563404	200025086	8	11.76	200811546	200529131	0	0.0
200565312	200280236	0	0.0	200820347	200421240	0	0.0
200570223	200280239	0	0.0	200820362	200623129	0	0.0
200575118	200375532	0	0.0	200828058	200480159	0	0.0
200575848	200180779	0	0.0	200828148	200563486	0	0.0
200580576	200025086	0	0.0	200835027	200421240	0	0.0
200623128	200126549	0	0.0	200840606	200555911	0	0.0
200623129	200126549	0	0.0	200841235	200555911	0	0.0
200639428	200126549	0	0.0	200851927	200563486	0	0.0
200639578	200438811	0	0.0	200856019	200555911	0	0.0
200653003	200126549	0	0.0	200856724	200755193	7	10.29
200653015	200480410	0	0.0	200856726	200522071	14	20.59
200653176	200480410	0	0.0	200856749	200755193	0	0.0
200653342	200126549	0	0.0	200863141	200275503	0	0.0
200653864	200240547	0	0.0	200863271	200421240	0	0.0
200661235	200126549	0	0.0	200863545	200555911	0	0.0

200724160	200470217	0	0.0	200863569	200555911	0	0.0
200724316	200234397	0	0.0	200870108	200522071	0	0.0
200727152	200529119	0	0.0	200870219	200555911	0	0.0
200727153	200529119	0	0.0	200875287	200421240	0	0.0
200727413	200275503	0	0.0	200880369	200522071	0	0.0
200727472	200275503	0	0.0				
200740102	200470024	0	0.0				
200740129	200458811	0	0.0				
200750095	200380115	0	0.0				
200750174	200126549	0	0.0				
200755193	200275503	0	0.0				
200765133	200327474	0	0.0				

In this table, there are nine genotypes mismatches out of the 185 parent-offspring pairs evaluated (4.9%) i.e. nine offspring didn't get their true fathers The nine wrong pairs are summarized below in Table 3.

Table 2: Paternity assigned offspring and probability of the assignment

OffspringID	FatherID	Prob	OffspringID	FatherID	Prob	OffspringID	FatherID	Prob
9121869	8923781	1	200623129	200126549	1	9867852	9483136	1
9225657	8927123	1	200755193	200275503	1	9880057	9680408	1
9260589	9061241	1	8967030	8676042	1	9920250	9626010	1
9540364	8927123	1	9067747	8670078	1	9920427	9521134	1
9680408	9483338	1	9070166	8575795	1	9920655	9626010	1
9727192	9427694	1	9113940	8765624	1	9929297	9626010	1
9920372	9626010	1	9121257	8665564	1	9940537	*2	NA
9955039	9757403	1	9170058	8970084	1	9950434	9525234	1
9970207	9626010	1	9170130	*1	NA	9955191	9525234	1
9981260	9525234	1	9260609	9061241	1	9955503	9626010	1
200025044	9828543	1	9326724	9129025	1	9955536	9626010	1
200040787	9540364	1	9413485	8927123	1	9955616	9525234	1
200059904	9626010	1	9424040	8927123	1	9955890	9626010	1
200061130	9513199	1	9427619	9225657	1	9956030	9525234	1
200061587	9626010	1	9525367	9326149	1	9965715	9660487	1
200180663	9981260	1	9537732	9129025	1	9975279	9626010	1
200225101	200025044	1	9621311	9323103	1	200025590	9764637	1
200241045	9626010	1	9670164	9470115	1	200040506	9680408	1
200275503	9970207	1	9713398	9413040	1	200061019	9727192	1
200280193	200180663	1	9713505	9513199	1	200061098	9427694	1
200341279	200075331	1	9713670	9313625	1	200075677	9626010	1
200358582	200040787	1	9723427	9521134	1	200075680	9626010	1
200380115	200280236	1	9740277	9121869	1	200113693	9626025	1
200421240	200040787	1	9758336	9260589	1	200122511	9920372	1
200438811	200040787	1	9812772	9427694	1	200137594	9660487	1
200458811	200040787	1	9813194	9613555	1	200156175	9626010	1
200470217	200040787	1	9858402	9427694	1	200175101	9655410	1
200480159	200040787	1	**9859266**	**9427584**	1	200175105	9655410	1
200480410	200280236	1	9859303	9427694	1	200175276	9975555	1
200522071	200061587	1	9859558	9655410	1	200180736	9513199	1
200529131	200126549	1	9862104	9427584	1	200225030	200025086	1
200555911	200280236	1	9862107	9660487	1	200240495	9680526	1
200240658	9626010	1	200541333	200341279	1	200765139	200327474	1
200256375	9955039	1	200541366	200061587	1	200770052	200570047	1
200320103	200040787	1	200555843	200025086	1	200775561	200275503	1
200320308	200040787	1	200555868	200358582	1	200780235	200460513	1
200340400	9981260	1	200555907	200180253	1	200811546	200529131	1
200353873	**9970207**	1	200555937	200061587	1	200820347	200421240	1
200354401	200040787	1	**200563404**	**200225058**	1	200820362	200623129	1
200364329	9981260	1	200565312	200280236	1	200828058	200480159	1
200421174	200075328	1	200570223	200280239	1	200828148	200563486	1
200421779	*1	NA	200575118	200375532	1	200835027	200421240	1

200428176	200180253	1	200575848	200180779	1	200840606	200555911	1
200433272	200025086	1	200580576	200025086	1	200841235	200555911	1
200440007	*3	NA	200623128	200126549	1	200851927	200563486	1
200440353	200240529	1	200639428	200126549	1	200856019	200555911	1
200440772	200241045	1	200639578	200438811	1	**200856724**	**200522071**	1
200441336	200059904	1	200653003	200126549	1	**200856726**	**200755193**	1
200450581	200040787	1	200653015	200480410	1	200856749	200755193	1
200458810	200040787	1	200653176	200480410	1	200863141	200275503	1
200458833	200025086	1	200653342	200126549	1	200863271	200421240	1
200458977	200075328	1	200653864	200240547	1	200863545	200555911	1
200459117	200075328	1	200661235	200126549	1	200863569	200555911	1
200460546	200025086	1	200724160	200470217	1	200870108	200522071	1
200475334	200025686	1	200724316	200234397	1	200870219	200555911	1
200480064	200040787	1	200727152	200529119	1	200875287	200421240	1
200480432	200280193	1	200727153	200529119	1	200880369	200522071	1
200480509	200025086	1	200727413	200275503	1			
200513424	200213379	1	200727472	200275503	1			
200522557	200025086	1	200740102	200470024	1			
200529095	200225101	1	200740129	200458811	1			
200529412	200327021	1	200750095	200380115	1			
200540022	200025086	1	200750174	200126549	1			

In assignment analysis, in addition to correct allocation of fathers to none excluded offspring in exclusion analysis, five out of the nine excluded offspring now get assigned to their true father (the five pairs are bolded in the table).

Table 3: Summary of wrongly matched parent-offspring pairs in both breeds and inferred father by the panel in the NWS

Individual ID	Father as in pedigree	Father inferred by the panel
For NWS:		
9170130	8575795	No assignment
9940537	9540149	" "
200421779	200061130	" "
200440007	200025686	" "
200563404	200025086	200225058
9859266	9427694	9427584
200353873	200040787	9970207
200856724	200755193	200522071
200856726	200522071	200755193
For Spael:		
810523	768161	Assignment analysis not done
9070476	8770533	" " "
200460668	200070442	" " "
200450202	200170423	" " "

Out of the nine mismatches detected by the exclusion analysis, five of them assigned to their true fathers, while assignments were not found for the remaining four individuals/offspring in the NWS. Assignment analysis was not conducted for the Spael breed due to time shortage.