Norwegian University of Life Sciences
Faculty of Veterinary Medicine and Biosciences
Department of Animal and Aquacultural Sciences

Master Thesis 2015
30 credits

DEVELOPING OF A SNP BASED PARENTAGE TESTING PANEL IN THE NORWEGIAN GOAT

CASANDRA RODRIGUEZ Y RAMIREZ

ACKNOLEDGMENTS

I am using this opportunity to express my gratitude to everyone who supported me throughout the process of my thesis. Hence, I start expressing my sincere gratitude to my supervisor Prof. Dag Inge Våge for his continuous support during this project, for his patience, motivation, and vast knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor for my master thesis.

Besides my supervisor, I want to express my gratitude to the CIGENE group who provided me with the facilities and the Norwegian University of Life Science who gave me the opportunity to have wonderful teachers that provided me with the knowledge required to conduct this project.

My sincere thanks also goes to the CIGENE/IHA team: Mariann Arnyasi who supported me with her knowledge and appreciated time during the entire Illumina genotyping work. Also, I am thankful to Kristil Kindem Sundsaasen who was really kind sharing her knowledge and taught me how to design accurately the primers for the 40 SNP multiplexes of Sequenom technology, and the last person from the group but not less important Kristina Vagonyte Hallan who guided me and explained me with detail and enthusiasm the whole process of Sequenom genotyping and its analysis. Also, I would like to show my gratitude to the PhD student Tesfaye Kebede Belay who helped me during the first part of my thesis and gave me valuable advises on how to continue my work. I am filled with gratitude for their friendly guidance, without their precious support it would not be possible to conduct this research.

I would like to thank the Norwegian Association of Sheep and Goat Breeders for the financial support and access to the DNA-samples used in this study.

Last but not the least, I would like to thank my lovely family: my parents that without their love, patience and support I would not be here, my aunt who has been always there, and to my brother and sister for supporting me spiritually throughout these two years far away from home. For being always together and have the precise words that encourage me to not give up.

ABBREVIATIONS

AI	Artificial insemination
CIGENE	Centre for Integrative Genetics
ddNTP	Dideoxynucleotide
EH	Expected heterozygosity
EPN	Expectation of probability of non-exclusion
FS	Full sibs
GWAS	Genome wide association study
HS	Half sibs
HWE	Hardy Weinberg Equilibrium
IBD	Identical by descendant
IBS	Identical by state
ID	Identity document
NSG	Norwegian Association of Sheep and Goat Breeders (Norsk Sau og Geit)
MAF	Minor allele frequency
OH	Observed heterozygosity
PCR	Polymerase chain reaction
PE	Probability of exclusion
PIC	Polymorphic information content
PN	Probability of non-exclusion
SNP	Single nucleotide polymorphism

NORWEGIAN UNIVERSITY OF LIFE SCIENCE

Department of Animal and Aquacultural Sciences

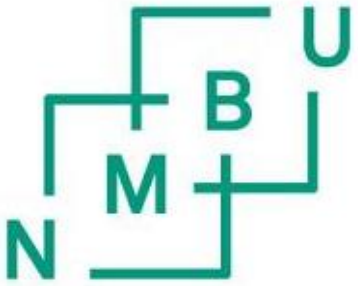
European Master in Animal Breeding and Genetics

THESIS (30 credits)

DEVELOPING OF A SNP BASED PARENTAGE TESTING PANEL IN THE NORWEGIAN GOAT

CASANDRA RODRÍGUEZ Y RAMÍREZ

AUGUST 2015



SUPERVISOR: PhD DAG INGE VÅGE

LOCATION: Centre for Integrative Genetics (CIGENE)
Department of Animal and Aquacultural Science
Norwegian University of Life Science, NMBU

CONTENT

ABSTRACT

A parentage test for Norwegian goats consisting of 59 highly informative SNPs has been developed. A panel of 48 AI bucks genotyped by the Illumina 50K SNP chip were used to select 80 highly informative SNPs for the parentage test. The 80 SNP parentage panel was evaluated by genotyping the 48 AI bucks together with 143 potential breeding bucks. According to the pedigree, 121 of the 143 potential breeding bucks were sired by one of the 48 AI bucks, allowing for parentage testing by the SNP-panel. The SNP based parentage analysis showed that 15% of the parentages were most likely wrongly assigned. Theoretical considerations show that the panel had a high power of non-exclusion (EPN= 0.004), given a MAF of 0.44 and only a single parent genotyped. We therefore conclude that this 59 based parentage SNP panel is an effective tool for excluding incorrect assigned sires in the Norwegian breeding system.

1. INTRODUCTION

1.1 Norwegian goat farming

Goats are considered as an important source of meat, milk, fibre and pelts, and have played important roles since early human civilization time in agricultural, economic, cultural and religious fields (Dong et al., 2013).

The Norwegian goat husbandry is relatively small when comparing to other livestock such as cattle. However, the milk production from goat still maintain a long and important tradition in Norway, with a total of goat milk production about 23 million litters. Goat farms are concentrated in the western and northern Norway, where milk production take place on small farms with an average size of 85 dairy goats (Tormod Ådnøy, 2014). Therefore, farmers have formed The Norwegian Association of Sheep and Goat Breeders (NSG) (in 1947), whose principal aim is to take responsibility for the nation-wide breeding programmes for sheep and goat. The Norwegian Goat breeding programme is used to select buck kids for AI to achieve their breeding goal (Tormod Ådnøy, 2014). The breeding values are evaluated based on information from sire and dam, as well as additional criteria. Thus, maximal genetic progress through the breeding programme is the main motivation to develop an accurate pedigree controls.

1.2 Importance of parentage control

Pedigree errors are considered to be a common problem in small ruminant breeding programs. These errors may occur when lambing happen in variable sized groups, making a significant risk of misallocation of the dams or an inadvertent misbreeding due to broken fences (Heaton et al., 2002). Other factors that may introduce pedigree errors are: Incorrect identification of semen samples; wrong males ID are entered into the insemination record; natural mating of sires with dams which were assumed to be pregnant by AI; interchange of offspring on the farm (Řehout et al., 2006, Israel et al., 2000, and Fisher et al., 2009).

Animal breeding programmes make use of different statistical models such as the animal model evaluation, which assumes that relationships between animals were properly

recorded (Hashemi M. et al., 2013). This assumption are not totally true however, and incorrect pedigree recording of livestock, especially sire identification has a major effect on breeding values estimation, genetic gain, as well as on inbreeding level. Therefore incorrect identification of sires bias heritability estimates, evaluations of breeding sires and causes direct negative effects on maternal genetic correlation (Heaton et al., 2002, Senneke et al., 2004, Hashemi, 2013 and Al-Atiyat et al., 2015). These biases will increase at the same time misidentification increases. In consequence, missing pedigree data would reduce the power of statistical models for genetic evaluations as accurate evaluations are based on correct pedigrees (Israel et al., 2000, Zhang et al., 2013 and Rosa et al., 2013).

1.3 DNA markers

DNA markers (e.g. Microsatellites or Single Nucleotide Polymorphisms) are very useful when performing parental verification (confirm relationships between two or more individuals), but also for parental identification (establish relationships between two or more individuals) (Ramos et al., 2009).

Microsatellites consist of repeated and highly polymorphic sequences dispersed through the genome, and are highly used for the analysis of relationships. Moreover, as they are highly informative, they have been the markers of choice for identity verification of livestock. Thus, their effectiveness depends on the level of information provided by the markers (María et al., 2013). Verification of relationships in segregating populations is generally based on the exclusion principle, which refers to a conflict at a specific locus between the genotypes of two individuals excluding the possibility that the alleged relationship is correct (Baruch et al., 2008 and Bolormaa et al., 2008). Some examples of parentage studies in livestock are described by Al-Atiyat (2015), where a panel of 28 microsatellites located on different chromosomes were developed to genotype DNA in Australian Merino sheep. This panel provided a probability of exclusion (PE) close to 99.06, ensuring wrongly assigned parents to be excluded. Similarly, a paternity test of 14 microsatellites developed in Australian goats gave a high level of exclusion; PE >99.70% (Bolormaa S. et al., 2008).

Recent advances in DNA sequencing, computer software and bioinformatics have made the use of SNP markers more common. Also, there is an increasing interest in using SNP markers for pedigree verification and identification purposes, due to easier automation and standardization between laboratories, as well as their low genotyping error rates, their abundance across the genomes, and their genetically stability (María et al. 2013).

However, due to the lower resolving power of biallelic SNP loci, SNP panels need to include more loci in order to obtain the same power of exclusion as microsatellites do (Baruch et al. 2008 and Heaton et al., 2014 & 2002). Therefore, an efficient SNP based identification systems must contain a minimal set of SNPs with sufficient power to identify individuals and their relationships for the common breeds and crossbred populations (María et al. 2013). Nevertheless, the information content in a SNP set may vary significantly between populations (Krawczak, 1999). Baruch et al. (2008) suggested that by using SNP marker panels with high minor allele frequencies values, the number of SNPs can be reduced.

Recently, high-density SNP chips have become available for different livestock species such as cattle, horse and dog (Ramos et al., 2009), and small ruminants like sheep and goats. However, the SNP panel for goats has not been available until recently, and population studies and paternity evaluations has been obtained from classical markers such as mitochondrial and microsatellite loci (Marcel, 2014). On the other hand, the sequencing of goat has generated a ~2.66Gb genome, in the framework of the International Goat Genome Consortium (http://www.goatgenome.org/) (Dong, 2013). Next generation sequencing has allow the development of large collections of SNPs, where a total of 60,000 SNPs were selected to generate a final cluster file of 53,347 SNPs. The goat 52K SNP chip developed by Illumina has already been used to e.g. carry out a genome wide association study (GWAS) for polledness in four different goat breeds. (Marcel, 2014 & Tosser-Klopp et al., 2014).

The main goal of the current study is to develop a SNP based parentage testing panel for Norwegian goats. The test is based on the principle that a parent and offspring must share an allele at every locus, and the probability of exclusion (PE) of an incorrectly assigned parent is dependent on density of markers, physical spacing, and Minor Allele Frequencies (MAF).

2. MATERIAL & METHODS

2.1 DNA samples and family structure

DNA was isolated from semen straws of 48 AI bucks born in the years 2009-2012. In addition, DNA was isolated from blood of 143 potential breeding bucks. DNA isolation was made by BioBank AS laboratories (http://www.biobank.no/index.php/no/). Relationships between individuals were determined by pedigree information provided by The Norwegian Association of Sheep and Goat Breeders (Norsk Sau og Geit, NSG www.nsg.no).

The majority of the 143 potential breeding bucks were sired by bucks found among the 48 AI bucks. Also, some close relationship was observed within the 48 AI bucks; 4 full sibs, 23 half sibs and 3 father-son relations (Table 1).

ID	Relationship	Individuals per related flock
2011255	FS	2
2011256	FS	
2011259	FS	2
2011260	FS	
2009251	HS	2
2009295	HS	
2009318	HS	
2009324	HS	3
2009344	HS	
2009259	HS	
2010139	HS	3
2010143	HS	
2009106	HS	2
2010173	HS	
2010166	HS	2
2010169	HS	
2009153	HS	2
2009245	HS	
2010220	HS	
2010251	HS	3
2010422	HS	
2009239	HS	
2009262	HS	
2011162	HS	6
2011171	HS	
2011181	HS	
2012151	HS	

Table 1. Number of full sibs (FS) and half sibs (HS) registered on the 48 animals pedigree obtained on The Norwegian Association of Sheep and Goat Breeders website (www.nsg.no). Highlighted IDs indicate the individuals that also sire other AI bucks in this panel.

2.2 Illumina genotyping and quality control

The 48 AI bucks were analysed by the Illumina 50K SNP array Infinium HD Assay Ultra protocol (Catalog # WG-901-4007). In total, 52025 markers were genotyped along all chromosomes. The locus discrimination occurs by the combination of sequence specific hybridization capture and array-based, single base primer extension (Illumina, 2010). When there is a perfect match, extension occurs and a signal is generated, and in contrary case when there is a mismatch the extension does not occurs and in consequence the signal is not generated. A laser excite the fluorophore of the single base extension product on the bead chips section, and high resolution images of the light emitted by the fluorophore are generated (Illumina, 2010). These reads are observed as green, red and yellow dots and are subsequently translated into genotypes.

The genotype data was analysed using the free open source Plink v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) (Purcell et al. 2007). Minor allele frequencies (MAF) was estimated and the Hardy Weinberg Equilibrium (HWE) test were performed in order to check the proportion of the genotypes.

Selection of highly informative SNP markers was done after the quality control performed by the Illumina GenomeStudio software (http://support.illumina.com/content/dam/illumina-marketing/documents/services/technote_infinium_genotyping_data_analysis.pdf). As mentioned previously, minor allele frequencies (MAF) and Hardy Weinberg Equilibrium (HWE) analysis were performed by Plink software subsequently to Illumina genotyping. The criteria for selecting a SNP marker for inclusion in the parentage test panel was a minor allele frequency in the range 0.45 to 0.5 and a physical distance of ≥20cM between individual markers on the same chromosome. The reason for a minimum threshold MAF of 0.45 was to obtain as informative SNPs as possible. The average MAF for selected SNP was 0.485.

The selected SNP markers were manually inspected with the GenomeStudio software results by revising the plots generated in this programme. A total of 128 SNP markers were selected as candidates for the parentage panel. These markers were subjected to primer design for iPLEX Gold genotyping (Sequenome). By mistake, two markers failing the HWE test were included in the primer design.

2.3 Primers designs for Sequenom genotyping

For each SNP marker, a set of amplification and extension primers were designed. The design was made using Assay Design Suite v 2.0 (Agena Bioscience) (http://agenabio.com/products/massarray-system/). The size of the amplicon that contain the polymorphic site was established following the standard conditions (max. 120 bp and min. 80 bp) and hybridization temperatures were used as indicated in the protocol. Besides, it was performed an analysis of possible primer dimer formations for the design in order to avoid interfere on the correct PCR reaction. Two 40 SNP multiplexes were generated based on the 128 candidate markers.

The correct design of extension primer it is of high importance due to the 3' end is located adjacent to the polymorphic site. Products of extension primers were analysed according to mass difference. In figure 2 it is shown the principle of the PCR reaction following the iPLEX Gold protocol.

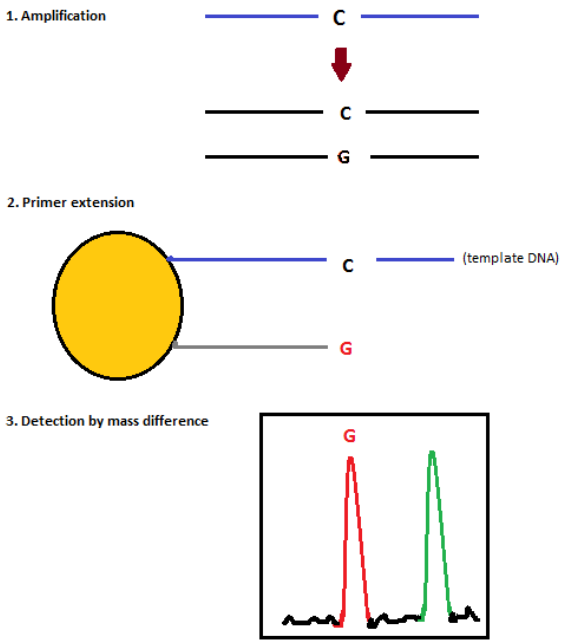


Fig. 2 Amplification and extension Illumina reactions. The first step include amplification of the fragment where the polymorphic site it is located. Amplification is followed by the annealing of extension primers. The extension reaction stops after adding one dideoxynucleotide (ddNTP) to the extension primer, allowing the detection of the polymorphic nucleotide by mass difference between the two extension primers that differs by one base.

2.4 Sequenom genotyping and relationship verification

Both the 48 AI bucks and the 143 potential breeding bucks were genotyped by 80 SNP markers using the Sequenom iPLEX® Gold mass array system protocol A. The assay was based on primers extension of the amplified goat DNA. SNPs were detected by MASSARRAY mass spectrometer which identified the different masses of the extended primers that contain the SNP allele. A software (SpectroTYPER) translated the mass of the observed primers into a genotype for each SNP.

In the first round, the 48 AI bucks were genotyped using the 80 SNP panel. These results were compared with the corresponding genotypes obtained with the Illumina SNP chip in order to avoid probable mismatches between the two genotyping technologies. In addition, the 80 SNPs panel were used for genotyping the 143 potential breeding bucks. Call rates and HWE were checked for each SNP after the genotype analysis. SNPs whose call rate were lower than 90% or that failed the HWE test were removed from the parentage based panel and not used for paternity testing.

The numbers of mismatches between putative parents and offspring were detected using an algorithm designed at the Centre for integrative Genetics (CIGENE), within Galaxy open source framework (https://usegalaxy.org/). The platform look for Mendelian errors by linking parentage and progeny genotypes with pedigree information. Parents with 0 or 1 SNP mismatch compared to offspring were accepted as parents, whilst parents with higher genotype mismatches were rejected.

2.5 Evaluation of exclusion power and marker informativeness

The statistical treatment of parentage control is commonly based on the probability of parentage exclusion. This approach is based on the principle that a parent and offspring must share at least one allele at every locus and the probability of exclusion (PE) is the probability that an alleged parent would be excluded from parentage (Heaton, 2014). Furthermore, achieving accurate parentage assignment without information of single parent

genotype while keeping the number of SNPs to a minimum, requires that each ‘‘parentage SNP’’ has a high PE value as possible (Heaton, 2014).

For the purpose of this study, probability of exclusion was estimated based on the assumption that the putative parent and the offspring genotypes follow Mendelian inheritance. Moreover, power of exclusion undertakes three scenarios proposed previously by Jamieson & Taylor, 1997 (Baruch et al., 2008):

1. Genotyping a progeny and a single assumed parent.

2. Genotyping a progeny and two putative parents.

3. Genotyping a progeny, one actual parent and one putative parent.

However, for the purpose of this study only scenario 1 was considered. The probability of exclusion (PE) is expressed as: $2(Pi)^2(1 - Pi)^2$, where Pi = MAF for marker i, and the probability of non-exclusion (PN) for a single marker is expressed as $1-2(Pi)^2(1 - Pi)^2$. Assuming that the allelic frequencies are unknown and can vary across markers and sets, the expectation of PN (EPN) is estimated as follow (Baruch et al., 2008):

$$\mathrm{EPN} = \left[\frac{7/15 - a + 2a^3/3 - a^4 + 2a^5/5}{0.5 - a}\right]^{\mathrm{N}}$$

Where N is equal to the number of markers, and a is equal to the MAF of each marker.

Theoretical simulation of EPN was performed using different MAF values: 0.1, 0.2, 0.3, 0.4 and 0.5, as well as different number of informative SNPs: 80, 70, 60, 50, 40, 30, 20, 10 and 0 in order to contrast the probability of non-exclusion along different situations.

The estimation genotype frequencies and observed heterozygosity (OH) from genotypes calculated previously on Plink v1.07 were checked by direct calculation in order to statistically characterise the 80 highly informative parentage SNPs used in this study. The expected heterozygosity (EH) was estimated using the formula of Nei et al. (1978).

$$2P_{ij}(1\text{-}P_{ij})2N_j/(2N_j\text{-}1)$$

In the formula, P_{ij} is the frequency of i^{th} allele at locus j^{th}, and N_j is the number of individuals for the locus j^{th}.

Moreover, polymorphic information content (PIC) for each marker was determined using the equation proposed by Botstein et al. (1980). This equation define the probability that a marker genotype of an offspring would allow to deduce which marker allele is inherited from a heterozygous father in Hardy Weinberg equilibrium situations (Robert, 2005). Theoretical PIC values range from 0 to 1, where values equal to 0, states that the marker has only one allele, on the contrary when PIC is equal to 1, the marker would have an infinite number of alleles. Therefore, values near to 1 are considered highly informative, however, for markers with only two alleles their maximum PIC is 0.375 (Hildebrand, 1992).

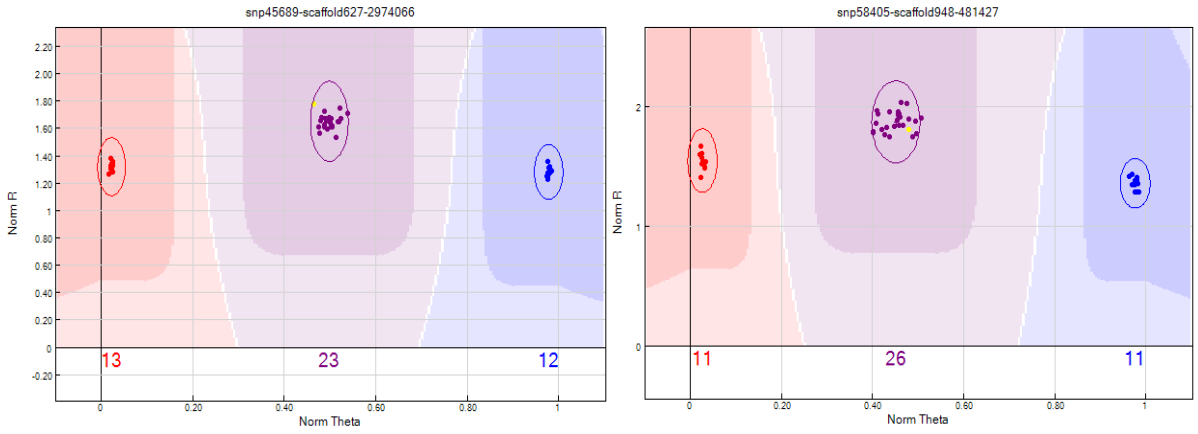
$$\text{PIC} = 1 - \sum_{i=l}^{i=n} p_i^2 - \sum_{i=l}^{i=n-1} \sum_{j=i+1}^{j=n} 2p_i^2 p_j^2$$

In the formula pi is equal to the frequency of the i^{th} allele, and n is the number of marker alleles (Botstein et al., 1980)

3. RESULTS

3.1 Illumina chip analysis (Genome Studio software)

From the analysis performed by the Illumina 50K SNP array on the 48 AI bucks samples, 52025 markers were genotyped. As described in methods, determination of genotypes were performed through GenomeStudio software. These results showed that a total of 1323 SNP had a call rate different from 100%, thus, a total of 50702 SNP were identified with a high sensitivity for each animal. Examples of genotype cluster plots for two SNPs are shown in figure 3. In each plot the different classes of genotypes are represented by different colours; homozygous genotypes in red and blue and heterozygous genotypes in purple. Each dot within the clusters represent one sample that was assigned into that group for a particular SNP marker. Signal intensity is represented by the Y axis (Norm R) and their allelic intensity ratios shown in the X axis (Norm Theta) (Illumina GenomeStudio Data Analysis Software, 2013 & 2014).



SNP ID	MAF
snp45689-scaffold627-2974066	0.4896

SNP ID	MAF
snp58405-scaffold948-481427	0.5

Fig. 3. The two plots show the heterozygous (purple dots) and homozygous (red and blue dots) samples for snp45689-scaffold627-2974066 and snp58405-scaffold948-481427. The number of individuals within each class is indicated below each class.

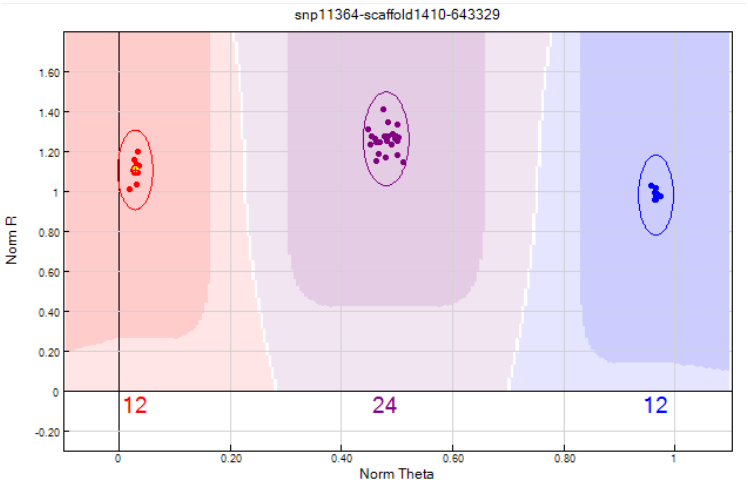
3.2 Estimation of MAF and HWE of SNP markers using Plink v 1.07

The Illumina genotypes of the 48 AI bucks were further analysed with PLINK v 1.07 open source (Purcell S et al. 2007). The Hardy Weinberg Equilibrium (HWE) test revealed that out of 52025 SNP markers genotyped, 373 SNPs failed the HWE test, leaving 51652 SNPs for further consideration. In addition, 1025 SNPs were tagged as X-chromosome located, thus, a total of 1398 SNPs were excluded. Therefore, the 128 candidate SNPs were selected from the remaining 50627 SNPs according to the given criteria (MAF ≥ 0.45, and physical distance ≥20cM between markers located in the same chromosome). Each SNP was visually examined by making use of GenomeStudio plots where signal intensities of the markers allowed the possibility to check clustering. For the clusters analysed from the 128 SNPs, all showed a call rate with high confidence (100%), where wrong calls were not found and the three clusters (homozygotes for one allele, heterozygotes and homozygotes for the other allele) are clearly identify. In table 2 and figure 4, are observed MAF and HWE values from one SNP markers obtained with Plink, and an intensity plot and MAF value obtained on Genome Studio respectively. The PIC values estimated for each marker gave an average of 0.3747. This estimated value is high since the maximum PIC value for a marker with two alleles is equal to 0.375. On the other hand the overall observed heterozygosity was relatively high than expected (OH=0.508, EH= 0.504) indicating efficiency in heterozygosity.

Hardy Weinberg equilibrium $p^2 + 2pq + q^2 = 1$

Chromosome	Allele frequency	Genotype	p	q	2pq	HWE	MAF
9	0.5	12/24/12	12/96=0.125	1-0.125= 0.875	2(0.125)(0.875) = 0.21875	$(0.125)^2+2(0.125)(0.875)+(0.875)^2=1$	AG: 24 + 2(12)=48 48/96=0.5

Table.2 Minor allele frequency (MAF) and Hardy Weinberg equilibrium (HWE) Plink outcomes (i.e. snp11364-scaffold1410-643329)



SNP ID	MAF
snp11364-scaffold1410-643329	0.5

Fig.4 On the plot is observed three distinct clusters represented by three different colours for the snp11364-scaffold1410-643329 marker assay. Homozygote genotypes are represented by red and blue dots, whilst heterozygote cluster is represented by purple dots. Each dot represent one sample called into the group. These dots are called by their signal intensity represented by the Norm R and their allelic intensity shown in the Norm Theta (Illumina GenomeStudio Data Analysis Software, 2013 & 2014).

3.3 SNP marker selection & primer design

Based on the 128 SNP candidate results, primers design was done as described in methods. The lower MAF obtained for the 80 informative SNP's was 0.4583. In figure 4 is shown the MAF rank for each of the 80 SNPs selected. The complete list of the SNPs selected is found in the appendix of this study.

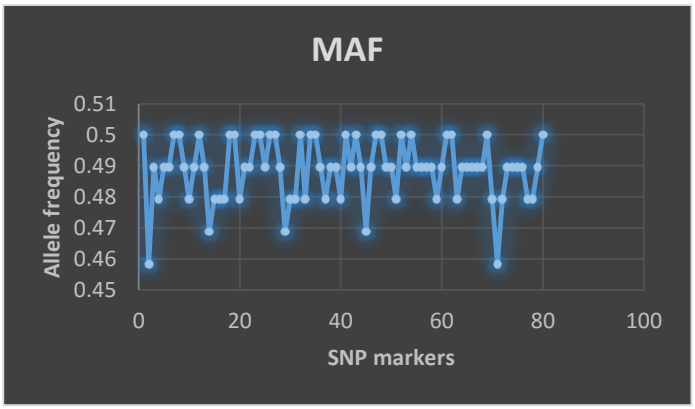


Fig.4 Minor allele frequency (MAF) distribution of the 80 SNP informative markers selected for the two 40 multiplex genotype assay. Threshold of MAF ≥ 0.45.

3.4 Sequenom genotyping

Performances of the 80 SNPs included in the parentage test (Sequenom iPLEX® Gold mass array system) were evaluated through the 191 animals (48 rams and 143 potential breeding bucks) in term of total genotyping efficiency primarily, i.e. the number of genotype called samples related to the total number of samples. Results showed that two individuals (14491367_5043 and 14491367_5045) had a bad DNA quality with only 3 and 4 informative sites respectively, however, by mistake they were not removed from the analysis. From the 80 SNP markers included in the parentage panel, 5 SNP markers had a call rate of 0, whilst 5 SNP markers had a call rate equal to or lower to 87% (Fig. 6a). For the remaining SNPs the total call rate was ≥ 90%. The SNP 58568-scf953-1146096 showed a particular distribution of genotypes, where 89% of the individuals were heterozygotes (AG) and only 5% were homozygotes (AA) and no animals were homozygous (GG). Hence, the number of genotyped individual and the genotype distributions were calculated for each SNP in order to avoid SNP errors that may bias the relationship analysis (Fig. 6b).

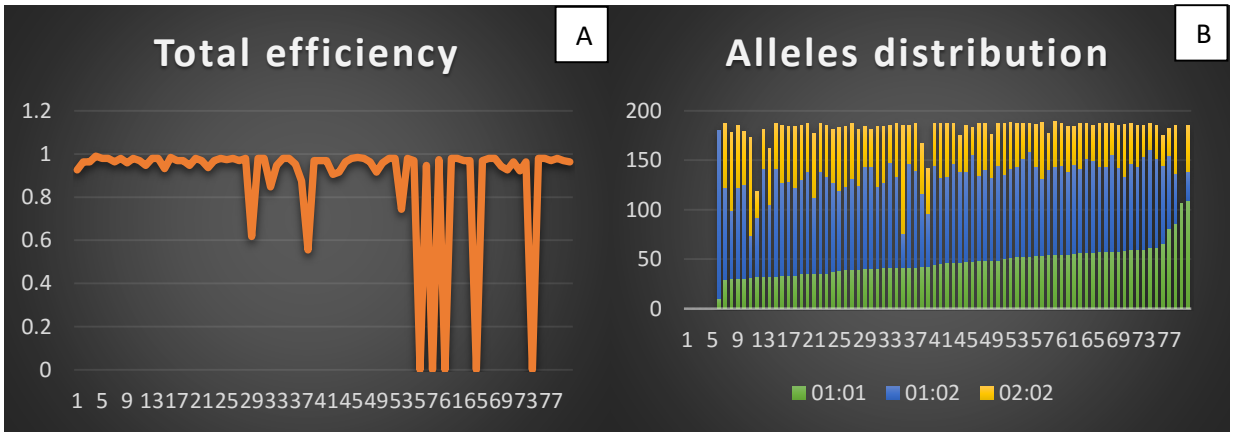


Fig.6 A) Evaluation of total efficiency of the 80 informative SNPs across the total population genotyped. B) Genotypes (heterozygous/homozygous) distribution of the 80 SNP among the 191 animals genotyped. Homozygote genotypes are represented as 01:01 and 02:02 tags, and heterozygote genotypes as 01:02.

The 48 AI bucks were genotyped by both the Illumina SNP chip and the Sequenome iPLEX mass array system, which allow for a direct comparison of genotypes determined by Illumina and Sequenom technologies. In total, 11 SNPs were found with genotype

discordances (Fig.7), showing an unusual high frequency of homozygotes when genotyped with Sequenom technology. For that reason, SNPs with a genotype discordance between the two technologies higher than 2 were removed from the SNP panel.

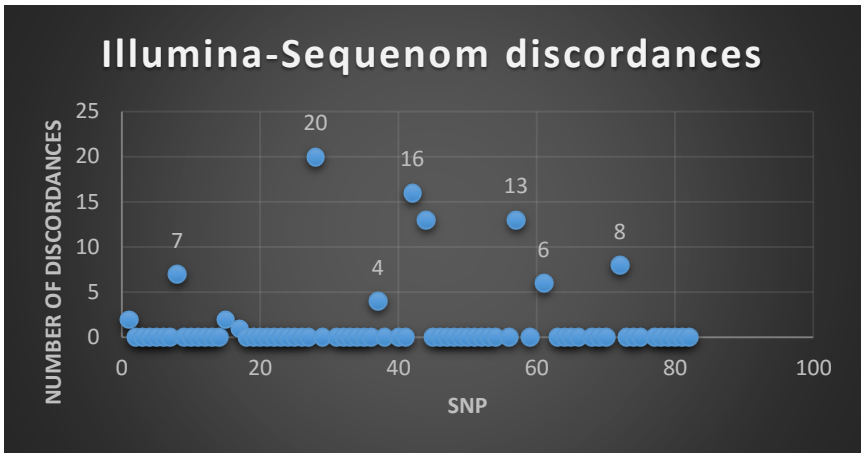


Fig.7 Genotype discordances between Illumina and Sequenom results in the 48 AI rams genotyped by both technologies.

Furthermore, 10 SNPs with a low total efficiency performance (call rate below 90%) were excluded from the parentage testing performed in Galaxy (Fig. 8). After withdrawing all these SNPs, the SNP parentage panel was built by 59 informative SNPs. Considering the SNPs 30627-scf339-2837646 and 33324-scf391-2614900 with 85% and 87% call rate respectively a panel of 61 SNP markers was also tested. Outcomes from 61 and 59 SNP panels were quite similar, however. In overall, 12.5% of the 80 SNP markers genotyped did not give satisfactory results, whilst 13.75% showed inconsistency in genotype technologies comparison study. SNPs excluded from the SNP parentage panel are shown in table 3.

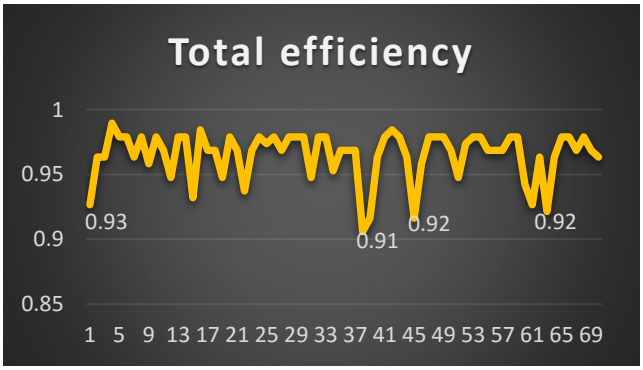


Fig.8 The graphic illustrate the rate of the SNPs efficiency after removing the 10 SNP markers with a call rate below 90%.

SNPs	Reason for elimination	Number of SNPs removed
10219-scf1368-489857 12947-scf1499-2414292 18180-scf185-15961025 18251-scf1850-107670 24678-scf2510-227964 32696-scf377-530544 36532-scf44-1147631 37317-scf455-1220426 45689-scf627-2974066 47853-scf674-504140 58755-scf959-748556	Discordant genotypes between Illumina/Sequenom genotyping technologies	11
46688-scf65-2255932 47454-scf668-2866943 50672-scf730-379018 5536-scf1195-7510 6908-scf1251-1645032	Cero call rate	5
28560-scf306-102662	Low call rate 62%	5
34377-scf404-62060	Low call rate 55% and "unexpected" distribution of genotypes	
45225-scf617-621933	Low call rate 74%	
30627-scf339-2837646	Low call rate 85 %	
33324-scf391-2614900	Low call rate 87%	

Table3. List of SNPs excluded from the 80 SNP panel used for analysis of relationships between parents and offspring, according to two selection criteria: Illumina-Sequenom technologies mismatches, and SNP efficiency (cero call rate/weak call rate <90%)

3.5 Verification of power panel

The practical power of the 80 SNP panel was verified in terms of the probability of non-exclusion. The analysis was based on the scenario where only one parent (father) and offspring genotypes are known. In total, 121 sire-offspring relationships were available for testing, among the 191 animals genotyped. The 80 SNP panel revealed 57 incompatible relationships of the 121 parent-offspring sets, ranking number of incompatibilities from 1 to 16 loci. However, from those 57 incompatibilities, 38 were avoided when the low-performing SNP were removed from the parentage based panel. Removing 21 SNPs (11 technology incompatible SNPs and 10 SNPs with zero/low call rate) reduced the number of relationship incompatibilities

to only one discordant locus significantly. As a result it was found 19 out of 121 father-son relationships (16 %) to be wrong in the tested material. Inspection of the genotype for the two animals (14491367_5045 and 14491367_5043) which number of disparities did not change after removing some SNPs from the panel showed that genotypes were based only on 3 and 4 informative sites respectively, explaining their incompatibilities in the relationship analyses. The additional incompatible relationships showed minuscule improvements giving 5 to 14 discordant genotypes (Table 4). Since at least a locus failing Mendelian segregation was considered acceptable, 15% incompatibilities were considered to have wrong parentage. Thus, the putative parents of the two animals with low number of informative sites and the other 15% could most likely be excluded from paternity for those animals. The complete list of parent-offspring relationships and discordances are presented in the appendix Table 2 and 4.

Number of incompatible loci detected in panels with different numbers of markers			
Parent - offspring	**74 SNPs**	**61 SNPs**	**59 SPNs**
1	16	14	13
2	15	11	10
3	14	12	11
4	14	12	11
5	13	10	9
6	13	9	9
⋮			
51	1	0	0
52	1	0	0
53	1	0	0
54	1	0	0
55	1	0	0
56	1	0	0
57	1	0	0

Table 4. Number of incompatible loci between parent-offspring sets concordance. The table only shows some of the discordances on the 121 relationships tested. Removing the low-performing SNPs from the panels reduced the number of loci conflicts, which are shown below each SNP. (Complete table is found in the Appendix)

Exclusion of the 21 low-performing markers from the original 80 SNP panel provided a non-exclusion probability lower than 0.01 with at least 1 locus conflict between father-offspring genotypes. The experimental and simulated results demonstrated that decreasing

the number of low-performing SNPs reduce the level of conflicts in the parentage – offspring correspondence. Moreover, a high minor allele frequency (≥0.45) plays an important role in reducing the number of SNPs required to achieve an accurate parentage without losing power of relationship correspondence (Fig.12).

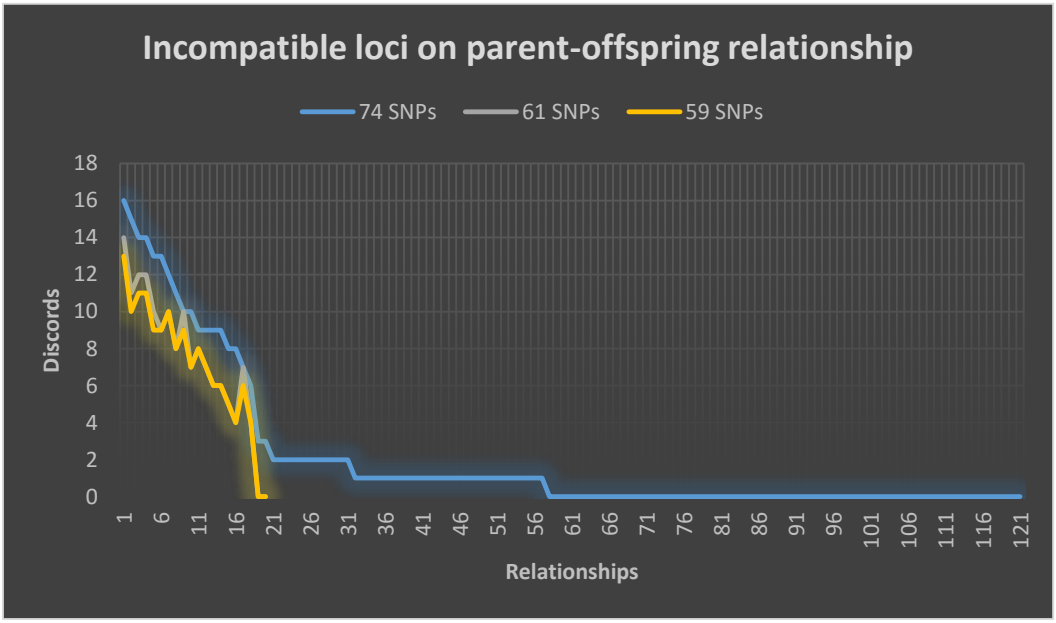


Fig. 12 Number of conflicting loci between parent – offspring, when only one parent and the offspring genotypes are known. Removing the low-performing SNPs from the panel (from 80 to 59 markers) reduced the number of conflicts. The panel with 59 SNPs drop drastically to 0 loci conflicts for the majority of the relationships after removing the 21 conflictive SNPs. The 18 discordant relationships are identified in all tested panels.

3.6 Probability of parentage exclusion analysis

The 80 SNPs panel was evaluated by estimating the theoretical probability of non-exclusion in paternity testing. This evaluation is based on the allele frequencies and number of markers. Thus, estimation of non-exclusion probabilities for 80 markers computed by Baruch equation when only one putative parent is given resulted in 0.003 for the whole panel and 0.000024 for a single SNP when MAF is equal or greater to 0.46. Moreover, simulated analyses with different MAF values (MAF; 0.1 to 0.5) was performed (Fig.9). In these simulations it was observed that circa 35 SNPs are needed to reach an EPN ≤0.01 when the MAF is equal to 0.5, whereas a probability of non-exclusion ≤0.01 is reached using a panel of

50 SNPs when MAF is equal to 0.1 (Fig. 10). Therefore, the number of SNPs required in the parentage panel is strongly dependent on the average MAF value of the markers. (Fig. 10)

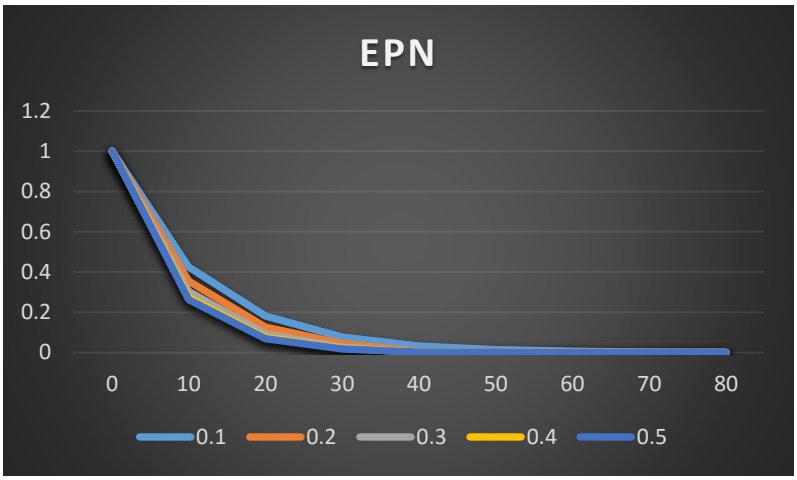


Fig.9 Simulation of the probability of non-exclusion (EPN) with different minor allele frequencies (0.1, 0.2, 0.3, 0.4 and 0.5). The higher the MAF the lower the probability of non-exclusion (EPN). On the x-axis are shown the values regarding the number of the SNPs in the panel, and on the y-axis the probability of non-exclusion values.

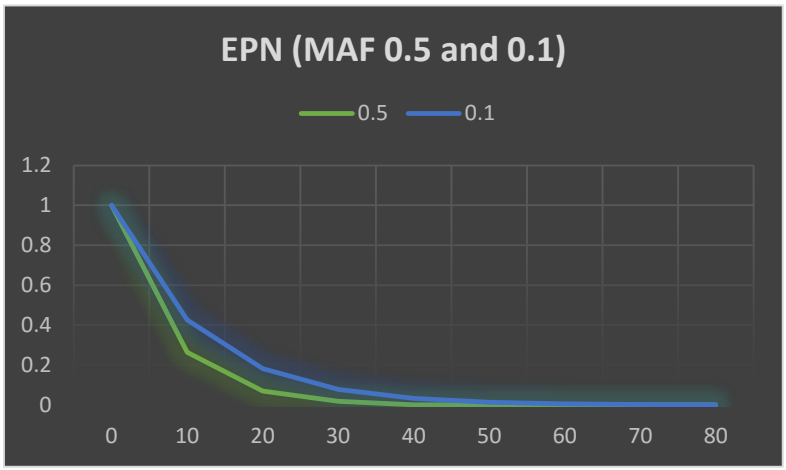


Fig.10 Effect of minor allele frequencies (MAF) on the probability of non-exclusion (EPN). When the MAF is low (0.1) a higher number of SNPs are required to obtain an EPN of 0.01, which is the opposite situation when MAF is high (0.5). On the x-axis are shown the values regarding the number of the SNPs in the panel, and on the y-axis the probability of non-exclusion values.

Using the true values of minor allele frequencies (≥0.364) of the population analysed (191 individuals), the estimation of non-exclusion power for different numbers of SNP markers (80 to 0) revealed that a probability of non-exclusion of 0.0051 is reached with 40 informative SNPs. It becomes more evident that effectiveness for non-exclusion probability

of the SNP parentage based panel depends closely on the minor allele frequencies. After excluding the 21 ineffective SNPs, 59 markers were used and obtained a probability of non-exclusion of 0.00041 for the whole panel, which is widely lower to 0.01 (Fig.11). The minor allele frequency on the population decreased after including the genotypes of the offspring (143 animals) as it was expected. However, the MAF of the whole population was 0.44 in average, being the lower MAF 0.364 and the higher 0.489. This difference slightly changed the probability of exclusion from 0.00038 to 0.00041 for the whole panel. Also it was reported that the probability of non-exclusion per SNP was 0.00056 when the MAF is 0.364, and 0.00037 when MAF is 0.489.

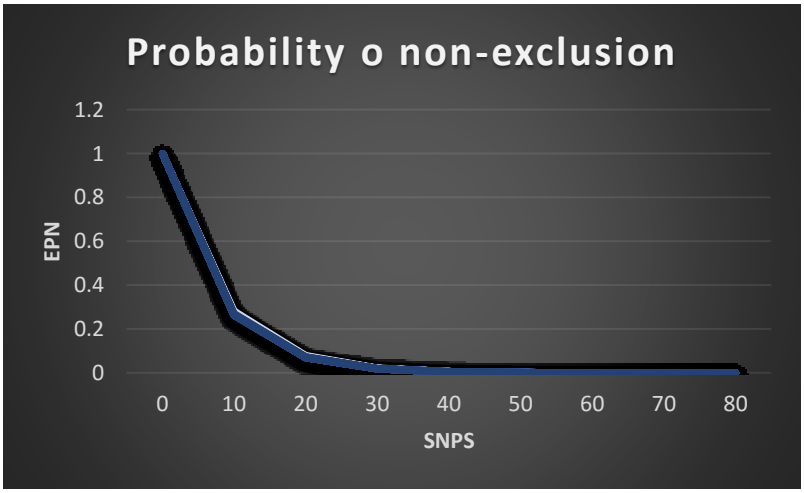


Fig.11 Probability of non-exclusion for a 59 SNP parentage panel. Probability of non-exclusion of the total panel when minor allele frequency is ≥ 0.364 was 0.00041 with a MAF of 0.44 in average. The probability of non-exclusion for the SNP with the lowest MAF (0.364) was 0.00056, and 0.00037 for the SNP with the highest MAF (0.489).

4. DISUSSION

4.1 Population size

In the present study 48 AI bucks were genotyped by the 50K Illumina SNP chip for goats in order to identify SNPs with a high minor allele frequency in the Norwegian goat population. The goal was to develop a parentage test array consisting of highly informative markers, and it was therefore necessary to estimate allele frequencies to select the most informative markers. A study by Marie et al. (2012) suggests that genotyping of 25 to 30 representative animals of a population give a reasonable estimate of the allele frequencies in the population. In this case 48 individuals were genotyped, which is higher than the suggested 25 to 30. On the other hand, the representativeness of the AI buck can be questioned. As shown in results section there is extensive relationship among the bucks, which in consequence will reduce the representativeness of these animals. However, semen from the AI bucks is extensively used in the Norwegian goat population. The alleles that is common among the AI buck will be therefore over time also common in the Norwegian goat population. In consequence this was the main reason for using these 48 bucks for estimating allele frequencies.

4.2 Parentage verification and effect of high MAF

The success of parentage testing depends on the level of information provided by the markers in the testing panel. A strict selection of highly informative SNPs are more likely to perform accurately, depending primarily on the MAF of the individual markers but also on independent segregation marker alleles, and Hardy Weinberg equilibrium. Blanchard et al. (2013) has mentioned that selection of high heterozygosity loci (a MAF near to 0.5) achieve a paramount power for parentage analysis. Hence, in this study selection of highly informative SNP markers resulted in a significant power of exclusion. For instance previous results demonstrated that a parentage SNP panel with a MAF equal to 0.44 in average require at least 35 SNPs to obtain a probability of non-exclusion equal to 0.01. Other criteria that

increases the accuracy of paternity test is the independent distribution of alleles. Therefore, markers whose segregation do not occur independently need to be excluded from the panel (Baruch et al., 2008). The average distance between each marker in this panel present an enough distance (≥20 Mb) to reduce allelic association (Hara et al, 2010). Since several loci are combined to obtain an accurate non-exclusion probability, markers whose segregation occurs independent will be more informative as it facilitates the detection of genotype differences. Related animals have more chances to share the same alleles, especially those who are closer related, therefore seems important to increase the number of highly informative SNPs in the panel with a good distance between each other in order to ensure independent assortment and reduce linkage between markers.

The most straightforward process for parentage identification is performed by the exclusion method. This method match all pairwise potential parent to each offspring comparing their genotypes, assuming that at least one allele is shared between parent and offspring at each marker locus. In this study one discordant genotype was allowed to account for possible genotyping errors. This power of parentage exclusion is preferred to be estimated on unrelated or distantly related individuals (Tesfaye, 2013 & Blanchard, 2013), due to the fact that sharing of alleles is more likely to occur between closely related individuals than in unrelated. Thus, close relationships between candidate parents will have a negative effect on power panel decreasing the probability of non-exclusion. This negative effect will depend on the level of relatedness between possible parents; e.g. full sibs or half sibs, and also the level of inbreeding between the animals. However, this non-exclusion probability can be reduced when SNP panels work with highly informative markers maintaining an even allele frequency of the two alleles at each loci (Blanchard, 2013).

In order to analyse the power of the panel as it is described in methodology, the effect of MAF lower than 0.5 was investigated (Figure 9). As expected, marker panels with MAF values ≥0.45 provide a lower non-exclusion probabilities (≥0.01) than those with lower MAF values. SNPs with high minor allele frequencies (≈0.5) therefore allow a lower number of markers in the panel (Baruch et al., 2008). Our analysis indicates that approximately 50 highly informative markers are needed to provide a probability of non-exclusion equal to 0.01 when MAF was equal to 0.1. In comparison, about 35 SNPs are needed to reach a probability of non-exclusion of 0.01 when MAF is 0.5. These results revealed that the present panel of 59 highly

informative markers should be sufficient even with lower MAF-values than those observed in this study. Remarkably, the probabilities of non-exclusion obtained in this study were similar to those SNP based parentage panels previously reported on livestock (Tesfaye, 2013 and María, 2013). Heaton et al. (2002) presented a non-exclusion probabilities of 0.0025 for pure breed cattle population for a panel of 32 SNPs with a MAF of 0.25, where we conclude from our results obtained that a panel with at least 59 SNP it is feasible to use for practical paternity testing with a MAF equal to 0.44 in average. Additionally, new recruitment of new informative SNPs that address the selection criteria may be possible to perform as desire to improve the current SNP performing panel.

4.3 Genetic relationship

The father- offspring relationships were determined based on pedigree records as well as by marker genotypes. As it was mentioned before, relationships may contain errors in the pedigree due to a variety of factors including incorrect parent-offspring bonding at birth or inadvertent misbreeding. Therefore, genetic markers seems to be an accurate alternative to trace parents of animals within the population. The 59 parentage based SNPs panel developed in this study allowed to test 121 potential sire-offspring relationships with high efficiency. However, when related individuals are used in the breeding system false paternity might be not be detected due to the increased sharing of alleles between relatives.

Since pedigree information of the 48 animals include related individuals (FS/HS), the verification of parents turn more difficult especially for full sibs (Herd 1: 2011255 and 2011256. Herd 2: 2011259 and 2011260) who are 50% related on average. However, the current panel demonstrated good efficiency of the power panel in front of half sib/full sib individuals, being capable to verify the paternity for individuals 2011259 and 2011260, whose parental relationships showed 0 loci conflict between sire-offspring genotypes. For the case of individual 201155 it was also possible to verify its paternity, however, it was excluded as father due to the 6 discordances loci found between father-offspring genotypes. Moreover, the relationship of the three animals described as half sibs in the pedigree (2009239, 2009153 and 2010422) which sire also some AI bucks within the 48 animals group were possible to verified revealing 0 loci conflicts between father-offspring genotypes. This last relationships

were corroborated by the database of the Norwegian Association of Sheep and Goat Breeders. Taking in account the level of probability of non-exclusion obtained for the current panel it is possible to conclude that the probability of no excluding the correct father it is most likely to perform.

4.4 Pedigree error rate

The percentage of rejected fathers found (15%) by the 59 SNP parentage panel is slightly high compared to the rate of paternity misidentification found in Bolormaa et al. (2008) (averaging around 12%) study of two Australian goat herds. However, this error rate is still within range of pedigree errors observed in sheep and in cattle (4.8–15.5% and 2–22% respectively) (Bolormaa et al., 2012), and since the potential breeding bucks (offspring) are recruited from ordinary production flocks, some parentage errors must be expected. From the 19 discordant relationships found, one proved to be related to a bad quality on animal genotype. In the study reported by Bolormaa et al. (2008) it was revealed that heritability increased in most of the cases when an accurate pedigree information is used. Furthermore, it was mentioned that the genetic gain is affected by 1% when one of ten sires used for breeding is absent from the analysis, therefore based on that it is expected that the error rate reported in the current study may have limited effect on the genetic evaluations.

4.5 Genotyping errors

Based on call rate, an overall of 74% of the initial 80 informative markers were performing technically well (≥90 call rate). Besides, two of the individuals show a null call rate for almost all SNPs. Lack of genotypes at many loci for an individual is a strong indication of bad DNA-quality. On the other hand, when the genotypes of the 48 AI bucks generated by the Illumina SNP chip was compared to the Sequenom data, some important genotype discrepancies were found (described in results). This seems to be a problem particularly related to the Sequenom genotypes, and the performance of the panel in terms of low

numbers of discordant genotypes were improved when these SNPs were removed. This action was possible due to the use of both genotyping technologies on a subset of the animals (the 48 AI bucks), and raise some serious questions about the reliability of apparently well performing SNPs. Without this possibility of comparing technologies, only deviations in HWE could identify this kind of poor SNPs performing.

4.6 Alternative statistical analysis for parentage test

In this study only the probability of non-exclusion has been considered. However, Baruch et al. (2008) and Anderson and Garza (2006) suggested that this method is not the only way of doing parentage analysis. Consequently, Baruch et al. (2008) proposed likelihood statistics as another option to determine parentage relationship. This likelihood estimation is defined in terms of evaluating hypotheses about data, hence, hypothesis is compared to genotypes giving a likelihood ratio. Therefore, the putative father that shares the allele with the lower frequency have a lower probability of matching by chance (IBD), and it is most likely to be the father (Blanchard, 2013 & Marshall 2007).

In the study by Ramos et al. (2009) it was considered the likelihood approach rather than the exclusion probability, which shows a major accuracy especially when the aim is to select the correct parent from a group of candidates. The likelihood confidence depends on the distribution of the log-likelihood ratio statistic, given by the allele frequencies. Although, likelihood method seem a good alternative, parentage determination turn difficult either when close relatives are considered as candidate parents. On study performed by Tesfaye (2013) it was possible to conduct an analysis were 97.3% of confidence was reached, and it was not necessary to improve the rate of relationship assignment by increasing the number of possible fathers. It is true that close relatives do lead to overestimation of confidence, however, this overestimation is pretty small under most common conditions and those overestimations can be reduced when SNPs used are highly informative as it was described in this study.

5. CONCLUSION

In this study a parentage test for goats was developed consisting of 80 potentially highly informative SNPs. After genotyping by the Sequenom iPLEX® Gold mass array system, some of the initial 80 markers were excluded from the analyses due to: low call rate, "unexpected" distribution of genotypes (HWE), and discordant genotypes between Illumina and Sequenom genotyping technologies. From the 48 AI parents and 143 potential breeding bucks genotyped in this study it was possible to evaluate 121 father-son relationships, according to the pedigree. Of these, 18 relationships (15 %) have 4 or more discordant genotypes and are considered to have wrong parentage. The percentage of parentage disproportions found in this study is relatively high, but since the potential breeding bucks are recruited from ordinary production flocks, some errors must be expected. Under theoretical consideration, the panel performed a high parentage probability of non-exclusion (EPN=0.004 with MAF of 0.44) with a single parent genotyped scenario. This 59 parentage based SNP panel could be seen as a simple but efficient exclusion parentage assignment for Norwegian goats and possibly also other breeds.

6. REFERENCES

1. Al-Atiyat, Raed M. "The power of 28 microsatellite markers for parentage testing in sheep." Electronic Journal of Biotechnology 18.2 (2015): 116-121.
2. Amills, M. (2014). The application of genomic technologies to investigate the inheritance of economically important traits in goats. Advances in Biology, 2014.
3. Ådnøy, T. (2014). The dairy goat industry in Norway: Challenges in a historical perspective. Small Ruminant Research, 122(1), 4-9.
4. Baruch, E., and J. I. Weller. "Estimation of the number of SNP genetic markers required for parentage verification." Animal genetics 39.5 (2008): 474-479.
5. Beadchips, D. N. a A. (2010). Interpreting Infinium Assay Data for Whole-Genome Structural Variation. *Analysis*, 0-9.
6. Belay Jr, T. K. (2013). Development of a SNP markers panel for parentage testing and individual identification in the Norwegian White Sheep (NWS) population.
7. Blanchard, K. A. (2013). Analysis of Single Nucleotide Polymorphism Panels for Bovine DNA Identification.
8. Bolormaa, S., et al. "DNA-based parentage verification in two Australian goat herds." Small Ruminant Research 80.1 (2008): 95-100.
9. Børsting, C., Sanchez, J. J., Hansen, H. E., Hansen, A. J., Bruun, H. Q., & Morling, N. (2008). Performance of the SNPforID 52 SNP-plex assay in paternity testing. Forensic Science International: Genetics, 2(4), 292-300.
10. Cutoff, G. S. Infinium ® Genotyping Data Analysis.
11. Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., ... & Wang, W. (2013). Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). Nature biotechnology, 31(2), 135-141.
12. Fernández, M. E., Goszczynski, D. E., Lirón, J. P., Villegas-Castagnasso, E. E., Carino, M. H., Ripoli, M. V., ... & Giovambattista, G. (2013). Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. Genetics and molecular biology, 36(2), 185-191.
13. Fisher, P. J., et al. "The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds." Journal of dairy science 92.1 (2009): 369-374.
14. Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies.
15. 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(2), 152-157.
16. Harlizius, B., et al. "A single nucleotide polymorphism set for paternal identification to reduce the costs of trait recording in commercial pig breeding." Journal of animal science 89.6 (2011): 1661-1668.

17. Hashemi, Mona, et al. "Validation of Pedigree Relationships Using a Multiplex Microsatellite Marker Assay in Iranian Holstein Cattle/Walidacja rodowodów irańskiego bydła holsztyńskiego przy użyciu analizy multipleks markerów mikrosatelitarnych." Annals of Animal Science 13.3 (2013): 481-493.
18. Hayes, B. J. "Technical note: Efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data." Journal of dairy science 94.4 (2011): 2114-2117.
19. Heaton, Michael P., et al. "Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle." *Mammalian Genome* 13.5 (2002): 272-281.
20. Heaton, M. P., Leymaster, K. A., Kalbfleisch, T. S., Kijas, J. W., Clarke, S. M., McEwan, J., ... & Chitko-McKown, C. G. (2014). SNPs for parentage testing and traceability in globally diverse breeds of sheep. PloS one, 9(4), e94851.
21. Hildebrand C.E., Torney D.C., and Wagner R.P. 1992. Informativeness of Polymorphic DNA Markers. Los Alamos Science 20, 100-102.
22. Illumina. (2010). Infinium HD Assay Methylation Protocol Guide. Retrieved from papers2://publication/uuid/9B92CF4C-DF14-4790-8B9E-495380DA3614
23. Illumina. GenomeStudio Data Analysis Software, 2013.
24. 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(1), 181-187.
25. iPLEX® Gold Application Guide July 22, 2009
26. Luna-González, A., et al. "Microsatellite loci and paternity analysis in Nubia and Boer goats." Arch Med Vet 44 (2012): 123-127.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.
28. Ramos, A. M., Crooijmans, R. P. M. A., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., ... & Groenen, M. A. (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. PloS one, 4(8), e6524.
29. Rosa, A. J. M., Sardina, M. T., Mastrangelo, S., Tolone, M., & Portolano, B. (2013). Parentage verification of Valle del Belice dairy sheep using multiplex microsatellite panel. Small Ruminant Research, 113(1), 62-65.
30. Řehout, V., E. Hradecká, and J. Čítek. "Evaluation of parentage testing in the Czech population of Holstein cattle." Czech J. Anim. Sci 51.12 (2006): 503-509.
31. Senneke, S. L., MacNeil, M. D., & Van Vleck, L. D. (2004). Effects of sire misidentification on estimates of genetic parameters for birth and weaning weights in Hereford cattle. *Journal of animal science*, *82*(8), 2307-2312.
32. Schütz, Ekkehard, and Bertram Brenig. "Analytical and statistical consideration on the use of the ISAG-ICAR-SNP bovine panel for parentage control, using the Illumina

BeadChip technology: example on the German Holstein population." Genetics Selection Evolution 47.1 (2015): 3.

33. Strucken, E. M., Gudex, B., Ferdosi, M. H., Lee, H. K., Song, K. D., Gibson, J. P., Kelly, M., et al. (2014). Performance of different SNP panels for parentage testing in two East Asian cattle breeds. Animal Genetics, 45(4), 572-575. Blackwell Publishing Ltd.
34. Tosser-Klopp, G., Bardou, P., Bouchez, O., Cabau, C., Crooijmans, R., Dong, Y., ... & International Goat Genome Consortium. (2014). Design and characterization of a 52K SNP chip for goats. PLoS One, 9(1).
35. Van Eenennaam, A. L., et al. "DNA-based paternity analysis and genetic evaluation in a large, commercial cattle ranch setting." Journal of Animal Science 85.12 (2007): 3159-3169.

APPENDIX

Table 1. List of the mismatching SNP markers from Illumina-Sequenom genotypes contrast. This table shows the 11 SNPs which were removed as consequence of genotype discordances between Illumina and Sequenom genotyping methods. In the first part is observed the SNP ID, followed by the 3 possible genotypes (e.g. AA, AG, GG) column for each marker. Subsequently it is shown the number of individuals for each genotype on Sequenom and Illumina. The column titled as DISC reveal the number of conflicts found between Illumina and Sequenom technologies.

SNP	Alleles	Geno (sum)	Sequenom	Illumina	PLINK	DISC	MAF
10219-scf1368-489857	AA	8	8	10	9/29/10	2	0.4896
	AG	29	31	29			
	GG	9	9	9			
12947-scf1499-2414292	CC	13	13	13	11/24/13	7	0.4896
	CT	17	17	24			
	TT	11	18	11			
18180-scf185-15961025	AA	12	14	12	12/24/12	2	0.5
	AC	21	21	24			
	CC	10	10	12			
18251-scf1850-107670	AA	13	13	13	13/22/13	1	0.5
	AG	20	20	22			
	GG	13	14	13			
24678-scf2510-227964	AA	8	8	8	8/31/9	20	0.4792
	AG	11	11	31			
	GG	9	29	9			
32696-scf377-530544	CC	9	9	10	10/27/11	4	0.5
	CT	23	23	27			
	TT	11	15	11			
36532-scf44-1147631	CC	12	28	12	12/24/12	16	0.5
	CT	8	8	24			
	TT	11	11	12			

37317-scf455-1220426	AA	10	10	11	11/25/12	13	0.4896
	AT	11	11	25			
	CC	12	25	12			
45689-scf627-2974066	CC	11	11	12	12/23/13	13	0.4896
	CT	10	10	23			
	TT	13	26	13			
47853-scf674-504140	AA	12	18	12	12/23/13	6	0.4896
	AC	17	17	23			
	CC	13	13	13			
58755-scf959-748556	AA	9	17	9	9/30/9	8	0.4896
	AG	20	20	30			
	GG	9	9	9			

Table 2. List of the number of discordances on the 121 sire-offspring relationships. In the table it is exposed three different panels (74 SNPs, 61 SNPs and 59 SNPs) used for sire-offspring relationship analysis. Under each panel it is expressed the number of loci conflicts per sire-offspring relationship found during the parentage analysis described on methods and results.

	74 SNPs	61 SNPs	59 SNPs
sire-offspring	discords	discords	discords
1	16	14	13
2	15	12	11
3	14	12	11
4	14	11	10
5	13	10	10
6	13	10	9
7	12	10	9
8	11	9	9
9	10	8	8
10	10	8	8
11	9	7	7
12	9	7	7
13	9	7	6
14	9	6	6
15	8	6	6
16	8	5	5

17	7	4	4
18	6	4	4
19	3	1	1
20	3	0	0
21	2	0	0
22	2	0	0
23	2	0	0
24	2	0	0
25	2	0	0
26	2	0	0
27	2	0	0
28	2	0	0
29	2	0	0
30	2	0	0
31	2	0	0
32	1	0	0
33	1	0	0
34	1	0	0
35	1	0	0
36	1	0	0
37	1	0	0
38	1	0	0
39	1	0	0
40	1	0	0
41	1	0	0
42	1	0	0
43	1	0	0
44	1	0	0
45	1	0	0
46	1	0	0
47	1	0	0
48	1	0	0
49	1	0	0
50	1	0	0
51	1	0	0
52	1	0	0
53	1	0	0
54	1	0	0
55	1	0	0
56	1	0	0
57	1	0	0
58	0	0	0
59	0	0	0
60	0	0	0
61	0	0	0
62	0	0	0

63	0	0	0
64	0	0	0
65	0	0	0
66	0	0	0
67	0	0	0
68	0	0	0
69	0	0	0
70	0	0	0
71	0	0	0
72	0	0	0
73	0	0	0
74	0	0	0
75	0	0	0
76	0	0	0
77	0	0	0
78	0	0	0
79	0	0	0
80	0	0	0
81	0	0	0
82	0	0	0
83	0	0	0
84	0	0	0
85	0	0	0
86	0	0	0
87	0	0	0
88	0	0	0
89	0	0	0
90	0	0	0
91	0	0	0
92	0	0	0
93	0	0	0
94	0	0	0
95	0	0	0
96	0	0	0
97	0	0	0
98	0	0	0
99	0	0	0
100	0	0	0
101	0	0	0
102	0	0	0
103	0	0	0
104	0	0	0
105	0	0	0
106	0	0	0
107	0	0	0
108	0	0	0

109	0	0	0
110	0	0	0
111	0	0	0
112	0	0	0
113	0	0	0
114	0	0	0
115	0	0	0
116	0	0	0
117	0	0	0
118	0	0	0
119	0	0	0
120	0	0	0
121	0	0	0

Table 3. List of 19 conflictive animals found during parentage-offspring concordance analysis. The list shows the 19 animals whose parentage exclusion results revealed more than 1 loci conflict between sire-offspring genotypes, and the animal whose DNA quality was not good (only 2 informative SNPs of 59) to determine a correct parentage analysis.

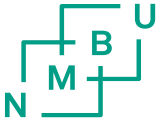
Animals' ID	Parent	Discords	Info. sites
14220021_5532	2012247	13	59
15251213_5127	2010169	11	59
14450495_5044	2010169	11	59
14450495_5051	2010169	10	59
15240744_5300	2012247	10	59
14450495_5047	2010169	9	59
14220021_5525	2011181	9	59
15190183_5121	2011181	9	59
15190183_5138	2012151	8	58
14290214_5127	**2011259**	8	59
15190183_5189	2010435	7	57
14290214_5036	2012188	7	58
15250427_5127	**2011255**	6	59
14450495_5048	2009106	6	59
14220021_5504	2011181	6	59
14450495_5041	2010390	5	58
14450495_5040	2010390	4	59
18710144_5492	2009106	4	59
14491367_5045	2010390	1	2

Table 4. List of the 121 sire-offspring relationships analysed. The table shows the father established according to the number of loci discordances between sire and offspring genotypes.

sire-offspring	74 SNPs		61 SNPs		59 SNPs	
	#ID	parent	#ID	parent	#ID	parent
1	14220021_5532	2012247	14220021_5532	2012247	14220021_5532	2012247
2	15240744_5300	2012247	15251213_5127	2010169	15251213_5127	2010169
3	15251213_5127	2010169	14450495_5044	2010169	14450495_5044	2010169
4	14450495_5044	2010169	15240744_5300	2012247	15240744_5300	2012247
5	15190183_5121	2011181	14220021_5525	2011181	14450495_5051	2010169
6	14450495_5047	2010169	14450495_5051	2010169	14220021_5525	2011181
7	14450495_5051	2010169	14450495_5047	2010169	15190183_5121	2011181
8	14290214_5127	2011259	15190183_5121	2011181	14450495_5047	2010169
9	14220021_5525	2011181	15190183_5138	2012151	15190183_5138	2012151
10	15190183_5189	2010435	14290214_5127	2011259	14290214_5127	2011259
11	14290214_5036	2012188	14290214_5036	2012188	14290214_5036	2012188
12	15190183_5138	2012151	15250427_5127	2011255	15190183_5189	2010435
13	14220021_5504	2011181	15190183_5189	2010435	15250427_5127	2011255
14	14450495_5048	2009106	14220021_5504	2011181	14220021_5504	2011181
15	14450495_5040	2010390	14450495_5048	2009106	14450495_5048	2009106
16	14450495_5041	2010390	14450495_5041	2010390	14450495_5041	2010390
17	15250427_5127	2011255	14450495_5040	2010390	14450495_5040	2010390
18	18710144_5492	2009106	18710144_5492	2009106	18710144_5492	2009106
19	14170177_5095	2011183	14491367_5045	2010390	14491367_5045	2010390
20	14210022_5308	2010435	14310299_5409	2012247	08343494_5073	2012247
21	12116050_5009	2011260	14310299_5408	2012247	08343494_5079	2012247
22	15250427_5124	2011259	14220021_5530	2012247	08343494_5080	2012247
23	04410327_5200	2010435	14210022_5338	2012247	12116050_5017	2012247
24	14210022_5337	2010435	12240963_5068	2012247	14210022_5338	2012247
25	12115681_5108	2010390	12240963_5066	2012247	12240963_5066	2012247
26	19330742_5111	2010390	12116050_5017	2012247	12240963_5068	2012247
27	15250427_5106	2010390	08343494_5080	2012247	14220021_5530	2012247
28	14491367_5045	2010390	08343494_5079	2012247	05200092_5046	2012247
29	15251213_5128	2010169	08343494_5073	2012247	05200092_5047	2012247
30	04410327_5196	2010166	05200092_5052	2012247	05200092_5051	2012247
31	14450495_5050	2010138	05200092_5051	2012247	05200092_5052	2012247
32	06190082_5049	2012151	05200092_5047	2012247	14310299_5408	2012247
33	15190183_5145	2012151	05200092_5046	2012247	14310299_5409	2012247
34	12115681_5112	2012130	08343494_5078	2012188	08343494_5078	2012188
35	14210134_5152	2012130	15190183_5145	2012151	06190082_5049	2012151
36	14220021_5536	2012130	14220021_5535	2012151	12116050_5010	2012151
37	14210134_5151	2012130	14210050_5011	2012151	14170177_5099	2012151
38	18710144_5490	2012130	14170177_5099	2012151	14210050_5011	2012151
39	18710144_5491	2011456	12116050_5010	2012151	14220021_5535	2012151
40	14290214_5105	2011259	06190130_5120	2012151	06190130_5120	2012151
41	15250479_5149	2011259	06190082_5049	2012151	15190183_5145	2012151

42	14290214_5126	2011183	18710144_5490	2012130	12115681_5112	2012130
43	04410327_5195	2010435	14220021_5536	2012130	12115681_5116	2012130
44	12115681_5107	2010435	14210134_5152	2012130	06190082_5050	2012130
45	14210022_5316	2010435	14210134_5151	2012130	14220021_5536	2012130
46	2011331	2010422	12115681_5116	2012130	14210134_5151	2012130
47	04410327_5197	2010390	12115681_5112	2012130	14210134_5152	2012130
48	14210080_5087	2010390	06190082_5050	2012130	18710144_5490	2012130
49	14491367_5047	2010169	18710144_5491	2011456	04371103_5352	2011456
50	15251213_5129	2010169	15190183_5186	2011456	14210022_5306	2011456
51	15250427_5133	2010169	14210022_5306	2011456	15190183_5186	2011456
52	12110191_5017	2010166	05440152_5082	2011456	05440152_5082	2011456
53	14450085_5042	2010166	04371103_5352	2011456	18710144_5491	2011456
54	15250311_5200	2010166	14290214_5088	2011260	12116050_5009	2011260
55	14450085_5048	2010166	12116050_5009	2011260	14290214_5088	2011260
56	05440152_5097	2010166	15250479_5149	2011259	15250427_5124	2011259
57	2011105	2010138	15250427_5141	2011259	15250427_5141	2011259
58	05200092_5046	2012247	15250427_5124	2011259	15250479_5149	2011259
59	05200092_5047	2012247	14290214_5105	2011259	14290214_5105	2011259
60	05200092_5051	2012247	15240744_5418	2011183	14170177_5098	2011183
61	05200092_5052	2012247	14290214_5126	2011183	14170177_5095	2011183
62	08343494_5073	2012247	14170177_5098	2011183	14290214_5126	2011183
63	08343494_5079	2012247	14170177_5095	2011183	15240744_5418	2011183
64	08343494_5080	2012247	06190130_5258	2011183	06190130_5258	2011183
65	12116050_5017	2012247	15251213_5112	2011181	15251213_5112	2011181
66	12240963_5066	2012247	15190183_5122	2011181	08341817_5009	2011181
67	12240963_5068	2012247	08341817_5009	2011181	15190183_5122	2011181
68	14210022_5338	2012247	15190183_5181	2010435	04410327_5194	2010435
69	14220021_5530	2012247	15190183_5108	2010435	04410327_5195	2010435
70	14310299_5408	2012247	15190183_5107	2010435	04410327_5200	2010435
71	14310299_5409	2012247	14210022_5337	2010435	12115681_5114	2010435
72	08343494_5078	2012188	14210022_5316	2010435	12115681_5115	2010435
73	06190130_5120	2012151	14210022_5308	2010435	12115681_5107	2010435
74	12116050_5010	2012151	14210022_5301	2010435	14210022_5301	2010435
75	14170177_5099	2012151	14170177_5097	2010435	14210022_5308	2010435
76	14210050_5011	2012151	12115681_5115	2010435	14210022_5316	2010435
77	14220021_5535	2012151	12115681_5114	2010435	14210022_5337	2010435
78	06190082_5050	2012130	12115681_5107	2010435	14170177_5097	2010435
79	12115681_5116	2012130	04410327_5200	2010435	15190183_5107	2010435
80	04371103_5352	2011456	04410327_5195	2010435	15190183_5108	2010435
81	05440152_5082	2011456	04410327_5194	2010435	15190183_5181	2010435
82	14210022_5306	2011456	2011331	2010422	2011331	2010422
83	15190183_5186	2011456	19330742_5111	2010390	15250427_5106	2010390
84	14290214_5088	2011260	15250427_5106	2010390	04410327_5197	2010390
85	15250427_5141	2011259	15190183_5103	2010390	12115681_5119	2010390
86	06190130_5258	2011183	15190183_5101	2010390	14450495_5042	2010390
87	14170177_5098	2011183	14450495_5042	2010390	12115681_5108	2010390

88	15240744_5418	2011183	14210080_5087	2010390	14210080_5087	2010390
89	08341817_5009	2011181	12115681_5119	2010390	19330742_5111	2010390
90	15190183_5122	2011181	12115681_5108	2010390	15190183_5101	2010390
91	15251213_5112	2011181	08334510_5004	2010390	15190183_5103	2010390
92	04410327_5194	2010435	04410327_5197	2010390	08334510_5004	2010390
93	12115681_5114	2010435	15251213_5129	2010169	15250427_5133	2010169
94	12115681_5115	2010435	15251213_5128	2010169	15251213_5128	2010169
95	14170177_5097	2010435	15250427_5133	2010169	15251213_5129	2010169
96	14210022_5301	2010435	14491367_5047	2010169	14450495_5045	2010169
97	15190183_5107	2010435	14491367_5043	2010169	14491367_5042	2010169
98	15190183_5108	2010435	14491367_5042	2010169	14491367_5043	2010169
99	15190183_5181	2010435	14450495_5045	2010169	14491367_5047	2010169
100	08334510_5004	2010390	14210134_5161	2010169	14210134_5161	2010169
101	12115681_5119	2010390	14210134_5153	2010169	14210134_5153	2010169
102	14450495_5042	2010390	05130050_5025	2010169	05130050_5025	2010169
103	15190183_5101	2010390	15250479_5126	2010166	15250479_5122	2010166
104	15190183_5103	2010390	15250479_5122	2010166	15250479_5126	2010166
105	05130050_5025	2010169	15250311_5200	2010166	04410327_5196	2010166
106	14210134_5153	2010169	14450085_5048	2010166	12110191_5017	2010166
107	14210134_5161	2010169	14450085_5042	2010166	14450085_5042	2010166
108	14450495_5045	2010169	12110191_5017	2010166	14450085_5048	2010166
109	14491367_5042	2010169	05440152_5097	2010166	15250311_5200	2010166
110	14491367_5043	2010169	05130050_5020	2010166	05440152_5097	2010166
111	05130050_5020	2010166	04410327_5196	2010166	05130050_5020	2010166
112	15250479_5122	2010166	14450495_5050	2010138	2011105	2010138
113	15250479_5126	2010166	14450495_5049	2010138	14450495_5050	2010138
114	05130050_5045	2010138	05130050_5049	2010138	14450495_5049	2010138
115	05130050_5048	2010138	05130050_5048	2010138	05130050_5045	2010138
116	05130050_5049	2010138	05130050_5045	2010138	05130050_5048	2010138
117	14450495_5049	2010138	2011105	2010138	05130050_5049	2010138
118	2010317	2009239	2010317	2009239	2010317	2009239
119	2012247	2009153	2012247	2009153	2012247	2009153
120	14450495_5043	2009106	15190183_5195	2009106	14450495_5043	2009106
121	15190183_5195	2009106	14450495_5043	2009106	15190183_5195	2009106



Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no