University of Natural Resources
and Applied Life Sciences, Vienna

Department of Sustainable
Agricultural Systems
Division of Livestock Sciences

Genetic Contribution of Important Ancestors in a Cattle Population

Dinesh Moorkattukara Thekkoot

Supervisor
Prof. Dr. Georg Thaller

Co - Supervisor
Prof. Dr. Johann Soelkner

Vienna, June 2009

Dedicated to my family,
who offered me unconditional love and support
throughout the course of this thesis

Acknowledgements

I am greatly indebted to ***Prof. Dr. Georg Thaller****, Professor and Head, Institute of Animal breding, Christian Albrechts University, Kiel and the Principal Supervisor of my thesis for his keen interest, excellent guidance and encouragement rendered during the course of this investigation.*

I wish to express my deep sense of gratitude to ***Prof. Dr. Johann Solkner****, Professor and Head, Department of Sustainable Agricultural Systems, University of Natural Resources and Applied Life Sciences, Vienna for the enthusiastic guidance, valuable suggestions and constructive criticism and motivation during the course of the study. His guidance helped me in all the time of research and writing of this thesis.*

My profound and sincere thanks to Dr. David Habier, Christian Albrechts University, Dr. Birgit Gredler, University of Natural Resources and Applied Life Sciences, Vienna and Dr. Christian Frst, ZuchtData EDV-Dienstleistungen GmbH for their useful suggestions and help rendered during this investigation.

My sincere thanks are due to the authorities and staff of Christian Albrechts University,Kiel, University of Natural Resources and Applied Life Sciences, Vienna and ZuchtData EDV-Dienstleistungen GmbH for providing all the facilities to carry out this study.

I am deeply grateful to Arun Kommadath, who informed me about the Erasmus Mundus scholarship and helped a lot during the process

of application. The financial support of the Erasmus Mundus Consortium is gratefully acknowledged.

I owe my gratitude to all the members of staff of the Wageningen University, EMABG Secretariat, Netherlands and Kerala Livestock Development Board, India for their whole hearted cooperation and assistance throughout the research.

My grateful thanks also goes to Dr. Jose James, General Manager, Kerala Livestock Development Board and Dr. C. P. Gopakumar, DLF Ranches Pvt. Ltd. for offering all the support and encouragement during the course of this thesis work.

During this work I have collaborated with many colleagues for whom I have great regard. I sincerely thank my beloved friends Binyam Simme, Syed Nurrdien, Astrid Koeck, Kahsay Geebramariyam, Gabor Meszaros, Luis Emmanuel, Esau Galukande, Mulindwa Henry, Bimal Bhaskaran, Biju Sivan and Roopika Menon for making my thesis program a memorable one.

I am heartily thankful Mr. Bakul Shah and many other unknown people from different parts of the world, who helped me a lot in my SAS programming.

I owe my loving thanks to my parents,my wife Smitha, my sons Pranav and Nikhil. They have lost a lot due to my research abroad. Without their encouragement and understanding it would have been impossible for me to finish this work.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the course of this thesis program.

Dinesh Moorkattukara Thekkoot

Abstract

Estimating genomic similarities among animals belonging to different generations is one of the fundamental challenges for animal breeders. However no empirical measure of his relatedness throughout the whole genome of cattle has yet been published. By utilising the Single Nucleotide Polymorphism information we, in this thesis, describes a genome wide and chromosome wide probability of sharing one of the allele to be IBD between the grandparents and grandsons. Genotypic information was collected from 1764 breeding bulls belonging to the Simmental breed using 50K chip from Illumina. The genome wide probability for one of the allele to be IBD didn't differed significantly from the expectations and the its correlation with the breeding values for different traits didn't manifest a particular trend. The results from the analysis performed to estimate the length, number and size of homozygous chromosomal segments in inbred and non inbred bulls shows a very strong inclination that these parameters are strongly influenced by inbreeding. Finally in this study, in order to estimate the correlation between pedigree relationship and genomic relationship, we compared the pair wise coefficient of coancestry calculated using pedigree relationship to the genomic parameters like IBD of 0, 1 and 2 allele, total proportion IBD and the number of non missing loci which are in identity by state for either 0, 1 or 2 allele. A medium level of correlation was observed for some of the genomic traits to the pedigree relationship.

Contents

List of Tables

Chapter 1

Introduction

Estimating genomic similarities among animals belonging to different generations is one of the fundamental challenges for animal breeders. A wide range of strategies have been developed by scientists for predicting the genomic share carried in a animal from its parent or grand parent. Pedigree analysis was one of the basic techniques used for this. If sufficient data is available in the pedigree, the contribution of the ancestor to the current generation is determined by means of estimating the genes that are passed from parents to progeny. The parents in one generation will pass 50 % of their genes to their offspring, but while passing one set of chromosomes, it will include a selection of ones he inherited from both parents. But here there is no guarantee that selection will be exactly equal. The estimates of contributions as therefor probabilities. The situation is complicated by the fact that most of the domestic animals have overlapping generations and complex pedigrees *i. e.* certain superior sires may appear more than once in the pedigree of an animal and in certain extreme cases some bulls may be found in the pedigree of more than 90% of the animals in the population.

The advancements in the field of molecular genetics helped scientists in solving this complex situation. Initially microsatellite markers were widely used and later with the advent of high density Single Nucleotide Polymorphism (SNP) chips these were used for this purpose. DNA polymorphism, the fundamental property of the markers, is used for genetic mapping studies. The same principle is applied, while using DNA markers to estimate the founder's genomic contribution to its

descendents. The entire genome of an individual is assumed to be a mixture of chromosomal segments each derived from different ancestors / founders.

Estimating the Identity by Descent(IBD) probabilities, *i.e.* a pair of related individuals that share 0,1 or 2 alleles IBD at a point along a chromosome, is another method for mapping ancestral contributions when two alleles at any given locus are inherited from a common ancestor, those genes are said to be in IBD. A parent and an offspring share exactly 50% of the autosomal loci but the proportion of share between a Grand parent and grand child need not be exactly 25%, as the parents will not transfer an exact proportion of what they received from their sire and dam.

Even though various scientists have explained the theoretical distribution of IBD sharing over the whole genome, especially in human beings, no detailed study of the IBD sharing, using molecular markers, among individuals belonging to different generations has yet been published. But recent advances in the field of micro array technology,like single nucleotide polymorphism chips has made this possible.

Using the data from Single Nucleotide Polymorphism (SNP) from Simmental bulls, we present here a genome wide and chromosome wide probability of sharing one of the allele to be IBD between the grandparents and grandsons along with their correlation to breeding values. In this study we also provide some basic information about the homozygous segments of various lengths in inbred and non inbred bulls. Finally we briefly asses the correlation between the Coefficient of Coancestry (estimated by pedigree relationship) with the genomic parameters like IBD and IBS for the entire group of genotyped bulls. Our aim is to furnish a baseline information about the variation of IBD and correlation between the pedigree relationship with the genomic relationship.

Chapter 2

Review of Literature

2.1 Identity by Descent

Identity by Descent (IBD) is a method for determining genomic similarities among relatives. When two alleles at a single locus are identical copies of the same allele in some earlier generation they are said to be identical by descent. In other words both are copies that arose by DNA replication from the same ancestral sequence with out any intervening mutation. This concept of IBD was introduced by Malecot (1941) and is widely used in many mapping studies in genetics. Gagnon et al. (2005) while quoting other authors claim that this concept has triggered the development of Quantitative Trait Loci (QTL) mapping, linkage analysis based on allele sharing and multipoint interval mapping. In 2002 Nagamine et al. introduced a simple and deterministic method for the detection of QTL and the evaluation of allelic effects via variance components method by utilising IBD coefficients in full sib and half sib families using simulated dataset. According to Cannings (2003) the probabilities of the various possible identity by descent states at a locus capture all the genealogical information for that locus for a set of individuals under consideration. IBD analysis involves investigating whether the proportion of affected individuals sharing 0, 1 or 2 alleles IBD at a marker locus differs significantly from expectations assuming a null model in which there is no linkage between the marker and the disease (Motro and Thomson, 1985). Hence IBD analysis is used mainly in epidemiological genetics in humans and for

tracing recessive traits in livestock.

2.2 IBD Analysis in Cattle

As in Human beings IBD studies in livestock are also mainly intended for studying recessive traits and also for fine mapping QTL, especially in combination with linkage information. Very few papers were found relating the IBD analysis and productivity taits in cattle. One of the first IBD related studies reported in the livestock was by Charlier et al. (1996). They used the Identity by Descent mapping strategies to map the gene causing bovine hereditary syndactyly to chromosome 15 of cattle. For this study they used a battery of 213 micro satellites spanning the 29 bovine autosomes. Later Riquet et al. (1999) used IBD analysis to investigate milk production of Holstein - Friesian cattle. They used this method to refine the map position of a Quantitative Trait Locus (QTL) present on bovine chromosome number 14. In continuation with the QTL studies Meuwissen and Goddard (2000) proposed a haplotype based method to fine map QTL using a mixed linear model on the basis of similarity of their marker haplotypes based on the assumption that individuals with similar marker haplotypes are likely to share QTL alleles that are in IBD. Grapes et al. (2006) developed IBD based methods to fine map QTL in a previously defined QTL region.

By 2008 with the availability of high density SNP chips, Druet et al. (2008) used IBD analysis to fine map fertility quantitative trait loci, which is of high interest in dairy cattle industry. Tarres et al. (2009)published a new strategy for QTL fine mapping by making use of statistical methods combining linkage and linkage disequilibrium analysis. This method estimated the IBD probabilities among the base haplotypes used for grouping.

2.3 Methods for estimating IBD probabilities

A large number of methods have been developed for estimating IBD probabilities. In the beginning it was estimated by analysing the pedigree data. Later

on with the discovery of markers algorithms were developed for the analysis of IBD utilising the marker information. One among the first stable algorithm developed was by Elston and Stewart in 1971 developed an approach that was later called as Elston-Stewart Algorithm. This algorithm was later used in the pedigree analysis software package VITESSE (O'Connell and Weeks, 1995). Even though it was appropriate for large pedigrees, there were limitations in the number of markers that can be used. To overcome these shortfalls another method based on Langer Green Algorithm was developed by Kruglyak et al. (1996) in 1996, in which multipoint linkage analysis was done using many markers by means of non parametric methods which takes into account the entire pedigree information. The algorithm had time and memory constraints proportional to the number of loci. Due to these limitations alternative methods using approximations were developed. The Markov Chain Monte Carlo method which is suitable for small or large pedigrees (Heath, 1997). These methods became computationally infeasible when pedigrees were large and complex and with many generations of individuals with no data (Abney, 2008).

In 2001 Meuwissen and Goddard published a new method for estimating the IBD probabilities at a given chromosomal location, utilising the data available on the haplotype of markers spanning that chromosomal region.The method developed was independent of the pedigree information, all the information come from the marker genotypes. This procedure was also applicable in a situation where some generations of unknown pedigree was followed by some generations where pedigree and marker genotypes were know, thereby making this algorithm highly flexible.Sobel et al. (2001) used an Markov Chain Monte Carlo algorithm for multipoint IBD probabilities at arbitrary positions among marker loci for general pedigrees. Their algorithm was used in the SIMWALK2 computer package. While validating their algorithm, they suggested that multipoint IBD estimation is much better than single point estimation

With the introduction of high density Single Nucleotide Polymorphism map the size of the data sets increased enormously and beyond the capabilities of many of the then existing computational tools. Abecasis et al. (2002) introduced a

highly efficient Langer -Green Algorithm for estimating IBD probabilities, allele sharing analysis and for haplotyping. Based on this algorithm they developed the state-of-art software called MERLIN (multi-point engine for rapid likelihood inference)for evaluating the IBD probabilities in small to medium sized pedigrees using high density marker data. The assumptions used in this package is that paternal haplotypes are drawn from a population in linkage equilibrium. *i.e.* the allele occurring at any given locus in the haplotype is assumed to be independent of the alleles occurring at all other loci. Later Merlin was modified by taking into account the linkage disequilibrium (Abecasis and Wigginton, 2005) while analysing multipoint analysis of pedigrees. The new version identifies marker groups that represent a haplotype block and it estimates the frequency of each in each group.

Keith et al. (2008) introduced a new algorithm and estimated the sharing of chromosome 15 between 169 monozygotic or dizygotic twins from Australian families. They compared this new algorithm with MERLIN and found that the accuracy of the results obtained was near to MERLIN. They concluded that ignoring linkage disequilibrium in founder haplotypes can cause errors in the calculation of IBD probabilities.

2.4 Genetic contribution

The estimation of founder contribution in population began with the introduction of theory of junctions by Fisher in in 1954. The theory of junctions traces parental chromosomal blocks in inbred populations. Later when artificial insemination(AI) became the most important tool for many breeding programs to improve beef and dairy cattle production, the genetic variability in the population decreased. The number of offsprings per male increased drastically and this in turn decreased the gene pool and genetic variability of the population. At this juncture scientists in the field of genetics, across the globe again became interested in estimating founder contribution. For cattle much of these kinds of work was based on analysing the pedigree of the concerned population. Lacy (1989) developed a founder equivalent approach which combine the information of the

founder animals contributing to the population under study and estimated the number of equally contributing ancestors that would provide the same level of genetic diversity.

Wray et al. (1990) estimated the long term genetic contribution *i. e.* ultimate proportional contribution of the ancestor to generations in the distant future using pedigree information. They concluded that after several generations the genetic contributions of ancestors stabilise and become equal for all individuals in that and subsequent generations of descendants, but values differ between ancestors.Bijma and Woolliams (1999) introduced a method to predict the long term genetic contribution of ancestors to future generations for a population with overlapping generations under mass or selection index. Roughsedge et al. (1999) described different techniques that can be used to quantify the genetic contribution to a population of UK Holstein - Friesian cattle by means of pedigree analysis. Baumung and Solkner (2003) described methods for estimating genetic variability by utilising marker and pedigree information. They calculated various genetic parameters for an endangered population using simulated data. Royo et al. (2007) estimated the genomic contributions of endangered Asturcn pony founders to the present generation by using pedigree information along with 15 microsatellite markers.

No many studies have been reported in cattle on genomic contributions studies using high density markers. Many studies of this kind are reported in human genetics. Gagnon et al. (2005) described a method to estimate the genetic similarity among the relatives. They analysed the proportion of alleles shared by siblings at highly polymorphic microsatellite loci present on the 22 autosomal chromosomes among eight *Centre d'etudes du polymorphisme Humain* (CEPH) families. They studied 88 sibpairs from these families to estimate how much genome that they shared among themselves were IBD. To account for the varying chromosomal lengths and recombination rates the analysis was performed at chromosomal level and marker level.

Williams and Reyes-Valds (2007) estimated the founder proportion of genome

in the second generation of outbred *Pinus taeda* pedigree using a donor recurrent method. The founder proportion ranged from 1.54% to 48.46% with a mean value of 17.59% well below the expected value of 25%.

2.5 Homozygous Chromosomal Segments

If both the alleles at a particular marker are identical, then that individual can be called as homozygous for that particular marker. When we look into a set of linked markers, if an individual is homozygous at a large number of continuous markers, it is likely that the individual is autozygous for that segment i. e., that the two chromosomal segments have a common origin (Broman and Weber, 1999). According to Clark (1999) the length of these homozygous segments depend on various factors in a complicated way. The factors include mutation rate, effective population size, effect of mutation on reproductive fitness, population subdivision and growth and on the patterns of inbreeding in the population. Gibson et al. (2006) concluded that the main reason for this phenomenon is inbreeding in which an individual inherits chromosomal segments that are identical by descent from each parent. Recombination can cause break up of chromosomal segments over generations. The longest tracts of these homozygous segments are therefore to be expected in populations with an appreciable degree of inbreeding.

Chapter 3

Materials and Methods

3.1 Data and Sampling

The bulls selected for this study were a part of the Genomic Selection Project of ZuchtData Corporation, the breeding organisation in Austria and from Genotrack, a research project on use of bovine high density SNP chips in Germany. As a part of two projects on genomic selection important Simmental bulls which are currently used and some prominent bulls which were used previously were selected and genotyped. In total 1764 bulls were genotyped in Austria and 480 bulls were analysed in Germany. The genotyping was done using 50 k SNP chip from Illumina and in total 54001 SNP's were genotyped. The genotyping of these bulls was done in two different laboratories, and as a result some genotype incompatibility was noticed. Due to time constraints, it was not possible to detect the incompatibility of these datasets, and hence for this study we used only the 1764 bulls genotyped for ZuchtData Corporation. As the bulls were selected for the Genomic Selection Project, the data set was not ideal enough for estimating IBD probabilities. Hence the analysis was done with the available set of bulls.

Five important Simmental bulls, namely *Haxl, Morello, Redad, Horror and Streif* were selected as important ancestors for estimating the genetic contribution to the current generation of bulls. During analysis it was noticed that Haxl was present in the pedigree of more than 90% of the bulls used and so it was dropped from the list. Bull Horror was genotyped by Genotrack and so it was not included

in this study. Finally only three bulls viz. Morello Streif and Redad were used for estimating the genetic contributions.

3.2 Selection of SNP's

54001 SNP's were genotyped in these animals. For the current study, all the yet unpositioned SNP's i.e., those SNP's with chromosome number and base pair position denoted as zero and those on the sex chromosome were removed, and there were 51515 SNP's remaining. Uninformative SNP's i.e., SNP's which are monomorphic in the population were also deleted. In total 3006 SNP's were deleted as they were monomorphic and finally 48509 SNP's placed on chromosome one to 29 were used for the current analysis.

3.3 IBD Estimation

The analysis was aimed to estimate the chromosome wide and genome wide IBD sharing between the important bulls and their grand sons. Bulls Morello and Streif were selected for this analysis. As none of the sons of Redad was genotyped, it was not used for the purpose of this study.

For Morello, 12 of its sons were genotyped and from among them, five had their sons (grand sons of Morello) also genotyped. In total 13 grand sons were genotyped. The IBD analysis the gnotypes of Morello, its three sons and 13 grand sons were used. The analysis was done by assigning different family numbers to these 13 bulls (grand sons of Morello). For each of these families contained seven animals i.e., each of these selected bulls had two parents, with father being genotyped and had four grand parents, with Morello one among them. All the cows in the pedigree were assigned with missing genotypes and the cows and bulls belonging to the grand parent generation i.e.,generation of Morello were assumed as founders. A similar method was adopted for the bull Streif and for it 10 sons and 45 grandsons were utilised for the analysis.For both these groups, genome wide and chromosome wide IBD was estimated using the 48509 SNP markers.

All IBD probabilities were estimated by the `-ibd` and `-extended` functions of MERLIN(Abecasis et al., 2002). MERLIN can rapidly solve for phase ambiguity by taking into account the information from the surrounding markers. The algorithm generates accurate probabilities that the individuals in the pedigree share 0, 1 or 2 markers at any locus. According to Gagnon et al. (2005) such "multipoint" analysis however can buffer out a sizeable amount of standard deviation of IBD sharing, which is undesirable. So they suggested that only pairs of individuals without any ambiguity regarding the phase of the markers should be selected. As the pairs of individuals available were extremely limited all the grand sons available were used in the current study.

3.4 Homozygous Chromosomal Segments

If both the alleles of a particular marker are identical, we call that marker as homozygous. In some individuals long tracts of homozygous markers can occur in an uninterrupted fashion while in certain others it can be in a fractured way. As a part of this analysis we estimated the number of homozygous segments and the number of markers present with in these homozygous stretches with in the genome of bulls which are related to the important common ancestor. With the term "homozygous segments",instead of just looking into "runs of homozygosity" we applied three conditions for defining this term, viz. both the copies of the markers are identical in the bull under consideration, at least one of the allele should be descended from the important ancestor and the continuous appearance of these markers on the genome. By this we assume that we can get some information about the stretches of homozygous segments descended from the important ancestor.

3.4.1 Bulls Related to the Important Ancestor Just Once

For this analysis first we selected bulls which were related to the important ancestor viz. Morello, Streif and Redad only once in their pedigree. Then these

bulls were divided into groups according to the generation in which the important ancestor appear in their pedigree, for example, the Bulls which have Morello once in second generation, third, fourth and fifth generation. Like this for each important ancestor four families were selected. So in total 12 families for the three important bulls. For this analysis also we used 48509 informative SNPs distributed across the 29 chromosomes.

Within each family the analysis was done in five different ways. We counted the number of "homozygous segments" which are longer than 10, 20, 30, 50 and 100 markers. The number of markers present in each of these "homozygous blocks" were also counted and this helps to get an idea about the total number of markers in these homozygous segments.

3.4.2 Bulls Inbred to the Important Ancestor

In this analysis those bulls which were inbred to the important ancestors i. e., Morello, Streif and Redad were selected. Only those animals which had these ancestors just once either once in the paternal and maternal side of the pedigree were selected. These bulls were then regrouped into different generations, based on the appearance of important ancestor either in the paternal side or maternal side. For example the bulls belonging to generation three has the important ancestor in the third generation either in its sire side or in its dam side. This type of classification will help to compare the results with those animals which have the important ancestor just once. All the rest of the analysis was done as described for those bulls which has the presence of important ancestor just once in their pedigree.

3.5 Correlation Between Pedigree and Genomic Relationship

In this section of the analysis, we tried to estimate the correlation between the coefficient of coancestry between bulls estimated by means of pedigree information with that of the genomic relationship values like Identity by Descent, and

Identity by State etc. For this analysis all the 1764 bulls genotyped by ZuchtData were utilised. The pair wise relationship based upon pedigree for this 1764 bulls were made available by ZuchtData Corporation. The pedigree data used for this analysis comprised the complete pedigree for all bulls. In total there were 1,554,966 ($1764 \times 1763/2$) pair wise relationships.

Using the software package PLINK (version 1.06) (Purcell et al., 2007) we estimated the genome wide Identity by Descent (IBD) and Identity by State (IBS) for the entire group of 1764 bulls. All the 48509 informative SNP's distributed on the 29 autosomes were used during this process. Following genomic parameters were estimated

1. Probability of none of the alleles IBD (IBD0)
2. Probability of one allele being IBD (IBD1)
3. Probability of both alleles being IBD (IBD2)
4. Proportion IBD (calculated as P(IBD=2)+0.5×P(IBD=1))
5. Number of IBS 0 non missing loci
6. Number of IBS 1 non missing loci
7. Number of IBS 2 non missing loci

The command line functions `--no-pheno --no-fid --no-parents --map3 --genome --genome-full` were used for estimating the above mentioned parameters using PLINK. The analysis was under the assumption that all the bulls are unrelated and were considered as founders. Software Plink requires either genotypes of all the animals in the pedigree or none of them. As we have the genotypes of only bulls in the pedigree, we decided to run the program by omitting the pedigree information.

Chapter 4

Results

4.1 Overall Distribution of IBD Sharing

4.1.1 Morello Family

Using the software MERLIN we estimated for each locus the wide probability of either zero, one or two alleles being IBD with the corresponding markers of a particular grandsire. This is represented as IBD0, IBD1 and IBD2 respectively. From among them, between the grand sire and grandson, IBD1 is most frequent. The average of this value was taken as an estimate of the genome wide probability of at least one of the allele being IBD with the alleles of grandparent.
The overall genome wide distribution of probability of IBD1 between Morello and its 13 grandsons was lower than the random expectation of 0.5. Thirteen pair wise analysis were done and the average probability of IBD1 obtained was 0.4626 with a standard deviation of 0.0752 The animal wise details are provided in table 4.1.

The probability of one of the allele being IBD shows a wide range of variation, the maximum probability for IBD1 among these 13 bulls was 0.6195 for a bull numbered 040000241087145, with name *Moretti* and the minimum of 0.3332 was observed for bull number 040000603564644 *Modello*.

Table 4.1: Genome wide Probability of IBD1 for the thirteen grandsons of Morello

Sl No	Grandson	Prob. of IBD1
1	040000125234833	0.4250
2	040000134360133	0.5430
3	040000199437133	0.5056
4	040000208397442	0.4302
5	040000241087145	0.6195
6	040000433229634	0.4373
7	040000486388444	0.4479
8	040000548502646	0.5392
9	040000603564644	0.3332
10	040000627000444	0.3979
11	040000729422942	0.3948
12	276000913087947	0.4769
13	276000919290678	0.4629

4.1.2 Streif Family

45 grand sons of this bull were genotyped and they were used for the estimating the genome wide probability of at least one of the allele being IBD. The average value for the genome wide probability of IBD1 was 0.4783, higher than that observed for Morello family. The standard deviation of this probability was 0.0580 lower than that obtained for Morello group. The animal wise details are provided in Tables 4.2, 4.3 and 4.4.

The highest IBD1 probability obtained was 0.6099 for the bull named *Stix* and the least IBD was obtained for the bull *Seemann*. The IBD probability of this bull with Streif was only 0.3430.

Table 4.2: Genome wide probability of IBD1 for the 45 grandsons of Streif

Sl No	Grandson	Prob. of IBD1
1	040000090129533	0.4851
2	040000092069733	0.4690
3	040000111377233	0.5679
4	040000113086433	0.4990
5	040000118011433	0.4742
6	040000119507633	0.4272
7	040000127285533	0.5531
8	040000158470533	0.4875
9	040000163458126	0.4073
10	040000169538126	0.4138
11	040000169542626	0.4182
12	040000246188933	0.5299
13	040000247106964	0.4947
14	040000249947533	0.4595
15	040000259180833	0.5332

4.2 IBD Distribution per Chromosome

In line with genome wide probability of IBD1 calculation, we also derived the chromosome wide probability for at least one of the allele being IBD with the important grandsire. The mean and standard deviation of these probabilities for the grandsons of Morello and Streif with their grandsire for all the 29 autosomes are given in table 4.5 and 4.6. The data shows that in both the families, the standard deviation is showing an increasing trend towards chromosome with higher chromosome number.

Among the grandsons of Morello the highest probability for at least one of the allele being IBD was around 0.607 and this was noticed in chromosome 14 and the lowest figure of 0.2479 was recorded for chromosome 20. While going through the chromosome wide IBD sharing among the grandsons of Streif, we can see

Table 4.3: Genome wide probability of IBD1 for the 45 grandsons of Streif (ctd)

Sl No	Grandson	Prob. of IBD
16	040000264244533	0.4147
17	040000269705433	0.4712
18	040000286713933	0.4284
19	040000297581433	0.4723
20	040000304754833	0.4755
21	040000319957262	0.6099
22	040000340040933	0.5732
23	040000343678833	0.3972
24	040000383426733	0.5390
25	040000392076333	0.4797
26	040000497333741	0.4660
27	040000501073266	0.3430
28	040000501601444	0.4929
29	040000513879644	0.4667
30	040000533500944	0.4925
31	040000534866233	0.4846

that the highest IBD1 probability of 0.5789 was found in chromosome 26 and the lowest probability of 0.3636 was recorded in chromosome 17.

4.3 Homozygous Chromosomal Segments

4.3.1 Bulls Related to the Important Ancestor Just Once

Three important bulls viz. Morello, Streif and Redad were selected for this study. From among the genotyped animals, those bulls were selected which were related to the important ancestor just once in their pedigree. They were again re grouped based on which generation the important ancestor appeared. The detailed results are summarised in table 4.7.

Table 4.4: Genome wide probability of IBD1 for the 45 grandsons of Streif (ctd.)

Sl No	Grandson	Prob. of IBD
32	040000543637644	0.5251
33	040000545844833	0.4862
34	040000547801646	0.4793
35	040000553305746	0.3765
36	040000567256644	0.4118
37	040000576300457	0.5763
38	040000587980644	0.469
39	040000729226747	0.4409
40	040000742134633	0.4074
41	040000742379333	0.5699
42	040000763230733	0.4366
43	040000789488533	0.5267
44	040000811315233	0.5364
45	276000910953855	0.4554

The generation numbers are given from two to five, and the generation two includes the grandsons of the important ancestor, and generation three includes the great grandsons etc. Except for two groups, the number of bulls per group was above 30. Only 17 bulls were analysed from Morello family belonging to the fifth generation and only seven bulls from second generation of Redad family. The highest number of bulls (187) were observed from the fourth generation of Redad family.

4.3.1.1 Number of Homozygous Segments

For the bulls in each generation the number of homozygous segments or homozygous marker clusters which exceeds 10, 20, 30, 50 and 100 markers were determined. The details are given in table 4.8, 4.9, 4.10, 4.11 and 4.12.

On an average there were 477.14 clusters which contain more than 10 consecutive

Table 4.5: Chromosome wide distribution of IBD1 probability

	Grandsons of Morello		Grandsons of Streif	
Chromosome	Mean	Std. Dev.	Mean	Std. Dev.
1	0.5177	0.3238	0.4393	0.3150
2	0.5098	0.3197	0.4891	0.2542
3	0.5830	0.2858	0.4417	0.2680
4	0.3892	0.2138	0.5677	0.2625
5	0.4094	0.2039	0.4670	0.2940
6	0.5393	0.2779	0.4786	0.2766
7	0.5449	0.3209	0.4452	0.2283
8	0.4172	0.2208	0.4631	0.2287
9	0.4580	0.3910	0.5170	0.3613
10	0.5706	0.2762	0.4571	0.2755
11	0.4365	0.2523	0.4853	0.2443
12	0.5124	0.2990	0.4986	0.2772
13	0.3424	0.2646	0.4632	0.2215
14	0.6070	0.2743	0.5240	0.2872
15	0.3716	0.2970	0.5087	0.3184

homozygous markers for the bulls in the second generation of Morello and 395.59 clusters for bulls in the fifth generation. The highest number of homozygous clusters were found in bulls belonging to Streif family and the lowest for bulls in the Redad family. A similar trend was observed in all the different groups of clusters, i. e., for clusters longer than 20, 30, 50 and 100 markers. For clusters bigger than 10 markers the rate of reduction across the different generation was different in different families. The lowest rate of reduction was observed in Redad family and highest for Morello family. A similar trend was found in clusters bigger than 10 markers.

For clusters longer than 20 markers the numbers varied between 51 and 72.06 in the second generation whereas it varied between 45.93 and 53.63 in fifth generation. Regarding the clusters longer than 30 markers the numbers decreased to

Table 4.6: Chromosome wide distribution of IBD1 probability(ctd.)

	Grandsons of Morello		Grandsons of Streif	
Chromosome	Mean	Std. Dev.	Mean	Std. Dev.
16	0.5588	0.2472	0.5496	0.2653
17	0.4498	0.4340	0.3636	0.3599
18	0.2798	0.2707	0.5031	0.2950
19	0.3926	0.4534	0.5752	0.3870
20	0.2479	0.2983	0.4155	0.2837
21	0.3927	0.3983	0.4116	0.2890
22	0.3822	0.4329	0.5475	0.3717
23	0.4191	0.3376	0.5330	0.3145
24	0.5014	0.4081	0.3907	0.3868
25	0.5390	0.3826	0.4386	0.3512
26	0.5119	0.4452	0.5789	0.4030
27	0.5038	0.3955	0.5426	0.3773
28	0.3213	0.4002	0.4135	0.4226
29	0.4369	0.4273	0.3907	0.3947

Table 4.7: Total Number of Bulls analysed for Homozygous Segments

	Number of Bulls in		
Generation No.	Morello Family	Streif Family	Redad Family
2	63	47	7
3	174	144	79
4	73	149	187
5	17	54	30

16.00 to 27.30 in second generation to 12.67 to 19.02 in fifth generation. Again when the analysis was repeated by increasing the cluster size to more than 50, the numbers varied from 7.52 to 4.14 in the second generation to 2.64 to 5.17 in the fifth generation. Finally a comparison was done between clusters bigger than 100

Table 4.8: Average Number of Homozygous Clusters Longer than 10 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	477.14	61.30	479.08	54.40	443.86	11.82
3	426.47	43.25	440.10	40.75	399.32	44.78
4	407.46	38.65	412.09	33.85	389.83	56.78
5	395.59	51.94	400.22	46.21	394.70	23.07

Table 4.9: Average Number of Homozygous Clusters Longer than 20 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	72.06	16.42	78.47	17.22	51.00	5.94
3	57.75	12.07	65.55	13.43	43.91	10.40
4	53.24	10.53	56.51	11.81	43.97	11.10
5	49.70	14.63	53.63	12.59	45.93	9.36

Table 4.10: Average Number of Homozygous Clusters Longer than 30 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	27.30	9.63	32.98	10.53	16.00	4.16
3	19.02	5.37	25.52	7.59	12.63	4.15
4	16.56	4.61	20.04	6.32	12.57	4.88
5	14.70	5.72	19.02	6.56	12.67	4.77

markers. At this stage majority of the animals didn't had any clusters. This is evident from table 4.12, which shows a high standard deviation than the average values. Hence the average figure observed was very low, around one cluster per animal.

Table 4.11: Average Number of Homozygous Clusters Longer than 50 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	7.52	4.92	11.64	5.37	4.14	2.34
3	4.77	2.40	8.24	4.16	2.59	1.77
4	4.23	2.38	5.90	3.19	2.49	1.82
5	2.64	1.90	5.17	2.88	2.90	2.97

Table 4.12: Average Number of Homozygous Clusters Longer than 100 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	1.12	2.43	2.46	1.90	0.86	1.21
3	0.33	0.53	1.60	1.58	0.25	0.52
4	0.20	0.47	0.95	1.01	0.24	0.59
5	0.12	0.33	0.54	0.82	0.50	1.66

4.3.1.2 Number of Markers Within the Homozygous Clusters

Along with counting different homozygous clusters, number of markers included in these clusters were also determined. This was also done for bulls included in three major families and for different generation and for different cluster size. The results are summarised in tables 4.13, 4.14, 4.15, 4.16 and 4.17.

The total number of markers included in clusters longer than 10 markers ranged between 5709.35 to 7832.49 with in different generations. When expressed as percentage of total markers included in the analysis this ranges from 11.77 to 16.15%. As observed in the case of clusters, the markers also showed a similar trend of distribution. The highest number of markers were found in bulls belonging to Streif family and the lowest among the Redad family. Similarly in clusters greater than 20 markers also the highest numbers were in Streif group and lowest

in Redad group. Percentage wise distribution ranged between 5.8% and 2.8% of the total markers were found to be homozygous.

For cluster size above 30 markers, the highest figure was 1768.87 (3.6%) and the lowest was 589.10 (1.21%) respectively from Streif and Redad family.When th e cluster size was increased to 50, the numbers reduced drastically and the highest figure was just 982.89 and the lowest was 171.94 markers. On repeating the analysis with more than 100 markers per cluster, we found that very few animals carried such clusters, and majority of them didn't had any. This is evident from the table 4.17, where the standard deviation is higher than the mean values.

The number of bulls which carried clusters longer than 100 markers

Table 4.13: Number of SNPs Included in Clusters Longer than 10 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	7406.43	1192.23	7832.49	1152.85	6468.57	149.25
3	6354.64	743.70	6903.08	866.55	5681.79	715.49
4	5992.49	648.23	6257.36	675.64	5582.06	910.20
5	5709.35	904.34	5991.39	802.45	5714.93	530.89

Table 4.14: Number of SNPs Included in Clusters Longer than 20 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	2324.84	748.88	2827.55	799.41	1582.43	222.98
3	1739.78	386.63	2224.90	596.39	1262.82	308.29
4	1568.75	339.08	1818.50	452.25	1276.04	387.56
5	1402.06	436.74	1671.70	417.11	1354.67	431.32

Table 4.15: Number of SNPs Included in Clusters Longer than 30 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	1286.08	646.11	1768.87	668.43	781.71	245.77
3	846.97	260.21	1292.15	491.38	540.14	184.34
4	728.62	226.36	970.91	360.87	550.88	244.95
5	606.47	254.88	865.59	305.93	589.10	392.90

Table 4.16: Number of SNPs Included in Clusters Longer than 50 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	561.13	500.81	982.89	506.77	346.43	198.70
3	329.82	173.68	664.90	398.56	181.62	129.15
4	280.70	157.74	457.97	269.23	184.33	166.64
5	171.94	134.95	357.68	217.04	233.86	360.66

Table 4.17: Number of SNPs Included in Clusters Longer than 100 Markers

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	148.41	356.30	364.60	293.97	116.29	147.95
3	39.27	63.23	228.30	243.11	29.93	61.89
4	24.44	56.47	133.95	147.11	36.65	103.52
5	13.65	38.58	68.83	110.68	76.3	289.29

4.3.2 Bulls Inbred to the Important Ancestor

In this section we analysed the length of the homozygous segments and the number of markers included in it, for those bulls which are inbred to the three important ancestor. Only those bulls which have the presence of important ancestor

just once in paternal and maternal side of the pedigree were selected. The bulls were grouped according to the presence of the important ancestor on the paternal side, just similar to that done for bulls related to them once, so that we can compare the results. The number of bulls analysed in each group are given in table 4.18. As compared to the section 4.3.1, in certain groups there weren't enough bulls for performing the analysis. There were no bulls in two groups and in one group had just one bull. The highest number of bulls (78) were present in the fourth generation of Streif family.

Table 4.18: Number of Inbred Bulls analysed for Homozygous Segments

	Number of Bulls in		
Generation No.	Morello Family	Streif Family	Redad Family
2	6	1	0
3	24	24	14
4	21	78	71
5	0	41	35

As in the case of analysis of bulls related to the important ancestor just once, here also we looked for the length of the homozygous segments, its number and the number of markers included in each group.

4.3.2.1 Number of Homozygous Segments

For each generation, in evry group we looked into the number of homozygous segments, which exceeded 10, 20, 30, 50 and 100 markers per cluster. The details are given in tables 4.19, 4.20, 4.21, 4.22 and 4.23. As there weren't enough bulls, out of the 12 class, the analysis was performed in only nine groups.
For bulls which were inbred to morello the number of clusters longer than 10 homozygous markers ranged between 507.67 in second generation to 417.81 in the fourth generation. For the bulls inbred to Streif, the figures ranged between 421.58 and 381.71. For Redad the highest was 423.86 in the third generation and lowest 400.37 in the fifth generation. The analysis of clusters longer than 20 markers showed that the Streif group had the highest number across different

Table 4.19: Average Number of Homozygous Clusters Longer than 10 Markers in Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	507.67	46.39				
3	419.71	65.87	421.58	89.29	423.86	26.37
4	417.81	62.14	402.41	87.24	401.17	67.55
5			381.71	96.26	400.37	34.36

Table 4.20: Average Number of Homozygous Clusters Longer than 20 Markers in Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	102.67	35.44				
3	58.13	19.28	67.92	25.84	58.79	12.13
4	60.71	19.47	60.74	23.28	48.46	15.41
5			52.59	22.1	47.14	11.31

Table 4.21: Average Number of Homozygous Clusters Longer than 30 Markers in Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	45.50	23.15				
3	21.25	9.74	30.25	15.21	21.00	7.53
4	25.05	11.54	24.88	12.53	16.34	7.11
5			20.02	10.99	15.46	5.67

generations. The distribution with in Morello group was showing an abnormal pattern i. e., the bulls belonging to the fourth generation had a higher number of

Table 4.22: Average Number of Homozygous Clusters Longer than 50 Markers in Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	18.00	10.49				
3	6.58	4.73	11.96	8.8	7.64	5.47
4	8.86	5.85	9.35	6.57	5.54	4.53
5			6.39	4.88	4.51	3.13

Table 4.23: Average Number of Homozygous Clusters Longer than 100 Markers in Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	5.50	4.42				
3	1.33	1.52	3.20	2.9	2.78	3.62
4	1.76	1.89	2.30	2.53	1.59	2.14
5			1.46	2.32	1.06	1.14

segments than those belonging to third generation. A similar pattern was found in other cluster sizes (homozygous segments longer than 30, 50 and 100 markers) also.

For clusters bigger than 30 markers also, the Streif group had the highest numbers in the respective classes except for the fourth generation. FOr the other two groups (clusters bigger than 50 and 100 markers) also, the Redad family dominated in all the classes. In the last class, i. e., clusters greater than 100 markers (table number 4.23) we can see that the standard deviation observed was very high.

4.3.2.2 Number of Markers Within the Homozygous Clusters

As done with the bulls with important ancestor appearing once in the pedigree, for inbred bulls also, the number of markers included in the homozygous clusters were also determined. The results are summarised in tables 4.24, 4.25, 4.26, 4.27 and 4.28.

Table 4.24: Number of SNPs Included in Clusters Longer than 10 Markers for Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	9182.83	1883.89				
3	6473.88	1306.16	7057.75	2015.62	6758.43	1066.08
4	6621.71	1374.86	6492.00	1800.75	6055.24	1289.91
5			5898.63	1796.78	5935.80	739.86

Table 4.25: Number of SNPs Included in Clusters Longer than 20 Markers for Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	4078.16	1841.70				
3	1957.17	763.89	2634.46	1344.73	2196.79	958.58
4	2169.48	878.70	2232.81	1073.05	1653.96	744.89
5			1801.00	923.10	1531.11	489.54

Through out the Morello group, the number of markers in the fourth generation was higher than that of the third. But in other two families the results were in the expected lines. As seen with the number of homozygous clusters, here also in general, Streif Family had the highest average number of markers in all the class across different groups.

Table 4.26: Number of SNPs Included in Clusters Longer than 30 Markers for Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	2738.83	1618.64				
3	1104.75	589.23	1758.33	1136.59	1320.21	877.50
4	1349.57	732.51	1396.78	859.46	911.41	591.04
5			1045.54	714.69	800.86	388.30

Table 4.27: Number of SNPs Included in Clusters Longer than 50 Markers for Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	1731.66	1221.79				
3	565.29	440.05	1081.67	899.94	829.36	809.64
4	752.19	555.37	823.58	666.88	518.92	517.54
5			547.95	531.63	400.40	308.62

Table 4.28: Number of SNPs Included in Clusters Longer than 100 Markers for Inbred Bulls

	Morello Family		Streif Family		Redad Family	
Generation No.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
2	870.67	792.51				
3	210.29	251.84	497.68	521.84	486.21	689.51
4	271.10	296.02	357.43	431.80	253.17	379.65
5			220.00	378.40	167.60	196.27

4.4 Correlation Between Pedigree and Genomic Relationship

The pair wide pedigree based relationship, coefficient of coancestory, obtained from ZuchtData corporation was compared with the genomic parameters estimated using the software Plink. 1,554,996 pair wise comparisons were generated from 1764 bulls. The software calculated the probability of Identity by Descent of none of the alleles, either one of the allele and both the alleles. Also we were able to estimate the overall proportion of IBD between different bulls. Apart from IBD, for each pair we estimated the number of loci which were in Identity by State (IBS) either for none of the markers, or for one marker or for both markers. The values obtained were then correlated with the pair wise coefficient of coancestory figures.

Considering all bulls

The analysis was done for different groups of bulls, based upon their pedigree relationship. At first the entire bulls were used for the analysis. The results are summarised in table 4.29.

Table 4.29: Correlation Between Coefficient of Coancestry and Genomic Parameters for All Bulls

	IBD-0	IBD-1	IBD-2	Prpor.IBD	IBS-0	IBS-1	IBS-2
Coancestry	-0.572	0.553	0.356	0.590	-0.452	-0.317	0.518
IBD-0		-0.998	-0.123	-0.998	0.797	0.795	0.088
IBD-1			0.059	0.992	-0.793	-0.785	-0.109
IBD-2				0.186	-0.133	-0.229	0.330
Propor.IBD					-0.798	-0.802	-0.065
IBS-0						0.525	-0.200
IBS-1							0.376

Among the different traits, the pedigree relation has a strong positive correlation of 0.59 with the trait overall proportion of IBD. This was followed by the correlation coefficient of 0.553 with the trait IBD-1 and 0.518 with IBS-2. The lowest correlation was observed with the parameter IBS-1, where the value observed was only -0.317.

With closely related Bulls

In the second stage those pairs of bulls whose coefficient of Cooancestry values were greater than 0.01 were selected for the study. Their correlation coefficient with the pairwise genomic traits were estimated and the results are described in table 4.30.

Table 4.30: Correlation Between Coefficient of Coancestry and Genomic Parameters for Closely Related Bulls

	IBD-0	IBD-1	IBD-2	Prpor.IBD	IBS-0	IBS-1	IBS-2
Coancestry	-0.626	0.605	0.306	0.642	-0.562	-0.354	0.594
IBD-0		-0.997	-0.089	-0.997	0.942	0.793	-0.019
IBD-1			0.016	0.989	-0.937	-0.779	-0.007
IBD-2				0.160	-0.115	-0.245	0.367
Propor.IBD					-0.941	-0.804	0.045
IBS-0						0.756	-0.051
IBS-1							0.269

In this study the correlation between Coefficient of coancestry and overall IBD increased to 0.642 and that with IBS2 increased to 0.594. In general the correlation increased either in a positive or in a negative direction.

Chapter 5

Discussion

5.1 Probability of Identity by Descent Sharing

The Mendelian Principles of inheritance explained using pea plants layed a solid foundation for our present day understanding of gene transmission. Even though this theory ignores several genetic mechanisms like dominance, epistasis, interaction etc, involved in the determination of traits, it could accurately describe the process of gene transmission at the locus level.Normally the genes are transmitted randomly from parent to off spring with a fair chance of 50%. Hence we obtain a probability of one for IBD sharing of least one of the allele of the offspring at any locus. On applying the same rule we can infer that probability of IBD sharing at least one of the allele in this generation will be 0.5. The results obtained for both the families are less than this expected value. For the morello family the average genome wide probability obtained was 0.4646, and for Streif family it was 0.4783. In this study we genotyped only the males and the estimation was based on the genotype of bulls present in the pedigree. This might be the reason for obtaining an average figure less than expected. Apart from that the software MERLIN is sensitive to the number of genotyped animals present in the pedigree of the individual analysed. Further verification of these probabilities are required by incorporating the genotypes of the females belonging to the pedigree of the bull in question. Only by doing that we can get the exact probability of IBD sharing for each marker locus

With in each family the probability of IBD 1 figures were showing a wide range of variation, i. e., between 0.6195 and 0.3332 for Morello group and between 0.6099 and 0.3430 for Streif family.

5.2 Chromosome Wide Probability of IBD1

the chromosome wide probability for the alleles being IBD1 was calculated for all the 29 autosomes. The calculations were done separately for the grandsons of Morello and Streif. The results didn't show any trend of similarity between these two families. The only trend that we could follow in this analysis was that the standard deviations of these figures were showing an increasing trend as we move from chromosome one to 29. A similar trend was also observed by Gagnon et al. (2005), while describing about the Identity by Descent sharing among the autosomes of *Centre d'etudes du polymorphisme humain*(CEPH) siblings.

5.3 Homozygous Chromosomal Segments

An individual is said to be homozygous if both alleles at that locus are identical. In some individuals we can see such long tracts of homozygous markers. Homozygosity mapping aims to identify such stretches of markers and it is commonly done for mapping recessive traits in consanguineous families. In this study we analysed the homozygous segments found in each generation of bulls. As described in Materials and Methods, here we looked for homozygous segments across different generations. With the term "homozygous segments",instead of just looking into "runs of homozygosity" we applied three conditions for defining this term,

1. Both the copies of the markers are identical in the bull under consideration
2. At least one of the allele should be descended from the important ancestor
3. The continuous appearance of these markers on the genome

By this kind of analysis we assume that we can get some information about the stretches of homozygous segments descended from the important ancestor. In

the present study we try to provide a glimpse of the variation of the number of homozygous clusters and the markers involved in those clusters in bulls which are related to one important ancestor just once and those which are inbred to that important ancestor. The general assumption is that the length of these segments and the distribution pattern will be different in animals with varying level of inbreeding coefficient.

5.3.1 Number of Homozygous Cluster

Here we are comparing the number of homozygous clusters found two groups of bulls, first group, those which are related to the important ancestor just once and the second which are inbred to the important ancestor. these two groups were further classified based on the generation in which the important ancestor appear. The results are summarised in tables 5.1, 5.2, 5.3, 5.4 and 5.5.

Table 5.1: Number of Homozygous Clusters Longer than 10 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	477.14	507.67	479.08		443.86	
3	426.47	419.71	440.10	421.58	399.32	423.86
4	407.46	417.81	412.09	402.41	389.83	401.17
5	395.59		400.22	381.70	394.70	400.37

From the tables 5.1, 5.2, 5.3, 5.4 and 5.5 we can see that except in certain cases, the number of clusters in inbred bulls are higher than non inbred bulls. The deviations from this is mainly found in smaller cluster sizes (cluster with 10 or more markers) and mainly in Streif group. This can be considered to be due to some sampling error. But it can be seen that as the cluster size increases the difference between the inbred and non inbred becomes more prominent. This trend is evident from tables 5.4 and 5.5, where we can see a significant difference between the inbred and non inbred bulls.

Table 5.2: Number of Homozygous Clusters Longer than 20 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	72.06	102.67	78.47		51.00	
3	57.75	58.13	65.55	67.92	43.91	58.79
4	53.24	60.71	56.51	60.74	43.97	48.46
5	49.70		53.63	52.59	45.93	47.14

Table 5.3: Number of Homozygous Clusters Longer than 30 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	27.30	45.50	32.98		16.00	
3	19.02	21.25	25.52	30.25	12.63	21.00
4	16.56	25.05	20.04	24.88	12.57	16.34
5	14.70		19.02	20.02	12.67	15.46

Table 5.4: Number of Homozygous Clusters Longer than 50 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	7.52	18.00	11.64		4.14	
3	4.77	6.58	8.24	11.96	2.59	7.64
4	4.23	8.86	5.90	9.35	2.49	5.54
5	2.64		5.17	6.39	2.90	4.51

Wang et al. (2009) stated that length of homozygous segments depends on the degree of parental consanguinity and so homozygous segments with short length can be found in outbred populations. The results from the current study shows that larger numbers of longer homozygous clusters can be found in inbred bulls

Table 5.5: Number of Homozygous Clusters Longer than 100 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	1.12	5.50	2.46		0.86	
3	0.33	1.33	1.60	3.20	0.25	2.78
4	0.20	1.76	0.95	2.30	0.24	1.59
5	0.12		0.54	1.46	0.50	1.06

when compared to non inbred ones.

5.3.2 Markers With in the Homozygous Cluster

A similar analysis was done for the markers included in these homozygous clusters. A comparison of inbred and non inbred bulls is given in tables 5.6, 5.7, 5.8, 5.9 and 5.10.

Table 5.6: Number of SNPs within Clusters Longer than 10 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	7406.43	9182.83	7832.49		6468.57	
3	6354.64	6473.88	6903.08	7057.75	5681.79	6758.43
4	5992.49	6621.71	6257.36	6492.00	5582.06	6055.24
5	5709.35		5991.39	5898.63	5714.93	5935.80

Unlike observed from the number of clusters, in this case in all the generation and in all the classes, the number of SNPs included in the cluster were higher for inbred bulls. The difference between these figures become more prominent as we look into big clusters. These are evident from tables 5.9 and 5.10, where the number of markers included in these clusters differ significantly. This result

Table 5.7: Number of SNPs within Clusters Longer than 20 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	2324.84	4078.16	2827.55		1582.43	
3	1739.78	1957.17	2224.90	2634.46	1262.82	2196.79
4	1568.75	2169.48	1818.50	2232.81	1276.04	1653.96
5	1402.06		1671.70	1801.00	1354.67	1531.11

Table 5.8: Number of SNPs within Clusters Longer than 30 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	1286.08	2738.83	1768.87		781.71	
3	846.97	1104.75	1292.15	1758.33	540.14	1320.21
4	728.62	1349.57	970.91	1396.78	550.88	911.41
5	606.47		865.59	1045.54	589.10	800.86

Table 5.9: Number of SNPs within Clusters Longer than 50 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	561.13	1731.66	982.89		346.43	
3	329.82	565.29	664.90	1081.67	181.62	829.36
4	280.70	752.19	457.97	823.58	184.33	518.92
5	171.94		357.68	547.95	233.86	400.40

is in agreement with the observation made by Wang et al. (2009), where in he concluded that an abundance of homozygous segments may significantly reduce the ability to fine map disease genes using association studies.

Table 5.10: Number of SNPs within Clusters Longer than 100 Markers

	Morello Family		Streif Family		Redad Family	
Gen. No.	Non Inbred	Inbred	Non Inbred	Inbred	Non Inbred	Inbred
2	148.41	870.67	364.60		116.29	
3	39.27	210.29	228.30	497.68	29.93	486.21
4	24.44	271.10	133.95	357.43	36.65	253.17
5	13.65		68.83	220.00	76.3	167.60

5.4 Correlation Between Coefficient of Coancestory and Genomic Relationship

In this section of the analysis we estimated the correlation coefficient between the Coefficient of Coancestory and Genomic parameters like Identity by Descent, Identity by State etc. Following parameters were estimated from the genomic data.

1. Probability of none of the alleles IBD (IBD0)
2. Probability of one allele being IBD (IBD1)
3. Probability of both alleles being IBD (IBD2)
4. Proportion IBD (calculated as P(IBD=2)+0.5×P(IBD=1))
5. Number of IBS 0 non missing loci
6. Number of IBS 1 non missing loci
7. Number of IBS 2 non missing loci

In the beginning we included all the bulls genotyped for the analysis. The results showed that the coefficient of coancestry has the highest level of correlation i. e.,0.59 with the parameter Probability of IBD. The software PLINK calculates the Probability of IBD by using the formula $0.5 \times IBD1 + IBD2$. So this parameter takes both probability of IBD1 and probability of IBD2 into consideration. The

results shows that the pedigree relationship has medium level of correlation with one of the most important genomic parameters, the over all proportion of IBD. The second most correlated parameter was the number of IBS2 non missing loci. The correlation coefficient was 0.518.when the same study was repeated for more related bulls, i.e., those bulls with coefficient of coancestry greater than 0.01 the correlation with overall IBD increased to 0.642 and IBS2 increased to 0.594.

Chapter 6

Conclusions

Identity by descent estimation

The study shows that the genome wide and chromosome wide probability for at least one of the allele to be IBD are close to the expected values. The values we calculated for IBD using the program MERLIN may not be exact, as only the bulls involved in the pedigree was genotyped. Also the number of bulls (grandson level) analyzed in this study was very less, 13 in one family and 45 in the second) and only the pedigree of the paternal side was analyzed in this study. The presence of these important bulls on the maternal side was not taken into consideration Hence further studies are required by incorporating the female genotypes for the exact calculations if chromosome wide and genome wide IBD.

Homozygous Segments

The study regarding the length, number and size of the homozygous segments clearly shows that these factors are influenced by inbreeding and the effects are more prominent in longer clusters (above 50 and 100 Markers) Further studies in this field are required for estimating the average length and position of these clusters. Generally these type of analysis are more suitable for Case Control type of study, especially for studying recessive genetic disorders.

Correlation between Coefficient of Coancestory and Breeding Values

Regarding this part of the study, we observed only a medium level of correlation (0.59) between the pedigree relationship and overall IBD probability. From among remaining the parameters IBS figures were showing more correlation than the IBD of 0, 1 or 2 alleles. The correlation observed with the number of IBS 2 non missing loci IBS2 was around 0.518

Chapter 7

Summary

The current study aimed at estimating the genomic contribution of some important bulls to a group of young bulls in the population of Simmental breed, used in Austria and Germany. For estimating this we utilised one of the most important development in the field of genomics, i. e., high density Single Nucleotide Polymorphism (SNP) chips. The bulls genotyped were a part of the genomic selection project running in Austria and Germany. The genome wide and chromosome wide probability of IBD of one of the allele estimated between the grandparents and grandsons didn't deviate much from the random expectation. The second analysis was regarding the length, number and size of homozygous chromosomal segments in two groups of bulls, i.e., inbred and non inbred bulls. The results clearly shows that inbreeding has a significant effect on these parameters, especially for those segments longer than 50 markers. Finally we studied the correlation between the genomic and pedigree relationship for all the genotyped bulls in this project. The results concluded that there exists a medium level of correlation between genomic traits like IBD and pedigree relationship.

Chapter 8

Zusammenfassung

Die aktuelle Studie zur Schtzung der genomischen Beitrag einige wichtige Bullen zu einer Gruppe von jungen Stieren in der Bevlkerung von Fleckvieh, die in sterreich und Deutschland. Fr die Abschtzung dieser genutzt werden wir eine der wichtigsten Entwicklungen im Bereich der Genomik, I. E., High-Density Single Nucleotide Polymorphism (SNP)-Chips. Die Bullen genotypisierter wurden ein Teil der genomischen Auswahl Projekt luft in sterreich und Deutschland. Die Genom-weite und breite Chromosom Wahrscheinlichkeit von IBD eines der Allel schtzungsweise zwischen den Groeltern und Enkel nicht wesentlich abweichen von den zuflligen Erwartung. Die zweite Analyse wurde in Bezug auf die Lnge, Anzahl und Gre der homozygot chromosomaler Segmente in zwei Gruppen von Bullen, dh inbred und nicht inbred Bullen. Das Ergebnis zeigt deutlich, dass Inzucht hat einen erheblichen Einfluss auf diese Parameter, insbesondere fr die Segmente mehr als 50 Marken. Schlielich haben wir die Korrelation zwischen der genomischen und Stammbaum Beziehung fr alle genotypisierter Bullen in diesem Projekt. Die Ergebnisse der Schluss gezogen, dass es eine mittlere Ebene der Zusammenhang zwischen der genomischen Eigenschaften wie IBD und Stammbaum Beziehung.

References

Abecasis, G. R., Cherny, S. S., Cookson, W. O. and Cardon, L. R. (2002), 'Merlin-rapid analysis of dense genetic maps using sparse gene flow trees', *Nat. Genet.* **30**, 97–101. 5, 11

Abecasis, G. R. and Wigginton, J. E. (2005), 'Handling marker-marker linkage disequilibrium: Pedigree analysis with clustered markers', *Am. J. Hum. Genet.* **77**(5), 754 – 767.
URL: *http://www.sciencedirect.com/science/article/B8JDD-4RDGV69-7/2/acf2e0cf75ff078d164f643ff248ee28* 6

Abney, M. (2008), 'Identity-by-Descent Estimation and Mapping of Qualitative Traits in Large, Complex Pedigrees', *Genetics* **179**(3), 1577–1590. 5

Baumung, R. and Solkner, J. (2003), 'Pedigree and marker information requirements to monitor genetic variability', *Genet. Sel. Evol.* **35**(4), 369–383.
URL: *http://dx.doi.org/10.1051/gse:2003029* 7

Bijma, P. and Woolliams, J. A. (1999), 'Prediction of Genetic Contributions and Generation Intervals in Populations With Overlapping Generations Under Selection', *Genetics* **151**(3), 1197–1210. 7

Broman, K. W. and Weber, J. L. (1999), 'Long homozygous chromosomal segments in reference families from the centre d'etude du polymorphisme humain', *Am. J. Hum. Genet.* **65**(6), 1493–1500. 8

Cannings, C. (2003), 'The identity by descent process along the chromosome', *Hum. Hered.* **56**(1-3), 126–130. 3

Charlier, C., Farnir, F., Berzi, P., Vanmanshoven, P., Brouwers, B., Vromans, H. and Georges, M. (1996), 'Identity-by-descent mapping of recessive traits in livestock: Application to map the bovine syndactyly locus to chromosome 15', *Genome Res.* **6**, 580–589. 4

Clark, A. G. (1999), 'The size distribution of homozygous segments in the human genome', *Am. J. Hum. Genet.* **65**(6), 1489 – 1492.
URL: *http://www.sciencedirect.com/science/article/B8JDD-4RDPT53-3/2/e6f1acea3f09f942b58aff8454ec729c* 8

Druet, T., Fritz, S., Boussaha, M., Ben-Jemaa, S., Guillaume, F., Derbala, D., Zelenika, D., Lechner, D., Charon, C., Boichard, D., Gut, I. G., Eggen, A. and Gautier, M. (2008), 'Fine Mapping of Quantitative Trait Loci Affecting Female Fertility in Dairy Cattle on BTA03 Using a Dense Single-Nucleotide Polymorphism Map', *Genetics* **178**(4), 2227–2235. 4

Elston, R. C. and Stewart, J. (1971), 'A general model for the genetic analysis of pedigree data', *Hum. Hered.* **21**, 523 –542. 5

Fisher, R. A. (1954), 'A fuller theory of junctions in inbreeding', *Heridity* **8**, 187–197. 6

Gagnon, A., Beise, J. and Vaupel, J. W. (2005), 'Genome-wide identity-by-descent sharing among ceph siblings', *Genet. Epidemiol.* **29**(3), 215–224. 3, 7, 11, 33

Gibson, J., Morton, N. E. and Collins, A. (2006), 'Extended tracts of homozygosity in outbred human populations', *Hum. Mol. Genet.* **15**(5), 789–795. 8

Grapes, L., Firat, M. Z., Dekkers, J. C. M., Rothschild, M. F. and Fernando, R. L. (2006), 'Optimal Haplotype Structure for Linkage Disequilibrium-Based Fine Mapping of Quantitative Trait Loci Using Identity by Descent', *Genetics* **172**(3), 1955–1965. 4

Heath, S. C. (1997), 'Markov chain monte carlo segregation and linkage analysis for oligogenic models', *Am. J. Hum. Genet.* **61**, 748–760. 5

Keith, J. M., McRae, A., Duffy, D., Mengersen, K. and Visscher, P. M. (2008), 'Calculation of ibd probabilities with dense snp or sequence data', *genet. Epidemiol.* **32**(6), 513–519. 6

Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. and S., L. E. (1996), 'Parametric and nonparametric linkage analysis: a unified multipoint approach', *Am. J. Hum. Genet.* **58**(6), 1347–1363. 5

Lacy, R. C. (1989), 'Analysis of founder representation in pedigrees: Founder equivalents and founder genome equivalents', *Zoo Biology* **8**(2), 111–123. 6

Malecot, G. (1941), 'Etude mathematique des populations mendeliennes (mathematical studies of mendalian populations)', *Science* **A-4**, 45–60. 3

Meuwissen, T. H. E. and Goddard, M. E. (2000), 'Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibria With Closely Linked Marker Loci', *Genetics* **155**(1), 421–430. 4

Meuwissen, T. H. E. and Goddard, M. E. (2001), 'Prediction of identity by descent probabilities from marker- haplotypes', *Genet. Sel. Evol.* **33**(6), 605–634. **URL:** *http://dx.doi.org/10.1051/gse:2001134* 5

Motro, U. and Thomson, G. (1985), 'The Affected Sib Method, 1. Stastical Features of teh Affected Sib Pair Method', *Genetics* **110**(3), 525–538. 3

Nagamine, Y., Knott, S. A., Visscher, P. M. and Haley, C. S. (2002), 'Simple deterministic identity-by-descent coefficients and estimation of qtl allelic effects in full and half sibs', *Genet. Res.* **80**(03), 237–243. 3

O'Connell, J. R. and Weeks, D. E. (1995), 'The vitesse algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance', *Nat Genet* **11**, 402–408. 5

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J. and Sham, P. C. (2007), 'Plink: A tool set for whole-genome association and population-based linkage analyses', *Am. J. Hum. Genet.* **81**(3), 559 – 575.

URL: *http://www.sciencedirect.com/science/article/B8JDD-4R29JC1-F/2/8ec497a5efa03cb1cb851829ae0841b9* 13

Riquet, J., Coppieters, W., Cambisano, N., Arranz, J.-J., Berzi, P., Davis, S. K., Grisart, B., Farnir, F., Karim, L., Mni, M., Simon, P., Taylor, J. F., Vanmanshoven, P., Wagenaar, D., Womack, J. E. and Georges, M. (1999), 'Fine-mapping of quantitative trait loci by identity by descent in outbred populations: Application to milk production in dairy cattle', *Proc. Natl. Acad. Sci. USA* **96**(16), 9252–9257. 4

Roughsedge, T., Brotherstone, S. and Visscher, P. M. (1999), 'Quantifying genetic contributions to a dairy cattle population using pedigree analysis', *Livest. Prod. Sci.* **60**(2-3), 359 – 369.
URL: *http://www.sciencedirect.com/science/article/B6T9B-487F5VS-X/2/624fb8c59bcdf5bcdd7227e56ec42e55* 7

Royo, L., lvarez, I., Gutirrez, J., Fernndez, I. and Goyache, F. (2007), 'Genetic variability in the endangered asturcn pony assessed using genealogical and molecular information', *Livestock Science* **107**(2-3), 162 – 169.
URL: *http://www.sciencedirect.com/science/article/B7XNX-4M4TW0K-1/2/1930b008081c6fba0f1ad48cc4f795b6* 7

Sobel, E., Sengul, H. and Weeks, D. (2001), 'Multipoint estimation of identity-by-descent probabilities at arbitrary positions among marker loci on general pedigrees', *Hum. Hered.* **52**, 121–131. 5

Tarres, J., Guillaume, F. and Fritz, S. (2009), 'A strategy for qtl fine-mapping using a dense snp map', *BMC Proceedings* **3**(Suppl 1), S3.
URL: *http://www.biomedcentral.com/1753-6561/3/S1/S3* 4

Wang, S., Haynes, C., Barany, F. and Ott, J. (2009), 'Genome-wide autozygosity mapping in human populations.', *Genetic epidemiology* **33**(2), 172–180.
URL: *http://dx.doi.org/10.1002/gepi.20344* 35, 37

Williams, C. G. and Reyes-Valds, M. H. (2007), 'Estimating a founder's genomic proportion for each descendant in an outbred pedigree', *Genome* **50**, 289–296. 7

Wray, N. R., Woolliams, J. A. and Thompson, R. (1990), 'Methods for predicting rates of inbreeding in selected populations', *Theor. Appl. Genet.* **80**, 503–512. 7