Molecular dissection of inbreeding depression for semen quality traits in cattle

Ferenčaković, Maja

University of Zagreb
FACULTY OF AGRICULTURE

Maja Ferenčaković

Molecular dissection of inbreeding depression for semen quality traits in cattle

DOCTORAL THESIS

Zagreb, 2015

Sveučilište u Zagrebu
AGRONOMSKI FAKULTET

Maja Ferenčaković

Molekularna disekcija inbriding depresije za svojstva kvalitete sperme kod goveda

DOKTORSKI RAD

Zagreb, 2015.

University of Zagreb
FACULTY OF AGRICULTURE

Maja Ferenčaković

Molecular dissection of inbreeding depression for semen quality traits in cattle

DOCTORAL THESIS

Supervisors: Professor Ino Čurik, PhD
Professor Johann Sölkner, PhD

Zagreb, 2015

Sveučilište u Zagrebu
AGRONOMSKI FAKULTET

Maja Ferenčaković

Molekularna disekcija inbriding depresije za svojstva kvalitete sperme kod goveda

DOKTORSKI RAD

Mentori: prof. dr. sc. Ino Čurik
Univ. Prof. DI. Dr. Johann Sölkner

Zagreb, 2015.

Supervisors:
prof. dr. sc. Ino Čurik
Department of Animal Science
Faculty of Agriculture, University of Zagreb
Svetošimunska cesta 25, 10000 Zagreb, Croatia
icurik@agr.hr

Univ. Prof. DI. Dr. Johann Sölkner
University of Natural Resources and Life Sciences, Vienna, Austria
johann.soelkner@boku.ac.at

This doctoral thesis was evaluated by the Dissertation Evaluation Committee members:

1. Miroslav Kapš, PhD

Full professor, Faculty of Agriculture, University of Zagreb, Croatia

2. Johannes Arjen Lenstra, PhD

Associate professor, Utrecht University, Institute for Risk Assessment Sciences, The Netherlands

3. Vlatka Čubrić Čurik, PhD

Assistant professor, Faculty of Agriculture, University of Zagreb, Croatia

Doctoral thesis was defended at Faculty of Agriculture, University of Zagreb on _____ 2015 in front of the Dissertation Defense Committee members:

1. Miroslav Kapš, PhD, _____

Full professor, Faculty of Agriculture, University of Zagreb, Croatia,

2. Johannes Arjen Lenstra, PhD, _____

Associate professor, Utrecht University, Institute for Risk Assessment Sciences, The Netherlands,

3. Vlatka Čubrić Čurik, PhD, _____

Assistant professor, Faculty of Agriculture, University of Zagreb, Croatia.

Acknowledgements

First of all, I would like to thank my mentor prof. dr. sc. Ino Čurik who guided me through the process of becoming a scientist in every sense of the word. I am grateful for your leadership, support, selflessly sharing knowledge, motivation and encouraging me to simply follow where the data lead. You helped with my personal and academic development. I cannot think of a better mentor to have. Truly, there are no words that could express my deep gratitude to you.

Many thanks to Univ. Prof. DI. Dr. Johann Sölkner for patience, encouragement and all the time dedicated to guide me in research. Thank you for giving me opportunity to work with you and for all your advices and encouragements in tough times.

I would like to express my deepest gratitude to doc. dr. sc. Vlatka Čubrić Čurik for support, caring for my future and for constantly reminding me that I have to finish my Ph.D. Thank you for being a good friend and excellent motivator.

Big thanks to prof. dr. sc. Miroslav Kapš for honest and unselfish support, sharing knowledge about modeling and all the help with solving statistical issues encountered during writhing this Thesis. You have provided invaluable perspectives on my research when most needed

I take this opportunity to express my gratitude to Dr. Johannes Lenstra for the useful comments and engagement through the process of writing of this thesis. Thank you for always finding time to propose consistently excellent improvements.

For funding I would like to acknowledge the Croatian Ministry of Science, Education and Sports (project 178-1780460-0546, 'Strategies for elimination of genetic defects in selected populations'), by a bilateral collaborative project involving the Croatian Ministry of Science, Education and Sports and the Austrian Agency for International Cooperation in Education and Research (OeADGmbH) ("Analysis of genetic diversity in cattle populations using dense sets of molecular markers"),"Ernst Mach grant – worldwide" financed by the Austrian Federal Ministry of Education and Science, Erasmus Mundus programme 'European Master in animal Breeding and Genetics', financed by the European Commission and European Science Foundation networking

program “Advances in Farm Animal Genomic Resources” (“GENOMIC-RESOURCES”) grant (EX/3733) for which I am deeply thankful.

I am very grateful to Zuchtdata EDV-Dienstleistungen GmbH and Geno SA, as well as Austrian Simmental association and Braunvieh Austria, Association of the Tyrolian Grey Cattle Breeders, Austrian association of the Pinzgauer cattle breeders and Federation of Austrian Cattle Breeders for provision of genotype and pedigree data.

I would like to thank Dipl. Ing. Peter Stückler from Besamungsanstalt Gleisdorf, Dr. Josef Miesenberger from Besamungsanstalt Hohenzell and Dr. Friedrich Führer from NÖ-Genetik Wieselburg for provision of semen quality records.

I would specially like to thank Golden Helix team for providing me free license for SNP & Variation Suite v7.6.8 Win64, which allowed me easy, efficient and headache-free work with genotype data. Thank you for making process of data preparation enjoyable experience.

Many thanks to my friend Miroslav Horvacki who selflessly helped me to solve all those mysteries of installing and working with various strange software and platforms. This thesis would not have been possible to write without your help.

A special mention for Dr. Félix Goyache; “Little by little…” was something I often heard from you and I continued practicing this philosophy permanently in my life. Thank you for lively discussions, encouraging emails, and sharing relevant publications.

I would like to thank my colleagues from Department of Animal Science for support and friendship and being there when I needed help.

To all my friends and dear people who patiently helped me to stay focused on my goal when I was exhausted and crazy and who kept reminding me there is a whole world outside of my PhD. Also, thank you for all your support, advice and belief that I was capable of taking on the PhD.

To my brother and his family who always jumped in and helped every time there were things to do at home when I was abroad. Knowing that you are there brought me peace

of mind and ability to easy concentrate to my work instead of worrying about all those domestic stuff, which are impossible to do if not present.

Finally I would like to thank my mother who has supported me throughout the entire process and who was silently proud at all my achievements. I will be grateful forever.

Thank you

Preface

This PhD thesis is based on four articles (Chapters 1, 3, 4 and 5), which are published in international journals, as well as study prepared for publication in which identification of genomic regions associated with inbreeding depression on semen quality traits is investigated (Chapter 5). A detailed literature review is presented in Chapter 1 in form of review article published in "Livestock Science" by Ino Čurik, Maja Ferenčaković and Johann Sölkner.

Chapter 2 gives the hypotheses stated in this thesis and the resulting objectives.

Chapter 3 was published in Journal of Animal Breeding & Genetics by Maja Ferenčaković, Edin Hamzić, Birgit Gredler, Trygve Solberg, Gunnar Klemetsdal, Ino Čurik and Johann Sölkner. This work is one of the first analyses of Runs of homozygosity (ROH) in cattle. It is important to notice that before this analysis similar studies were performed only in humans with exception of an EAAP conference paper in 2010 (Sölkner *et al*. 2010) and one paper published 2011, both by the same group of authors (article is in the supplementary material). This pioneer research was published in Agriculturae Conspectus Scientificus and this chapter is extension and confirmation of this research on larger sample and in more cattle populations.

Chapter 4 was published in "Genetics Selection Evolution" by Maja Ferenčaković, Johann Sölkner and Ino Čurik. This article represents first comprehensive analysis of the issues that can arise in process of ROH determination and F_{ROH}.

Chapter 5 present material and methods and results and discussion of unpublished part of this Thesis. This part shows how ROH approach can be used for estimating inbreeding depression. It clearly shows that simple attempt of finding regions that are responsible for changes in semen quality traits when they are autozygous, can yield in finding genes that are proven to affect fertility, as well as some whose function highly suggests important role for fertility traits.

In Chapters 6 and 7 general discussion and conclusions are presented

Copies of publications which I co-authored and which are directly associated with this Thesis are attached as Supplementary material

A hundred times every day I remind myself that my inner and outer life depend on the labors of other men, living and dead, and that I must exert myself in order to give in the same measure as I have received and am still receiving.

Albert Einstein, (1879 - 1955)

ABSTRACT

Runs Of Homozygosity (ROH) are a newly introduced approach for identifying inbreeding in diploid individuals. This approach is more reliable and available than pedigree data, but the lack of universal standards about ROH definition and identification introduces serious bias in ROH studies. In this thesis, ROH were analyzed in five cattle breeds (Brown Swiss, Fleckvieh, Norwegian Red, Pinzgauer and Tyrol Grey). The effects of SNP chip density and genotyping errors were tested on Brown Swiss, Pinzgauer and Tyrol Grey in order to establish the most optimal settings for precise estimation of levels of autozygosity. Data from the 50 k chip led to an overestimation of the number of ROH shorter than 4 Mb, since the analysis could not identify heterozygous SNPs present on the denser chip. Conversely, data from the denser chip underestimated the number of ROH longer than 8 Mb, unless the presence of a small number of heterozygous SNP genotypes was allowed within a ROH. Using this options ROH were identified in Brown Swiss, Fleckvieh, Norwegian Red and Tyrol Grey. Levels of autozygosity were calculated and compared with pedigree inbreeding coefficients. For all four breeds, population inbreeding levels estimated by the genomic inbreeding coefficients $F_{ROH > 8 Mb}$ and $F_{ROH > 16 Mb}$ were similar to the levels estimated from pedigrees. In contrast, inbreeding estimates based on $F_{ROH > 1 Mb}$ and $F_{ROH > 2 Mb}$ were considerably higher than pedigree-derived estimates. Pearson correlations between F_{ROH} and F_{PED} ranged from 0.50 to 0.72, as dependent on pedigree depth. In the analysis of inbreeding depression a significant influence of F_{PED}, $F_{ROH2 - 4 Mb}$ and $F_{ROH > 2 Mb}$ on total number of spermatozoa in 554 Fleckvieh bulls was found. Exact autozygous regions that influence this trait were detected on chromosomes 7, 10, 17, 20, 22 and 27 containing 41 genes. Five obvious candidate genes were found which are known to be directly associated with spermatogenesis, energy levels in spermatozoa and osmotic balance of the sperm. In conclusion, genotyping errors and SNP chip density do affect estimates of autozygosity from ROH, ROH distributions (number and size) enables precise estimation of autozygosity at individual and population levels in cattle and genomic autozygosity does have influence on bull semen quality.

Key words: inbreeding, inbreeding depression, SNP chip data, cattle, bull semen quality

SAŽETAK

"Runs Of Homozygosity" (ROH) nov su pristup utvrđivanja inbridinga kod diploidnih organizama i smatraju se pouzdanijim i više dostupnim od rodovnika, no nedostaju jedinstveni standardi za njihovu uporabu. U ovoj disertaciji ROH su analizirani kod pet pasmina goveda (Brown Swiss, Fleckvieh, Norwegian Red, Pinzgauer i Tyrol Grey). Utjecaj gustoće genomske informacije (SNP chipa) i utjecaj genotipskih pogrešaka na detekciju ROH-ova istražen je na pasminama Brown Swiss, Pinzgauer i Tyrol Grey. SNP chip manje gustoće sustavno je precjenjivao broj ROH-ova <4 Mb, što je uzrokovalo i precjenjivanje inbridinga. Gušći SNP chip podcjenjivao je velike segmente osim u slučaju kada se dozvoljavao određen broj heterozigotnih genotipova. Koristeći ova saznanja procijenjen je ROH inbriding (F_{ROH}) za populacije Brown Swiss, Fleckvieh, Norwegian Red i Tyrol Grey, te je uspoređen s koeficijentima inbridinga iz rodovnika (F_{PED}). Za sve četiri populacije vrijednosti $F_{ROH > 8 Mb}$ i $F_{ROH > 16 Mb}$ bile su slične vrijednosti F_{PED} dok su $F_{ROH > 1 Mb}$ i $F_{ROH > 2 Mb}$ vrijednosti inbridinga bile mnogo veće od F_{PED}. Pearsonovi korelacijski koeficijenti između F_{ROH} i F_{PED} bili su rasponu od 0.50 do 0.72, a vrijednost je ovisila o dubini rodovnika. Kod analize inbriding depresije uočen je značajan utjecaj F_{PED}, $F_{ROH2 - 4 Mb}$ i $F_{ROH > 2 Mb}$ na ukupan broj spermatozoida kod bikova Fleckvieh pasmine. Detekcija autozigotnih regija koje utječu na ovo svojstvo otkrila je regije na kromosomima 7, 10, 17, 20, 22 i 27 u kojima je pronađen 41 gen od kojih je pet izglednih kandidata, jer su povezani sa spermatogenezom, razinom energije spermatozoida i osmotskom ravnotežom spermatozoida i sjemene tekućine. Sve navedeno dovodi sljedećih zaključaka; greške genotipizacije i gustoća SNP chipa imaju utjecaj na identifikaciju ROH-ova; ROH metodom procijenjena autozigotnost predstavlja dobar pokazatelj stupnja inbridinga; autozigotnost genoma je povezana s plodnošću bikova.

Ključne riječi: inbriding, inbriding depresija, SNP chip, goveda, kvaliteta sjemena bikova

TABLE OF CONTENTS

GENERAL INTRODUCTION

Inbreeding depression (harmful effect of inbreeding) is a well known phenomenon recognized as a problem very early in human history. It is characterized in reduced fitness of progenies of related individuals. Harmful effects of close inbreeding were widely recognized well before any formal scientific investigation. Indeed, in humans about 42% of offspring from sister-brother marriages die before they reach reproductive age and most cultures have strong traditions with respect to avoidance of incest. In spite of its importance the genetic basis of inbreeding depression is still unclear and inbreeding depression is still often estimated simply by regression of a specific trait on the inbreeding coefficient calculated from pedigree (F_{PED}). However, F_{PED} has several disadvantages. First, F_{PED} fails to capture the influence of relatedness among founders from the base population. Second, F_{PED} is equal to the expected proportion of the genome that is IBD and does not take into account the stochastic nature of recombination. Third, several studies confirm that errors in cattle pedigrees are common due to misinterpretation, misidentification and incorrect recording. Finally, F_{PED} assumes that the entire genome is selection neutral and does not account for potential bias resulting from selection, i.e. assumes equal levels of autozygosity over whole genome.

Recent development of high-density, genome-wide single nucleotide polymorphism chips has made the calculation of individual inbreeding coefficients from molecular data feasible. Among many methods proposed, one very simple and straightforward method seems to become the method of choice. It is called Runs of homozygosity (ROH). ROH are long stretches of homozygous segments that reflect autozygosity and its age. The principle of the approach is to consider a continuous length of homozygous loci corresponding to haplotype transmission from parent to offspring. The length of ROH is affected by recombination events, or put otherwise, by the number of generations from the common ancestor. Consequently, ROH and their appearance in the genome have a clear biological interpretation. Moreover, it is easy to derive inbreeding coefficients (F_{ROH}) from ROH. F_{ROH} is defined by the sum of lengths of all ROH of a specific minimum length divided by the total autosomal genome length. This measure has many advantages compared to inbreeding coefficients derived from pedigree (F_{PED}). It is sensitive to distant inbreeding and directly quantifies homozygosity. It also accounts for the stochastic nature of inheritance on individual level, it is applicable when pedigree data are not available, and it allows dissection of inbreeding on the chromosome level and even at individual marker level.

These allow better precision in estimating levels of inbreeding, and consequently, better monitoring and possible conservation of population, and for estimation of inbreeding depression.

The use of ROH is well established in humans, while this is not always the case for cattle and other animal populations. Furthermore, because of the lack of universal standards about ROH definition and identification, interpretations of the results from ROH studies are somewhat diverse not only for animal species but also for humans.

This thesis is therefore intended to contribute to both human and animal genetics in establishing the optimal way to estimate autozygosity by ROH. Cattle data from several breeds were used to prove efficiency of ROH inbreeding coefficients and to demonstrate their practical use. A possible association between bull fertility and genome wide autozygosity as well as dissecting this association down to genomic regions and individual genes could make inbreeding depression better understood, as well as it may be helpful in future human fertility studies.

CHAPTER 1 INBREEDING AND RUNS OF HOMOZYGOSITY: A POSSIBLE SOLUTION TO AN OLD PROBLEM

Ino Curik[1], Maja Ferenčaković[1], Johann Sölkner[2]

[1]Department of Animal Science, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25,10000 Zagreb, Croatia.
[2]Division of Livestock Sciences, Department of Sustainable Agricultural Systems,University of Natural Resources and Life Sciences Vienna, Gregor Mendel Str.33, A-1180 Vienna, Austria

Abstract

Quantifying human, plant and livestock inbreeding has been an important goal of evolutionary biologists and agricultural scientists for nearly a century, and methods to do so continue to evolve. This review examines current approaches for estimating inbreeding at individual and population levels based on genetic information. Central to this approach is the detection of Runs of Homozygosity (ROH), long stretches of homozygous genome that most likely arise when the individual is the offspring of related individuals. When related individuals mate, the offspring carry long sections of the genome that are homozygous and identical by descent (IBD). Long ROH are most likely derived from a recent ancestor; shorter ones, from a more distant ancestor. Calculating how much an individual's genome occurs as ROH of particular lengths (e.g. 41 Mb, 42 Mb, and 44 Mb) provides information about levels of inbreeding relative to reference populations specific numbers of generations ago. Although identifying and quantifying ROH can be complicated by genotyping errors and undetected heterozygosity within apparently continuous ROH, inbreeding estimates based on ROH clearly indicate that inbreeding levels in bovine and porcine populations are much higher than those in human populations. Frequencies of ROH vary widely within and across chromosomes, with chromosomes exhibiting ROH hotspots or “islands” as well as coldspots or “deserts”. The reasons for this variation are unclear and are attracting growing interest. Next-generation sequencing may improve our understanding of ROH and their usefulness as a tool in inbreeding research. We argue for combining ROH analysis and other genomic estimators unrelated to haplotype length in order to better define the inbreeding reference population.

Keywords: *Genomic inbreeding coefficients, Runs of homozygosity*, *Individual heterozygosity, Autozygosity*, *Livestock populations*

1. Introduction

Inbreeding refers to mating by parents who share one or more ancestors. Therefore the concept of inbreeding is closely connected to the idea of relatedness, and both play crucial roles in evolutionary genetics (Charlesworth & Charlesworth 1987) forensic science (Weir 1994), plant breeding (Hallauer & Miranda Filho 1981), animal breeding(Pirchner 1985; Kristensen & Sorensen 2005), biomedical research (Festing 1979), human health and genetics (Rudan *et al.* 2003a; Rudan *et al.* 2003b; Bittles & Black 2010) and conservation biology (Ballou 1997; Hedrick & Kalinowski 2000).

Despite the importance of inbreeding, the term is often misused, even by geneticists (Jacquard 1975; Templeton & Read 1994). This is primarily because of disagreement over where to draw the line between related and unrelated, given that any two individuals in an empirical population share at least one ancestor or over what we consider as a reference population where from we quantify changes. One approach to avoiding this ambiguity is to define inbreeding as a consequence of mating between two individuals in a population who are more related to each other than the average relatedness for that population. Note that in this definition the average relatedness in a population is considered as a reference population. As suggested by Lush *et al*. (1994): "The most satisfactory basis for defining inbreeding as a breeding policy or mating system is in terms of choices actually open to the breeder, or to the individual plant or animal in nature".

Another reason for confusion over the concept of inbreeding is that it is invoked to explain a wide range of genetic phenomena, including decreases in genetic diversity of finite populations, changes in inbreeding effective population size, genetic drift, changes in population structure, deviations from Hardy–Weinberg equilibrium, and decreases in population means. In simplest terms, inbreeding changes genotype frequencies by increasing homozygosity at the expense of heterozygosity while leaving gene (allele) frequencies unaffected. This can lead to redistribution of the genetic variations within and between populations (Fernandez *et al.* 1995), reduction in the population mean for traits closely related to fitness (Charlesworth & Willis 2009), higher incidence of homozygous recessive defects (Arcos-Burgos & Muenke 2002; Alvarez *et al.* 2009) and a decrease in homeostasis (Lerner 1954). Falconer & Mackay (1996) discuss these consequences in greater detail.

Even the earliest human civilizations were aware of the negative consequences of inbreeding, based on their observations of breeding in domestic animal populations, though precise attitudes towards inbreeding vary across religions and continents (Bittles & Black 2010). Despite this wide spread recognition of the negative effects of inbreeding on health and performance, it has often been used when creating livestock breeds as a way to ensure uniformity. In this way, inbreeding is a double-edged sword that, together with genetic drift and selection forces, can lead to improvement or deformations of individuals or the population.

The importance of inbreeding for understanding the evolution of plants and animals and for monitoring the quality of commercially important livestock breeds has led researchers to develop several genomic methods to estimate inbreeding. This review examines the state of the art in this field based on studies related to livestock, wildlife and human populations. The concept of identity by descent (IBD) is explained, and pre genomic approaches to estimating inbreeding based on pedigree and molecular markers are explored. Finally runs of homozygosity (ROH) are defined and discussed in detail, since ROH-based metrics form the foundation of many post-genomic studies aimed at quantifying and understanding inbreeding.

2. Identity by descent (IBD)

Identity by descent (IBD), coined by Crow (1954), links inbreeding and relatedness: two alleles (haplotypes) are IBD if they have been inherited from the same ancestral haplotype, either parental or maternal, in the absence of recombination or mutation. When inbreeding is calculated from the pedigree information IBD status can be interpreted through probability or path correlation and the value obtained is always determined by the pedigree pattern. This differs from the molecular approach typically used in simulations aimed at determining all IBD segments. Such exhaustive determination is impossible in the real world because not all IBD segments can be identified for all shared ancestors over a sufficiently long timescale. Complete phased sequence information is rarely available for all members of even restricted pedigrees.

To overcome this problem, researchers rely on additional information that directly assesses, or correlates with, autozygosity. For example, the chance that two identical-by- state (IBS) haplotypes are also IBD is low if their population frequency is high, and vice versa. One approach to determining whether two haplotypes are IBD is to define a suitable frequency threshold. Another approach is to take into account the effects of linkage on inbreeding level by defining a function that relates the expected genotypic frequencies for

inbreed populations to the expected identity disequilibrium (Cockerham & Weir 1968; Weir & Cockerham 1969b, a). For example, in order to estimate population-level inbreeding (*f*) at a biallelic locus (represented by alleles A_i and A_j with frequencies p_i and p_j, respectively), we can use the expected probabilities (*Pr*) of homozygote genotypes (A_iA_i, A_jA_j) and heterozygote genotype (A_iA_j, $i \neq j$), defined as follows:

$$Pr(A_iA_i) = fp_i + (1-f)p_i^2$$

$$Pr(A_iA_j) = 2(1-f)p_ip_j$$

$$Pr(A_jA_j) = fp_j + (1-f)p_j^2$$

Current methods for detecting IBD segments are substantially more complex than this simple example (Browning & Browning 2012; Thompson 2013).

Yet another approach to compensating for the lack of complete phased information is to exploit information on haplotype length: the longer homozygous haplotypes are, the more likely they are to be IBD (Figure. 1). This is because the longer the haplotype, the more likely it will undergo recombination or mutation as it segregates through the generations. As a result, the expected frequency of untouched haplotypes decreases with increasing length, making it more probable that two identical haplotypes are indeed IBD.

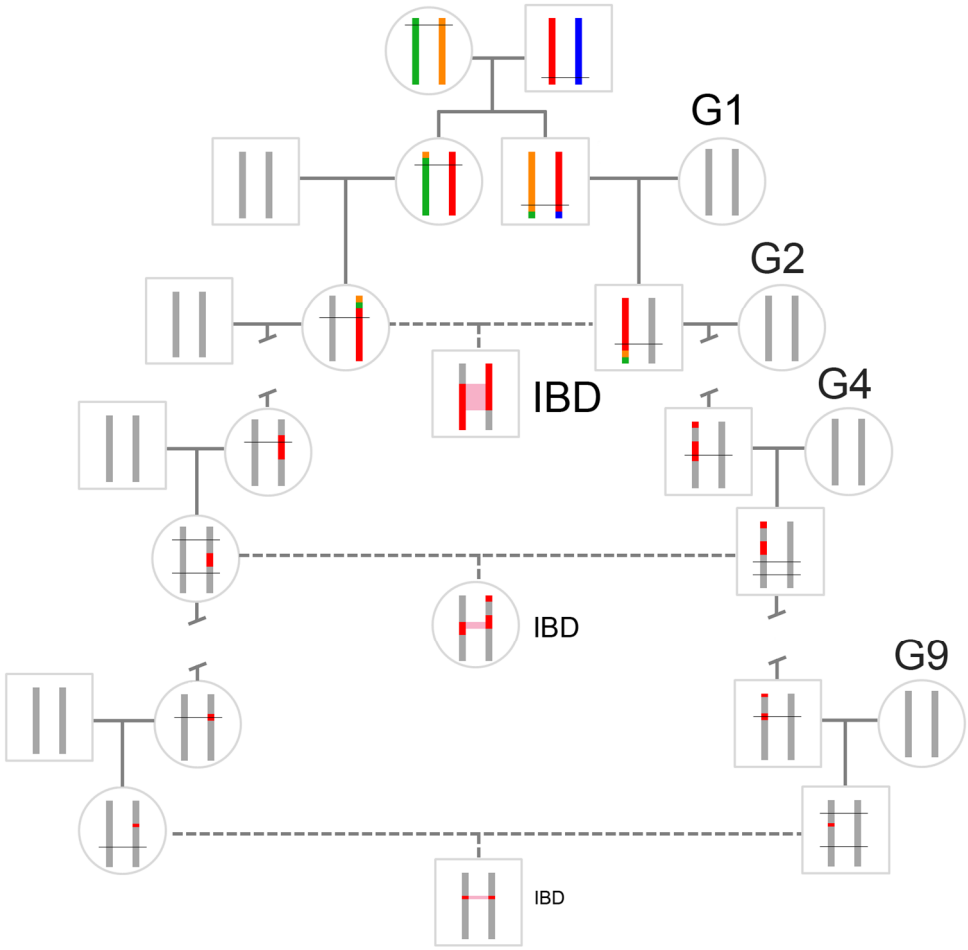


Figure 1. Hypothetical multi-generational pedigree illustrating the relationship between the lengths of IBD segments in an individual and remoteness from the common ancestor.
Founder chromosomes are shown in color (green, orange, red, and blue), while chromosomes of other individuals are shown in gray. Red shadowing marks IBD regions in homologous chromosomes. Chromosomes are assumed to be 100Mb long and to undergo recombination at 1.0cM/Mb.

3. Pedigree inbreeding coefficient (F_{PED})

Raymond Pearl, in a series of papers published between 1913 and 1917 made the first attempts to quantify inbreeding based on pedigree information. A few years later, Wright (1922) developed what would become a widely applied approach based on an inbreeding coefficient calculated using the path coefficients technique from the correlation between arbitrary values assigned to the union of all possible gametes. Although this coefficient was straightforward to calculate, its biological meaning was difficult to interpret, particularly for individuals with arbitrary pedigrees.

Malécot (1948) developed a more intuitive and interpretable definition of inbreeding based on the probability that a genotype is autozygous or that two haplotypes within a locus are IBD. Malécot defined the inbreeding coefficient as the probability that two haplotypes at

any locus randomly sampled among all loci in the genome are IBD. In the absence of mutations and selection, all loci are assumed to segregate under the same genealogical pattern and are therefore expected to have the same inbreeding coefficient, called the pedigree inbreeding coefficient (F_{PED}). In this way, F_{PED} is equivalent to the average genome-wide autozygosity of an individual or to the proportion of autozygosity of that individual.

Inbreeding estimates based on IBD must be referenced to an ancestral population in which none of the members are related. When calculating F_{PED}, for example, individuals not present in the pedigree are considered unrelated, and the reference population for the coefficient is taken to be the founding members. As originally defined, F_{PED} does not take into account the stochastic nature of inheritance resulting from the finite number of chromosomes and the small number of recombination events during meiosis. Therefore, F_{PED} for all offspring of a mating of first cousins is always the same (6.25×10^{-2}), despite the expected variance of autozygosity, estimated to be 5.90×10^{-4} in humans (Carothers *et al.* 2006). In reality, this variance increases with each meiosis, such that, for example, autozygosity can be higher in the offspring of third-cousin mating than in that of second-cousin mating.

F_{PED} estimates also assume neutrality. A simulation study by Curik *et al.* (2002) suggests that estimating inbreeding coefficients from pedigree leads to biased values for "true" or "realized" autozygosity. The magnitude of the bias depends on the selection intensity and genetic model of the traits under selection. Pedigree-based estimates of autozygosity will be lower than estimates based on loci featuring additive and partial dominance, and higher than estimates based on loci featuring overdominance. The effects of selection were ignored in the pre-genomic era as inbreeding was incorporated into quantitative genetics in the form of the infinitesimal model assuming selection does not affect autozygosity of a locus. These standings were nicely defended by Wray *et al.* (1990); 'if the selected trait is assumed to be controlled by many unlinked loci, each of small additive effect (the infinitesimal model),then the rate of inbreeding at selected loci is expected to be the same as at neutral loci'.

The inbreeding coefficient, as conceived by both Wright and Malécot, can be extended from the pedigree level to population level simply by averaging the coefficients of the individual pedigrees. Population-level coefficients can be compared with one another but if the pedigrees differ in their depth – i.e. in the numbers of complete equivalent generations – then the estimates will refer to different reference populations. Adjusting for differences in pedigree depth may improve the accuracy of inter-population comparisons. This adjustment

relies on the concept of equivalent complete generations, which was originally used to estimate inbreeding effective population size (Gutierrez *et al.* 2008; Gutierrez *et al.* 2009). More recently, Nagy *et al.* (2010) and Leroy *et al.* (2013) have used scatter plots of inbreeding level and numbers of complete equivalent generations to compare inbreeding in different livestock populations.

For calculating the inbreeding coefficient, Wright's approach remains the standard when it refers to a single individual with a simple pedigree. This is the method presented in most genetics textbooks. In contrast, the tabular method has proven quite efficient for rapidly calculating inbreeding coefficients for all members of a population, even those with extremely large and complex pedigrees (Tier 1990; VanRaden 1992; Aguilar & Misztal 2008). In other cases, the gene dropping approach may be the most appropriate for obtaining unbiased estimates, such as when calculating ancestral inbreeding coefficients (Suwanlee *et al.* 2007).

4. Genomic estimators of individual autozygosity unrelated to haplotype length

The first progress towards using molecular information to estimate inbreeding and individual multilocus heterozygosity (*MLH*), as a closely related metric, came from theoretical studies at the population level (Li & Horvitz 1953; Curie-Cohen 1982) and computer simulations at the individual level (Bereskin *et al.* 1969, 1970; Mitton & Pierce 1980). Inspired by the tabular method to calculate pedigree inbreeding coefficients using an additive relationship matrix, (Caballero & Toro 2002) defined the molecular inbreeding coefficient (F_{Mi}) as $F_{Mi}=2f_{Mii}-1$,where f_{Mii} is the molecular self coancestry coefficient derived from IBS similarity. The theory behind this method of calculating F_{Mi} is straightforward, and the value obtained is equal to the proportion of individual homozygosity (Saura *et al.* 2013). F_{Mi}, also equal to 1-*MLH*, forms the basis of several metrics to estimate inbreeding using information from the growing array of molecular technologies that take advantage of microsatellites as well as high-throughput genotypes that show co-dominance or high polymorphism. These metrics can be highly effective for analyzing wild life population genetics, where pedigrees are impossible to construct.

Instead of pedigree inbreeding coefficients, molecular techniques use several measures of individual multilocus heterozygosity. The most frequent ones are (a) individual multilocus heterozygosity (*MLH*), which measures the proportion of heterozygous loci (Coulson *et al.* 1998; Slate & Pemberton 2002); (b) mean d^2, defined as the squared difference in the numbers of repeat units between two alleles at a microsatellite locus, averaged over all typed loci (Coltman *et al.* 1998; Coulson *et al.* 1998; Pemberton *et al.* 1999); and (c) internal

relatedness (*IR*), defined as $IR = (2H - \sum h_i)/(2L - \sum h_i)$, where H is the number of homozygous loci, L is the number of all loci and h_i is the frequency of the i-th allele in the genotype (Amos *et al.* 2001). Some of less frequently used measures are described Amos *et al.* (2001), Aparicio *et al.* (2006) and Coltman and Slate (2003).

A wake-up call for the use of microsatellite-derived inbreeding metrics came in 2004 and 2005, when empirical studies (Slate *et al.* 2004), computer simulations (Balloux *et al.* 2004) and theoretical analyses (DeWoody & DeWoody 2005) showed that the numbers of microsatellite markers typically used, ranging between 15 and 50, are far too small to accurately estimate genome-wide heterozygosity or show any correlation with pedigree inbreeding coefficients. Now that genome-wide data on numerous single-nucleotide polymorphisms (SNPs) can be obtained simultaneously with high-density BeadChip technology, interest has grown for using molecular markers to estimate inbreeding and *MLH*.

These research efforts have focused on estimating *MLH* or standardized *MLH* (*sMLH*) using a growing array of methods in populations of humans (Carothers *et al.* 2006; Polasek *et al.* 2010), wild (Santure *et al.* 2010) and domestic animals (Curik *et al.* 2010; Saura *et al.* 2013; Silió *et al.* 2013). One single-point approach to calculating human inbreeding that does not include information on marker dependences is the PLINK genomic inbreeding coefficient (F_{PLINK}) (Purcell *et al.* 2007), defined as $F_{PLINK} = (O_i - E_i)/(L_i - E_i)$. If individual i has L_i genotyped autosomal loci, O_i and E_i denote the number of observed and expected homozygous genotypes, respectively. Purcell *et al.* (2007) noted that E_i must be adjusted when allele frequencies are not known but instead estimated from the sample. Another single-point approach that does not include information on marker dependences is F_{ADC} (Carothers *et al.* 2006) defined as, $F_{ADC} = \sum_k \alpha_k$, where $\alpha_k = 1$ if the genotype is homozygous or $\alpha_k = 1 - 1/(1 - \sum_l p_{kl^2})$ if it is heterozygous, and p_{kl} is the frequency of allele l at locus k. F_{ADC} provides an unbiased estimate of an individual's inbreeding at each locus α_k if p_{kl} is known.

FEstim (Leutenegger *et al.* 2003a) is a multi-point estimator of human inbreeding; it is calculated using a maximum likelihood approach that takes into account the frequencies of certain marker-alleles and inter-marker genetic distances. Observed marker genotypes are modeled using a hidden Markov chain that depends on inbreeding level and the rate of change in IBD status per cM. The reliability of *FEstim* depends on the density and heterozygosity rate of the markers used; long homozygous stretches aid substantially in the estimation of inbreeding, while diminishing the impact of isolated homozygous markers. Unfortunately, calculating *FEstim* requires a preliminary step of data "pruning", in which a subset of markers

in approximate linkage equilibrium with one another is carefully selected. This is necessary to avoid overestimating inbreeding when loci that deviate from linkage equilibrium are included.

Polasek *et al.* (2010) discussed in greater detail the calculations and properties of F_{PLINK}, F_{ADC} and *FEstim*. Several authors have examined other recently described methods for estimating inbreeding or the broader concept of relatedness between individuals (Caballero & Toro 2002; Browning & Browning 2012; Thompson 2013).

Inbreeding metrics based on multilocus microsatellite data have been applied almost exclusively to wildlife. They have rarely been applied to domestic animal populations (Curik *et al.* 2003; Slate *et al.* 2004), for which pedigree information continues to be widely used for estimating inbreeding depression. Nevertheless, approaches have been developed to estimate the inbreeding of domestic populations based on SNPs. One method relies on variance of genotype values (VanRaden 2007, 2008), another relies on a combination of variance of genotype values and levels of homozygosity (Yang *et al.* 2010) Both methods are conceptually related to the molecular inbreeding coefficient defined by Caballero and Toro (2002).

Inbreeding estimates can take on negative values only when reference population is referred to the populations that show Hardy–Weinberg equilibrium, indicating the occurrence of mating in which inbreeding is avoided. In such situations, the inbreeding coefficient should be interpreted through the correlation concept (Wang 2014).

5. Runs of homozygosity (ROH)

5.1. History and applications

Broman and Weber (1999b) were the first to recognize that long stretches of homozygous segments in human populations, later referred to as runs of homozygosity (ROH), most likely reflect autozygosity and may have far-reaching implications for human health. Gibson *et al.* (2006) were perhaps the first to fully appreciate the importance of this finding, and they further developed it by analyzing the lengths, numbers and distribution of ROH in apparently outbreed HapMap populations. Lencz *et al.* (2007) validated the prediction of Broman and Weber (1999b) by showing that ROH can be systematically used to map genes linked to diseases such as schizophrenia. Lencz *et al.* (2007) were the first to coin the term "ROH", defining it as a window of ≥100 consecutive SNPs on a single chromosome that does not receive a heterozygous call when uncalled SNPs are permitted.

McQuillan *et al.* (2008) provided perhaps the strongest basis for using ROH in population

genetics when they performed a comprehensive analysis of European populations, including island isolates within Croatia and Scotland. They defined a new genomic inbreeding coefficient (F_{ROH}), actually comprising the three coefficients $F_{ROH0.5}$, $F_{ROH1.5}$ and $F_{ROH5.0}$, based on ROH with respective lengths of 0.5, 1.5 and 5.0 Mb. Those authors showed that F_{ROH} correlates with F_{PED}, F_{PLINK} and *MLH*, with correlation coefficients ranging from 0.74–0.82. The strongest correlation was obtained between F_{ROH} and F_{PED}. The authors also confirmed that hemizygote deletions and variations in copy number have a detectable but negligible impact on ROH identification.

Since this seminal work, the ROH concept has been applied to population genomics and demography (Kirin *et al.* 2010; Nothnagel *et al.* 2010; Palamara *et al.* 2012), inbreeding depression (Keller *et al.* 2011; McQuillan *et al.* 2012), disease-linked genes (Nalls *et al.* 2009; Keller *et al.* 2012; Wang *et al.* 2013), and recombination (Bosse *et al.* 2012). Recent developments in microarray platforms and other methods for inexpensive genotyping of $>10^6$ SNPs has led to exponential growth in the numbers of ROH-based publications.

The ROH has even begun to find wide acceptance in studies of domestic animals. To the best of our knowledge, Sölkner *et al.* (2010) and (Ferenčaković *et al.* 2011) were the first to extend the ROH concept to cattle, and they were followed closely by Purfield *et al.* (2012) and (Ferenčaković *et al.* (2013a); Ferenčaković *et al.* (2013b)). A focus of these studies has been to compare FROH and F_{PED} in terms of ROH length, pedigree depth and quality, calculation algorithms and marker density. This work provides strong evidence that F_{ROH} is a better estimator of individual autozygosity than F_{PED}. Building on this work, researchers have begun to apply the ROH concept to the estimation of inbreeding depression in cattle (Bjelland *et al.* 2013); Curik *et al.* 2012) and pigs (Silió *et al.* 2013). (Kim *et al.* 2013) linked ROH to signatures of selection in Holstein cattle; they compared three groups of animals whose ancestors had been exposed to different selection pressures to optimize milk production.

5.2. The inbreeding coefficient FROH: concept and assumptions

McQuillan *et al.* (2008) introduced FROH as a genomic measure of individual autozygosity and defined it as the proportion of the autosomal genome lying in ROH of a certain minimal length relative to the overall genome in the area of interest. In this way, F_{ROH} ignores ROH on sex chromosomes in females, since they have different IBD distribution pattern, and in regions around centromeres, since including long genomic stretches devoid of SNPs may lead to biased estimates. The general formula for calculating F_{ROH} from chip data is $F_{ROH} = \sum L_{ROH} / L_{AUTOSOME}$ where $\sum L_{ROH}$ is the total length of all ROH in the genome of

an individual, where the regions contain the minimum specified number of successive homozygous SNPs, and $L_{AUTOSOME}$ refers to the specified length of the autosomal genome covered by SNPs on the chip. Previous studies have used $L_{AUTOSOME}$ values of 2 673 768 kb for the human genome and either 2 543 177 kb (Ferenčaković *et al.* 2011) or 2,500,265 kb (Purfield *et al.* 2012) for the cattle genome. Variations on these values are possible. For example, (Bjelland *et al.* 2013) define F_{ROH} on all 30 cattle chromosomes, covering a total length of 2 612 820 kb (Zimin *et al.* 2009).

F_{ROH} has an easy biological interpretation, and it can be conveniently partitioned into values for individual chromosomes (F_{ROH_Ch1}, F_{ROH_Ch2}, F_{ROH_Ch3},…, F_{ROH_Chn}) or even for specific chromosomal segments. Another advantage of F_{ROH} is that the reference population is clear: it is based on the expectation that two related individuals, or two gametes uniting in an individual, will share identical chromosomal segments (haplotypes) of a certain length, assuming they are IBD. While defining the reference population is straightforward enough, determining the number of generations back to the shared ancestry is not trivial when the genealogy is complex. In principle, this requires analyzing the distributions of the numbers and lengths of shared IBD haplotypes as a function of the number of generations back to the reference population; Browning and Browning (2012) and Thompson (2013) have written more on this topic. One way around this problem is to assume that the expected length of an IBD haplotype ($L_{IBD\text{-}H}|gcA$) follows an exponential distribution, the mean of which equals $100/(2\ gcA)$ cM, where gcA is the number of generations from the common ancestor. Assuming that $E(L_{IBD\text{-}H}|gcA)$=100/(2 gcA) and that 1 cM ≈ 1 Mb, we would expect ROH that are 16.6, 10.0 or 5.0 Mb long to come from a common ancestor occurring, respectively, three generations back (six meioses), five generations (10 meioses) or 10 generations (20 meioses). F_{ROH} values calculated for the sex chromosome would not reflect the same number of meioses since the common ancestor as the corresponding F_{ROH} values for autosomal chromosomes.

When converting ROH lengths to gcA, the assumption 1 cM ≈ 1 Mb is frequently taken, but the relationship between recombination rate and physical distance varies across species and chromosomes. For example, a more precise relationship of 1 cM ≈ 1.28 Mb has been determined for 29 autosomes in cattle (Arias *et al.* 2009), giving ROH lengths of 13.0, 7.8 and 3.9 Mb for reference populations occurring three, five and 10 generations ago. In contrast, analysis of four pig populations led to a quite different relationship of 1 cM ≈ 0.76 Mb (Herrero-Medrano *et al.* 2013).

5.3. Software and technical options

Tools frequently used to identify ROH segments in SNP chip data are PLINK v1.07 ((Purcell *et al.* 2007); http://pngu.mgh.harvard.edu/purcell/plink/) and the Golden Helix SNP and Variation suite (SVS; www.goldenhelix.com). Differences between the two approaches are readily visible when portrayed graphically (Figure. 2A); however, the two programs generate ROH-based inbreeding coefficients with correlation coefficients > 0.99. It is important to note that neither these nor other software tools takes into account the possibility of adjacent heterozygous SNPs or heterozygous SNPs lying close together in an ROH (Ferenčaković *et al.* 2013b). Such events are less likely to reflect sequencing errors and more likely to indicate that the region is actually heterozygous. Until ROH detection can be made more robust to this and other issues, visual analysis of ROH segments remains the only way to exclude spurious ROH.

5.3.1. PLINK

PLINK uses a sliding window approach to define an ROH as a stretch including a minimum specified number of homozygous SNPs within a specified kb distance. The software supports only basic ROH detection: the - -homozyg command defines ROH segments using a sliding window that searches along SNP data to detect homozygous stretches. PLINK first determines whether a given SNP may lie within an ROH by calculating the proportion of completely homozygous windows in which that SNP occurs. Using the default window threshold of 0.05 means that if 5% of these windows are completely homozygous, then the SNP is included in the ROH.

ROH can also be called according to the number of SNPs (- -homozyg-snp) or the minimum segment length (- -homozyg-kb). In either case, the sliding window size can be specified when calculating proportions using the options - -homozyg-window-snp or - -homozyg-window-kb.

Window size should not be larger than the desired number of SNPs, otherwise the program will fail to detect segments smaller than the window size.

PLINK also provides for specifying the maximum gap between two homozygous SNPs (- -homozyg-gap) and a minimum density threshold for ROH (- -homozyg-density). (Bjelland *et al.* 2013) provide more detailed description of the ROH algorithms implemented in PLINK.

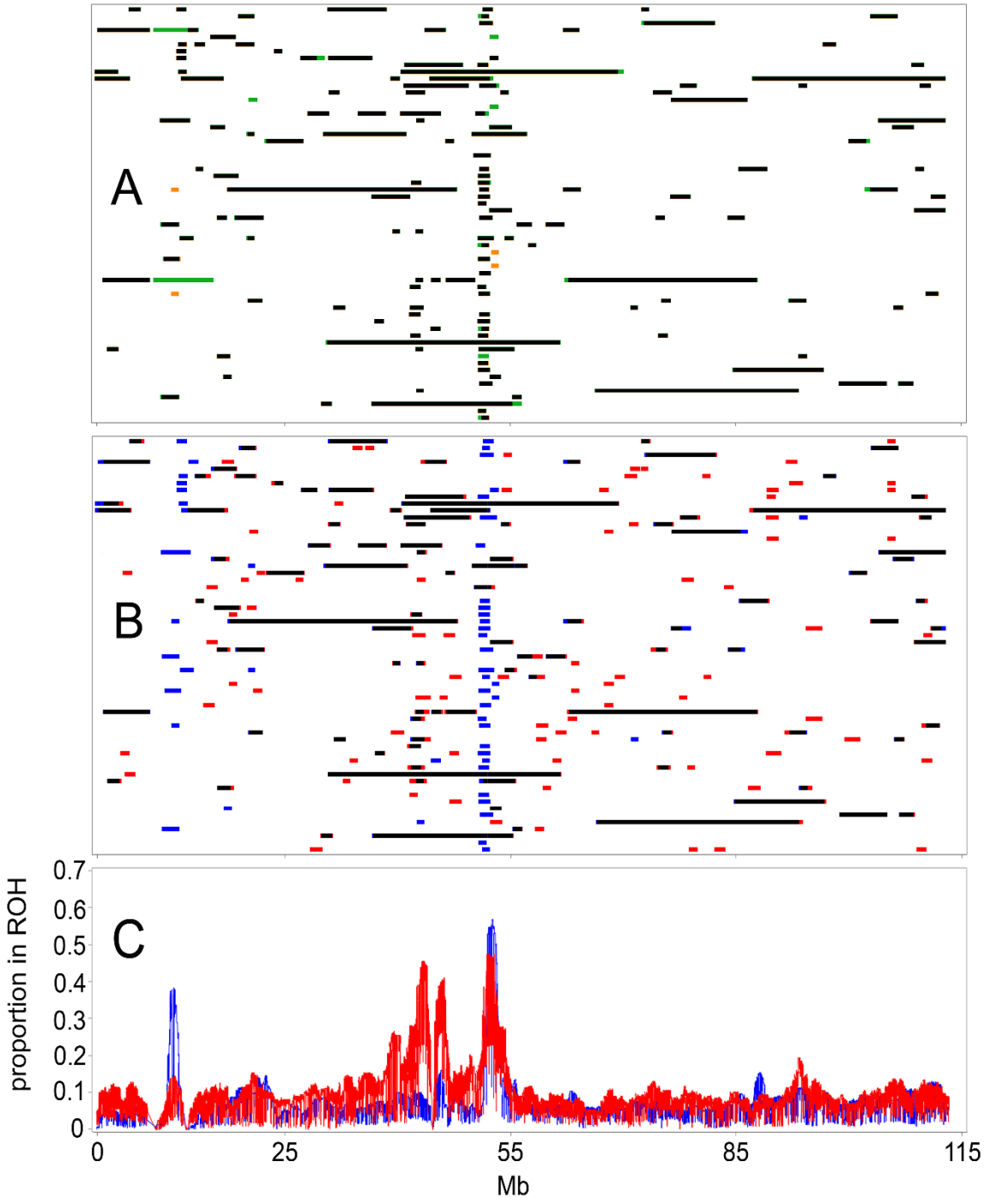


Figure. 2. Graphical presentation of ROH patterns on BTA 7. (A) *Individual barplots of Tyrol Grey, Brown Swiss and Pinzgauer animals (n=20 each), generated using PLINK (green) and SVS (orange) software. ROH that do not contribute to the difference are shown in black.* ***(B)*** *Individual barplots of Tyrol Grey, Brown Swiss and Pinzgauer animals (n=20 each) generated from datasets using an Illumina BovineSNP50 Genotyping BeadChip (50 k, red) or an Illumina BovineHD Genotyping BeadChip (HD, blue). ROH that do not contribute to the difference are shown in black.* ***(C)*** *Proportions of animals showing the specified numbers of SNPs in ROH among taurine breeds (108 Angus, 46 Brown Swiss and 97 Fleckvieh; red) and among indicine breeds (101 Brahman, 101 Gir, 134 Nelore; blue).*

5.3.2. SVS

The ROH module of SVS software does not rely on sliding windows to identify runs of consecutive homozygous SNPs. Instead the algorithm works continuously across an entire chromosome, examining every possible run for a match with user-specified input parameters. These parameters include minimum ROH length, minimum number of SNPs in the ROH, minimum density, maximum gap and maximum number of heterozygous and missing calls allowed. The algorithm applies the limit on heterozygous and missing calls strictly across the entire run. It considers every homozygous SNP as the potential start of a new ROH. Each SNP is then classified as homozygous, in which case the run is extended, or as heterozygous or missing, in which case the appropriate number of calls is increased.

5.3.3. Effect of SNP chip density on ROH identification

The density of the SNP chip used to identify ROH segments strongly influences the efficiency of detection. Purfield *et al.* (2012) investigated differences in ROH segments identified by the two SNP chips most frequently used in cattle: the Illumina BovineSNP50 Genotyping BeadChip with 54,001 SNPs (50 k), and the Illumina BovineHD Genotyping BeadChip with 777,972 SNPs (HD). The results suggested that the 50 k chip is appropriate only for identifying ROH longer than 5 Mb. In addition, lower-density chips can fail to detect heterozygous SNP genotypes within observed ROH (Ferenčaković *et al.* 2013b). As a result, analyses based on such chips can over-estimate the number of segments less than 4 Mb long. The 50 k chip lacks the sensitivity to precisely determine small segments, while the HD chip can fail to reveal certain ROH patterns (Figure. 2B).

5.4. Empirical studies of F_{ROH}

An overview of ROH inbreeding levels and mean F_{ROH} for cattle, pig and human populations is given in Table S1. F_{ROH} is generally much lower for humans than for livestock except for very isolated populations (see Table S1 in the Supplementary material). The much higher F_{ROH} for livestock, often >15%, reflects artificial selection and small effective population size. We found it difficult to compare F_{ROH} levels across species because of differences in chromosomal architecture and recombination rate.

5.5. Genomic landscape – "islands" and "deserts"

ROH are not uniformly distributed across the genome but are more prevalent in some regions, termed ROH islands by (Nothnagel *et al.* 2010) and ROH hotspots by (Pemberton *et*

al. 2012) Conversely, they are rare in so-called ROH deserts or coldspots. Among European human populations, chromosomes 3, 4 and 14 were found to contain an abundance of ROH (Nothnagel *et al.* 2010). When Pemberton *et al.* (2012) analyzed ROH patterns in 64 populations worldwide, they found distinct continental patterns. The two sets of studies overlapped in identifying hotspots on chromosomes 4 and 10, and these cannot be explained by linkage disequilibrium or local recombination alone. Many such regions harbor genes known to been affected by selection, and some of these genes have even become fixed. In contrast to ROH hotspots, cold spots are likely to be regions enriched for loci associated with a critical function (Pemberton *et al.* 2012). We have started to explore ROH islands in cattle populations, as have others (Karimi 2013). Figure 2C compares ROH frequencies at BTA 7 in taurine breeds (Angus, Brown Swiss, Fleckvieh) and indicine breeds (Brahman, Gir, Nelore). As in human populations, the different cattle breeds show similarities and differences along the chromosome. It may be fruitful to search within these regions for genes under selection (Kim *et al.* 2013) or genes linked to disease (Wang 2014). The results may also be used in meta-analyses that use multiple methods to search for selection signatures (Grossman *et al.* 2010; Utsunomiya *et al.* 2013).

6. Conclusions and research possibilities

The emergence of next-generation sequencing at the start of the 21st century has provided the technological basis for genotyping numerous loci (SNPs) at an affordable price. This has spurred growing interest in developing molecular metrics of inbreeding that will allow more accurate research and animal management. The ROH concept appears particularly well suited for estimating inbreeding at the individual and population levels but it needs to be improved. We foresee several key directions for future research. Efforts are needed to improve the estimation of inbreeding from next-generation sequencing data, particularly in order to reduce the effects of sequencing errors. Better insights are needed into recombination rates and their connection to ROH patterns; work by Bosse *et al.* (2012) on ROH and recombination in the porcine genome is a step in this direction. Analysis of the lengths and distribution of ROH may help reveal a population's demo-graphic history. MacLeod *et al.* (2013) have made progress in this direction by using whole-genome sequence data from two Holstein bulls to infer ancestral demography in terms of effective population size. Future studies should build on the approach of Szpiech *et al.* (2013), who searched the sequenced human exome for detrimental variation in order to gain insights into inbreeding depression. Definition of ROH length should be systematically researched and perhaps standardized, since current estimates

of inbreeding are based on arbitrarily defined threshold lengths, making it difficult to use data from shorter autozygous ROH. Recent developments in sequence analysis, such as automatic phasing of paternal/ maternal haplotype origin, will no doubt expand our ability to estimate inbreeding. At the same time, the field may easily take unexpected turns as the technology develops rapidly. We are curious about these future unexpected developments.

Conflict of interest statement

None.

Acknowledgments

Research time and visits required were funded by the Croatian Ministry of Science, Education and Sports (MZOS, Project 178-1780460-0546, 'Strategies for elimination of genetic defects in selected populations'), and by a bilateral collaborative project involving the MZOS and the Austrian Agency for International Cooperation in Education and Research (OeAD GmbH, project "Analysis of genetic diver-sity in cattle populations using dense sets of molecular markers"). We thank the Golden Helix team for providing a free license for SNP and Variation Suite v7.6.8 Win64. We are very grateful to Zuchtdata EDV-Dienstleistungen GmbH for providing pedigree data as well as Fleckvieh SNP chip data.

References

Aguilar I. & Misztal I. (2008) Technical Note: Recursive Algorithm for Inbreeding Coefficients Assuming Nonzero Inbreeding of Unknown Parents. *Journal of Dairy Science* **91**, 1669-72.

Alvarez G., Ceballos F.C. & Quinteiro C. (2009) The role of inbreeding in the extinction of a European royal dynasty. *PLoS ONE* **4**, e5174.

Amos W., Wilmer J. & Fullard K. (2001) The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society of London Series B-Biological Sciences* **268**, 2021 - 7.

Aparicio J.M., Ortego J. & Cordero P.J. (2006) What should we weigh to estimate heterozygosity, alleles or loci? *Molecular Ecology* **15**, 4659-65.

Arcos-Burgos M. & Muenke M. (2002) Genetics of population isolates. *Clinical genetics* **61**, 233-47.

Arias J., Keehan M., Fisher P., Coppieters W. & Spelman R. (2009) A high density linkage map of the bovine genome. *BMC genetics* **10**, 18.

Ballou J. (1997) Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. *Journal of Heredity* **88**, 169 - 78.

Balloux F., Amos W. & Coulson T. (2004) Does heterozygosity estimate inbreeding in real populations? *Molecular Ecology* **13**, 3021 - 31.

Bereskin B., Shelby C.E. & Hazel L.N. (1969) Monte Carlo Studies of Selection and Inbreeding in Swine I. Genetic and Phenotypic Trends. *Journal of Animal Science* **29**, 678-86.

Bereskin B., Shelby C.E. & Hazel L.N. (1970) Monte Carlo Studies of Selection and Inbreeding in Swine. II. Inbreeding Coefficients. *Journal of Animal Science* **30**, 681-9.

Bittles A.H. & Black M.L. (2010) Consanguinity, human evolution, and complex diseases. *Proceedings of the National Academy of Sciences* **107**, 1779-86.

Bjelland D., Weigel K., Vukasinovic N. & Nkrumah J. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science* **96**, 4697 - 706.

Bosse M., Megens H.-J., Madsen O., Paudel Y., Frantz L., Schook L., Crooijmans R. & Groenen M. (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *PLoS Genetics* **8**, e1003100.

Broman K.W. & Weber J.L. (1999) Long Homozygous Chromosomal Segments in Reference Families from the Centre d'Étude du Polymorphisme Humain. *American Journal of Human Genetics* **65**, 1493-500.

Browning S.R. & Browning B.L. (2012) Identity by Descent Between Distant Relatives: Detection and Applications. *Annual Review of Genetics* **46**, 617-33.

Caballero A. & Toro M.A. (2002) Analysis of genetic diversity for the management of conserved subdivided populations. *Conservation Genetics* **3**, 289-99.

Carothers A., Rudan I. & Kolcic I. (2006) Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. *Annals of Human Genetics* **70**, 666 - 76.

Charlesworth D. & Charlesworth B. (1987) Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology, Evolution, and Systematics* **18**, 237 - 68.

Charlesworth D. & Willis J. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783 - 96.

Cockerham C.C. & Weir B.S. (1968) Sib mating with two linked loci. *Genetics* **60**, 629-40.

Coltman D. & Slate J. (2003) Microsatellite measures of inbreeding: A meta-analysis. *Evolution* **57**, 971 - 83.

Coltman D.W., Bowen W.D. & Wright J.M. (1998) Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 803-9.

Coulson T.N., Pemberton J.M., Albon S.D., Beaumont M., Marshall T.C., Slate J., Guinness F.E. & Clutton-Brock T.H. (1998) Microsatellites reveal heterosis in red deer. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 489-95.

Crow J.F. (1954) Breeding structure of populations. II. Effective population number. pp. 543-56. The Genetical Society of Great Britain.

Curie-Cohen M. (1982) Estimates of inbreeding in a natural population: a comparison of sampling properties. *Genetics* **100**, 339-58.

Curik, I., Ferencakovic, M., Gredler, B. & Sölkner, J. (2010) Genome-wide heterozygosity and pedigree inbreeding coefficients in Simmental cattle population. In: *Proceedings - 9th World Congress on Genetics Applied to Livestock Production / German Society for Animal Science (ed.). Gießen: 9th World Congress on Genetics Applied to Livestock Production*. 1-4.

Curik, I; Ferenčaković, M., Sölkner, J. (2012) Modeling perspectives in the estimation of inbreeding depression based on genomic information: lessons from the bull fertility. In: *Genetika 2012: book of abstracts/ 6th Congress of the Genetic Society of Slovenia [and] 6th Meeting of the Slovenian Society for Human Genetics with International Participation/Uroš Potočnik (Ed.). Ljubljana: Genetic Society of Slovenia, 2012.* 82-82.

Curik I., Solkner J. & Stipic N. (2002) Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. *Journal of Animal Breeding & Genetics* **119**, 101-15.

Curik I., Zechner P., Solkner J., Achmann R., Bodo I. & Dovc P. (2003) Inbreeding, microsatellite heterozygosity, and morphological traits in Lipizzan horses. *Journal of Heredity* **94**, 125-32.

DeWoody Y.D. & DeWoody J.A. (2005) On the estimation of genome-wide heterozygosity using molecular markers. *Journal of Heredity* **96**, 85-8.

Falconer D. & Mackay T. (1996) Introduction to Quantitative Genetics. Harlow: Longman Group Ltd.

Ferenčaković M., Hamzic E., Gredler B., Curik I. & Solkner J. (2011) Runs of homozygosity reveal genome-wideautozygosity in the Austrian fleckvieh cattle. *Agriculturae Conspectus Scientificus* **76**, 325 - 8.

Ferenčaković M., Hamzić E., Gredler B., Solberg T.R., Klemetsdal G., Curik I. & Sölkner J. (2013a) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding & Genetics* **130**, 286-93.

Ferenčaković M., Solkner J. & Curik I. (2013b) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution* **45**, 42.

Fernandez A., Toro M.A. & Lopez-Fanjul C. (1995) The effect of inbreeding on the redistribution of genetic variance of fecundity and viability in Tribolium castaneum. *Heredity* **75**, 376-81.

Festing M.F.W. (1979) *Inbred strains in biomedical research*. Macmillan.

Gibson J., Morton N. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789 - 95.

Grossman S.R., Shylakhter I., Karlsson E.K., Byrne E.H., Morales S., Frieden G., Hostetter E., Angelino E., Garber M., Zuk O., Lander E.S., Schaffner S.F. & Sabeti P.C. (2010)

A Composite of Multiple Signals Distinguishes Causal Variants in Regions of Positive Selection. *Science* **327**, 883-6.

Gutierrez J., Cervantes I. & Goyache F. (2009) Improving the estimation of realized effective population sizes in farm animals. *Journal of Animal Breeding & Genetics* **126**, 327 - 32.

Gutierrez J., Cervantes I., Molina A., Valera M. & Goyache F. (2008) Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genetics Selection Evolution* **40**, 359 - 78.

Hallauer A.R. & Miranda Filho J.B. (1981) *Quantitative genetics in maize breeding*. Iowa State University Press, Ames.

Hedrick P.W. & Kalinowski S.T. (2000) Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics* **31**, 139-62.

Herrero-Medrano J., Megens H.-J., Groenen M., Ramis G., Bosse M., Perez-Enciso M. & Crooijmans R. (2013) Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC Genetics* **14**, 106.

Jacquard A. (1975) Inbreeding - One Word, Several Meanings. *Theoretical Population Biology* **7**, 338-63.

Karimi, S., 2013. Runs of Homozygosity Patterns in Taurine and Indicine Cattle Breeds (Master thesis). *BOKU-University of Natural Resources and Life Sciences, Vienna*, 53pp.

Keller M., Visscher P. & Goddard M. (2011) Quantification of Inbreeding Due to Distant Ancestors and its Detection Using Dense SNP Data. *Genetics* **189**, 237 - 49.

Keller M.C., Simonson M.A., Ripke S., Neale B.M., Gejman P.V., Howrigan D.P., Lee S.H., Lencz T., Levinson D.F., Sullivan P.F. & The Schizophrenia Psychiatric Genome-Wide Association Study C. (2012) Runs of Homozygosity Implicate Autozygosity as a Schizophrenia Risk Factor. *PLoS Genetics* **8**, e1002656.

Kim E.-S., Cole J.B., Huson H., Wiggans G.R., Van Tassell C.P., Crooker B.A., Liu G., Da Y. & Sonstegard T.S. (2013) Effect of Artificial Selection on Runs of Homozygosity in U.S. Holstein Cattle. *PLoS ONE* **8**, e80813.

Kirin M., McQuillan R., Franklin C., Campbell H., McKeigue P. & Wilson J. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996.

Kristensen T. & Sorensen A. (2005) Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Science* **80**, 121 - 33.

Lencz T., Lambert C., DeRosse P., Burdick K., Morgan T., Kane J., Kucherlapati R. & Malhotra A. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19942 - 7.

Lerner I.M. (1954) Genetic homeostasis. *Genetic homeostasis*.

Leroy G., Mary-Huard T., Verrier E., Danvy S., Charvolin E. & Danchin-Burge C. (2013) Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genetics Selection Evolution* **45**, 1.

Leutenegger A., Prum B., Genin E., Verny C., Lemainque A., Clerget-Darpoux F. & Thompson E. (2003) Estimation of the inbreeding coefficient through use of genomic data. *American Journal of Human Genetics* **73**, 516 - 23.

Li C. & Horvitz D. (1953) Some methods of estimating the inbreeding coefficient. *American Journal of Human Genetics* **5**, 107.

Lush J.L., Chapman A.B., Shrode R.R. & Crow J.F. (1994) *The Genetics of Populations*. Iowa Agriculture and Home Economics Experiment Station, College of Agriculture, Iowa State University.

MacLeod I.M., Larkin D.M., Lewin H.A., Hayes B.J. & Goddard M.E. (2013) Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors. *Molecular biology and evolution* **30**, 2209-23.

Malécot G. (1948) *Les mathématiques de l'hérédité*. Masson.

McQuillan R., Eklund N., Pirastu N., Kuningas M., McEvoy B.P., Esko T., Corre T., Davies G., Kaakinen M., Lyytikäinen L.-P., Kristiansson K., Havulinna A.S., Gögele M., Vitart V., Tenesa A., Aulchenko Y., Hayward C., Johansson Å., Boban M., Ulivi S., Robino A., Boraska V., Igl W., Wild S.H., Zgaga L., Amin N., Theodoratou E., Polašek O., Girotto G., Lopez L.M., Sala C., Lahti J., Laatikainen T., Prokopenko I., Kals M., Viikari J., Yang J., Pouta A., Estrada K., Hofman A., Freimer N., Martin N.G., Kähönen M., Milani L., Heliövaara M., Vartiainen E., Räikkönen K., Masciullo C., Starr J.M., Hicks A.A., Esposito L., Kolčić I., Farrington S.M., Oostra B., Zemunik T., Campbell H., Kirin M., Pehlic M., Faletra F., Porteous D., Pistis G., Widén E., Salomaa V., Koskinen S., Fischer K., Lehtimäki T., Heath A., McCarthy M.I., Rivadeneira F., Montgomery G.W., Tiemeier H., Hartikainen A.-L., Madden P.A.F., d'Adamo P., Hastie N.D., Gyllensten U., Wright A.F., van Duijn C.M., Dunlop M., Rudan I., Gasparini P., Pramstaller P.P., Deary I.J., Toniolo D., Eriksson J.G., Jula A., Raitakari O.T., Metspalu A., Perola M., Järvelin M.-R., Uitterlinden A., Visscher

P.M., Wilson J.F. & on behalf of the R.C. (2012) Evidence of Inbreeding Depression on Human Height. *PLoS Genetics* **8**, e1002655.

McQuillan R., Leutenegger A., Abdel-Rahman R., Franklin C., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A., Farrington S., Rudan P., Hayward C., Vitart V., Rudan I., Wild S., Dunlop M., Wright A., Campbell H. & Wilson J. (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics* **83**, 359 - 72.

Mitton J.B. & Pierce B.A. (1980) The distribution of individual heterozygosity in natural populations. *Genetics* **95**, 1043-54.

Nagy I., Curik I., Radnai I., Cervantes I., Gyovai P., Baumung R., Farkas J. & Szendro Z. (2010) Genetic diversity and population structure of the synthetic Pannon White rabbit revealed by pedigree analyses. *Journal of Animal Science* **88**, 1267 - 75.

Nalls M., Guerreiro R., Simon-Sanchez J., Bras J., Traynor B., Gibbs J., Launer L., Hardy J. & Singleton A. (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics* **10**, 183 - 90.

Nothnagel M., Lu T.T., Kayser M. & Krawczak M. (2010) Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human Molecular Genetics* **19**, 2927-35.

Palamara P.F., Lencz T., Darvasi A. & Pe'er I. (2012) Length distributions of identity by descent reveal fine-scale demographic history. *American Journal of Human Genetics* **91**, 809-22.

Pemberton J.M., Coltman D.W., Coulson T.N. & Slate J. (1999) Using microsatellites to measure the fitness consequences of inbreeding and outbreeding. pp. 151-64.

Pemberton T., Absher D., Feldman M., Myers R., Rosenberg N. & Li J. (2012) Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics* **91**, 275 - 92.

Pirchner F. (1985) Genetic structure of populations. In: *General and quantitative genetics* (ed. by Chapman AB), pp. 227-50. World Animal Science A4, Elsevier, Amsterdam.

Polasek O., Hayward C., Bellenguez C., Vitart V., Kolcic I., McQuillan R., Saftic V., Gyllensten U., Wilson J., Rudan I., Wright A., Campbell H. & Leutenegger A.-L. (2010) Comparative assessment of methods for estimating individual genome-wide homozygosity-by-descent from human genomic data. *BMC Genomics* **11**, 139.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M., Bender D., Maller J., Sklar P., De Bakker P., Daly M. & Sham P. (2007) PLINK: a tool set for whole-genome

association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559 - 75.

Purfield D., Berry D., McParland S. & Bradley D. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* **13**, 70.

Rudan I., Rudan D. & Campbell H. (2003a) Inbreeding and risk of late onset complex disease. *Journal of Medical Genetics* **40**, 925 - 32.

Rudan I., Smolej-Narancic N., Campbell H., Carothers A., Wright A., Janicijevic B. & Rudan P. (2003b) Inbreeding and the genetic complexity of human hypertension. *Genetics* **163**, 1011-21.

Santure A., Stapley J., Ball A., Birkhead T., Burke T. & Slate J. (2010) On the use of large marker panels to estimate inbreeding and relatedness: empirical and simulation studies of a pedigreed zebra finch population typed at 771 SNPs. *Molecular Ecology* **19**, 1439 - 51.

Saura M., Fernández A., Rodríguez M.C., Toro M.A., Barragán C., Fernández A.I. & Villanueva B. (2013) Genome-Wide Estimates of Coancestry and Inbreeding in a Closed Herd of Ancient Iberian Pigs. *PLoS ONE* **8**, e78314.

Silió L., Rodríguez M.C., Fernández A., Barragán C., Benítez R., Óvilo C. & Fernández A.I. (2013) Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal of Animal Breeding & Genetics* **130**, 349 - 60

Slate J., David P., Dodds K., Veenvliet B., Glass B., Broad T. & McEwan J. (2004) Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: theoretical expectations and empirical data. *Heredity* **93**, 255 - 65.

Slate J. & Pemberton J. (2002) Comparing molecular measures for detecting inbreeding depression. *Journal of Evolutionary Biology* **15**, 20 - 31.

Sölkner, J., Ferencakovic, M., Gredler, B. & Curik, I. (2010) Genomic metrics of individual autozygosity, applied to a cattle population. In: *Book of Abstracts of the 61st Annual Meeting of the European Association of Animal Production*, / *Wageningen Academic Publishers (ed.). Wageningen Academic Publishers*. 306-307.

Suwanlee S., Baummung R., Solkner J. & Curik I. (2007) Evaluation of ancestral inbreeding coefficients: Ballou's formula versus gene dropping. *Conservation Genetics* **8**, 489 - 95.

Szpiech Z.A., Xu J., Pemberton T.J., Peng W., Zollner S., Rosenberg N.A. & Li J.Z. (2013) Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics* **93**, 90-102.

Templeton A.R., Read B., (1994) Inbreeding: One Word, Several Meanings, Much Confusion. In: Loeschcke, V., Tomiuk, J., Jain, S.K., (Eds.), *Conservation Genetics*. Birkhäuser–Verlag, Basel, Switzerland, pp. 91-106.

Thompson E.A. (2013) Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics* **194**, 301-26.

Tier B. (1990) Computing Inbreeding Coefficients Quickly. *Genetics Selection Evolution* **22**, 419-30.

Utsunomiya Y.T., Pérez O'Brien A.M., Sonstegard T.S., Van Tassell C.P., do Carmo A.S., Mészáros G., Sölkner J. & Garcia J.F. (2013) Detecting Loci under Recent Positive Selection in Dairy and Beef Cattle by Combining Different Genome-Wide Scan Methods. *PLoS ONE* **8**, e64280.

VanRaden P. (2007) Genomic measures of relationship and inbreeding. *INTERBULL bulletin*, 33.

VanRaden P. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414 - 23.

VanRaden P.M. (1992) Accounting for Inbreeding and Crossbreeding in Genetic Evaluation of Large Populations. *Journal of Dairy Science* **75**, 3136-44.

Wang C., Xu Z., Jin G., Hu Z., Dai J., Ma H., Jiang Y., Hu L., Chu M. & Cao S. (2013) Genome-wide analysis of runs of homozygosity identifies new susceptibility regions of lung cancer in Han Chinese. *Journal of Biomedical Research* **27**, 208.

Wang J. (2014) Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. *Journal of Evolutionary Biology* **27,** 518-530.

Weir B.S. (1994) The Effects of Inbreeding on Forensic Calculations. *Annual Review of Genetics* **28**, 597-621.

Weir B.S. & Cockerham C.C. (1969a) Group inbreeding with two linked loci. *Genetics* **63**, 711-42.

Weir B.S. & Cockerham C.C. (1969b) Pedigree mating with two linked loci. *Genetics* **61**, 923-40.

Wray N.R., Woolliams J.A. & Thompson R. (1990) Methods for predicting rates of inbreeding in selected populations. *Theoretical and Applied Genetics* **80**, 503-12.

Wright S. (1922) Coefficients of inbreeding and relationship. *The American Naturalist* **56**, 330 - 8.

Yang J., Benyamin B., McEvoy B., Gordon S., Henders A., Nyholt D., Madden P., Heath A., Martin N. & Montgomery G. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565 - 9.

Zimin A.V., Delcher A.L., Florea L., Kelley D.R., Schatz M.C., Puiu D., Hanrahan F., Pertea G., Van Tassell C.P., Sonstegard T.S., Marcais G., Roberts M., Subramanian P., Yorke J.A. & Salzberg S.L. (2009) A whole-genome assembly of the domestic cow, Bos taurus. *Genome Biology* **10**, R42.

Appendix A. Supplementary material

Table S.1. *ROH inbreeding level, mean $F_{ROH,}$ in populations of domestic animals and humans*

Species	Population	No. individuals	Mean (SD) F_{ROH} %	ROH length	Chromosomes	No. of SNPs	Software	Reference
Cattle	Holstein USA	5853	3.8 (2.1)	>5.0 Mb	30	7997[P] (54001, 61%)	PLINK	Bjelland et al., (2013)
	Simmental AUT	500	9.0 (2.2)	>1.0	29	41733 (54001, 87%)	PLINK	Ferenčaković et al. (2011)
		500	5.4 (2.0)	>2.0	29	41733 (54001, 87%)	PLINK	Ferenčaković et al. (2011)
		500	3.2 (1.9)	>4.0	29	41733 (54001, 87%)	PLINK	Ferenčaković et al. (2011)
		500	2.1 (1.7)	>8.0	29	41733 (54001, 87%)	PLINK	Ferenčaković et al. (2011)
		500	1.8 (1.3)	>16.0	29	41733 (54001, 87%)	PLINK	Ferenčaković et al. (2011)
	Brown Swiss AUT	304	15.6 (3.2)	>1.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
			12.9 (3.2)	>2.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
			7.4 (2.9)	>8.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
			3.7 (2.2)	>16.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al.

							(2013a)
Brown Swiss AUT	46	15.1 (3.6)[a]	>1.0	29	615618 (777972, 79%)	SVS	Ferenčaković et al., (2013b)
		13.2 (3.7)[a]	>2.0	29	615618 (777972, 79%)	SVS	Ferenčaković et al., (2013b)
		10.9 (3.5)[a]	>4.0	29	615618 (777972, 79%)	SVS	Ferenčaković et al., (2013b)
		7.9 (3.2)[a]	>8.0	29	615618 (777972, 79%)	SVS	Ferenčaković et al., (2013b)
		4.2(2.2)[a]	>16.0	29	615618 (777972, 79%)	SVS	Ferenčaković et al., (2013b)
Norwegian Red	498	9.9 (2.5)	>1.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
		7.4 (2.5)	>2.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
		3.5 (2.1)	>8.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
		2.0 (1.5)	>16.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
Pinzgauer	118	6.2 (2.4)[a]	>1.0	29	606120 (777972, 78%)	SVS	Ferenčaković et al., (2013b)
		5.2 (2.3)[a]	>2.0	29	606120 (777972, 78%)	SVS	Ferenčaković et al., (2013b)
		4.2 (2.2)[a]	>4.0	29	606120 (777972, 78%)	SVS	Ferenčaković et al., (2013b)
		3.0 (1.9)[a]	>8.0	29	606120 (777972, 78%)	SVS	Ferenčaković et al.,

								(2013b)
			1.7 (1.5)[a]	>16.0	29	606120 (777972, 78%)	SVS	Ferenčaković et al., (2013b)
	Tyrol Grey	117	8.7 (3.1)	>1.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
			6.9 (3.2)	>2.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
			3.6 (3.1)	>8.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
			2.5 (2.5)	>16.0	29	36273 (54001, 67%)	SVS	Ferenčaković et al. (2013a)
	Tyrol Grey	113	6.6 (3.2)[a]	>1.0	29	684172 (777972, 88%)	SVS	Ferenčaković et al., (2013b)
			5.2 (3.1)[a]	>2.0	29	684172 (777972, 88%)	SVS	Ferenčaković et al., (2013b)
			4.2 (3.0)[a]	>4.0	29	684172 (777972, 88%)	SVS	Ferenčaković et al., (2013b)
			3.0 (2.7)[a]	>8.0	29	684172 (777972, 88%)	SVS	Ferenčaković et al., (2013b)
			1.7 (2.1)[a]	>16.0	29	684172 (777972, 88%)	SVS	Ferenčaković et al., (2013b)
Pigs	Torbiscal Iberian	64	6.7 (2.8)	>1.0	18	20652 (62163, 33.2%)	SVS	Silio et al., (2013)
		64	4.4 (3.0)	>5.0	18	20652 (62163, 33.2%)	SVS	Silio et al., (2013)
Humans	Orcadians	249	3.9 (1.3)[b]	>0.5	22	289738 (351454, 82%)	PLINK	McQuillan et al. (2008)
		249	1.0 (1.1)[b]	>1.5	22	289738 (351454, 82%)	PLINK	McQuillan et al.

							(2008)
	249	0.5 (0.8)[b]	>5.0	22	289738 (351454, 82%)	PLINK	McQuillan et al. (2008)
Oceania	.	11.7 (.)	>0.5	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
	.	0.4 (.)	>5.0	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
America	.	17.2 (.)	>0.5	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
	.	3.9 (.)	>5.0	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
Africa	.	2.4 (.)	>0.5	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
	.	0.3 (.)	>5.0	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
Central/South Asia	207	6.0 (.)	>0.5	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
		1.5 (.)	>5.0	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
West Asia	176	5.9 (.)	>0.5	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
		1.4 (.)	>5.0	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
East Asia	235	5.3 (.)	>0.5	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
		0.2 (.)	>5.0	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
Europe	160	4.6 (.)	>0.5	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
		0.2 (.)	>5.0	22	415130 (644258, 64%)	PLINK	Kirin et al. (2010)
Korčula Island, Croatia	866	0.7 (0.9)[c]	>1	22	48168[p] (351454, 13.7%)	PLINK	McQuillan et al. (2012)
	866	1.3 (1.4)[c]	>1.5	22	318448 (351454, 90.6%)	PLINK	McQuillan et al. (2012)
City of Split, Croatia	499	0.1 (0.2)[c]	>1	22	33718[p] (351454, 9.6%)	PLINK	McQuillan et al. (2012)
	499	0.7 (0.7)[c]	>1.5	22	325070 (351454, 92.5%)	PLINK	McQuillan et al. (2012)
Vis Island, Croatia	778	0.5 (0.7)[c]	>1	22	47802[p] (351454, 13.6%)	PLINK	McQuillan et al.

							(2012)
	778	0.9 (1.0)[c]	>1.5	22	299337 (351454, 85.2%)	PLINK	McQuillan et al. (2012)
Estonians	2395	0.1 (0.0)[c]	>1	22	33852[p] (351454, 9.6%)	PLINK	McQuillan et al. (2012)
	2395	0.6 (0.5)[c]	>1.5	22	321859 (351454, 91.6%)	PLINK	McQuillan et al. (2012)
Rucphen, Netherlands	789	0.5 (0.7)[c]	>1	22	43019[p] (351454, 12.2%)	PLINK	McQuillan et al. (2012)
	789	1.1 (1.1)[c]	>1.5	22	307909 (351454, 87.6%)	PLINK	McQuillan et al. (2012)
Rotterdam, Netherlands	5737	0.0 (1.2)[c]	>1	22	49162[p] (351454, 14.0%)	PLINK	McQuillan et al. (2012)
	5737	0.3 (0.4)[c]	>1.5	22	307042 (351454, 87.4%)	PLINK	McQuillan et al. (2012)
FINRISK study, Finland	1884	0.2 (0.4)[c]	>1	22	45433[p] (351454, 12.9%)	PLINK	McQuillan et al. (2012)
	1884	0.8 (0.7)[c]	>1.5	22	300312 (351454, 85.4%)	PLINK	McQuillan et al. (2012)
Health2000 Survey, Finland	2101	0.2 (0.2)[c]	>1	22	45159[p] (351454, 12.8%)	PLINK	McQuillan et al. (2012)
	2101	0.8 (0.5)[c]	>1.5	22	300493 (351454, 85.5%)	PLINK	McQuillan et al. (2012)
Helsinki, Finland	1721	0.1 (0.2)[c]	>1	22	45479[p] (351454, 12.9%)	PLINK	McQuillan et al. (2012)
	1721	0.6 (0.4)[c]	>1.5	22	298835 (351454, 85.0%)	PLINK	McQuillan et al.

							(2012)
Northern Finland	4988	0.3 (0.4)[c]	>1	22	44560[p] (351454, 12.7%)	PLINK	McQuillan et al. (2012)
	4988	1.0 (0.7)[c]	>1.5	22	302524 (351454, 86.1%)	PLINK	McQuillan et al. (2012)
Young Finns, Finland	2437	0.2 (0.3)[c]	>1	22	44890[p] (351454, 12.8%)	PLINK	McQuillan et al. (2012)
	2437	0.8 (0.5)[c]	>1.5	22	299112 (351454, 85.1%)	PLINK	McQuillan et al. (2012)
Carlatino, Italy	430	0.4 (1.1)[c]	>1	22	48204[p] (351454, 13.7%)	PLINK	McQuillan et al. (2012)
	430	0.8 (1.4)[c]	>1.5	22	300235 (351454, 85.4%)	PLINK	McQuillan et al. (2012)
Val Borbera, Italy	961	0.4 (0.6)[c]	>1	22	47960[p] (351454, 13.6%)	PLINK	McQuillan et al. (2012)
	961	0.8 (0.8)	>1.5	22	305451 (351454, 86.9%)	PLINK	McQuillan et al. (2012)
Friuli-Venezia-Giulia-Genetic Park, Italy	1661	0.9 (1.2)[c]	>1	22	47960[p] (351454, 13.6%)	PLINK	McQuillan et al. (2012)
	1661	1.5 (1.9)[c]	>1.5	22	300648 (351454, 85.5%)	PLINK	McQuillan et al. (2012)
South Tyrol, Italy	1079	0.5 (0.7)[c]	>1	22	47118[p] (351454, 13.4%)	PLINK	McQuillan et al. (2012)
	1079	0.9 (1.0)[c]	>1.5	22	307473 (351454, 87.5%)	PLINK	McQuillan et al. (2012)
Lothian Region, Scotland	512	0.1 (0.5)	>1	22	46827[p] (351454, 13.3%)	PLINK	McQuillan et al.

							(2012)
	512	0.3 (0.6)	>1.5	22	297795 (351454, 84.7%)	PLINK	McQuillan et al. (2012)
Lothian Region, Scotland	1005	0.0 (0.2)	>1	22	47139[p] (351454, 13.4%)	PLINK	McQuillan et al. (2012)
	1005	0.3 (0.3)	>1.5	22	297795 (351454, 84.7%)	PLINK	McQuillan et al. (2012)
Orkney Islands, Scotland	697	0.4 (0.5)	>1	22	45208[p] (351454, 12.9%)	PLINK	McQuillan et al. (2012)
	697	0.8 (0.8)	>1.5	22	306689 (351454, 87.3%)	PLINK	McQuillan et al. (2012)
Nation collection, Scotland	842	0.0 (0.1)	>1	22	46781[p] (351454, 13.3%)	PLINK	McQuillan et al. (2012)
	842	0.3 (0.3)	>1.5	22	306310 (351454, 87.2%)	PLINK	McQuillan et al. (2012)
Northern Sweden, Sweden	638	1.2 (1.3)	>1	22	34917[p] (351454, 9.9%)	PLINK	McQuillan et al. (2012)
	638	2.8 (2.4)	>1.5	22	303583 (351454, 86.4%)	PLINK	McQuillan et al. (2012)
Australia	3925	0.0 (0.1)	>1	22	31760[p] (351454, 9.0%)	PLINK	McQuillan et al. (2012)
	3925	0.2 (0.3)	>1.5	22	295000 (351454, 83.9%)	PLINK	McQuillan et al. (2012)

[a]SD values additionally calculated, [b]SD estimated from SE, [c] SD estimated from CI, [p] SNP data LD pruned

CHAPTER 2 AIM OF THE THESIS AND HYPOTHESES

Aim of this thesis was to analyze the identification of ROH of different length categories and the estimation of genomic inbreeding coefficients based on ROH in cattle breeds taking into account effects of chip density (777 972 versus 54 001 SNPs) and genotyping errors. Furthermore, ROH inbreeding coefficients were compared with pedigree inbreeding coefficients in various cattle populations in order to prove their efficiency and simplicity. Finally, the association with semen quality traits in cattle and inbreeding at individual SNPs was assessed to find genomic regions that influence those traits when autozygous.

HYPOTHESES

1. ***Genotyping errors and SNP chip density affect estimates of autozygosity from ROH.***

2. ***Knowledge about ROH distribution (number and size) enables precise estimation of autozygosity at individual and population levels in cattle.***

3. ***Genomic autozygosity affects bull semen quality.***

OBJECTIVES

- to compare ROH identified with SNP chips of different densities for cattle and by allowing different rates of genotyping errors

- to determine ROH from SNP chip data in cattle and to estimate levels of autozygosity derived from ROH

- to use ROH mapping to investigate whether specific areas of the genome have an influence on sperm quality traits in bulls when autozygous

CHAPTER 3 ESTIMATES OF AUTOZYGOSITY DERIVED FROM RUNS OF HOMOZYGOSITY: EMPIRICAL EVIDENCE FROM SELECTED CATTLE POPULATIONS

M. Ferenčaković[1], E. Hamzić[2,3], B. Gredler[4], T.R. Solberg[5], G. Klemetsdal[3], I. Curik[1] & J. Sölkner[2]

[1]Department of Animal Science, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

[2]Division of Livestock Sciences, Department of Sustainable Agricultural Systems, University of Natural Resources and Life Sciences Vienna, Vienna, Austria

[3]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, Ås, Norway

[4]Qualitas AG, Zug, Switzerland

[5]GENO SA, Hamar, Norway

Summary

Using genome-wide SNP data, we calculated genomic inbreeding coefficients ($F_{ROH > 1\ Mb}$, $F_{ROH > 2\ Mb}$, $F_{ROH > 8\ Mb}$ and $F_{ROH > 16\ Mb)}$ derived from runs of homozygosity (ROH) of different lengths (>1, >2, >8 and > 16 Mb) as well as from levels of homozygosity (F_{HOM}). We compared these values of inbreeding coefficients with those calculated from pedigrees (F_{PED}) of 1422 bulls comprising Brown Swiss (304), Fleckvieh (502), Norwegian Red (499) and Tyrol Grey (117) cattle breeds. For all four breeds, population inbreeding levels estimated by the genomic inbreeding coefficients $F_{ROH > 8\ Mb}$ and $F_{ROH > 16\ Mb}$ were similar to the levels estimated from pedigrees. The lowest values were obtained for Fleckvieh (F_{PED} = 0.014, $F_{ROH > 8\ Mb}$ = 0.019 and $F_{ROH > 16\ Mb}$ = 0.008); the highest, for Brown Swiss (F_{PED} = 0.048, $F_{ROH > 8\ Mb}$ = 0.074 and $F_{ROH > 16\ Mb}$ = 0.037). In contrast, inbreeding estimates based on the genomic coefficients $F_{ROH > 1\ Mb}$ and $F_{ROH > 2\ Mb}$ were considerably higher than pedigree-derived estimates. Standard deviations of genomic inbreeding coefficients were, on average, 1.3–1.7-fold higher than those obtained from pedigrees. Pearson correlations between genomic and pedigree inbreeding coefficients ranged from 0.50 to 0.62 in Norwegian Red (lowest correlations) and from 0.64 to 0.72 in Tyrol Grey (highest correlations). We conclude that the proportion of the genome present in ROH provides a good indication of inbreeding levels and that analysis based on ROH length can indicate the relative amounts of autozygosity due to

recent and remote ancestors.

Keywords: *Cattle, genome inbreeding, pedigree, runs of homozygosity*

Introduction

Mating of related individuals results in inbred off- spring. In closed and selected populations, inbreeding is unavoidable. Increasing inbreeding reduces genetic variation and leads to inbreeding depression. The individual inbreeding coefficient (F) is defined as the proportion of an individual's genome that is autozygous, that is, that has homozygous "identical by descent" (IBD) status, or equivalently the probability of a randomly sampled locus in the genome to be autozygous. The average of all individual values of F represents the inbreeding level of a population.

Traditionally, inbreeding coefficients are calculated from pedigree records (F_{PED}) using path coefficient methodology first proposed by Wright (1922). When pedigrees are not available, inbreeding coefficients can be derived from genotypic data examining the difference between observed and expected multilocus heterozygosity (e.g. Polasek *et al.* (2010)). The recent development of high-density, genome-wide single- nucleotide polymorphism (SNP) bead chips has revived interest in the calculation of individual inbreeding coefficients from molecular information. Inbreeding levels are being calculated from SNPs using variance of genotype values (VanRaden 2008) or using a combination of variance of genotype values and levels of homozygosity (Yang *et al.* 2010). A higher level of inbreeding, that is, proportion of genome that is IBD, brings more chance for homozygous deleterious recessives. These are considered to be a main cause of inbreeding depression. To avoid inbreeding depression, accurate and sensitive estimation of inbreeding is very important. Keller *et al.* (2011) recently concluded that inbreeding coefficients derived from runs of homozygosity (ROH) are optimal for the estimation of genome-wide autozygosity and for detecting inbreeding effects. In the whole-genome sequence, ROH are defined as continuous and uninterrupted stretch of DNA sequence without heterozygosity in diploid state. When using SNP data, ROH may be defined as long stretches of homozygous SNPs. Broman and Weber (1999a) first recognized that ROH are highly likely to be autozygous. Because recombination events interrupt long chromosome segments, over time very long ROH are expected to be autozygous segments originated from recent common ancestors. On the other side, shorter ROH likely originated from more remote ancestors, but they can also include some non-IBD segments. ROH length can give insight into the age of inbreeding. The expected length of an autozygous segment follows an exponential distribution with mean equal to $1/2\,g$ Morgans, where g is the number

of generations as the common ancestor (e.g. Howrigan *et al.* (2011)). ROH did not receive serious attention until the first study using an SNP array was carried out by Gibson *et al.* (2006). This was followed by a number of population genetics studies analyzing ROH in humans (McQuillan *et al.* 2008; Kirin *et al.* 2010; Nothnagel *et al.* 2010) and in cattle (Sölkner *et al.* 2010; Ferenčaković *et al.* 2011), as well as by the association studies examining the relationship between ROH and complex diseases and traits (Lencz *et al.* 2007).

Molecular approaches based on ROH and SNPs may help avoid several drawbacks of using pedigrees to analyze inbreeding. First, F_{PED} describes IBD status with respect to a rather poorly defined founder generation considered to be unrelated. This approach fails to capture the influence of relatedness among founders from the base population. Second, F_{PED} is the expected proportion of the genome that is IBD and does not take into account the stochastic nature of recombination. For example, F_{PED} resulting from the mating of first cousins is always the same (0.0625), while the average F_{ROH} from the same parents would be 0.0625 with a standard deviation of 0.0243 (Carothers *et al.* 2006). This variance increases with each meiosis and it is even possible for offspring of third cousins to be more autozygous than offspring of second cousins (McQuillan *et al.* 2008). Third, several studies confirm that errors in cattle pedigrees are common due to misinterpretation, misidentification and incorrect recording (e.g. Ron *et al.* (1996)). Finally, F_{PED} assumes that the entire genome is selection-neutral and does not account for potential bias resulting from selection (Curik *et al.* 2002). The main aim of this study was to analyze estimates of inbreeding derived from ROH using different ROH lengths and to compare them with those estimated using pedigree data in four cattle breeds: Brown Swiss, Fleckvieh, Norwegian Red and Tyrol Grey. All four breeds have deep and complete pedigrees, but different breed histories. This allowed thorough comparative analysis of inbreeding coefficients based on the traditional pedigree and on ROH. While such validation has been performed in various human populations, it has not been reported in domestic animal breeds undergoing strong artificial selection and exhibiting higher levels of inbreeding and significant gametic disequilibria.

Materials and methods

All analyses were performed on a sample of 1421 bulls comprising Brown Swiss (304), Fleckvieh (502), Norwegian Red (498) and Tyrol Grey (117) cattle breeds. Cattle were born between 1996 and 2006, except for Norwegian Red, which were born between 1996 and 2004, and those animals are representative of their populations. Individual pedigree inbreeding coefficients were calculated from all available pedigree data (F_{PED}) and from

pedigree data restricted to five generations (F_{PED5}). Quality of pedigrees was evaluated via complete generation equivalents (Boichard *et al.* 1997; Solkner *et al.* 1998). The complete generation equivalent is computed as the sum over all known ancestors of the terms computed as the sum of *(1/2)n* where *n* is the number of generations separating the individual to each known ancestor (Maignel *et al.* 1996). Only animals with at least five complete generation equivalents were considered. Calculations were performed using the *ngen.f* and *vanrad.f* routines in PEDIG software (Boichard 2002). Genotyping was performed using Illumina Bovine SNP 50k bead chip technology (Illumina Inc., San Diego, CA). Genotype data of each breed contained different number of SNP markers, and consensus data set including autosomal SNP markers common for all four cattle breeds was com- posed. Markers with a GenCall score lower than 0.2 and markers unassigned to a chromosome were excluded. Using SNP & Variation Suite v7.6.8 Win64 (Golden Helix, Bozeman, MT, USA www.goldenhelix. com) we also excluded animals with more than 5% of missing genotypes, SNP markers with a minor allele frequency (MAF) lower than 1%, SNP markers with more than 5% of missing genotypes and SNP markers assigned to X chromosome. This left a total of 36 273 SNPs for analysis. According to analysis on humans (Kirin *et al.* 2010), we did not prune our data for markers in linkage disequilibrium (LD). Setting minimum ROH length to 1 Mb (representing 50 generations), we avoided occurrence of short ROHs which existence is due to LD. ROH were detected by SNP & Variation Suite v7.6.8 Win64 using the following settings: minimum number of SNPs needed to define a segment as an ROH, 15; number of missing calls allowed, 5; number of heterozygous calls allowed, 0; maximum gap between consecutive homozygous SNPs, 1 Mb; minimum length to define ROH >1, >2, >8 and >16 Mb, representing up to ≈ 50, ≈ 25, ≈ 6 and ≈3 generations from common ancestor, respectively. The inbreeding coefficient based on ROH (F_{ROH}) was defined as the length of the genome present in ROH, divided by the overall length of the genome covered by SNPs (Leutenegger *et al.* 2003b). This overall length was taken to be 2 543 177 bp, based on the consensus map. For each bull, we calculated genomic inbreeding coefficients ($F_{ROH > 1\ Mb}$, $F_{ROH > 2\ Mb}$, $F_{ROH > 8\ Mb}$ and $F_{ROH > 16\ Mb}$) derived from ROH with different lengths (>1, >2, >8 or >16 Mb).

We determined the distribution of ROH across the following five length (Mb) categories: 1–2, 2–4, 4–8, 8–16 and >16. We also created four categories of animals according to their F_{PED} values (very low inbreeding 0–0.0025, inbreeding representing mating of half cousins 0.031–0.034, inbreeding representing mating of cousins 0.06–0.07 and animals with highest F_{PED} in data set going from 0.08 up to 0.17) and visualize their variation in FROH > 8 Mb.

We calculated the genomic inbreeding coefficient (F_{HOM}) based on the difference between observed and expected numbers of homozygous genotypes using SNP & Variation Suite v7.6.8 Win64. This inbreeding coefficient is equivalent to Wright's within-subpopulation fixation index, F_{is}, here used as the actual measure of inbreeding among individuals, because it is measured against others who are in the same subpopulation.

Results obtained from SNP & Variation Suite v7.6.8 Win64 were analyzed using SAS software version 9.2 (SAS 2009).

Results

The average complete generation equivalent was 8.84 for Brown Swiss, 7.30 for Fleckvieh, 9.02 for Norwegian Red and 7.31 for Tyrol Grey, indicating good pedigree depth and completeness for all breeds. Distributions and descriptive statistics for pedigree and ROH-based inbreeding coefficients are presented in Figure 1. The highest inbreeding level was observed in the Brown Swiss breed for all genomic inbreeding coefficients, which ranged from 0.037 for $F_{ROH > 16\ Mb}$ to 0.156 for $F_{ROH > 1\ Mb}$, and the same was true for the pedigree inbreeding coefficients F_{PED} and F_{PED5}. The smallest level of inbreeding for most genomic and both pedigree inbreeding coefficients was observed for Fleckvieh, with values ranging from 0.009 for F_{PED5} to 0.088 for $F_{ROH > 1\ Mb}$. The levels of inbreeding in Norwegian Red and Tyrol Grey breeds were intermediate. Population means of F_{PED} values were between $F_{ROH > 16\ Mb}$ and $F_{ROH > 8\ Mb}$ for all four breeds (Figure 1). Observed levels of $F_{ROH > 1\ Mb}$ and $F_{ROH > 2\ Mb}$ were much higher compared to F_{PED} and F_{PED5}. This is due to their ability to capture both recent and distant IBD segments. Still, there is a possibility, especially for $F_{ROH >1\ Mb}$ that some non-IBD segments were also included. F_{HOM} was 0.002 for Brown Swiss, -0.013 for Fleckvieh, 0.006 for Norwegian Red and 0.001 for Tyrol Grey. Standard deviations of genomic inbreeding coefficients were, on average, 1.3–1.7-fold higher than those obtained from pedigrees. Pearson correlations between F_{PED} and F_{ROH} estimates ranged from r = 0.50 to r = 0.72 (Table 1). Further, F_{ROH} showed a higher correlation with F_{PED} than with F_{PED5} particularly for breeds with deeper pedigrees (Norwegian Red and Brown Swiss). This is especially for segments up to 16 Mb, while for segments >16 Mb, correlations were lowest in these two breeds. High correlations (r > 0.89) were found between F_{HOM} and coefficients based on short ROH ($F_{ROH > 1\ Mb}$ and $F_{ROH > 2\ Mb}$) for all breeds except Fleckvieh (r = 0.86 and 0.82), while correlations between F_{HOM} and F_{PED} and between F_{HOM} and F_{PED5} ranged, respectively, from 0.55 to 0.64 and from 0.48 to 0.65.

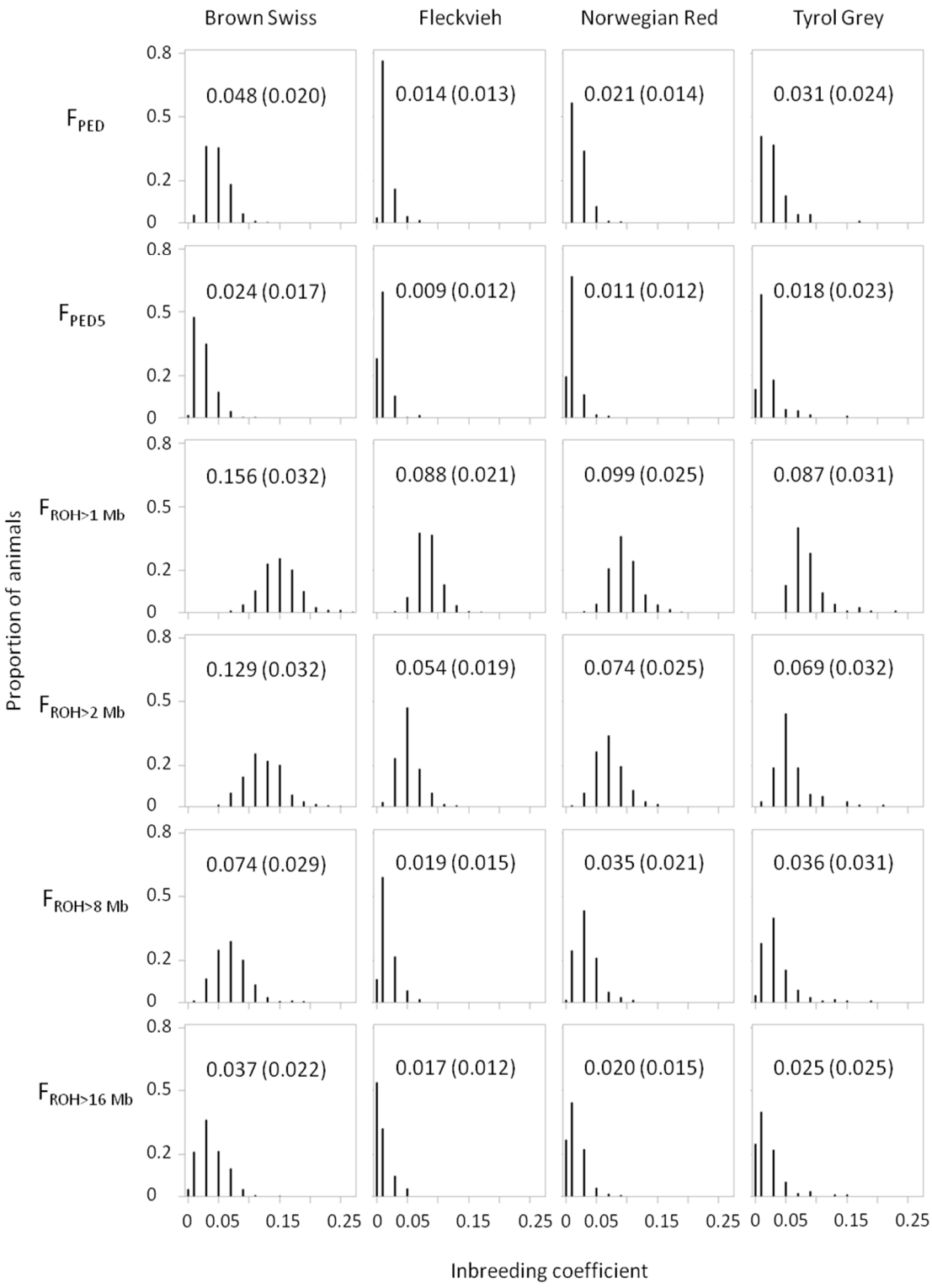


Figure 1. *Distributions and descriptive statistics; mean (standard deviation) for inbreeding coefficients based on pedigree and based on ROH.*

Table 1. *Pearson correlation coefficients between genomic inbreeding coefficients ($F_{ROH>1}$, $F_{ROH > 2\ Mb}$, $F_{ROH > 8\ Mb}$, $F_{ROH>16\ Mb}$ and F_{HOM}) and pedigree inbreeding coefficients (F_{PED} and F_{PED5}) across four cattle breeds*

Breed	Inbreeding coefficient	$F_{ROH > 1\ Mb}$	$F_{ROH > 2\ Mb}$	$F_{ROH > 8\ Mb}$	$F_{ROH>16\ Mb}$	F_{HOM}
Brown	F_{PED}	0.66	0.67	0.60	0.50	0.63
Swiss	F_{PED5}	0.60	0.61	0.58	0.50	0.58
Fleckvieh	F_{PED}	0.66	0.69	0.70	0.64	0.55
	F_{PED5}	0.63	0.67	0.68	0.63	0.52
Norwegian	F_{PED}	0.61	0.61	0.62	0.53	0.58
Red	F_{PED5}	0.50	0.50	0.55	0.50	0.48
Tyrol	F_{PED}	0.71	0.72	0.71	0.70	0.64
Grey	F_{PED5}	0.71	0.71	0.72	0.72	0.65

The Brown Swiss breed had the highest average number of ROH (98.9) and the longest segments (4.01 Mb) of all breeds (Table 2 and Figure 2). The shortest average ROH length (2.36 Mb) was found in Fleckvieh, which had a high average number of ROH 94.5). Figure 2 illustrates the numbers and total lengths of ROH in the four breeds. This illustration gives insight into ROH content of total length of ROH for these breeds. Total length of ROH for Fleckvieh is composed mostly of high number of shorter ROH segments, while for Brown Swiss, the total length of ROH is composed of a lower number of large segments. Norwegian Red here shows similar pattern as Fleckvieh, while in Tyrol Grey, some extreme animals are observed. Those animals had low number of segments (from 65 up to 80) covering 400 to more than 630 Mb of genome. If ROH lengths are divided into categories (Figure 3), Fleckvieh had the highest number and total length of ROH 1–2 Mb long, amounting to over 90 Mb of genome. For all other length categories, Brown Swiss had the longest total length of ROH.

Table 2. *Descriptive statistics for the number of ROH and total length (Mb) of genome in ROH for four cattle breeds*

Parameter	Brown Swiss	Fleckvieh	Norwegian red	Tyrol Grey
Number of ROH				
mean	98.9	94.5	80.8	72.3
std	10.2	13.1	10.3	10.3
min	67.0	42.0	46.0	52.0
max	121.0	135.0	108.0	103.0
Total length of genome in ROH >1 Mb				
mean	396.8	223.1	253.4	221.0
std	82.9	52.5	63.4	78.9
min	177.7	71.6	68.7	117.1
max	690.3	453.0	481.3	634.2

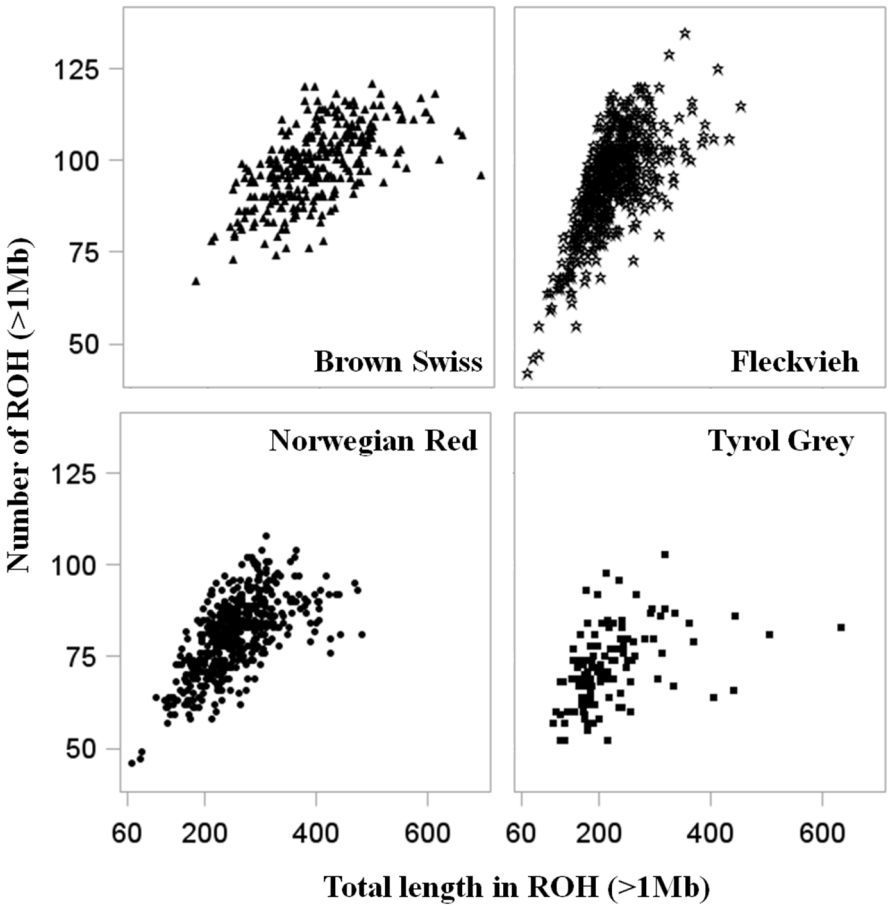


***Figure 2**. Relationship between the number of ROH >1 Mb and the total length (Mb) of genome in such ROH for individuals from each breed.*

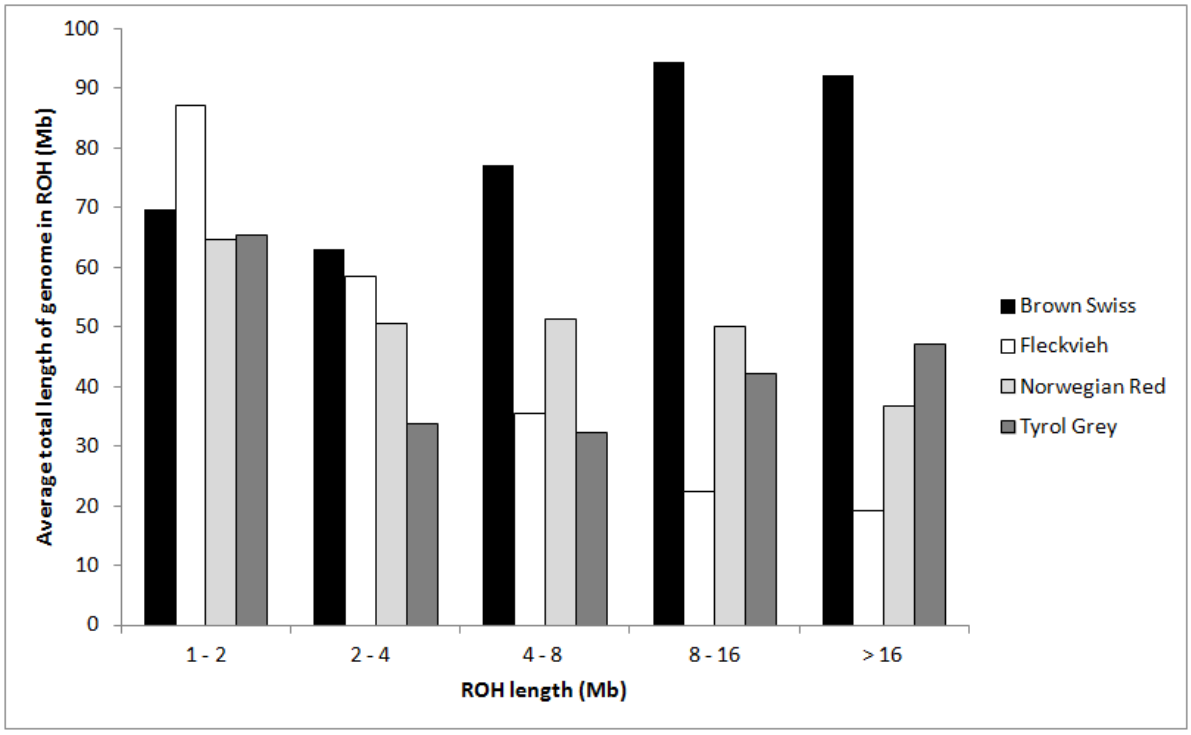


***Figure 3.** Interbreed differences in average total length of genome in ROH of different length categories, by ROH length. Bars indicate mean values together with ± 2 standard errors; thus, non-overlapping bars are significantly different from each other.*

Comparison of animals having similar F_{PED} values with $F_{ROH > 8\ Mb}$, which are most likely to be true IBD segments and represent inbreeding occurred up to ≈6 generations ago (Figure 4), shows spread of $F_{ROH > 8\ Mb}$ values present in all four breeds. With increase in F_{PED} spreading of $F_{ROH > 8\ Mb}$ values on y-axis for very similar values of F_{PED} are greater.

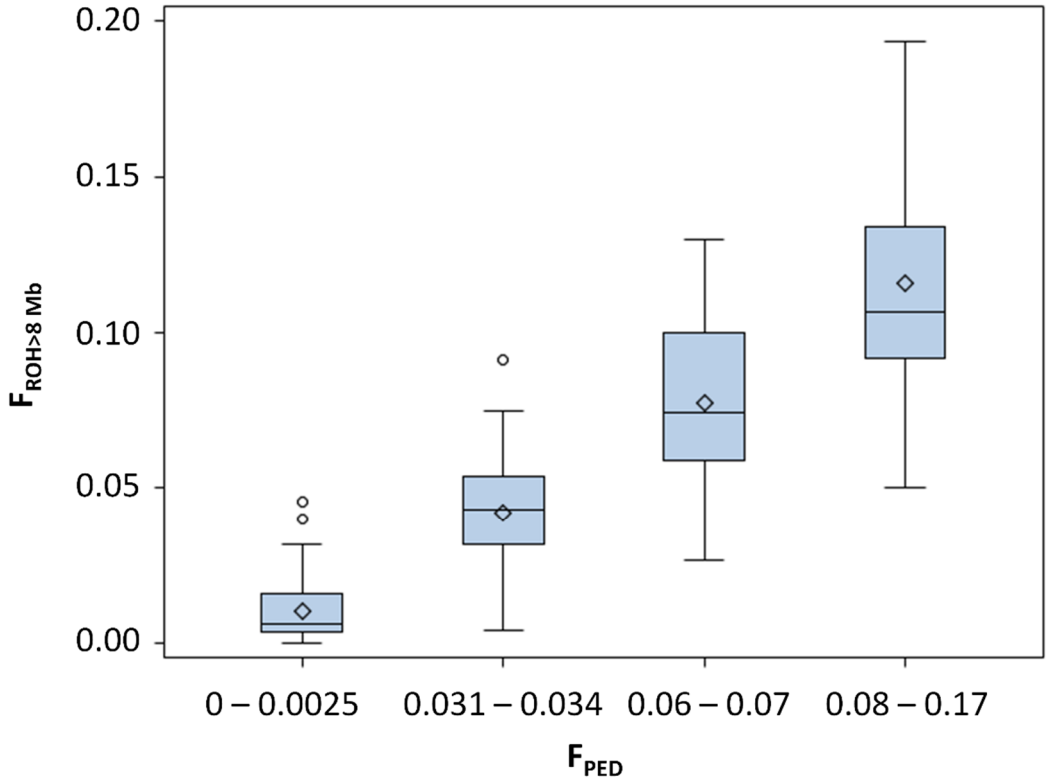


Figure 4. *Box and whisker plots representing the variability in $F_{ROH>8\ Mb}$ for four groups of animals with similar F_{PED} values, representing unrelated, offspring of half cousins, offspring of cousins and the group of animals with highest levels of F_{PED} in analyzed populations.*

Discussion

We analyzed animals from four cattle breeds with different inbreeding backgrounds to derive the levels of autozygosity based on ROH (F_{ROH}). We also correlated F_{PED} and F_{ROH} values and obtained moderate to relatively high correlations, indicating that FROH provides a good indication of individual levels of inbreeding. Trend was that, within breed, $F_{ROH > 1\ Mb}$, $F_{ROH > 2\ Mb}$, $F_{ROH > 8\ Mb}$ gave similar correlations with F_{PED} (Table 1). In breeds with deeper pedigrees (Norwegian Red and Brown Swiss), we observed a drop in correlations for $F_{ROH > 16\ Mb}$. Very long runs represent recent inbreeding (16 Mb segments are expected mean after ≈3 generations), so part of autozygosity that is due to more distant common ancestors is not covered with them. Correlations between F_{ROH} from different lengths are linked with depth of pedigree. Overall correlations of F_{ROH} estimates based on ROH of different lengths with F_{PED} or F_{PED5} did not differ substantially. This is consistent with our previous study on Fleckvieh (Ferenčaković *et al.* 2011). VanRaden (2008) correlated estimates of inbreeding levels based on SNP variance and estimates based on pedigrees and found the former to be higher in Holstein Friesian and Jersey breeds. Applying methods of VanRaden (2008) and Yang *et al.* (2010), Sölkner *et al.* (2010) obtained much lower correlations for Fleckvieh, while those of FROH were similar to those presented here. A study of inhabitants of Orkney Islands reported a correlation of r = 0.86 between inbreeding estimates based on the proportion of ROH longer than 1.5 Mb and estimates from pedigrees (McQuillan *et al.* 2008). This correlation is considerably higher than that of F_{PED} or F_{PED5} with our estimates based on ROH in similar length categories ($F_{ROH > 1\ Mb}$, $F_{ROH > 2\ Mb}$). The strongest correlation was obtained for $F_{ROH > 1\ Mb}$ in Tyrol Grey (F_{PED} = 0.71, F_{PED5} = 0.71), while the lowest was in Norwegian Red (F_{PED} = 0.61, F_{PED5} = 0.50). These different results may be attributed to differences in population structure.

F_{PED} reflects recent inbreeding, while inbreeding coefficients based on ROH can capture both recent and distant inbreeding. We consider ROH 2–4 Mb long (25–12.5 generations from common ancestor) to correspond mostly to IBD segments from the past that we usually will not be able to capture with available pedigree information (CGE from 7.3 to 9.0), although they may also contain some ROH that are IBS without being IBD. In contrast, ROH >8 Mb long are likely to be autozygous segments of recent origin and are extremely unlikely to be non-IBD. F_{PED} also does not account for stochastic nature of recombination, while F_{ROH} is sensitive to it (Figure 4). The spread of values from $F_{ROH > 8\ Mb}$ for groups of animals with similar F_{PED} clearly shows advantage of using F_{ROH}.

Studies of outbred human populations have reported co-occurrence of ROH in regions

with extended linkage disequilibrium and low recombination rates (Gibson *et al.* 2006; Curtis *et al.* 2008), so common extended haplotypes may partly contribute to high F_{ROH} estimates based on shorter ROH cut-off. Kirin *et al.* (2010) used minimum length of ROH of 500 kb to avoid very short ROH that can occur due to LD, while from the same reason, we used minimum of 1 Mb because it is known that cattle have longer range LD. We would also like to point out that F_{ROH} estimates can also depend on the program used to calculate them, because different software packages for analyzing inbreeding using ROH do not give identical results (Howrigan *et al.* 2011). Parameter settings, for example, the minimum number of SNPs in an ROH or tolerance of a small number of heterozygous SNPs, may also considerably influence F_{ROH} estimates.

F_{ROH} estimates confirmed that Brown Swiss bulls show comparatively high levels of inbreeding. While most of these bulls are of Austrian origin, much of their pedigrees can be traced back to the US Brown Swiss population, from which semen was imported into Europe as early as the early the 1970s (Solkner *et al.* 1998). The US Brown Swiss population is genetically small, mostly derived from 21 male and 169 female animals imported into the United States between 1869 and 1906 (Yoder & Lush 1937). Unfortunately, information about large sections of the US pedigrees tracing ancestry further back was unavailable. The relatively high levels of pedigree inbreeding and of long ROH observed here (Table 2, Figure 2) are consistent with the import of semen from a small number of US bulls and subsequent interbreeding. On the basis of ROH > 1 Mb long, the average level of autozygosity of the Brown Swiss population was 0.151. This is much higher than the level of 0.048 estimated from available pedigree information, but consistent with the origin of the breed in a small number of animals imported into the United States 100–150 years ago.

Fleckvieh is a breed with a larger effective population size, partly due to the fact that breeding was until recently carried out by independent regional associations (Solkner *et al.* 1998). Using microsatellite markers, Medugorac *et al.* (2009) found an effective population size of 410 for this breed. This is consistent with the small proportion of autozygous genome estimated here. The large number of short segments 1–2 Mb long (Figure 3) indicates that the Fleckvieh breed originated as a relatively homogeneous population with a small effective population size.

Tyrol Grey is a local breed with a small population size of <5000 registered cows. The breeding program involves a bull-testing scheme with artificial insemination and natural mating. Solkner *et al.* (1998) and subsequent analyses by (Fürst & Fürst-Waltl 2009) indicated substantially more inbreeding in Tyrol Grey than in Fleckvieh for concurrent

reference populations. Medugorac *et al.* (2009) also analyzed Tyrol Grey and found the effective population size to be 200. ROH analysis in the present study confirms that the level of inbreeding is relatively low considering the very small population size.

Norwegian Red is known for its high heterogeneity as a result of historic admixture (Sodeland *et al.* 2011). In 2008, the effective population size was 173 (Garmo 2010). Large effective population size has been maintained through the control of inbreeding and gene flow by importing sires from other Nordic countries.

We conclude that levels of autozygosity derived from ROH provide a very good indication of individual inbreeding levels, as well as additional information about inbreeding due to remote ancestors. The observational approach of ROH, in contrast to the probabilistic approach of pedigree analysis, which does not take stochastic variations into account, most likely gives more precise information about levels of autozygosity. Higher standard deviations of inbreeding coefficients derived from ROH then those derived from pedigree are suggesting more power for estimating inbreeding depression. Performing analyses with ROH of different lengths allows estimation of the distance of the current population from the base population.

Acknowledgements

We would specially like to thank Golden Helix team for providing us free license for SNP & Variation Suite v7.6.8 Win64, which allowed us easy and efficient work with genotype data. We are very grateful to Zuchtdata EDV-Dienstleistungen GmbH and Geno SA for provision of genotype and pedigree data. Maja Ferenčaković is supported by an 'Ernst Mach grant world-wide' financed by the Austrian Federal Ministry of Education and Science. Edin Hamzic worked on this study while enrolled in the Erasmus Mundus programme 'European Master in animal Breeding and Genetics', financed by the European Commission. This research was partially funded by the Croatian Ministry of Science, Education and Sports (project 178-1780460-0546, 'Strategies for elimination of genetic defects in selected populations'). Finally, we gratefully acknowledge the editor and anonymous reviewers. Their expertise contributed to improve quality of the manuscript.

References

Boichard D. (2002) PEDIG: a fortran package for pedigree analysis suited for large populations. *Proceedings of the 7th World Congress of Genetics Applied to Livestock Production: 19-23 August 2002; Montpellier*.

Boichard D., Maignel L. & Verrier E. (1997) The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selelection Evolution* **29**, 5 - 23.

Broman K. & Weber J. (1999) Long homozygous chromosomal segments in reference families from the centre d'etude du polymorphisme humain. *American Journal of Human Genetics* **65**, 1493 - 500.

Carothers A., Rudan I. & Kolcic I. (2006) Estimating human inbreeding coefficients: Comparison of genealogical and marker heterozygosity approaches. *Annals of Human Genetics* **70**, 666 - 76.

Curik I., Solkner J. & Stipic N. (2002) Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. *Journal of Animal Breeding & Genetics* **119**, 101-15.

Curtis D., Vine A. & Knight J. (2008) Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. *Annals of Human Genetics* **72**, 261-78.

Ferenčaković M., Hamzic E., Gredler B., Curik I. & Solkner J. (2011) Runs of homozygosity reveal genome-wideautozygosity in the Austrian fleckvieh cattle. *Agriculturae Conspectus Scientificus* **76**, 325 - 8.

Fürst C. & Fürst-Waltl B. (2009) Inzucht & Co–Aktuelle Auswertungen zur genetischen Vielfalt. In: *Kunterbunte rinderwelt Genetische vielfalt-Erhaltung und züchtung, Seminar des Ausschusses für Genetik der ZAR*.

Garmo R.T. (2010) The Significance of Sustainable Breeding and Management Programs on Reproductive Performance in Norwegian Red Cows. Norwegian School of Veterinary Science.

Gibson J., Morton N. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789 - 95.

Howrigan D., Simonson M. & Keller M. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* **12**, 460.

Keller M., Visscher P. & Goddard M. (2011) Quantification of Inbreeding Due to Distant Ancestors and its Detection Using Dense SNP Data. *Genetics* **189**, 237 - 49.

Kirin M., McQuillan R., Franklin C., Campbell H., McKeigue P. & Wilson J. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996.

Lencz T., Lambert C., DeRosse P., Burdick K., Morgan T., Kane J., Kucherlapati R. & Malhotra A. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19942 - 7.

Leutenegger A., Prum B., Genin E., Verny C., Lemainque A., Clerget-Darpoux F. & Thompson E. (2003) Estimation of the inbreeding coefficient through use of genomic data. *American Journal of Human Genetics* **73**, 516 - 23.

Maignel L., Boichard D. & Verrier E. (1996) Genetic variability of French dairy breeds estimated from pedigree information. *INTERBULL bulletin*, 49.

McQuillan R., Leutenegger A., Abdel-Rahman R., Franklin C., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A., Farrington S., Rudan P., Hayward C., Vitart V., Rudan I., Wild S., Dunlop M., Wright A., Campbell H. & Wilson J. (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics* **83**, 359 - 72.

Medugorac I., Medugorac A., Russ I., VEIT□KENSCH C.E., Taberlet P., Luntz B., Mix H.M. & Foerster M. (2009) Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Molecular ecology* **18**, 3394-410.

Nothnagel M., Lu T.T., Kayser M. & Krawczak M. (2010) Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human Molecular Genetics* **19**, 2927-35.

Polasek O., Hayward C., Bellenguez C., Vitart V., Kolcic I., McQuillan R., Saftic V., Gyllensten U., Wilson J., Rudan I., Wright A., Campbell H. & Leutenegger A.-L. (2010) Comparative assessment of methods for estimating individual genome-wide homozygosity-by-descent from human genomic data. *BMC Genomics* **11**, 139.

Ron M., Blanc Y., Band M., Ezra E. & Weller J. (1996) Misidentification rate in the Israeli dairy cattle population and its implications for genetic improvement. *Journal of Dairy Science* **79**, 676-81.

SAS (2009) SAS/STAT® User's Guide Version 9.2. SAS Institute Inc.Cary, NC, USA.

Sodeland M., Kent M., Hayes B.J., Grove H. & Lien S. (2011) Recent and historical recombination in the admixed Norwegian Red cattle breed. *BMC Genomics* **12**, 33.

Sölkner, J., Ferencakovic, M., Gredler, B. & Curik, I. (2010) Genomic metrics of individual autozygosity, applied to a cattle population. In: *Book of Abstracts of the 61st Annual Meeting of the European Association of Animal Production*, / *Wageningen Academic Publishers (ed.). Wageningen Academic Publishers*. 306-307.

Solkner J., Filipcic L. & Hampshire N. (1998) Genetic variability of populations and similarity of subpopulations in Austria cattle breeds determined by analysis of pedigrees. *Animal Science* **67**, 249 - 56.

VanRaden P. (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**, 4414 - 23.

Yang J., Benyamin B., McEvoy B., Gordon S., Henders A., Nyholt D., Madden P., Heath A., Martin N. & Montgomery G. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565 - 9.

Yoder D.M., Lush J.L. (1937) A genetic history of the Brown Swiss cattle in the United States. *Journal of Heredity* **28**, 154–160.

CHAPTER 4 ESTIMATING AUTOZYGOSITY FROM HIGH-THROUGHPUT INFORMATION: EFFECTS OF SNP DENSITY AND GENOTYPING ERRORS

Maja Ferenčaković[1], Johann Sölkner[2] and Ino Curik[1]

[1]Department of Animal Science, Faculty of Agriculture, University of Zagreb, Svetošimunska cesta 25,10000 Zagreb, Croatia.

[2]Division of Livestock Sciences, Department of Sustainable Agricultural Systems,University of Natural Resources and Life Sciences Vienna, Gregor Mendel Str.33, A-1180 Vienna, Austria

Abstract

Background: Runs of homozygosity are long, uninterrupted stretches of homozygous genotypes that enable reliable estimation of levels of inbreeding (i.e., autozygosity) based on high-throughput, chip-based single nucleotide polymorphism (SNP) genotypes. While the theoretical definition of runs of homozygosity is straightforward, their empirical identification depends on the type of SNP chip used to obtain the data and on a number of factors, including the number of heterozygous calls allowed to account for genotyping errors. We analyzed how SNP chip density and genotyping errors affect estimates of autozygosity based on runs of homozygosity in three cattle populations, using genotype data from an SNP chip with 777 972 SNPs and a 50 k chip.

Results: Data from the 50 k chip led to overestimation of the number of runs of homozygosity that are shorter than 4 Mb, since the analysis could not identify heterozygous SNPs that were present on the denser chip. Conversely, data from the denser chip led to underestimation of the number of runs of homozygosity that were longer than 8 Mb, unless the presence of a small number of heterozygous SNP genotypes was allowed within a run of homozygosity.

Conclusions: We have shown that SNP chip density and genotyping errors introduce patterns of bias in the estimation of autozygosity based on runs of homozygosity. SNP chips with 50 000 to 60 000 markers are frequently available for livestock species and their information leads to a conservative prediction of autozygosity from runs of homozygosity longer than 4 Mb. Not allowing heterozygous SNP genotypes to be present in a homozygosity run, as has been advocated for human populations, is not adequate for livestock populations because they have much higher levels of autozygosity and therefore longer runs of homozygosity. When

allowing a small number of heterozygous calls, current software does not differentiate between situations where these calls are adjacent and therefore indicative of an actual break of the run versus those where they are scattered across the length of the homozygous segment. Simple graphical tests that are used in this paper are a current, yet tedious solution.

Background

Runs of homozygosity (ROH) are continuous stretches of homozygous genotypes without heterozygosity in the diploid state. Although ROH can arise by different mechanisms,(Gibson *et al.* 2006) the primary cause is believed to be inbreeding (Broman & Weber 1999a) Long ROH are most likely the result of recent inbreeding, where recombination events do not shorten identical haplotypes inherited from the common ancestor. Short ROH, in contrast, suggest more ancient inbreeding. The ability of ROH to reveal information about ancient and re- cent genetic events makes them useful tools to analyze population history (Kirin *et al.* 2010) inbreeding levels (Keller *et al.* 2011) and effects of inbreeding on complex traits and congenital disorders (Lencz *et al.* 2007)

While ROH from high-throughput genotyping analyses have been studied extensively in humans, such analyses are rare in cattle and other livestock species (Ferenčaković *et al.* 2011; Purfield *et al.* 2012; Bjelland *et al.* 2013; Ferenčaković *et al.* 2013a; Silió *et al.* 2013). The lack of standards for ROH definition and identification may intro- duce bias in ROH-based estimates of autozygosity. Howrigan *et al.* (2011) found that the numbers and sizes of ROH that are identified in genotyping data can strongly depend on certain parameters and thresholds imposed during sequence analysis. In addition, pruning single nucleotide polymorphisms (SNPs) that show low minor allele frequency (MAF), that deviate from Hardy-Weinberg equilibrium (HWE), or that show high linkage disequilibrium (LD), can affect the results (Wigginton *et al.* 2005; Albrechtsen *et al.* 2010)

The density of the SNP chip used to generate the data for ROH identification is another factor that strongly affects autozygosity estimates. Purfield *et al.* (2012) compared estimates obtained using the two SNP chips most frequently used in cattle: the Illumina BovineSNP50 Genotyping BeadChip with 54 001 SNPs (50 k) and the Illumina BovineHD Genotyping BeadChip with 777 972 SNPs (HD). They concluded that the 50 k chip is appropriate only for identifying ROH longer than 5 Mb. Indeed, analyses based on lower-density chips can fail to detect heterozygous SNP genotypes that are present in observed ROH.

The frequency of SNP genotyping errors is another factor that can affect ROH-based estimates of autozygosity. Since this frequency usually varies between 0.2% and 1.0% (Rabbee & Speed 2006; Howrigan *et al.* 2011) it may affect identification of very long ROH that contain numerous SNPs. In fact, any genotyping error, whether homozygote to heterozygote or vice versa, can affect the determination of ROH. A potential solution is to allow a certain number of SNPs to be heterozygous (Gibson *et al.* 2006) but whether this compromises the reliability of ROH analyses has not been systematically analyzed.

The aim of this study was to analyze the identification of ROH of different length categories and the estimation of genomic inbreeding coefficients based on ROH in three cattle breeds (Brown Swiss, Pinzgauer, Tyrol Grey). Our study focused on the effects of chip density (777 972 versus 54 001 SNPs) and genotyping errors. Results demonstrate, both graphically and statistically, that density of SNP chips affects ROH detection and subsequent estimation of inbreeding levels. The optimal number of heterozygous SNPs allowed during ROH analysis was found to depend on chip density and ROH length.

Methods

Genotype data and quality control

The semen samples of the animals included in this study used for DNA extraction and genotyping were obtained from AI centers through their routine practice in the framework of breeding programs. Therefore, no ethical approval was required for sampling of biological material. DNA samples were obtained from 277 bulls of three breeds: Brown Swiss, 46; Pinzgauer, 118; and Tyrol Grey, 113. Mean pedigree-based inbreeding coefficients (and ranges) were as follows: Brown Swiss, 0.033 (0.009- 0.096); Pinzgauer, 0.019 (0–0.088); and Tyrol Grey, 0.022(0–0.169). The mean complete generation equivalent (see e.g., (Solkner *et al.* 1998) was highest for Brown Swiss (7.32 generations) and lowest for Pinzgauer (5.32 generations). DNA samples were genotyped using the BovineHD Bead Chip (Illumina Inc., San Diego, CA), which contains 777 972 SNPs; this data set is referred to hereafter as the high- density (HD) panel. For comparison, we extracted and retained SNPs from this panel that were common to both the HD panel and the bovine SNP50 Beadchip v1 (Illumina Inc., San Diego, CA), which contains 54 001 SNPs and which will be referred to in the remainder as the 50 k panel.

Data extraction and quality control were performed separately for each breed. We excluded all SNPs that had not been assigned to a chromosome or that had been assigned to

chromosomes X or Y or to the mitochondrial genome. We also excluded SNPs for which more than 10% of genotypes were missing and SNPs with an Illumina GenCall score ≤ 0.7 or an Illumina GenTrain score ≤ 0.4. Two Tyrol Grey bulls were excluded from further analysis because more than 5% of their genotypes were missing. In doing this, our objective was to exclude poorly performing loci and minimize risk of genotyping errors. After quality control, the numbers of SNPs in the HD and 50 k panels were as follows for each breed: Brown Swiss, 615 618 and 38 710; Pinzgauer, 606 120 and 38 198; and Tyrol Grey, 684 172 and 42 997.

Although it is customary in genome-wide association studies and ROH analyses to exclude SNPs with low MAF or high LD with neighboring SNPs or that deviate from HWE, we did not apply such exclusion criteria in our study. Instead we relied on Illumina quality scores (GenCall, GenTrain) to reduce genotyping problems. We also defined the minimum ROH length as 1 Mb to exclude short, common ROH arising from LD (Kirin *et al.* 2010; Purfield *et al.* 2012) ROH calling options ROH were identified in every individual using the SNP& Variation Suite (v7.6.8 Win64; Golden Helix, Bozeman, MT, USA www.goldenhelix.com). This algorithm is designed to find stretches of consecutive homozygous SNPs; it works continuously across an entire chromosome, examining every possible run that matches the user-specified parameters. We chose this software instead of the PLINK ROH algorithm (Purcell *et al.* 2007) which uses a sliding window that may introduce artificial runs and fail to identify segments shorter than the window.

ROH exceeding the allowed number of heterozygotes or missing SNPs were checked automatically to deter- mine whether they should be removed based on their length, SNP density, and user-specified parameters. ROH were called if 15 or more consecutive homozygous SNPs (Powell *et al.* 2010) were present at a density of at least 1 SNP every 100 kb, with gaps of no more than 1000 kb between them. These density and gap thresholds were applied to SNPs in both the HD and 50 k panels to ensure comparability of the results.

Five categories of ROH length (in Mb) were defined: [1,2], (2, 4], (4, 8], (8, 16], and >16. The number of heterozygous SNPs allowed was set to different values for different length categories. First, we called ROH without allowing any heterozygous calls, and we obtained the average numbers of SNPs in each length category (Table 1). We then assumed a genotype error rate of 0.25%, recalculated the numbers of heterozygote calls allowed, and rounded the number of heterozygous SNPs allowed to the nearest whole number. This approach led to the following numbers of heterozygous SNPs allowed for each length category (in Mb) in the HD panel: [1,2], one heterozygeous SNP; (2, 4], two heterozygous SNPs; (4, 8], four heterozygous SNPs; (8, 16], eight heterozygous SNPs; and >16, 16

heterozygous SNPs (Table 2, class C). In the case of the 50 k panel, we allowed one heterozygous SNP for length category >16, and no heterozygous SNPs for the other categories (Table 2, class A).

Like the number of heterozygous SNPs, we set the number of missing SNPs allowed to different values for different length categories. First, we determined ROH allowing any number of missing SNPs and then used the results to set limits. This approach led to the following limits for missing SNPs for each ROH length category (in Mb) in the HD and 50 k panels, respectively: [1,2], four or no missing SNPs; (2, 4], eight or no missing SNPs; (4, 8], 16 or one missing SNP; (8, 16], 32 or two missing SNPs; and > 16, 64 or four missing SNPs.

Calculating inbreeding coefficients from runs of homozygosity (F_{ROH})

Statistically F_{ROH} is defined as the length of the autosomal genome present in ROH, divided by the overall length of the autosomal genome covered by the SNPs (Leutenegger *et al.* 2003b) For each bull, we calculated F_{ROH}>1 Mb, F_{ROH}>2 Mb, F_{ROH}>4 Mb, F_{ROH}>8 Mb and F_{ROH}>16 Mb based on ROH of different minimum lengths (>1, >2, >4, >8 or >16 Mb). F_{ROH} was calculated for different minimum ROH lengths because lengths of autozygous segments in a genome are predicted to show an exponential distribution, with a mean length equal to 1/2 g Morgan, where g is the number of generations since the common ancestor (e.g. (Howrigan *et al.* 2011) If the genome of an individual contains segments as short as 1 Mb, we can conclude that the individual's autozygosity originated from common ancestors up to 50 generations in the past. Based on the F_{ROH} values across all ROH lengths, detected with both 50 k and HD panel, correlations with pedigree inbreeding coefficients were calculated in order to investigate their relationships.

Table 1. *Summary statistics for the numbers of SNPs in ROH of different length categories*

Panel	Statistic	ROH length category (in Mb)				
		[1, 2]	(2, 4]	(4, 8]	(8, 16]	>16
50 k panel	mean	21.69	45.13	90.95	178.77	399.39
	std	5.68	12.90	23.31	43.17	156.13
	min	15.00	21.00	44.00	92.00	210.00
	max	49.00	98.00	195.00	354.00	1360.00
HD panel	mean	291.29	694.25	1432.46	2856.02	6385.90
	std	138.56	207.76	361.29	633.38	2377.02
	min	15.00	31.00	90.00	1834.00	3617.00
	max	808.00	1353.00	2668.00	4825.00	20325.00

Summary statistics were calculated from ROH identified when no heterozygous calls are allowed.

Table 2. *Definition of classes according to the maximum number of heterozygous SNPs allowed (values in columns) within ROH length categories*

Panel	Class	ROH length category (in Mb)				
		[1, 2]	(2, 4]	(4, 8]	(8, 16]	>16
50 k panel	A	0	0	0	0	1
	B	.	.	.	.	0
HD panel	C	1	2	4	8	16
	D	0	1	2	4	8
	E	.	0	1	2	4
	F	.	.	0	1	2
	G	.	.	.	0	1
	H	.	.	.	.	0

Dots indicate that the value of 0 (no heterozygous allowed) for the given length category was reached in a previous class of the same panel and information is not repeated.

Identifying significant differences in autozygosity estimates based on the number of heterozygous calls allowed

Mean values of F_{ROH} were calculated within classes (scenarios) in which different numbers of heterozygous SNPs were allowed in each ROH length category. Eight classes (A to H) were defined, two (A and B) for the 50 k panel and six (C-H) for the HD panel (Table 2). Numbers of heterozygous SNPs allowed within a class were based on the average numbers of SNPs in a length category and an assumed genotyping error rate of 0.25% for classes A

and C. The other classes were formed by successively halving the allowed number of heterozygotes and only considering longer segments (see Table 2).

Mean F_{ROH} values obtained when allowing different numbers of heterozygous SNPs were compared within the same length category using paired t-tests. In addition, *F*ROH values were compared between the 50 k and HD panels. The SAS 9.3 (Institute 2011) procedure TTEST with the PAIRED statement was used to generate p values. The step-down Bonferroni method of Holm (Holm 1979) using the MULTTEST procedure and the HOLM statement was used to adjust the p values of the 186 comparisons.

Results and discussion

Impact of SNP chip density on ROH identification

Across all three cattle breeds, we identified 19 392 ROH segments using the 50 k panel and 14 148 ROH segments using the HD panel (Table 3). For all three breeds, analysis with the 50 k panel identified more ROH >1 Mb than the HD panel. The two panels gave similar numbers of ROH >4 Mb. As ROH length increased, the HD panel yielded a higher number of ROH than the 50 k panel (Figure 1). The 50 k panel revealed an abundance of small segments and overestimated the numbers of segments 1–4 Mb long, suggesting that it is not sensitive enough for the precise determination of small segments.

The 50 k panel did, however, prove suitable for detect-ng segments longer than 4 Mb. This finding is consistent with that of Purfield *et al.* (2012) who concluded that the 50 k panel recognizes only segments longer than 5 Mb as well as the HD panel does.

The 50 k and HD panels gave noticeably different distributions and mean values of ROH length within each length category (Figure 2). Differences were greatest for the (Broman & Weber 1999a; Gibson *et al.* 2006) length category, and then gradually disappeared as ROH length increased. These findings provide further evidence that data from the 50 k panel lead to imprecise determination of short ROH and overestimation of $F_{ROH.}$

Impact of genotyping errors on autozygosity estimates

To our knowledge, a simulation study by Howrigan *et al.* (2011) is the only source of recommendations on the number of heterozygous calls allowed in ROH. They suggested allowing no heterozygous calls. However, since genotyping errors in SNP chip data do occur, it seems more reasonable to allow some heterozygous calls, particularly for ROH >8 Mb on dense SNP chips. These long segments are much more frequent in cattle populations than in

human populations, even for population isolates (e.g. (Esko *et al.* 2013)). We determined the numbers of SNPs in ROH of specific lengths and assumed a 0.25% rate of genotyping errors in order to define the number of heterozygous genotypes allowed separately for each ROH length category. Then, we determined mean F_{ROH} values for the classes defined in Table 2 for different allowed numbers of heterozygous calls. Paired t-tests were conducted within the eight classes (A-H) within the same length category and within each cattle breed (Table 4). The 50 k and HD panel data gave significantly different mean F_{ROH}>1 Mb values in Pinzgauer and Tyrol Grey cattle, and significantly different mean F_{ROH}>4 Mb and F_{ROH}>8 Mb values in the Brown Swiss and Pinzgauer breeds. For all three breeds, mean F_{ROH} >16 Mb based on the 50 k panel differed significantly depending on whether one (class A) or no (class B) heterozygous calls were allowed. These differences had important effects on estimates of inbreeding levels. For each breed, inbreeding levels based on F_{ROH} >16 Mb based on the HD panel differed by approximately 1.7-fold, depending on whether 16 or no heterozygous calls were allowed (Table 4). In fact, inbreeding coefficients derived from ROH > 16 Mb with no allowance for heterozygous calls were lower than inbreeding coefficients estimated from pedigrees. These findings suggest that for such long ROH, which can have more than 5000 to 6000 SNPs, some heterozygous calls must be allowed due to the possibility of genotyping errors.

At the same time, the number of allowable heterozygous calls should be limited. On the one hand, SNP data from chromosome 20 in the 46 Brown Swiss cattle (Figure 3) shows clearly that single, potentially miscalled heterozygous SNPs would interrupt ROH segments if such SNPs were not allowed. On the other hand, the figure also shows that allowing certain minimum numbers of heterozygous SNPs leads to inaccurate ROH calling at the ends of ROH. Such inaccurate calling is also likely to be a problem in individual ROH, since we sometimes observed multiple heterozygous SNPs close together within a ROH, not only when using the SNP & Variation software suite but also when using other programs (PLINK; (Purcell *et al.* 2007) cgaTOH; (Zhang *et al.* 2013) data not shown). In any event, ROH identification software should be improved so that instances of multiple heterozygous SNPs very close to one another should automatically lead the program to define separate ROH. Until such an improvement is made, we recommend careful visual analysis of ROH segment structure in order to exclude spurious ROH.

Table 3. *Summary statistics per breed for the numbers of ROH of different minimum lengths*

Breed	ROH length (Mb)	Panel	Mean	std	min	max
Brown Swiss	**>1**	50 k	94.76	14.55	66.00	136.00
		HD	82.02	15.48	60.00	150.00
	>2	50 k	47.98	11.66	27.00	81.00
		HD	46.59	9.88	31.00	81.00
	>4	50 k	24.85	6.64	11.00	42.00
		HD	25.93	6.63	13.00	40.00
	>8	50 k	11.50	4.54	3.00	25.00
		HD	12.48	4.66	2.00	23.00
	16	50 k	3.96	1.89	0.00	8.00
		HD	4.33	2.01	0.00	9.00
Pinzgauer	**>1**	50 k	59.96	9.91	33.00	84.00
		HD	43.26	9.97	19.00	95.00
	>2	50 k	19.44	6.01	5.00	34.00
		HD	19.08	6.66	5.00	46.00
	>4	50 k	8.85	3.93	2.00	20.00
		HD	9.47	4.48	1.00	22.00
	>8	50 k	4.09	2.55	0.00	11.00
		HD	4.41	2.67	0.00	12.00
	16	50 k	1.36	1.37	0.00	6.00
		HD	1.36	1.39	0.00	6.00
Tyrol Grey	**>1**	50 k	70.86	9.51	52.00	102.00
		HD	44.94	12.14	24.00	100.00
	>2	50 k	21.08	7.94	4.00	55.00
		HD	18.72	7.24	6.00	50.00
	>4	50 k	9.99	5.19	1.00	33.00
		HD	9.60	5.00	1.00	31.00
	>8	50 k	4.43	3.34	0.00	20.00
		HD	4.65	3.29	0.00	21.00
	16	50 k	1.64	1.90	0.00	12.00
		HD	1.70	1.87	0.00	12.00

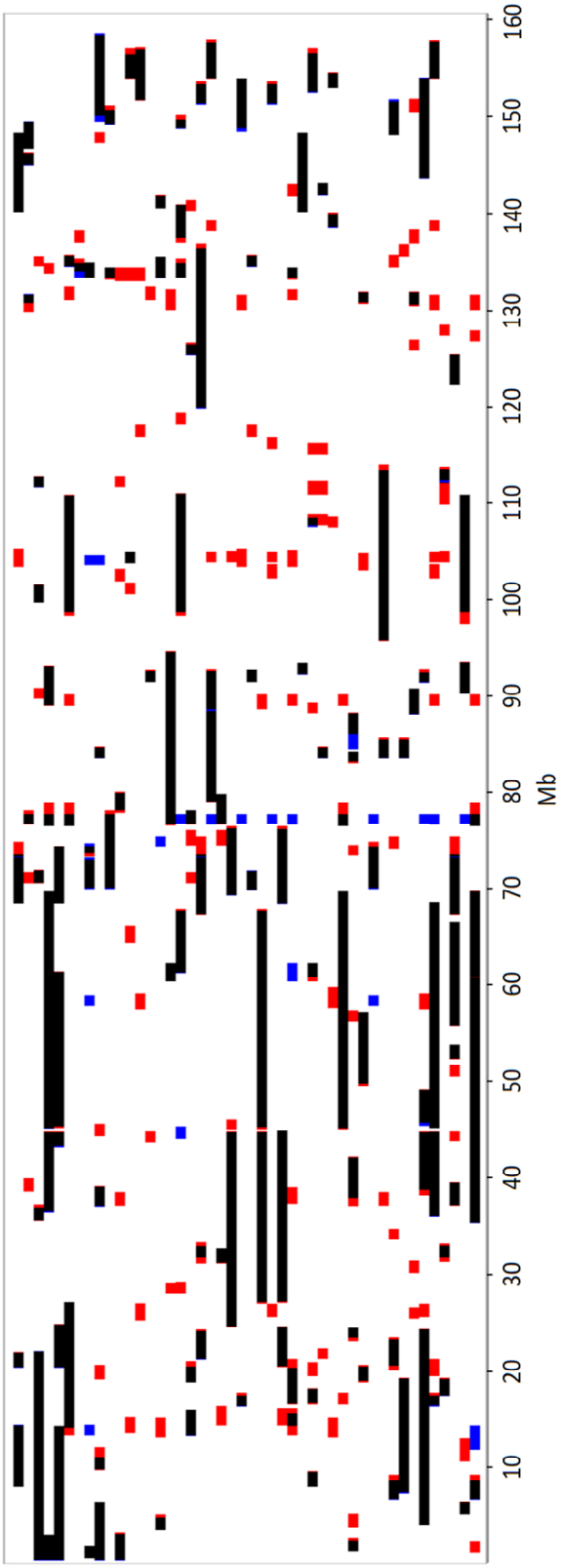


Figure 2 *Overlay of ROH identified on chromosome 1 in Brown Swiss animals. ROH were identified using 50 k and HD panel data and then overlaid; each row represents one animal, and different colors were used to indicate whether ROH segments were identified using both the 50 k and HD panel (black), only the 50 k panel (red), or only the HD panel (blue).*

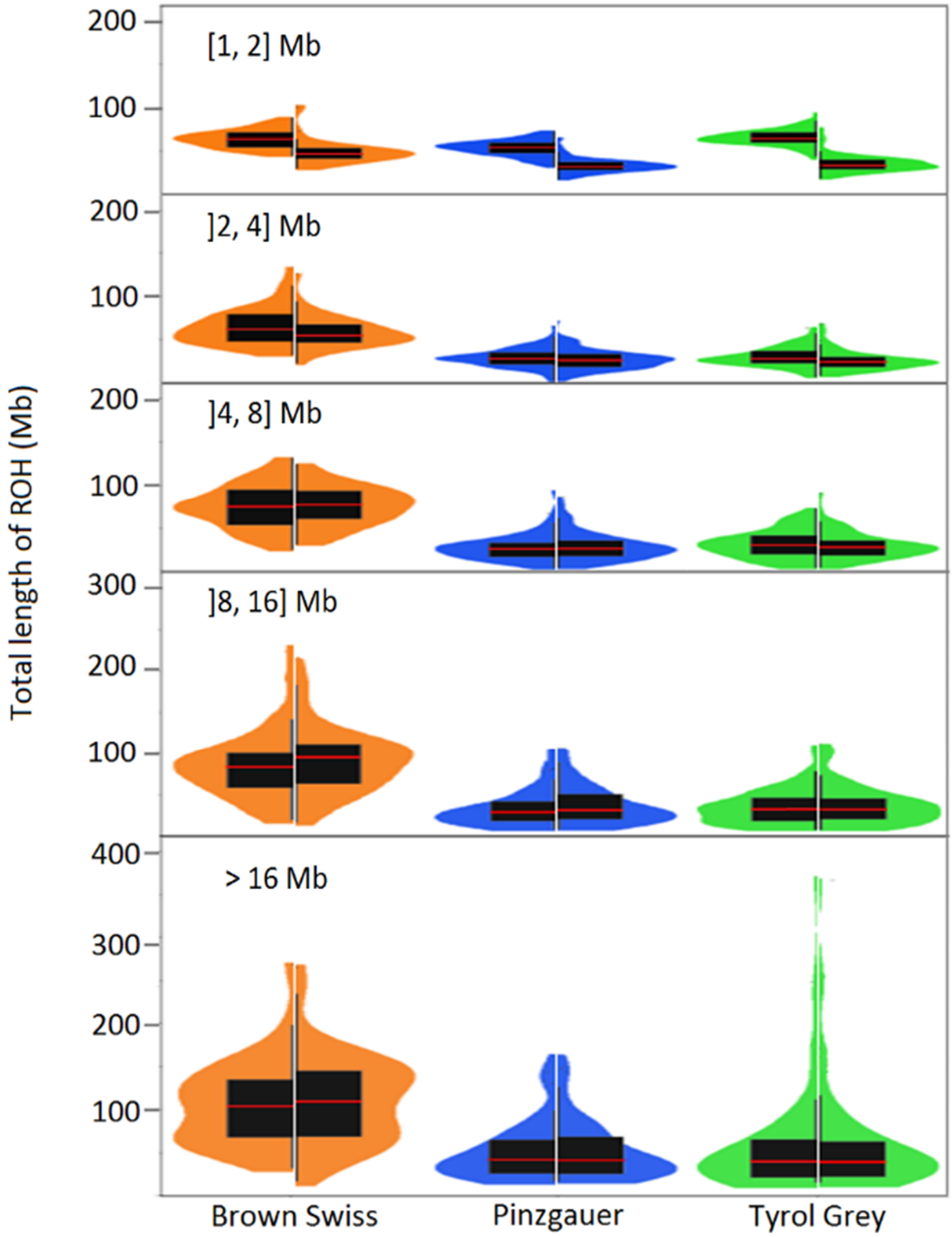


Figure 3 *Overlay of box plots and kernel density plots. Overlay of box plots and kernel density plots that show the distribution of the total ROH length among all Brown Swiss bulls (orange), Pinzgauer bulls (blue) and Tyrol Grey bulls (green) for five ROH length categories; box plots (black) are shown inside the density plots, and horizontal red lines indicate mean values; the left half of each density and box plot was obtained from the 50 k panel data, while the right half was obtained from the HD panel data.*

Table 4. *Comparison of F_{ROH} values obtained by allowing different numbers of heterozygous SNPs*

Breed	Source	Class	$F_{ROH>1}$	$F_{ROH>2}$	$F_{ROH>4}$	$F_{ROH>8}$	$F_{ROH>16}$
Brown Swiss	50 k	A	0.154^{b}	0.129^{ab}	0.103^{a}	0.073^{ab}	0.039^{dfhi}
		B	.	.	.	.	0.039^{bceg}
	HD	C	0.151^{b}	0.132^{b}	0.109^{b}	0.079^{c}	0.042^{i}
		D	0.147^{a}	0.128^{a}	0.105^{a}	0.076^{b}	0.040^{gh}
		E	.	0.129^{ab}	0.105^{a}	0.075^{b}	0.038^{ef}
		F	.	.	0.105^{a}	0.071^{a}	0.035^{cd}
		G	.	.	.	0.068^{a}	0.033^{b}
		H	.	.	.	.	0.028^{a}
Pinzgauer	50 k	A	0.069^{c}	0.048^{ab}	0.037^{a}	0.026^{bc}	0.014^{f}
		B	.	.	.	.	0.013^{ed}
	HD	C	0.062^{b}	0.049^{b}	0.039^{b}	0.027^{d}	0.014^{fe}
		D	0.060^{a}	0.048^{ab}	0.038^{ab}	0.026^{dc}	0.013^{fe}
		E	.	0.048^{a}	0.037^{a}	0.026^{bc}	0.012^{d}
		F	.	.	0.036^{a}	0.025^{ab}	0.012^{c}
		G	.	.	.	0.024^{a}	0.011^{b}
		H	.	.	.	.	0.008^{a}
Tyrol Grey	50 k	A	0.080^{c}	0.054^{a}	0.042^{a}	0.029^{abc}	0.017^{df}
		B	.	.	.	.	0.015^{ce}
	HD	C	0.066^{b}	0.052^{a}	0.042^{a}	0.030^{c}	0.017^{f}
		D	0.063^{a}	0.051^{a}	0.041^{a}	0.029^{b}	0.016^{d}
		E	.	0.051^{a}	0.040^{a}	0.029^{b}	0.016^{ed}
		F	.	.	0.040^{a}	0.028^{ab}	0.015^{c}
		G	.	.	.	0.026^{a}	0.013^{b}
		H	.	.	.	.	0.010^{a}

Definition of Class is according to the number of heterozygous SNPs allowed within ROH length categories (Table 2.). F_{ROH} values were obtained by allowing different numbers of heterozygous SNPs in each ROH length category; different letters indicate statistical significance within the same column and breed ($P < 0.05$, paired t-test). P values were corrected for multiple test using step down Bonferroni method of Holm (Holm 1979)

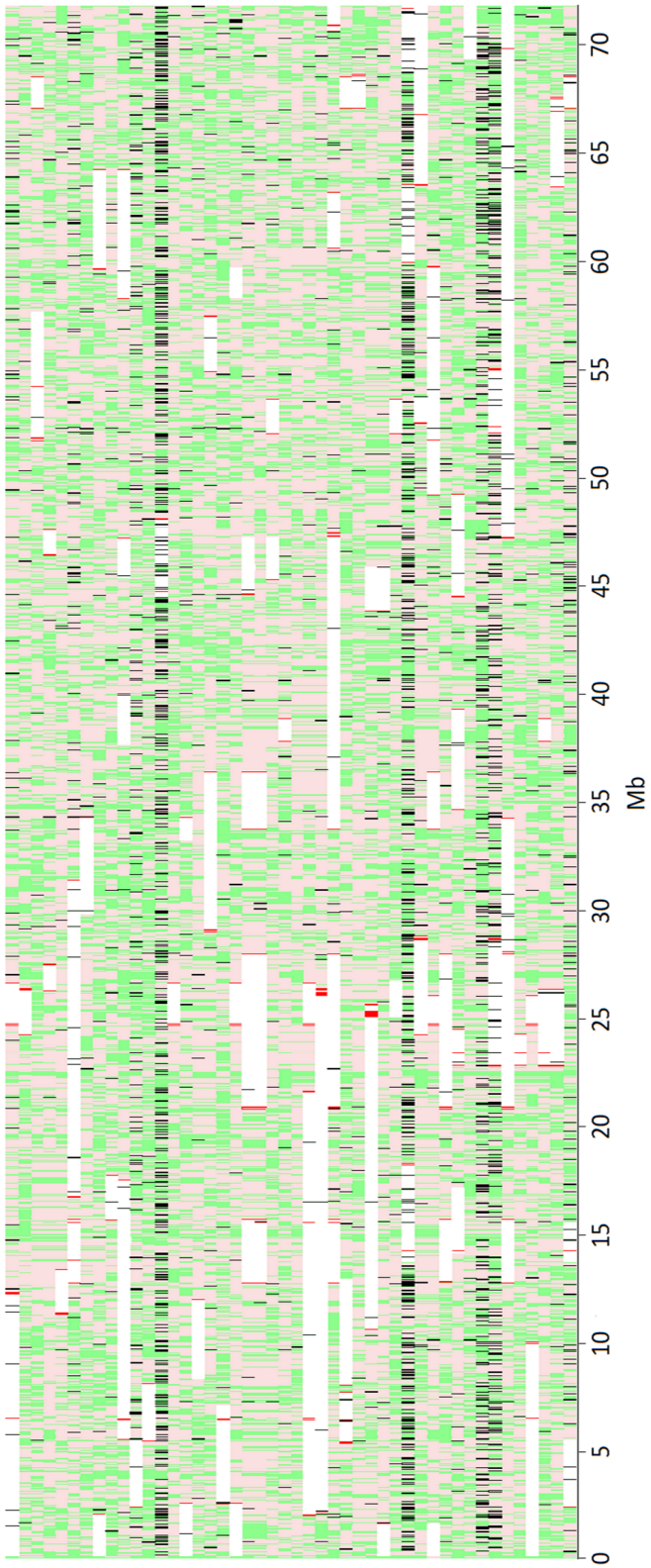


Figure 4 *Visualization of SNP data of chromosome 20 in Brown Swiss animals. Light pink and light green colors represent homozygous and heterozygous SNPs, respectively; ROH are represented by white blocks, while missing SNPs are indicated in black; red lines within ROH indicate the presence of heterozygous SNPs; each column represents one animal.*

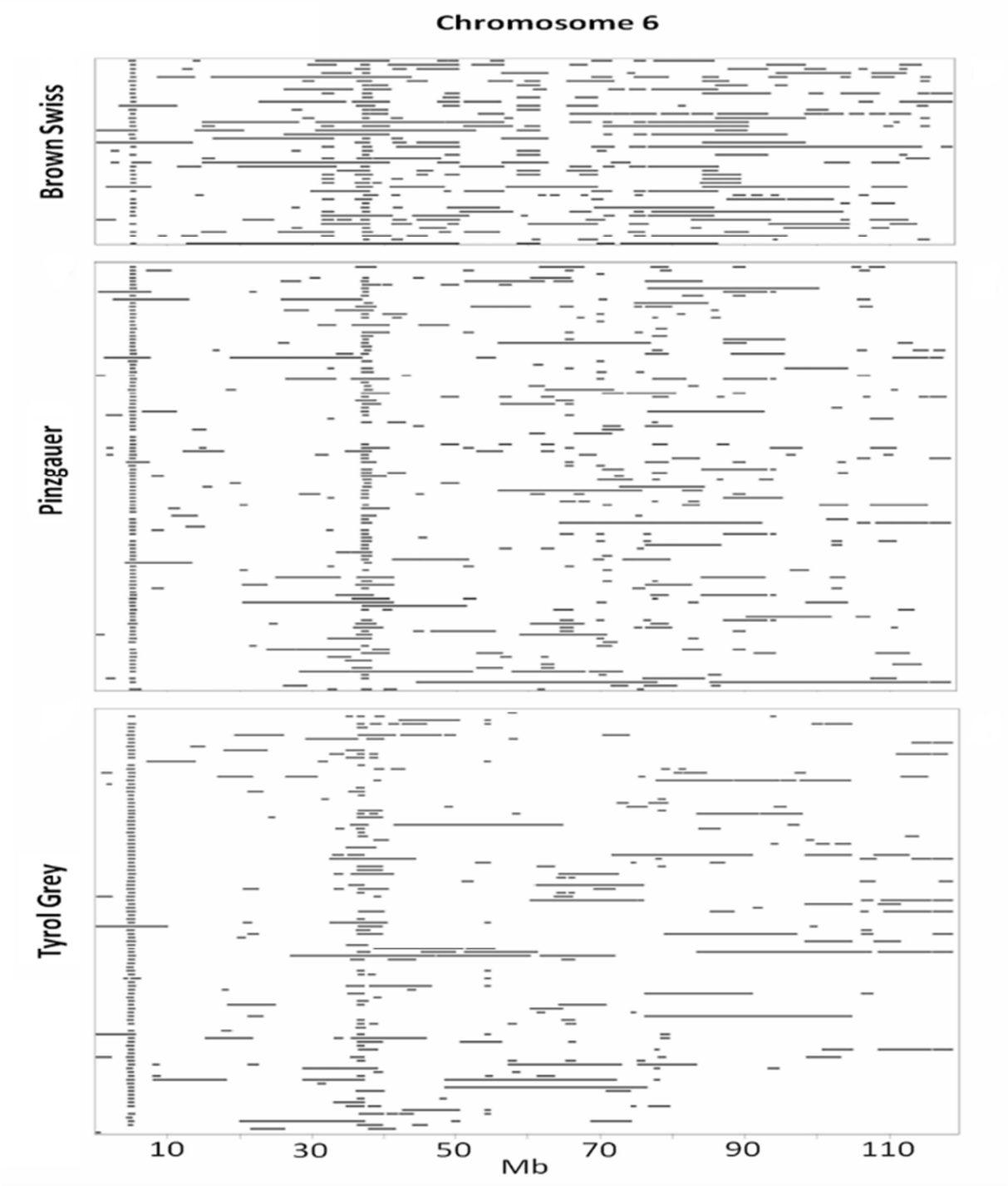


Figure 5 ROH patterns on chromosome 6.

ROH on chromosome 6 from Brown Swiss, Pinzgauer and Tyrol Grey bulls identified using HD panel data; each row represents one animal

Inbreeding coefficients estimated from ROH and ROH distribution

The HD panel gave the following mean F_{ROH} values across all ROH lengths: Brown Swiss, 0.151; Pinzgauer, 0.062; and Tyrol Grey, 0.066. Short ROH, i.e. 1 to 2 Mb long, covered an average of 36.7 Mb of the 2.3 Gb of the autosomal cattle genome covered with SNPs (Figure 2), with the highest short-ROH coverage observed in Brown Swiss and the lowest in Pinzgauer, the total genome length covered by all ROH > 1 Mb was 24.5% for one Brown Swiss bull and 23.0% for one Tyrol Grey bull. ROH > 16 Mb covered an average of 66.1 Mb of genome, although this number varied widely from animal to animal and between breeds. The highest long ROH coverage was observed in Brown Swiss and the lowest in Pinzgauer cattle. Some animals lacked such long ROH, whereas others showed a few that covered more than 300 Mb. The greatest genome coverage by long ROH was observed in a Tyrol Grey bull, in which 12 long ROH segments covered 368.6 Mb, corresponding to an average segment length of ≈30 Mb. The length of an autozygous segment indicates its age; since haplotypes are broken up by meiotic recombination, a short autozygous region is likely to have an ancient origin, while a long one probably arose recently (Broman & Weber 1999a; Keller *et al.* 2011) These findings suggest that the Brown Swiss breed experienced both recent and ancient inbreeding events to a higher degree than the two other breeds.

Correlations of F_{ROH} values across all ROH lengths with pedigree inbreeding coefficients were similar to those previously reported by Ferencakovic *et al.* (2013). Correlations for the 50 k panel were 0.62, 0.65 and 0.77 for Brown Swiss, Pinzgauer and Tyrol Grey, respectively, and corresponding values were 0.61, 0.62 and 0.75 for the HD panel. Differences in correlations between panels within breeds were not statistically significant. Variation of these values is most likely due to the fact that pedigree-based inbreeding coefficients do not account for variation in meiosis, inheritance of segments of chromosomes and LD.

The genomic distribution of ROH based on the HD panel data shows that 99.98% of SNPs occurred within an ROH of at least one individual. However, the frequency with which different SNPs occurred within ROH was not uniform across the genome, revealing genomic regions with abundant ROH, called ROH hotspots, which are also often detected in human populations (Nothnagel *et al.* 2010; Pemberton *et al.* 2012). Several ROH hotspots were common to all three breeds. For example, two hotspots were identified on chromosome 6 in all three breeds: one at 5.3-6.3 Mb and another at 38.4-39.5 Mb (Figure 4). Why these hotspots occur, and how they compare among cattle breeds and with other animal species, are questions currently under investigation.

Conclusions

ROH identification in cattle is usually performed with the Illumina BovineSNP50 Genotyping BeadChip (50 k panel) or the Illumina BovineHD Genotyping BeadChip (HD panel). Here, we report that data from the 50 k panel do not represent the true state of autozygosity well for short ROH segments, while it is as reliable as the HD panel data for ROH > 4 Mb. When shorter segments are included with the 50 k panel, F_{ROH} is systematically overestimated. The bias due to potential genotyping errors depends on the allowance of heterozygous genotypes in a ROH calling software. While not allowing for heterozygous calls often just splits a very long ROH in two shorter ones that are still recognized and therefore the level of autozygosity of an individual is virtually unaffected, there are many cases where the shorter part of the split does not reach the minimum size of a ROH and the level of autozygosity of an individual is underestimated. Allowing many heterozygous calls in an ROH adds many short segments that are most likely not autozygous to the terminal regions of ROH. Our aim was to provide guidelines to identify ROH from high-throughput SNP genotype data. First, quality control should be performed by removing SNPs based on strict limits on genotype quality scores provided to reduce genotyping errors. Second, the number of heterozygous SNPs allowed should be determined separately for each ROH length of interest and for each SNP density, as suggested here. Third, if multiple heterozygous SNPs are allowed within the same ROH, adjacent heterozygous SNPs should be treated differently from heterozygous SNPs that are further apart. Because no current ROH identification software takes care of adjacent heterozygous SNPs, careful visual inspection of ROH segments should be applied to exclude spurious ROH called by the software.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MF participated in the design of the study, identified ROH, performed all data and statistical analyses, and drafted the manuscript. JS and IC conceived the study, helped to prepare figures and tables, and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the Golden Helix team for providing a free license for SNP & Variation Suite v7.6.8 Win64. We are very grateful to ZuchtdataEDV-Dienstleistungen GmbH for provision of pedigree data. This research was funded by the Croatian Ministry of Science, Education and Sports (project 178-1780460-0546, 'Strategies for elimination of genetic defects in selected populations'), by a bilateral collaborative project involving the Croatian Ministry of Science, Education and Sports and the Austrian Agency for International Cooperation in Education and Research (OeAD- GmbH) ("Analysis of genetic diversity in cattle populations using dense sets of molecular markers"), and by grant EX/3733 from the European Science Foundation networking program "Advances in Farm Animal Genomic Resources" ("GENOMIC-RESOURCES"). Finally, we gratefully acknowledge the editor and anonymous reviewers. Their expertise contributed to improve the quality of the manuscript.

References

Albrechtsen A., Nielsen F. & Nielsen R. (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution* **27**, 2534 - 47.

Bjelland D., Weigel K., Vukasinovic N. & Nkrumah J. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science* **96**, 4697 - 706.

Broman K. & Weber J. (1999) Long homozygous chromosomal segments in reference families from the centre d'etude du polymorphisme humain. *American Journal of Human Genetics* **65**, 1493 - 500.

Esko T., Mezzavilla M., Nelis M., Borel C., Debniak T., Jakkula E., Julia A., Karachanak S., Khrunin A., Kisfali P., Krulisova V., Ausrele Kucinskiene Z., Rehnstrom K., Traglia M., Nikitina-Zake L., Zimprich F., Antonarakis S., Estivill X., Glavac D., Gut I., Klovins J., Krawczak M., Kucinskas V., Lathrop M., Macek M., Marsal S., Meitinger T., Melegh B., Limborska S. & Lubinski J. (2013) Genetic characterization of northeastern Italian population isolates in the context of broader European genetic diversity. *European Journal of Human Genetics* **21**, 659 - 65.

Ferenčaković M., Hamzic E., Gredler B., Curik I. & Solkner J. (2011) Runs of homozygosity reveal genome-wideautozygosity in the Austrian fleckvieh cattle. *Agriculturae Conspectus Scientificus* **76**, 325 - 8.

Ferenčaković M., Hamzić E., Gredler B., Solberg T.R., Klemetsdal G., Curik I. & Sölkner J. (2013) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding & Genetics* **130**, 286-93.

Gibson J., Morton N. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789 - 95.

Holm S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**, 65 - 70.

Howrigan D., Simonson M. & Keller M. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* **12**, 460.

Keller M., Visscher P. & Goddard M. (2011) Quantification of Inbreeding Due to Distant Ancestors and its Detection Using Dense SNP Data. *Genetics* **189**, 237 - 49.

Kirin M., McQuillan R., Franklin C., Campbell H., McKeigue P. & Wilson J. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996.

Lencz T., Lambert C., DeRosse P., Burdick K., Morgan T., Kane J., Kucherlapati R. & Malhotra A. (2007) Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 19942 - 7.

Leutenegger A., Prum B., Genin E., Verny C., Lemainque A., Clerget-Darpoux F. & Thompson E. (2003) Estimation of the inbreeding coefficient through use of genomic data. *American Journal of Human Genetics* **73**, 516 - 23.

Nothnagel M., Lu T.T., Kayser M. & Krawczak M. (2010) Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human Molecular Genetics* **19**, 2927-35.

Pemberton T., Absher D., Feldman M., Myers R., Rosenberg N. & Li J. (2012) Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics* **91**, 275 - 92.

Powell J., Visscher P. & Goddard M. (2010) Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics* **11**, 800 - 5.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M., Bender D., Maller J., Sklar P., De Bakker P., Daly M. & Sham P. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559 - 75.

Purfield D., Berry D., McParland S. & Bradley D. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* **13**, 70.

Rabbee N. & Speed T. (2006) A genotype calling algorithm for affymetrix SNP arrays. *Bioinformatics* **22**, 7 - 12.

SAS Institute. (2011) *What's New in SAS 9.3*. SAS Institute.

Silió L., Rodríguez M.C., Fernández A., Barragán C., Benítez R., Óvilo C. & Fernández A.I. (2013) Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *Journal of Animal Breeding & Genetics* **130**, 349 - 60.

Solkner J., Filipcic L. & Hampshire N. (1998) Genetic variability of populations and similarity of subpopulations in Austria cattle breeds determined by analysis of pedigrees. *Animal Science* **67**, 249 - 56.

Wigginton J., Cutler D. & Abecasis G. (2005) A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics* **76**, 887 - 93.

Zhang L., Orloff M., Reber S., Li S., Zhao Y. & Eng C. (2013) cgaTOH: extended approach for identifying tracts of homozygosity. *PLoS ONE* **8**, e57772.

CHAPTER 5 MOLECULAR DISSECTION OF INBREEDING DEPRESSION FOR SEMEN QUALITY TRAITS IN AUSTRIAN FLECKVIEH CATTLE

The inbreeding depression (Darwin 1876) defines the reduction in fitness or in the mean value of the phenotype of the offspring derived from mating between relatives (inbreeding). The inbreeding depression can have various manifestations, from major abnormalities (mutant phenotypes lethal early in life, genetic diseases), to less serious conditions (decrease in performance, in growth, reproduction and viability) (Charlesworth & Willis 2009). Inbreeding depression is major concern in many fields of biology, for example, in agriculture, conservation biology, human health etc. Still, the genetic basis of inbreeding depression is unclear.

Overdominance and partial dominance are the two main hypotheses used for explaining inbreeding depression. In the overdominance hypothesis (East 1908), the heterozygous genotype is supposed to be superior to both homozygous genotypes. The loss of heterozygosity through inbreeding leads to a decrease in mean values of traits associated with fitness and consequently to inbreeding depression (Lynch & Walsh 1998). The partial dominance hypothesis (Davenport 1908) supposes that most mutations are neutral or deleterious and recessive in general (Mackay 2001). The increase of the proportion of homozygotes caused by inbreeding increases the probability of expression of these deleterious alleles (Charlesworth & Charlesworth 1999), consequently leading to inbreeding depression (Keller & Waller 2002). While changes in genotypic value due to inbreeding are related to the existence of directional dominance interaction (Falconer & Mackay 1996), there is still an ongoing discussion as to which of the two previously described hypotheses explains the decline in fitness associated with inbreeding (Ritland 1996; Karkkainen *et al.* 1999; Roff 2002) The two hypotheses have different evolutionary implications as well as implications for animal breeding but they are not mutually exclusive (Kristensen & Sorensen 2005). Evidence supports both models (Hughes 1995; Li *et al.* 2001; Carr & Dudash 2003), but modern molecular approaches suggest that inbreeding depression is predominantly caused by the presence of recessive deleterious mutations in populations i.e. dominance hypothesis (Charlesworth & Willis 2009). On the other side, some authors suggest that inbreeding

depression may also be explained, at some part, by the epistatic interactions between genes (Templeton & Read 1984, 1994; Curik *et al.* 2001).

Underlying molecular mechanisms of inbreeding depression, including gene pathways or number of loci involved, are also unknown. However, because most of the important traits have a polygenic nature it was confirmed by Ayroles *et al.* 2009 that a large proportion of the genome (i.e. large number of genes) is involved in the expression of inbreeding depression. It is crucial find the causal locus or loci in order to precisely determine the part of the genome that is identical by descent (IBD). Keller *et al.* (2011 & 2012) proposed runs of homozygosity (ROH), regions of the genome without heterozygosity in diploid state (Gibson *et al.* 2006) as a reliable estimate of genome autozygosity as well as a suitable tool for estimating inbreeding depression. Bjelland *et al.* (2013) used inbreeding coefficients derived from ROH (F_{ROH}) and other molecular inbreeding estimators for detection of inbreeding depression on lactation performance and reproductive traits in Holstein cattle and concluded that only F_{ROH} could distinguish between markers that were IBD and markers that were identical by state (IBS). Recently, Pryce *et al.* (2014) successfully used ROH for pinpointing specific regions of the genome that were associated with inbreeding depression in fertility and milk production.

Inbreeding depression for cattle fertility reported by Bjelland *et al.* (2013) and Pryce *et al.* (2014) was analyzed only on traits measured on cows. However, male fertility, i.e. sperm structural defects of bulls, are also important. Inbreeding negatively affects reproductive performance and that high levels of inbreeding could be the major cause of poor semen quality (Wildt *et al.* 1982; Smith *et al.* 1989; Margulis & Walsh 2002; Aurich *et al.* 2003; Van Eldik *et al.* 2006; Maximini *et al.* 2011). Moreover, Maximini *et al.* (2011) showed that inbreeding depression affected male fertility of Simmental bulls using pedigree and sperm quality data. They analyzed five qualitative semen quality traits (volume, concentration, total number of spermatozoa, percent of alive spermatozoa and motility) of which all traits but concentration were affected by inbreeding.

The aim of this study is to use ROH obtained from bovine SNP50 Beadchip v1 (Illumina Inc., San Diego, CA) to detect autozygous regions of the genome that influence sperm quality in Austrian Fleckvieh bulls. Possible associations between bull fertility and genome wide autozygosity could reveal genes responsible for inbreeding depression of bull fertility.

Materials and methods

In total, 1799 Austrian Fleckvieh bulls were genotyped with bovine SNP50 Beadchip v1 (Illumina Inc., San Diego, CA), containing 54 001 SNPs. Pedigree data that traced back to the 1930s consisted of 41090 animals and were obtained from Zuchtdata EDV-Dienstleistungen GmbH. The pedigree was checked and recoded using the CFC program (Sargolzaei *et al.* 2006). From the pedigree data the equivalent complete generations and pedigree inbreeding coefficients for full pedigree (F_{PED}) and for five generations (F_{PED}) were calculated using ENDOG v4.8 (Gutierrez & Goyache 2005).

The phenotype data related to sperm quality were obtained from three Austrian artificial insemination (AI) stations. From Gleisdorf, the station located in Styria, 7704 ejaculates from 301 bulls were collected from 2000 to 2010. From Hohenzell, the station located in Upper Austria, 16671 ejaculates from 309 bulls were collected from 2000 to 2009. From Wieselburg, the station located in Lower Austria, 15514 ejaculates from 293 bulls were collected from 2000 to 2009. All the three AI stations keep their bulls in tie-stalls and collect semen several times a week, using a dummy or teaser animal and an artificial vagina. The traits recorded routinely for every ejaculate were volume, sperm concentration, percentage of viable spermatozoa, and motility, except in AI Gleisdorf where the motility was not recorded. The semen collectors were also recorded. Stations Hohenzell and Wieselburg routinely collect ejaculate two or three times per day from the same bull, while in the Gleisdorf station only one ejaculate is collected per day from the same bull.

Taking account of the SNP density of bovine SNP50 Beadchip v1 (Illumina Inc., San Diego, CA), quality control was performed and ROH were determined following based on the settings proposed by Ferenčaković *et al.* (2013b). Regarding allowed missing SNPs, zero were allowed for ROH categories [1,2] and (2, 4], one for category (4, 8] Mb, two for category (8, 16] Mb and finally, four for category >16Mb. Only in category >16 Mb also one heterozygous call was allowed. From determined ROH segments F_{ROH} estimates were calculated following McQuillan *et al.* (2008) and Ferenčaković *et al.* (2013b). Additionally, partial F_{ROH} coefficients were estimated by using only ROH segments of a particular size. The coefficients for ROH segments were estimated in range of (2, 4] Mb; (F_{ROH2-4}), (4, 8] Mb; (F_{ROH4-8}) and for ROH segments in range of (8, 16] Mb; ($F_{ROH8-16}$). Using SNP & Variation Suite v7.6.8 Win64 (Golden Helix, Bozeman, MT, USA www.goldenhelix. com) the molecular inbreeding coefficient was calculated for every individual. This inbreeding coefficient is equivalent to Wright's within population fixation index F_{is} (Wright 1922).

Chromosome inbreeding coefficients ($F_{ROH\_Chr1,}$ F_{ROH_Chr2} ... F_{ROH_Chr29}) were also calculated based on ROH segments greater than 2 Mb. This was done in the same manner as for the genome $F_{ROH,}$ with the exception that here the total length of the specific chromosome arranged in segments >2 Mb was divided by the total length of the chromosome covered by SNPs. Correlations of the estimated inbreeding coefficients were calculated and visualized with SAS 9.3 (SAS Institute 2011).

The data editing was performed in the following steps. The sperm data where the volume was not between 1 to 25 ml, concentration was not in the range of 0.1 to 3×10^9 ml^{-1} or where values were missing, were removed. Records from second and third jump in the same day were also removed. The total number of spermatozoa, calculated from volume and concentration was used in further analyses in this study because it was considered to be the most objective trait.

Bulls with less than 10 observations were not included in analyses. The ejaculates collected from the same bull in a period shorter than four days were also not included. The analysis of the distribution of the total number of spermatozoa showed that the trait was not normally distributed and that transformation was needed. The transformation in which transformed total number of spermatozoa = (total number of spermatozoa$^{0.3}$-1)/0.3 was determined with TRANSREG procedure and BOXCOX options (SAS Institute 2011) (Box & Cox 1964). After transformation, observations beyond the range mean ±2.5 standard deviations were excluded from analysis. Also, the influence of each single observation was checked in the model of transformed number of spermatozoa by including the following independent variables: age of bull, semen collector, month and year of collection, station and the interval of days between two ejaculates. The age of bull was defined as a categorical variable in three categories (<16 months, 16 to 72 and >72 months). The period between two successive ejaculates was defined in three categories (4 to 7 days, 7 to 9, and >9 days). The months of collection (season) was defined in three categories (February-March-April-May, June-July-August-September, October-November-December-January.) To fit the model and detect outliers, MIXED procedure (SAS Institute 2011) and INFLUENCE options were used. After extensive data cleaning and quality control the final data set consisted of 19720 ejaculate records from 554 bulls.

The models for estimating inbreeding depression using F_{ROH} and F_{PED} were the following:

$$y_{ijklmnop} = \mu + \alpha_i + \text{age}_j + \text{month}_k + \text{year}_l + \text{collection\_interval}_m + \text{semen\_colector}_n + \text{station}_o + bF_{ijklmnop} + \varepsilon_{ijklmnop}$$

where $bF_{ijklmnop}$ is the regression coefficient on $F_{ijklmnop}$, which were various measures of inbreeding (F_{PED}, F_{PED5}, $F_{GENOMIC}$, $F_{ROH>1\ Mb}$, $F_{ROH>2\ Mb}$, $F_{ROH>4\ Mb}$, $F_{ROH>8\ Mb}$, $F_{ROH>16\ Mb}$, or $F_{ROH1\text{-}2\ Mb}$, $F_{ROH1\text{-}4\ Mb}$, $F_{ROH2\text{-}4\ Mb}$, $F_{ROH4\text{-}8\ Mb}$, $F_{ROH8\text{-}16\ Mb}$) creating 13 different models, $y_{ijklmnop}$ is the individual observation, μ the overall mean, α_i the random effect of animal i assumed to follow the distribution N(0, **G**σ2), where **G** is the genomic relationship matrix calculated using IBD algorithms as devised in Oliehoek *et al.* (2006), age_j the fixed effect of age class j, $month_k$ the fixed effect of month class k, $year_l$ the fixed effect of year class l, $collection\_interval_m$ the fixed effect of interval in days since last collection, $semen\_colector_n$ the fixed effect of semen collector n, and $\varepsilon_{ijklmnop}$ is the random error associated with the observation.

G matrix was obtained using JMP Genomics 5.0 (SAS Institute Inc, Cary, NC, USA) while all other analysis were made with PROC MIXED implemented in SAS 9.3 (SAS Institute 2011). Suitability of inbreeding coefficients for detection of inbreeding depression was tested in a way that AIC was obtained for every model regarding the used measure of inbreeding and the best were chosen for further analysis (Burnham & Anderson 2002).

In order to investigate the effect of inbreeding of a single SNP on the total number of spermatozoa, which should help in identification of regions associated with inbreeding depression, a model similar to the one described in previous paragraph was set.

The model was defined as:

$$y_{ijklmnop} = \mu + \alpha_i + age_j + month_k + year_l + collection\_interval_m + semen\_colector_n + station_o + b_1SNP_{ijklmnop} + b_2ROH_{ijklmnop} + \varepsilon_{ijklmnop}$$

where $b_1SNP_{ijklmnop}$ is the regression coefficient which was to correct for the additive effect of the SNP and $b_2ROH_{ijklmnop}$ is the regression coefficient on $ROH_{ijklmnop}$. At each SNP position, $SNP_{ijklmnop}$ was coded 0, 1 or 2 for homozygous, heterozygous and alternative homozygous configurations while $ROH_{ijklmnop}$ was coded as 1 if SNP was in the run of homozygosity, and 0 if it was not while all the rest remains the same like in the previous model. From the analysis of suitable inbreeding coefficients from the previous paragraph $F_{ROH>2\ Mb}$ and $F_{ROH2\text{-}4}$ were taken into account. These two separate models were applied on every autosomal SNP in the data set, which means 42 817 runs were performed per model.

Adjustment of p values was done with multiple testing correction *simpleM*, which uses composite linkage disequilibrium (CLD) to create the correlation matrix of SNPs and *MeffG* to calculate the effective number of independent tests (Gao *et al.* 2008). Effective number of independents was then implemented in the Bonferroni correction formula (Holm 1979). Data analysis, corrections and visualizations were made with SAS 9.3.

SNPs with significant effect on the total number of spermatozoa and annotated genes in the vicinity of those SNPs (± 1 Mb from the signal) were analyzed using Ensembl BioMart MartView (http://asia.ensembl.org/biomart/martview/). Only genes with known function described in UniProt (http://www.uniprot.org) and/or GeneCards (http://www.genecards.org) were further investigated.

Results & Discussion

Data, pedigree and ROH inbreeding coefficients

The overall statistics of the trait transformed total number of spermatozoa was as follows: mean; 2.71, standard deviation (SD); 0.87, Range; 0.09 – 5.16. Genotype data consisted of 42 817 autosomal SNPs. Mean complete generations equivalent for the 554 bulls was 6.82 (SD; 0.63) ranging from 4.58 to 8.32. Overall statistics of pedigree inbreeding coefficients, genomic and ROH inbreeding coefficients are presented in Table 1. As expected, the highest values of inbreeding were estimated using $F_{ROH\ >1\ Mb}$ and this is consistent with our previous findings (Ferenčaković *et al.* 2011; Ferenčaković *et al.* 2013a; Ferenčaković *et al.* 2013b). However, this value is not a reliable estimator of autozygosity for this SNP chip because it systematically overestimates number of ROH segments (Purfield *et al.* 2012; Ferenčaković *et al.* 2013b). Inbreeding coefficients $F_{ROH2\text{-}4\ Mb}$, $F_{ROH4\text{-}8\ Mb}$ and $F_{ROH8\text{-}16\ Mb}$ had values very close to values obtained from F_{PED} and F_{PED5}. The chromosome inbreeding coefficients were not uniformly distributed across genome (Figure 1). This information proves the complex nature of genome wise landscape of autozygosity and emphasizes that some parts of genome are more IBD than others. The correlations between all estimated inbreeding coefficients are shown in Table 2.

Table 1. *Descriptive statistics for the various inbreeding coefficients*

Inbreeding coefficient	Mean	SD	Min	Max
F_{PED}	0.01269	0.01112	0.00000	0.07535
F_{PED5}	0.00960	0.01029	0.00000	0.07031
F_{is}	-0.00452	0.02681	-0.08613	0.12013
$F_{ROH >1 Mb}$	0.08545	0.01799	0.04007	0.17790
$F_{ROH >2 Mb}$	0.04711	0.01644	0.01048	0.14552
$F_{ROH >4 Mb}$	0.02730	0.01517	0.00241	0.12130
$F_{ROH >8 Mb}$	0.01415	0.01323	0.00000	0.10195
$F_{ROH >16 Mb}$	0.00636	0.00934	0.00000	0.07131
$F_{ROH1-2 Mb}$	0.03834	0.00542	0.01939	0.05535
$F_{ROH1-4 Mb}$	0.05815	0.00806	0.03332	0.08637
$F_{ROH2-4 Mb}$	0.01980	0.00505	0.00591	0.03535
$F_{ROH4-8 Mb}$	0.01315	0.00595	0.00000	0.03418
$F_{ROH8-16 Mb}$	0.00779	0.00708	0.00000	0.05020

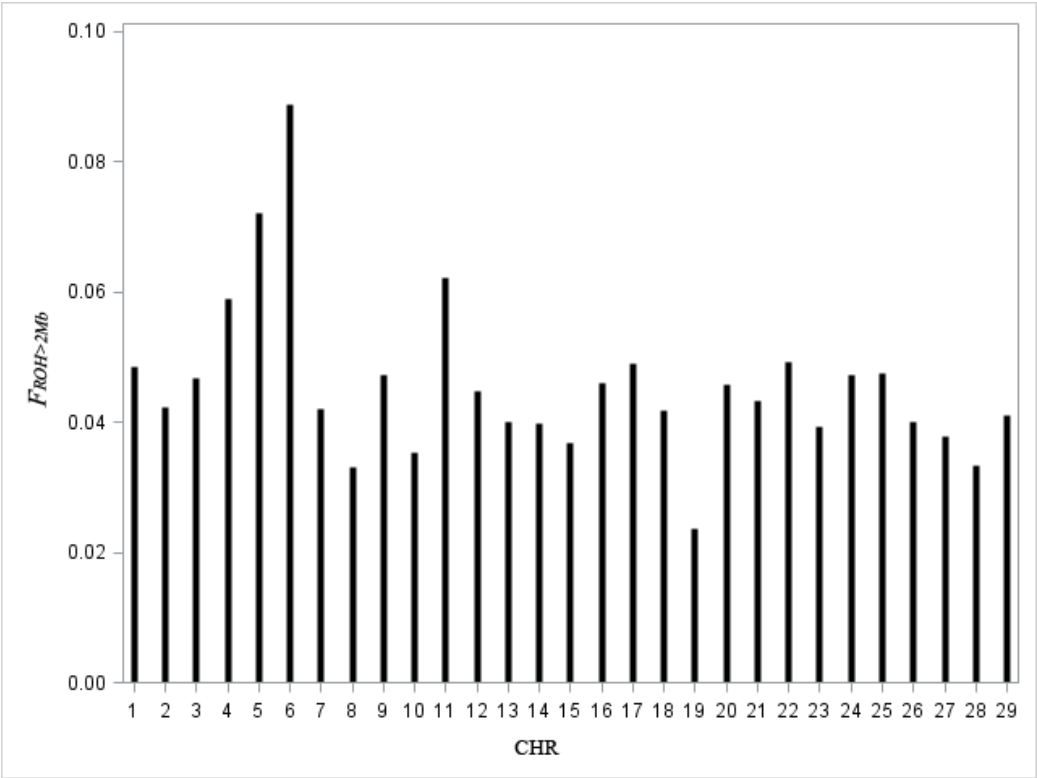


Figure 1. *Levels of autozygosity observed on 29 bovine chromosomes derived from ROH segments >2 Mb.*

Table 2. Pearson correlation coefficients between pedigree inbreeding coefficients (F_{PED} and F_{PED5}) and genomic inbreeding coefficients (F_{HOM}, $F_{ROH>1}$, $F_{ROH > 2\ Mb}$, $F_{ROH > 4\ Mb}$, $F_{ROH > 8\ Mb}$, $F_{ROH>16\ Mb}$, $F_{ROH1-2\ Mb}$, , $F_{ROH1-4\ Mb}$, $F_{ROH2-4\ Mb}$, $F_{ROH4-8\ Mb}$, and $F_{ROH8-16\ Mb}$).

Inbreeding coefficient	F_{PED}	F_{PED5}	F_{is}	F_{ROH} >1 Mb	F_{ROH} >2 Mb	F_{ROH} >4 Mb	F_{ROH} >8 Mb	F_{ROH} >16 Mb	F_{ROH} 1-2 Mb	F_{ROH} 1-4 Mb	F_{ROH} 2-4 Mb	F_{ROH} 4-8 Mb	F_{ROH} 8-16 Mb
F_{PED5}	0.96												
F_{is}	0.47	0.43											
$F_{ROH\ >1\ Mb}$	0.55	0.53	0.89										
$F_{ROH\ >2\ Mb}$	0.58	0.57	0.81	0.95									
$F_{ROH\ >4\ Mb}$	0.59	0.60	0.75	0.90	0.95								
$F_{ROH\ >8\ Mb}$	0.59	0.60	0.66	0.82	0.88	0.92							
$F_{ROH\ >16\ Mb}$	0.56	0.57	0.55	0.70	0.76	0.79	0.86						
$F_{ROH1-2\ Mb}$	0.05	0.03	0.50	0.42	0.14	0.09	0.05	0.02					
$F_{ROH1-4\ Mb}$	0.05	0.03	0.50	0.55	0.34	0.12	0.10	0.08	0.79				
$F_{ROH2-4\ Mb}$	0.11	0.07	0.38	0.42	0.39	0.09	0.11	0.11	0.18	0.75			
$F_{ROH4-8\ Mb}$	0.20	0.19	0.45	0.46	0.46	0.50	0.13	0.11	0.11	0.07	0.00		
$F_{ROH8-16\ Mb}$	0.36	0.37	0.50	0.61	0.64	0.68	0.74	0.29	0.07	0.08	0.06	0.09	

In Figure 2. the correlations between F_{PED} and chromosome inbreeding coefficients (F_{ROH_Chr1}, F_{ROH_Chr2} ... F_{ROH_Chr29}) estimated from segments > 2 Mb are presented via radar plot while Figure 3. presents the correlations between $F_{ROH\ >2\ Mb}$ and chromosome inbreeding coefficients.

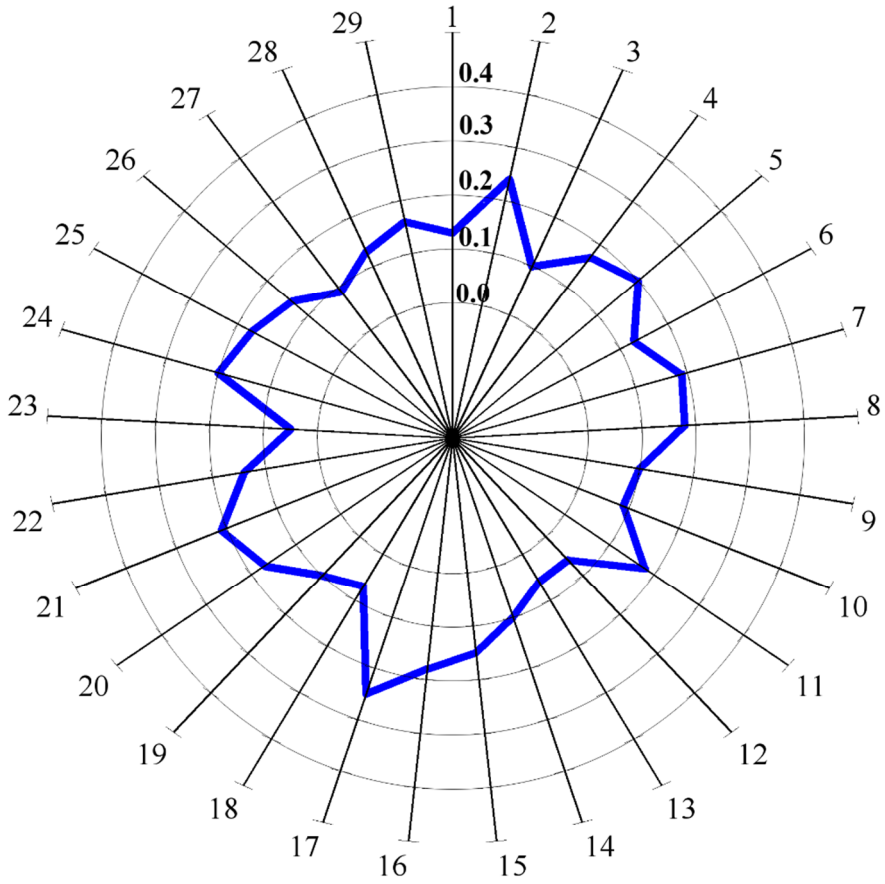


Figure 2. Radar plot of Pearson correlation coefficients between pedigree inbreeding coefficients (F_{PED}) and chromosome inbreeding coefficients (F_{ROH_Chr1}, F_{ROH_Chr2} ... F_{ROH_Chr29}) estimated from segments >2 Mb.

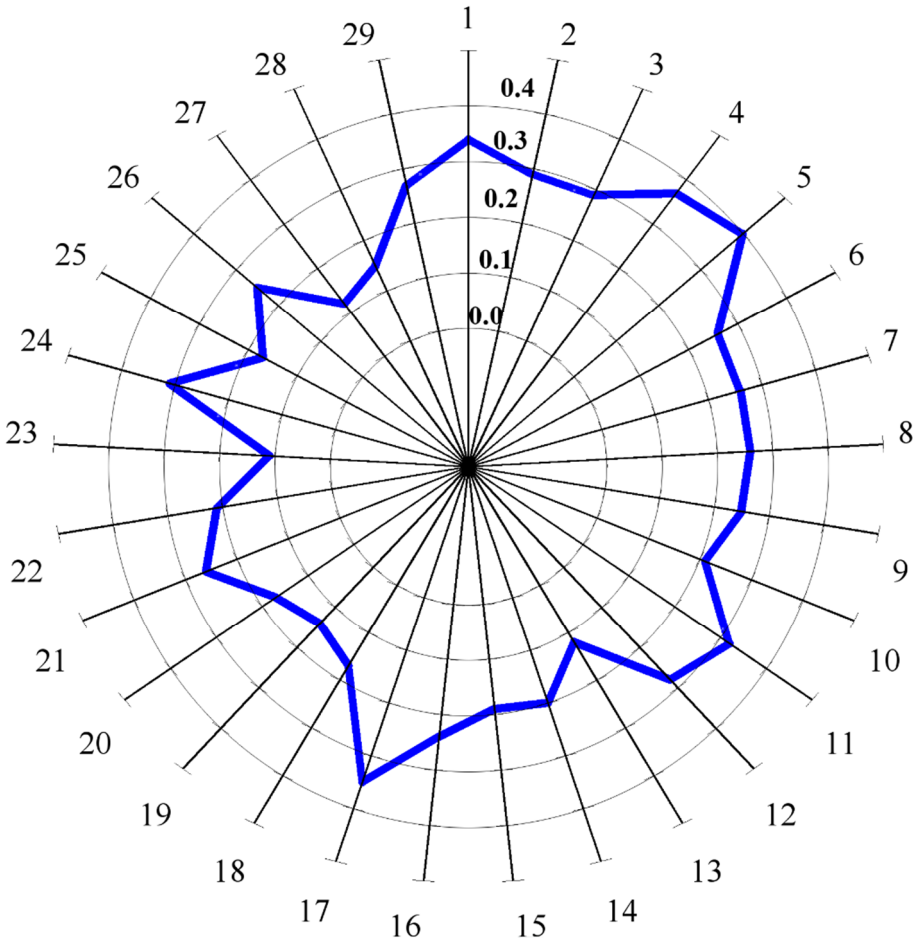


Figure 3. Radar plot of Pearson correlation coefficients between $F_{ROH\ >2\ Mb}$ and chromosome inbreeding coefficients (F_{ROH_Chr1}, F_{ROH_Chr2} ... F_{ROH_Chr29}) estimated from segments >2 Mb.

Inbreeding depression on total number of spermatozoa

A significant ($p < 0.05$) inbreeding depression on transformed value of total number of spermatozoa was detected using the pedigree inbreeding coefficient (F_{PED}), the inbreeding coefficient derived from ROH segments >2 Mb ($F_{ROH\ >2\ Mb}$) and partial inbreeding coefficients $F_{ROH2\text{-}4\ Mb}$ and $F_{ROH8\text{-}16\ Mb}$ (Table 3). Maximini *et al.* (2011) also reported the inbreeding depression on total number of spermatozoa when using F_{PED}. The use of various inbreeding measures was also tested using Burnham & Anderson (2002) model selection. By using ΔAIC, F_{PED} and $F_{ROH2\text{-}4\ Mb}$ were found to be the best measures of inbreeding in detecting inbreeding depression. ΔAIC between those two models was smaller than 2 indicating that there is no significant difference between them. Models with ΔAIC between 3 and 7 indicate that they have considerably less support, while those with ΔAIC >7 indicate that the model is not very likely (Burnham & Anderson 2002). This brings us to the conclusion that the cause for inbreeding depression on the total number of spermatozoa is in the segments from 2 to 4 Mb. Assuming that the expected length of an IBD haplotype follows an exponential distribution, the mean of which equals *100/(2 gcA)* cM, where *gcA* is the number of generations from the common ancestor, suggested age of inbreeding causing inbreeding depression on the total number of spermatozoa is 25 to 12.5 generations ago. F_{is} is a measure often used for estimating relatedness in population. This measure is implemented in two most used programs for analysis of genotype data PLINK v1.07 (http://pngu.mgh.harvard.edu/~purcell/plink/) and SNP & Variation Suite (Golden Helix, Bozeman, MT, USA www.goldenhelix. com). F_{is} performed very poorly in estimating inbreeding depression. Low performance of F_{ROH} from large segments (>8 Mb) is most likely due to the inability of large segments to explain older relatedness, which is still explained by the pedigree. The efficiency of F_{ROH} from segments of specific size will depend on the age of inbreeding in the population of interest. Very small segments were also bad in estimating inbreeding depression. This is likely due to the inability of SNP50 Beadchip v1 (Illumina Inc., San Diego, CA) to precisely estimate autozygosity. A denser SNP chip could improve performance of short (1 Mb) segments (Ferenčaković *et al.* 2013b).

Table 3 *Effect of increase in inbreeding (F) on transformed value of total number of spermatozoa*

Inbreeding coefficient	Regression coefficient	SE	AIC	$_{\Delta}$AIC
F_{PED}	**-5.8047	**1.9367**	**43907.5**	**0.0**
$F_{ROH2\text{-}4\ Mb}$	**-11.4029	**3.9858**	**43908.3**	**0.8**
$F_{ROH>2\ Mb}$	**-2.7150	**1.2153**	**43911.4**	**3.9**
$F_{ROH8\text{-}16\ Mb}$	-5.8106	2.9330	43912.5	5.0
$F_{ROH>1\ Mb}$	-2.7131	1.2204	43913.2	5.7
$F_{ROH>4\ Mb}$	-1.9121	1.3082	43914.0	6.5
F_{PED5}	-2.8128	2.0251	43914.5	7.0
$F_{ROH>8\ Mb}$	-1.9073	1.5079	43914.8	7.3
$F_{ROH1\text{-}4\ Mb}$	-3.2766	2.6573	43914.9	7.4
$F_{ROH1\text{-}2\ Mb}$	4.1089	4.10242	43915.4	7.9
$F_{ROH4\text{-}8\ Mb}$	-3.3663	3.4881	43915.5	8.0
F_{is}	-0.4858	0.7704	43916.0	8.5
$F_{ROH>16\ Mb}$	-0.7760	2.1609	43916.3	8.8

SE, standard error; Significance levels: *p < 0.05; **p < 0.01

Detection of the regions of the genome affected by autozygosity

The analysis of genomic regions associated with inbreeding depression was performed for every autosomal SNP. A total of 24 significant signals associated with the total number of spermatozoa were observed on chromosomes 7, 10, 17, 20, 22 and 27. Annotated genes in the vicinity of those signals (±1 Mb from the signal) are presented in Table 4. Among the 41 genes found, five promising candidate genes were observed. On chromosome 10 *Ribosomal Protein L10-Like (RPL10L)* gene may play a role in compensating for the inactivated X-linked gene during spermatogenesis (http://www.uniprot.org). On chromosome 17 *Solute Carrier Family 25 (Mitochondrial Carrier; Adenine Nucleotide Translocator), Member 31* also known as *Sperm Flagellar Energy Carrier Protein (SLC25A31)* gene catalyzes the

exchange of cytoplasmic adenosine diphosphate (ADP) with mitochondrial adenosine triphosphate (ATP) across the mitochondrial inner membrane. It may serve to mediate energy generating and energy consuming processes in the distal flagellum, possibly as a nucleotide shuttle between flagellar glycolysis, protein phosphorylation and mechanisms of motility (http://www.genecards.org). The signal on chromosome 20 was close to the *Cadherin 18, Type 2 (CDH18)* gene which encodes a type II classical cadherin from the cadherin superfamily of integral membrane proteins that mediate calcium-dependent cell-cell adhesion. In study of Pacheco *et al.* (2011) this gene had significant influence on motility of spermatozoa. The signal on chromosome 22 revealed many genes from which *Nuclear Receptor Subfamily 2, Group C, Member 2 (NR2C2)* is also known as *Testicular Nuclear Receptor 4* and it is required for normal spermatogenesis (http://www.uniprot.org). Finally, on chromosome 27, the signal was in vicinity of *Potassium Channel, Subfamily U, Member 1* (*KCNU1)* gene. This gene codes Testis-specific potassium channel activated by both intracellular pH and membrane voltage that mediates export of K (+) and therefore it may represent the primary spermatozoa K(+) current and is critical for fertility. It also may play an important role in sperm osmoregulation required for the acquisition of normal morphology and motility when faced with osmotic challenges, such as those experienced after mixing with seminal fluid and entry into the vagina (http://www.uniprot.org).

The results obtained allowed dissection of inbreeding effects on SNP level and proved ROH as a suitable method for finding autozygous regions affecting bull fertility. Further investigation of these regions, using a denser SNP chip or even next generation sequence data could help in better understanding the molecular background of inbreeding depression and male fertility.

Table 4 *Genes present in the chromosome (Chr) regions where significant effect of SNP on transformed number of spermatozoa was detected*

Chr	SNPs (position in bp)	Genes in the region	Gene start (bp)	Gene end (bp)	Role of the genes in biological process*
7	ARS-BFGL-NGS-70114 (100787445)	*RGMB*	100424425	100446937	Development of nervous system
		CHD1	100492794	100565687	ATP binding, chromatin remodeling,
10	BTA-20229-no-rs (39486849), ARS-BFGL-NGS-39082 (39940420)	*CCNDBP1*	38507733	38518176	Cell cycle
		EPB42	38519703	38538816	Protein glutamine gamma-glutamyltransferase activity
		RPL10L	39735371	39736135	Spermatogenesis
		MDGA2	39914871	40445335	Pattern specification process
17	BTB-00101072 (28653636), BTB-00101173 (28673548), BTB-00101401 (28779152), BTA-47941-no-rs (28816998), Hapmap38834-BTA-40794 (29018684), ARS-BFGL-NGS-24203 (29318680), ARS-BFGL-NGS-6317 (29398648), ARS-BFGL-NGS-30118 (29440222)	*SCLT1*	29190572	29354595	Centriole, extracellular vesicular exosome
		JADE1	29368416	29421827	Apoptotic process, regulation of transcription,
		COMMD6	29777299	297775656	Cytoplasm, inhibits TNF-induced NFKB1 activation
		PGRMC2	29872406	29890856	Membrane receptor
		LARP1B	29938416	30073786	Poly(A) RNA binding
		MFSD8	30144105	30181831	Transmembrane transport
		PLK4	30185756	30202777	Serine/threonine protein kinase, centriole replication
		HSPA4L	30228693	30285560	ATP binding
		SLC25A31	30291318	30319495	Transmembrane transport, sperm flagellar energy carrier
		INTU	30324842	30404702	Ciliogenesis, embryonic development
20	BTB-01281598 (54965589)	*CDH18*	53409876	54014021	Calcium ion binding
22	ARS-BFGL-NGS-106885	*PLXND1*	56774467	56774467	Semaphorin receptor activity, important role in cell-cell

	(57730808)				signaling, and in regulating the migration of a wide spectrum of cell types
		H1FOO	56829078	56860068	Role in the control of gene expression during oogenesis and early embryogenesis
		RHO	56867887	56872817	Photoreceptor required for image-forming vision at low light intensity
		IFT122	56881079	56948817	Required for cilia formation during neuronal patterning
		MBD4	56949105	56958155	Base-excision repair
		CAND2	56990323	57017221	Role in the cellular repertoire of SCF complexes
		TMEM40	57027958	57060239	transmembrane protein
		RAF1	57122412	57204951	Role as a switch determining cell fate decisions including proliferation, differentiation, apoptosis, survival and oncogenic transformation.
		MKRN2	57210750	57236946	Poly(A) RNA binding
		SYN2	57571207	57625307	Encode neuronal phosphoproteins
		ZFYVE20	57836761	57857029	Role in the lysosomal trafficking of CTSD/cathepsin D from the Golgi to lysosomes
		MRPS25	57865203	57909688	Structural constituent of ribosome
		NR2C2	57916365	57951507	Required for normal spermatogenesis
		FGD5	58037191	58148608	May play a role in regulating the actin cytoskeleton and cell shape
		SLC6A6	58448092	58494800	Sodium-dependent taurine and beta-alanine transporter
		LSM3	58693672	58701742	Ribonucleoprotein, important for pre-mRNA splicing
		XPC	58702024	58724970	Damaged DNA binding
27	Hapmap22787-BTA-103450 (32184354), ARS-BFGL-NGS-39651 (32230922), ARS-BFGL-NGS-17420	*KCNU1*	31849127	31993694	Testis specific potassium channel. Critical for fertility
		ZNF703	32683065	32685824	Regulation of cell adhesion, migration and proliferation
		ERLIN2	32721072	32738208	ER-associated ubiquitin –dependent protein catabolic process
		PROSC	32746961	32756493	Pyridoxal phosphate binding
		GPR124	32774355	32797592	G-protein coupled receptor activity

(32281266),	*BRF2*	32799190	32803514	General activator of RNA polymerase III transcription
Hapmap40574-BTA-68418 (32303544),	*RAB11FIP1*	32823609	32863935	Coding Rab11 effector protein responsible for endosomal recycling process
ARS-BFGL-NGS-23068 (32328719),	*GOT1L1*	32885370	32890975	Coding putative aspartate aminotransferase, catalytic activity L-aspartate + 2-oxoglutarate = oxaloacetate + L-glutamate
ARS-BFGL-NGS-119875	*ADRB3*	32912525	32915668	Regulation of lipolysis and thermogenesis
(32405436),	*EIF4EBP1*	32951594	32973435	Mediates the regulation of protein translation
BTA-119240-no-rs	*ASH2L*	32989769	33014982	Histone methyltransferase activity, response to estrogen
(32488843),	*STAR*	33016930	33024353	Key role in in steroid hormone synthesis
BTB-01759195	*LSM1*	33042449	33051041	Role in replication dependent histone mRNA degradation
(32513708), ARS-BFGL-NGS-112547	*DDHD2*	33110521	33140951	Membrane trafficking between the endoplasmatic reticulum and Golgi body
(32541258), UA-IFASA-1808	*FGFR1*	33250534	33291989	Essential role in the regulation of embryonic development, cell proliferation, differentiation and migration.
(32561963), ARS-BFGL-NGS-111566 (32671451),	*TACC1*	33597619	33659545	Promotion of cell division prior to the formation of differentiated tissues

*Gene function described in UniProt (http://www.uniprot.org) and/or GeneCards (http://www.genecards.org)

Conclusions

F_{ROH} performed equally well as F_{PED} for detection of inbreeding depression on total number of spermatozoa. Using ROH length categories we can estimate the age of inbreeding causing inbreeding depression to be ≈ 25 to 12.5 generations from common ancestor. The detection of the homozygous regions responsible for the decline in the total number of spermatozoa yielded in 24 significant signals in 6 regions on chromosomes 7, 10, 17, 20, 22 and 27. In total 41 gene was detected from which *Ribosomal Protein L10-Like (RPL10L), Sperm Flagellar Energy Carrier Protein (SLC25A31), Cadherin 18, Type 2 (CDH18), Testicular Nuclear Receptor 4 (NR2C2)* and *Potassium Channel, Subfamily U, Member 1* (*KCNU1)* are the most promising candidates. Use of next generation sequence data for these genes for subsets of bulls included in this study with the gene being in ROH and not in ROH is future analysis may reveal causative mutations for deficiencies in sperm quality. Regions of significant signals should be further explored in order to examine the possible involvement of other detected genes on sperm quality.

References

Akaike H. (1974) A new look at the statistical model identification. *Automatic Control, IEEE Transactions on* **19**, 716-23.

Aurich C., Achmann R. & Aurich J.E. (2003) Semen parameters and level of microsatellite heterozygosity in Noriker draught horse stallions. *Theriogenology* **60**, 371-8.

Ayroles J.F., Hughes K.A., Rowe K.C., Reedy M.M., Rodriguez-Zas S.L., Drnevich J.M., Cáceres C.E. & Paige K.N. (2009) A genomewide assessment of inbreeding depression: gene number, function, and mode of action. *Conservation Biology* **23**, 920-30.

Bjelland D., Weigel K., Vukasinovic N. & Nkrumah J. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science* **96**, 4697 - 706.

Box G.E. & Cox D.R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-52.

Burnham K.P. & Anderson D.R. (2002) *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer Science & Business Media, New York 2002.

Carr D.E. & Dudash M.R. (2003) Recent approaches into the genetic basis of inbreeding depression in plants. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **358**, 1071-84.

Charlesworth B. & Charlesworth D. (1999) The genetic basis of inbreeding depression. *Genetics Research* **74**, 329 - 40.

Charlesworth D. & Willis J. (2009) The genetics of inbreeding depression. *Nature Reviews Genetics* **10**, 783 - 96.

Curik I., Sölkner J. & Stipic N. (2001) The influence of selection and epistasis on inbreeding depression estimates. *Journal of Animal Breeding & Genetics* **118**, 247-62.

Curik I., Solkner J. & Stipic N. (2002) Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. *Journal of Animal Breeding & Genetics* **119**, 101-15.

Darwin C. (1876) *The Effects of Cross and Self Fertilisation in the Vegetable Kingdom*. London: J. Murray, UK, 1876.

Davenport C.B. (1908) Degeneration, albinism and inbreeding. *Science (New York, NY)* **28**, 454.

East E.M. (1908) Inbreeding in corn. *Connecticut Agricultural Experiments Station Report* **1907**, 419-28.

Falconer D. & Mackay T. (1996) Introduction to Quantitative Genetics. Harlow: Longman Group Ltd.

Ferenčaković M., Hamzic E., Gredler B., Curik I. & Solkner J. (2011) Runs of homozygosity reveal genome-wideautozygosity in the Austrian fleckvieh cattle. *Agriculturae Conspectus Scientificus* **76**, 325 - 8.

Ferenčaković M., Hamzić E., Gredler B., Solberg T.R., Klemetsdal G., Curik I. & Sölkner J. (2013a) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding & Genetics* **130**, 286-93.

Ferenčaković M., Solkner J. & Curik I. (2013b) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics Selection Evolution* **45**, 42.

Fuerst-Waltl B., Schwarzenbacher H., Perner C. & Sölkner J. (2006) Effects of age and environmental factors on semen production and semen quality of Austrian Simmental bulls. *Animal Reproduction Science* **95**, 27-37.

Gao X., Starmer J. & Martin E.R. (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology* **32**, 361-9.

Gibson J., Morton N. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789-95.

Gutierrez J. & Goyache F. (2005) A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding & Genetics* **122**, 172-6.

Holm S. (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**, 65-70.

Hughes K.A. (1995) The inbreeding decline and average dominance of genes affecting male life-history characters in Drosophila melanogaster. *Genetical research* **65**, 41-52.

Karkkainen K., Kuittinen H., van Treuren R., Vogl C., Oikarinen S. & Savolainen O. (1999) Genetic basis of inbreeding depression in *Arabis petraea*. *Evolution*, 1354-65.

Keller L. & Waller D. (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution* **17**, 230-41.

Keller M., Visscher P. & Goddard M. (2011) Quantification of Inbreeding Due to Distant Ancestors and its Detection Using Dense SNP Data. *Genetics* **189**, 237-49.

Kristensen T. & Sorensen A. (2005) Inbreeding - lessons from animal breeding, evolutionary biology and conservation genetics. *Animal Science* **80**, 121-33.

Li Z.-K., Luo L., Mei H., Wang D., Shu Q., Tabien R., Zhong D., Ying C., Stansel J. & Khush G. (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. *Genetics* **158**, 1737-53.

Lynch A. & Walsh B. (1998) *Genetics and Analysis of Quantitative Traits. Vol. 1*. Sunderland: Sinauer Associates, Massachusetts 1998.

Mackay T.F. (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics* **35**, 303-39.

Margulis S.W. & Walsh A. (2002) The effects of inbreeding on testicular sperm concentration in Peromyscus polionotus. *Reproduction, Fertility and Development* **14**, 63-7.

Maximini L., Fuerst-Waltl B., Gredler B. & Baumung R. (2011) Inbreeding Depression on Semen Quality in Austrian Dual-Purpose Simmental Bulls. *Reproduction in Domestic Animals* **46**, e102-e4.

McQuillan R., Leutenegger A., Abdel-Rahman R., Franklin C., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A., Farrington S.,

Rudan P., Hayward C., Vitart V., Rudan I., Wild S., Dunlop M., Wright A., Campbell H. & Wilson J. (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics* **83**, 359-72.

Nalls M., Guerreiro R., Simon-Sanchez J., Bras J., Traynor B., Gibbs J., Launer L., Hardy J. & Singleton A. (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics* **10**, 183-90.

Oliehoek P.A., Windig J.J., Van Arendonk J.A. & Bijma P. (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. *Genetics* **173**, 483-96.

Pacheco S.E., Houseman E.A., Christensen B.C., Marsit C.J., Kelsey K.T., Sigman M. & Boekelheide K. (2011) Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm. *PLoS ONE* **6**, e20280.

Pryce J.E., Haile-Mariam M., Goddard M.E. & Hayes B.J. (2014) Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution* **46**, 71.

Purfield D., Berry D., McParland S. & Bradley D. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* **13**, 70.

Ritland K. (1996) Inferring the genetic basis of inbreeding depression in plants. *Genome* **39**, 1-8.

Roff D.A. (2002) Inbreeding depression: tests of the overdominance and partial dominance hypotheses. *Evolution* **56**, 768-75.

Sargolzaei M., Iwaisaki H. & Colleau J. (2006) CFC: a tool for monitoring genetic diversity. In: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Minas Gerais, Brazil, 13-18 August, 2006.*, pp. 27-8. Instituto Prociência.

SAS Institute (2011) *What's New in SAS 9.3*. SAS Institute.

Smith B.A., Brinks J.S. & Richardson G.V. (1989) Relationships of Sire Scrotal Circumference to Offspring Reproduction and Growth. *Journal of Animal Science* **67**, 2881-5.

Templeton A. & Read B. (1984) Factors eliminating inbreeding depression in a captive herd of Speke's gazelle. *Zoo Biology* **3**, 177-99.

Templeton A.R., Read B., (1994) Inbreeding: One Word, Several Meanings, Much Confusion. In: Loeschcke, V., Tomiuk, J., Jain, S.K., (Eds.), *Conservation Genetics*. Birkhäuser–Verlag, Basel, Switzerland, pp. 91-106

Van Eldik P., Van Der Waaij E., Ducro B., Kooper A., Stout T. & Colenbrander B. (2006) Possible negative effects of inbreeding on semen quality in Shetland pony stallions. *Theriogenology* **65**, 1159-70.

Wildt D., Baas E., Chakraborty P., Wolfle T. & Stewart A. (1982) Influence of inbreeding on reproductive performance, ejaculate quality and testicular volume in the dog. *Theriogenology* **17**, 445-52.

Wright S. (1922) Coefficients of inbreeding and relationship. *The American Naturalist* **56**, 330-8.

CHAPTER 6 GENERAL DISCUSSION

THE INFLUENCE OF SNP CHIP DENSITY AND GENOTYPING ERRORS ON ROH DETECTION: AN OFTEN OVERLOOKED ISSUE

The emergence of next-generation sequencing has provided the technological basis for genotyping numerous loci (SNPs), which arouse interest in developing molecular measures of inbreeding. Those measures overcome problems like missing pedigree, pedigree errors, stochastic nature of recombination, selection, relatedness in founder population, etc., that are often connected with inbreeding coefficients from pedigree data. Runs of homozygosity are proven to be an elegant solution for these problems (Keller *et al.* 2011), however, so far the precise definition is missing and was not systematically checked (Ku *et al.* 2010). In the study of Howrigan *et al.* (2011), the authors approached ROH segments only as a number of homozygous SNPs, without the lengths of genome, and they performed linkage disequilibrium (LD) cleaning on initial data and some other steps in quality control of the data which led to loss of information on true autozygosity. Purfield *et al.* (2012) in their study addressed the influence of SNP chip density on ROH determination, however, again their quality control most likely removed part of information. Another issue is use of software for ROH detection. Most of authors for detection use PLINK (Purcell *et al.* 2007), but unfortunately algorithm for ROH detection in this software cannot properly control the number of heterozygous calls nor number of missing calls allowed by the user, leading to spurious results.

In this thesis, the effects of approaches used for ROH determination, SNP chip density and genotyping errors (described as number of heterozygous calls allowed in ROH) were investigated. Software can introduce serious bias in determination of ROH segments. The PLINK and SNP & Variation Suite (SVS) (Golden Helix, Bozeman, MT, USA www.goldenhelix. com) were analyzed and compared in this thesis. Because of dissimilarity of algorithms for detection of ROH segments, the differences can be quite large and can influence observed chromosomal levels of autozygosity. Moreover, in this research, an additional question was raised. Neither PLINK nor SVS take into account the appearance of adjacent heterozygous SNPs or heterozygous SNPs lying close to each other in ROH segment. Such events are less likely to reflect sequencing errors and more likely to wrongly suggest that the region is actually heterozygous. Adjacent or very close heterozygous calls inside ROH segment should split the segment in two.

The comparison of SNPs chips of different density revealed that 50 k panel overestimate the numbers of segments 1–4 Mb long, suggesting that it is not sensitive enough for the precise determination of small segments. Purfield *et al.* (2012) reported similar conclusion, that is, that the 50 k panel recognizes only segments longer than 5 Mb equally well as the HD panel. This can also be seen in this study from distributions of ROH lengths from both panels (Chapter 4 Figure 2). The differences were largest for [1,2] length category, and then gradually disappeared as ROH length increased. These provide further evidence that data from the 50 k panel lead to imprecise determination of short ROH and overestimation of F_{ROH}. Along with overestimation of F_{ROH}, the use of 50 k panel did not reveal ROH regions in the genome that are very usual within population (Chapter 4 Figure 5). The ROH were not uniformly distributed across the genome and they have higher frequencies in some regions. These regions were called ROH islands by Nothnagel *et al.* (2010) and ROH hotspots by Pemberton *et al.* (2012). Conversely, ROH frequencies are rare in so-called ROH deserts or coldspots. In European human populations, chromosomes 3, 4 and 14 were found to contain abundance of ROH (Nothnagel *et al.* 2010). When Pemberton *et al.* (2012) analyzed ROH patterns in 64 populations worldwide, they found distinct continental patterns. The two sets of studies overlapped in identifying hot spots on chromosomes 4 and 10, and these cannot be explained by linkage disequilibrium or local recombination alone. Many such regions harbor genes known to been affected by selection, and some of these genes have even become fixed. In contrast to ROH hotspots, coldspots are likely to be conserved regions associated with a critical physiological function (Pemberton *et al.* 2012). Why these hotspots and coldspots occur among cattle breeds and within other animal species are questions currently of interest (Sölkner *et al*. 2014).

In this thesis, the number of heterozygous calls allowed proved to influence F_{ROH}. In long ROH (which can have more than 5000 to 6000 SNPs) some heterozygous calls must be allowed due to the possibility of genotyping errors, however, the number of allowable heterozygous calls should be limited. On the one hand, SNP data from chromosome 20 in the 46 Brown Swiss cattle (Chapter 4 Figure 4) shows clearly that single, potentially miscalled heterozygous SNPs would interrupt ROH segments if such SNPs were not allowed.

To achieve better precision and make ROH studies comparable, algorithms for ROH detection should be improved and definition of ROH length should be systematically researched and standardized.

ESTIMATES OF AUTOZYGOSITY DERIVED FROM RUNS OF HOMOZYGOSITY: A BETTER WAY TO ESTIMATE INBREEDING

ROH is the method of choice for estimation of inbreeding coefficients in human populations. However, suitability of this method is still not investigated in cattle populations. In this thesis, animals from four cattle breeds with different inbreeding backgrounds were analyzed in order to compare levels of autozygosity derived from ROH (F_{ROH}) and from pedigree records (F_{PED}). The correlations of F_{ROH} and F_{PED} were moderate to relatively high, indicating that F_{ROH} is suitable for measuring individual levels of inbreeding. Within breed, $F_{ROH >1 Mb}$, $F_{ROH >2 Mb}$, $F_{ROH >8 Mb}$ gave similar correlations with F_{PED} (Chapter 3 Table 1). In breeds with deeper pedigrees (Norwegian Red and Brown Swiss), there was a decrease in correlations for $F_{ROH >16 Mb}$.

Correlations between F_{ROH} from different lengths are linked with depth of pedigree. Overall, correlations of F_{ROH} estimates based on ROH of different lengths with F_{PED} or F_{PED5} did not differ substantially. VanRaden (2008) reported the correlations between the estimates of inbreeding levels based on SNP variance and the estimates based on pedigrees. Applying the proposed methods of VanRaden (2008) and Yang *et al.* (2010), Sölkner *et al.* (2010) reported much lower correlations between inbreeding levels based on SNP variance and F_{PED} for Fleckvieh cattle than those reported by VanRaden (2008), while correlations of F_{ROH} and F_{PED} were similar to those in this thesis. A study by McQuillan *et al.* 2008 on the population of Orkney Islands reported a correlation of $r = 0.86$ between inbreeding estimates based on the proportion of ROH longer than 1.5 Mb and estimates from pedigrees. That correlation is considerably higher than that of F_{PED} or F_{PED5} with estimates presented in this thesis and based on ROH in similar length categories ($F_{ROH >1 Mb}$, $F_{ROH >2 Mb}$). The strongest correlation was estimated for $F_{ROH >1 Mb}$ in Tyrol Grey ($F_{PED} = 0.71$, $F_{PED5} = 0.71$), while the lowest was in Norwegian Red ($F_{PED} = 0.61$, $F_{PED5} = 0.50$). The different estimates may be attributed to differences in population structure: F_{PED} reflects recent inbreeding, and inbreeding coefficients based on ROH can capture both recent and distant inbreeding.

The ROH 2–4 Mb long (25–12.5 generations from common ancestor) corresponds mostly to identical by descent (IBD) segments from the past that usually could not be captured with available pedigree information (CGE from 7.3 to 9.0), although they may also contain some ROH that are identical by state (IBS) without being IBD In contrast, ROH >8 Mb long are more likely to be autozygous segments of recent origin and are extremely unlikely to be non-IBD. F_{PED} also does not account for stochastic nature of recombination,

while F_{ROH} is sensitive to it (Keller *et al.* 2011). The range of values from $F_{ROH\ >8\ Mb}$ for groups of animals with similar F_{PED} clearly shows advantage of using F_{ROH}.

The studies on outbred human populations reported co-occurrence of ROH in chromosome regions with extended linkage disequilibrium and low recombination rates (Gibson *et al.* 2006; Curtis *et al.* 2008). They conclude that common extended haplotypes may partly contribute to high F_{ROH} estimates based on shorter ROH. Kirin *et al.* (2010) used minimum length of ROH of 500 kb to avoid very short ROH that can occur due to LD. In this thesis for the same reason, the minimum of 1 Mb was used, because it is known that cattle have longer range LD.

RUNS OF HOMOZYGOSITY AS A TOOL FOR INDICATING REGIONS INFLUENCED BY INBREEDING

The use of artificial insemination (AI) in cattle breeding programs resulted in fewer bulls producing larger numbers of offspring. In such systems, frequent usage of high valued bulls leads to reduction of genetic variability, and to inbreeding and inbreeding depression. In bulls, low concentration of spermatozoa, low volume of sperm, low progressive motility and low number of alive spermatozoa, are considered to be under influence of inbreeding depression. Maximini *et al.* (2011) using pedigree and sperm quality data reported that inbreeding depression affected male fertility in Simmental bulls.

In this thesis the influence of inbreeding on total number of spermatozoa using pedigree inbreeding coefficient was confirmed. However, a significant influence was found only when minimum length of a segment was set to 2 Mb. It was expected that different ROH lengths will yield in different effects because length also discovers age of inbreeding (Keller *et al.* 2011). The results presented in this study sugested that this particular influence originates 25 to 12.5 generations ago, because this is the age of inbreeding expected from segments between 2 to 4 Mb.

The usage of ROH status was explored as potential tool for pinpointing the genome regions that influence quantitative traits of interest. The genomic regions associated with inbreeding depression were shown in 24 significant signals on chromosomes 7, 10, 17, 20, 22 and 27. In total, 41 genes were detected, of which *Ribosomal Protein L10-Like (RPL10L), Sperm Flagellar Energy Carrier Protein (SLC25A31), Cadherin 18, Type 2 (CDH18), Testicular Nuclear Receptor 4 (NR2C2)* and *Potassium Channel, Subfamily U, Member 1*

(*KCNU1)* were very promising candidates. Thus, it can be confirmed that the inbreeding effects on bull semen quality were detected at SNP level.

References

Bjelland D., Weigel K., Vukasinovic N. & Nkrumah J. (2013) Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *Journal of Dairy Science* **96**, 4697 - 706.

Curik I., Solkner J. & Stipic N. (2002) Effects of models with finite loci, selection, dominance, epistasis and linkage on inbreeding coefficients based on pedigree and genotypic information. *Journal of Animal Breeding & Genetics* **119**, 101-15.

Curtis D., Vine A. & Knight J. (2008) Study of regions of extended homozygosity provides a powerful method to explore haplotype structure of human populations. *Annals of Human Genetics* **72**, 261-78.

Gibson J., Morton N. & Collins A. (2006) Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**, 789 - 95.

Howrigan D., Simonson M. & Keller M. (2011) Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* **12**, 460.

Keller M., Visscher P. & Goddard M. (2011) Quantification of Inbreeding Due to Distant Ancestors and its Detection Using Dense SNP Data. *Genetics* **189**, 237 - 49.

Kirin M., McQuillan R., Franklin C., Campbell H., McKeigue P. & Wilson J. (2010) Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **5**, e13996.

Ku C., Naidoo N., Teo S. & Pawitan Y. (2010) Regions of homozygosity and their impact on complex diseases and traits. *Human Genetics* **129**, 1 - 15.

Maximini L., Fuerst-Waltl B., Gredler B. & Baumung R. (2011) Inbreeding Depression on Semen Quality in Austrian Dual-Purpose Simmental Bulls. *Reproduction in Domestic Animals* **46**, e102-e4.

McQuillan R., Leutenegger A., Abdel-Rahman R., Franklin C., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A., Farrington S., Rudan P., Hayward C., Vitart V., Rudan I., Wild S., Dunlop M., Wright A., Campbell H. & Wilson J. (2008) Runs of homozygosity in European populations. *American Journal of Human Genetics* **83**, 359 - 72.

Nalls M., Guerreiro R., Simon-Sanchez J., Bras J., Traynor B., Gibbs J., Launer L., Hardy J. & Singleton A. (2009) Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics* **10**, 183 - 90.

Nothnagel M., Lu T.T., Kayser M. & Krawczak M. (2010) Genomic and geographic distribution of SNP-defined runs of homozygosity in Europeans. *Human Molecular Genetics* **19**, 2927-35.

Pemberton T., Absher D., Feldman M., Myers R., Rosenberg N. & Li J. (2012) Genomic patterns of homozygosity in worldwide human populations. *American Journal of Human Genetics* **91**, 275 - 92.

Pryce J.E., Haile-Mariam M., Goddard M.E. & Hayes B.J. (2014) Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genetics Selection Evolution* **46**, 71.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M., Bender D., Maller J., Sklar P., De Bakker P., Daly M. & Sham P. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559 - 75.

Purfield D., Berry D., McParland S. & Bradley D. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* **13**, 70.

Sölkner J., Ferencakovic M., Gredler B. & Curik I. (2010) Genomic metrics of individual autozygosity, applied to a cattle population. In: *Proceedings of the 61st Annual Meeting of the European Association of Animal Production, Heraklion, Crete Island, Greece*.

Sölkner, J., Ferenčaković, M., Karimi, Z., Perez O'Brien, A.M., Mészáros, G., Eaglen, S., Boison, S.A., & Curik, I. (2014). Extremely Non-uniform: Patterns of Runs of Homozygosity in Bovine Populations. In: *Proceedings - 10th World Congress on Genetics Applied to Livestock Production / American Society of Animal Science (ed) Champaign; 10th World Congress on Genetics Applied to Livestock Production.* 1-4.

VanRaden P. (2008) Efficient methods to compute genomic predictions. *Journal Dairy Science* **91**, 4414 - 23.

Yang J., Benyamin B., McEvoy B., Gordon S., Henders A., Nyholt D., Madden P., Heath A., Martin N. & Montgomery G. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42**, 565 - 9.

CHAPTER 7 GENERAL CONCLUSIONS

The data from the 50 k bovine SNP panel do not present true state of autozygosity for short ROH segments, while they are as reliable as the HD panel data for ROH > 4 Mb. When shorter segments are included with the 50 k panel, F_{ROH} is systematically overestimated. The errors due to potential genotyping errors depend on allowing of heterozygous genotypes in a ROH detection. While not allowing for heterozygous calls often splits a very long ROH in two shorter ones that are still recognized and therefore the level of autozygosity of an individual is virtually unaffected, there are many cases where the shorter part of the split does not reach the minimum size of a ROH and the level of autozygosity of an individual is underestimated. This thesis showed that the identification of ROH should be done in the following steps: (1) Reduction of genotyping errors should be performed by removing SNPs using strict limits on genotype quality scores provided. (2) Number of heterozygous SNPs allowed should be determined separately for each ROH length of interest and for each SNP density. (3) If multiple heterozygous SNPs are allowed within the same ROH, adjacent heterozygous SNPs should be treated differently from heterozygous SNPs that are further apart.

The observational approach of ROH, in contrast to the probabilistic approach of pedigree analysis, which does not take stochastic variations into account, gives more precise estimate of levels of autozygosity. Performing analyses with ROH of different lengths allows estimation of the distance of the current population from the base population.

Runs of homozygosity are a good approach for estimation of inbreeding depression on total number of spermatozoa. ROH inbreeding coefficients calculated from ROH segments between 2 and 4 Mb, and segments >2 Mb showed significant effect on number of spermatozoa in bulls. Assuming that mean length of IBD segments equals $100/(2\ gcA)$ cM, where gcA is the number of generations from the common ancestor, inbreeding causing decrease in the mean value of the trait originates from 25 to 12.5 generations from the common ancestor.

Six genomic regions on chromosomes 7, 10, 17, 20, 22 and 27 containing 41 genes were shown to influence the total number of spermatozoa in Simmental bulls. Five of those genes are already known to be directly associated with spermatogenesis, energy levels in spermatozoa and osmotic balance of the sperm.

Results of this thesis supported the proposed hypotheses, leading to the following final conclusions:

(1) Genotyping errors and SNP chip density does affect estimates of autozygosity from ROH.

(2) Knowledge about ROH distribution (number and size) allows a precise estimation of autozygosity at individual.

(3) Population inbreeding levels in cattle and genomic autozygosity does have influence on bull semen quality.

(4) ROH enable identification of narrow chromosomal regions where inbreeding has impact on quantitative trait

SUPPLEMENTARY MATERIAL

Copies of publications which are directly associated with this Thesis but are not part of it

Ferenčaković M., Hamzic E., Gredler B., Curik I. & Solkner J. (2011) Runs of homozygosity reveal genome-wideautozygosity in the Austrian fleckvieh cattle. *Agriculturae Conspectus Scientificus* 76, 325-8.

Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle

Maja FERENCAKOVIC [1] (✉)
Edin HAMZIC [2]
Birgit GREDLER [2]
Ino CURIK [1]
Johann SÖLKNER [2]

Summary

Runs of homozygosity (ROH) are recognized as potential inbreeding measure in studies on humans. Inbreeding coefficients derived from ROH (F_{ROH}) measure proportion of the genome arranged in long homozygous segments and highly correlate with those derived from pedigree (F_{ped}). From that we assumed that ROH represent an alternative to pedigree inbreeding levels in studies on animals too, because pedigree can be incorrect, incomplete and can not fully explain what happened in meiosis. To confirm our premise we used pedigree and genotype data from 500 Austrian dual purpose Simmental bulls to determine correlation between F_{ROH} and F_{ped}. ROH were obtained using Fortran 90 software created by the authors. Proportions of genome in ROH were calculated for lengths of ROH of >1, >2, >4, >8 and >16 Mb. Pedigree data were analyzed and inbreeding coefficients for complete pedigree (F_{pedT}) and five generations (F_{ped5}) were calculated using ENDOG software. We found low F_{pedT} and F_{ped5} (means of 1.5% and 0.9%) while F_{ROH} for segments >1Mb suggested much higher values (9.0%) indicating old inbreeding that can not be traced using pedigree. The highest correlations were found between F_{ROH} calculated from ROH of length >4Mb and F_{pedT} (0.68) that is consistent with studies on humans. We conclude that inbreeding coefficients derived from ROH are useful for measuring levels of inbreeding in cattle, because ROH are not subject to mistakes as pedigrees and calculations made from those.

Key words

inbreeding, runs of homozygosity, genome-wide autozygosity, pedigree, cattle

[1] University of Zagreb, Faculty of Agriculture, Department of Animal Science, Svetošimunska 25, 10000 Zagreb, Croatia
✉ e-mail: mferencakovic@agr.hr
[2] University of Natural Resources and Life Sciences Vienna, Department of Sustainable Agricultural Systems, Division of Livestock Sciences, Gregor Mendel Str. 33, A-1180 Vienna, Austria

Received: June 1, 2011 | Accepted: July 14, 2011

ACKNOWLEDGEMENTS

This research was partially funded by Scientific project of Ministry of Science Education and Sports no. 178-1780460-0546 (Strategies for elimination of genetic defects in selected populations). We gladly acknowledge support in part by an "Ernst Mach grant - worldwide" financed by the Austrian Federal Ministry of Education and Science. We are very grateful to Zuchtdata EDV-Dienstleistungen GmbH and the Association of Austrian Simmental breeders (AGÖF) for provision of genotype and pedigree data.

Introduction

Mating of related individuals results with inbred offspring that are generally less viable, less fertile or/and smaller than the population mean. The phenomenon is also known as inbreeding depression and occurs regularly in animal and plant breeding, in small natural populations (Pirchner, 1985; Charlesworth and Charlesworth, 1987) and in humans (Schull and Neel, 1965). The inbreeding coefficient is a measure of inbreeding defined as the proportion of an individual's genome that is autozygous, relative to that of a poorly characterized founder generation. Since it was developed (Wright, 1922) inbreeding coefficient has mainly been estimated from the pedigree information, here denoted as F_{ped}. The advent of high throughput methods enabled genotyping of individual (animals) for a large number of molecular markers spread all over the genome, and further stimulated development of molecular measures that estimate autozygosity of an individual (Leutenegger et al., 2003; Carothers et al., 2006; Polasek et al., 2010). Runs of homozygosity (ROH) were recently proposed as a measure quantifying individual autozygosity (McQuillan et al., 2008; Nalls et al., 2009). A ROH is a continuous or uninterrupted part of genome without heterozygosity in the diploid state. As recombination interrupts long chromosome segments over the time, it is expected that long identical segments come, through the parents, from the same haplotype of their common ancestor. Furthermore, the number of segregations to the common ancestors is lower for long segments in comparison to the shorter homozygous segments.

In addition, in an inbred population we expect to find more and longer homozygous segments than in outbreed populations (Gibson et al., 2006). Human genome studies have also shown that individuals born in consanguineous unions (marriages between close relatives), have levels of homozygosity that are even higher than were expected (Woods et al., 2006; Broman and Weber, 1999) form pedigree information. Precise estimating inbreeding coefficients from pedigrees do not cover ancient relatedness and correct pedigrees. Even if the pedigree is well known and correct the estimates of inbreeding for single individuals can differ from expectation due to the stochastic pattern of inheritance. The mean inbreeding coefficient of the offspring of the first cousins is 0.0625 with a standard deviation of 0.0243 (Carothers et al., 2006). This variance increases with each meiosis, so it is possible for offspring of the third cousins to be more autozygous than offspring of second cousins (McQuillan et al., 2008). The availability of genome scan technology, which can genotype individual at large number of markers, provides us with the opportunity to "observe" levels of true inbreeding. Thus, distribution and size of ROH can provide information for calculating true individual level of autozygosity.

Aim of this study was to compare pedigree inbreeding coefficients with measures derived from ROH information for a 500 Austrian dual purpose Simmental bulls. We will also compare our results with the similar studies obtained in human populations and provide information for utility of ROH as a tool for measuring inbreeding coefficients from molecular data in cattle.

Materials and methods

Overall 1837 Austrian dual purpose Simmental, 447 Brown Swiss and 217 Tyrol Grey bulls were genotyped using the Illumina 50K bovine SNP chip (San Diego, CA, USA). All markers with unknown position and/or chromosome assignment as well as with GC-score lower than 0.2 were removed before preparing input files for PLINK software (Purcell et al., 2007). After the application of PLINK software (Purcell et al., 2007), by applying parameters -- mind 0.05, --maf 0.001 and --geno 0.25, 42262 markers were left for analyses. Additionally we excluded 529 SNP assigned to X chromosome. Final data set was including 41733 SNP on 29 chromosomes and they cover 2557.47 Mb of genome. For this pilot study we only used the 500 youngest Austrian Simmental bulls (born in 2001 to 2004) available in the data set.

For the analyses of Austrian Simmental population the pedigree included 41090 animals. From the pedigree data we calculated the equivalent complete generations and pedigree inbreeding coefficients referred to all (F_{pedT}) and five generation long pedigree (F_{ped5}) using ENDOG v4.8 (Gutiérrez and Goyache, 2005).

ROH were determined using Fortran 90 software developed by authors. The software simply counts homozygous SNP along chromosome and by their bp-positions providing information on length of ROH within given parameters. Depending on the minimum length of ROH in which no heterozygote SNP were allowed, we calculated ROH1, ROH2, ROH4, ROH8 and ROH16 according to the size of ROH being 1, 2, 4, 8 and 16 Mb long, respectively. Every ROH was required to have a minimum of 15 SNP. We also calculated molecular inbreeding coefficients based on ROH. Depending on the ROH size molecular inbreeding coefficients F_{ROH1}, F_{ROH2}, F_{ROH4}, F_{ROH8} and F_{ROH16} were calculated by dividing the sum distances covered by the ROH per individual by length of genome covered by SNP as described in Leutenegger et al. (2003). All statistical analyses and figures were done with SAS software v9.2 (SAS, 2009)

Results and discussion

On the population of 500 genotyped Austrian Simmental we observed average complete generation equivalent of 7.30 (±0.41; range of 5.91 to 8.32). The maximum number of generations tracked in a pedigree was 17. Descriptive statistics of the pedigree inbreeding coefficient estimations is given in Table 2. All animals (except one) were inbred for all generations period while 74.6% were inbred for five generations period. Both average pedigree inbreeding coefficients were low (up to 1.50 % and 0.9%) that was consistent with previously reported levels of inbreeding in this population (Maximini et al., 2011). Descriptive statistics of total length and number of determinate ROH in 500 Austrian Simmental bulls is given in Table 1, while descriptive statistics of molecular inbreeding coefficients calculated from ROH are given in Table 2.

ROH greater than 1Mb cover on average 9.0 % of the genome while pedigree inbreeding indicates a proportion of only 1.5 %. The similar level of autozygosity was also estimated by ROH greater than 16 Mb, thus, indicating recent inbreeding. Difference is due to "old" inbreeding that can not be traced using pedigree data but can be with short ROH. This is confirmed by the observation that correlations of F_{ROH} and inbreeding coefficients (F_{pedT} and F_{ped5}) are higher for ROH greater than 4Mb than those for F_{ROH1} and F_{ROH2} (Table 3). Studies on humans (McQuillan et al., 2008) give similar information. Low level of autozygosity from pedigree data (0.38%) was confirmed with low level of

Table 1. Descriptive statistics for the total length (in Mb) and number (in brackets) of runs of homozygosity (ROH) of different size (>1, 2, 4, 8 and 16 Mb) in 500 Austrian Simmental bulls

Total length of ROHs (Number of ROHs)	Mean	Standard deviation	Minimum	Maximum
>1 Mb	229.25 (96.79)	55.02 (13.37)	81.87 (48)	498.86 (135)
>2 Mb	139.17 (30.49)	52.20 (6.99)	24.15 (7)	419.65 (56)
>4 Mb	82.91 (9.65)	48.24 (3.99)	5.11 (1)	358.04 (29)
>8 Mb	52.86 (3.50)	42.79 (2.35)	8.09 (1)	290.70 (17)
>16 Mb	47.16 (1.89)	33.03 (1.23)	16.03 (1)	181.99 (7)

Table 2. Descriptive statistics of the pedigree and molecular, derived from runs of homozygosity (ROH) of different size (1, 2, 4, 8 and 16 Mb), inbreeding coefficients in 500 Austrian Simmental bulls

Inbreeding coefficient	Mean	Standard deviation	Minimum	Maximum
F_{pedT}	0.015	0.013	0.000	0.090
F_{ped5}	0.009	0.012	0.000	0.085
F_{ROH1}	0.090	0.022	0.032	0.195
F_{ROH2}	0.054	0.020	0.009	0.164
F_{ROH4}	0.032	0.019	0.002	0.140
F_{ROH8}	0.021	0.017	0.003	0.114
F_{ROH16}	0.018	0.013	0.006	0.071

Table 3. Correlations between pedigree and molecular, derived from runs of homozygosity (ROH) of different size (1, 2, 4, 8 and 16 Mb), inbreeding coefficients in 500 Austrian Simmental bulls

Inbreeding coefficient	F_{ROH1}	F_{ROH2}	F_{ROH4}	F_{ROH8}	F_{ROH16}	F_{pedT}
F_{ROH2}	0.969					
F_{ROH4}	0.920	0.965				
F_{ROH8}	0.866	0.914	0.950			
F_{ROH16}	0.755	0.800	0.839	0.890		
F_{pedT}	0.644	0.674	0.683	0.682	0.632	
F_{ped5}	0.613	0.648	0.663	0.671	0.625	0.973

autozygosity from ROH of minimum 5Mb (0.45%). Very small segments (minimum 0.5Mb), which could be observation of very old inbreeding give higher values (3.9%). The highest correlation between pedigree inbreeding coefficient and those derived from ROH in Orcadian is for length of ROH of minimum 1.5 Mb and it is 0.86. In our case of somewhat more inbred animals genome is covered with larger autozygous segments (4-8 Mb).

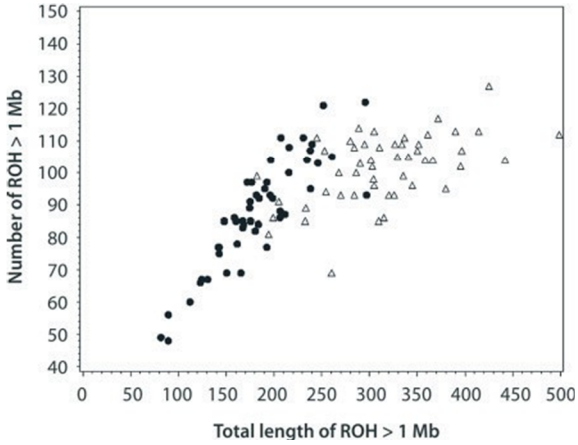


Figure 1. Relationship between number and total length of ROH, runs of homozigosity (ROH) greater than 1 Mb, in 50 animals with low (black dots) and in 50 animals with high (white triangles) total pedigree inbreeding coefficient

Number and length of ROH are getting higher as inbreeding coefficient grows. In animals with high inbreeding, the number of ROH is more constant while length of genome covered with ROH is different, suggesting very large segments. This can be observed from Figure 1 where we have chosen 50 animals with lowest F_{pedT} (0 - 0.005) and 50 animals with highest F_{pedT} (0.033 – 0.09).

Conclusion

We conclude that inbreeding coefficients derived from ROH are a very useful tool for indicating levels of inbreeding, especially if pedigree data are missing or pedigrees are not correct. They are also giving a better picture about inbreeding in the ancestral population that we are usually not able to track. Information that we receive from ROH not only provides information on inbreeding level but also on its age. Using the observational approach rather than the probabilistic one applied in pedigree analysis, most likely provides more accurate information about the state of autozygosity of individuals.

References

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: 1493–1500.

Carothers A. D., Rudan I., Kolcic I., Polasek O., Hayward C., Wright A. F., Campbell H., Teague P., Hastie N. D. and Weber J. L. (2006). Estimating Human Inbreeding Coefficients: Comparison of Genealogical and Marker Heterozygosity Approaches. Ann Hum Genet 70: 666–676.

Charlesworth D. and Charlesworth B., (1987). Inbreeding depression and its evolutionary consequences. Annu Rev Ecol Syst 18: 237-268.

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

Gutiérrez J. P. and Goyache F. (2005). A note on ENDOG: a computer program for analysing pedigree information. J Anim Breed Genet 122: 172-176.

Leutenegger A.-L., Prum B., Genin E., Verny C., Lemainque A., Clerget-Darpoux F., Thompson E.A.(2003). Estimation of the inbreeding coefficient through use of genomic data. Am J Hum Genet 73: 516-523.

Maximini L., Fuerst-Waltl B., Gredler B. and Baumung R. (2011). Inbreeding Depression on Semen Quality in Austrian Dual-Purpose Simmental Bulls. Reprod Domest Anim 46: e102–e104.

McQuillan R., Leutenegger A.-L., Abdel-Rahman R., Franklin C.S., Pericic M., Barac-Lauc L., Smolej-Narancic N., Janicijevic B., Polasek O., Tenesa A., Macleod A.K., Farrington S.M., Rudan P., Hayward C., Vitart V., Rudan I., Wild S.H., Dunlop M.G., Wright A.F., Campbell H., Wilson J.F. (2008). Runs of homozygosity in European populations. Am J Hum Genet 83: 359–372.

Nalls M.A., Guerreiro R.J., Simon-Sanchez J., Bras J.T., Traynor B.J., Gibbs J.R., Launer L., Hardy J., Singleton A.B. (2009). Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. Neurogenetics 10: 183-190.

Polasek O., Hayward C., Bellenguez C., Vitart V., Kolcic I., McQuillan R., Saftic V., Gyllensten U., Wilson J.F., Rudan I., Wright A.F., Campbell H., Leutenegger A.L. (2010). Comparative assessment of methods for estimating individual genome-wide homozygosity-by-descent from human genomic data. BMC Genom 11: 139.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., Sklar P., de Bakker P.I.W., Daly M.J., Sham P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 81: 559-575.

Pirchner F. (1985). Genetic structure of population. 1. Closed populations or matings among related individuals. In: General and Quantitative Genetics (A.B. Chapman, ed.), World Animal Science A4 Elsevier, Amsterdam. 227-250.

SAS (2009). SAS/STAT® User's Guide Version 9.2. SAS Institute Inc. Cary, NC, USA.

Schull W. J., Neel J. V. (1965). The Effect of Inbreeding on Japanese Children. Harper & Row, New York, USA.

Woods C.G., Cox J., Springell K., Hampshire D.J., Mohamed M.D., McKibbin M., Stern R., Raymond F.L., Sandford R., Malik Sharif S., Karbani G., Ahmed M., Bond J., Clayton D., Inglehearn C.F. (2006). Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. Am J Hum Genet 78: 889–896.

Wright S. (1922). Coefficients of inbreeding and relationship. Am Nat 56: 330–338.

acs76_63

Zavarez L.B., Utsunomiya Y.T., Carmo A.S., Neves H.H.R., Carvalheiro R., Ferencakovic M., Pérez O'Brien A.M., Curik I., Cole J.B., Van Tassell C., Silva M.V., Sonstegard T.S., Sölkner J. & Garcia J.F. (2015) Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. *Frontiers in Genetics* **6**, 1-8.

frontiers in
GENETICS

ORIGINAL RESEARCH ARTICLE
published: 29 January 2015
doi: 10.3389/fgene.2015.00005

Assessment of autozygosity in Nellore cows (*Bos indicus*) through high-density SNP genotypes

Ludmilla B. Zavarez[1], Yuri T. Utsunomiya[1], Adriana S. Carmo[1], Haroldo H. R. Neves[2,3], Roberto Carvalheiro[3], Maja Ferenčaković[4], Ana M. Pérez O'Brien[5], Ino Curik[4], John B. Cole[6], Curtis P. Van Tassell[6], Marcos V. G. B. da Silva[7], Tad S. Sonstegard[6], Johann Sölkner[5] and José F. Garcia[1,8]*

[1] *Departamento de Medicina Veterinária Preventiva e Reprodução Animal, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Jaboticabal, São Paulo, Brazil*
[2] *GenSys Consultores Associados, Porto Alegre, Rio Grande do Sul, Brazil*
[3] *Departamento de Zootecnia, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Jaboticabal, São Paulo, Brazil*
[4] *Department of Animal Science, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia*
[5] *Division of Livestock Sciences, Department of Sustainable Agricultural Systems, BOKU - University of Natural Resources and Life Sciences, Vienna, Austria*
[6] *Animal Genomics and Improvement Laboratory, United States Department of Agriculture, Agricultural Research Service, Beltsville, MD, USA*
[7] *Bioinformatics and Animal Genomics Laboratory, Embrapa Dairy Cattle, Juiz de Fora, Minas Gerais, Brazil*
[8] *Laboratório de Bioquímica e Biologia Molecular Animal, Departamento de Apoio, Produção e Saúde Animal, Faculdade de Medicina Veterinária de Araçatuba, UNESP – Univ Estadual Paulista, Araçatuba, São Paulo, Brazil*

Edited by:
Paolo Ajmone Marsan, Università Cattolica del S. Cuore, Italy

Reviewed by:
Yniv Palti, United States Department of Agriculture, USA
Nicolo Pietro Paolo Macciotta, University of Sassari, Italy

***Correspondence:**
José F. Garcia, Laboratório de Bioquímica e Biologia Molecular Animal, Departamento de Apoio, Produção e Saúde Animal, Faculdade de Ciências Agrárias e Veterinárias, UNESP – Univ Estadual Paulista, Rua Clóvis Pestana 793, Araçatuba, São Paulo, 16050-680, Brazil
e-mail: jfgarcia@fmva.unesp.br

The use of relatively low numbers of sires in cattle breeding programs, particularly on those for carcass and weight traits in Nellore beef cattle (*Bos indicus*) in Brazil, has always raised concerns about inbreeding, which affects conservation of genetic resources and sustainability of this breed. Here, we investigated the distribution of autozygosity levels based on runs of homozygosity (ROH) in a sample of 1,278 Nellore cows, genotyped for over 777,000 SNPs. We found ROH segments larger than 10 Mb in over 70% of the samples, representing signatures most likely related to the recent massive use of few sires. However, the average genome coverage by ROH (>1 Mb) was lower than previously reported for other cattle breeds (4.58%). In spite of 99.98% of the SNPs being included within a ROH in at least one individual, only 19.37% of the markers were encompassed by common ROH, suggesting that the ongoing selection for weight, carcass and reproductive traits in this population is too recent to have produced selection signatures in the form of ROH. Three short-range highly prevalent ROH autosomal hotspots (occurring in over 50% of the samples) were observed, indicating candidate regions most likely under selection since before the foundation of Brazilian Nellore cattle. The putative signatures of selection on chromosomes 4, 7, and 12 may be involved in resistance to infectious diseases and fertility, and should be subject of future investigation.

Keywords: *Bos indicus*, runs of homozygosity, selection, cattle, fertility, disease resistance

INTRODUCTION

Autozygosity is the homozygote state of identical-by-descent alleles, which can result from several different phenomena such as genetic drift, population bottleneck, mating of close relatives, and natural and artificial selection (Falconer and Mackay, 1996; Keller et al., 2011; Curik et al., 2014). In the past 20 years, the heavy use of relatively low number of sires in Brazilian Nellore breeding programs (*Bos indicus*) is deemed to have mimicked all these triggers of autozygosity, especially considering the increasing use of artificial insemination over the decades. As inbreeding has been incriminated in reduced fitness and reproductive performance in other cattle populations under artificial selection (Bjelland et al., 2013; Leroy, 2014), avoidance of mating of close relatives is a typical practice of many Nellore breeders. Therefore, there is a growing interest in characterizing and monitoring autozygosity in this breed to preserve genetic diversity and allow the long-term sustainability of breeding programs in Brazil.

Evidence from whole-genome sequencing studies in humans indicate that highly deleterious variants are common across healthy individuals (MacArthur et al., 2012; Xue et al., 2012), and although no such systematical survey has been conducted in cattle to the present date, it is highly expected that unfavorable alleles also segregate in cattle populations. Therefore, the use of ever-smaller numbers of animals as founders is expected to inadvertently increase autozygosity of such unfavorable alleles (Szpiech et al., 2013), potentially causing economic losses.

Recently, the use of high-density single nucleotide polymorphism (SNP) genotypes to scan individual genomes for contiguous homozygous chromosomal fragments has been proposed as a proxy for the identification of identical-by-descent haplotypes (Gibson et al., 2006; Lencz et al., 2007). As the length

of autozygous chromosomal segments is proportional to the number of generations since the common ancestor (Howrigan et al., 2011), the identification of runs of homozygosity (ROH) can reveal recent and remote events of inbreeding, providing invaluable information about the genetic relationships and demographic history of domesticated cattle (Purfield et al., 2012; Ferenčaković et al., 2013a; Kim et al., 2013). Also, given the stochastic nature of recombination, the occurrence of ROH is highly heterogeneous across the genome, and hotspots of ROH across a large number of samples (hereafter referred as common ROH) may be indicative of selective pressure. Moreover, the fraction of an individual's genome covered by ROH can be used as an estimate of its genomic autozygosity or inbreeding coefficient (McQuillan et al., 2008; Curik et al., 2014).

Here, we investigated the occurrence of ROH in high-density SNP genotypes in order to characterize autozygosity in the genomes of a sample of 1,278 Nellore cows under artificial selection for weight, carcass and reproductive traits. We aimed at characterizing the distribution of ROH length and genome-wide levels of autozygosity, as well as detecting common ROH that may be implicated in past events of selection.

MATERIALS AND METHODS

ETHICAL STATEMENT

The present study was exempt of the local ethical committee evaluation as genomic DNA was extracted from stored hair samples of animals from commercial herds.

GENOTYPING AND DATA FILTERING

A total of 1,278 cows were genotyped with the Illumina® BovineHD Genotyping BeadChip assay (HD), according to the manufacturer's protocol (http://support.illumina.com/array/array_kits/bovinehd_dna_analysis_kit.html). These animals comprised part of the genomic selection reference population from a commercial breeding program. These dams were born between 1993 and 2008, being under routine genetic evaluation for weight, carcass and reproductive traits by the DeltaGen program, an alliance of Nellore cattle breeders from Brazil. Data filtering was performed using *PLINK v1.07* (Purcell et al., 2007), and markers were removed from the dataset if GenTrain score lower than 70% or a call rate lower than 98% was observed. All genotyped samples exhibited call rates greater than 90%, thus no animals were filtered from further analyses. Minor allele frequency (MAF) was not used as an exclusion criterion in this analysis, so that the detection of homozygous segments was not compromised. Both autosomal and X-linked markers were included.

ESTIMATES OF GENOMIC INDIVIDUAL AUTOZYGOSITY

Genomic autozygosity was measured based on the percentage of an individual's genome that is covered by ROH. Stretches of consecutive homozygous genotypes were identified for each animal using *SNP & Variation Suite v7.6.8* (Golden Helix, Bozeman, MT, USA http://www.goldenhelix.com), and chromosomal segments were declared ROH under the following criteria: 30 or more consecutive homozygous SNPs, a density of at least 1 SNP every 100 kb, gaps of no more than 500 kb between SNPs, and no more than 5 missing genotypes across all individuals. In order to account for genotyping error and avoid underestimation of long ROH (Ferenčaković et al., 2013b), heterozygous genotype calls were allowed under conditions where there were 2 heterozygous genotypes for ROH $\geq$ 4 Mb, or no heterozygous genotypes for ROH $<$ 4 Mb. Autozygosity was estimated according to McQuillan et al. (2008):

$$F_{ROH} = \frac{\sum_{j=1}^{n} L_{ROH_j}}{L_{total}}$$

Where L_{ROH_j} is the length of ROH j, and L_{total} is the total size of the genome covered by markers, calculated from the sum of inter-marker distances in the UMD v3.1 assembly. In order to facilitate comparisons with other studies, F_{ROH} was calculated using both the genome size based on autosomal and autosomal + X chromosomes. For each animal, F_{ROH} was calculated based on ROH of different minimum lengths: 0.5, 1, 2, 4, 8 or 16 Mb, representing autozygosity events that occurred approximately 100, 50, 25, 13, 6, and 3 generations in the past, respectively (Howrigan et al., 2011; Ferenčaković et al., 2013b). Additionally, chromosome-wise F_{ROH} was also computed.

An alternative measure of autozygosity was obtained by computing the diagonal elements of a modified realized genomic relationship matrix (VanRaden, 2008; VanRaden et al., 2011), calculated as:

$$G = \frac{ZZ'}{2\sum_{l=1}^{n} p_l(1-p_l)}$$

Where Z is a centered genotype matrix and p_l is the reference allele frequency at locus l. Matrix Z is obtained by subtracting from the genotype matrix M (with genotype scores coded as 0, 1 or 2 for alternative allele homozygote, heterozygote, and reference allele homozygote, respectively) the matrix P, whose elements of column l are equal to $2p_l$. The diagonal elements of G ($G_{i,i}$) represent the relationship of an animal with itself, and thus encapsulate autozygosity information. Following VanRaden et al. (2011), $G_{i,i}$ can provide a more suitable proxy for the pedigree-based inbreeding coefficient when assuming $p_l = 0.5$, rather than using base population allele frequencies estimates (which could be difficult to estimate especially in absence of complete pedigree data). Thus, matrix G was computed using allele frequencies fixed at 0.5.

DETECTION OF COMMON RUNS OF HOMOZYGOSITY

Chromosomal segments presenting ROH hotspots were defined as ROH islands or common ROH. In order to identify such genomic regions, we used two different strategies. First, we used the clustering algorithm implemented in *SNP & Variation Suite v7.6.8*, which identifies clusters of contiguous set of SNPs with size $> s_{min}$, where every SNP has at least n_{min} samples presenting a run. Clusters were identified based on a fixed minimum cluster size of $s_{min} = 0.5$ Mb for varying minimum number of samples: 127 (10%), 255 (20%), 319 (25%), and 639 (50%). In order to assess the sensitivity of the algorithm to parameter settings in

ROH detection, we repeated the analysis using minimum numbers of 30 or 150 SNPs in a run, maximum gap sizes of 100 kb or 500 kb, and 0 or 2 heterozygous genotypes as variable parameters.

Alternatively, we calculated locus autozygosity (F_L) following Kim et al. (2013). Briefly, for each SNP, animals were scored as autozygous (1) or non-autozygous (0) based on the presence of a ROH encompassing the SNP. Then, the locus autozygosity was simply computed as:

$$F_L = \frac{\sum_{i=1}^{n} S_i}{n}$$

where S_i is the autozygosity score of individual i, and n is the number of individuals. In essence, F_L represents the proportion of animals with scores equal to 1 (i.e., that present a ROH enclosing the marker), thus it summarizes the level of local autozygosity in the sample.

RESULTS AND DISCUSSION

DISTRIBUTION OF ROH LENGTH

After filtering, 668,589 SNP marker genotypes across 1,278 animals were retained for analyses. The average, median, minimum and maximum ROH length detected across all chromosomes were 1.26, 0.70, 0.50, and 70.91 Mb, respectively, suggesting this specific Nellore cattle population experienced both recent and remote autozygosity events. Segments as large as 10 Mb are traceable to inbreeding that occurred within the last five generations (Howrigan et al., 2011), and a total of 942 samples (73.7%) presented at least one homozygous fragment larger than 10 Mb. Therefore, it is likely that these long ROH are signatures of the extended use of recent popular sires.

DISTRIBUTION OF GENOME-WIDE AUTOZYGOSITY

The distributions of $G_{i,i}$ and F_{ROH} based on autosomal ROH of different minimum lengths (>0.5, >1, >2, >4, >8 or >16 Mb) are shown in **Figure 1**. Although the inclusion of the X chromosome did not cause substantial differences in the calculation of genome-wide F_{ROH} (Supplementary Figure S1), we focused on the estimates using only autosomes for the ease of comparison with other studies. The skewness of the autosomal F_{ROH} distribution increased as the minimum fragment length increased, ranging from 1.56 for $F_{ROH>0.5Mb}$ to 3.98 for $F_{ROH>16Mb}$. The number of animals with $F_{ROH} = 0$ also increased as the minimum ROH length increased, starting at 12 (0.94%) for $F_{ROH>2Mb}$ and increasing to 827 (64.71%) for $F_{ROH>16Mb}$. Under the assumption of the relationship between ROH length and age of autozygosity, these findings show that varying the minimum ROH length in the calculation of F_{ROH} can be useful to discriminate animals with recent and remote autozygosity.

As shown in **Figure 2**, the correlation between autosomal $F_{ROH>1Mb}$ and $G_{i,i}$ ($r = 0.69$) was close to the ones reported by Ferenčaković et al. (2013b) for the comparison between $F_{ROH>1Mb}$ derived from the HD panel and pedigree estimates in Brown Swiss ($r = 0.61$), Pinzgauer ($r = 0.62$), and Tyrol Gray ($r = 0.75$). Similar correlations were observed when the X chromosome was included in the analysis (Supplementary Figure S2). McQuillan et al. (2008) also reported correlations between F_{ROH} and pedigree estimates in human European populations ranging from 0.74 to 0.82. Considering that VanRaden (2008) proposed G as a proxy for a numerator relationship matrix obtained from highly reliable and recursive pedigree data, we expect that the correlations found for $G_{i,i}$ are fair approximations to the ones we would have found if complete pedigree data was available.

In the present study, correlations between F_{ROH} and $G_{i,i}$ decreased as a function of different ROH length (**Figure 2**). This may be due to the properties of the G matrix, which is based on individual loci, whereas F_{ROH} is based on chromosomal segments. Ferenčaković et al. (2013b) showed that medium density SNP panels, such as the Illumina® BovineSNP50, systematically overestimate F_{ROH} when segments shorter than 4 Mb are included in the calculations, while the Illumina® BovineHD panel is robust for the detection of shorter segments. Hence, although the inclusion of short length ROH in the calculation of F_{ROH} may be desirable for autozygosity estimates accounting for remote inbreeding, there is a compromise between SNP density, minimum ROH length and false discovery of ROH. Since the HD panel allows for the detection of short ROH, in this section we focused on the results obtained with $F_{ROH>1Mb}$ as it presented the second highest correlation with $G_{i,i}$ and is comparable with previous studies.

The minimum, average, median, and maximum autosomal $F_{ROH>1Mb}$ across all animals were 0.43, 4.79, 4.58, and 18.55%, respectively. The animal presenting the highest autozygosity value (18.55%) exhibited 69 ROH > 1 Mb encompassing 465.66 Mb of the total autosomal genome extension covered by markers (2.51 Gb), with a mean ROH length of 6.75 ± 9.20 Mb, and a maximum segment length of 43.79 Mb. The least inbred animal presented 8 ROH > 1 Mb, summing up only 10.72 Mb, with an average length of 1.34 ± 0.46 Mb and a maximum of 2.43 Mb.

The coefficient of variation (here denoted as the ratio of the standard deviation to the mean) of the $F_{ROH>1Mb}$ distribution was 37.5%, indicating moderate variability in autozygosity levels in this sample. In spite of the average genome coverage by ROH of 4.58% may seem to indicate moderate inbreeding levels for classical standards, it has to be considered that incomplete pedigree data usually fails to capture remote inbreeding, so that traditional inbreeding estimates based on pedigree are only comparable with F_{ROH} calculated over large ROH lengths, which in the present study were close to 0%.

Compared to other cattle populations, this sample of Nellore cows presented a lower average autozygosity. For instance, Ferenčaković et al. (2013b) reported average autosomal $F_{ROH>1Mb}$ of 15.1%, 6.2%, and 6.6% for samples of the *Bos taurus* breeds Brown Swiss, Pinzgauer, and Tyrol Gray, respectively. Also, the effective population size estimated for this Nellore sample was approximately 362 animals (Supplementary Material), which is consistent with the low genome average LD reported by other studies (McKay et al., 2007; Espigolan et al., 2013; Pérez O'Brien et al., 2014) and indicative of a non-inbred population.

DISTRIBUTION OF CHROMOSOME-WISE AUTOZYGOSITY

The averages of the chromosome-wise $F_{ROH>0.5Mb}$ across samples are shown in **Figure 3**. Chromosome X exhibited a substantially higher average autozygosity when compared to the autosomes. Importantly, we found no evidence for a smaller effective population size for the X chromosome in comparison

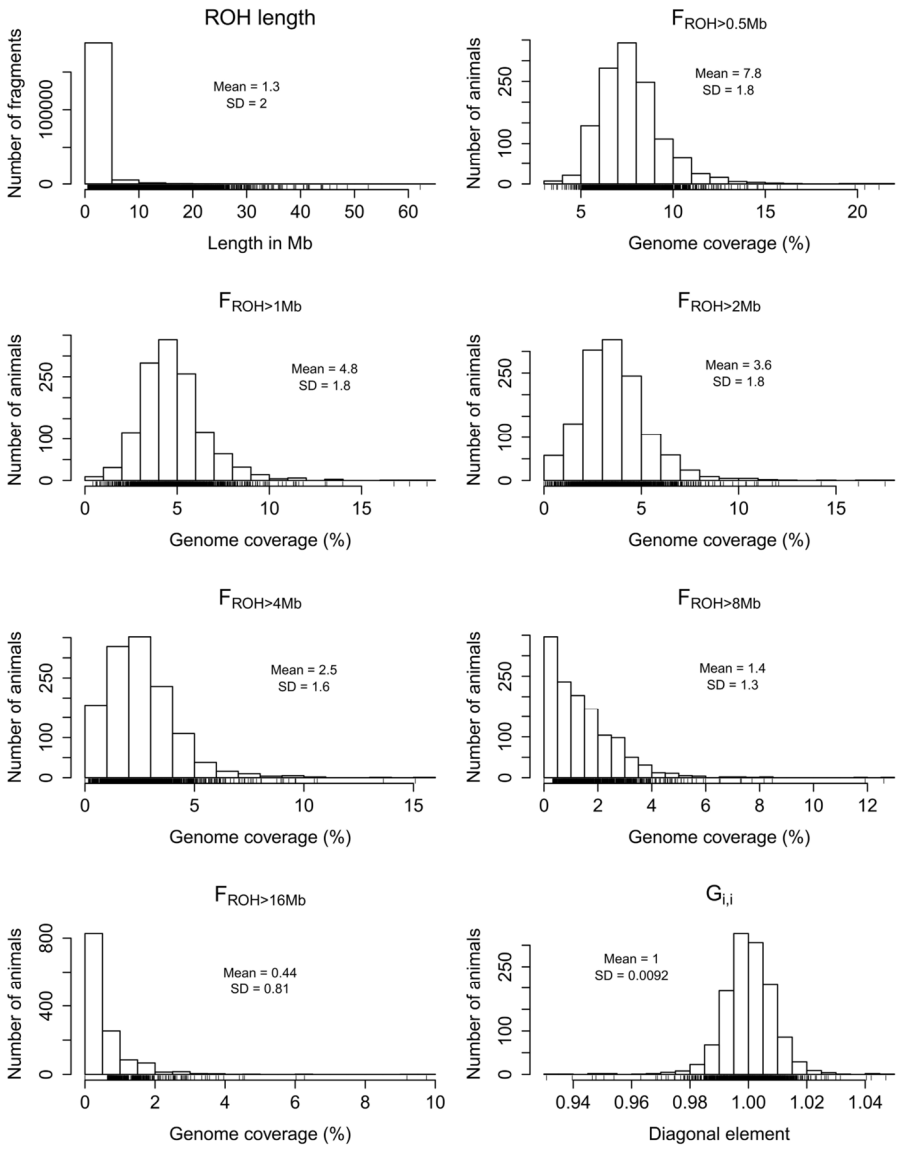


FIGURE 1 | Frequency distributions of all detected runs of homozygosity (ROH) across samples, percentage of autosomal genome coverage by ROH (F_{ROH}) of different minimum lengths (>0.5, >1, >2, >4, >8, and >16 Mb), and diagonal elements of the realized genomic relationship matrix ($G_{i,i}$).

to the autosomal genome (Supplementary Material). This may be due to the mode of inheritance of the X chromosome, which is hemizygous in the male lineage and therefore more susceptible to bottlenecks and drift even under assumptions of balanced numbers of males and females (Gottipati et al., 2011).

An alternative explanation is that the gene content and the sex-specific copy number of the X chromosome is under stronger selective pressure in comparison to autosomal DNA (Hammer et al., 2010; Deng et al., 2014). In both hypotheses, this higher autozygosity may reflect historical and demographical events. In the early 20th century, when more frequent importation of Nellore cattle to Brazil was initiated, the indigenous herds mainly consisted of descendants from taurine (*Bos taurus*) cattle imported since the late 15th century after the discovery of America (Ajmone-Marsan et al., 2010). In spite of the use of taurine dams for breeding during the early establishment of Nellore cattle in Brazil, the decades that followed were marked by intense backcrossing to Nellore bulls, causing most of the taurine contribution to be swept out from the Nellore autosomal genome (Utsunomiya et al., 2014). However, it is well-established that taurine mitochondrial DNA is prevalent in Nellore cattle, as it is a strict maternal contribution (Meirelles et al., 1999). Therefore, the X chromosome may have experienced a greater drift than the autosomal genome due to limited number of founders. The levels

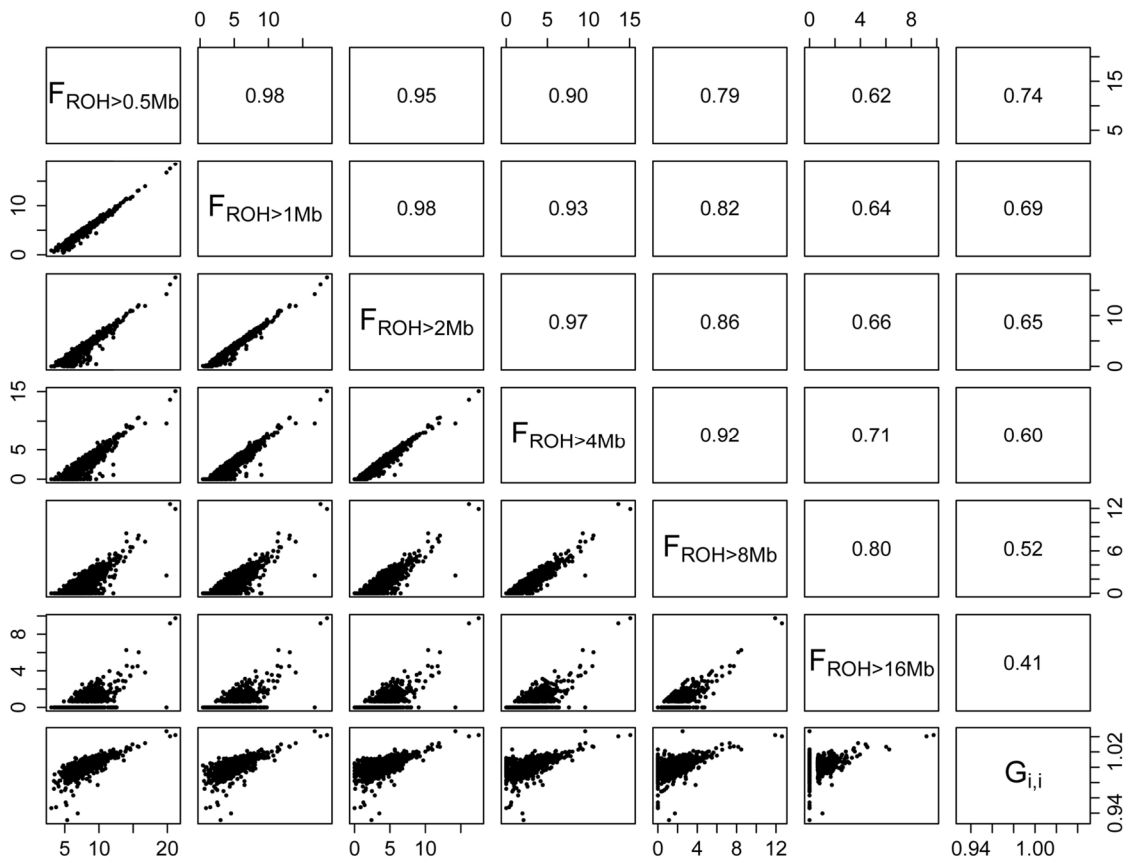


FIGURE 2 | Scatterplots (lower panel) and correlations (upper panel) of percentage of autosomal genome coverage by runs of homozygosity (F_{ROH}) of different minimum lengths (>0.5, >1, >2, >4, >8, and >16 Mb) and diagonal elements of the realized genomic relationship matrix ($G_{i,i}$). The last column of panels on the right indicates that the correlation between F_{ROH} and $G_{i,i}$ decreases as a function of minimum fragment size.

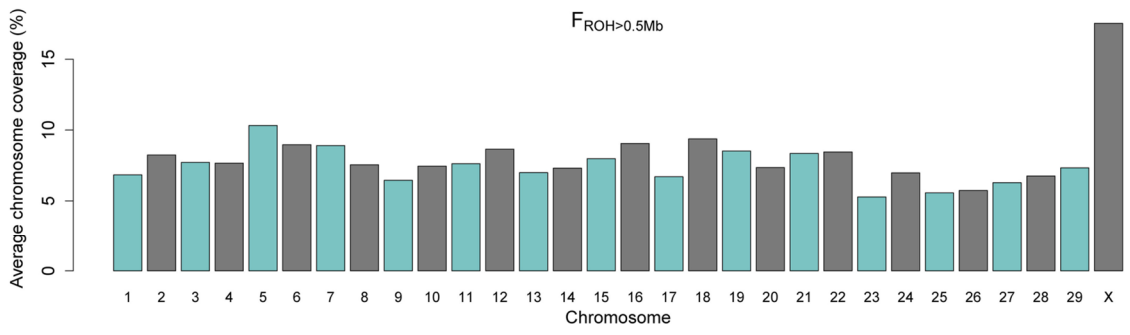


FIGURE 3 | Barplot of average percentage of chromosome coverage by runs of homozygosity (F_{ROH}) of minimum length of 0.5 Mb.

of taurine introgression still segregating in the X chromosome in this herd remain unclear.

IDENTIFICATION OF COMMON ROH

Table 1 presents the results obtained from the ROH clustering analysis. The algorithm was robust in respect to gap size between SNPs, but substantial differences were observed when the number of consecutive SNPs and the number of heterozygous genotypes were modified. Few common ROH were identified even when the minimum number of samples in the cluster was 10%, indicating that ROH distribution is not uniform across the genome. In fact, despite of the occurrence of 99.98% of the SNPs within a ROH of at least one individual, only 19.37% markers were encompassed by ROH observed in 10% or more of the samples. This finding is similar to that reported by Ferenčaković et al. (2013b), and is consistent with the stochastic nature of meiotic recombination. This

suggests that the ongoing selection for weight, carcass and reproductive traits in this population has not yet created detectable ROH-based selection signatures related to production.

The calculations of locus autozygosity were consistent with the cluster analysis using 150 SNPs and 2 heterozygous genotypes, regardless of permitted gap size (**Figure 4**). Seven distinct genomic regions, four of them on chromosome X, presented strong hotspots of autozygosity, where over half of the samples ($n = 639$) contained a ROH. The common ROH on the X chromosome are difficult to be discussed as they span several millions of bases, encompassing hundreds of genes and making functional explorations unfeasible. Besides, the assembly status of X chromosome is poorer than the autosomal ones. Hence, we focused on the three autosomal regions on chromosomes 4, 7, and 12. The three regions were relatively short, ranging from 0.73 to 1.43 Mb. For this range of ROH length, the expected number of generations since the common ancestor is estimated between 35 and 69 (Howrigan et al., 2011). Assuming a cattle generation interval of 5 years, these inbreeding events may have occurred between 175 and 345 years ago. Although this estimate does not account for birth date and overlapping generations, these remote autozygosity events are likely to predate the foundation of the Nellore breeding programs, and therefore expected to be related to natural selection, random drift or population bottlenecks.

The most autozygous locus was found at chromosome 7:51605639-53035752. This region was previously reported in genome-wide scans for signatures of selection in cattle through the comparison of *Bos taurus* and *indicus* breeds via F_{ST} analysis (Bovine HapMap Consortium, 2009; Porto-Neto et al., 2013) and was detected as a ROH hotspot in an analysis of three taurine and indicine breeds each (Sölkner et al., 2014). This region has been implicated in the control of parasitemia in cattle infected by *Trypanosoma congolense* (Hanotte et al., 2003), and is orthologous to the human chromosome segment 5q31-q33, known as the Th2 cytokine gene cluster, which has been shown to be implicated in the control of allergy and resilience against infectious diseases such as malaria (Garcia et al., 1998; Rihet et al., 1998; Flori et al., 2003; Hernandez-Valladares et al., 2004) and leishmaniasis (Jeronimo et al., 2007). The region also flanks *SPOCK1*, a candidate gene for puberty both in humans (Liu et al., 2009) and cattle (Fortes et al., 2010). Although fertility and resistance to infectious diseases are candidate biological drivers of this ROH hotspot, the gene and the phenotype underlying this putative selection signature are unknown.

The common ROH at 12:28433881-29743057 identified in the present study also overlaps a common ROH hotspot (Sölkner et al., 2014) and a region of divergent selection between *Bos taurus* and *Bos indicus* cattle (Gautier et al., 2009; Porto-Neto et al., 2013), and the segment encompasses the human ortholog *BRCA2*, involved in Fanconi anemia in humans (Howlett et al., 2002). A signature of selection nearby the 4:46384250-47113352 region detected here has also been reported by Gautier and Naves (2011), but the genes involved and the selective pressure remain uncharacterized.

Table 1 | Detection of common runs of homozygosity according to different number of consecutive SNPs, percentage of animals, gap size, and number of heterozygous genotypes.

Gap size	30 SNPs				150 SNPs				Heterozygotes
	10%	20%	25%	50%	10%	20%	25%	50%	
100 Kb	437	106	57	9	186	29	12	1	0
	471	288	183	29	365	91	47	7	2
500 Kb	479	126	76	13	193	32	14	1	0
	768	334	214	45	375	96	50	7	2

CONCLUSIONS

We used high-density SNP genotypes to successfully characterize autozygosity in Nellore cows under artificial selection for reproductive, carcass and weight traits. We have shown that, although the massive use of relatively few sires and artificial insemination has generated long stretches of homozygous haplotypes in the genomes of over 70% of these animals, inbreeding levels were considerably low in this population. We also found few genomic regions with high homozygosity across individuals, suggesting that the ongoing selection for reproductive, weight and carcass traits in this population is not very intensive or too recent to have left selection signatures in the form of ROH. Furthermore, the current common breeding practices of avoiding inbreeding in the mating schemes are antagonistic to additive trait selection,

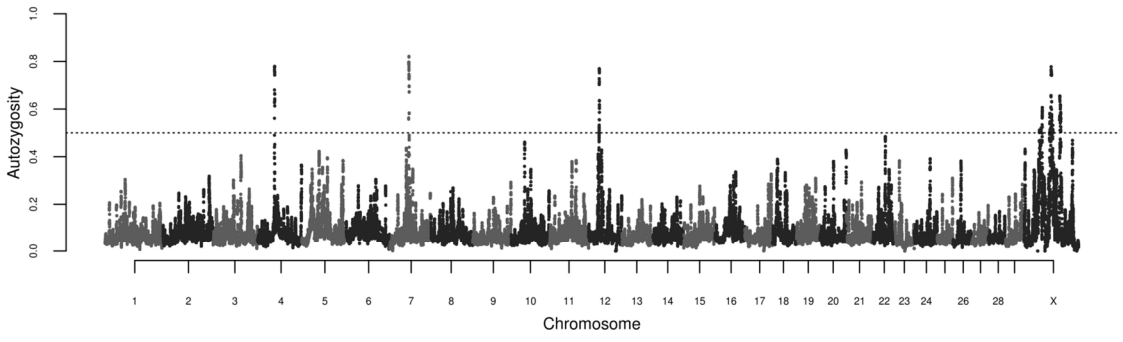


FIGURE 4 | Manhattan plot of genome-wide locus autozygosity in Nellore cows. The dashed line represents the 50% threshold.

making it hard to maintain ROH signatures in the herds. The three candidate regions under selection identified herein were likely to be contributions from remote ancestors, predating the foundation of the Nellore breeding programs. The selective pressure effects and the genes involved in these regions should be subject of future investigation.

ACKNOWLEDGMENTS

We thank Guilherme Penteado Coelho Filho and Daniel Biluca for technical assistance in sample acquisition. We also thank to Fernando Sebastian Baldi Rey for the manuscript revision and pertinent suggestions. This research was supported by: National Counsel of Technological and Scientific Development (CNPq - http://www.cnpq.br/) (process 560922/2010-8 and 483590/2010-0); and São Paulo Research Foundation (FAPESP - http://www.fapesp.br/) (process 2013/15869-2 and 2014/01095-8). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Mention of trade name proprietary product or specified equipment in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the authors or their respective institutions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fgene.2015.00005/abstract

REFERENCES

Ajmone-Marsan, P., Garcia, J. F., Lenstra, J. A., and The Globaldiv Consortium (2010). On the origin of cattle: how aurochs became cattle and colonized the world. *Evol. Anthropol.* 19, 148–157. doi: 10.1002/evan.20267

Bjelland, D. W., Weigel, K. A., Vukasinovic, N., and Nkrumah, J. D. (2013). Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* 96, 4697–4706. doi: 10.3168/jds.2012-6435

Bovine HapMap Consortium (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 324, 528–532. doi: 10.1126/science.1167936

Curik, I., Ferenčaković, M., and Sölkner, J. (2014). Inbreeding and runs of homozygosity: a possible solution to an old problem. *Livest. Sci.* 166, 26–34. doi: 10.1016/j.livsci.2014.05.034

Deng, X., Berletch, J. B., Nguyen, D. K., and Disteche, C. M. (2014). X chromosome regulation: diverse patterns in development, tissues and disease. *Nat. Rev. Genet.* 15, 367–378. doi: 10.1038/nrg3687

Espigolan, R., Baldi, F., Boligon, A. A., Souza, F. R., Gordo, D. G., Tonussi, R. L., et al. (2013). Study of whole genome linkage disequilibrium in Nellore cattle. *BMC Genomics.* 14:305. doi: 10.1186/1471-2164-14-305

Falconer, D. S., and Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*, 4 Edn. (Harlow, UK: Addison-Wesley Longman), 464.

Ferenčaković, M., Hamziæ, E., Gredler, B., Solberg, T. R., Klemetsdal, G., Curik, I., et al. (2013a). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *J. Anim. Breed. Genet.* 130, 286–293. doi: 10.1111/jbg.12012

Ferenčaković, M., Sölkner, J., and Curik, I. (2013b). Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genet. Sel. Evol.* 45:42. doi: 10.1186/1297-9686-45-42

Flori, L., Kumulungui, B., Aucan, C., Esnault, C., Traore, A. S., Fumoux, F., et al. (2003). Linkage and association between *Plasmodium falciparum* blood infection levels and chromosome 5q31-q33. *Genes Immun.* 4, 265–268. doi: 10.1038/sj.gene.6363960

Fortes, M. R. S., Reverter, A., Zhang, Y., Collis, E., Nagaraj, S. H., Jonsson, N. N., et al. (2010). Association weight matrix for the genetic dissection of puberty in beef cattle. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13642–136427. doi: 10.1073/pnas.1002044107

Garcia, A., Marquet, S., Bucheton, B., Hillaire, D., Cot, M., Fievet, N., et al. (1998). Linkage analysis of blood *Plasmodium falciparum* levels: interest of the 5q31-q33 chromosome region. *Am. J. Trop. Med. Hyg.* 58, 705–709.

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., et al. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics.* 10:550. doi: 10.1186/1471-2164-10-550

Gautier, M., and Naves, M. (2011). Footprints of selection in the ancestral admixture of a New World Creole cattle breed. *Mol. Ecol.* 20, 3128–3143. doi: 10.1111/j.1365-294X.2011.05163.x

Gibson, J., Morton, N. E., and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Hum. Mol. Genet.* 15, 789–795. doi: 10.1093/hmg/ddi493

Gottipati, S., Arbiza, L., Siepel, A., Clark, A. G., and Keinan, A. (2011). Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. *Nat. Genet.* 43, 741–743. doi: 10.1038/ng.877

Hammer, M. F., Woerner, A. E., Mendez, F. L., Watkins, J. C., Cox, M. P., and Wall, J. D. (2010). The ratio of human X chromosome to autosome diversity is positively correlated with genetic distance from genes. *Nat. Genet.* 42, 830–831. doi: 10.1038/ng.651

Hanotte, O., Ronin, Y., Agaba, M., Nilsson, P., Gelhaus, A., Horstmann, R., et al. (2003). Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7443–7448. doi: 10.1073/pnas.1232392100

Hernandez-Valladares, M., Rihet, P., ole-MoiYoi, O. K., and Iraqi, F. A. (2004). Mapping of a new quantitative trait locus for resistance to malaria in mice by a comparative mapping approach with human chromosome 5q31-q33. *Immunogenetics* 56, 115–117. doi: 10.1007/s00251-004-0667-0

Howlett, N. G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., De Die-Smulders, C., et al. (2002). Biallelic inactivation of *BRCA2* in Fanconi anemia. *Science* 297, 606–609. doi: 10.1126/science.1073834

Howrigan, D. P., Simonson, M. A., and Keller, M. C. (2011). Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 12:460. doi: 10.1186/1471-2164-12-460

Jeronimo, S. M., Holst, A. K., Jamieson, S. E., Francis, R., Martins, D. R., Bezerra, F. L., et al. (2007). Genes at human chromosome 5q31.1 regulate delayed-type hypersensitivity responses associated with *Leishmania chagasi* infection. *Genes Immun.* 8, 539–551. doi: 10.1038/sj.gene.6364422

Keller, M. C., Visscher, P. M., and Goddard, M. E. (2011). Quantification of inbreeding due to distant ancestors and its detection using dense single nucleotide polymorphism data. *Genetics* 189, 237–249. doi: 10.1534/genetics.111.130922

Kim, E. S., Cole, J. B., Huson, H., Wiggans, G. R., Van Tassell, C. P., Crooker, B. A., et al. (2013). Effect of artificial selection on runs of homozygosity in u.s. Holstein cattle. *PLoS ONE* 8:e80813. doi: 10.1371/journal.pone.0080813

Lencz, T., Lambert, C., DeRosse, P., Burdick, K. E., Morgan, V., Kane, J. M., et al. (2007). Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19942–19947. doi: 10.1073/pnas.0710021104

Leroy, G. (2014). Inbreeding depression in livestock species: review and meta-analysis. *Anim. Genet.* 45, 618–628. doi: 10.1111/age.12178

Liu, Y. Z., Guo, Y. F., Wang, L., Tan, L. J., Liu, X. G., Pei, Y. F., et al. (2009). Genome-wide association analyses identify SPOCK as a key novel gene underlying age at menarche. *PLoS Genet.* 5:e1000420. doi: 10.1371/journal.pgen.1000420

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823. doi: 10.1126/science.1215040

McKay, S. D., Schnabel, R. D., Murdoch, B. M., Matukumalli, L. K., Aerts, J., Coppieters, W., et al. (2007). Whole genome linkage disequilibrium maps in cattle. *BMC Genet.* 8:74. doi: 10.1186/1471-2156-8-74

McQuillan, R., Leutenegger, A. L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., et al. (2008). Runs of homozygosity in European populations. *Am. J. Hum. Genet.* 83, 359–372. doi: 10.1016/j.ajhg.2008.08.007

Meirelles, F. V., Rosa, A. J. M., Lobo, R. B., Garcia, J. M., Smith, L. C., and Duarte, F. A. M. (1999). Is the American zebu really *Bos indicus*? *Genet. Mol. Biol.* 22, 543–546. doi: 10.1590/S1415-47571999000400013

Pérez O'Brien, A. M., Mészáros, G., Utsunomiya, Y. T., Sonstegard, T. S., Garcia, F. J., Van Tassell, C. P., et al. (2014). Linkage disequilibrium levels in *Bos indicus* and *Bos taurus* cattle using medium and high density SNP chip data and different minor allele frequency distributions. *Livest. Sci.* 166, 121–132. doi: 10.1016/j.livsci.2014.05.007

Porto-Neto, L. R., Sonstegard, T. S., Liu, G. E., Bickhart, D. M., Da Silva, M. V., Machado, M. A., et al. (2013). Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* 14:876. doi: 10.1186/1471-2164-14-876

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Purfield, D. C., Berry, D. P., McParland, S., and Bradley, D. G. (2012). Runs of homozygosity and population history in cattle. *BMC Genet.* 13:70. doi: 10.1186/1471-2156-13-70

Rihet, P., Traor,eì, Y., Abel, L., Aucan, C., Traoreì-Leroux, T., and Fumoux, F. (1998). Malaria in humans: *Plasmodium falciparum* blood infection levels are linked to chromosome 5q31-q33. *Am. J. Hum. Genet.* 63, 498–505. doi: 10.1086/301967

Sölkner, J., Ferenčaković, M., Karimi, Z., Perez O'Brien, A. M., Mészáros, G., Eaglen, S., et al. (2014). "Extremely non-uniform: patterns of runs of homozygosity in bovine populations," in *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production* (Vancouver, BC).

Szpiech, Z. A., Xu, J., Pemberton, T. J., Peng, W., Zollner, S., Rosenberg, N. A., et al. (2013). Long runs of homozygosity are enriched for deleterious variation. *Am. J. Hum. Genet.* 93, 90–102. doi: 10.1016/j.ajhg.2013.05.003

Utsunomiya, Y. T., Bomba, L., Lucente, G., Colli, L., Negrini, R., Lenstra, J. A., et al. (2014). Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle. *BMC Genet.* 15:47. doi: 10.1186/1471-2156-15-47

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423. doi: 10.3168/jds.2007-0980

VanRaden, P. M., Olson, K. M., Wiggans, G. R., Cole, J. B., and Tooker, M. E. (2011). Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94, 5673–5682. doi: 10.3168/jds.2011-4500

Xue, Y., Chen, Y., Ayub, Q., Huang, N., Ball, E. V., Mort, M., et al. (2012). Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing. *Am. J. Hum. Genet.* 91, 1022–1032. doi: 10.1016/j.ajhg.2012.10.015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 October 2014; paper pending published: 03 December 2014; accepted: 07 January 2015; published online: 29 January 2015.

Citation: Zavarez LB, Utsunomiya YT, Carmo AS, Neves HHR, Carvalheiro R, Ferenčaković M, Pérez O'Brien AM, Curik I, Cole JB, Van Tassell CP, da Silva MVGB, Sonstegard TS, Sölkner J and Garcia JF (2015) Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. Front. Genet. ***6****:5. doi: 10.3389/fgene.2015.00005*

This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics.

Curik I., Ferenčaković M., Karapandza N., Cubric Curik V. & Sölkner J. (2014) Estimation of inbreeding and effective population size in Istrian cattle using molecular information. *Acta Agraria Kaposváriensis* **18**, 30-4.

Acta Agraria Kaposváriensis (2014) Vol 18 Supplement 1, 30-34.
Kaposvári Egyetem, Agrár- és Környezettudományi Kar, Kaposvár
Kaposvár University, Faculty of Agricultural and Environmental Sciences, Kaposvár

Estimation of inbreeding and effective population size in Istrian cattle using molecular information

I. Curik[1], M. Ferenčaković[1], N. Karapandza[1], V. Cubric Curik[1], J. Sölkner[2]

[1]University of Zagreb, Faculty of Agriculture, Svetošimunska 25, 10000 Zagreb, Croatia
[2]University of Natural Resources and Life Sciences Vienna, Department of Sustainable Agricultural Systems Division of Livestock Sciences, Gregor Mendel Str. 33, A-1180 Vienna, Austria

ABSTRACT

To provide preliminary insight in the conservation risk status in Istrian cattle we analysed ROH inbreeding and effective population size in 15 individuals, mostly bulls, using BovineSNP50K BeadChip. We obtained very high inbreeding level, although with broad confidence interval, and very low effective population size. While the results obtained are preliminary (small sample size) and should be treated with caution, the high recent inbreeding and small effective population size suggest additional monitoring of the conservation risk status of the Istrian cattle.
(Keywords: Istrian cattle, Inbreeding, Runs of homozygosity, effective population size, single nucleotide polymorphism)

INTRODUCTION

Istrian cattle, colloquially called Boškarin, is the autochthonous breed spread mainly over the Istrian peninsula. The breed belongs to the group of grey cattle breeds that are scattered over the Balkan and neighboring countries (Croatia, Bulgaria, Greece, Hungary, Italy, Romania, Serbia, Turkey and Ukraine) and that are considered as direct descendants from the Auroch (Bos primigenius). In the last 50 years the number of Istrian cattle individuals has been reduced dramatically. Inbreeding level and effective population size (*Ne*) are among the most important conservation genetic parameters. Classical inbreeding and *Ne* estimates rarely work well in real populations as they are mostly based on inaccurate pedigree records or, in case of *Ne* estimation, on robust demographic parameters that do not completely recognize the history of the population (bottlenecks, preferential mating or population subdivision). The rapid development of new molecular technologies enabled high-throughput genotyping of individual animals at available prices. Consequently, those technological achievements provide new views on old problems and reinforce estimation of inbreeding and *Ne* from molecular markers. Runs of homozygosity (ROH) were recently proposed as a useful concept in quantifying individual inbreeding in humans (*McQuillan et al.,* 2008), cattle (*Ferenčaković et al.,* 2011; *Purfield et al.,* 2012) and pigs (*Bosse et al.,* 2012), performing even better than traditional estimates calculated from the pedigree. *Sved* (1971) and *Hill* (1981) showed that linkage disequilibrium (*LD*) could be used to estimate *Ne*. While theoretical basis has been established before, the practical use of *LD* in estimating *Ne* started by *Hayes et al.*, (2003) and, further, continued by *Tenesa et al.,* (2007); *Qanbari et al.,* (2009).

The aim of this study was, based on high-throughput genotypes (BovineSNP50K BeadChip), to estimate inbreeding level and effective population size in Istrian cattle. The results obtained will contribute to the conservation management strategy of the Istrian cattle.

MATERIAL AND METHODS

Samples (15) representing Istrian cattle population, mostly bulls, were either taken from the blood (randomly chosen from several private farms in Istria, or were obtained as semen straws (three bulls) from CRSH d.o.o. in Krizevci (www.crsh.hr). As the number of Istran bulls is extremely small we have considered our sample as representative, although, we are aware that larger sample would be more adequate.

After ROH calculation quality control that was performed according to *Ferenčaković et al.* (2013b) we proceed with analyses including information from 42265 SNPs (%), placed on 29 autosomes and with average distance of 59 kb between adjacent SNPs. ROH segments were identified as a part of the genome in which 15 or more consecutive homozygous SNPs at a density of one SNP on every 100 kb are not more than one Mb apart. ROH calculations were done by SNP & Variation Suite (v7.6.8 Win 64; Golden Helix, Bozeman, MT, USA www.goldenhelix.com). The general formula for calculating F_{ROH} from chip data is $F_{ROH}= L_{ROH}/L_{AUTOSOME}$, where L_{ROH} is the total length of all ROH in the genome of an individual while $L_{AUTOSOME}$ refers to the specified length of the autosomal genome covered by SNPs on the chip (here 2,543,177 kb). For each bull, we calculated three inbreeding coefficients ($F_{ROH>4Mb}$, $F_{ROH>8Mb}$ and $F_{ROH>16Mb}$) based on ROH of different minimum lengths (>4, >8 or >16). Different ROH inbreeding coefficients are expected to have differently remote common ancestors (for details see *Curik et al*., 2014).

Effective population size (Ne) was estimated following the approach described in *Flury et al.* (2010) respecting functional relationship of *Ne* with correlation r^2 and recombination rate (*c*), here inter-marker genetic distance between two considered loci with assumption that 1 Mb = 1 cM. Two slightly different formulas were used, one described in Sved (1971) where $r^2=1/(1+4\cdot c\cdot Ne_1)$ and the other described in Weir and Hill (1980) where $r^2=1/(1+4\cdot c\cdot Ne_2)+(1/n)$ with n=2·number of animals (bulls) used in the calculation as a correction factor for a sample size induced LD. Only SNPs with adjacent r^2 values from 0.01 to 0.99 were used in the calculation by *Uimari and Tapio* (2011). Finally, time defined effective population size Ne_T was derived from 40 marker distance derived categories as described in *Flury et al.* (2010). Current effective population size was predicted based on the regression analysis of estimated values in previous 150 generations. LD (r^2) was estimated using SNP & Variation Suite (v7.6.8 Win 64, Golden Helix, Bozeman, MT, USA www.goldenhelix.com). Data manipulations, numerical calculations and graphical visualisations were done by procedures included in SAS 9.3 (*SAS Institute*, 2011).

RESULTS AND DISCUSSION

Summary statistics of the ROH estimated inbreeding level ($F_{ROH>4Mb}$, $F_{ROH>8Mb}$ and $F_{ROH>16Mb}$) in 15 Istrian cattle bulls are presented in *Table 1*. The estimates obtained (mean and standard deviations) were much higher than those obtained in Brown Swiss, Fleckvieh, Norwegian Red and Tyrol Grey by *Ferenčaković et al.* (2013a) or in Pinzgauer by *Ferenčaković et al.* (2013b). However, one should be aware that the

confidence limits are very broad with values comparable to any population studied so far. One individual had extremely high close inbreeding ($F_{ROH>8Mb}$=0.351 and $F_{ROH>16Mb}$=0.287) indicating the absence of mating strategy respecting avoidance of close inbreeding.

Table 1

Summary statistics of inbreeding calculated from ROH with different lengths (>4 Mb, >8 MB and >16 MB) based on Illumina BovineSNP50K BeadChip in 15 Istrian cattle bulls

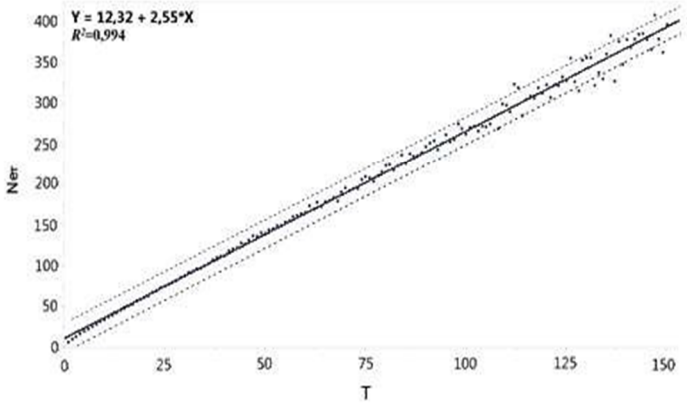
Inbreeding coefficient	Mean	Lower 95% CI	Upper 95% CI	Standard deviation	Range
$F_{ROH>4Mb}$	0.093	0.039	0.147	0.092	0.002-0.368
$F_{ROH>8Mb}$	0.081	0.029	0.133	0.091	0.000-0.351
$F_{ROH>16Mb}$	0.075	0.014	0.096	0.078	0.000-0.287

CI = Confidence interval

Historical estimates of the effective population size (Ne_T) during last 150 generations showed rather linear decrease of 2.55 individuals per generation while predicted current generation effective population size (Ne_0) was equal to 12.32 with 95% confidence interval ranging from 9.58 to 15.06 individuals (*Figure 1*).

Figure 1

Linear regression with 95% confidence interval presenting relationship between historical effective population size (Ne_T) and number of generations in the past (T) while Ne_T values were previously estimated from genomic data of 15 Istrian cattle individuals



Thus, the linear regression function was Ne_T=12.32+2.55·T with extremely high coefficient of determination (R^2=0.994). The obtained prediction for the current effective population size of Istrian cattle was surprisingly small. According to the Croatian Agricultural Agency report (2013) the breed status is highly endangered with Ne estimated to 151.59 (721 cows and 40 bulls) when calculated from the sex ratio

[Ne = (4·Nm·Nf) / (Nm+Nf), where Nm and Nf represent the number of breeding males and females, respectively].

Although, the sample size was very small, historical estimates of effective population size do represent large number of chromosomal segments originating from much larger number of individuals and, thus, should be less sensitive to the sample size. Still, the interpretation of the results should be considered with caution as we are not fully aware of the magnitude of potential bias resulting from one individual being highly inbred.

CONCLUSIONS

Although, the results obtained are preliminary (small sample size) and should be treated with caution, the appearance of high recent inbreeding in some individuals and small effective population size require additional monitoring of the conservation risk of Istrian cattle population.

ACKNOWLEDGEMENTS

We thank to Istrian cattle breeders, Agency for rural development of Istria Ltd., Pazin (AZZRI) and Centre for reproduction and livestock production in Croatia (CRSH d.o.o., Križevci) for their help in providing semen samples. We also thank to Dr. Silvio Vince for taking the blood samples. Parts of this research were funded by the German Federal Ministry of Education and Research (BMBF) within the project FUGATO-plus GenoTrack (FKZ0315134C).

REFERENCES

Bosse, M., Megens, H.J., Madsen, O., Paudel, Y., Frantz, L., Schook, L., Crooijmans, R., Groenen, M. (2012). Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. PLoS Genetics, 8. e1003100.

Croatian Agricultural Agency (2014). Cattle breeding, Annual report 2013, Barač Z. (Editor). Croatian Agricultural Agency, Križevci, 2014.

Curik, I., Ferenčaković, M., Sölkner, J. (2014). Inbreeding and Runs of homozygosity: a possible solution to an old problem. Livestock Science, Available online 9 June 2014, DOI: 10.1016/j.livsci.2014.05.034

Ferenčaković, M., Hamzic, E., Gredler, B., Curik, I., Solkner, J. (2011). Runs of homozygosity reveal genome-wideautozygosity in the Austrian Fleckvieh cattle. Agriculturae Conspectus Scientificus. 76. 325-328.

Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I., Sölkner, J. (2013a). Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. Journal of Animal Breeding and Genetics. 130. 286-293.

Ferenčaković, M., Solkner, J., Curik, I. (2013b). Estimating autozygosity from high- throughput information: effects of SNP density and genotyping errors. Genetics Selection Evolution. 45. 42.

Hayes, B.J., Visscher, P.M., McPartlan, H.C., Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Research. 13. 635-643.

Hill, W.G. (1981). Estimation of effective population size from data on linkage disequilibrium. Genetics Research. 38. 209-216.

McQuillan, R., Leutenegger, A., Abdel-Rahman, R., Franklin, C., Pericic, M., Barac- Lauc, L., Smolej-Narancic, N., Janicijevic, B., Polasek, O., Tenesa, A., Macleod, A., Farrington, S., Rudan, P., Hayward, C., Vitart, V., Rudan, I., Wild, S., Dunlop, M., Wright, A., Campbell, H., Wilson, J. (2008). Runs of homozygosity in European populations. American Journal of Human Genetics. 83. 359-372.

Purfield, D., Berry, D., McParland, S., Bradley, D. (2012). Runs of homozygosity and population history in cattle. BMC Genetics. 13. 70.

Qanbari, S., Pimentel, EC., Tetens, J., Thaller, G., Lichtner, P., Sharifi, AR., Simianer, H. (2010b). The pattern of linkage disequilibrium in German Holstein cattle. Animal Genetics. 41. 346-356.

Sved, J.A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations, Theoretical Population Biology. 2. 125-141.

Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E., Visscher, P.M. (2007). Recent human effective population size estimated from linkage disequilibrium. Genome Research. 17. 520-526.

Uimari, P., Tapio, M. (2011). Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds. Journal of Animal Science. 89. 609-614.

Weir, B.S., Hill, W.G. (1980). Effect of mating structure on variation in linkage disequilibrium. Genetics. 95. 477-488.

Corresponding author:

Ino Curik
University of Zagreb, Faculty of Agriculture Svetosimunska 25, 10000 Zagreb, Croatia
Phone: +386-239-4010
E-mail: icurik@agr.hr

Curik, I., Ferencakovic, M., Gredler, B. & Sölkner, J. (2010) Genome-wide heterozygosity and pedigree inbreeding coefficients in Simmental cattle population. In: *Proceedings - 9th World Congress on Genetics Applied to Livestock Production / German Society for Animal Science (ed.). Gießen: 9th World Congress on Genetics Applied to Livestock Production*. 1-4.

Genome-wide heterozygosity and pedigree inbreeding coefficients in Simmental cattle population

I. Curik[*], M. Ferencakovic[*], B. Gredler[†] and J. Sölkner[‡]

Introduction

The reduction of the population mean for a quantitative trait such as size, fertility, vigour, yield, and fitness is a negative consequence of inbreeding known as inbreeding depression. The phenomenon has been experimentally observed in numerous wild and domestic animal species (Falconer and Mackay 1996; Lynch and Walsh 1997). The inbreeding coefficient is a quantitative measure of inbreeding defined by Wright (1922) and by Malécot (1948), in terms of correlation and probability, respectively. Until recently inbreeding coefficients have been mostly estimated from pedigree information. As pedigree inbreeding coefficient (F_{PED}) refers to the expected value, there is no sampling variation in its value for individuals with the same pedigree. An additional assumption is that there are no systematic changes in allele frequencies due to selection (Wright 1951, 1965). Thus, consequently, it is appropriate for traits controlled by a number of loci that are close to infinity i.e. expected to be the same as at neutral loci. As the main effect of inbreeding is to render the population homozygous at the cost of decrease in heterozygosity, decline in heterozygosity is expected to be correlated with increase of inbreeding coefficient on, both, individual and population level. Berskin et al. (1970) and Groen et al. (1995) on population level and Curik et al. (2002) on individual level showed, by Monte Carlo simulations, that increase in F_{PED} do not correspond to the expected heterozygosity decline (inbreeding coefficient derived from the heterozygosity) when selection is affecting a trait controlled by a finite number of loci. Contradictory results were obtained in the first empirical analyses (based on less than 30 microsatellite loci) of the correlation between F_{PED} and individual heterozygosity by Ellegren (1999) and Curik et al. (2003) as correlation coefficients were -0.82 ($P<0.0001$) and -0.03 ($P=0.526$), respectively. Balloux et al. (2004) and Slate et al. (2004) have shown that the number of markers required to estimate genome-wide heterozygosity should be much higher than commonly applied (20 to 50 markers).

The main goal of this study was to analyze relationship between individual genome-wide heterozygosity and F_{PED} in artificially selected population, here Simmental (Flechvieh) bulls, as well as to analyze the trend in heterozygosity and inbreeding over a period of 30 years period.

[*] University of Zagreb, Faculty of Agriculture, Svetosimunska 25, 10000 Zagreb, Croatia;

[†] University of Natural Resources and Applied Life Sciences, Gregor-Mendel-Strasse 33, A-1180 Vienna, Austria.

[‡] University of Natural Resources and Applied Life Sciences, Gregor-Mendel-Strasse 33, A-1180 Vienna, Austria.

Material and methods

Data. The analyses performed relate to 1851 dual purpose Simmental (Fleckvieh) bulls born from 1975 to 2004.

Heterozygosity and inbreeding coefficients. Individual heterozygosity was calculated as the number of heterozygous loci divided by the total number of analyzed loci. Calculations were done for all loci ($pHt_{1\text{-}29}$) and for each chromosome separately (pHt_1, pHt_2, pHt_3, …, pHt_{29}). Inbreeding coefficients (F_{PED}) were calculated for each bull, from the pedigree data file consisting of 24071 animals, by the tabular method using the algorithm of Van Raden (1992). We also calculated inbreeding coefficients with restricted pedigree information in terms of maximum number of generations included (F_{PED3}, F_{PED4}, F_{PED5}, …, F_{PED20}). The
discrete generation equivalent (EqG) is computed for each individual as the sum of $(1/2)^n$, where n is the number of generations separating the individual from each known ancestor (Boichard et al., 1997).

Genotyping. Genotyping was performed for 54001 SNPs using the Illumina Bovine SNP50^{TM} Beadchip. After excluding SNPs with allele frequencies less than 1%, SNPs that strongly deviated from Hardy-Weinberg equilibrium and SNPs on Y chromosome, we considered 42198 SNPs for analysis.

Statistical analyses. Data manipulation, descriptive statistics, and simple statistics were obtained using various SAS procedures (SAS Institute, Cary, NC). Existence of population structure can influence estimation of correlations. To reduce potential bias we repeated all analyses over population structures with respect to birth year and graphical illustration obtained from PCA analysis using EIGENSTRAT software (Patterson et al. 2006).

Results and discussion

Descriptive statistics of F_{PED} and $pHt_{1\text{-}29}$, as well as their correlations, $r(F_{PED}, pHt_{1\text{-}29})$, with respect o birth year and position in the population are provided in Table 1. Calculated correlations were higher for subpopulation defined as Bulls born after 1998 versus Bulls born before 1999 and for subpopulation defined as Peripheral population versus Central population which is indication that population structure, here birth year and PCA based position, does influence correlations between F_{PED} and $pHt_{1\text{-}29}$. Although, EqG is expected to have influence on the estimation of correlations studied. Mean values of EqG for Bulls born after 1998 were 3.976±0.107, for Bulls born before 1999 were 3.925±0.206, for Central population were 4.000±0.000 and for Peripheral population were 3.749±0.301, and we think did not contribute strongly to the differences in observed correlations. Thus, we also observed somewhat lower values, -0.331 ($P<0.0001$) for $r(F_{PED3}, pHt_{1\text{-}29})$ and -0.491 ($P<0.0001$) for $r(F_{PED8}, pHt_{1\text{-}29})$.

Unfortunately, there are no similar analyses performed for comparison, so it is difficult to evaluated are correlations obtained within expectable range. Correlations between F_{PED} and individual chromosomal heterozygosities varied from -0.201 for $r(F_{PED}, pHt_8)$ up to -0.0789 for $r(F_{PED}, pHt_{12})$. The magnitude of correlations was strongly influenced by the number of SNPs genotyped per chromosome. Thus, among the first four highest correlations all chromosomes had more than 1800 SNPs genotyped. In contrast, among the first lowest correlations only one chromosome had 1356 SNPs genotypes while three other chromosomes had less than 870 SNPs genotyped.

Table 1: Descriptive statistics of pedigree inbreeding coefficient (F_{PED}), individual heterozygosity (pHt_{1-29}) and their Pearson correlation coefficients[$], $r(F_{PED}, pHt_{1-29})$, in Simmental cattle with respect to birth year and PCA defined population structure.

Population structure	Variable	N	Mean (Std) / Correlation[$]	Range
Whole population	F_{PED}	1851	0.011 (0.012)	0.000; 0.091
	pHt_{1-29}	1851	0.338 (0.009)	0.295; 0.382
	$r(F_{PED}, pHt_{1-29})$	1851	-0.491 (P<0.0001)	
Bulls born after 1998	F_{PED}	919	0.014 (0.001)	0.000; 0.091
	pHt_{1-29}	919	0.338 (0.009)	0.295; 0.382
	$r(F_{PED}- pHt_{1-29})$	919	-0.555 (P<0.0001)	
Bulls born before 1999	F_{PED}	932	0.008 (0.011)	0.000; 0.071
	pHt_{1-29}	932	0.338 (0.008)	0.299; 0.370
	$r(F_{PED}, pHt_{1-29})$	932	-0.451 (P<0.0001)	
Central population[#]	F_{PED}	1278	0.011 (0.012)	0.000; 0.076
	pHt_{1-29}	1278	0.338 (0.008)	0.295; 0.382
	$r(F_{PED}, pHt_{1-29})$	1278	-0.475 (P<0.0001)	
Peripheral population[#]	F_{PED}	574	0.006 (0.009)	0.000; 0.063
	pHt_{1-29}	574	0.340 (0.009)	0.317; 0.382
	$r(F_{PED}, pHt_{1-29})$	574	-0.530 (P<0.0001)	

[#] Central population was defined with respect to PCA1 [-0.04, 0.01] and PCA2 [-0.2, 0.2] intervals, while all other bulls were considered as members of Peripheral population.

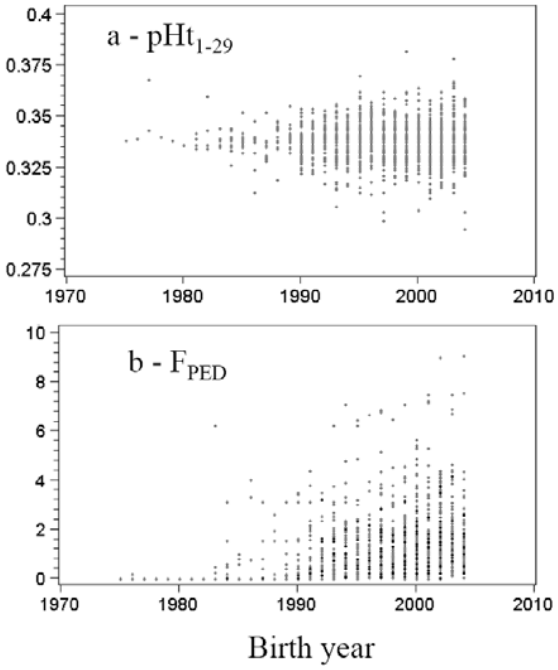


Figure 1: Trend for individual heterozygosity (pHt_{1-29}) and pedigree inbreeding coefficients (F_{PED}) over birth years of the bulls.

This suggests that for the population analyzed much large number of SNPs would be required to obtaine stable estimates of $r(F_{PED}, pHt_i)$ i.e. the estimates not infuluenced by the number of SNPs genotyped. While we expected that F_{PED} will increase during 30 year period, we were surprised by the stability of pHt_{1-29} value which remained constant around 0.338, (see Figure 1a. and 1b.). We are interested if the same pattern would be present in a population with higher inbreeding.

Conclusion

For the artificially selected population, here Simmental cattle with mean inbreeding of 1.1% and mean discrete generation equivalent (EqG) equal to 3.951, genotyped for 42198 SNPs correlation between F_{PED} and pHt_{1-29} was not high (-0.491; $P<0.0001$)) but varied across different population structures. When correlations were estimated chromosome-wise the number of SNPs (range from 804 to 2739) strongly affected estimates. While inbreeding coefficients increased over period of 30 years, the values obtained for individual heterozygosity were rather stable. Similar analyses on populations of various structures are required to for comparisons and better understanding of inbreeding depression.

References

Balloux, F., Amos, W., and Coulson, T. (2004). Does *Mol. Ecol.*, 13:3021–3031.

Berskin B, Shelby, C.,E., and Hazel, L.N. (1970). J. Anim. Sci. 30: 681–689.

Boichard, D., Maignel, L., and Verrier, E. (1997). *Genet. Sel. Evol.*, 29:5–23. Curik, I.,

Stipic, N., and Sölkner, J. (2002). *J. Anim. Breed. Genet.*, 119:101–115. Curik, I.,

Zechner, P., Solkner, J., and *et al.* (2003). *J. Hered.,* 94:125–132.

Ellegren, H. (1999). *Hereditas,* 130:239-244.

Falconer, D.S., and Mackay, T.F.C. (1996). *Introduction to Quantitative Genetics*. Harlow, UK: Longman.

Groen A.B., Kennedy, B.W. and Eisen, J.,. (1995). *Theor. Appl. Genet.,* 91:665–671. Lynch, M.,

and Walch, B. (1997). *Fundamentals of Quantitative Genetics*, vol. 1. Sunderland, Mas.: Sinauer Asociates.

Malécot, G. (1948). *Les Mathématiques de L'hérédité.* Masson et Cie, Paris, France. Van

Raden, P.M. (1992). *J. Dairy. Sci.*, 75:3136–3144.

Patterson, N., Price, A.L, and Reich, D. (2006). *PLOS Genet.* 2:e190. Slate, J,

David, P., Dodds, K.G., and *et al.* (2004). *Heredity*, 93:255–265. Wright, S.

(1922). *Am. Nat.,* 56:330–338.

Wright (1951). *Ann. Eugen.*, 15:323–354.

Wright (1965). *Evolution,* 19:395–420.

Sölkner, J., Ferenčaković, M., Karimi, Z., Perez O'Brien, A.M., Mészáros, G., Eaglen, S., Boison, S.A., & Curik, I. (2014). Extremely Non-uniform: Patterns of Runs of Homozygosity in Bovine Populations. In: *Proceedings - 10th World Congress on Genetics Applied to Livestock Production / American Society of Animal Science (ed) Champaign; 10th World Congress on Genetics Applied to Livestock Production.* 1-4.

Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Extremely Non-uniform: Patterns of Runs of Homozygosity in Bovine Populations

J. Sölkner[1], **M. Ferenčaković [2], Z. Karimi**[13], **A.M. Perez O'Brien**[1] **G. Mészáros**[14], **S. Eaglen**[14], **S.A. Boison**[1] **and I. Curik**[2]
[1]University of Natural Resources and Life Science, Vienna, Austria,
[2]University of Zagreb, Zagreb, Croatia, [3]University of Guelph, Guelph, Canada., [3]Roslin Institue, Edinburgh, U.K.

ABSTRACT: Runs of homozygosity (ROH) are regions of the genome that are contiguously homozygous because the parents of an individual have transmitted identical haplotypes. Long ROH indicate a recent common ancestor while very short ones are indicators of demographic history of a population. The distribution of ROH along the genome has been shown to be extremely non-uniform in human populations. We have analyzed patterns of runs of homozygosity from the Illumina BovineHD BeadChip, featuring more than 700,000 autosomal SNPs for the taurine cattle breeds Angus, Brown Swiss and Fleckvieh as well as for the indicine breeds Brahman, Gir und Nelore. ROH of lengths >1 Mb were accepted.
The average proportion of individuals for which any SNP was in a ROH across the genome was .081 (.069 for indicine breeds and .096 for taurine breeds), 99.0 and 99.9 percentiles were .183 (.165, .241) and .275 (.330, .406), respectively. ROH islands, i.e. regions with highest incidences (>.40) were partly overlapping and partly distinct between subspecies and breeds within subspecies. ROH islands were surprisingly gene rich compared to equally sized regions in distant regions of the same chromosomes. The genesis and function of ROH islands is still unclear and worth pursuing. Analysis of whole genome sequence data will add new challenges of analysis and prospects of better understanding the biology of ROH islands.

Keywords: inbreeding; cattle; SNP; runs of homozygosity; pattern

Introduction

High throughput genotyping allows a new and more accurate view on levels and effects of inbreeding in livestock (Bjelland et al., 2013, Ferenčaković et al., 2013a, Purfield et al., 2012). Runs of Homozygosity (termed by Lencz et al., 2007) are contiguous regions of the genome in homozygous state. ROH are due to both parents transmitting identical haplotypes from an ancestor to the emerging offspring, their lengths indicate how recent the common ancestor was, with longer ROH being derived from more recent ancestors. Length and configuration of haplotypes are determined by recombination events during meiosis. With high density SNP chip data, it is possible to accurately determine whether a particular SNP is part of a ROH or not. Howrigan et al. (2011) and Ferenčaković et al. (2013b) indicated ways of dealing with heterozygous calls due to genotyping errors for human and bovine populations, respectively. In this way, each SNP in the autosome of an individual can be marked as being part of a ROH or not. From there, proportions of individuals for which a SNP is in a ROH may be calculated, representing the level of local (SNP-wise) autozygosity of a population. Studies of human populations (McQuillan et al., 2008; Nothnagel et al., 2010; Pemberton et al., 2012) have indicated that the patterns of local levels of autozygosity is not uniform at all, calling regions of extreme high frequency of ROH as ROH islands. The causes for these differences are not well established yet, and while levels of linkage disequilibrium play a role, they explain only a relatively small part of ROH variation (Nothnagel et al., 2010).

In this study we explore heterogeneity of ROH levels in three taurine and three indicine cattle populations. We investigate the occurrence of genes within ROH islands and compare the numbers of genes found in these islands with those in regions of equal size 10 and 20 Mb downstream the chromosome.

Data and Methods

Breeds investigated. Illumina BovineHD BeadChip (777 K) data of the taurine breeds Angus (107 individuals), Brown Swiss (46), Fleckvieh (96), and the indicine breeds Brahman (100), Gir (100) and Nelore (133) were used. Genotypes were provided by Zebu Genomic Consortium - Brazil for Nelore, Embrapa - Brazil for Gir, Zuchtdata GmbH - Austria for Fleckvieh, by AGBU (Animal Genetics and Breeding Unit) University of New England - Australia for Angus and Brahman as well as BOKU University - Austria for Brown Swiss.

ROH detection. Data extraction and quality control were performed following Ferenčaković et al. (2013b). Data from the three taurine breeds were merged and only SNPs that were present in all three breeds were retained (555 609 autosomal SNPs). The same was done with the three indicine breeds, resulting in 649 218 autosomal SNPs. ROH were detected as in Ferenčaković et al. (2013b) using the SNP&Variation suite (SVS) from Golden Helix (www.goldenhelix.com), with exceptions for minimum number of SNP required to call a ROH that was here set to 30, maximum gap 250 kb and minimum density of 1 SNP/50 kb.

Comparative analysis. Proportions of SNPs being in a ROH were calculated per breed and per subspecies. Percentiles (50.0%, 99.0%, 99.9%) were calculated to detect ROH islands. Top regions with frequencies >40% within either subspecies were selected for further inspection. The width of an island region was determined by the positions of the leftmost and rightmost SNPs surpassing the 40% limit.

Genes located within the pattern boundaries and the orthologs in human were found using the Ensemble Genome Biomart tool (WTSI/EBI), while the gene functions were searched using the quickGO (EMBL-EBI), Ensembl release 72 - June 2013. http://www.ensembl.org.

Results and Discussion

The median proportion of individuals for which any SNP was in a ROH across the genome was .081 (.069 for indicine breeds and .096 for taurine breeds), 99.0 and 99.9 percentiles were .183 (.165, .241) and .275 (.330, .406), respectively. As visible from Figures 1 and 2, ROH frequencies vary vastly across positions in the genome.

Figure 1 provides a genome wide views of ROH islands for taurine and indicine cattle. Figure 2 gives the ROH patterns along BTA 12 and 21 in greater detail. ROH islands are partly coinciding across subspecies, partly they are private. Exploration of individual breeds (not shown here, see Karimi, 2013) also indicated ROH islands that are private to breeds.

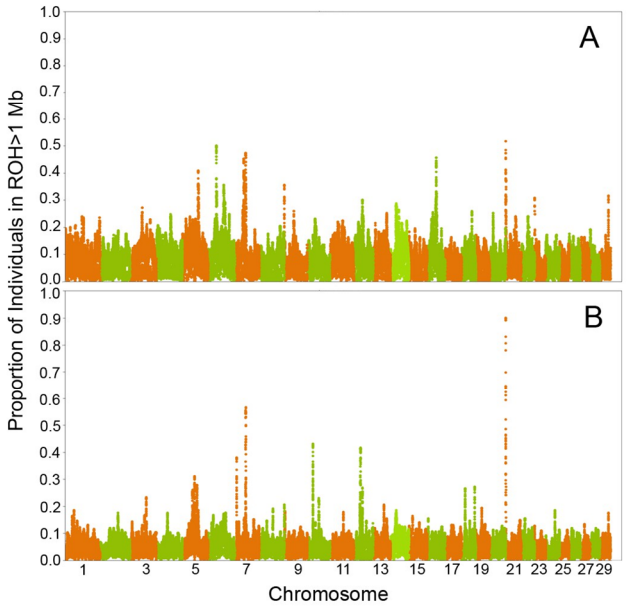


Figure 1: Manhattan plots for proportion of individual in ROH>1Mb for (A) taurine and (B) indicine pools of individuals

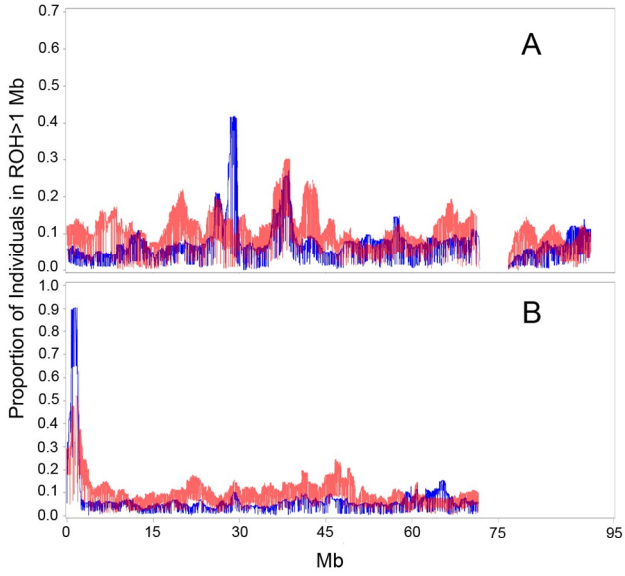


Figure 2: Local ROH frequencies at (A) BTA 12 and at (B) BTA 21. Red and blue color indicates taurine and indicine populations, respectively. The gap around Mb 75 at BTA12 is due to a genotype gap not allowing reconstruction of ROH in this region.

The regions surpassing frequencies of 40% in either taurine or indicine pools of individuals are shown in Table 1. These are four regions each, with two regions on BTA7 and BTA21 being more frequent than 40% in both subspecies. The sizes of ROH islands with borders as defined in Data and Methods were in the range of 0.5 to 1.2 Mb (Table 2). Kim et al. (2013) presented local patterns of autozygosity in three groups of Holstein Friesian cattle. The region on BTA7 seems to coincide with an ROH island in that paper.

Table 1: ROH islands, defined as regions with ROH frequencies >40% per SNP, in taurine and indicine cattle breeds

ROH island location	Physical Position(bp)	Frequency (%)	
		taurine	indicine
BTA6	38,268,200 : 39,451,000	54	-
BTA7	51,502,500 : 52,353,000	47	58
BTA10	24,575,700 : 25,619,800	-	45
BTA12	28,434,000 : 29,628,100	-	42
BTA16	43,802,200 : 44,968,700	44	-
BTA21	1,360,390 : 1,853,150	53	93

Table 2: Number of annotated elements (NCBI, Feb 2014) inside ROH islands and downstream neighbouring (+10Mb and +20Mb) regions of equal size.

CHR	Region size, kb	ROH Island	10Mb downstream	20Mb downstream
BTA6	1183	10	2	5
BTA7	851	17	5	1
BTA10	1044	4	0	3
BTA12	1194	7	0	1
BTA16	1166	20	0	5
BTA21	493	8	0	0

.The comparatively narrow region of ~850 Kb on BTA7 indicated in our study is hosting 17 annotated genes (Table 2), see Table 3 for acronyms of these genes. It was notable that the six regions inspected in detail were all gene-rich compared to regions of equal size 10 or 20 Mb downstream along each of the involved chromosomes (Table 2). Two ribosomal pseudogenes on BTA21were found to be orthologous with genes on the human genome (HAS3) that are located in one of the three peak ROH regions in the study of Nothnagel et al. (2010).

Table 3: Annotated genes within ROH islands.

CHR	Genes
BTA6	*MEPE, IBSP, LAP3, BT.29898, FAM184B, BT.100379, LCORL, BT.94996*
BTA7	*HSPA9, BT.63787, LRRTM2, SIL1, GPX4, BT.71626, PAIP2, SLC23A1, PACAP,SPATA24, DNAJC18, ECSCR, C5ORF65, 5S_rRNA, SNORA74, SNORA74*
BTA10	*TRAV14DV4, BT.64165, BT.101619*
BTA12	*PDS5B, N4BP2L2, N4BP2L1, BRCA2, ZAR1L, FRY, RXFP2*
BTA16	*BT.104317, BT.103198, DFFA,CORT,APITD1,PGD,KIF1B, UBE4B, RBP7, NMNAT1, CTNNBIP1*
BTA21	*OR5D13, U6, U6, 5S_rRNA, 5S_rRNA*

Checks of patterns of linkage disequilibrium in the ROH island regions indicated elevated LD in most of these regions but not to a degree that would explain the high incidence of ROH alone. This was similar to the findings of Nothnagel et al. (2010). ROH islands have been implicated with signals of strong selection (Nothnagel et al., 2010, Pemberton et al., 2012, Kim et al., 2013). While we found overlaps with QTL regions for all regions except for the one on BTA21, we did not explore this option in great detail.

Conclusions

Runs of homozygosity are obviously very unevenly distributed along the bovine genome. Extreme ROH islands (or hotspots), involving a large part of a population, appear in breeds and even across breeds within subspecies. The causes for such islands and their biological significance are still largely untapped (Wang et al., 2013). It is worthwhile to explore links with different types of indicators of selection (Utsunomiya et al., 2013) and the connection with patterns of linkage disequilibrium (Nothnagel et al., 2010). Next generation sequence data will require different types of ROH analysis due to rates of sequencing errors. First attempts have been performed by Bosse et al. (2013) in pigs and McLeod et al. (2013) in cattle, yet the field is open for exploration.

Literature cited

Bjelland, D.W., Weigel, K.A., Vukansinovic, N. et al. (2013). J. Dairy Sci. 96:4697-4706.

Bosse, M., Megens, H-J., Madsen, O., et al. (2013) PloS Genetics. 8:e1003100

Ferenčaković, M., Hamzić, E., Gredler, B., et al. (2013a). J. Anim. Breed. Genet. 130:286-293.

Ferenčaković, M. Sölkner, J. and Curik, I. (2013b). Genet. Sel. Evol. 45: 42.

Howrigan D.P., Simonson M.A., Keller M.C. (2011). BMC Genomics. 12: 460.

Karimi, Z. (1013). Runs of Homozygosity patterns in Taurine and Indicine cattle breeds. Master Thesis. BOKU – University of Natural Resources and Life Sciences, Vienna, Austria.

Kim, E-S., Cole, J.B., Huson, H. et al. (2013) PLoS one 8: e80813.

Lencz, T., Lambert, C., DeRosse, P. et al. (2007). PNAS 104:19942-19947.

McLeod, I.M., Larkin, D.M., Lewin, H.A. (2013). Mol Biol Evol 30:2209-2223.

McQuillan, R., Leutenegger, A.L., Abdel-Rahman, R. et al. (2008). Am. J. Hum. Genet. 83:359–372.

Nothnagel, M., Lu, T.T., Kayser, M. et al. (2010). Hum. Mol. Genet. 19:15.

Pemberton, T.J., Absher, D., Feldman, M.W. et al. (2012). Am. J. Hum. Genet. 91:275-292.

Purfield, D.C., Berry, D.P., McParland, S. et al. (2012). BMC Genetics 13:70.

Utsunomiya, Y.T., Perez O'Brien, A.M., Sonstegard, T.S. et al. (2013). PloS one 8:e64280

Wang, C., Xu, Z., Jin, G. et al. (2013). J. Biomed. Res. 27:208-214.

CURICULUM VITAE

Maja Ferenčaković was born on March 28, 1983 in Zagreb. She finished Master of Fisheries and Game Management at University of Zagreb, Faculty of Agriculture in March 2009, and in July of the same year she was accepted as a junior researcher at the Department of Animal Science, Faculty of Agriculture, on a scientific project "Strategies for elimination of genetic defects in selected populations". Since 2014 she has been employed as an assistant researcher at the same institution. In 2008 she received Rector's award for student scientific research. From 2009 until today she won few international scholarships and awards that allowed her 18 months of scientific specializations in Spain and Austria, as well as participation at international conferences. She published five articles indexed in a1 databases (three as the first author), four in a2 databases and two in a3 database. She participated in 16 scientific conferences in Croatia and abroad. Presently, she is a teaching assistant at several undergraduate and graduate courses and she is involved as a researcher in international and domestic scientific projects.

List of publications according to CROSBI (*https://bib.irb.hr/lista-radova?autor=320876*)

Journal articles and review articles in CC journals (5)

Curik, I., Ferenčaković, M. & Sölkner, J. (2014) Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock science.* **166**, 1-9.

Ferenčaković, M**.**, Curik, I., Perez-Pardal, L., Royo, L.J., Čubrić Čurik, V., Fernandez, I., Alvarez, I., Kostelić, A., Šprem, N., Krapinec, K. & Goyache, F. (2013) Mitochondrial DNA and Y-chromosome diversity in East Adriatic sheep. *Animal genetics.* **44**, 184-192.

Ferenčaković, M**.,** Hamzić, E., Gredler, B., Solberg, T.R.; Klemetsdal, G., Curik, I. & Sölkner, J. (2013) Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. *Journal of Animal Breeding & Genetics.* **130,** 286-293.

Ferenčaković, M**.**, Soelkner J. & Curik I. (2013) Estimating autozygosity from high-throughput information: effects of SNP density and genotyping errors. *Genetics selection evolution*. **45**, 42.

Čubrić Čurik, V., Feligini, M., Ferenčaković, M., Džidić, A., Salajpal, K., Ambriović-Ristov, A., Ćetković, H., Majhen, D. & Curik, Ino. (2009) Sequence polymorphism of PrP exon 3 gene in Istrian and crossbred sheep. *Italian journal of animal science*. **8**, 86-88.

Scientific papers in other journals (4)

Zavarez L.B., Utsunomiya Y.T., Carmo A.S., Neves H.H.R., Carvalheiro R., Ferencakovic M., Pérez O'Brien A.M., Curik I., Cole J.B., Van Tassell C., Silva M.V., Sonstegard T.S., Sölkner J. & Garcia J.F. (2015) Assessment of autozygosity in Nellore cows (Bos indicus) through high-density SNP genotypes. *Frontiers in Genetics* **6**.

Curik I., Ferenčaković M., Karapandza N., Cubric Curik V. & Sölkner J. (2014) Estimation of inbreeding and effective population size in Istrian cattle using molecular information. *Acta Agraria Kaposváriensis* **18**, 30-4.

Ribarić, J., Lončar D., Balinović I.; Ferenčaković, M**.** & Salajpal, K. (2012) Effect of season on variation in somatic cell count in milk. *Stockbreeding: journal for the improvement of the stockbreeding*. **66,** 137-147.

Ferencakovic, M., Hamzic, E., Gredler, B., Curik, I. & Sölkner, J. (2011) Runs of Homozygosity Reveal Genome-wide Autozygosity in the Austrian Fleckvieh Cattle. *Agriculturae conspectus scientificus*. **76,** 325-328.

Conference reports (abstracts) in other journals (4)

Ferencakovic, M**.**, Hamzic, E., Fuerst, C., Schwarzenbacher, H., Gredler, B., Solberg, T.R., Curik, I. & Sölkner, J. (2011) Runs of homozygosity and levels of inbreeding in cattle breeds. In: *Book of Abstracts of the 62st Annual Meeting of the European Federation of Animal Science*, / *Wageningen Academic Publishers (ed.). Wageningen Academic Publishers*. 176-176.

Cubric Curik, V., Frkonja, A., Kostelic, A., Bokor, A., Ferencakovic, M**.**, Tariba, B., Druml, T. & Curik, I. (2010) Revitalization of Medimurje horse: mitochondrial DNA evidence. In: *Book of Abstracts of the 61st Annual Meeting of the European Association of Animal Production*, / *Wageningen Academic Publishers (ed.). Wageningen Academic Publishers*. 152-152.

Cubric-Curik, V., Kostelic, A., Feligini, M**.,** Ferencakovic, M., Ambriovic-Ristov, A., Cetkovic, H. & Curik, I. (2010) Sequencing of prion protein gene (PRNP) in East

Adriatic sheep populations revealed new nonsynonymous polymorphisms linked to ARQ allele. In: *Book of Abstracts of the 61st Annual Meeting of the European Association of Animal Production, / Wageningen Academic Publishers (ed.). Wageningen Academic Publishers*. 349-349.

Sölkner, J., Ferencakovic, M., Gredler, B. & Curik, I. (2010) Genomic metrics of individual autozygosity, applied to a cattle population. In: *Book of Abstracts of the 61st Annual Meeting of the European Association of Animal Production*, / *Wageningen Academic Publishers (ed.). Wageningen Academic Publishers*. 306-307.

Scientific conference papers with international peer-review (3)

Sölkner, J., Ferenčaković, M., Karimi, Z., Perez O'Brien, A.M., Mészáros, G., Eaglen, S., Boison, S.A., & Curik, I. (2014). Extremely Non-uniform: Patterns of Runs of Homozygosity in Bovine Populations. In: *Proceedings - 10th World Congress on Genetics Applied to Livestock Production* / *American Society of Animal Science (ed) Champaign; 10th World Congress on Genetics Applied to Livestock Production.* 1-4.

Frkonja A., Kostelić A., Čurik I., Bokor A., Druml T., Belčić M., Prvanović N., Ferenčaković M., & Čubrić Čurik V. (2011) Phylogenetic analisys of Medjimurje horse. In: *Book of Abstracts of the 46th Croatian & 6th International Symposium on Agriculture* / *Milan Pospišil, (ed) Zagreb: University of Zagreb, Faculty of Agriculture.* 846-849.

Curik, I., Ferencakovic, M., Gredler, B. & Sölkner, J. (2010) Genome-wide heterozygosity and pedigree inbreeding coefficients in Simmental cattle population. In: *Proceedings - 9th World Congress on Genetics Applied to Livestock Production / German Society for Animal Science (ed.). Gießen: 9th World Congress on Genetics Applied to Livestock Production*. 1-4.

Abstracts in Book of abstracts (10)

Ferenčaković, M**.**, Curik, I. & Sölkner, J. (2014) Modeling genome wise dissection of inbreeding depression for bull fertility traits. In: *Program and Abstracts of Livestock Genomic Resources in a Changing World Conference / Bruford, Michael W. ; Orozco-terWengel, Pablo ; Costa, Mafalda, (ed.). Cardiff,* 17-17.

Curik, I., Ferencakovic, M. **&** Soelkner, J. (2012) Genomic dissection of inbreeding depression: Lessons from the bull fertility traits. In: *Book of Abstracts of the Fourth International Conference on Quantitative Genetics*, Edinburg, Scotland.

Curik, I., Ferenčaković, M**.** & Sölkner J. (2012) Modeling perspectives in the estimation of inbreeding depression based on genomic information: lessons from the bull fertility. In: *Genetika 2012: book of abstracts, 6th Congress of the Genetic Society of Slovenia [and] 6th Meeting of the Slovenian Society for Human Genetics with International Participation / Uroš Potočnik (ed.). Ljubljana: Genetic Society of Slovenia*. 82-82.

Ferencakovic, M**.**, Curik, I., & Soelkner, J. (2012) Estimation of genomic inbreeding in cattle: the impact of density. In: *Book of Abstracts of the 63rd Annual Meeting of the European Federation of Animal Science / Wageningen Academic Publishers (ed.). Wageningen : Wageningen Academic Publishers*, 12-12.

Ferenčaković, M., Curik, I. & Soelkner, J. (2012) The impact of high density SNP chip on the estimation of genomic inbreeding in cattle In: *Book of Abstracts of the 47th Croatian & 7th International Symposium on Agriculture / Milan Pospišil (ed.). Zagreb: University of Zagreb, Faculty of Agriculture*.

Ferenčaković, M**.**, Curik, I. & Sölkner, J. (2012) Estimation of genomic inbreeding using runs of homozygosity - cattle experience. In: *Book of Abstracts of the 3rd Congress of Croatian Geneticists with international participation / Jasna Franekić, Verica Garaj-Vrhovac (ed.). Zagreb: Croatian Genetic Society,* 75-75.

Ferenčaković, M., Sölkner, J. & Curik, I. (2012) Genomic analysis of inbreeding depression on sperm quality traits in the Austrian Fleckvieh cattle. In: *Book of Abstracts of the 47th Croatian & 7th International Symposium on Agriculture / Milan Pospišil (ed.). Zagreb : University of Zagreb, Faculty of Agriculture*.

Ferenčaković, M., Sölkner, J. & Curik, I. (2012) Genomic analysis of inbreeding depression: Lessons from bull fertility traits. In: *Book of Abstracts of the 3rd Congress of Croatian Geneticists with international participation / Jasna Franekić, Verica Garaj-Vrhovac (ed.). Zagreb: Croatian Genetic Society,* 70-70.

Ferenčakovic, M., Royo, L.J., Pérez-Pardal, L., Čurik, I., Fernández, I., Álvarez, I., Čubrić Čurik, V., Kostelić, A., Sprem, N., Krapinec, K. & Goyache, F. (2011) Genetic characterization of Croatian autochthonous sheep populations: Analysis of mtDNA and Y chromosome. In: *46th Croatian & 6nd International Symposium on Agriculture: book of abstracts / Pospišil, Milan (ed.). Zagreb: Agronomski fakultet*, 223-224.

Ferenčaković, M., Royo, L.J., Pérez-Pardal, L., Čurik, I., Fernández, I., Álvarez, I., Čubrić Čurik, V., Kostelić, A., Šprem, N., Krapinec, K. & Goyache, F. (2010) Y chromosome and mtDNA polymorphism in Eastern Adriatic sheep populations. In: *Abstracts of the 32nd International Conference on Animal Genetics,* 72-72.