Master Thesis

Genome-wide association study and genomic prediction of sperm quality traits in Austrian pigs

Submitted by

Martin BAUER, BSc

in the framework of the Master programme

Nutztierwissenschaften

in partial fulfilment of the requirements for the academic degree

Diplom-Ingenieur

Vienna, December 2021

Supervisors:

Univ.-Prof. Dr. Johann Sölkner
Division of Livestock Sciences
Department of Sustainable Agricultural Systems

Dr. Christina Pfeiffer, PIG Austria GmbH

Master thesis

Genome-wide association study and genomic prediction of sperm quality traits in Austrian pigs

Martin Bauer

11770627

Livestock Sciences UH 066 456

Supervisors:

Univ.-Prof. Dr. Johann Sölkner
Dr. Christina Pfeiffer

Division of Livestock Sciences

Department of Sustainable Agricultural Systems

University of Natural Resources and Life Sciences Vienna
Gregor-Mendel-Straße 33
1180 Wien
Austria

E-Mail: martin.bauer@boku.ac.at

Statutory Declaration

I hereby declare that I prepared this thesis independently. No assistance other than that which is permitted has been used. Ideas and quotes taken directly or indirectly from other sources are identified as such and are declared in the text or duly acknowledged. This written work has not yet been submitted in any part.

Signature: _____   Date: _____

Abstract

In the present study a genome-wide association study (GWAS) for three Austrian pig populations (Landrace, Large White and Pietrain) was carried out. Genetic parameters and breeding values were estimated for the Pietrain population for semen quality and quantity traits. The traits inferred were total number of sperm, motility, volume and density.

Trait recording was done using a CASA (computed aided sperm analysis) system from 2011 to 2021 and delivered data of 127,544 ejaculations from 2,276 boars, of which the majority belonged to Pietrain breed. The animals were kept at three Austrian AI stations. Genomic data was assessed using Illumina 60k and 80k SNP arrays, yielding 43,430 markers from 1,233 individuals after quality control. The genome-wide association was based on a single-breed mixed linear model association with phenotypic pre-correction. Genetic parameters and breeding values were estimated for the Pietrain breed using a multivariate single-step GBLUP (ssGBLUP) procedure.

The GWAS revealed some possible associations for total number of sperm in Landrace on chromosome 1 in the region of the AGPAT4 gene and in Pietrain on chromosome 1, 8, 14 and 16, pointing to two genes from the cadherin family (PCDH15, CDH12) as well as others.

Estimated heritabilities for total number of sperm and motility were 24.7 % and 13.1 %, respectively. We discovered a positive genetic correlation (0.205) between motility and sperm density, but negative genetic correlation (-0.307) between motility and ejaculate volume.

The breeding value estimation provided proof of the advantage of ssGBLUP compared to pedigree-based BLUP in this study as reliabilities (r²) for young boars without phenotypes were substantially higher (0.369 vs. 0.217 for total number of sperm, 0.270 vs. 0.161 for motility). The overall reliabilities were relatively high for pig evaluations. This implies high selection response if included in routine selection. A negative genetic trend for total number of sperm over the last ten years was observed, indicating the need of monitoring in the future.

Keywords

Overview

Abbreviations

A: numerator relationship matrix (pedigree-based)

AI: artificial insemination

AI-REML: average information restricted maximum likelihood

EM-REML: expectation-maximization restricted maximum likelihood

G: genomic relationship matrix (if not defined otherwise)

GWAS: Genome-wide association study

H: combined relationship matrix in single-step GBLUP, created from G and A

LD: Linkage disequilibrium

LR: Landrace (pig breed), here referring to the Austrian Landrace population

LW: Large White (pig breed), here referring to the Austrian Large White population

PI: Pietrain (pig breed), here referring to the Austrian Pietrain population

QC: quality control

REML: restricted maximum likelihood

SEP: standard error of prediction

SNP: single nucleotide polymorphism, a type of genetic marker

1. Introduction

In Austria, the majority of piglets are produced using artificial insemination (AI). This applies to crossbreeding systems for pork production as well as purebreeding. Currently more than 90 % of sows are inseminated artificially. In the past years, insufficient semen quality proved to be the top reason for eliminating boars from the use in artificial insemination, as data from AI station Steinhaus showed. This has manifested throughout all breeds, only in Landrace low semen quality and low breeding value have been sharing rank one for the most important elimination causes (Pfeiffer, 2020).

While female fertility in pigs has been in the focus of breeding programs for decades, male fertility (especially sperm quality and quantity) has not received much attention yet.

Phenotypic data on porcine male fertility traits (for example, volume and cell density of ejaculates, sperm motility) were collected systematically for many years in Austrian AI stations. Genotypic data have become available from 2011 on for many pigs in the breeding populations. Since 2016, when routine genomic selection was introduced, the number of genotyped pigs increased tremendously. While data are available, no genetic or genomic analyses of these traits have been carried out in Austria so far.

2. Aims of the thesis

The study examined the genetic background of sperm quality and quantity traits in boars and aimed to transfer this knowledge for practical breeding decisions. This should be carried out based on all available AI and genome data for Austrian pig populations. As far as possible, the study aimed to examine all major pig breeds (Large White, Landrace, Pietrain) in Austria.

The main objectives were:

- To conduct a genome-wide association study to find out which regions of the genome have a major influence on sperm quality/quantity traits.
- To estimate genetic parameters (variance components, heritabilities, phenotypic and genotypic correlations) of sperm quality/quantity traits.
- To predict (genomic) breeding values and assess genetic trends of sperm quality/quantity traits.

3. Literature review

3.1.   Pig production in Austria

In 2019 2,77 million pigs were kept on about 21.000 farms. Most of the pigs were fattening pigs and producing sows. The nucleus breeding populations were rather small comprising 7,515 sows and 889 boars belonging to Pietrain (PI), Large White

(LW) and Landrace (LR) (BMLRT - Ministry of Agriculture Regions and Tourism, 2020) The small proportion of the breeding population compared to the total population is a result of the dominant mating systems in pig production: In Austria fattening pigs are mostly crosses between at least two breeds. Most piglets are produced from a three-way cross of an LW x LR sow inseminated with a PI boar. The most relevant alternative is a single-cross of a pure LW sow and a PI boar. Other breeds like Duroc and Schwäbisch-Hällisch only play a minor role.

This leads to a few small purebreeding populations responsible for achieving selection response, the largest ones being the mentioned breeds PI, LW and LR. In total 134 farms were recognised as herdbook breeders in 2019 (BMLRT - Ministry of Agriculture Regions and Tourism, 2020).

3.2.   Organisation of pig breeding and insemination

The formerly independent pig breeding organisations of Upper Austria, Lower Austria and Styria fused into one new breeding organisation 'Schweinezucht Österreich eGen' in 2019. It owns the subsidiary company PIG Austria GmbH, which is the operative platform for both breeding and AI (Pig Austria GmbH, 2021a). PIG Austria is responsible for the entire breeding programs, including performance testing activities like data collection on farms and station, breeding value estimation, selection, sale of breeding stock and production of semen doses for artificial insemination (Pig Austria GmbH, 2021b).

The production of artificial insemination doses is carried out on three stations: Steinhaus in Upper Austria, Hohenwarth in Lower Austria and Gleisdorf in Styria. All three stations keep boars, do the quality examination in the laboratory and produce semen doses on site (Pig Austria GmbH, 2021b).

3.3.   Sperm quality assessment and requirements for porcine sperm in artificial insemination

The routine assessment of boar ejaculates in Austrian AI recorded several values: date, volume, cell density (sperm cells per millilitre), total motility. These were used to calculate the dilution to a given number of sperm cells per dose, yielding different numbers of doses per ejaculate. These assessments were done manually under a microscope for decades but have been replaced by automated systems (CASA – *computer aided sperm analysis*) in recent years in all stations. A main advantage of using CASA systems is objectivity and comparability of the outcome, although this applies only when using the same CASA system. The other reason for widespread use is the economic advantage as CASA systems save expensive manual labour (Amann & Waberski, 2014).

Beside just counting sperm cell numbers, CASA systems are able to assess many movement parameters. These parameters are used to distinguish single cells into "motile/immotile" or "progressively motile/non-linear motile/immotile". The

corresponding percentages are then displayed as "motility" or "progressive motility", respectively. The detailed movement parameters typically include curvilinear velocity, average path velocity, straight-line velocity, amplitude of lateral head displacement and beat-cross frequency. These are briefly explained in Figure 1. Thresholds for these measurements are then used to distinguish sperm cells into the mentioned groups. For more details see e.g. Amann & Waberski (2014).
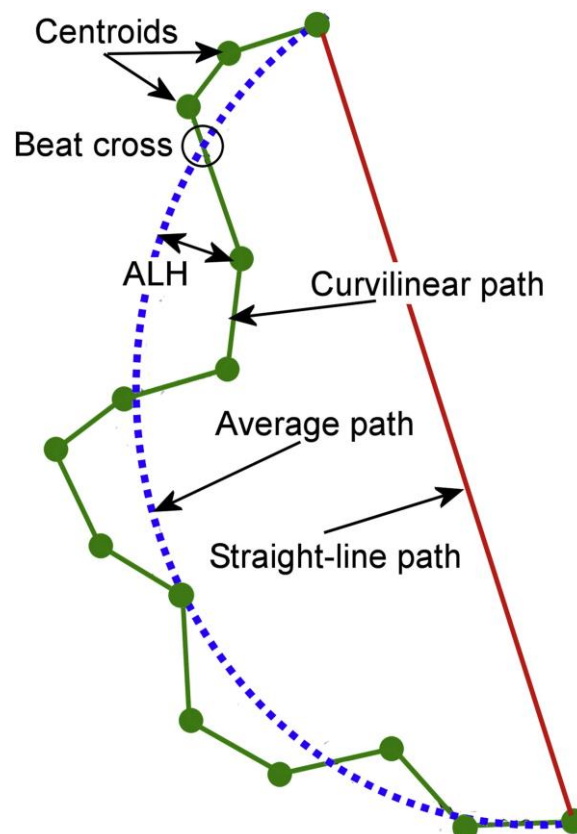
Figure 1: Schematic illustration of CASA motion measurements: the green dots represent the head of a sperm cell, the green line is its actual path (the velocity along this path is the curvilinear velocity, VCL; µm/s), the blue dotted line is the computed average path (the velocity along this is the average path velocity, VAP; µm/s), the red line represents the straight path (the velocity along this is the straight-line velocity, VSL; µm/s), the number of crossings across the average path is called beat-cross frequency (BCF) and the average deviation from the average path is the amplitude of lateral head displacement (ALH; µm) (Amann & Waberski, 2014)

Note: By Amann & Waberski (2014).

In Austrian pig production, the PIG Austria stations guarantee a set of quality criteria for each insemination dose (Pfeiffer & Kreiner, 2020):

- At least 1.8 billion sperm cells (total) per dose (85 ml)
- Max. 25 % morphologically abnormal sperm (including cytoplasmic droplets)
- Max. 15 % cytoplasmic droplets
- Motility after 24 h > 75 %
- Motility after 72 h > 65 % (also used as criterion for maximum storage time)

The focus on cytoplasmic droplets is due to studies that have shown clear negative correlations between cytoplasmic droplet rates and farrowing rates in pigs (Lovercamp et al., 2007), as well as negative effects on fertility in other species (Cooper, 2011).

Although these associations have been known for a long time and it has been recommended to keep the total rate of cytoplasmic droplets low, e.g. below 20 % (Althouse, 1998), there is little knowledge about the mechanisms behind this abnormality.

Other authors report dosage in a similar range compared to Austrian standards. In two papers $4*10^9$ sperm cells (Althouse et al., 1998) or $3*10^9$ sperm cells (Tsakmakidis et al., 2010) were used for a dose in artificial insemination for pigs. In general, numbers between 0.5 and 5 billion sperm cells have been used in various papers and recommendations.

Note that a more specific examination on morphological abnormalities and other traits was carried out for every boar at the beginning of its service at the AI station. This was carried out at the laboratory of the University of Veterinary Medicine Vienna. These data were not used in the present study.

3.4. Spermatogenesis in domestic pigs (*Sus scrofa domesticus*)

The formation of haploid gametes (oocytes and spermatozoa) is an essential precondition for the sexual reproduction cycle of most eukaryotic lifeforms, in which haploid gametes fuse to a diploid zygote becoming the organism that later in life again produces haploid gametes. In male mammals the process of forming gametes is called spermatogenesis and takes place inside the seminiferous tubules of the testes. This process consists of three main parts:

- Mitotic proliferation of spermatogonia (the cells that later undergo meiosis), to reproduce these stem cells and keep spermatogenesis possible throughout the life.
- Meiosis, the process of meiotic division and recombination of the chromosome set to form four haploid spermatids from one diploid spermatogonium.
- Spermiogenesis, the process in which spermatids differentiate into finished spermatozoa (Hale, 1996).

3.4.1. Details of the male reproductive tract

The seminiferous tubules are small tubular structures filling up most of the testis volume, for example 70 % in stallions' testes (Johnson et al., 1997). They are consisting of spermatogonia, spermatids and their further stages (meiotic cells), and somatic Sertoli cells that are building the structure of the tubules attached to a basal lamina. While the meiotic cell lines are mostly small and round shaped, the Sertoli cells are large cells with irregular shape. In figure 2 they can be seen as long, narrow structures between the round meiotic cell types. They bind the meiotic cells during their development from spermatogonia to fertile spermatozoa and deliver nutrients and resources. The Sertoli cells therefore have to maintain flexible binding structures to the meiotic cells throughout the whole process. This function is necessary as the Sertoli

cells form the blood-testis barrier by tight junctions between them, separating the inside of the seminiferous tubules totally from the blood stream (Hale, 1996). A secondary function is the recycling of resources, as they phagocytose the surplus cytoplasm (called "residual body") of finished spermatozoa as well as all kinds of defective sperm cells. This phagocytotic function proved to be important, as individuals with inhibited phagocytosis produced lower numbers of viable sperm (Nakanishi & Shiratsuchi, 2004).

Outside, the seminiferous tubules are surrounded by Leydig cells which have hormonal activity, muscle tissue to allow contractive movement and blood capillaries supplying resources. The inner part of the tubules is the lumen, it stores and transports the finished spermatozoa (Hale, 1996).
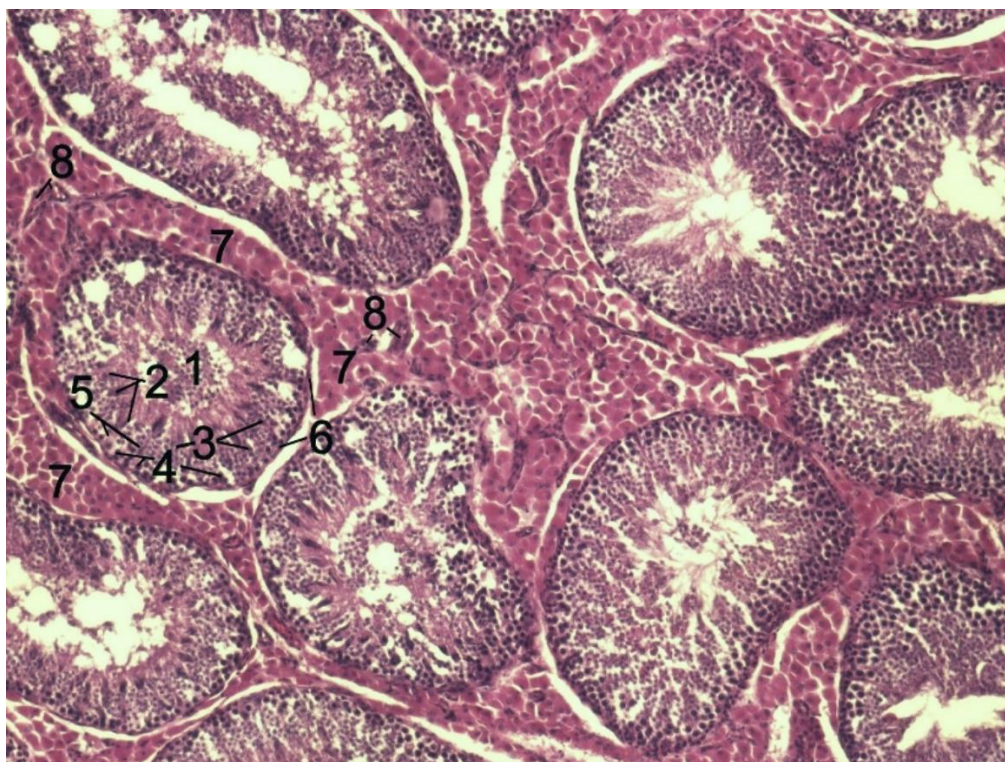
Figure 2: A section of the testis parenchyma of boars: 1 - seminiferous tubule lumen; 2 - spermatids; 3 - spermatocytes; 4 - spermatogonia; 5 - Sertoli cells; 6 - myofibroblasts; 7 - Leydig cells; 8 – capillary

Note: By Gille (2006), Wikimedia Commons (https://commons.wikimedia.org/wiki/File:Testicle-histology-boar.jpg?uselang=de). GFDL 3.0.

The spermatozoa are then stored in the epididymis, a coiled structure outside the testis, until ejaculation (Hale, 1996). During this storage period the sperm cells gain their motility and develop the typical swimming pattern (Yeung & Cooper, 2002).

3.4.2. Mitotic proliferation

The first step in spermatogenesis is the continuous mitosis of spermatogonia which are located in the base of the tubules' structure, still outside the tight junctions (Hale, 1996). They reproduce themselves to ensure a constant supply of these stem cells,

and part of them wanders inside and becomes primary spermatocytes. Different types of spermatogonia have been described (Johnson et al., 1997).

3.4.3. Meiosis

The meiosis is the core process halving the diploid genome to form haploid gametes. Before meiosis, the primary spermatocyte (a direct mitosis product from one spermatogonium) goes through the synthesis (S) phase of the cell cycle to become a double-chromatid diploid cell. In meiosis I, it divides into two haploid, double-chromatid cells that are now called secondary spermatocytes. It is this process (more exactly: the prophase of meiosis I) where recombination takes place by crossing-over between the two homologous chromosomes/four chromatids, creating most of genetic variability. Secondary spermatocytes are rarely seen in microscopic analyses, as they rapidly undergo meiosis II. Meiosis II is similar to mitosis, separating the sister chromatids of the two secondary spermatocytes to create four haploid, single-chromatid cells. These are known as spermatids and are still round in shape. (Hale, 1996)

3.4.4. Spermiogenesis

In spermiogenesis, the spermatid cells transform to the well-known spermatozoa with their unique shape and function, shown in Figure 3. The nucleus becomes massively compacted, rendering the genome transcriptionally almost inactive. To achieve this, histones are replaced by protamines, small (49-63 amino acids in mammals), arginine-rich, highly basic proteins (Balhorn, 2007). One of the centrioles forms the flagellum, the other disappears. Around the base of the flagellum mitochondria are concentrated to deliver energy directly to the main energy consumer. This structure is known as the mid piece and is usually not transmitted to the zygote in fertilization (causing the maternal inheritance of mitochondrial DNA). The head of the sperm cell gets covered in the acrosome, built from the cell's Golgi apparatus. It contains the enzymes necessary for entering the oocyte membrane. (Hale, 1996). As the mature sperm head is much smaller than the spermatid cell before, much of the cytoplasm is removed as a "residual body" that is later phagocytosed by a Sertoli cell. Until its release the spermatozoon is still connected to this residual body with a "stalk" of cytoplasm. When this connecting stalk is not removed, it forms a cytoplasmic droplet at the mid piece, a well-known abnormality of sperm cells (Johnson et al., 1997).
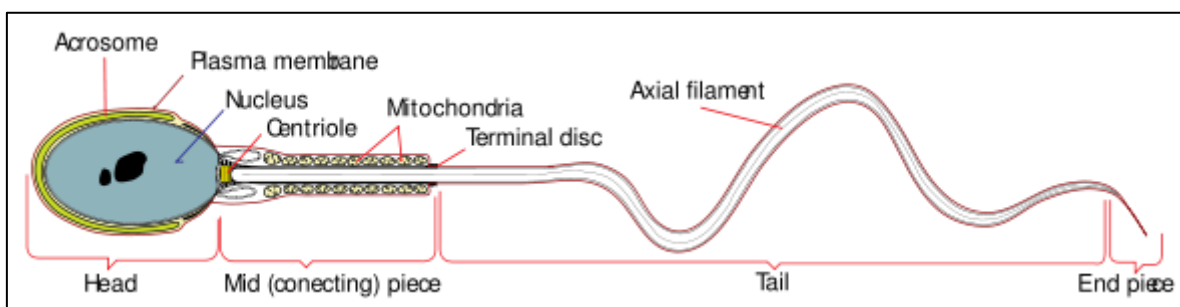

Figure 3: A simple diagram of a spermatozoon.

Note 2: The nucleus is the large black body in the middle (arrow points at the wrong spot), "conecting" should mean "connecting".

3.5. Environmental effects on male fertility

Several environmental effects on sperm quality and fertilizing ability are known. High temperatures and/or high humidity cause heat stress which is known to reduce sperm motility and increase the percentage of various abnormalities. However, pigs can adapt quite well to constant high temperatures, but fail to adapt to high temperature differences between day and night (Kunavongkrit et al., 2005). This causes problems especially in the summer months in temperate climates. Schulze et al. (2014) reported significant effects of season (winter/spring/summer/autumn) on total number of sperm, with spring and summer effecting it negatively compared to the other two seasons. However, they did not find any significant effect on motility or abnormalities.

A photoperiodic effect has been reported, as wild boars are known to have a seasonally changing fertility depending on day length. A study stated that constant 18 h artificial lighting was able to neutralise negative fertility effects on sows in September/October in a European pig housing (Kunavongkrit et al., 2005).

Diseases can influence sperm quality. Schulze et al. (2013) found that a PRRS infection increased the percentage of non-linear motile sperm and acrosome-defective sperm cells, but did not change volume, density or total motility of the ejaculate.

Breed has a large impact on sperm traits and has to be considered when comparing sperm quantity and quality results. Significantly different total numbers of sperm and various abnormality rates have been reported between Landrace, Yorkshire, Large White, Duroc and Pietrain boars (Schulze et al., 2014).

3.6. Quantitative genetics

Previous studies estimated low to moderate heritabilities for quantitative semen traits. Marques et al. (2018) estimated heritabilities for motility and total number of sperm per ejaculate for LW and LR population. Authors reported heritabilities of 21 and 12 % for motility for LW and LR and 10 and 13 % for total number for LW and LR of sperm per ejaculate, respectively. In an earlier publication heritabilities in the same range were mentioned (Wolf, 2010): 10 % for motility and total number of sperm cells, 18 % for sperm concentration and 20 % for sperm volume in both LW and LR. A study from Smital et al. (2005) estimated higher heritabilities: 38 % for motility and 42 % for total number of sperm. It should be mentioned that the authors did not use a repeatability model, but averaged the values over the whole lifetime of each boar, eliminating the variance between ejaculations. (Smital et al., 2005)

The heritabilities for some distinct morphological abnormalities seem to be higher, as Zhao et al. (2020) found values of 2,9 % for coiled tail, 13,7 % for bent tail, 24,4 % for proximal droplet, 29,5 % for distal droplet and 26,7 % for distal midpiece reflex.

Genetic correlations between sperm traits and to other production traits have been reported. Wolf (2010) found high negative correlations between sperm volume and density, as well as between morphological abnormalities and motility. In the same study some unexpected results were found: positive genetic correlations between percentage of abnormal sperm and litter size in Large White sows, but negative genetic correlations between these traits in Landrace. Broekhuijse et al. (2012) found significant positive genetic correlations between sperm motility parameters and farrowing rate/total number of pigs born, although AI stations correct for the motility via semen dilution.

3.7. Genomic research

In recent times more extensive research has been done to understand the genetic basis of spermatogenesis and sperm quality. Some abnormalities with massive effects could be traced down to causative mutations, like a deletion in DNAH17 (dynein axonemal heavy chain 17) on chromosome 12 that caused defective flagella, leading to infertility of eight inbred boars (Nosková et al., 2021). In mice, a large number of loci are known to be linked to infertility, mostly by interrupting spermatogenesis, e.g. mutations in c-KIT, mHR6B, BMP8B, RXRβ, dhh, CREM, Hsp70-2, PMS2, MLH1 and ATM (Elliott & Cooke, 1997).

Aside from such massive impacts, the individual differences in sperm abnormality rates seem to have a polygenic background in pigs (Zhao et al., 2020). This suggests that other semen assessment parameters like motility and sperm cell number are also of highly polygenic nature. Godia et al. (2020) reported a large number of possible causative regions for sperm quality traits. A study from Marques et al. (Marques et al., 2018) used an approach to find regions that explain most of the genetic variance and found candidate genes SCN8A, PTGS2, PLA2G4A, DNAI2, IQCG, LOC102167830 in Large White and NME5, AZIN2, SPATA7, METTL3 and HPGDS in Landrace within these regions. These genes were completely different between two closely related white pig breeds and the study could not explain any reason for that.

4. Animals, materials and methods

4.1. Animals, materials and data

4.1.1. Analysis equipment

Sperm quality assessment was carried out using the CASA (computer aided sperm analysis) system *AndroVision* from German manufacturer Minitüb GmbH. These semi-automatic systems are in use at all three stations and were upgraded to a fully automatic system with the *eFlow* chambers at slightly different times. These upgrades

were put in effect at Gleisdorf on July 21, 2020, at Hohenwarth on May 12, 2020 and at Steinhaus on July 1, 2020.

4.1.2. Sperm quality and quantity traits

The routine assessment consisted of three parameters: ejaculate volume (in ml), ejaculate density (in billion cells per ml) and total motility (in %). As volume and density were varying, we used their product, the total number of sperm cells per ejaculate (in billion cells) as main quantity parameter. Semen doses were diluted to a constant number of sperm cells, so this trait directly reflects the productivity of an ejaculate. The motility measured was the percentage of motile sperm cells in the ejaculate (not corrected for direction or speed of motion, like e.g. progressive motility).

Additionally, the maximum storage time was recorded. This was defined as the number of days before the motility of the diluted doses falls below 65 %, leading to them not being saleable anymore. This value was recorded by hand (microscopic evaluation of motility), not automatically, which can cause less objective or reliable measuring. The storage time records have not been carried out for the whole time, so data were not complete for this trait. Of total 127,544 ejaculates, 28,552 did not have storage time records.

4.1.3. Descriptive statistics

Prior to analysis data were filtered to remove measurement errors (e.g., records with zero volume, decimal point errors, etc), unplausible values and records from boars which are not closely related to the Austrian populations. The restrictions applied were:

- no unrelated boars from imports (as mentioned above)

- ejaculation interval between 3 and 30 days

- age of the boar at ejaculation ≥ 200 days

- total number of sperm in an interval of ± 3 standard deviations from the mean

- motility > 0,1 % quantile of the data

These restrictions reduced the total number of records from 150,966 to 127,544. The data was collected from ejaculates from January 2012 to January 2021 from all three breeds. The three stations differ in their contribution to the data set as their size was different and the CASA systems were introduced at different time points: Steinhaus data were available for the entire time period (Jan 2012 to Jan 2021), Hohenwarth included data from Jan 2018 to Jan 2021 and Gleisdorf from Jan 2017 to Dec 2020, respectively. An overview of the distribution on AI stations and breeds is given in Table 1:

Table 1: Number of observations (ejaculates) on AI stations Gleisdorf, Hohenwarth and Steinhaus for Large White (LW), Landrace (LR) and Pietrain (PI)

AI Station	Breed	Number of observations	
Gleisdorf	LW	2,524	
	LR	363	17,572
	PI	14,685	
Hohenwarth	LW	1,018	
	LR	1,020	23,775
	PI	21,737	
Steinhaus	LW	4,543	
	LR	6,576	86,197
	PI	75,078	

The data set consists of the three breeds Pietrain, Large White and Landrace. Most of the boars were Pietrain boars (78.8 %) for piglet production, leading to Pietrain delivering the largest part of the records (87.4 %).

Table 2 shows the mean and standard deviation of the five traits and the two regression covariates, sorted by breed:

Table 2: Means and standard deviations of total number of sperm [bn], motility [%], volume [ml], density [ml^{-1}], storage time [d], age [d], ejaculation interval [d] and number of observations per breed for Large White (LW), Landrace (LR) and Pietrain (PI).

Breed	Total number of sperm [bn]	Motility [%]	Volume [ml]	Density [bn * ml^{-1}]	Storage time [d]	Age [d]	Ejaculation interval [d]	Number of obs.
LW	88.04 +-31.83	84.66 +-8.41	229.2 +-97.1	0.4346 +-0.2082	3.737 +-4.684	725.8 +-361.4	9.150 +-5.429	8,085
LR	89.60 +-32.47	88.30 +-7.75	248.5 +-105.5	0.4066 +-0.1922	4.685 +-2.662	667.6 +-369.0	10.544 +-5.787	7,959
PI	83.45 +-30.00	87.30 +-8.14	257.3 +-97.5	0.3601 +-0.1709	4.102 +-3.286	844.2 +-501.0	7.268 +-3.200	111,500
Total	84.12	87.20	255.0	0.3677	4.115	825.7	7.592	127,544

Pietrain also contributes most of the individuals (boars), shown in Table 3.

Table 3: Numbers of breeders and boars of Large White (LW), Landrace (LR) and Pietrain (PI)

Breed	Number of breeders	Number of boars
LW	28	209
LR	33	272
PI	38	1,795
Total	78[a]	2,276

[a] *Note:* The total number of breeders (farms) was smaller than the sum of the three values above as some herdbook breeders keep more than one breed.

4.1.4.   Pedigree data

We used the total available pedigree for each breed, consisting of 110,885 pigs (92,656 females, 18,229 males) for LW, 79,357 animals (62,394 females, 16,963 males for LR and 66,825 animals (35,227 females, 31,598 males) for PI.

4.1.5.   Genomic data

In total, 1,233 genotypes of boars of all breeds were available, resulting in 155 LW, 169 LR and 909 PI boars. The genotyping was carried out using either the Illumina Porcine SNP60 v2.0 BeadChip or the Illumina GeneSeek Genomic Profiler (GGP) 80k array. Both chips were used in all three breeds. The 60k chip consisted of 59,319 SNPs, the 80k of 77,122 SNPs. In all data sets the base-pair positions were updated to the most recent pig genome build *Sscrofa 11.1* (Warr et al., 2020) using the publicly available Illumina manifest on the SNP60 v2.0 BeadChip (Illumina Inc., 2021). It has to be noted that not all the SNPs on the smaller 60k chip were on the larger 80k chip. For consistency the three breeds were merged to one large data set to apply the same quality control to get the same set of SNPs in all breeds. No SNP imputation was performed.

After merging, quality control (QC) was carried out applying following restrictions:

- Only autosomal SNPs
- Only SNPs with known physical (base-pair) position and chromosome

- Minor allele frequency > 1.0 %
- Minimum SNP genotyping rate 90.0 % (SNPs with > 10.0 % missing genotypes are sorted out)

A set of 43,430 markers with an overall genotyping rate of 0.999 was left after quality control and used for all analyses. The reduction in marker numbers was not only due to QC criteria, but largely due to the different marker sets of the two chips.

4.2.   Software

Data management and preparation was carried out via bash and AWK commands or in R using RStudio version 1.4.1717 (RStudio PBC, 2021) and packages *dplyr*, *tidyr*, readr, *ggplot2*, *lme4* (Bates et al., 2015) and *qqman* (Turner, 2018). The graphics were created with R base functions and mentioned packages *ggplot2* and *qqman*. Tables were created in Microsoft Excel (Microsoft Corporation, 2021).

Genomic data management was done mostly in PLINK 1.9 (Chang et al., 2015; Purcell & Chang, 2021; Purcell et al., 2007), to some smaller extent with custom bash and awk commands.

The genome-wide association study was carried out using the software package GCTA (Yang et al., 2011) and its mixed linear model association tools MLMA and MLMA-LOCO (Yang et al., 2014).

Estimation of genetic parameters was carried out using ASReml 3.0 (Gilmour et al., 2009) and REMLF90/AIREMLF90 from the BLUPF90 family of programs (Misztal et al., 2018).

The conventional and single-step BLUP breeding value estimation were carried out using RENUMF90, PREGSF90 and BLUPF90 (Misztal et al., 2018). For result processing and creation of graphics, RStudio (RStudio PBC, 2021) and the same packages as mentioned above were used.

4.3.   Genome-wide association study

Our approach for a genome-wide association analysis accounted for the high number of repeated observations per individual. Unlike similar studies that use only one phenotypic measure or create one pseudo-phenotype per individual (Diniz et al., 2014), we tried to avoid this source of bias and use the observations directly in the model. The GWAS was carried out for all three breeds (LW, LR, PI). It was carried out for each breed separately, as well as for each trait (univariate). All five traits (total number of sperm, motility, ejaculate volume, sperm cell density, storage time) were used for completeness, although volume and density play a minor role as they were coerced to the total number of sperm.

4.3.1. Data transformation and GWAS methodology

Collected phenotypes were not normally distributed, therefore transformations were applied for total number of sperms and motility. Volume and density were not transformed, because of their lower relevance (as they have been coerced to total number of sperm), and storage time was not transformed as it was not possible to find a meaningful method to transform such (highly discrete) values.

The applied transformations were:

- For total number of sperm: $y = (x^{0,3} - 1) / 0,3$
- For motility: $y = log_{10}(101-x)$

Although transformation was applied, total number of sperm and motility were still not completely normally distributed.

The main principle we used in the GWAS was a mixed linear model association. For the correction of repeated measures, a repeatability model with a permanent environmental random effect were used. Due to the inability of GCTA to incorporate this second random effect, beside the additive genetic effect, a two-step procedure had to be used, where phenotypic values were pre-corrected for fixed effects and repeated measurements.

In the first step, phenotypic values were corrected using a univariate mixed linear model in R (using package *lme4*) with fixed effects, a permanent environmental effect and a residual random effect. Following model was applied:

$$y = Xb + Wp + e \qquad (1)$$

y is the vector of all trait observations, **b** is the vector of fixed effect estimates, **p** denotes the vector of random permanent environmental effects and **e** is the vector of residual effects. The fixed effects in **b** were:

- Breeder (cross-classified, 28/33/38 effect groups)
- Station * year * month (cross-classified, 277 effect groups)
- Age of the boar (continuous covariate, linear and quadratic regression coefficients)
- Ejaculation interval (continuous covariate, linear and quadratic regression coefficients)

The permanent environmental effect **p** was modelled as a normally distributed random intercept per individual: $p \sim N(0, I*\sigma^2_{pe})$, where **I** is an identity matrix of size m x m (m = number of animals) and σ^2_{pe} is the estimated permanent environmental variance. The residual random effect was assumed to be normally distributed and independent between observations: $e \sim N(0, I*\sigma^2_e)$, where **I** is an identity matrix of size n x n (n = number of observations = length of vector y). **X** and **W** are incidence matrixes linking observations to the effect groups. The estimation of parameters and effects was carried out using the standard REML algorithm of *lme4*.

The residuals **e** were then used as response variables (**y***) in the second mixed model for the genome-wide association, carried out in GCTA:

$$y^* = snp_i * b + Wa + e \qquad (2)$$

y* is the vector of "response variables", in this case residuals from the first model, **snp$_i$** is the vector of the SNP's allele content, coded as 0, 1 or 2, with b as its fixed effect estimate. **a** is the individual's additive genetic effect, modelled as random effect with $a \sim N(0, G_{ex}*\sigma^2_a)$. **G$_{ex}$** is the genomic relationship matrix calculated according to the first method from VanRaden (2008), but excluding the SNP's own chromosome. The reason for computing the G matrix without the chromosome where the marker sits on is a possible loss of power. When using the whole genome including the marker itself, the marker is fitted twice, once as fixed effect and once in the G matrix, leading to bias and decreased power of the following hypothesis test. This model was estimated once per SNP and fitted via GCTA's AI-REML algorithm. This approach has been called MLMA-LOCO ("leave-one-chromosome-out") inside the GCTA package (Yang et al., 2014). **W** is the incidence matrix as described above.

The p-value for the *SNPi* effect was calculated using the LRT (Likelihood-Ratio test). The p-values of all SNPs were displayed in Manhattan plots (see Results section). Note that the displayed arbitrary threshold of $p < 10^{-5}$ would correspond to a Bonferroni-corrected p-value of 0,05 for 1,000 independent markers and has to be interpreted only as indicative. Another similar GWAS in LW pigs estimated an equivalent of 6,993 independent markers (by LD-based pruning) for the pig genome (Wang et al., 2018), so this threshold was set comparatively high.

4.3.2. Gene identification

The regions around SNPs (±1 MB) with the lowest p-values were looked up in the NCBI online database (NCBI - National Center for Biotechnology Information, 2021) to find known genes related to spermatogenesis or male fertility. The genome build used was *Sscrofa 11.1* of the Swine Genome Sequencing Consortium (Warr et al., 2020). The corresponding regions on the human genome were also looked up for possible associations.

4.4. Genetic parameter estimation and single-step genomic breeding value estimation (ssGBLUP)

The second set of methods used was a multivariate single-step GBLUP to compute breeding values for the traits. Note that storage time was not included here as its additive genetic variance component was close to zero or could not even be estimated. The estimation was only carried out for PI, because the other two populations were too small.

Untransformed phenotypic values were used for easier interpretation, being aware that this introduces a deviation of the phenotypes (and likely also the residuals) from normal distribution.

4.4.1. Animal model

The observed phenotypes were described by a multivariate repeatability animal model, which can be written as

$$y = Xb + Wp + Wu + e \qquad (3)$$

y is the vector of all trait observations, **b** is the vector of fixed effect estimates, **p** is the vector of random permanent environmental effects and **u** is the vector of additive genetic effects. **X** and **W** are incidence matrices linking observations to fixed or random effects, respectively. **p** and **u** both have length m (number of animals). The description of the fixed effects is identical to them in the GWAS from equation (1). The random permanent environmental effect **p** was modelled as a normally distributed random intercept per individual: $p \sim MVN(0, I*\sigma^2_{pe})$, where **I** is an identity matrix of size m x m (m = number of animals) and σ^2_{pe} is the estimated permanent environmental variance.

The additive genetic effect **u** was assumed to have following distribution: $u \sim MVN(0, H*\sigma^2_u)$. σ^2_u is the estimated additive genetic variance. **H** is the combined relationship matrix; its inverse is easily computed from the genomic relationship matrix **G** and the numerator relationship matrix **A**:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22}^{-1} \end{bmatrix} \qquad (4)$$

The G matrix (**G**$_b$) was directly set up from all SNPs of the genotyped individuals according to the first proposal of VanRaden (2008). As it is often singular, a weighted version with

$$G = 0{,}95G_b + 0{,}05A_{22} \qquad (5)$$

was used, causing negligible bias but making the **G** matrix invertible.

A^{-1} was set up from pedigree records using the well-known rule from Henderson (1976). The term **A**$^{-1}_{22}$ reflects the inverted numerator relationship matrix for only the genotyped animals.

4.4.2. Genetic parameter estimation

The genetic variance-covariance matrices of the additive genetic, random permanent environmental and random residual effects were estimated using the EM-REML algorithm of REMLF90. Convergence criterium was set to 10^{-10}. The parameter estimation was carried out on single-step genomic basis, not in a conventional pedigree-based model.

For speeding up the convergence a pre-estimate from a bivariate pedigree-based estimation was used, carried out in ASReml 3.0 (Gilmour et al., 2009). These values were used as starting values for REMLF90, making the final estimation possible in a finite amount of time (around 40 h and 145 iterations).

The displayed heritabilities, genetic and phenotypic correlations of the four traits were calculated from the multivariate parameter estimates. For estimating the standard errors of prediction (S.E.P.) of these parameters, the AI-REML algorithm of AIREMLF90 was used, because EM-REML does not provide a method for their computation. The S.E.P. were computed using a Monte-Carlo sampling method (Meyer & Houle, 2013), included in AIREMLF90 with *OPTION se_covar_function*. This estimates the standard error of variance component functions by averaging the standard deviation of many simulated samples, in this study 10,000 samples per parameter (default in the program).

4.4.3. Single-step genomic BLUP

Using the model (3) above and the estimated variance-covariance structures, the effect sizes for fixed (BLUE) and random (BLUP) effects were estimated in BLUPF90. The options used included:

- the calculation of standard errors of prediction (S.E.P.) (*OPTION sol se*)
- the solving method FSPAK (automatically with the option given above).

The BLUP estimates in natural units were standardised to relative breeding values with mean 100 and a genetic standard deviation of 20, e.g. a boar with a relative breeding value of 140 has a BLUP estimate two genetic standard deviations above the mean. This method of displaying breeding values is used in routine genetic evaluation of the Austrian pig breeding program and is well-known among breeders. The values were displayed directly and not corrected to any "base" level like it is done in routine estimation.

The reliability (r^2) of breeding values (r^2 between estimated breeding values and true breeding values) for individuals was calculated from the S.E.P. as follows (Misztal et al., 2018):

$$r^2 = 1 - \frac{SEP^2}{\sigma_A^2} \qquad (6)$$

4.4.4. Validation of ssGBLUP breeding values

First the correlation coefficient of ssGBLUP and conventional BLUP breeding value estimates was computed to get a quick validation. This normally should be in the range of 0.90 – 1.00, being closer to 1.00 for traits with high heritability.

To quantify the advantage of ssGBLUP in terms of reliability (r^2) compared to pedigree-based BLUP, a validation was carried out. For that, all boars born in years 2019 or 2020 got their records removed. Therefore, their breeding values were estimated like young individuals without records. These were 177 boars in total, 144 born in 2019 and 33 in 2020, respectively. Their BLUPs and r^2 for all four traits were estimated using either ssGBLUP or conventional pedigree-based BLUP.

The average relative breeding values of all individuals, sorted by year of birth were used to display the genetic trend of the traits. The purpose of this method was to monitor medium- and long-term evolution of the genetic values of the population.

5. Results and discussion

5.1.   Environmental effects from the GWAS model

For setting up the model we tested fixed effects on the sperm parameters in all three breeds (LW, LR, PI) in a univariate model. The fixed effects of breeders and station/year/month plus linear and quadratic regressions for age of the boar and ejaculation interval were highly significant (P < 0.001) and were thus included in the model. The following results originate only from the Pietrain data as the data sets were rather small for the other two breeds.

5.1.1.   Breeder effects

We found quite extreme effects of some breeders on total number of sperm and sperm motility, as shown in Table 4.

Table 4: Ten highest and lowest breeders' effects for total number of sperm (left, in billions) and motility (right, in %)

Breeder (anonymised)	Effect [billions]	Breeder (anonymised)	Effect [%]
ZNR100387	35.37	ZNR114	2.8
ZNR9399	33.25	ZNR897	1.3
ZNR812	16.53	ZNR158	1.1
ZNR114	12.94	ZNR100043	0.7
ZNR722	6.88	ZNR146	0.7
ZNR897	2.69	ZNR251	0.6
ZNR9815	-1.86	ZNR370	0.5
ZNR9961	-3.94	ZNR722	0.4
ZNR148	-4.78	ZNR89	0.3
ZNR106	-5.00	ZNR17	0.1
ZNR35	-15.24	ZNR194	-0.9
ZNR87	-16.85	ZNR100387	-0.9
ZNR194	-17.81	ZNR627	-0.9
ZNR750	-17.90	ZNR750	-1.1
ZNR845	-20.54	ZNR87	-1.1
ZNR627	-21.70	ZNR812	-1.2
ZNR370	-22.79	ZNR48	-1.5
ZNR155	-27.32	ZNR155	-1.7
ZNR48	-28.27	ZNR9399	-2.0
ZNR100043	-35.37	ZNR9815	-2.8

Surprisingly, high breeders' effects were found, ranging from +35.37 to -35.37 billion sperm cells per ml and from +2.8 to -2.8 % motility.

Antohi et al., (2011) could show a negative effect of antibiotic treatments on sperm motility in rats. Other research has found a negative influence not directly on spermatogenesis, but on the function of the epididymis that could possibly harm the sperm cells (Rosenfeld et al., 2018). The concern has been brought up that prophylactic use of relatively high antibiotic treatments by breeders' veterinarians before boars are quarantined by PIG Austria GmbH (Pfeiffer, 2021). It is unclear whether this effect is still noticeable, considering the long quarantine period of around two month before boars start to produce semen for sale. Another explanation for the high breeder effects may be feed-caused, as e.g. an influence of trace elements like zinc on sperm quality is well-known (Fallah et al., 2018). Again it has to be assumed, that the long feeding period during quarantine, where all boars are fed equally, should compensate for any nutrient deficiencies a young boar might have. To verify this assumption further research is required.

5.1.2.   Station and seasonal (year/month) effects

The station/year/month effect was introduced not only to take account of seasonal and geographical differences, but also for management-related, staff-related and technical issues. As any changes from one of these sources (an upgrade to the CASA system at one station, some new workers at a laboratory, etc) have an influence on traits of interest, the station/year/month effect would show these at least in monthly resolution. The following graphs (Figure 4) show the expected marginal means (means corrected for all effects except the station/year/month effect) for total sperm number in the three different AI stations.
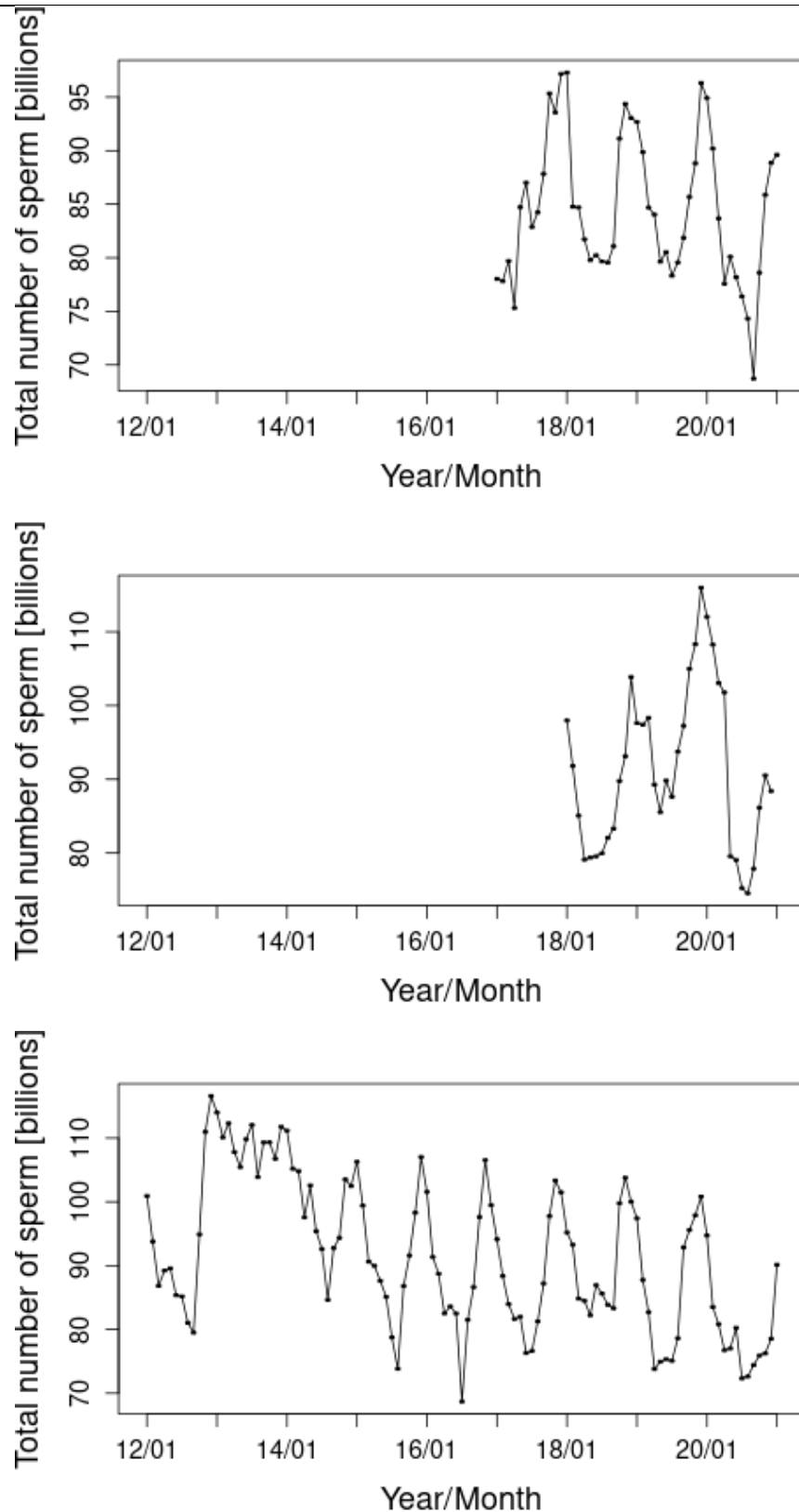
Figure 4: Expected marginal means over time for total number of sperm at AI stations Gleisdorf, Hohenwarth and Steinhaus (top to bottom)

As shown in Figure 4, the total number of sperm tends to decrease during summer months in almost all years and all stations. This finding is consistent with literature.

Especially high temperature and humidity had a negative effect on sperm numbers and sperm motility (Johnson et al., 1997; Kunavongkrit et al., 2005). Although the AI stations use methods to cool down the air during summer and increase air flow rates, one cannot completely neutralise the seasonal effects.

Here technical issues can also be traced. The introduction of the CASA *eFlow* system in mid-2020 took some time for correct adjustment which influenced measuring results. The large decreases in total numbers of sperm at Gleisdorf and Hohenwarth in mid-2020 may be the result of the CASA upgrade, as the station manager reported problems in adjusting the new equipment (Doppelhofer, 2021).

Motility did not show a strong response on seasonal effects, as shown in Figure 5:
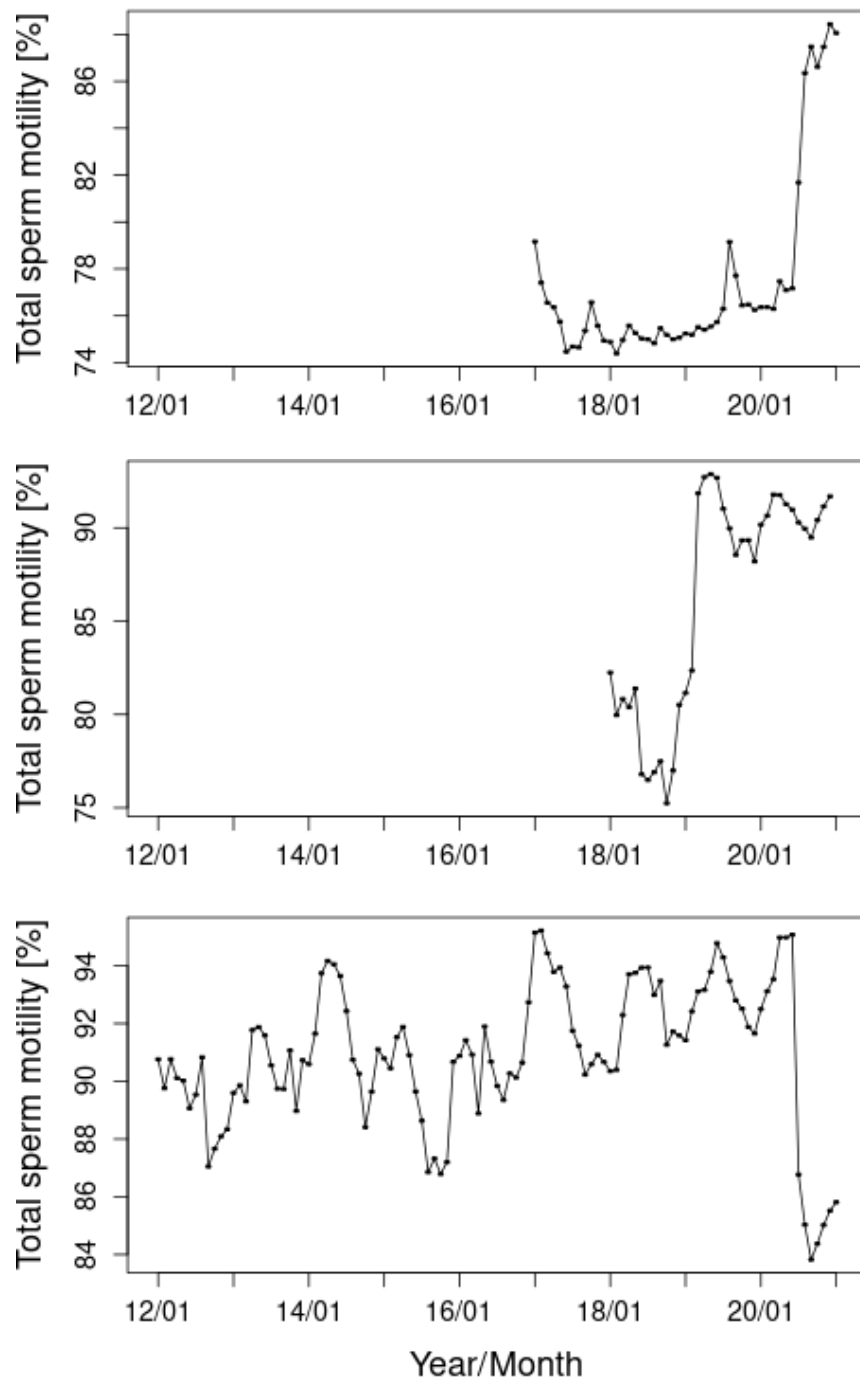
Figure 5: Expected marginal means over time for motility at AI stations Gleisdorf, Hohenwarth and Steinhaus (top to bottom)

The upgrade to the eFlow system in 2020 also showed its effects in motility. The numbers at Gleisdorf increased from ~76 % to ~ 86 % while the motility at Steinhaus decreased from ~92 % to ~86 %. The motilities after the upgrade were much closer to each other which might also be caused by using the same system in all stations.

5.1.3.   Age and ejaculation interval

The effects of the age of the boar and the ejaculation interval were modelled as quadratic polynomials. Both the age of the boar (here shown in the interval of 200 to 2,000 days) and the time since last ejaculation (ejaculation interval, shown from 3 to 30 days) had a large effect on the total number of sperm, but an almost negligible effect on motility (Figure 6).
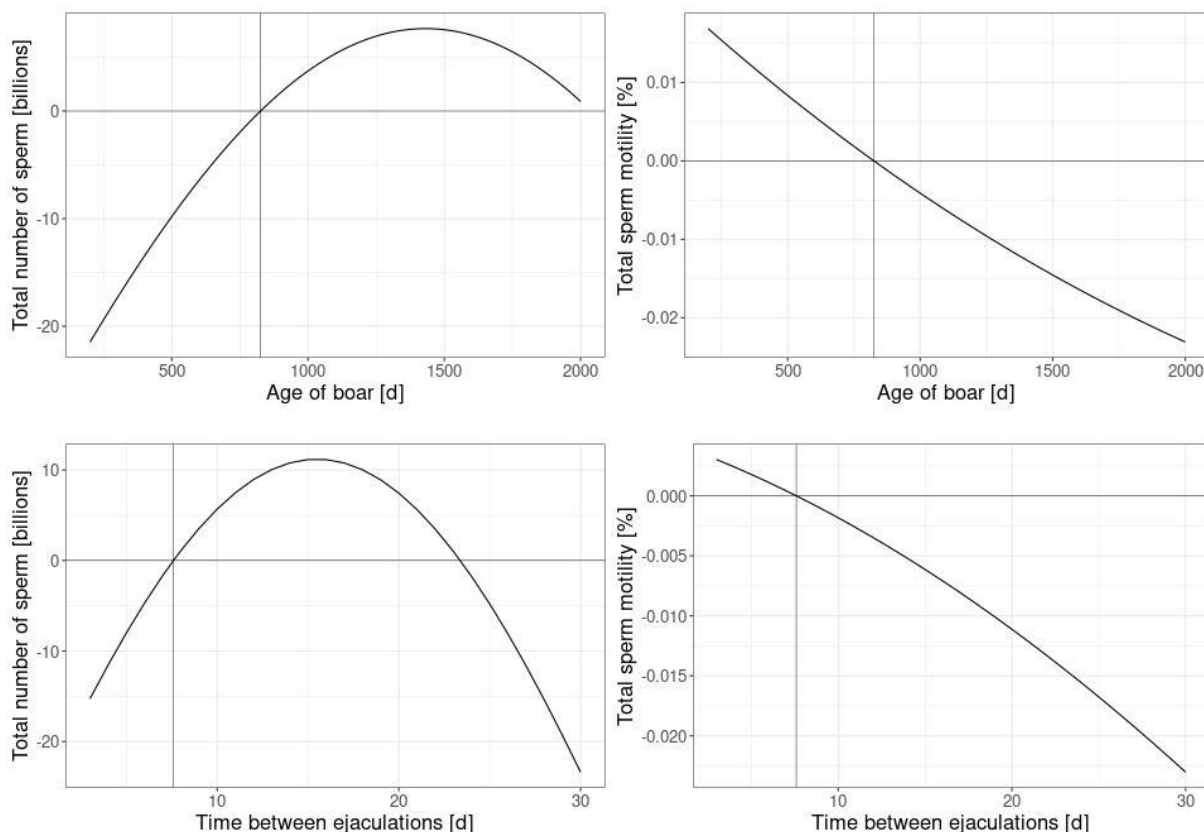

Figure 6: Effect of the boar's age and ejaculation interval (in days) on total number of sperm and motility in Pietrain boars

An effect of the age of the boar on total number of sperm was observed, but this cannot be changed easily in the daily routine of the AI stations. The elimination of boars on the stations is much more dependent on sperm abnormalities, motility, breeding values and economic demand so its productivity does not play a major role (Pfeiffer, 2020).

The effect of ejaculation interval on total number of sperm is of practical relevance. The mean ejaculation interval in this study was around 7.6 days (marked in the lower left plot with a vertical, black line). An increase to 10 days would increase the sperm number (mean: 84.1 bn) by roughly 5 billion (5.9 %) but lowers the number of ejaculations in a given amount of time by 24 %. Overall, this would cause a lower total productivity. In the other direction, a reduction to 5 days would decrease the sperm number by 8 bn (9.5 %), but increase the number of ejaculates by 52 %, causing a higher total productivity. If the ejaculation interval does not cause any other problems,

e.g., libido problems, reduced motility, or increasing abnormalities, a shorter interval than the average of 7.6 days may be applied.

The upper half of the curve, which implies falling sperm numbers above 15.5 days interval, has to be interpreted with care. The number of observations with such large intervals was low in the data set, so these values are not as reliable as for the lower half.

Effects of boars' age and time between ejaculation sampling on motility were small and may be neglected in practice, since they were only in a range of 0.00 - 0.02 %.

5.2. Genome-wide association study

The results of the genome-wide associations were displayed in Manhattan plots, where single dots represent one SNP, the x-axis shows its position on a chromosome and the y-axis the p-value of the hypothesis test for significant influence on the trait of interest. As in almost all other GWAS studies, the scale of -log(p-value) was chosen to present significant, e.g. small p-values, high up on the y-axis. An indicative threshold of 10^{-5} was set. The GWAS has been carried out separately for all three breeds (LW, LR and PI) because the genetic background and population-wise allele frequencies of markers may affect the results.

In the current study repeated measurements for all traits of interest were used. Therefore, a repeatability model using a random permanent environmental effect was applied. Due to computational limitations, this permanent environmental random effect had to be used in the R-based pre-correction model, while the additive genetic effect is included in the GWAS model. The estimation of two separate mixed models may not be optimal as the permanent environmental and additive genetic effect might be correlated, being both distributed with the same incidence matrix (**W**), only differing in their variance-covariance structure: $\mathbf{I}*\sigma^2_{pe}$ vs. $\mathbf{G_{ex}}*\sigma^2_a$. The extent to which this has an influence on the results is hard to evaluate.

5.2.1. Total number of sperm

The following sections represent the results of the genome-wide association study for each trait, starting with total number of sperm, carried out separately for LR, LW and PI in univariate models.
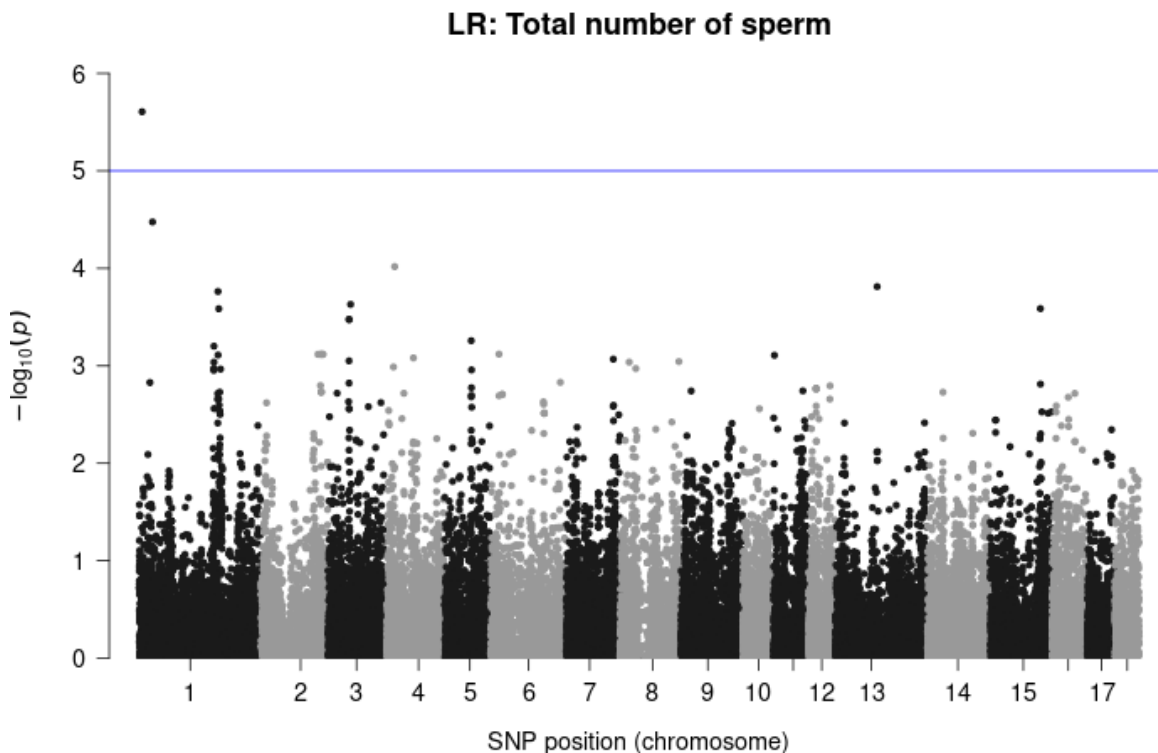
Figure 7: Manhattan plot for total number of sperm in Landrace pigs

In LR, a single associated SNP on chromosome 1 (ALGA0000593, position 6,532,869 bp) linked to either the PRKN gene (5,698,508 – 6,731,132 bp), coding for parkin RBR E3 ubiquitin protein ligase, or AGPAT4 (6,761,734 – 6,883,555 bp), coding for acylglycerol-3-phosphate O-acyltransferase 4 (NCBI - National Center for Biotechnology Information, 2021).

PRKN is involved in the ubiquitin-proteasome protein recycling system for misfolded or damaged proteins (Imai et al., 2001). In humans, mutations of this gene cause an early form of Parkinson's disease (Bjerre et al., 2006). However, little research has been done on the porcine homolog. Currently, no causal connection to male fertility is known. c

AGPAT4 is in involved in the membrane fatty acid metabolism and is broadly expressed in meiotic and somatic testis cells. Defects of its isoenzymes (other AGPAT variants) caused severe male and female fertility problems in mice. Male individuals with knocked-out AGPAT1 showed an interruption of spermatogenesis at spermatid stage, rendering these individuals infertile (Agarwal et al., 2017). This seems to be a promising candidate gene, as an influence of another acylglycerolphosphate acyltransferase on spermatogenesis is likely.

It remains unclear why this signal could not be observed in any of the other breeds. The fact that there is a single marker without any surrounding markers at similar significance levels, which would be expected in theory due to LD at this marker density, may also imply an artefact.
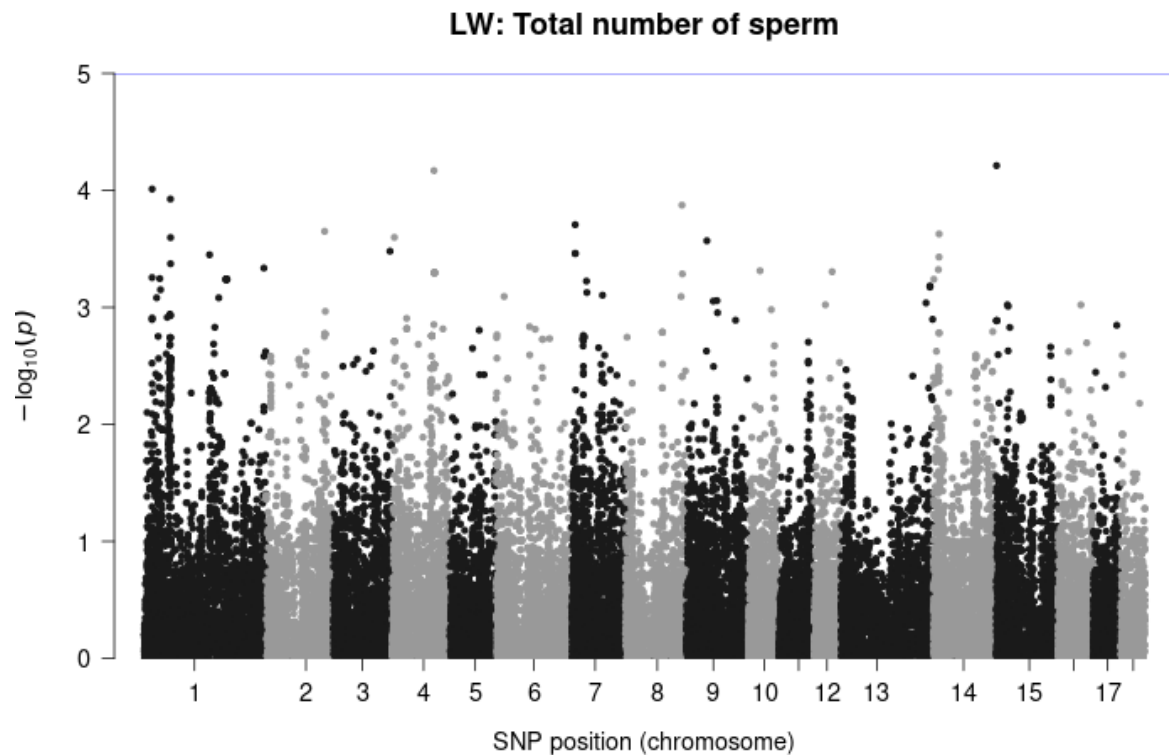
Figure 8: Manhattan plot for total number of sperm in Large White pigs

As shown in Figure 8, LW had no promising associations for total number of sperm, as most SNPs are below or around a p-value of 10^{-4}.
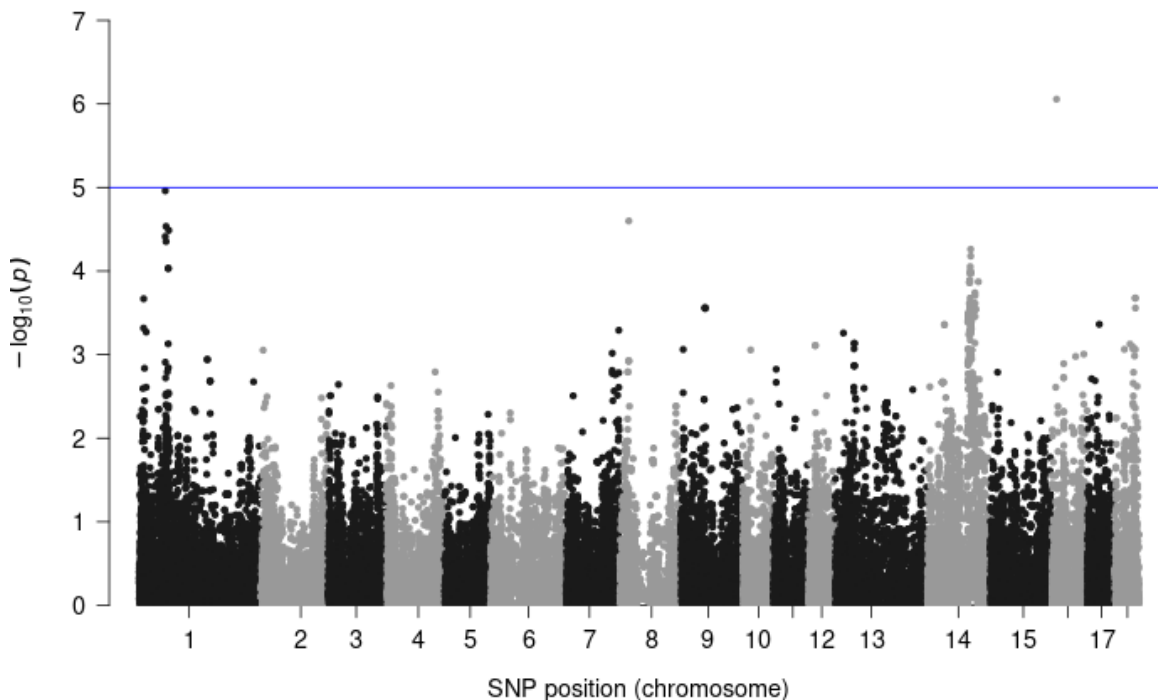
Figure 9: Manhattan plot for total number of sperm in Pietrain pigs

Figure 9 for total number of sperm in PI showed four possible associations at chromosomes 1, 8, 14 and 16:

- Chromosome 1: around SNP ALGA0003762 (58,760,643 bp)
- Chromosome 8: around SNP MARC0077431 (21,004,503 bp)
- Chromosome 14: around SNP CASI0007598 (97,292,020 bp)
- Chromosome 16: only a single SNP DRGA0015821 (10,165,906 bp)

The peak on chromosome 1 seems to be more concise than it actually is, as the SNPs with high -$\log_{10}$(p-values) were scattered along a rather large distance, around 58.7 Mbp, 60.7 MBp and 65 – 66 Mbp. This was not further verified, since none of the locations were clearly connected to any known fertility-related gene.

Also, in the region around 21 Mbp on chromosome 8 there were no well investigated genes. Within ± 0.5 Mbp there were only a few coding sequences (STIM2, LOC11026069, LOC11026070) where no connections to male fertility are known so far.

The peak on chromosome 14 consisted of a large number of SNPs which implied some recent signature through linkage between these markers. From the top 50 SNPs, 35 were found in this region, spanning from 94.5 – 97.8 Mbp. The region featured 4 genes: PCDH15, MBL2, DKK1, PRKG1 (NCBI - National Center for Biotechnology Information, 2021). PCDH15 codes for protocadherin 15, a large protein involved in calcium-dependent cell-cell adhesion. It is well known to be essential for hearing, as mutations of it cause Usher syndrome type 1F, a hereditary disease with deafness and blindness in humans (Choudhary et al., 2020). MBL2 and DKK1 are involved in the

immune system and in embryonal development, respectively (NCBI - National Center for Biotechnology Information, 2021). PRKG1 codes for type 1 cGMP-dependent protein kinase, involved in the cGMP signalling pathway in humans (NCBI - National Center for Biotechnology Information, 2021). No known connection to male fertility in pigs was found.

The marker at chromosome 16 is located within the transcribed region of CDH12 gene (9,985,759 – 10,975,489), coding for cadherin 12 (NCBI - National Center for Biotechnology Information, 2021). This protein is, like other cadherins, involved in calcium-dependent cell-cell adhesion (Alvarez-Rodriguez et al., 2020). A direct connection of this gene to spermatogenesis has not been reported to our knowledge.

There may be an influence of cadherins on spermatogenesis, as the spermatocytes/spermatids have to maintain an adhesive, flexible binding system to their Sertoli cells that has not been investigated well (Hale, 1996). Although, both protocadherin 15 and cadherin 12 belong to the family of N-cadherins (*neuronal*) that are known for cell-cell adhesion processes in the nerval system, there are reports showing that N-cadherins also mediate the adhesion process of spermatogenic cells to Sertoli cells (Newton et al., 1993; Wu et al., 2020).

5.2.2.   Motility

The plots for motility did not show clear signals in any of the three breeds. Most SNPs were far below the threshold, as shown in Figure 10 to Figure 12. The interesting peak in LR on chromosome 3 and LW on chromosome 7 were consisting of only a single or small number of SNPs, which might imply an artefact. The fact that no signals were found in Pietrain, being the largest population in this study, did not allow any reliable interpretation of other signals on this trait.
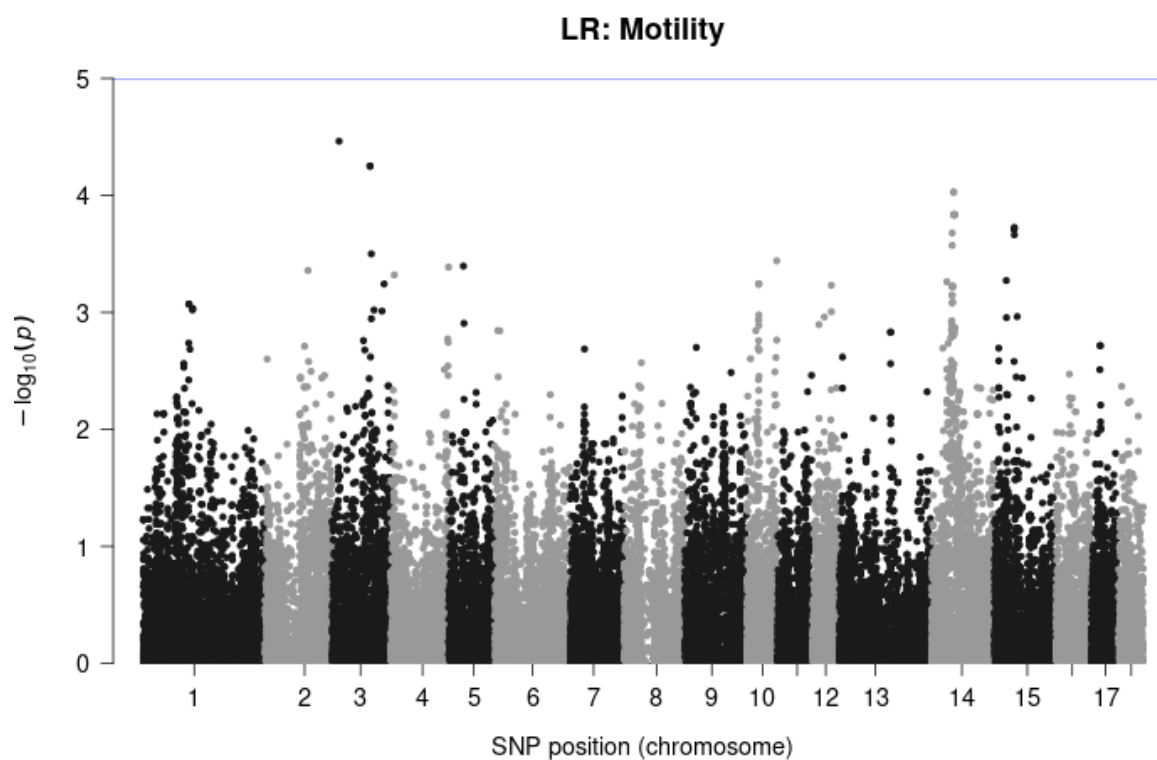
Figure 10: Manhattan plot for motility in Landrace pigs
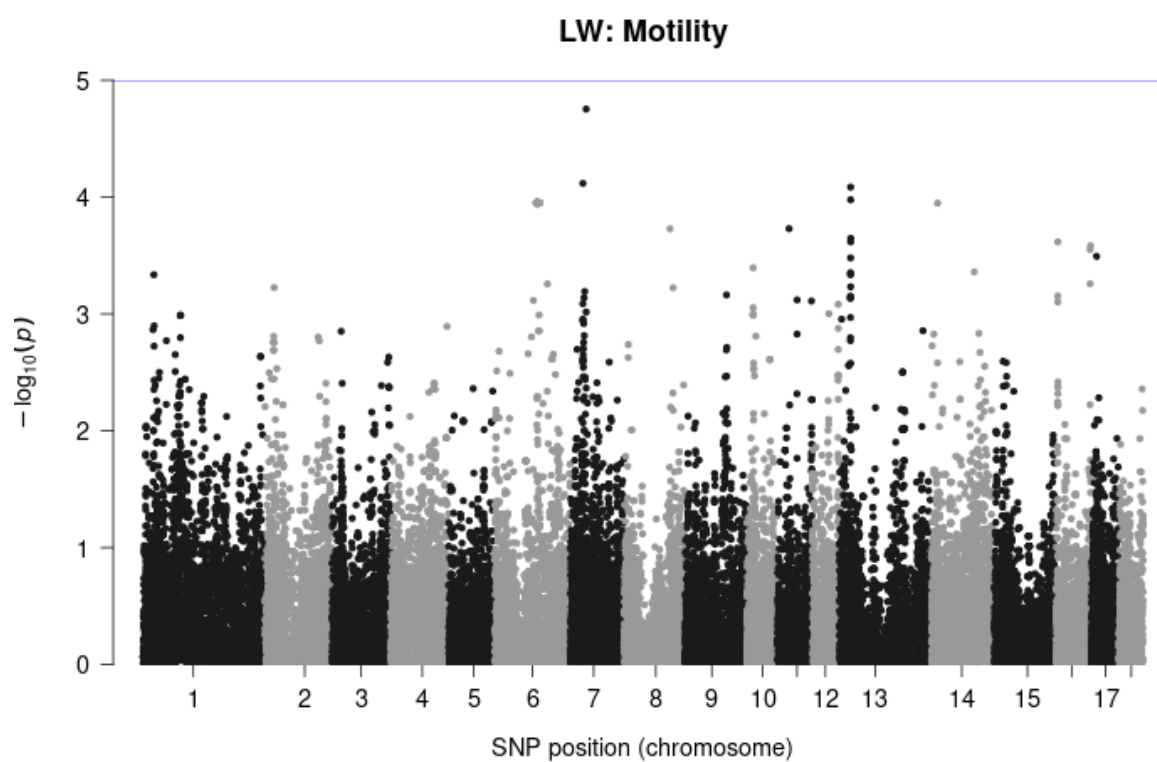
Figure 11: Manhattan plot for motility in Large White pigs
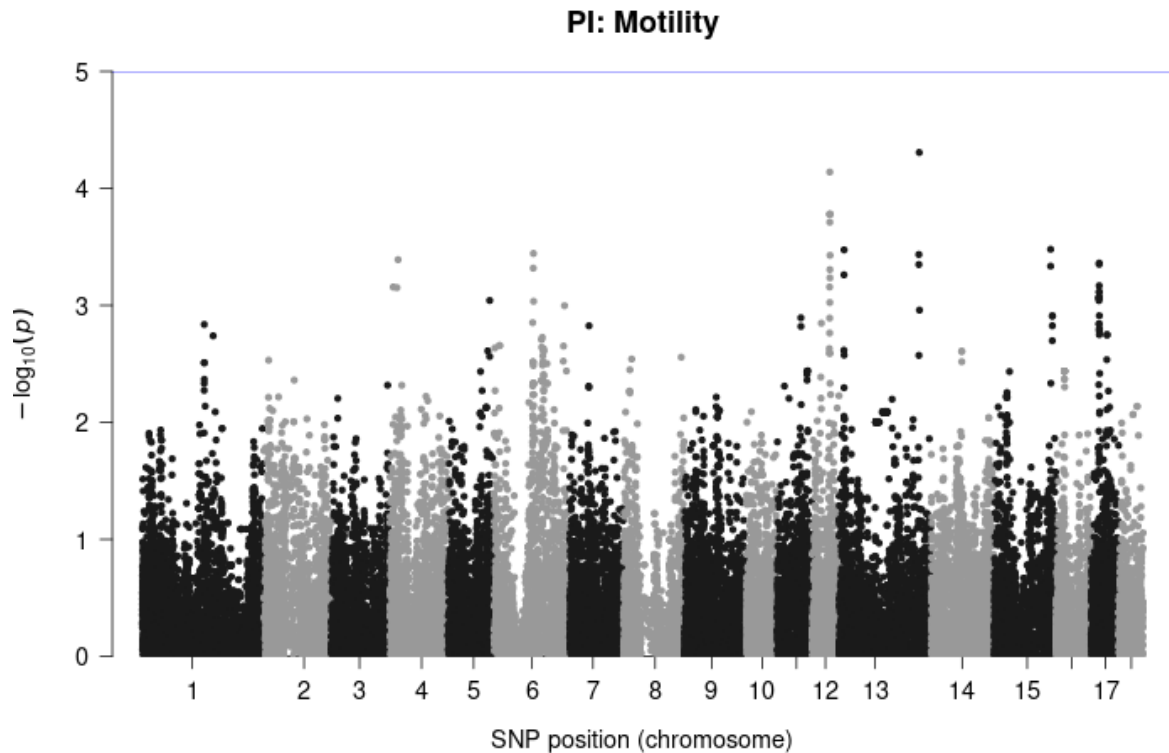
Figure 12: Manhattan plot for motility in Pietrain pigs

5.2.3. Ejaculate volume and density

The results for volume and density did not show any promising markers (Figure 13 to Figure 18). Most SNPs were far from significance. The single SNPs near or above the threshold could not be traced to any ejaculate-relevant gene.
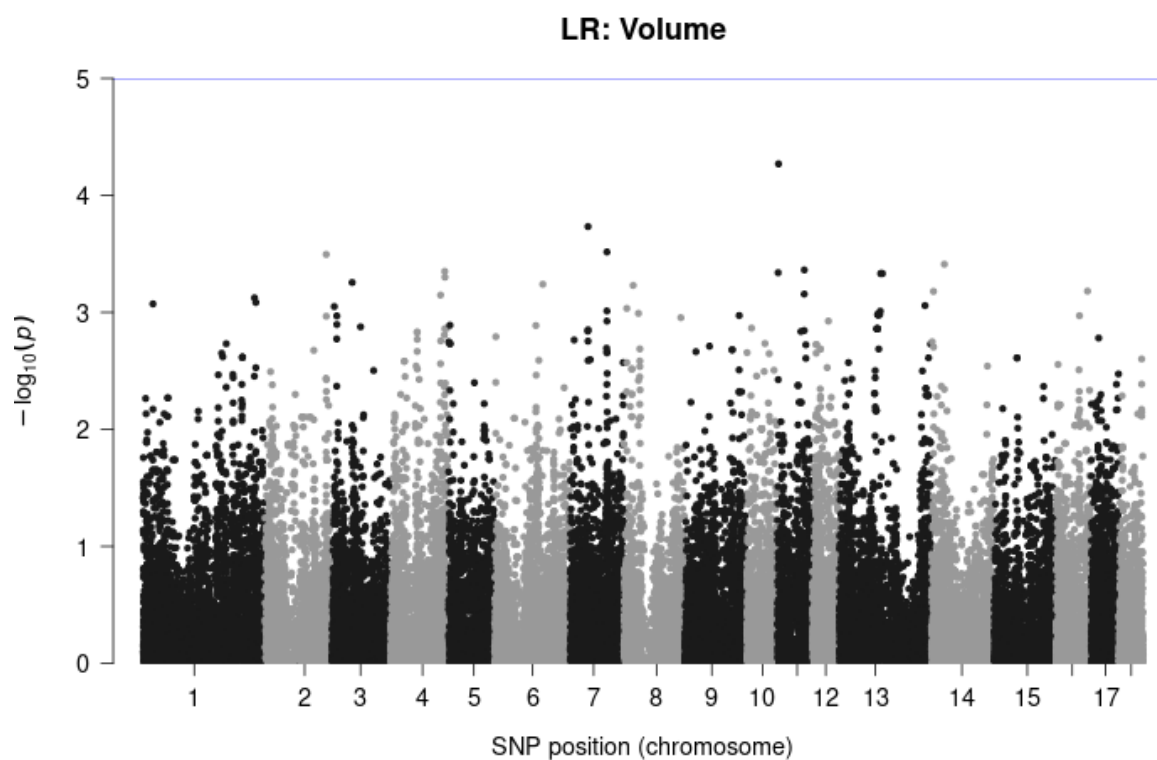
LR: Volume

Figure 13: Manhattan plot for ejaculate volume in Landrace pigs
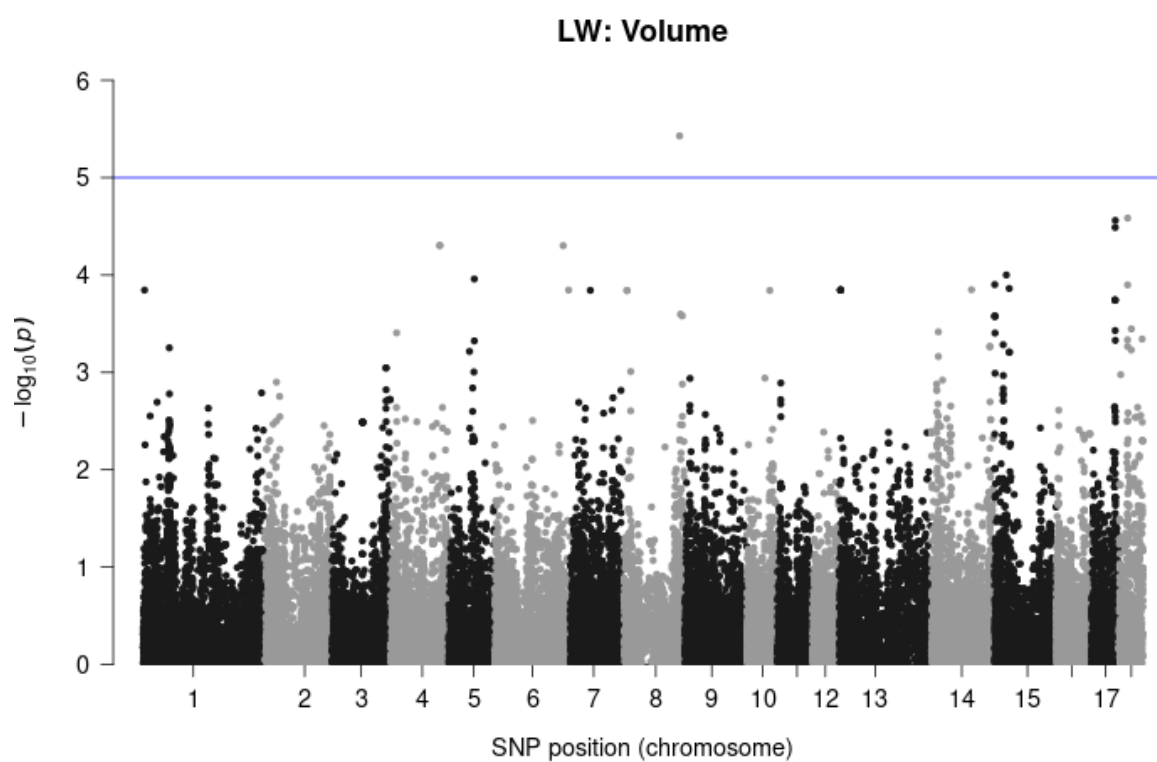
LW: Volume

Figure 14: Manhattan plot for ejaculate volume in Large White pigs
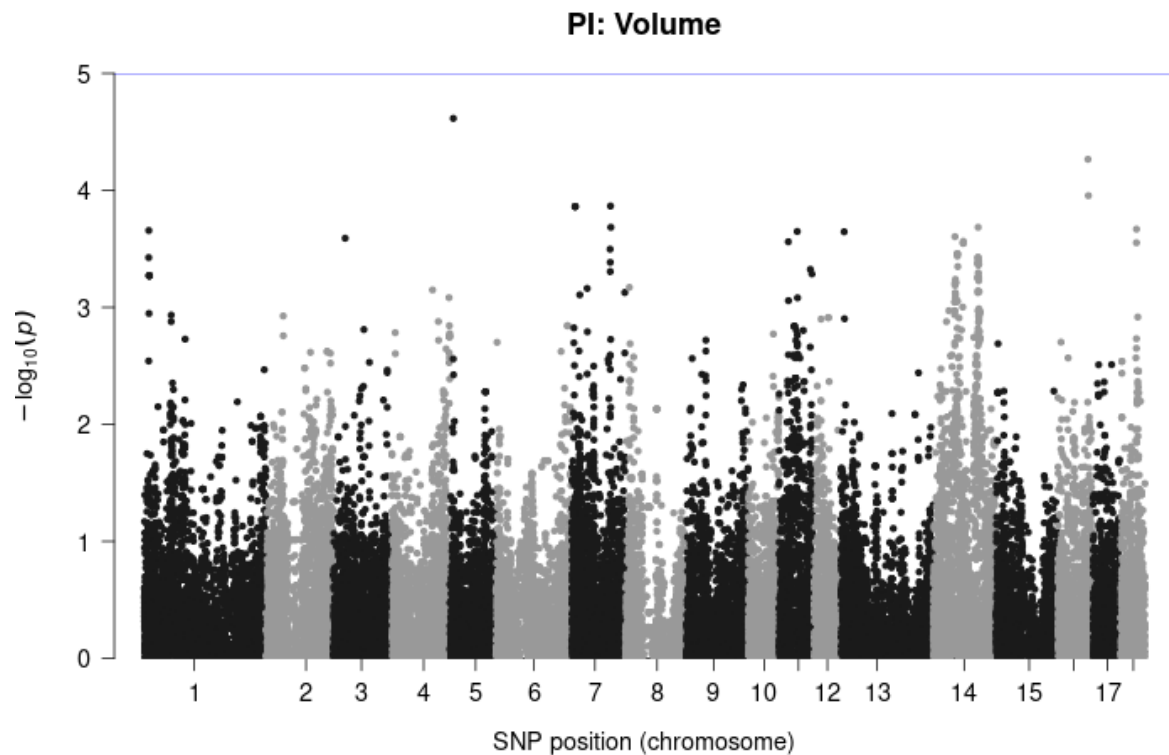
PI: Volume

Figure 15: Manhattan plot for ejaculate volume in Pietrain pigs
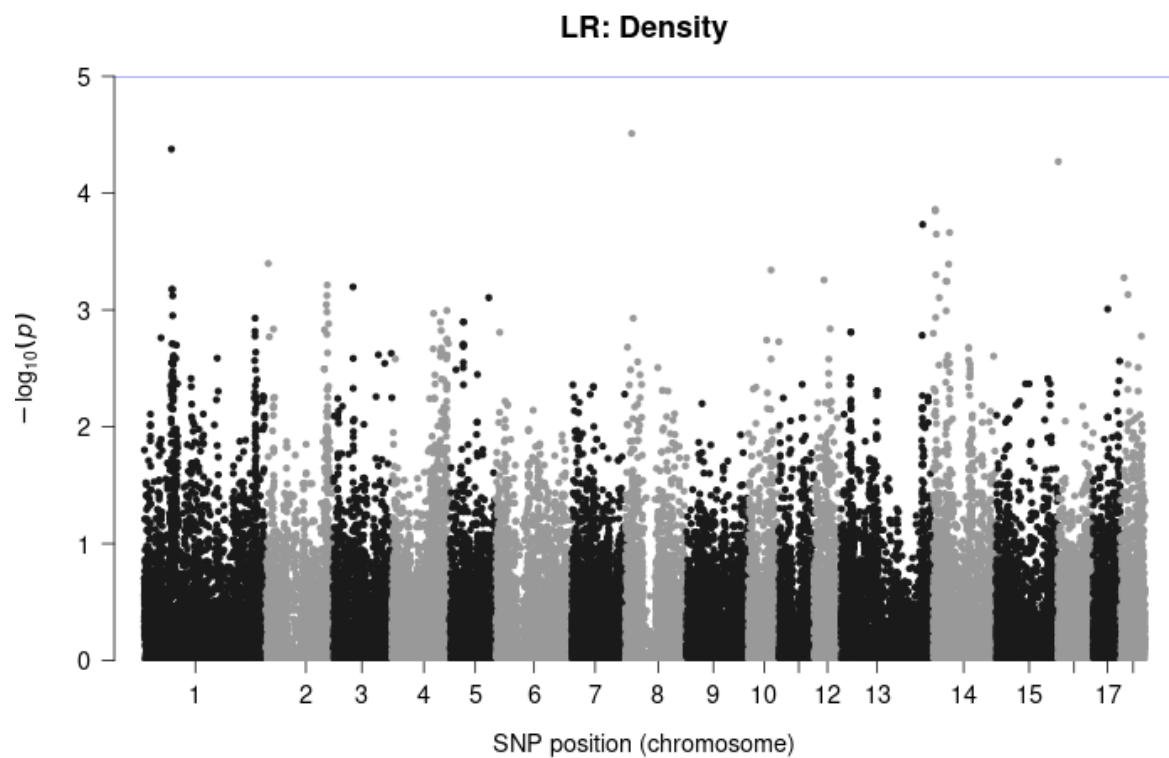
LR: Density

Figure 16: Manhattan plot for sperm cell density in Landrace pigs

LW: Density

Figure 17: Manhattan plot for sperm cell density in Large White pigs

PI: Density

Figure 18: Manhattan plot for sperm cell density in Pietrain pigs

It has to be mentioned that any associations of SNPs on our traits of interest have to be interpretated with care. Ejaculate volume is not dependent on sperm production, but mostly on prostate and epididymis secrete (Hale, 1996). Hence, this trait resembles the fluid productivity of these organs, not the reproductive productivity (sperm cell numbers) or the fertilizing ability of the ejaculate. Density i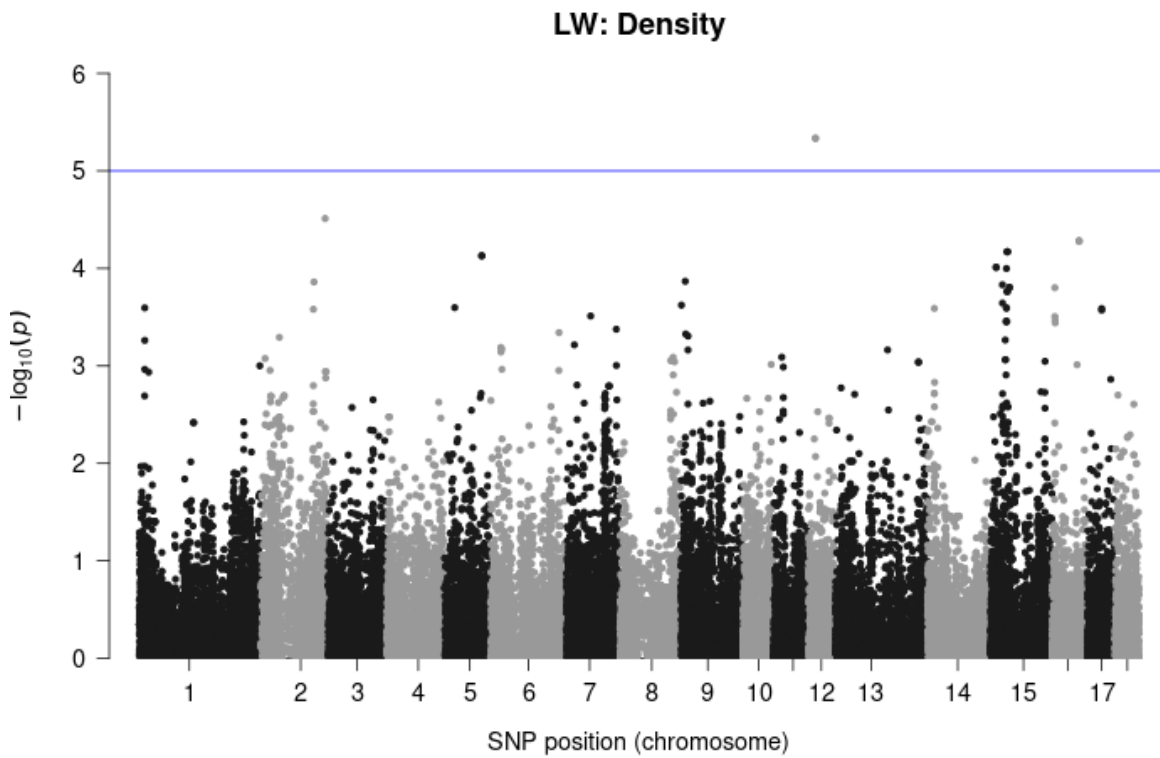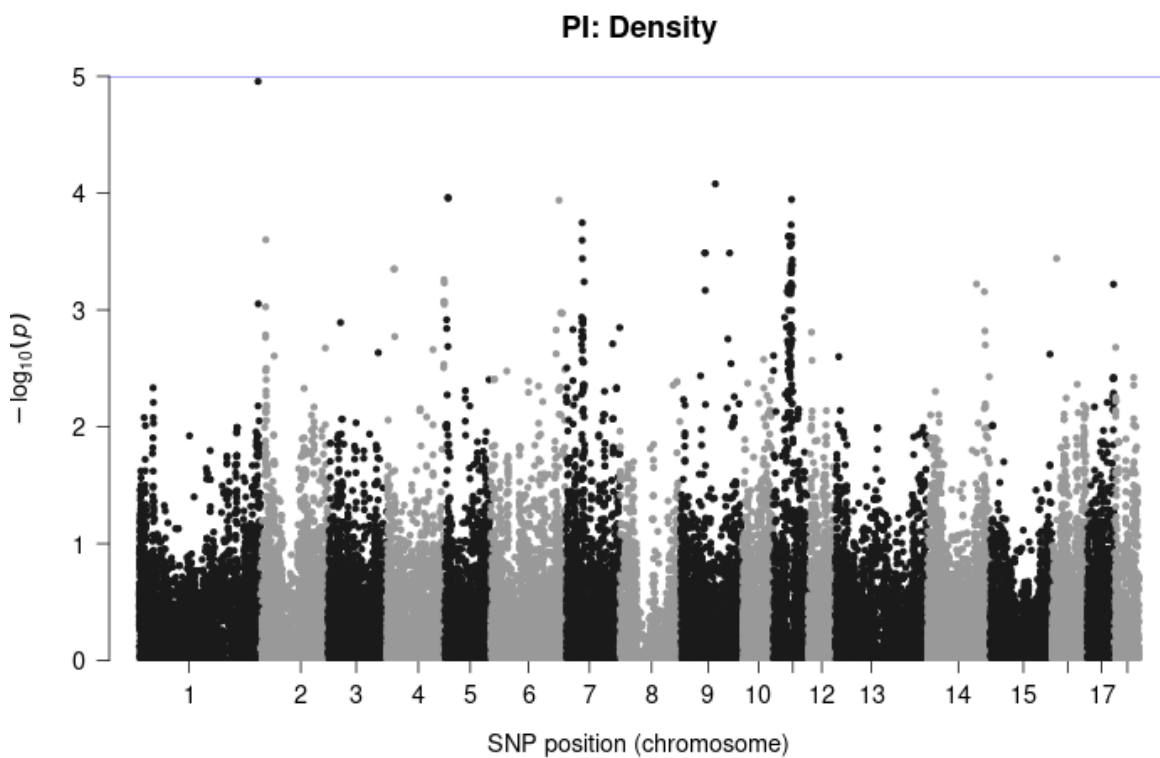s, from a biological point of view, only the ratio of the number of sperm and the fluid volume delivered by the prostate, epididymis and other organs.

5.3.   Genetic parameters

After running a 5-variate estimation, the genetic variance (and so the heritability) of storage time proved to be close to 0. Reasons for almost zero heritability of storage time are noted in the Methods section. Therefore, we reduced the model to a 4-variate model. Estimated genetic parameters are shown in Table 5. Heritabilities ranged from 0.340 to 0.000 for density and storage time, respectively. Genetic correlations ranged between -0.759 between volume and density and 0.547 between total number of sperms and density. Phenotypic correlations ranged between -0.557 between volume and density and 0.486 between total number of sperm and density. All genetic parameters were significant (to $p = 0.05$) from zero, except for the genetic correlation (-0.089) between total number of sperm and motility and the genetic correlation (0.121) between total number of sperm and volume.

Table 5: Heritabilities, phenotypic and genetic correlations of sperm traits (diagonal: heritabilities, upper triangular matrix: phenotypic correlations, lower triangular matrix: genetic correlations), with their estimated standard errors of prediction

	Total number of sperm	Motility	Volume	Density
Total number of sperm	**0.247** ±0.027	0.153 ±0.013	0.304 ±0.014	0.486 ±0.012
Motility	-0.089 ±0.112	**0.131** ±0.022	-0.124 ±0.013	0.237 ±0.013
Volume	0.121 ±0.080	-0.305 ±0.094	**0.304** ±0.027	-0.557 ±0.010
Density	0.547 ±0.059	0.206 ±0.096	-0.759 ±0.038	**0.340** ±0.028

The observed heritabilities were in accordance with other studies (Marques et al., 2017; Wolf, 2010), although ours were higher for total number of sperm (24,7 % vs. 10 or 13 %). We found a small and not significant negative genetic correlation (-0.089±0.112) between total number of sperm and motility, while the phenotypic correlation was positive and significant (0.153±0.013). The positive correlations between total number of sperm and volume/density were not surprising as these two values were multiplied to get total number of sperm, so they were highly dependent.

It is remarkable that motility has a negative genetic correlation (-0.305±0.094) to volume, but a positive genetic correlation (0.206±0.096) to density, both being significant. As motility is also linked to fertilizing abilities (Broekhuijse et al., 2012) and

other sperm quality parameters (Marques et al., 2017), selection on boars with higher sperm cell density (and avoiding higher volume) may provide an opportunity for improving quality parameters indirectly. A possible reason may be provided by the fact that prostate inflammations in humans reduce cell density, motility and fertility while increasing semen volume (Rusz et al., 2012). Therefore, the observed genetic correlation could be caused by a genetic predisposition to prostate inflammations.

5.4. Genomic breeding value estimation and genetic trends

For comparison and validation purposes the ssGBLUP solutions were compared with conventional pedigree-based BLUP solutions. We calculated correlation coefficients between ssGBLUP estimates and conventional BLUP estimates, shown in Table 6.

Table 6: Pearson's correlation coefficients (r) between single-step genomic BLUP and conventional BLUP breeding value estimates, for all four traits

Trait	Correlation coefficient (r)
Total number of sperm	0.981
Motility	0.945
Volume	0.982
Density	0.982

The lower correlation (0.945) of motility compared to the other traits reflects the lower heritability of this trait. The range of the correlation coefficients is remarkably high, implying a correct application of the method and consistent pedigree recordings. For illustration, Figure 19 shows the estimated breeding values using either method in total number of sperm.
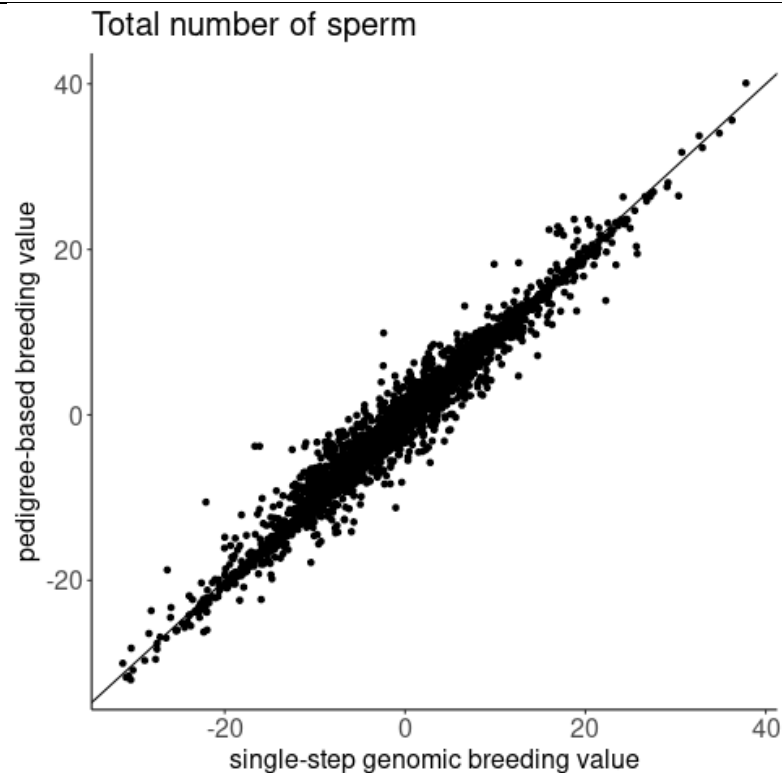
Figure 19: Single-step genomic and pedigree-based conventional breeding value estimates (in natural units, i.e. billion sperm cells) for total number of sperm in Pietrain pigs, the line resembles x = y

For a quantitative validation of the advantage of ssGBLUP compared to conventional BLUP estimation, an estimation was carried out using either method, with the boars born in 2019 and 2020 (n = 144 + 33) having their phenotypes removed. This is done to estimate their GEBVs like on young boars without phenotypic measures. The average reliabilities (r^2) of the traits' estimated breeding values are shown in Table 6.

Table 7: Mean reliabilities of four sperm quality traits in 2019/2020 born Pietrain boars with no own recordings, compared between BLUP and ssGBLUP

Trait	Mean reliability (r^2) in	
	pedigree-based BLUP	single-step GBLUP
Total number of sperm	0.217	0.369
Motility	0.161	0.270
Ejaculate volume	0.231	0.393
Sperm density	0.249	0.427

Using the genomic information in a single-step model caused a large increase of r^2 for young boars. The increase in reliability for total number of sperm and motility was substantial compared to the reliability of pedigree-based conventional BLUP values of

young animals. The higher r^2 might increase selection response and therefore make selection of young boars without phenotypes feasible (therefore reducing generation interval), like many other studies on genomic breeding value estimation have shown, e.g. VanRaden (2008).

This methodological comparison was not a main objective of the study but showed the effectiveness of the single-step method also in male fertility traits.

Another remarkable result were relatively high overall reliabilities, as boars with some decent recordings were in the range of 0.50 – 0.65 for total number of sperm and 0.40 – 0.50 for motility. Top reliabilities were over 0.70. These were high r^2 for pig evaluations, being in a range one would expect from dairy cattle. They are likely caused by the large number of repeated observations (ejaculations) per individual, similar to milk recordings in dairy cattle, as well as comparatively high heritabilities.

Table 8 and Table 9 show the top ten boars with the highest and lowest breeding values, respectively for total number of sperm and motility to the date of the estimation (data until 01/2021).

Table 8: Relative breeding values and reliabilities for total number of sperm of the ten highest and lowest ranked individuals

ID	Year of birth	Relative breeding value	Reliability	Genotyped
0004132574	2016	152.9	0.671	no
0002886579	2010	150.7	0.529	no
0002817814	2009	148.9	0.509	no
0003547246	2013	146.2	0.540	no
0004414015	2017	145.7	0.634	yes
0004425097	2017	142.9	0.589	yes
0004367468	2017	142.4	0.645	yes
0002651856	2008	140.9	0.549	no
0003553556	2013	140.8	0.515	no
0002400936	2007	138.6	0.611	no
0003671282	2013	61.2	0.594	no
0004504431	2017	60.6	0.678	yes
0004099336	2016	60.2	0.706	yes
0003660883	2013	59.8	0.579	no
0003705039	2014	57.9	0.590	no
0003933976	2015	57.8	0.654	yes
0004698087	2018	57.7	0.635	yes
0003490096	2013	57.3	0.578	no
0003185933	2011	57.0	0.577	no
0004557056	2018	56.4	0.585	no

Table 9: Relative breeding values and reliabilities for motility of the ten highest and lowest ranked individuals

ID	Year of birth	Relative breeding value	Reliability	Genotyped
0004601045	2018	128.6	0.526	yes
0004485365	2017	128.0	0.333	yes
0004127810	2016	126.5	0.500	yes
0003408063	2012	126.4	0.362	no
0004762261	2019	126.0	0.447	yes
0002494217	2008	125.4	0.313	no
0004511930	2017	125.1	0.362	no
0003407886	2012	124.8	0.379	no
0004504094	2017	124.8	0.433	yes
0004859812	2019	124.6	0.450	yes
0004376959	2017	48.1	0.533	yes
0004196263	2016	47.8	0.490	yes
0003231918	2012	46.9	0.561	no
0003612946	2013	45.8	0.387	no
0004461905	2017	43.3	0.418	no
0002911893	2010	40.8	0.344	no
0004561950	2018	34.8	0.451	yes
0004461600	2017	28.6	0.411	no
0004461601	2017	26.0	0.418	no
0003158348	2011	10.1	0.335	no

The existence of quite large deviations from the mean (100) in recent individuals implies a large selection potential.

5.4.1.   Genetic trend

We used the average breeding value of each birth-year's individuals as a measure for the genetic trend. This is expected to be positive or at least neutral: There is some positive natural selection on these traits because males with poor fertility will have less offspring or will be eliminated from AI stations earlier. The results for 2010 – 2019 are shown in the following graphs (Figure 20 and Figure 21):
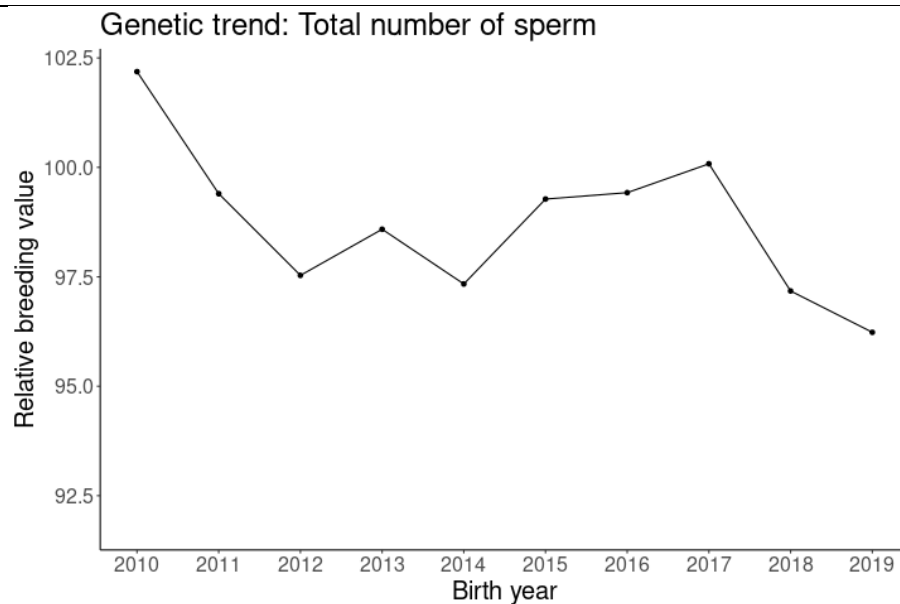
Figure 20: Genetic trend (average of relative breeding values per birth year) for total number of sperm for Pietrain boars
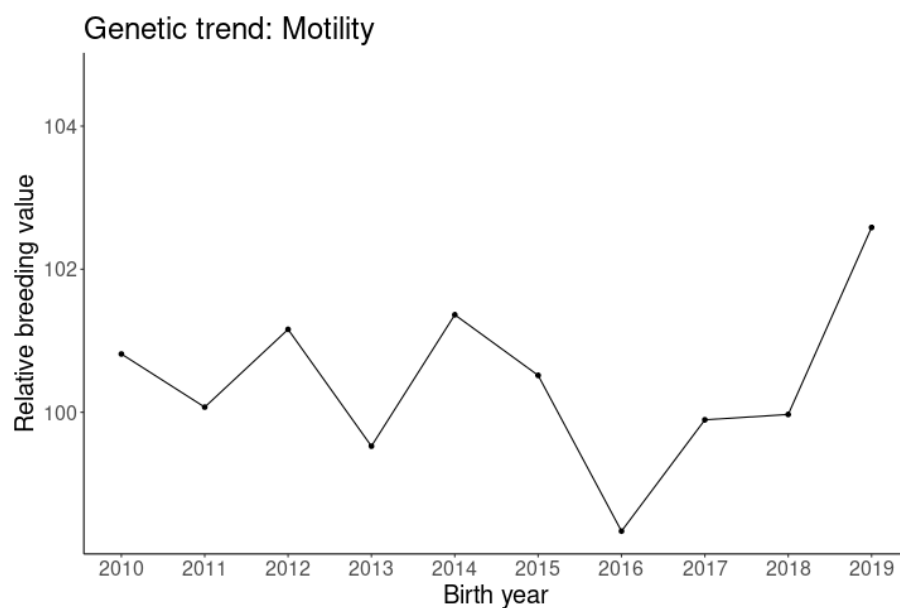
Figure 21: Genetic trend (average of relative breeding values per birth year) for motility for Pietrain boars

The data of 2020 consisted of only 40 boars, while the years before there were between 186 and 402 boars each year. Due to that the year 2020 is not shown in the figure above.

While the trend for total number of sperm dropped almost six points within the last nine years, the values for motility only increased by roughly two points. The reason for this is unknown. These findings are consistent with subjectively higher numbers of boars rejected recently due to sperm quality/quantity issues (Pfeiffer, 2021), though the changes are too small in absolute numbers to make conclusions.

These findings need to be monitored in future. Re-evaluations of the genetic trend of sperm numbers must be done to conclude whether there is an emerging problem. It may also help to evaluate other measures of sperm quality like morphological abnormalities to get a clearer view on the current development. The results of this study are going to be used as a starting point to monitor the genetic basis of sperm quality traits at porcine AI stations in Austria. The use of these traits for selection decisions may help to avoid future problems with sperm quality and reduce the elimination rate of boars.

6. Conclusions

The present study was able to answer part of its research questions. The genome-wide associations delivered some possible loci connected to spermatogenesis and sperm quality traits, but the methodology used should be refined. This may provide more concise signals and prove or falsify some of the findings here. Associations to the cadherin superfamily of proteins (CDH12, PCDH15) showed up two times, suggesting a role in spermatogenesis, if these associations prove right. The cytological background would support this, and some research has found N-type cadherins involved in spermatogenic cell adhesion.

The genetic parameter estimation resulted in heritabilities within the range known from literature. Genetic correlations showed a negative correlation of motility with semen volume, but positive with cell density which may provide an indirect selection opportunity on motility. Further research, especially correlations with distinct morphological traits (abnormalities) and more detailed motility parameters (progressive motility, etc) seems promising. This could find more concise connections to use in semen quality assessment and boar selection.

The single-step genomic BLUP evaluation proved its methodological superiority to pedigree-based BLUP regarding reliabilities of breeding values. Semen traits were shown to reach high reliabilities, normally not known from pig breeding value evaluations. This implies quick and precise response on selection if semen traits would be used as additional information in routine boar selection procedures. A slightly negative genetic trend for total number of sperm showed up in the last few years. This should be consequently monitored, and the selection strategy should be adapted if these findings prove true. Although production and vitality criteria will always be the most important traits in a pig selection index, the additional use of breeding values for semen traits could prove useful in boar selection for artificial and natural insemination.

7. Conflicts of interest

The author declares no conflict of interest regarding the research topic. Co-supervisor C. Pfeiffer is employed at PIG Austria GmbH which delivered the data and owns the boars.

8. References

Agarwal, A. K., Tunison, K., Dalal, J. S., Nagamma, S. S., Hamra, F. K., Sankella, S., Shao, X., Auchus, R. J., Garg, A. (2017). Metabolic, Reproductive, and Neurologic Abnormalities in Agpat1-Null Mice. *Endocrinology, 158*(11), 3954-3973. https://doi.org/10.1210/en.2017-00511

Althouse, G. C. (1998). Cytoplasmic droplets on boar sperm cells. *Swine Health and production, 6*(3), 128.

Althouse, G. C., Wilson, M. E., Kuster, C., & Parsley, M. (1998). Characterization of lower temperature storage limitations of fresh-extended porcine semen. *Theriogenology, 50*(4), 535-543. https://doi.org/https://doi.org/10.1016/S0093-691X(98)00159-9

Alvarez-Rodriguez, M., Martinez, C., Wright, D., Barranco, I., Roca, J., & Rodriguez-Martinez, H. (2020). The Transcriptome of Pig Spermatozoa, and Its Role in Fertility. *International Journal of Molecular Sciences, 21*(5). https://doi.org/10.3390/ijms21051572

Amann, R. P., & Waberski, D. (2014). Computer-assisted sperm analysis (CASA): Capabilities and potential developments. *Theriogenology, 81*(1), 5-17.e13. https://doi.org/https://doi.org/10.1016/j.theriogenology.2013.09.004

Balhorn, R. (2007). The protamine family of sperm nuclear proteins. *Genome biology, 8*(9), 227-227. https://doi.org/10.1186/gb-2007-8-9-227

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bjerre, D., Madsen, L. B., Bendixen, C., & Larsen, K. (2006). Porcine Parkin: Molecular cloning of PARK2 cDNA, expression analysis, and identification of a splicing variant. *Biochemical and Biophysical Research Communications, 347*(3), 803-813. https://doi.org/https://doi.org/10.1016/j.bbrc.2006.06.167

BMLRT - Ministry of Agriculture Regions and Tourism. (2020). *Grüner Bericht 2020*. Vienna: self-publication. Retrieved from www.gruenerbericht.at

Broekhuijse, M. L. W. J., Šoštarić, E., Feitsma, H., & Gadella, B. M. (2012). Application of computer-assisted semen analysis to explain variations in pig fertility1. *Journal of Animal Science, 90*(3), 779-789. https://doi.org/10.2527/jas.2011-4311

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience, 4*(1), s13742-13015.

Choudhary, D., Narui, Y., Neel, B. L., Wimalasena, L. N., Klanseck, C. F., De-la-Torre, P., Chen, C., Araya-Secchi, R., Tamilselvan, E., Sotomayor, M. (2020). Structural determinants of protocadherin-15 mechanics and function in hearing and balance perception. *Proceedings of the National Academy of Sciences, 117*(40), 24837. https://doi.org/10.1073/pnas.1920444117

Cooper, T. G. (2011). The epididymis, cytoplasmic droplets and male fertility. *Asian journal of andrology, 13*(1), 130-138. https://doi.org/10.1038/aja.2010.97

Diniz, D. B., Lopes, M. S., Broekhuijse, M. L. W. J., Lopes, P. S., Harlizius, B., Guimarães, S. E. F., Duijvesteijn, N., Knol, E. F., Silva, F. F. (2014). A genome-wide association study reveals a novel candidate gene for sperm motility in pigs. *Animal Reproduction Science, 151*(3), 201-207. https://doi.org/https://doi.org/10.1016/j.anireprosci.2014.10.014

Doppelhofer, V. (2021). Personal communication.

Elliott, D. J., & Cooke, H. J. (1997). The molecular genetics of male infertility. *Bioessays*, *19*(9), 801-809.

Fallah, A., Mohammad-Hasani, A., & Colagar, A. H. (2018). Zinc is an Essential Element for Male Fertility: A Review of Zn Roles in Men's Health, Germination, Sperm Quality, and Fertilization. *Journal of reproduction & infertility*, *19*(2), 69-81.

Gille, U. (2006). *Histological section of the testicular parenchyma of wild boars* Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Testicle-histology-boar.jpg?uselang=de

Gilmour, A. R., Gogel, B. J., Cullis, B. R., & Thompson, R. (2009). ASReml user guide release 3.0 VSN International Ltd. *Hemel Hempstead, HP1 1ES, UK www. vsni. co. uk*.

Pig Austria GmbH (2021a). *Programm*. Retrieved August 31, 2021 from https://www.pig.at/programm-12998.html

Pig Austria GmbH (2021b). *Unternehmen*. Retrieved August 31, 2021 from https://www.pig.at/unternehmen-12997.html

Godia, M., Reverter, A., Gonzalez-Prendes, R., Ramayo-Caldas, Y., Castello, A., Rodriguez-Gil, J. E., . . . Clop, A. (2020). An integrative systems biology approach to identify the molecular basis of sperm quality in swine.

Hale, D. W. (1996). Mammalian Spermatogenesis. In R. S. Verma (Ed.), *Advances in Genome Biology* (Vol. 4, pp. 249-304). JAI. https://doi.org/https://doi.org/10.1016/S1067-5701(96)80011-8

Henderson, C. R. (1976). A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics*, 69-83.

Illumina Inc. (2021). Porcine SNP60 v2.0 Manifest File.

Imai, Y., Soda, M., Inoue, H., Hattori, N., Mizuno, Y., & Takahashi, R. (2001). An Unfolded Putative Transmembrane Polypeptide, which Can Lead to Endoplasmic Reticulum Stress, Is a Substrate of Parkin. *Cell*, *105*(7), 891-902. https://doi.org/https://doi.org/10.1016/S0092-8674(01)00407-X

Johnson, L., Blanchard, T. L., Varner, D. D., & Scrutchfield, W. L. (1997). Factors affecting spermatogenesis in the stallion. *Theriogenology*, *48*(7), 1199-1216. https://doi.org/https://doi.org/10.1016/S0093-691X(97)00353-1

Kunavongkrit, A., Suriyasomboon, A., Lundeheim, N., Heard, T. W., & Einarsson, S. (2005). Management and sperm production of boars under differing environmental conditions. *Theriogenology*, *63*(2), 657-667. https://doi.org/https://doi.org/10.1016/j.theriogenology.2004.09.039

Lovercamp, K. W., Safranski, T. J., Fischer, K. A., Manandhar, G., Sutovsky, M., Herring, W., & Sutovsky, P. (2007). High Resolution Light Microscopic Evaluation of Boar Semen Quality Sperm Cytoplasmic Droplet Retention in Relationship with Boar Fertility Parameters. *Archives of Andrology*, *53*(4), 219-228. https://doi.org/10.1080/01485010701426463

Marques, D., Lopes, M., Broekhuijse, M., Guimarães, S., Knol, E., Bastiaansen, J., Silva, F., Lopes, P. (2017). Genetic parameters for semen quality and quantity traits in five pig lines. *Journal of Animal Science*, *95*. https://doi.org/10.2527/jas2017.1683

Marques, D. B. D., Bastiaansen, J. W. M., Broekhuijse, M. L. W. J., Lopes, M. S., Knol, E. F., Harlizius, B., Guimarães, S. E. F., Silva, F. F., Lopes, P. S. (2018). Weighted single-step GWAS and gene network analysis reveal new candidate

genes for semen traits in pigs. *Genetics Selection Evolution*, *50*(1), 40. https://doi.org/10.1186/s12711-018-0412-z

Meyer, K., & Houle, D. (2013). Sampling based approximation of confidence intervals for functions of genetic covariance matrices. Proc. Assoc. Advmt. Anim. Breed. Genet,

Microsoft Corporation (2021). Microsoft Excel.

Misztal, I., Tsuruta, S., Lourenco, D. A. L., Masuda, Y., Aguilar, I., Legarra, A., & Vitezica, Z. (2018). Manual for BLUPF90 family programs. In: University of Georgia.

Nakanishi, Y., & Shiratsuchi, A. (2004). Phagocytic Removal of Apoptotic Spermatogenic Cells by Sertoli Cells: Mechanisms and Consequences. *Biological and Pharmaceutical Bulletin*, *27*(1), 13-16. https://doi.org/10.1248/bpb.27.13

NCBI - National Center for Biotechnology Information. (2021).

Newton, S. C., Blaschuk, O. W., & Millette, C. F. (1993). N-cadherin mediates sertoli cell-spermatogenic cell adhesion. *Developmental Dynamics*, *197*(1), 1-13. https://doi.org/https://doi.org/10.1002/aja.1001970102

Nosková, A., Hiltpold, M., Janett, F., Echtermann, T., Fang, Z.-H., Sidler, X., Selige, C., Hofer, A., Neuenschwander, S., Pausch, H. (2021). Infertility due to defective sperm flagella caused by an intronic deletion in DNAH17 that perturbs splicing. *Genetics*, *217*(2). https://doi.org/10.1093/genetics/iyaa033

Pfeiffer, C. (2020). Häufigkeiten der Abgangsursachen in % nach Rassen.

Pfeiffer, C. (2021). Personal communication.

Pfeiffer, C., & Kreiner, T. (2020). *Umrauscher reduzieren - Merkblatt PIG Austria.*

Purcell, S., & Chang, C. (2021). *PLINK 1.9.* Retrieved September 20 from www.cog-genomics.org/plink/1.9/

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. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics*, *81*(3), 559-575.

Rosenfeld, C. S., Javurek, A. B., Johnson, S. A., Lei, Z., Sumner, L. W., & Hess, R. A. (2018). Seminal fluid metabolome and epididymal changes after antibiotic treatment in mice. *Reproduction (Cambridge, England)*, *156*(1), 1-10. https://doi.org/10.1530/rep-18-0072

RStudio PBC. (2021). *RStudio*. In (Version 1.4.1717) https://www.rstudio.com/

Ruiz, M. (2009). *Simplified spermatozoon diagram* Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Simplified_spermatozoon_diagram.svg

Rusz, A., Pilatz, A., Wagenlehner, F., Linn, T., Diemer, T., Schuppe, H. C., Lohmeyer, J., Hossain, H., Weidner, W. (2012). Influence of urogenital infections and inflammation on semen quality and male fertility. *World Journal of Urology*, *30*(1), 23-30. https://doi.org/10.1007/s00345-011-0726-8

Schulze, M., Buder, S., Rüdiger, K., Beyerbach, M., & Waberski, D. (2014). Influences on semen traits used for selection of young AI boars. *Animal Reproduction Science*, *148*(3), 164-170. https://doi.org/https://doi.org/10.1016/j.anireprosci.2014.06.008

Schulze, M., Revilla-Fernández, S., Schmoll, F., Grossfeld, R., & Griessler, A. (2013). Effects on boar semen quality after infection with porcine reproductive and

respiratory syndrome virus: a case report. *Acta Veterinaria Scandinavica*, *55*(1), 16. https://doi.org/10.1186/1751-0147-55-16

Smital, J., Wolf, J., & Sousa, L. L. D. (2005). Estimation of genetic parameters of semen characteristics and reproductive traits in AI boars. *Animal Reproduction Science*, *86*(1), 119-130. https://doi.org/https://doi.org/10.1016/j.anireprosci.2004.05.023

Tsakmakidis, I. A., Lymberopoulos, A. G., & Khalifa, T. A. A. (2010). Relationship between sperm quality traits and field-fertility of porcine semen. *Journal of veterinary science*, *11*(2), 151-154.

Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, *3*. https://doi.org/10.21105/joss.00731

VanRaden, P. M. (2008). Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, *91*(11), 4414-4423. https://doi.org/https://doi.org/10.3168/jds.2007-0980

Wang, Y., Ding, X., Tan, Z., Xing, K., Yang, T., Sun, D., & Wang, C. (2018). Genome-wide association study for reproductive traits in a Large White pig population [https://doi.org/10.1111/age.12638]. *Animal Genetics*, *49*(2), 127-131. https://doi.org/https://doi.org/10.1111/age.12638

Warr, A., Affara, N., Aken, B., Beiki, H., Bickhart, D. M., Billis, K., . . . Archibald, A. L. (2020). An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience*, *9*(6). https://doi.org/10.1093/gigascience/giaa051

Wolf, J. (2010). Heritabilities and genetic correlations for litter size and semen traits in Czech Large White and Landrace pigs1. *Journal of Animal Science*, *88*(9), 2893-2903. https://doi.org/10.2527/jas.2009-2555

Wu, S., Yan, M., Ge, R., & Cheng, C. Y. (2020). Crosstalk between Sertoli and Germ Cells in Male Fertility. *Trends in Molecular Medicine*, *26*(2), 215-231. https://doi.org/https://doi.org/10.1016/j.molmed.2019.09.006

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76-82.

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics*, *46*(2), 100-106.

Yeung, C.-H., & Cooper, T. G. (2002). Acquisition and development of sperm motility upon maturation in the epididymis. *The epididymis: from molecules to clinical practice*, 417-434.

Zhao, Y., Gao, N., Li, X., El-Ashram, S., Wang, Z., Zhu, L., Jiang, W., Peng, X., Zhang, C., Chen, Y., Li, Z. (2020). Identifying candidate genes associated with sperm morphology abnormalities using weighted single-step GWAS in a Duroc boar population. *Theriogenology*, *141*, 9-15. https://doi.org/https://doi.org/10.1016/j.theriogenology.2019.08.031

9. Tables

Table 10: Number of observations (ejaculates) on AI stations Gleisdorf, Hohenwarth and Steinhaus for Large White (LW), Landrace (LR) and Pietrain (PI)

AI Station	Breed	Number of observations	
Gleisdorf	LW	2,524	
	LR	363	17,572
	PI	14,685	
Hohenwarth	LW	1,018	
	LR	1,020	23,775
	PI	21,737	
Steinhaus	LW	4,543	
	LR	6,576	86,197
	PI	75,078	

Table 11: Means and standard deviations of total number of sperm, motility, volume, density, storage time, age, ejaculation interval and number of observations per breed for Large White (LW), Landrace (LR) and Pietrain (PI).

Breed	Total number of sperm [bn]	Motility [%]	Volume [ml]	Density [bn * ml^{-1}]	Storage time [d]	Age [d]	Ejaculation interval [d]	Number of obs.
LW	88.04 +-31.83	84.66 +-8.41	229.2 +-97.1	0.4346 +-0.2082	3.737 +-4.684	725.8 +-361.4	9.150 +-5.429	8,085
LR	89.60 +-32.47	88.30 +-7.75	248.5 +-105.5	0.4066 +-0.1922	4.685 +-2.662	667.6 +-369.0	10.544 +-5.787	7,959
PI	83.45 +-30.00	87.30 +-8.14	257.3 +-97.5	0.3601 +-0.1709	4.102 +-3.286	844.2 +-501.0	7.268 +-3.200	111,500
Total	84.12	87.20	255.0	0.3677	4.115	825.7	7.592	127,544

Table 12: Numbers of breeders and boars of Large White (LW), Landrace (LR) and Pietrain (PI)

Breed	Number of breeders	Number of boars
LW	28	209
LR	33	272
PI	38	1,795
Total	78[a]	2,276

[a] Note: The total number of breeders (farms) was smaller than the sum of the three values above as some herdbook breeders keep more than one breed.

Table 13: Ten highest and lowest breeders' effects for total number of sperm (left, in billions) and motility (right, in %)

Breeder (anonymised)	Effect [billions]	Breeder (anonymised)	Effect [%]
ZNR100387	35.37	ZNR114	2.8
ZNR9399	33.25	ZNR897	1.3
ZNR812	16.53	ZNR158	1.1
ZNR114	12.94	ZNR100043	0.7
ZNR722	6.88	ZNR146	0.7
ZNR897	2.69	ZNR251	0.6
ZNR9815	-1.86	ZNR370	0.5
ZNR9961	-3.94	ZNR722	0.4
ZNR148	-4.78	ZNR89	0.3
ZNR106	-5.00	ZNR17	0.1
ZNR35	-15.24	ZNR194	-0.9
ZNR87	-16.85	ZNR100387	-0.9
ZNR194	-17.81	ZNR627	-0.9
ZNR750	-17.90	ZNR750	-1.1
ZNR845	-20.54	ZNR87	-1.1
ZNR627	-21.70	ZNR812	-1.2
ZNR370	-22.79	ZNR48	-1.5
ZNR155	-27.32	ZNR155	-1.7
ZNR48	-28.27	ZNR9399	-2.0
ZNR100043	-35.37	ZNR9815	-2.8

Table 14: Heritabilities, phenotypic and genetic correlations of sperm traits (diagonal: heritabilities, upper triangular matrix: phenotypic correlations, lower triangular matrix: genetic correlations), with their estimated standard errors of prediction

	Total number of sperm	Motility	Volume	Density
Total number of sperm	**0.247** ±0.027	0.153 ±0.013	0.304 ±0.014	0.486 ±0.012
Motility	-0.089 ±0.112	**0.131** ±0.022	-0.124 ±0.013	0.237 ±0.013
Volume	0.121 ±0.080	-0.305 ±0.094	**0.304** ±0.027	-0.557 ±0.010
Density	0.547 ±0.059	0.206 ±0.096	-0.759 ±0.038	**0.340** ±0.028

Table 15: Pearson's correlation coefficients between single-step genomic BLUP and conventional BLUP breeding value estimates, for all four traits

Trait	Correlation coefficient (r)
Total number of sperm	0.981
Motility	0.945
Volume	0.982
Density	0.982

Table 16: Mean reliabilities of four sperm quality traits in 2019/2020 born Pietrain boars with no own recordings, compared between BLUP and ssGBLUP

Trait	Mean reliability (r²) in	
	pedigree-based BLUP	single-step GBLUP
Total number of sperm	0.217	0.369
Motility	0.161	0.270
Ejaculate volume	0.231	0.393
Sperm density	0.249	0.427

Table 17: Relative breeding values and reliabilities for total number of sperm of the ten highest and lowest ranked individuals

ID	Year of birth	Relative breeding value	Reliability	Genotyped
0004132574	2016	152.9	0.671	no
0002886579	2010	150.7	0.529	no
0002817814	2009	148.9	0.509	no
0003547246	2013	146.2	0.540	no
0004414015	2017	145.7	0.634	yes
0004425097	2017	142.9	0.589	yes
0004367468	2017	142.4	0.645	yes
0002651856	2008	140.9	0.549	no
0003553556	2013	140.8	0.515	no
0002400936	2007	138.6	0.611	no
0003671282	2013	61.2	0.594	no
0004504431	2017	60.6	0.678	yes
0004099336	2016	60.2	0.706	yes
0003660883	2013	59.8	0.579	no
0003705039	2014	57.9	0.590	no
0003933976	2015	57.8	0.654	yes
0004698087	2018	57.7	0.635	yes
0003490096	2013	57.3	0.578	no
0003185933	2011	57.0	0.577	no
0004557056	2018	56.4	0.585	no

Table 18: Relative breeding values and reliabilities for motility of the ten highest and lowest ranked individuals

ID	Year of birth	Relative breeding value	Reliability	Genotyped
0004601045	2018	128.6	0.526	yes
0004485365	2017	128.0	0.333	yes
0004127810	2016	126.5	0.500	yes
0003408063	2012	126.4	0.362	no
0004762261	2019	126.0	0.447	yes
0002494217	2008	125.4	0.313	no
0004511930	2017	125.1	0.362	no
0003407886	2012	124.8	0.379	no
0004504094	2017	124.8	0.433	yes
0004859812	2019	124.6	0.450	yes
0004376959	2017	48.1	0.533	yes
0004196263	2016	47.8	0.490	yes
0003231918	2012	46.9	0.561	no
0003612946	2013	45.8	0.387	no
0004461905	2017	43.3	0.418	no
0002911893	2010	40.8	0.344	no
0004561950	2018	34.8	0.451	yes
0004461600	2017	28.6	0.411	no
0004461601	2017	26.0	0.418	no
0003158348	2011	10.1	0.335	no