B. Czech, E. Dobkowski, M. Mielczarek, M. Frąszczak, J. Szyda
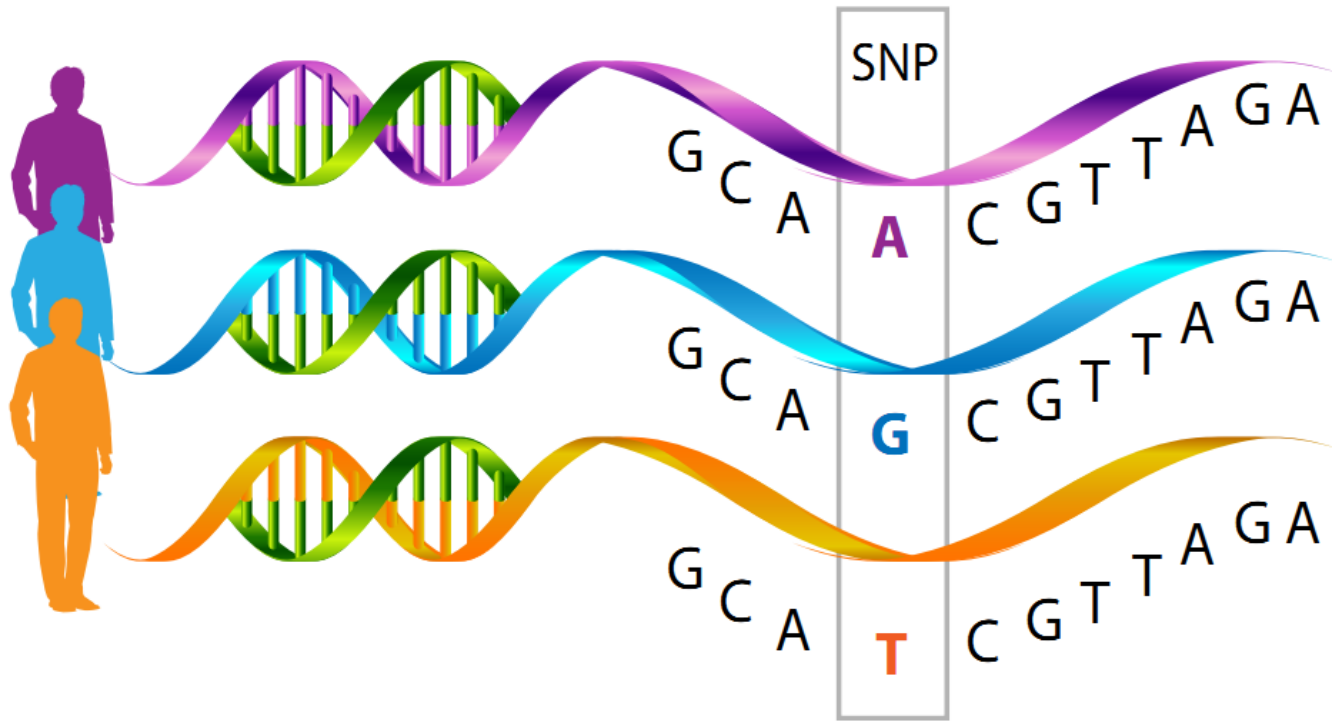
THETA
Statistical Genetics Group
Institute of Animal Genetics

The construction of *Bos taurus* breed specific reference genomes for Fleckvieh, Simmental, Guernsey and Brown Swiss

Introduction

- Reference genomes
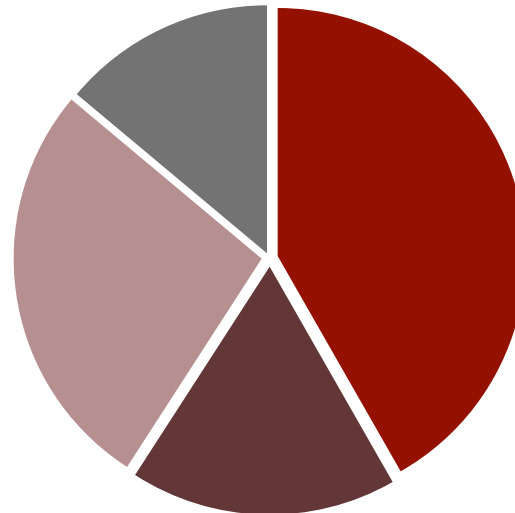- SNPs – genetic marker
- Bioinformatics and data mining

Objectives

- Creating breed specific reference genomes
- Analysis and characterization of selected variants

Data set

- Whole genome DNAseq – 115 bulls
- Illumina HiSeq 2000 next-generation sequencing platform
- 7-28 x genome average coverage
- Reference genome UMD 3.1
- 4 breeds

● Brown Swiss   ● Guernsey
● Fleckvieh   ● Simmental

Methods

BWA-MEM

Aligment to the reference genome

Picard, SAMtools

Post-aligment processes

VCF Tools

SNPs filtering

GATK

SNPs detection

Methods

BWA-MEM

Aligment to the reference genome

Picard, SAMtools

Post-aligment processes

VCF Tools

SNPs filtering

GATK

SNPs detection

7% missing genotypes, AF = 1.0

Methods

BWA-MEM

Aligment to the reference genome

Picard, SAMtools

Post-aligment processes

VCF Tools

SNPs filtering

GATK

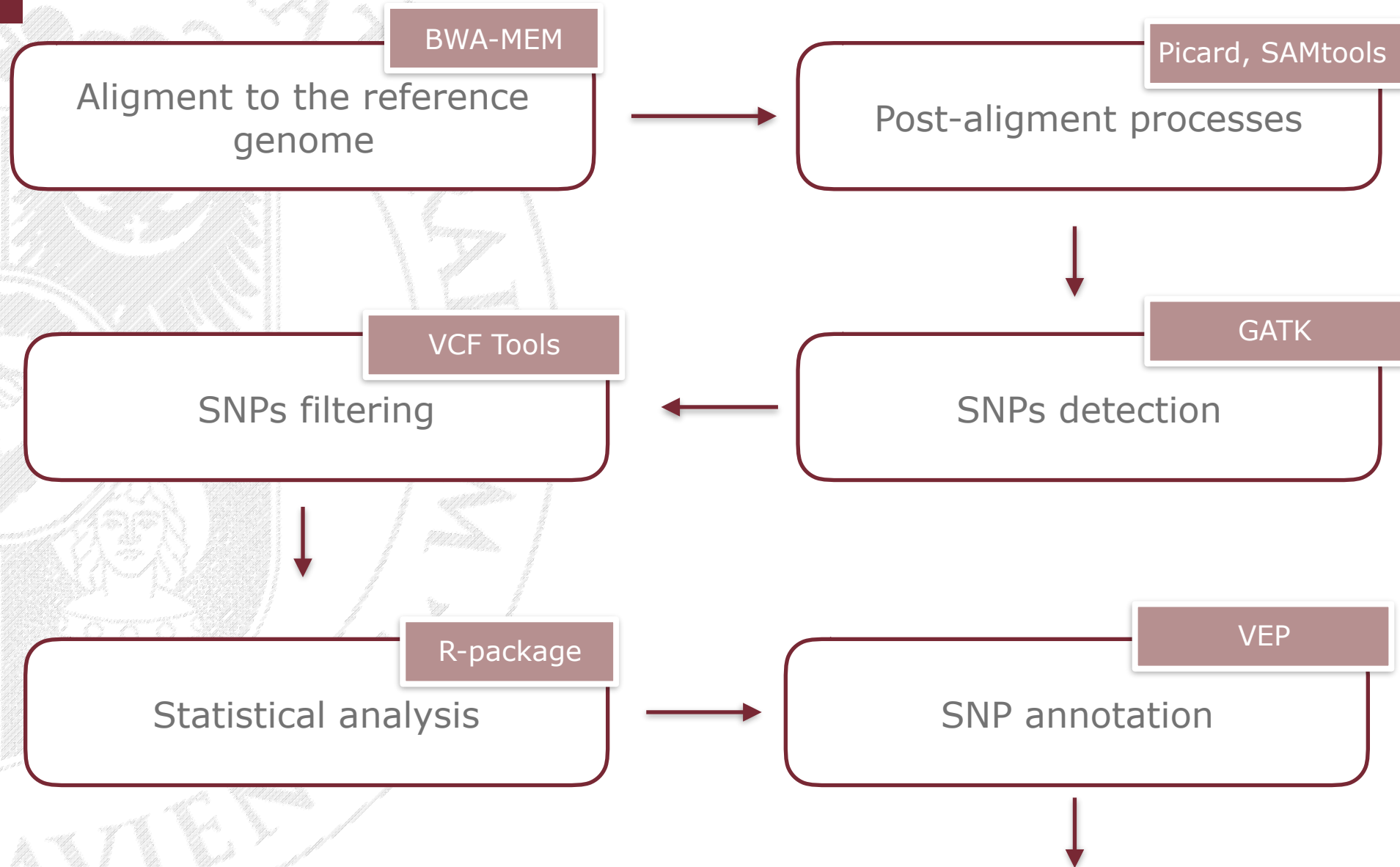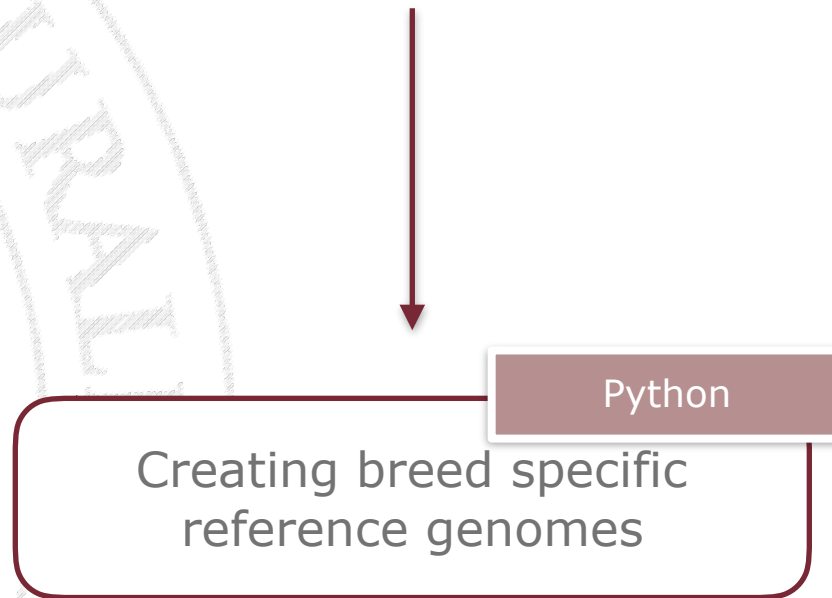SNPs detection

R-package

Statistical analysis

VEP

SNP annotation

Methods – final step

Python

Creating breed specific
reference genomes
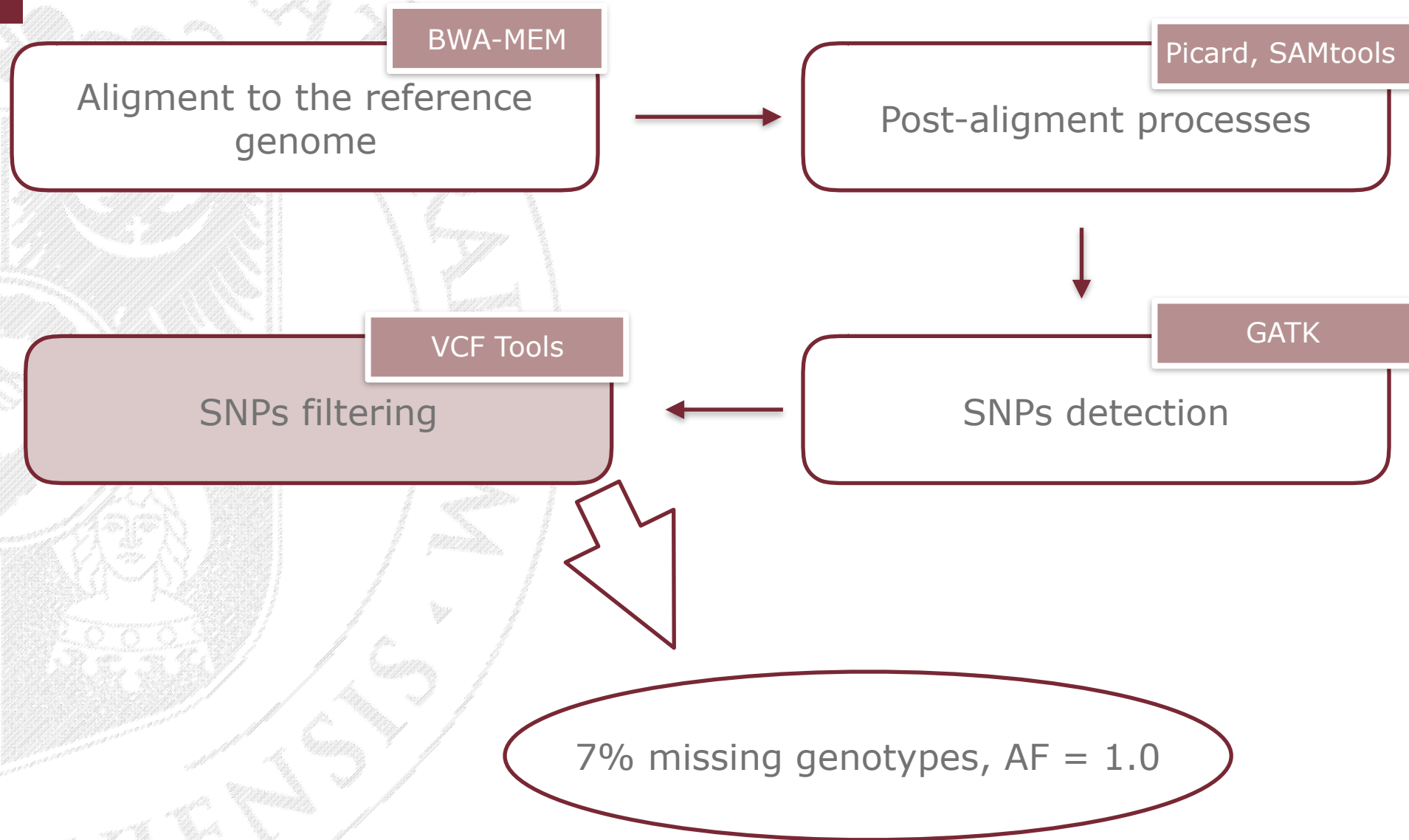
Results – identified and filtered SNPs

Breed	Total number of SNPs	Number of SNPs after filtration	Percent
Brown Swiss	15 615 013	371 988	2.382%
Fleckvieh	17 401 625	182 612	1.049%
Guernsey	14 259 793	467 338	3.277%
Simmental	15 181 467	354 343	2.334%

Results – Venn diagram
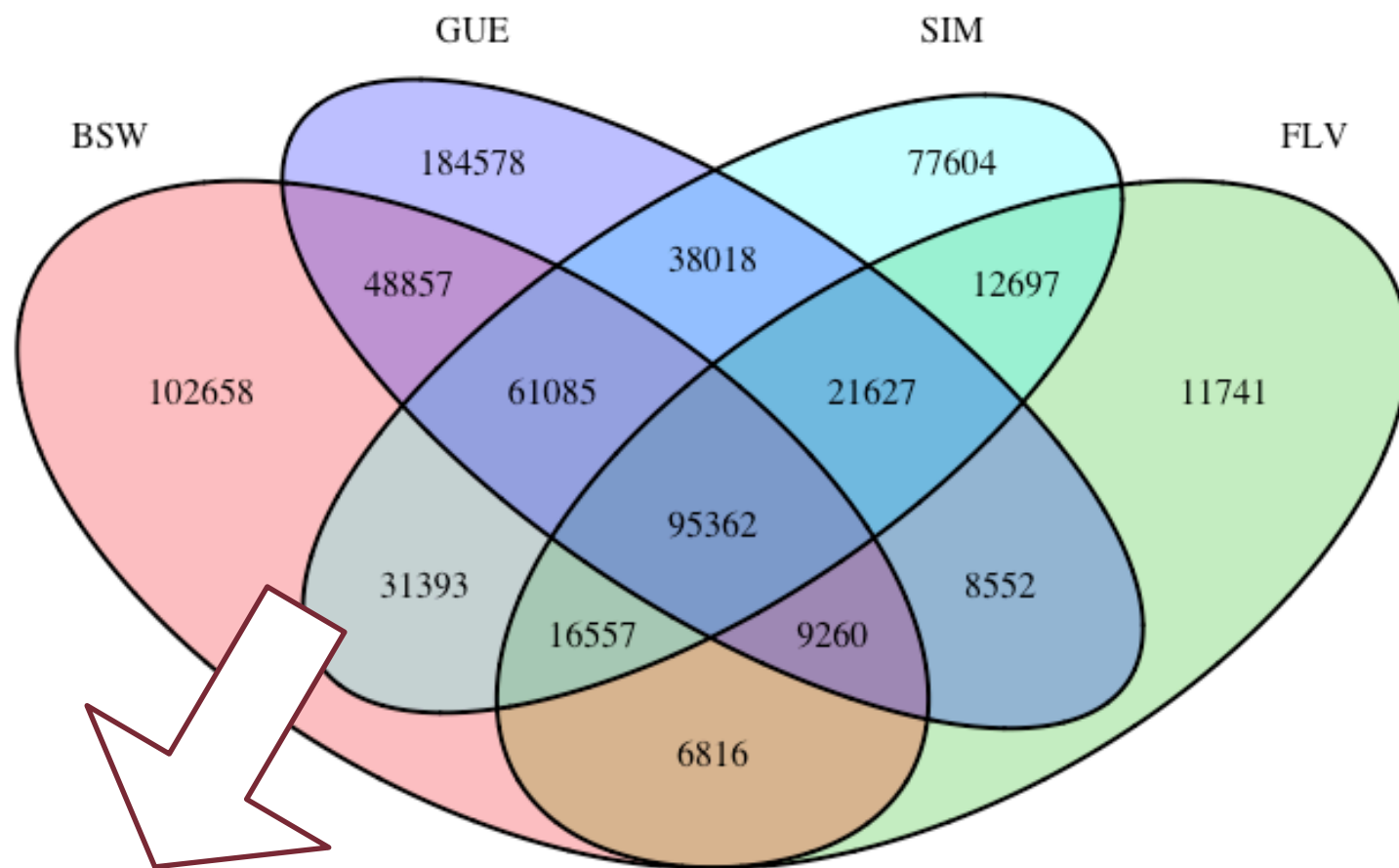
Unique SNPs:
Brown Swiss = 102,658   Simmental = 77,605
Guernsey   = 184,578   Fleckvieh   = 11,741

Results – annotation (all consequences) – Brown Swiss

- ⬤ intergenic_variant: 64%
- ⬤ intron_variant: 27%
- ⬤ downstream_gene_variant: 4%
- ⬤ upstream_gene_variant: 4%
- ⬤ synonymous_variant: 0%
- ⬤ missense_variant: 0%
- ⬤ 3_prime_UTR_variant: 0%
- ⬤ splice_region_variant: 0%
- ⬤ 5_prime_UTR_variant: 0%
- ⬤ Others

Results – annotation (coding consequences) – Brown Swiss

synonymous_variant: 57%
missense_variant: 43%
coding_sequence_variant: 0%
stop_lost: 0%
start_lost: 0%
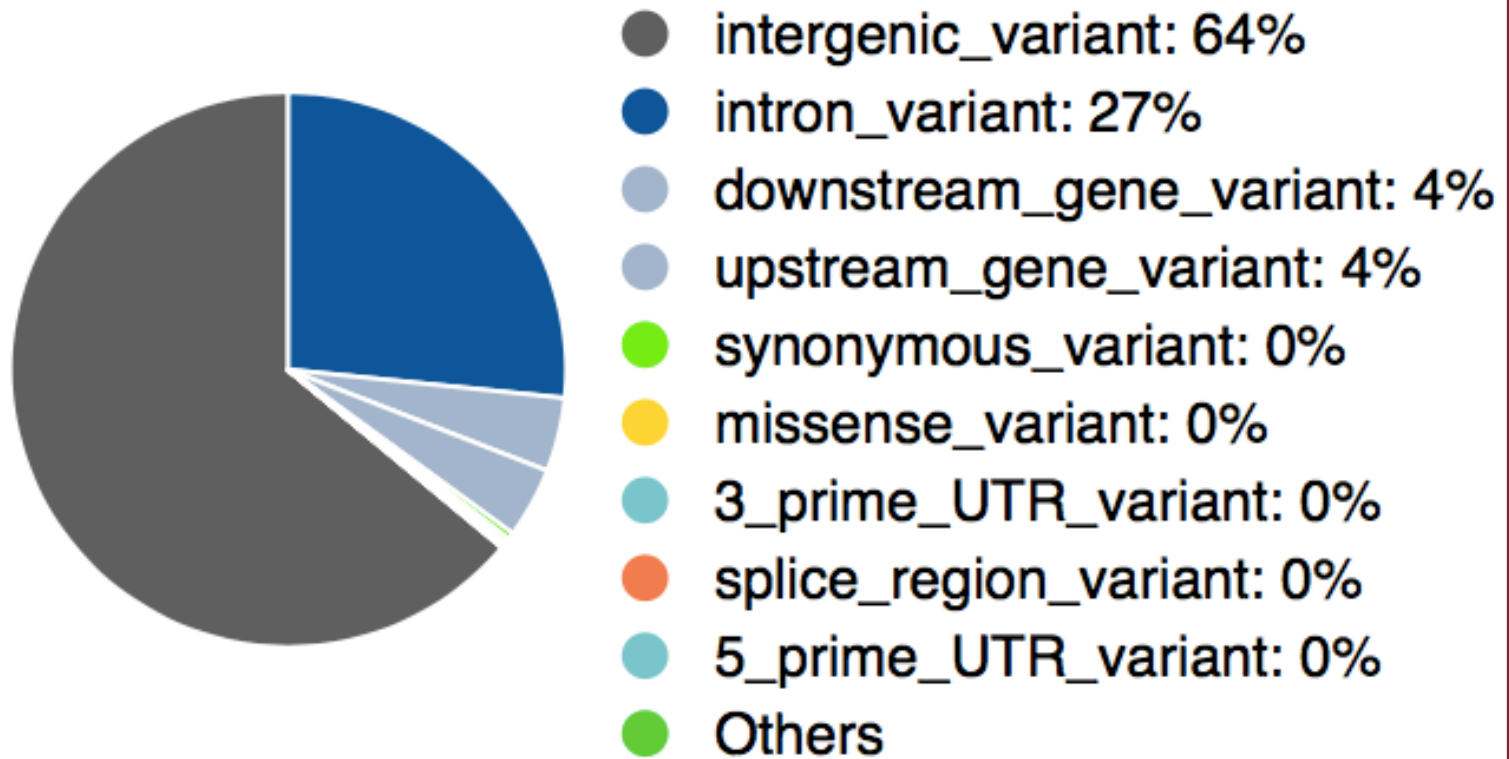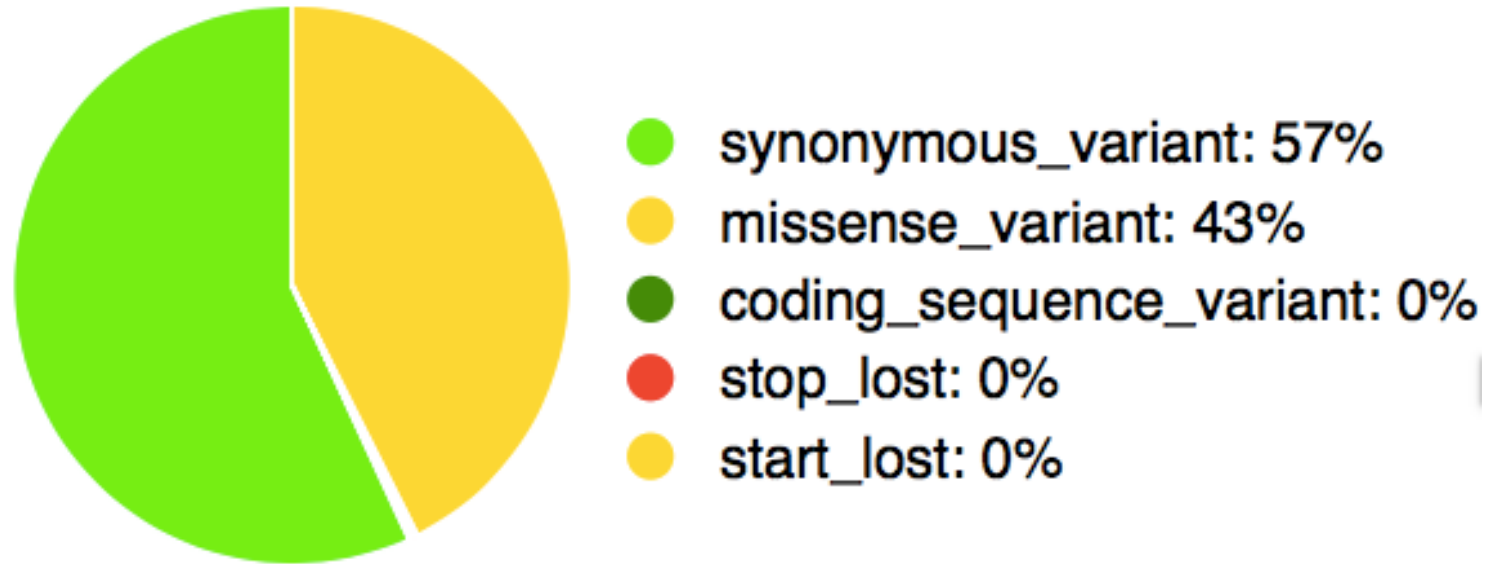
Number of genes = 445

Results – annotation (all consequences) – Fleckvieh

- intergenic_variant: 59%
- intron_variant: 29%
- upstream_gene_variant: 6%
- downstream_gene_variant: 5%
- synonymous_variant: 1%
- missense_variant: 0%
- 3_prime_UTR_variant: 0%
- 5_prime_UTR_variant: 0%
- splice_region_variant: 0%
- Others

Results – annotation (coding consequences) – Fleckvieh

- synonymous_variant: 52%
- missense_variant: 48%
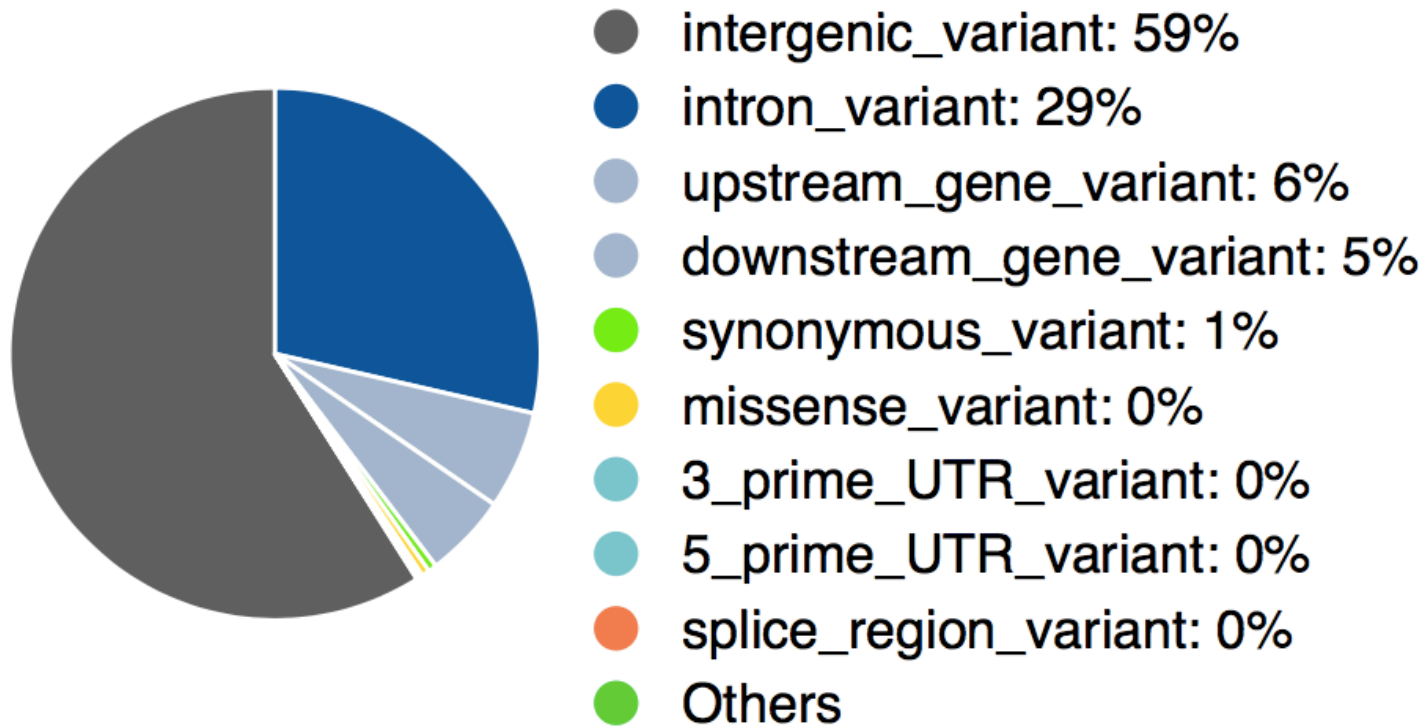
Number of genes = 80

Results – annotation (all consequences) – Guernsey

- intergenic_variant: 64%
- intron_variant: 27%
- downstream_gene_variant: 4%
- upstream_gene_variant: 4%
- synonymous_variant: 0%
- missense_variant: 0%
- 3_prime_UTR_variant: 0%
- splice_region_variant: 0%
- 5_prime_UTR_variant: 0%
- Others

Results – annotation (coding consequences) – Guernsey

- synonymous_variant: 58%
- missense_variant: 41%
- stop_gained: 0%
- coding_sequence_variant: 0%
- start_lost: 0%
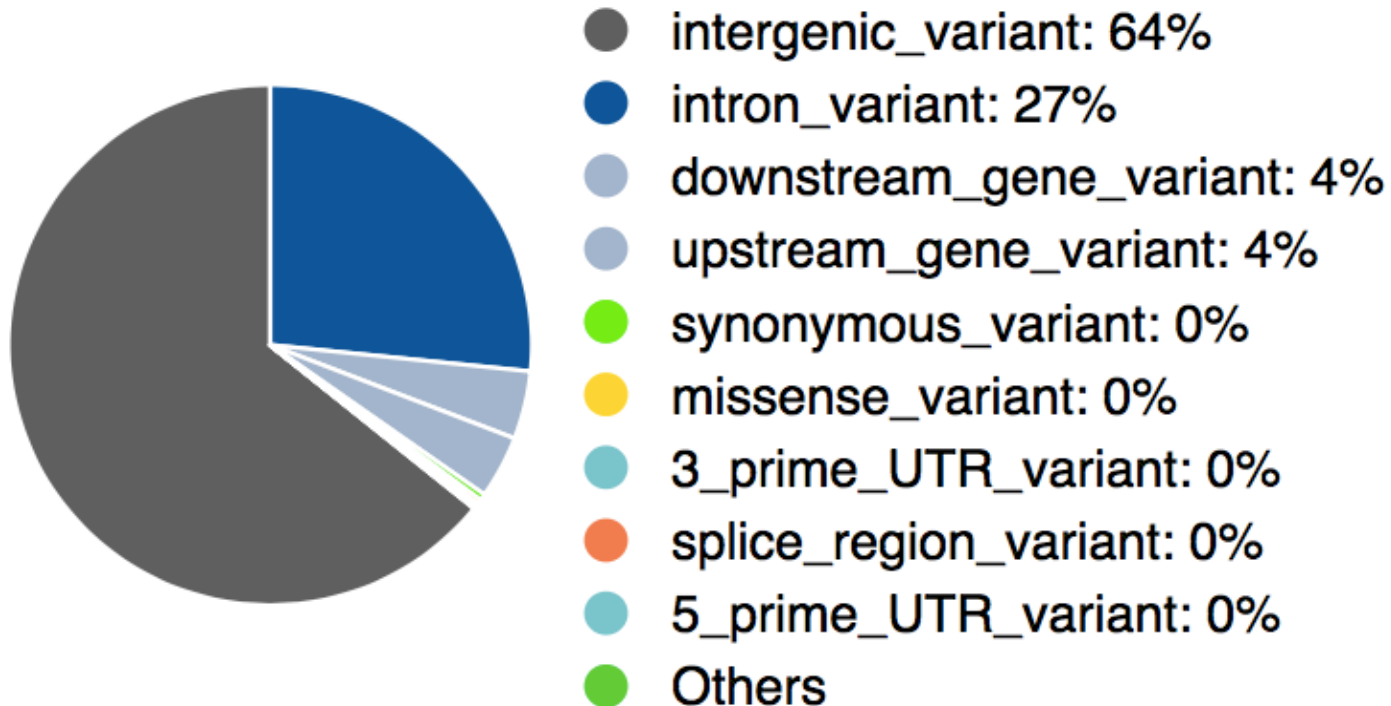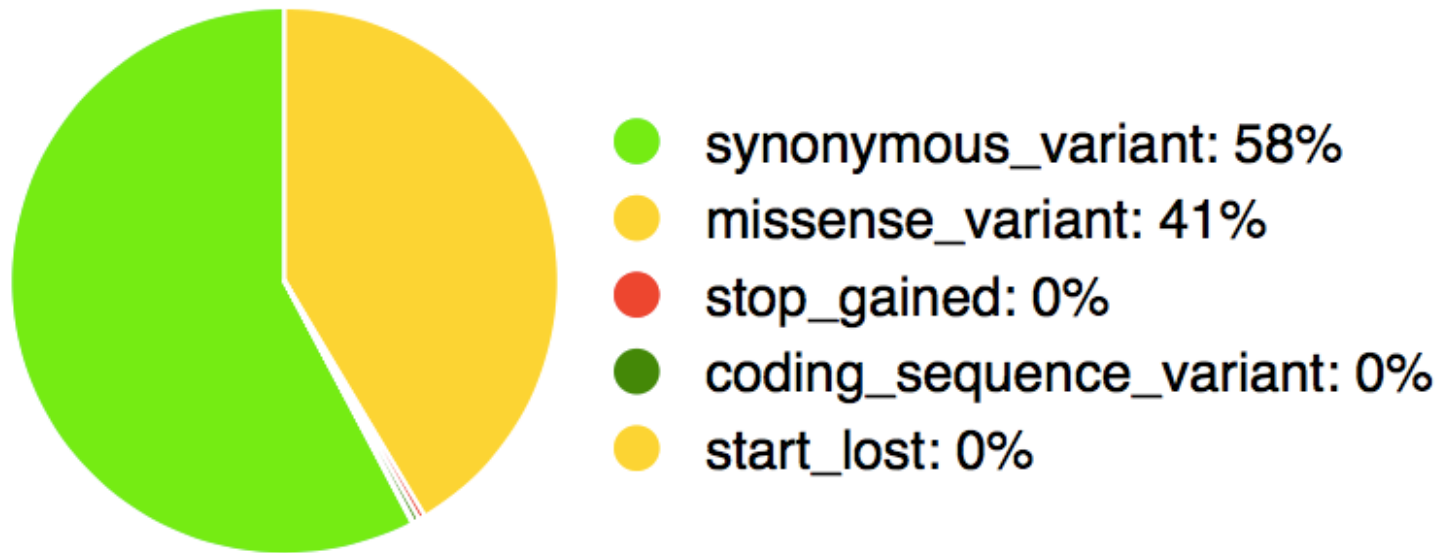
Number of genes = 752

Results – annotation (all consequences) – Simmental

- intergenic_variant: 63%
- intron_variant: 26%
- downstream_gene_variant: 5%
- upstream_gene_variant: 5%
- synonymous_variant: 0%
- missense_variant: 0%
- 3_prime_UTR_variant: 0%
- splice_region_variant: 0%
- 5_prime_UTR_variant: 0%
- Others

Results – annotation (coding consequences) – Simmental

- synonymous_variant: 59%
- missense_variant: 39%
- coding_sequence_variant: 2%
- start_lost: 0%
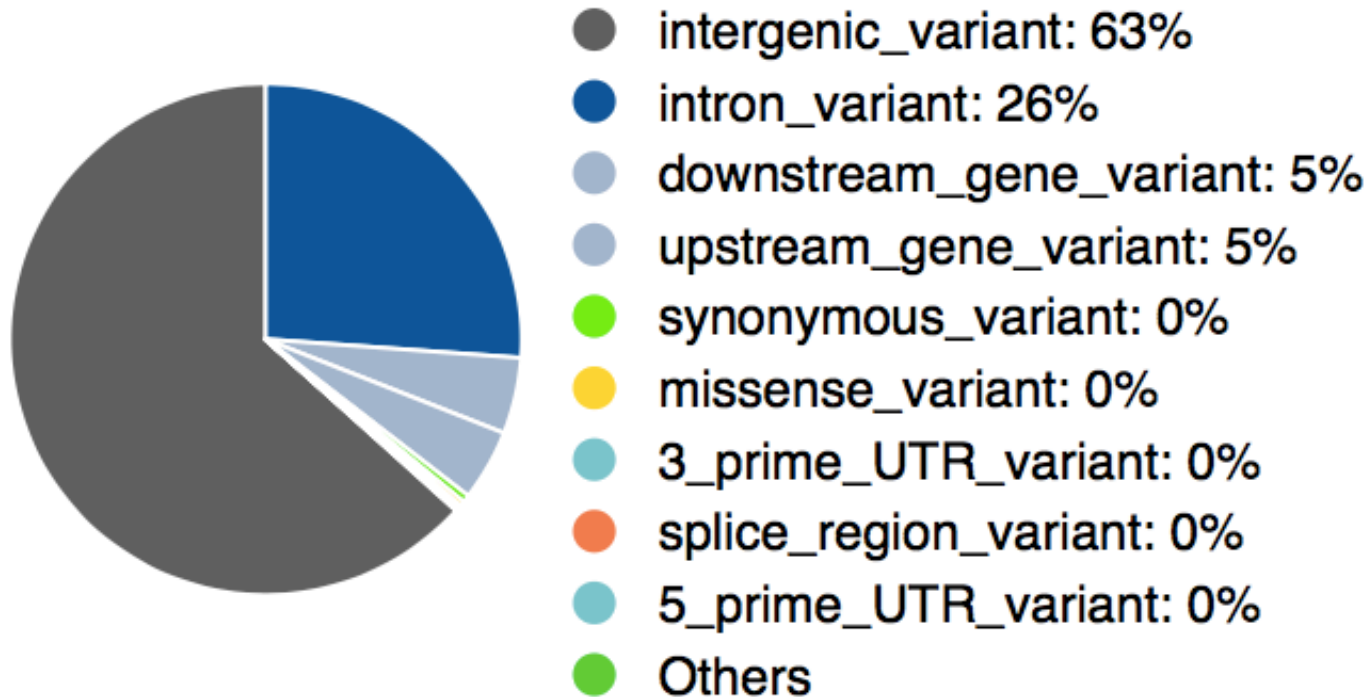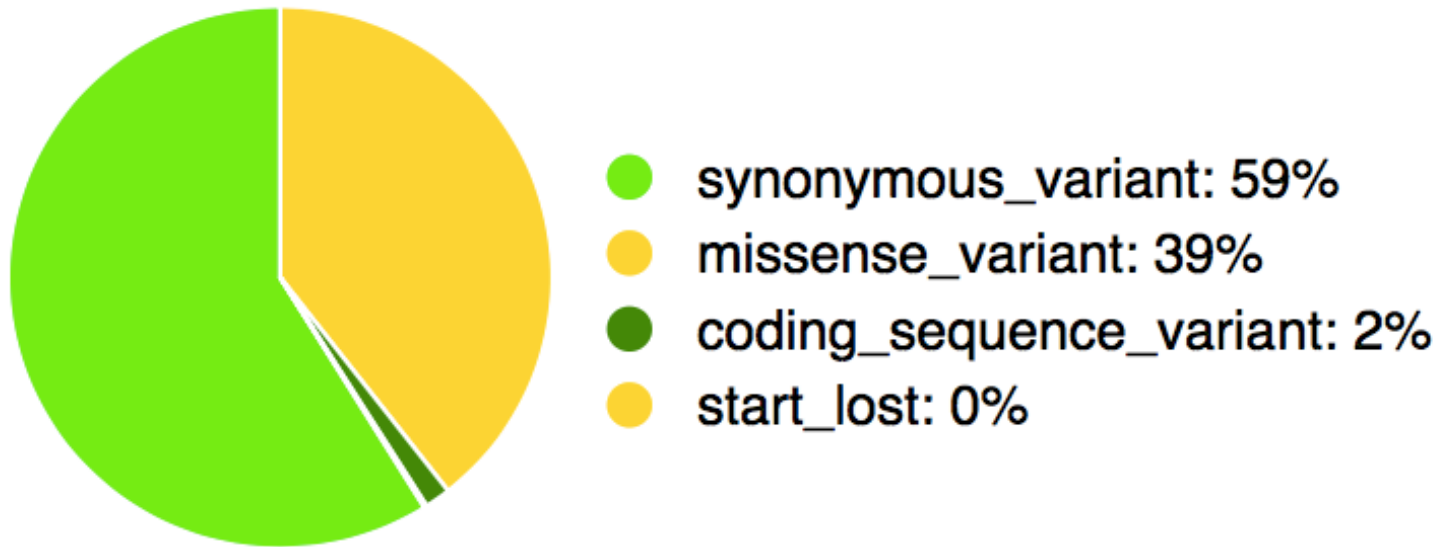
Number of genes = 383

Breed specific mutations – highlighted genes

Brown Swiss

OR1L1 (start lost)

Guernsey

IGFN1 (stop gained)

Simmental

MED25 (splice acceptor variant & missense variant)

Fleckvieh

Not identified

Conclusions

- Chosen variants are potentially characteristic for the breeds

- Created genomes could be the basis for further research on the breed

- Reference genomes may provide the help in breed-tailored GWAS

Thank you for attention

WROCŁAW UNIVERSITY
OF ENVIRONMENTAL
AND LIFE SCIENCES

www.theta.edu.pl