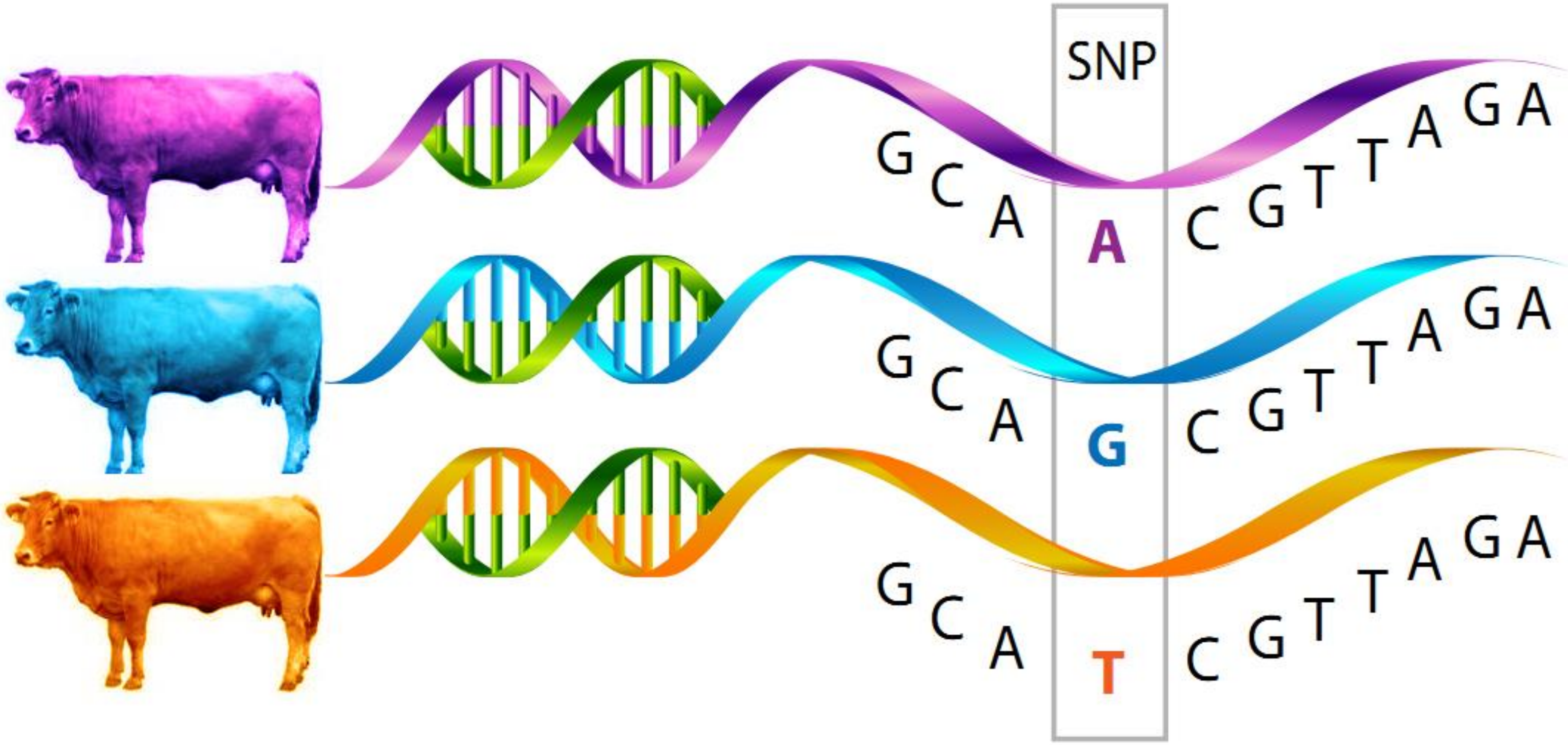
Breed specific reference genomes in cattle

Bartosz Czech[1], Magda Mielczarek[1,2], Magdalena Frąszczak[1], Joanna Szyda[1,2]

[1]Wrocław University of Environmental and Life Sciences, Biostatistics Group
[2]National Research Institute of Animal Production

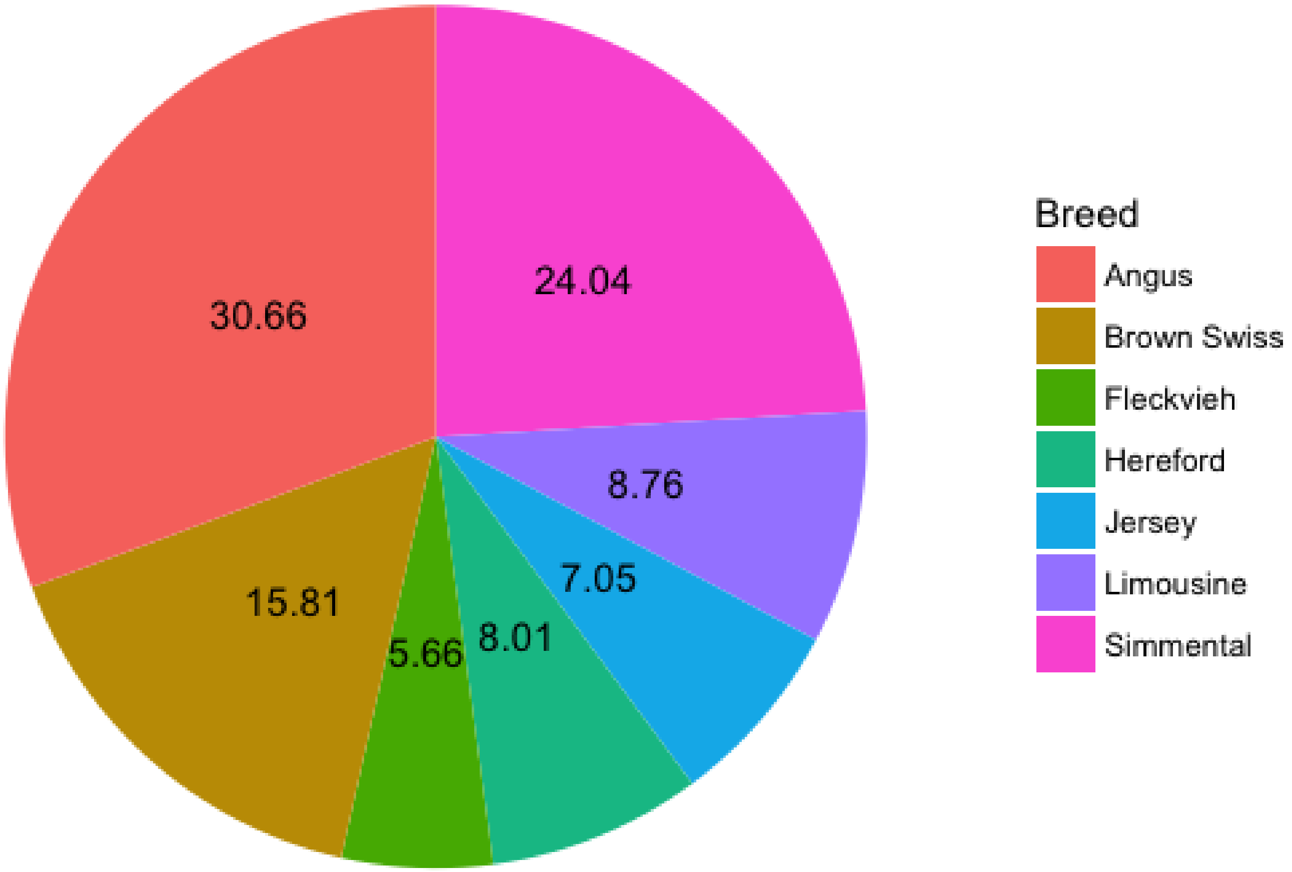
Objectives

- Identification of breed-specific SNPs
- Creating breed-specific reference genomes
- Annotation of breed-specific SNPs



Material

- Whole genome DNA-seq – 936 individuals
- Reference genome UMD 3.1

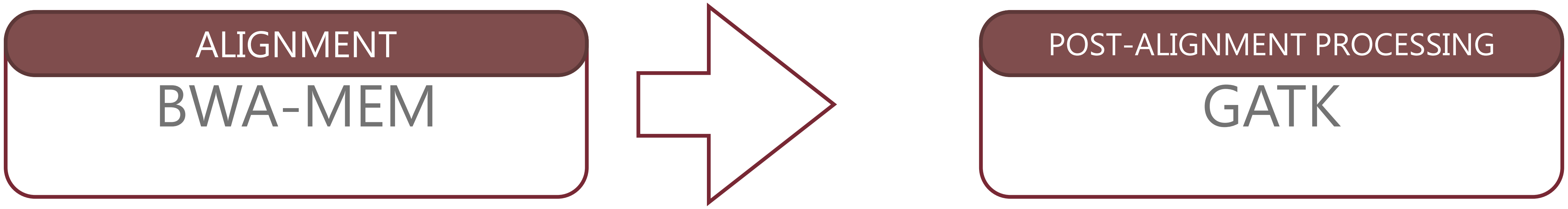


Methods

ALIGNMENT

BWA-MEM

Methods



ALIGNMENT

BWA-MEM

POST-ALIGNMENT PROCESSING

GATK

Methods

ALIGNMENT

BWA-MEM

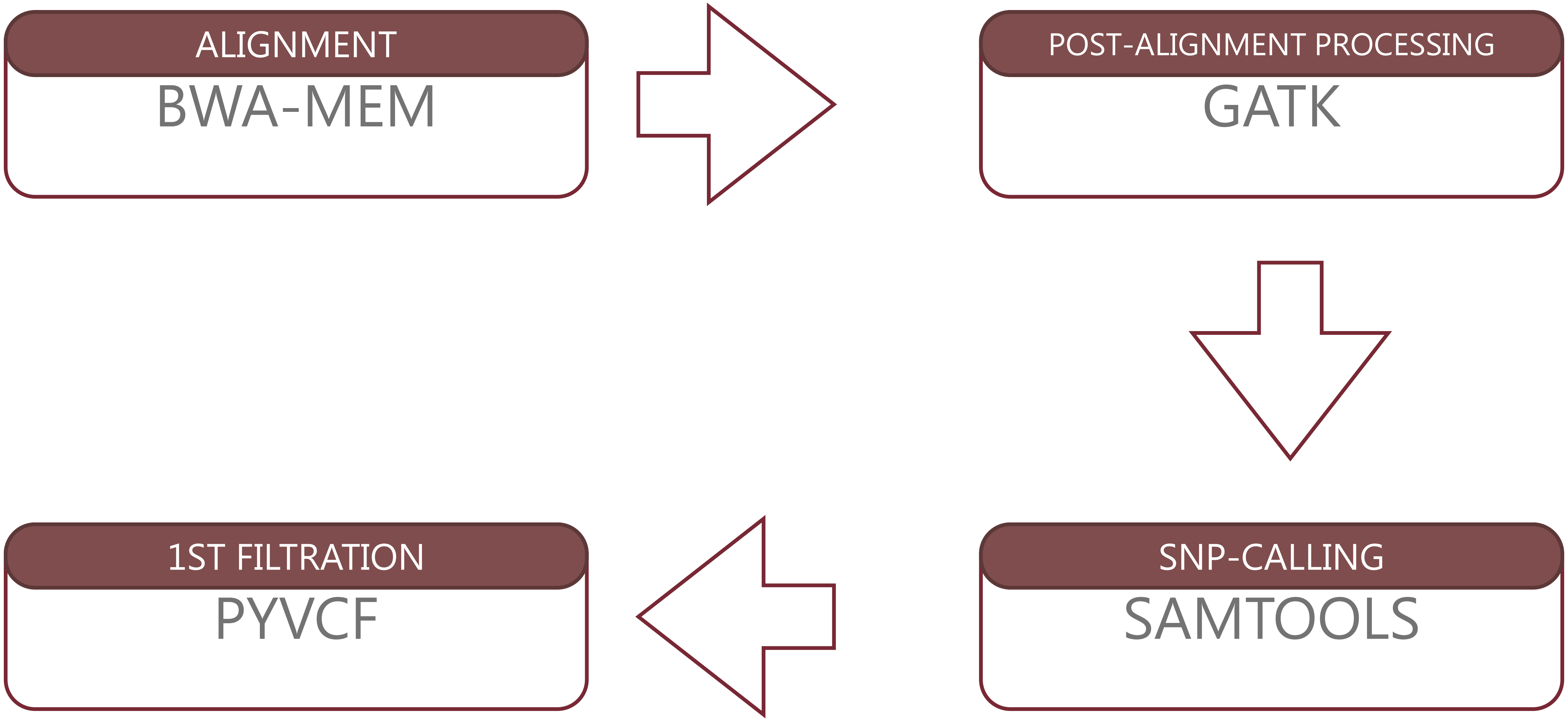
POST-ALIGNMENT PROCESSING

GATK

SNP-CALLING

SAMTOOLS

Methods



ALIGNMENT

BWA-MEM

POST-ALIGNMENT PROCESSING

GATK

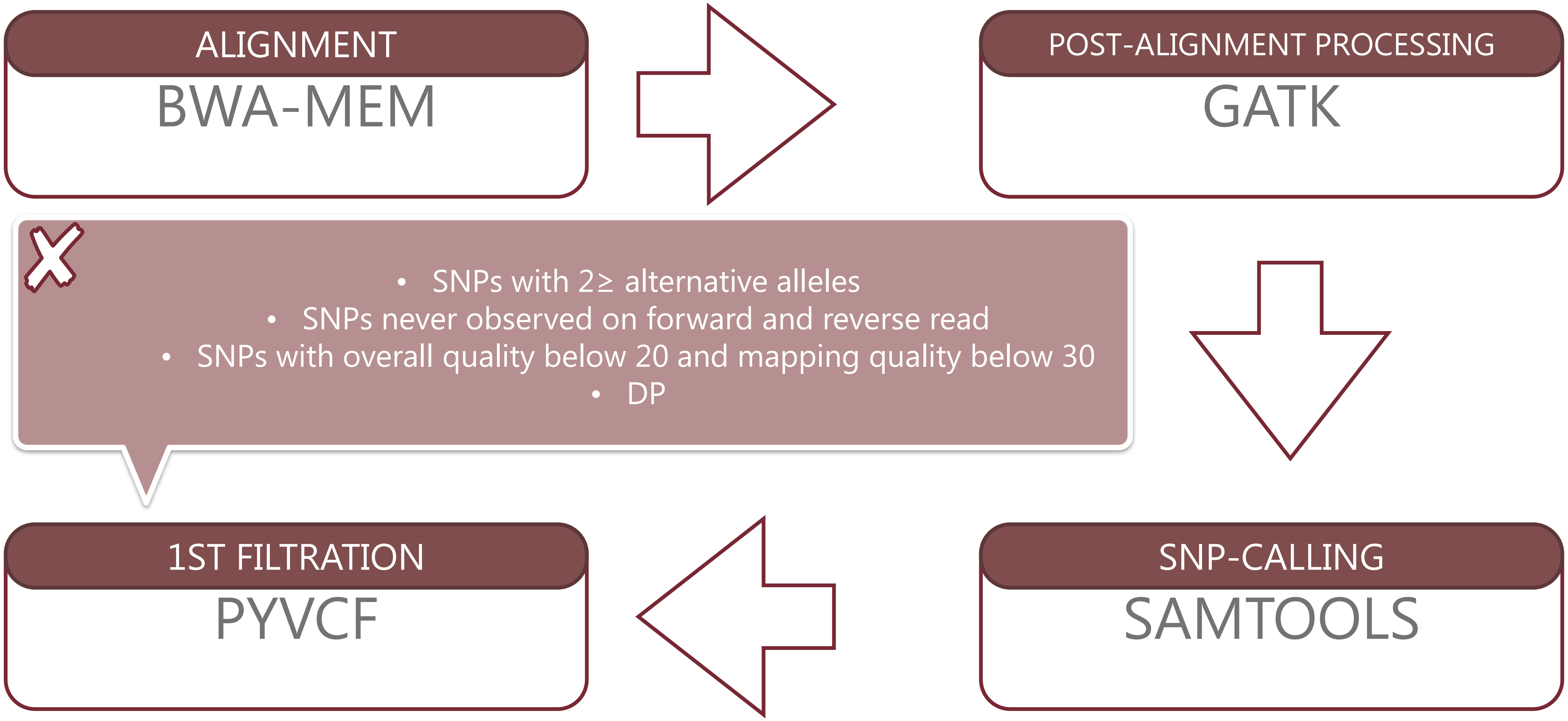
SNP-CALLING

SAMTOOLS

1ST FILTRATION

PYVCF

Methods



ALIGNMENT

BWA-MEM

POST-ALIGNMENT PROCESSING

GATK

SNP-CALLING

SAMTOOLS

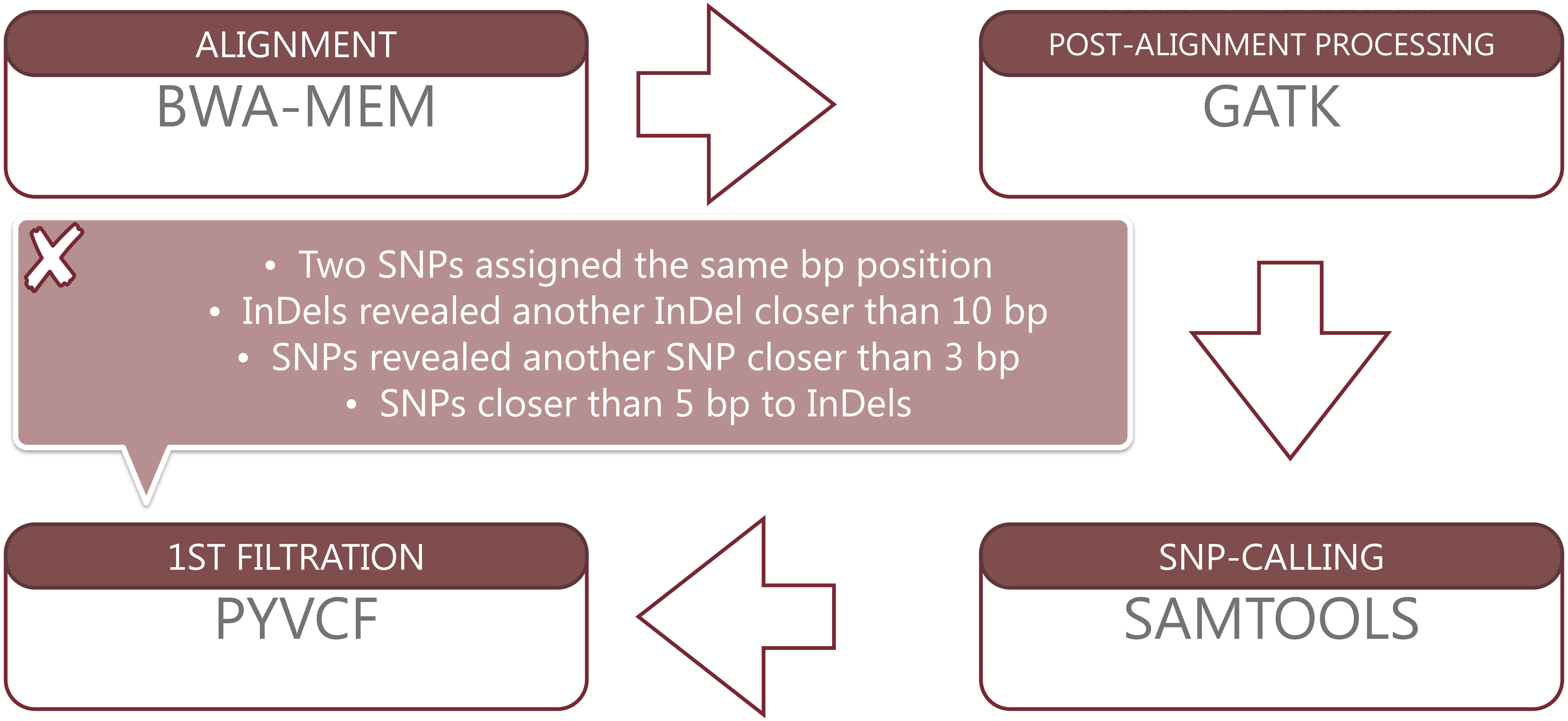
1ST FILTRATION

PYVCF

✗

- SNPs with 2≥ alternative alleles
- SNPs never observed on forward and reverse read
- SNPs with overall quality below 20 and mapping quality below 30
- DP

Methods



ALIGNMENT
BWA-MEM

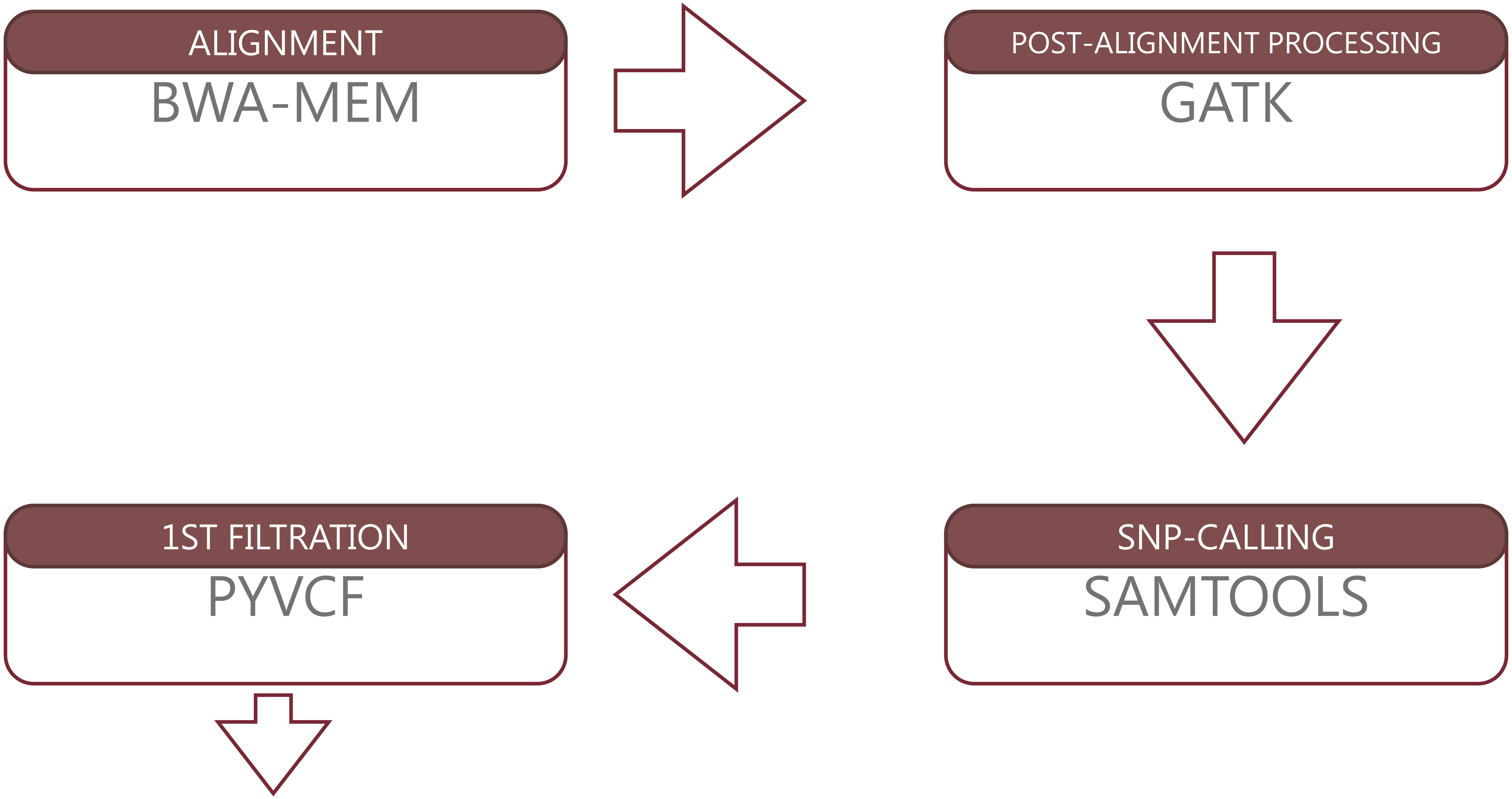
POST-ALIGNMENT PROCESSING
GATK

SNP-CALLING
SAMTOOLS

1ST FILTRATION
PYVCF

- Two SNPs assigned the same bp position
- InDels revealed another InDel closer than 10 bp
- SNPs revealed another SNP closer than 3 bp
- SNPs closer than 5 bp to InDels

Methods



ALIGNMENT

BWA-MEM

POST-ALIGNMENT PROCESSING

GATK

SNP-CALLING

SAMTOOLS

1ST FILTRATION

PYVCF

Methods

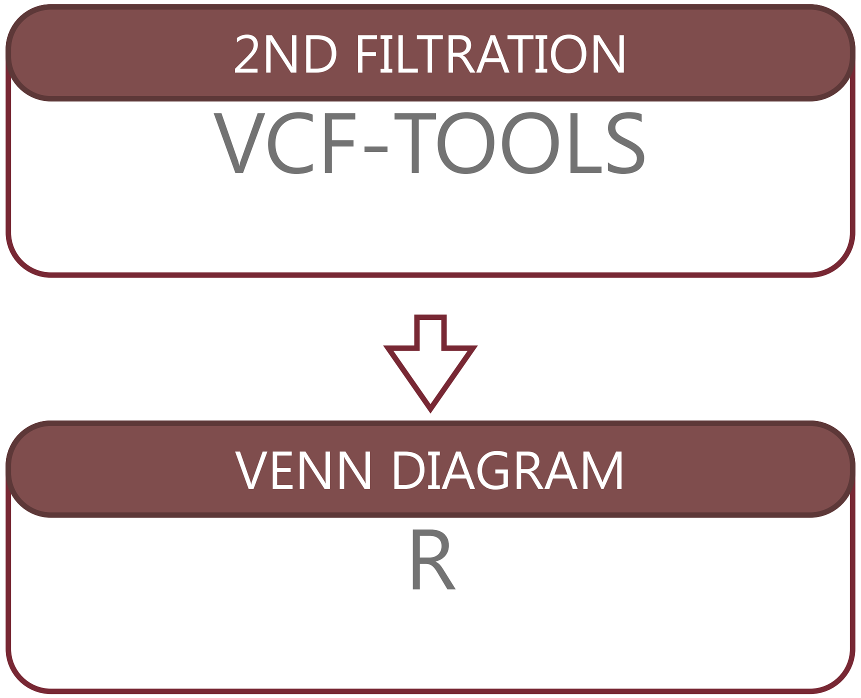
✗

- AF ≠ 1
- Missing genotypes >7%

2ND FILTRATION

VCF-TOOLS

Methods



2ND FILTRATION

VCF-TOOLS

VENN DIAGRAM

R

Methods

2ND FILTRATION

VCF-TOOLS

VENN DIAGRAM

R

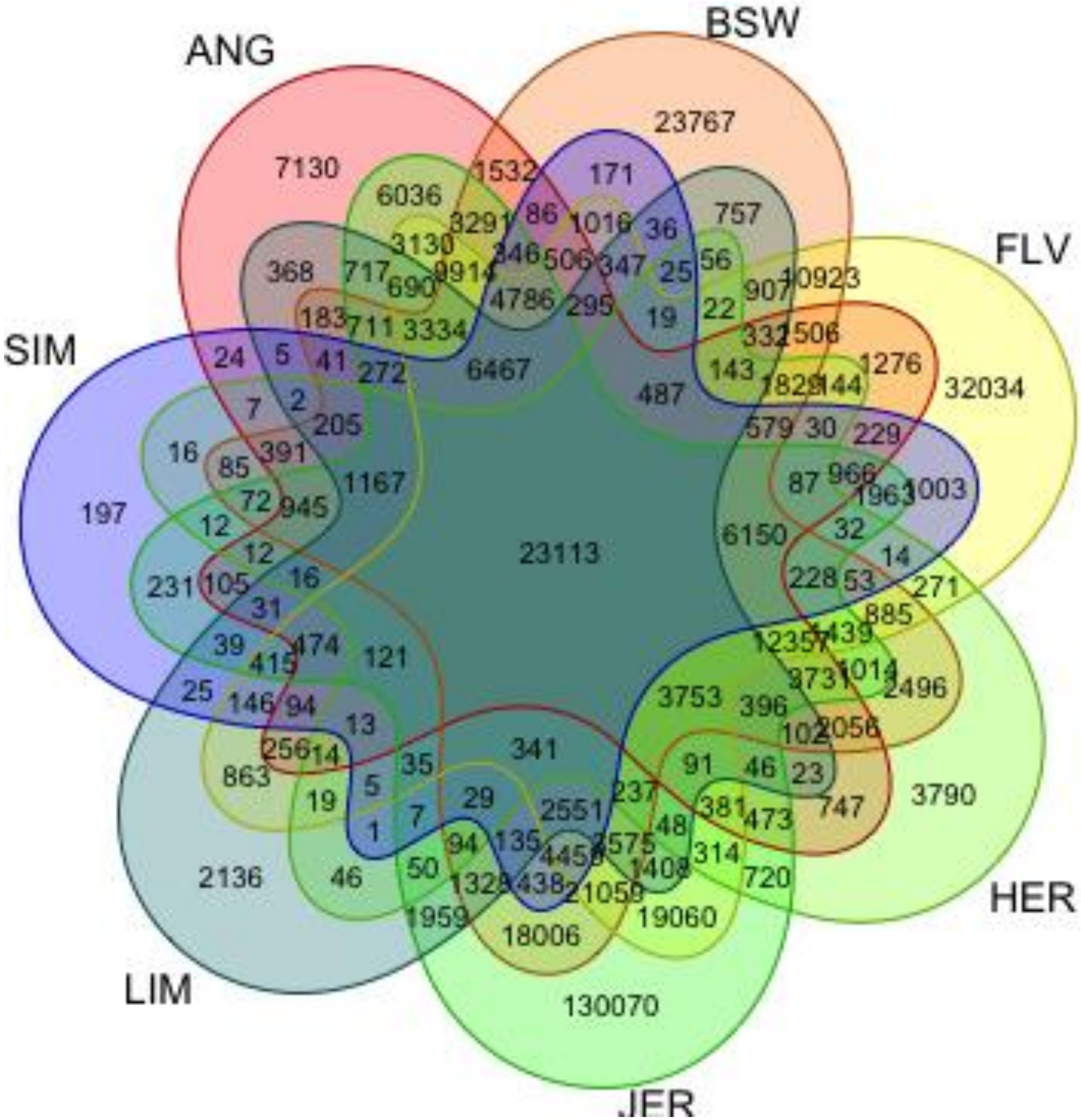
ANNOTATION

VEP, KOBAS, BioMart, AnimalQTLdb

Results

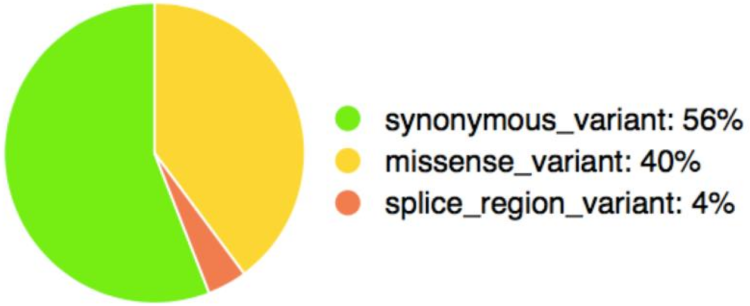
Breed	Total number of SNPs	Number of breed-specific SNPs	% of breed-specific SNPs
Angus	61 823 312	117 744	0.186
Brown Swiss	61 814 873	187 576	0.303
Fleckvieh	61 780 912	189 238	0.306
Hereford	61 813 028	73 772	0.119
Jersey	61 797 641	305 477	0.494
Limousin	61 820 287	61 623	0.100
Simmental	61 824 209	60 582	0.098

Results

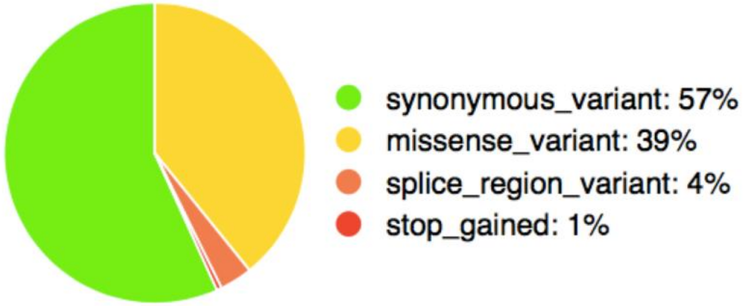

ANG
BSW
FLV
HER
JER
LIM
SIM
7130
23767
32034
3790
130070
2136
197
23113
1532
171
6036
3291
86
1016
36
757
3130
346
506
347
25
56
368
717
690
9914
4786
295
19
22
907
10923
183
711
3334
332
1506
24
5
41
272
6467
487
143
1829
144
1276
7
2
205
579
30
229
16
85
391
1167
87
966
1963
1003
72
945
12
6150
32
12
14
231
105
16
228
53
271
31
885
12357
1439
39
474
121
3731
1014
415
2496
25
146
94
3753
396
13
102
2056
341
256
14
863
35
91
46
23
19
5
29
2551
237
381
473
747
1
7
48
94
135
4458
2575
314
50
1328
438
21059
1408
720
46
1959
18006
19060

Results

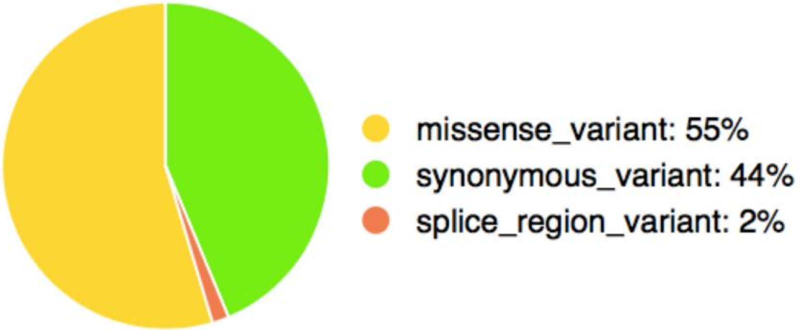
ANGUS



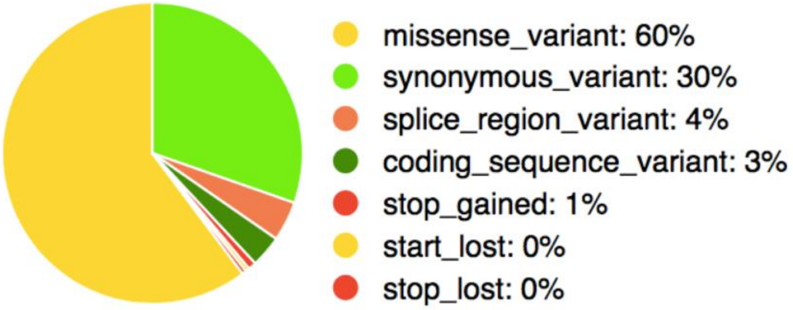
BROWN SWISS



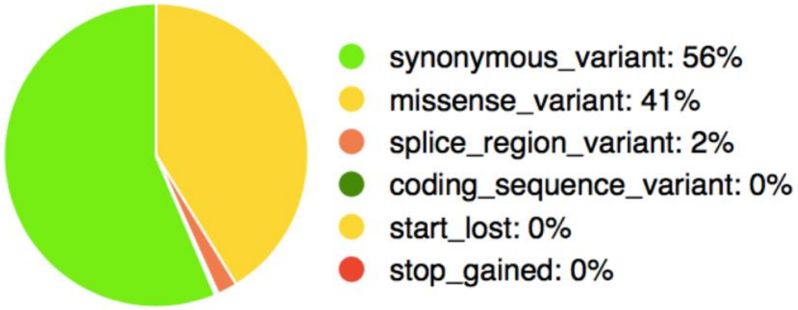
FLECKVIEH



HEREFORD



LIMOUSINE



SIMMENTAL

Results

biomart

ANGUS

- *SLC39A4* (BTA14) – *Acrodermatitis enteropathica*

BROWN SWISS

- *SLC4A2* (BTA4) – *Osteopetrosis*
- *LAMA3* (BTA24) – *Epidermolysis bullosa*

JERSEY

- *GON4L* (BTA3) – dwarfism
- *APOB* (BTA11) – cholesterol deficiency
- *TG* (BTA14) – familial goitre
- *COL7A1* (BTA22) - *Epidermolysis bullosa*

LIMOUSINE

- *PFAS* (BTA19) – abortion

Results

AnimalQTLdb

ANGUS

- iron content in mussels

BROWN SWISS

- milk characteristics: milk solids, overall proteins, alpha-casein, alpha-lactalbumin, kappa-casein, beta-casein percentages, milk lactose content and yield
- somatic cell score
- growth traits: body weight, longissimus muscle area and gastrointestinal nematode burden

Conclusions

- Differences observed between breeds express the dynamics of changes in the *Bos taurus* genome

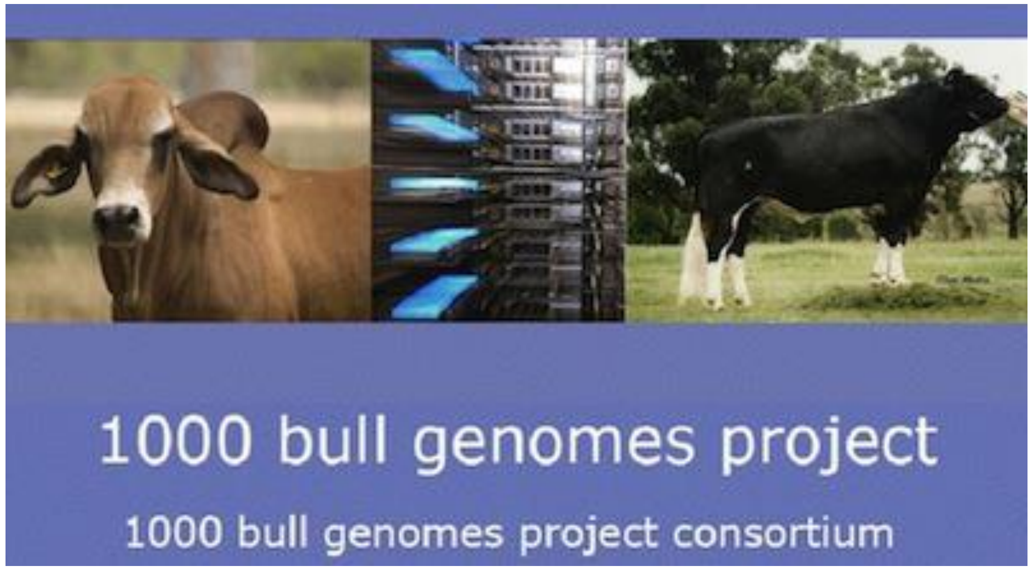
- Breed specific reference genomes increase the accuracy of GWAS and SNP genotypes imputation

Acknowledgement

NATIONAL SCIENCE CENTRE
POLAND

KNOW | The Leading National Research Centre

Wroclaw Centre of Biotechnology 2014-2018



1000 bull genomes project

1000 bull genomes project consortium

PSNC

Thank you for your attention!

Czech B, Frąszczak M, Mielczarek M, Szyda J (2018) Identification and annotation of breed-specific single nucleotide polymorphisms in *Bos taurus* genomes. PLoS ONE 13(6): e0198419.
https://doi.org/10.1371/journal.pone.0198419