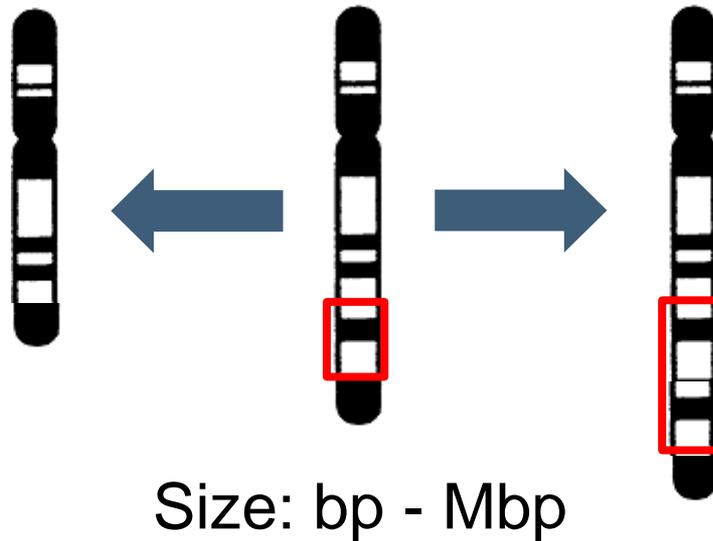


Objectives

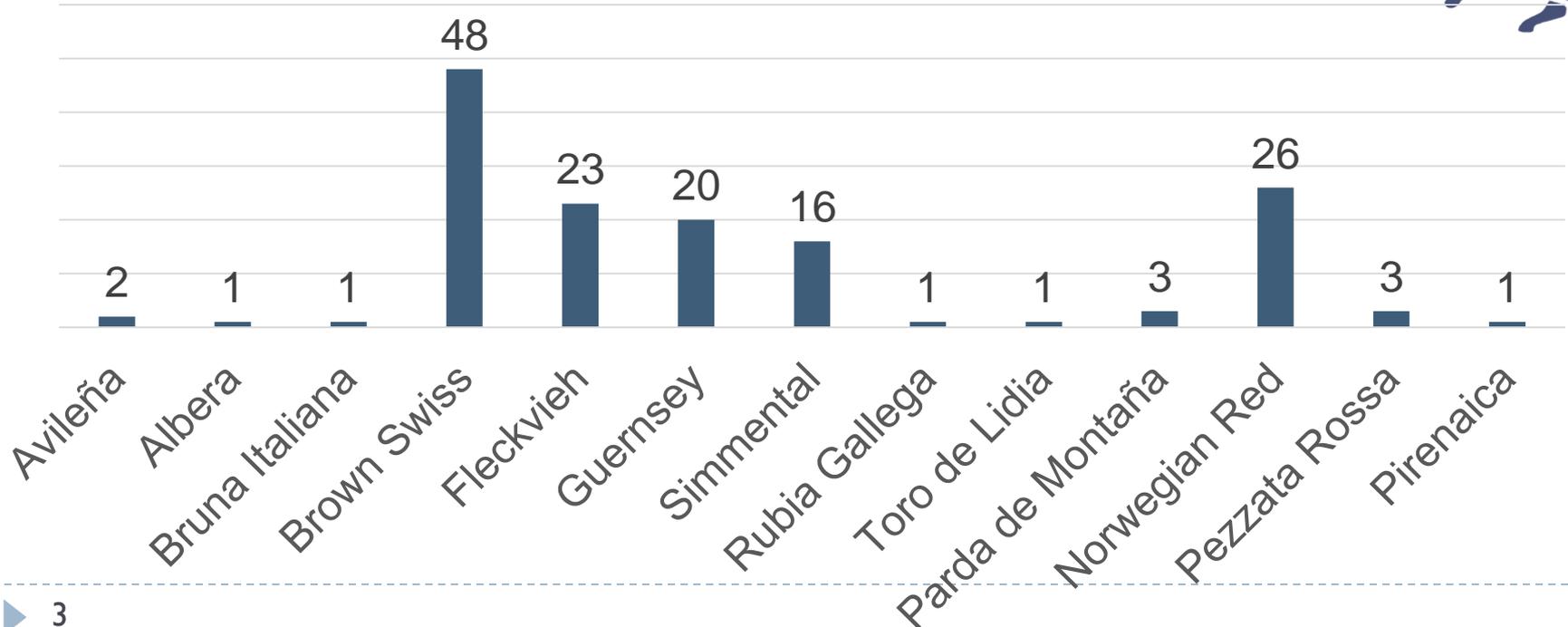
- ▶ detecting of CNV polymorphisms



- ▶ describing the **inter-individual** and **inter-breed** variation

Dataset

- ▶ Whole genome DNaseq – **146** bulls
- ▶ Illumina HiSeq next-generation sequencing platform
- ▶ 7-28 x genome average coverage
- ▶ 13 breeds / 5 breeds



Methods

Alignment to the reference genome

BWA

Post-alignment processes

Picard,
SAMtools

CNV detection

CNVnator

CNV detection

Pindel

Consensus CNV set constructing

CNV annotation

VEP

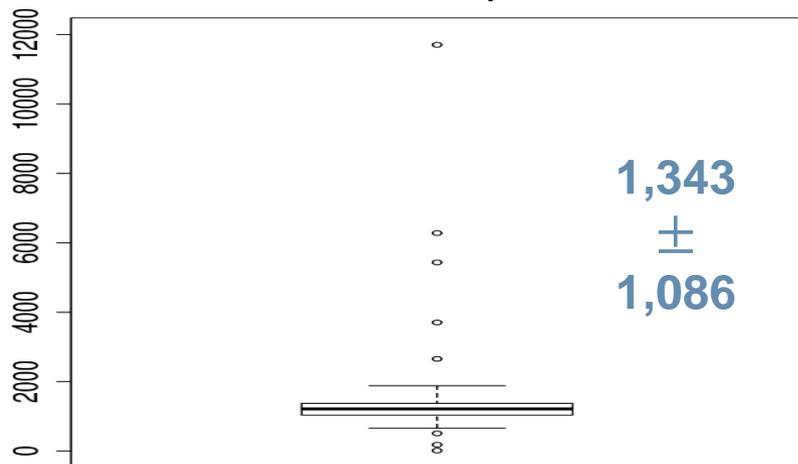
Statistical analysis



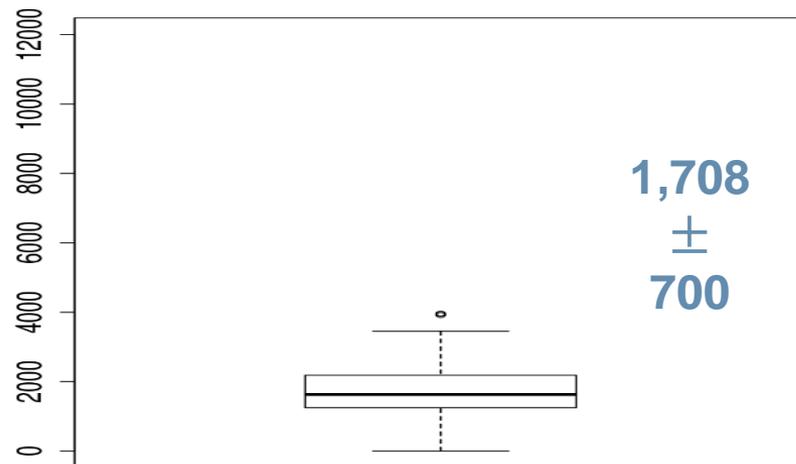
Results

Total **number**/**length** of dup/del called

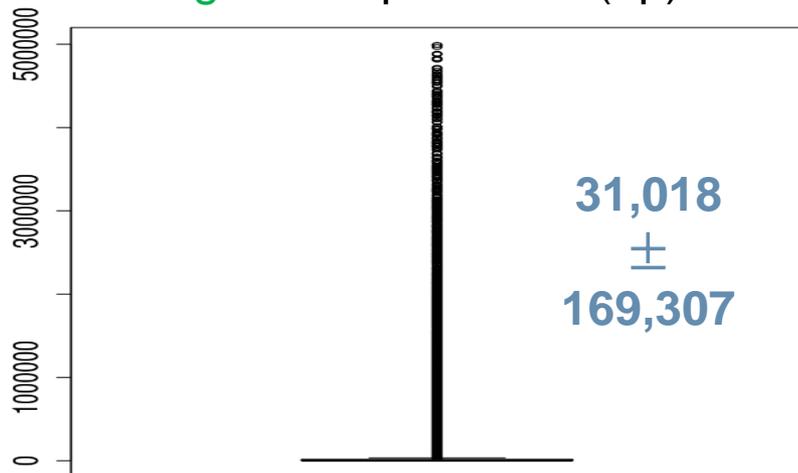
number of duplications



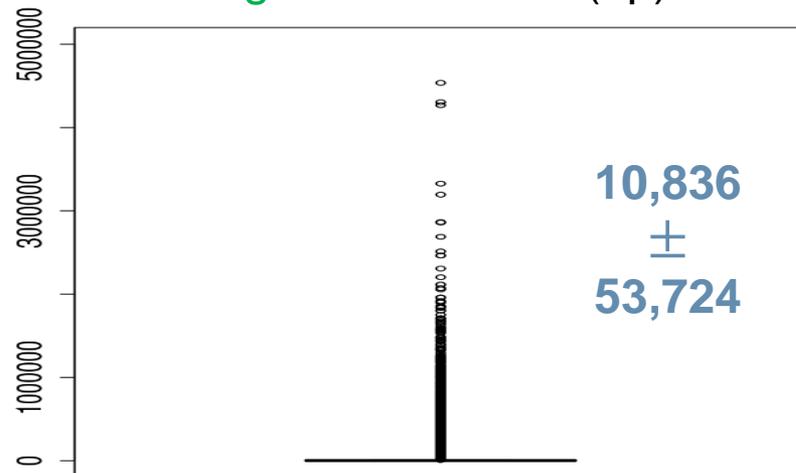
number of deletions



length of duplications (bp)



length of deletions (bp)

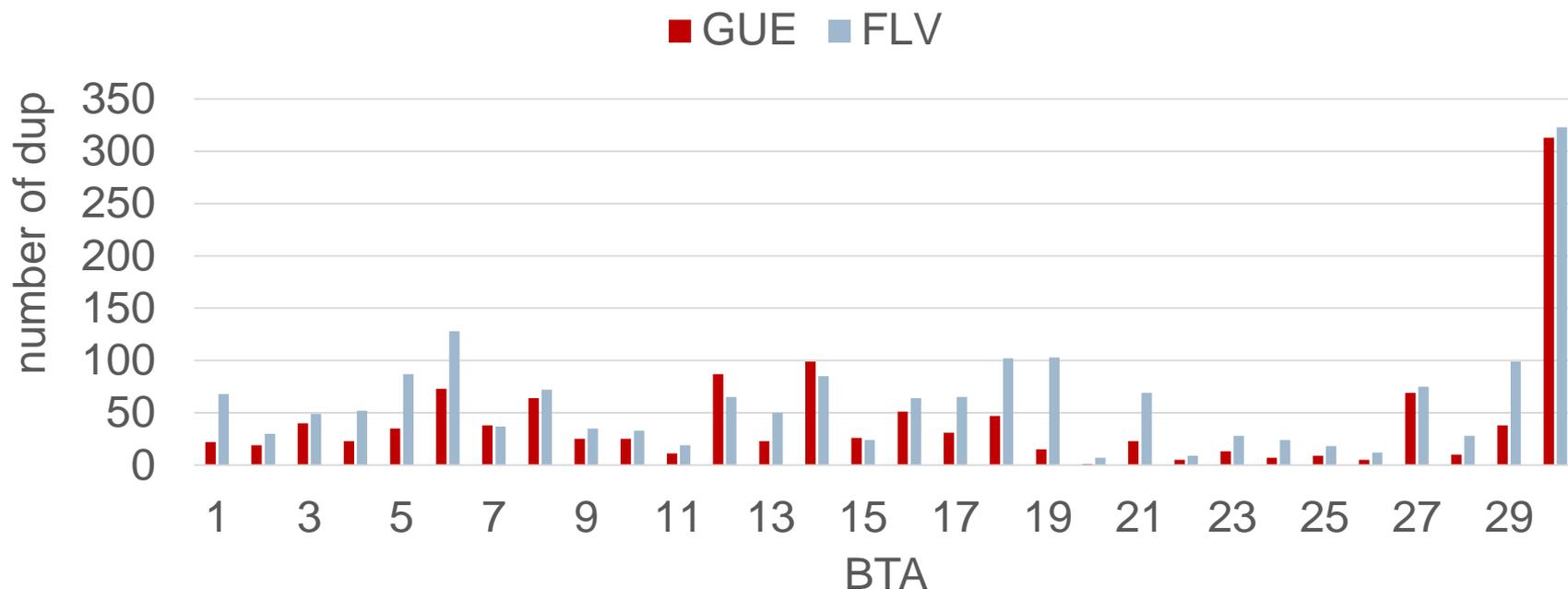


Results

Number of dup/del

- ▶ not uniformly distributed across 146 individuals

dup: $P < 2.20 \cdot 10^{-16}$, del: $P < 2.20 \cdot 10^{-16}$



Results

Number of dup/del

- ▶ not uniformly distributed across 146 individuals
dup: $P < 2.20 \cdot 10^{-16}$, del: $P < 2.20 \cdot 10^{-16}$
- ▶ not uniformly distributed across individuals within each breed
dup: $P < 2.20 \cdot 10^{-16}$, del: $P < 2.20 \cdot 10^{-16}$
(for each breed)
- ▶ variability between breeds exists
dup: $P=0.01932$, del: $P=0.01006$ ($\alpha=0.05$)

Results

Length of dup/del

- ▶ distribution not the same for 146 individuals

$$\text{dup: } P < 2.20 \cdot 10^{-16}, \text{ del: } P < 2.20 \cdot 10^{-16}$$

- ▶ distribution not the same across individuals within each breed

$$\text{dup: } P \in (3.02 \cdot 10^{-94}, 0.1 \cdot 10^{-12})$$

$$\text{del: } P \in (1.23 \cdot 10^{-192}, 0.1 \cdot 10^{-12})$$

- ▶ variability between breeds exists

$$\text{dup: } P < 2.20 \cdot 10^{-16}, \text{ del: } P < 2.20 \cdot 10^{-16}$$

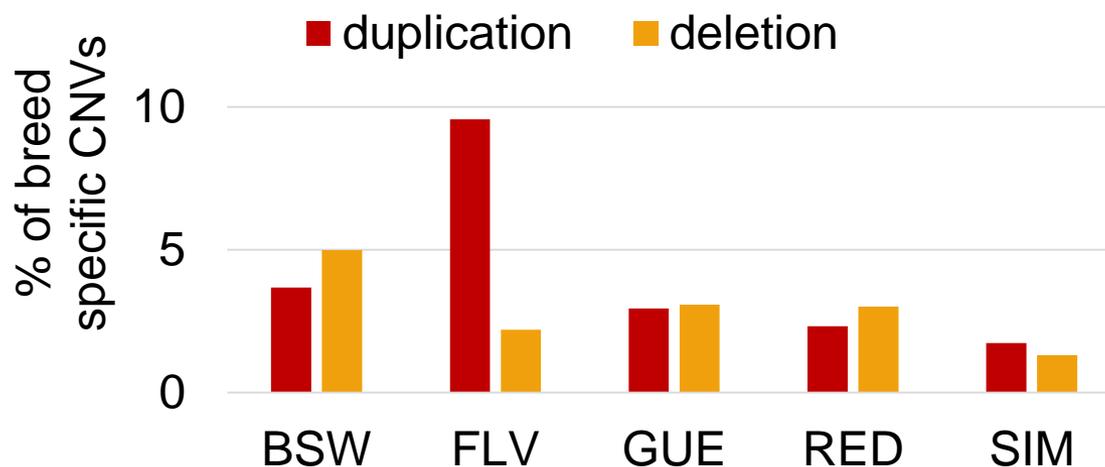
Results

Unique, shared and breed specific duplications

▶ **Unique** (1/146): dup 84.85 %, del 77.22 %

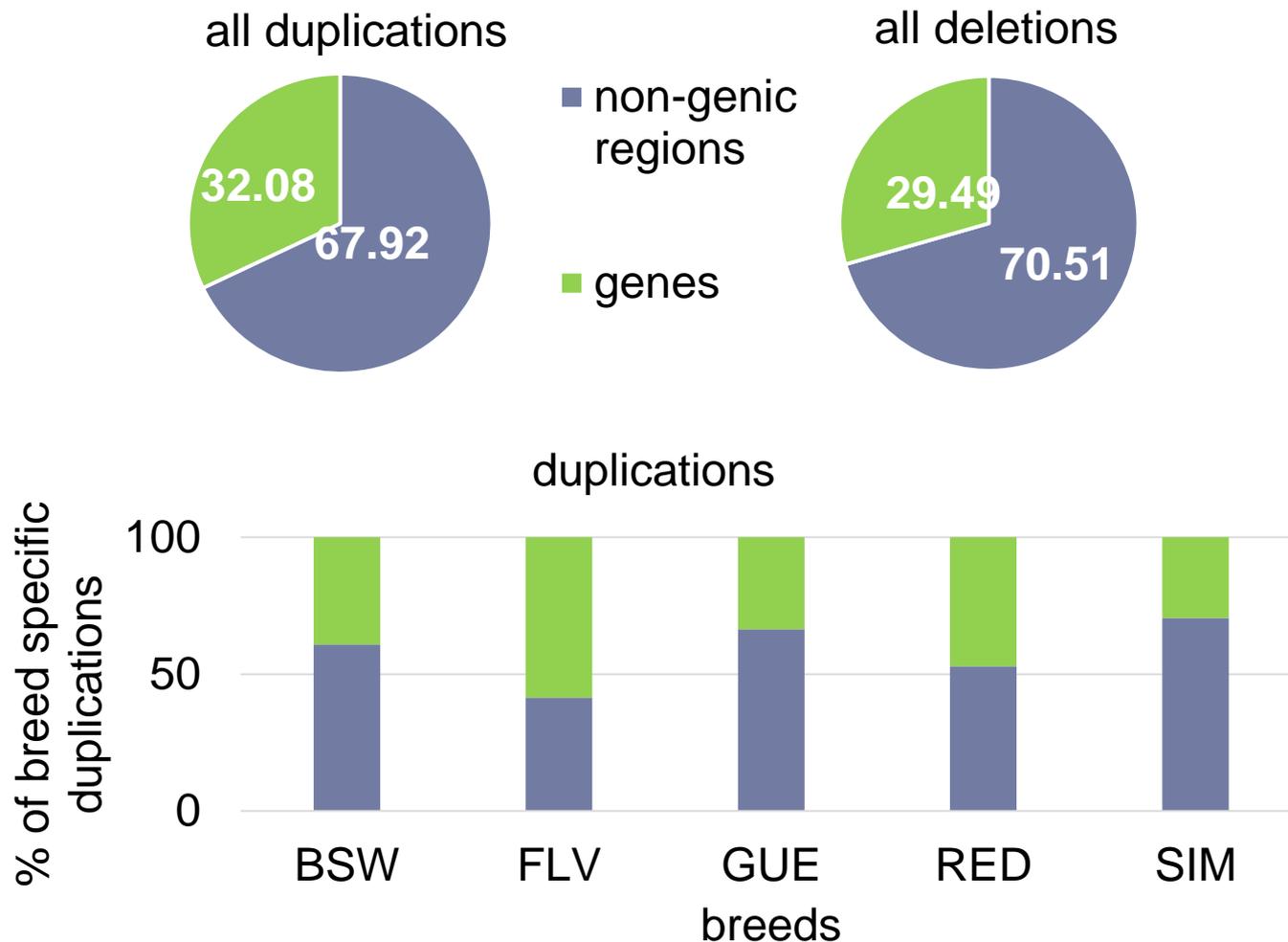
▶ Max **shared**: dup 117/146, del - 140/146

▶ **Breed specific**: min $2/N_{\text{breed}}$



Results

Functional CNV annotation

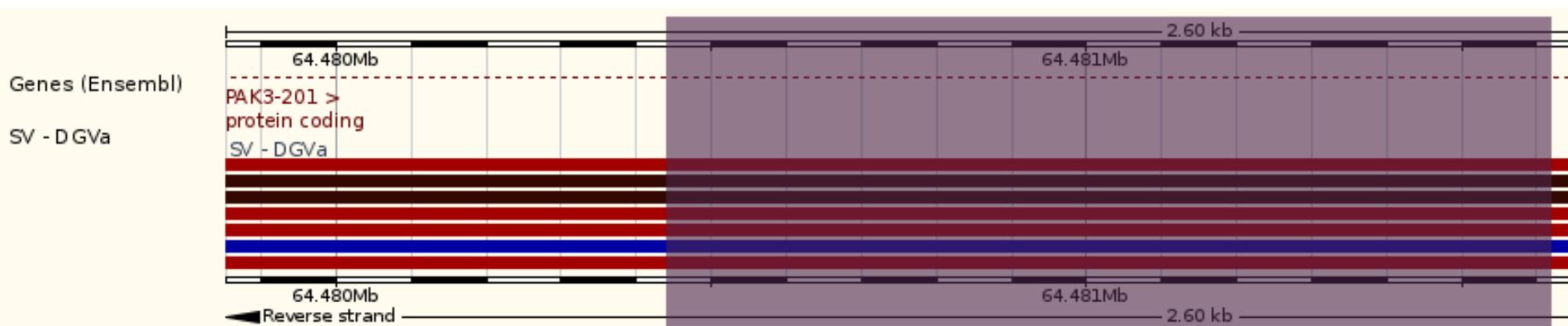


Results

20 most common dup (117-74 bulls) and del (140-117 bulls)

	intergenic	intron	protein coding	non-protein coding
dup	13	4	1	2
del	19	1	-	-

Common variants overlapp with DGVa



- gain (Bickhart *et al.* 2012)
- tandem duplication (Boussaha *et al.* 2015)
- deletion (Liu *et al.* 2010, Boussaha *et al.* 2015)

Conclusions

- ▶ high complexity of a CNV landscape in *Bos taurus* genomes
- ▶ varying density of CNV depends on genome function
- ▶ the breed-specific phenomenon in the Fleckvieh
- ▶ deletion events in coding regions less evolutionary accepted than duplications

Acknowledgements

National Science Centre in Poland:

Describing the genome-wide distribution of copy number variations in various breeds of domestic cattle based on the next-generation sequencing data of 121 individuals

National Science Centre in Poland:

Resolving genetic predisposition to clinical mastitis based on whole genome sequences of 32 cows

European Union's Seventh Framework Programme

Poznan Supercomputing and Networking Center



Conclusions

- ▶ high complexity of a CNV landscape in *Bos taurus* genomes
- ▶ varying density of CNV depending on genome function
- ▶ the breed-specific phenomenon in the Fleckvieh
- ▶ deletion events in coding regions less evolutionary accepted than duplications



Thank you!



▶ Additional slides

Additional slides

number of CNVs

χ^2 test of goodness of fit

- ▶ H_0 : the number of dup/del uniformly distributed across 146 individuals

$$\text{dup } P < 2.20 \cdot 10^{-16}, \text{ del } P < 2.20 \cdot 10^{-16}$$

$$\chi^2 = \sum_{i=1}^m \frac{(O_i - E)^2}{E},$$

where O_i denoted the number of duplications/deletions for i -th bull, E was the average number of identified deletions/duplications in the whole dataset and m denoted the number of bulls. The test statistic is χ^2 -distributed with $m-1$ degrees of freedom.

Additional slides number of CNVs

χ^2 test of goodness of fit

- ▶ H_0 : the number of dup/del uniformly distributed across individuals within each breed

$$\text{dup } P < 2.20 \cdot 10^{-16}, \text{ del } P < 2.20 \cdot 10^{-16}$$

for each breed

$$\chi^2 = \sum_{i=1}^m \frac{(O_i - E)^2}{E},$$

where O_i denoted the number of duplications/deletions for i -th bull, E was the average number of identified deletions/duplications for the breed and m denoted the number of bulls. The test statistic is χ^2 -distributed with $m-1$ degrees of freedom.

Additional slides

number of CNVs

Kruskal-Wallis test

- ▶ H_0 : the variability in the number of dup/del between breeds does not exist

dup $P=0.01932$, del $P=0.01006$ ($\alpha=0.05$)

$$H = \frac{12}{k(k+1)} \sum_{i=1}^m \frac{R_i^2}{k_i} - 3(m+1)$$

where k_i was the number of bulls belonging to i -th breed, and $k = \sum_{i=1}^n k_i$, m was the number of breeds and R_i denoted the sum of ranks for deletion/duplication number corresponding to i -th breed. The test statistic is approximately χ^2 - distributed with $k-1$ degrees of freedom

- ▶ H_0 : lengths of dup/del are normally distributed

$$W = \frac{[\sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} a_i(n)(X_{(n-i+1):n} - X_{i:n})]^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

where a_i were constant from Shapiro-Wilk tables, n denoted the number of CNVs, $X_{i:n}$ was the length of i -th variant sorted by length. The null hypothesis is rejected when the value of the test statistic is lower than the quantile of Shapiro-Wilk distribution

Additional slides length of CNVs

Kruskal-Wallis test

- ▶ H_0 : the distribution of dup/del lengths is the same for 146 individuals

$$\text{dup } P < 2.20 \cdot 10^{-16}, \text{ del } P < 2.20 \cdot 10^{-16}$$

$$H = \frac{12}{k(k+1)} \sum_{i=1}^m \frac{R_i^2}{k_i} - 3(m+1)$$

where k_i was the number of duplications or deletions for i -th bull, and $k = \sum_{i=1}^n k_i$, m was the number of bulls and R_i denoted the sum of ranks for deletion/duplication length corresponding to i -th bull. The test statistic is approximately χ^2 - distributed with $k-1$ degrees of freedom

Additional slides

length of CNVs

Kruskal-Wallis test

- ▶ H_0 : the distribution of dup/del lengths is the same across individuals within each breed

$$\text{dup } P = 3.02 \cdot 10^{-94} - 0.1 \cdot 10^{-12}$$

$$\text{del } P = 1.23 \cdot 10^{-192} - 0.1 \cdot 10^{-12}$$

(min and max P-values for particular breeds)

$$H = \frac{12}{k(k+1)} \sum_{i=1}^m \frac{R_i^2}{k_i} - 3(m+1)$$

where k_i was the number of duplications or deletions for i -th bull, and $k = \sum_{i=1}^n k_i$, m was the number of bulls and R_i denoted the sum of ranks for deletion/duplication length corresponding to i -th bull. The test statistic is approximately χ^2 -

- ▶ 21 distributed with $k-1$ degrees of freedom

Additional slides

length of CNVs

Kruskal-Wallis test

- ▶ H_0 : the variability in the length of dup/del **between breeds does not exist**

$$\text{dup } P < 2.20 \cdot 10^{-16}, \text{ del } P < 2.20 \cdot 10^{-16}$$

$$H = \frac{12}{k(k+1)} \sum_{i=1}^m \frac{R_i^2}{k_i} - 3(m+1)$$

where k_i was the number of bulls belonging to i -th breed, and $k = \sum_{i=1}^n k_i$, m was the number of breeds and R_i denoted the sum of ranks for deletion/duplication length corresponding to i -th breed. The test statistic is approximately χ^2 - distributed with $k-1$ degrees of freedom