Copy number variations associated with insect bite hypersensitivity in Friesian horses

67th Annual Meeting of the European Federation of Animal Science

Anouk Schurink
August 30, 2016

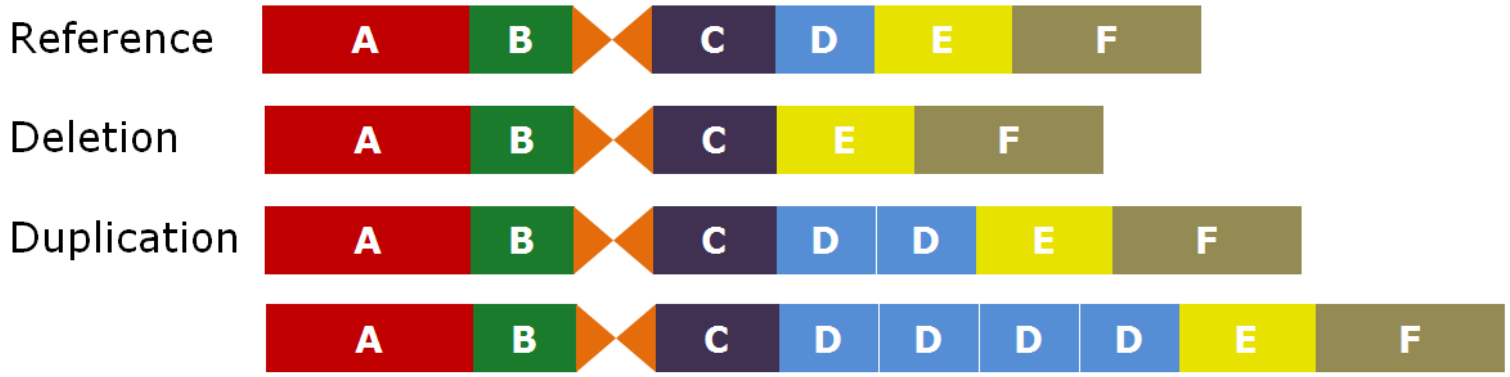
Vinicius da Silva - Bart Ducro - Bert Dibbits - Richard Crooijmans
Brandon Velie - Gabriella Lindgren - Sarah Blott - Nadine Buys
Liesbeth François - Anneleen Stinckens

WAGENINGEN UR
For quality of life

KFPS • ROYAL FRIESIAN
KONINKLIJK FRIESCH PAARDEN-STAMBOEK

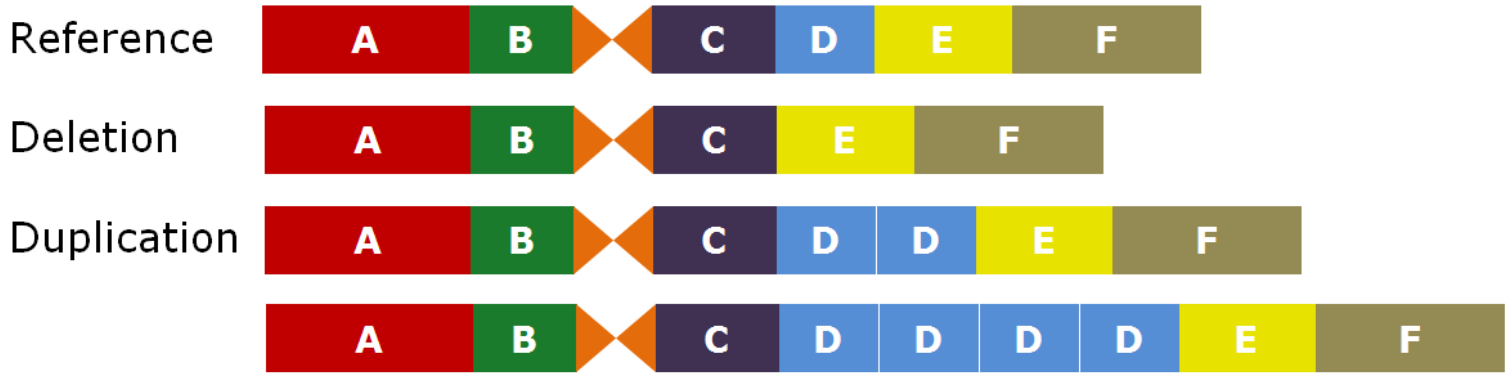
Horsegene

Introduction – copy number variation (CNV)



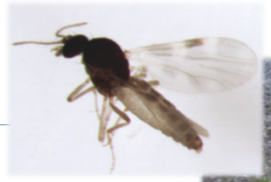
- Change in number of copies
- Genomic region of reasonable size (≥1kb)
- Largest source of genetic variation in many genomes

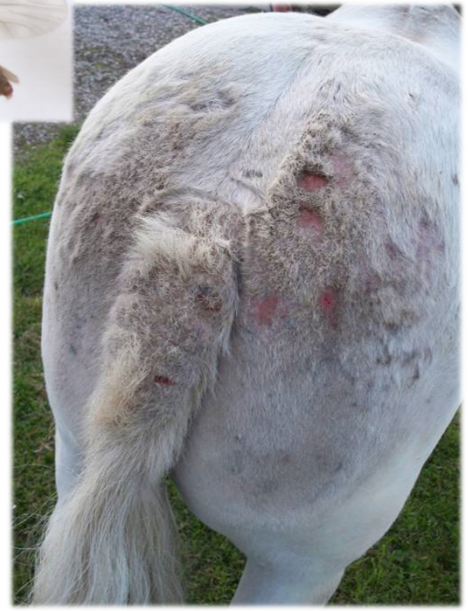
Introduction – copy number variation (CNV)



- Facilitate our understanding of the genome and its expression → associations with our traits of interest?
- In horses: CNV (association) studies limited!

Introduction





Insect bite hypersensitivity (IBH)

- Seasonal allergy to *Culicoides* spp.
- Many breeds affected worldwide
- Intense itch → self-inflicted trauma

Genetics

- Multifactorial and polygenic in nature ($h^2 \sim 0.2$)
- Across-breed associations: MHC region (ECA20)

Introduction

Friesian horses

- Common native breed
- ~20% affected with IBH

Our aims

- Identify CNVs in Friesian horses
- Perform a CNV-based GWAS to identify genomic regions associated with IBH

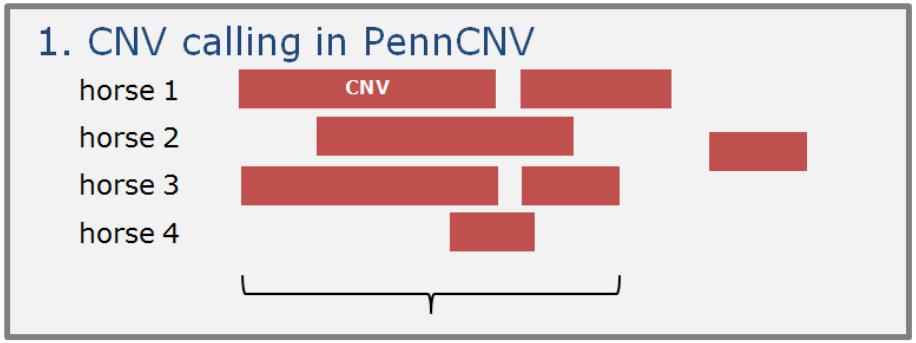
Materials and Methods

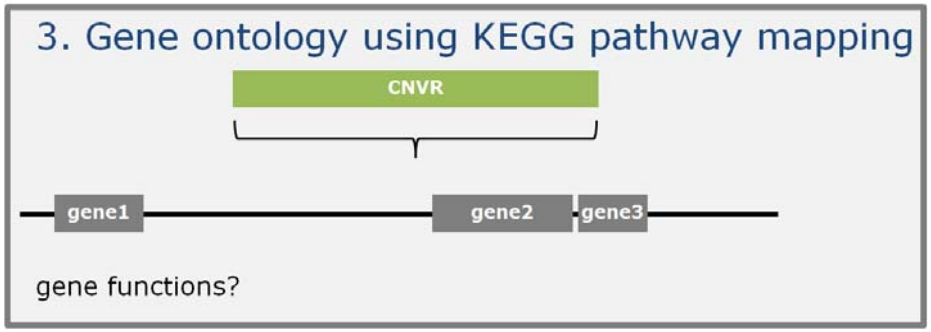
Materials

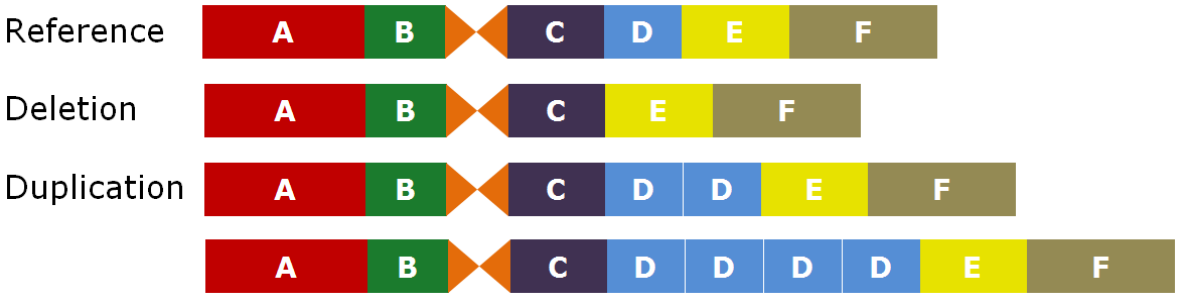
- Case-control approach with strict protocol (n=280)
- Axiom® Equine Genotyping Array (670k)

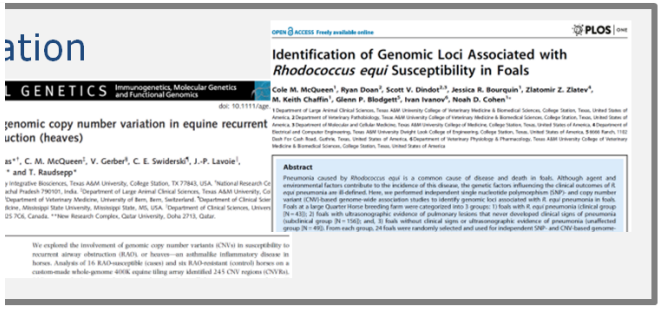
Materials and Methods

Methods



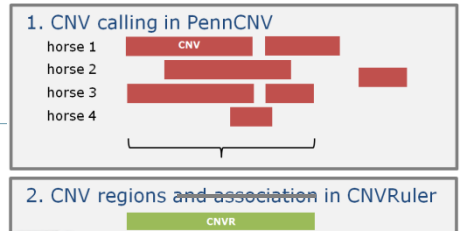






WAGENINGEN UR
For quality of life

CNVs and CNV regions



Identified CNVs

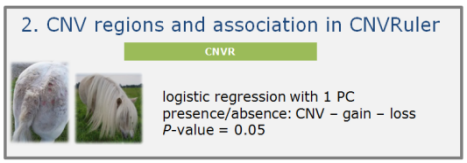
- 15,041 CNVs: 18 to 262 per horse (mean=67.8)

✓ Results partly comparable to previous equine CNV studies

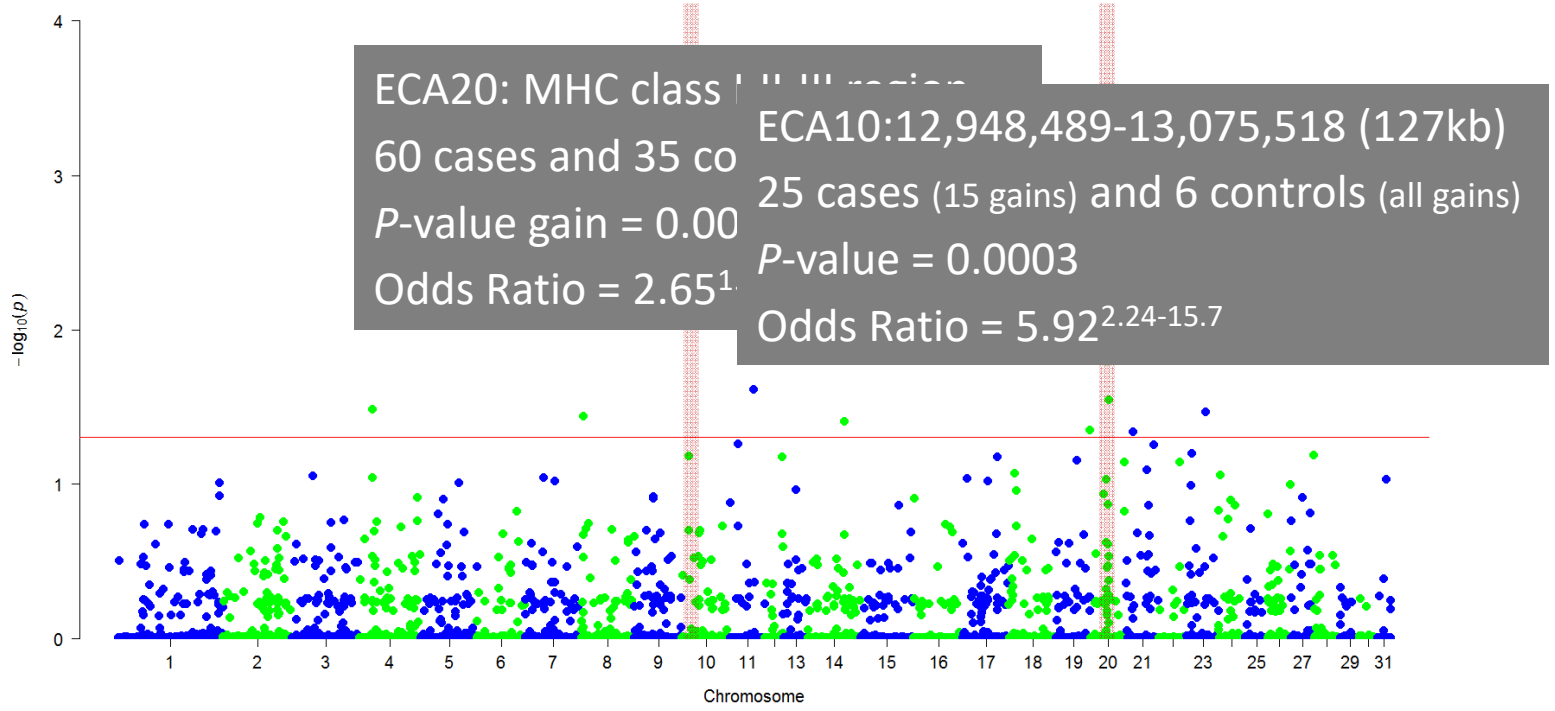
✓ Diversity between studies likely due to experimental set-up

- 5,350 CNV regions
- Genome coverage was 11.2%

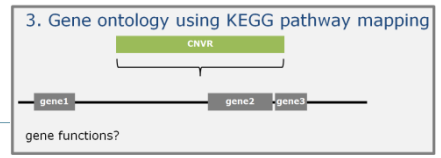
CNV region association



19 CNVRs significantly associated with IBH (*P*-value CNV <0.05)



Gene ontology



GO analysis CNVRs

- 43.7% CNVRs involved genes

✓ Enriched pathways comparable to previous (equine) CNV studies

✓ Candidate genes within the MHC region

- Strong candidate region: MHC
- Enriched for immunity related genes
- No genome-wide significance

WAGENINGEN UR
For quality of life

CNV validation

Based on literature

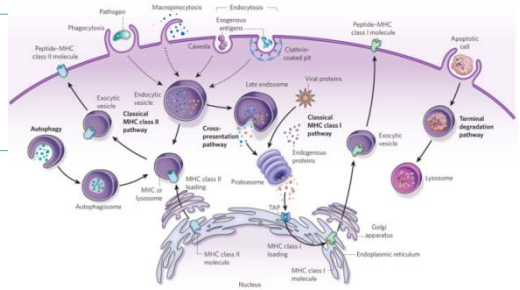
- 42.0% of CNVRs in Friesian horses validated in other breed(s)
- 84.2% of CNVRs associated with IBH validated

✓ Reasonable percentage of CNVRs validated

✓ Breed-specific CNV(R)s are to be expected

WAGENINGEN UR
For quality of life

Discussion



MHC extremely polymorphic for a reason!

CNVs: structural variations in the genome

- Responsible for more heritable sequence differences between individuals than SNPs
- Might contribute to variation in phenotypic expression of complex traits

→ More complex structures underlying phenotypic variation

CNVRs in the MHC class I-II-III region on chromosome 20 are associated with insect bite hypersensitivity in Friesian horses

Our study contributes to the understanding of the equine genome and its expression

WAGENINGEN UR
For quality of life

KFPS • ROYAL FRIESIAN
KONINKLIJK FRIESCH PAARDEN-STAMBOEK

Horsegene

PennCNV: CNV detection with LRR and BAF

Wang *et al.*, 2007

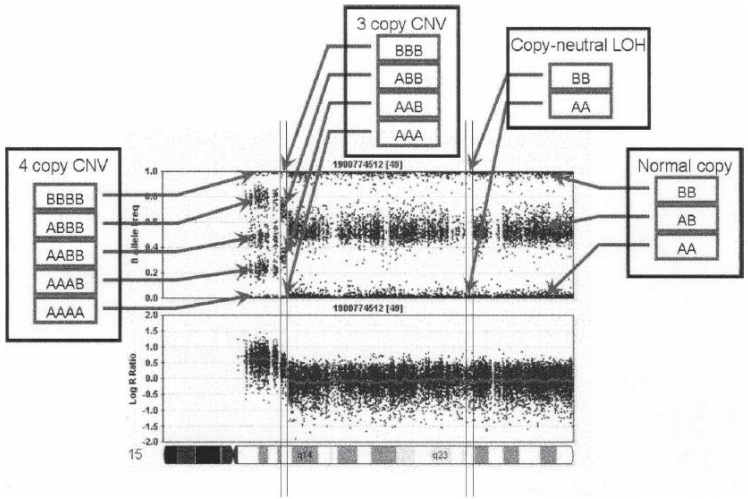


Figure 1. An illustration of log R Ratio (LRR) and B Allele Freq (BAF) values for the chromosome 15 q-arm of an individual. A normal chromosome region has three BAF genotype clusters, as represented as AA, AB, and BB genotypes in boxes, and with LRR values centered around zero. The copy-neutral LOH region has normal LRR values, but without the AB genotype cluster. The increased copy number for a CNV region can be detected based on an increased number of peaks in the BAF distribution, as well as increased LRR values. The patterns of LRR and BAF for different CNV regions, normal regions, and copy-neutral LOH regions are distinct from each other, thus the combination of LRR and BAF can be used to generate CNV calls.

Inference of log R Ratio (LRR) and B Allele Frequency (BAF)

For each SNP, its two alleles are referred to as the A and B alleles using a set of specific naming rules (see http://www.illumina.com/downloads/TopBot_TechNote.pdf). The raw signal intensity values measured for the A and B alleles are then subject to a five-step normalization procedure using the signal intensity of all SNPs (see Illumina white paper at https://icom.illumina.com/icom/software.ilmn). This procedure produces the X and Y values for each SNP, representing the experiment-wide normalized signal intensity on the A and B alleles, respectively. Two additional measures are then calculated for each SNP, where $R = X + Y$ refers to the total signal intensity, and $\theta = \arctan(Y/X)/(\pi/2)$ refers to the relative allelic signal intensity ratio.

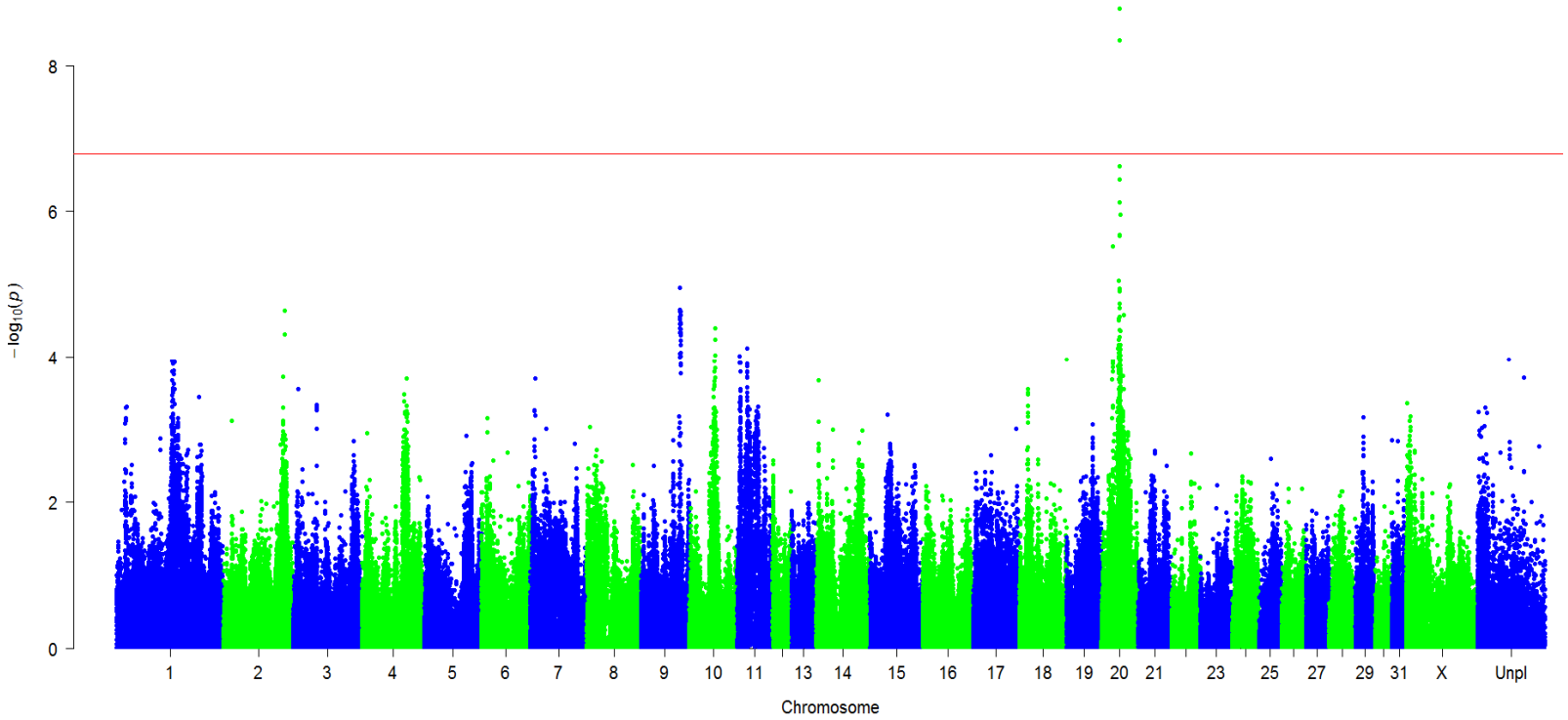
As a normalized measure of total signal intensity, the log R Ratio (LRR) value for each SNP is then calculated as $\mathrm{LRR} = \log_2(R_{\mathrm{observed}}/R_{\mathrm{expected}})$, where R_{expected} is computed from linear interpolation of canonical genotype clusters (Peiffer et al. 2006). The B Allele Frequency (BAF) is a somewhat confusing term that actually refers to a normalized measure of relative signal intensity ratio of the B and A alleles:

$$BAF = \begin{cases} 0, \text{ if } \theta < \theta_{AA} \\ 0.5(\theta - \theta_{AA})/(\theta_{AB} - \theta_{AA}), \text{ if } \theta \le \theta < \theta_{AB} \\ 0.5 + 0.5(\theta - \theta_{AB})/(\theta_{BB} - \theta_{AB}), \text{ if } \theta_{AB} \le \theta < \theta_{BB} \\ 1, \text{ if } \theta \ge \theta_{BB} \end{cases} \quad (1)$$

where θ_{AA}, θ_{AB}, and θ_{BB} are the θ values for three canonical genotype clusters generated from a large set of reference samples. The transformation from θ to BAF values adjusts for different chemical characteristics of each SNP so that values for different SNPs are more comparable to each other.

WAGENINGEN UR
For quality of life

SNP-based GWAS results



WAGENINGEN UR
For quality of life