High resolution copy number variation analysis using two cattle genome assemblies

70th Annual Meeting of the European Federation of Animal Science
(26th August 2019)

Young-Lim Lee

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

Acknowledgement

Aniek C. Bouwman

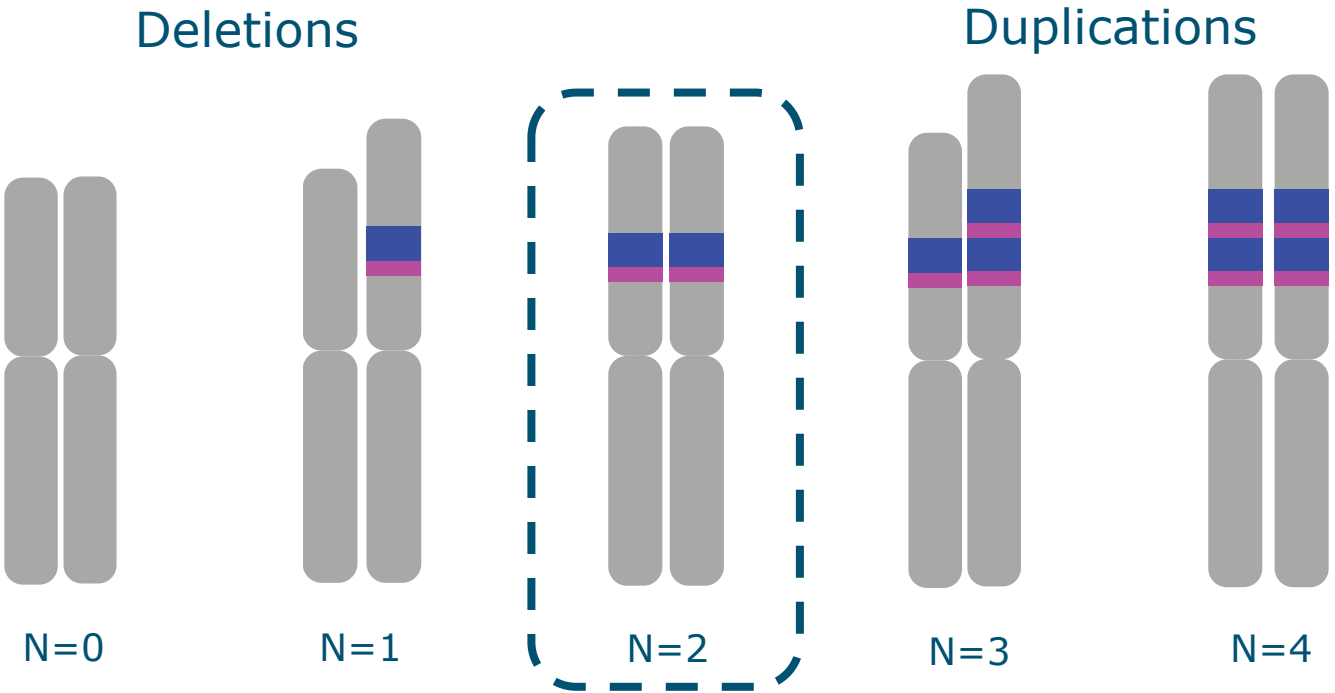
Martien A. M. Groenen

Erik Mullaart

Roel F. Veerkamp

Mirte Bosse

Copy number variations

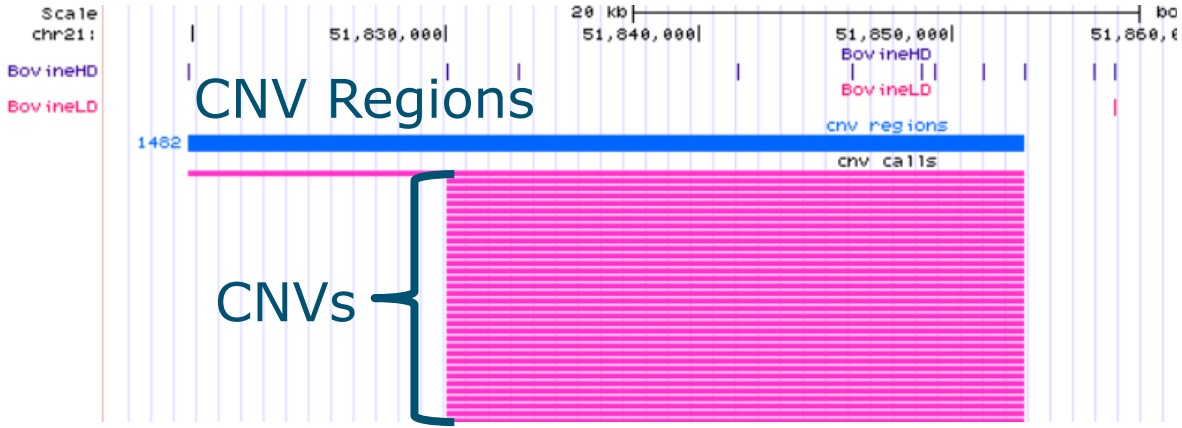


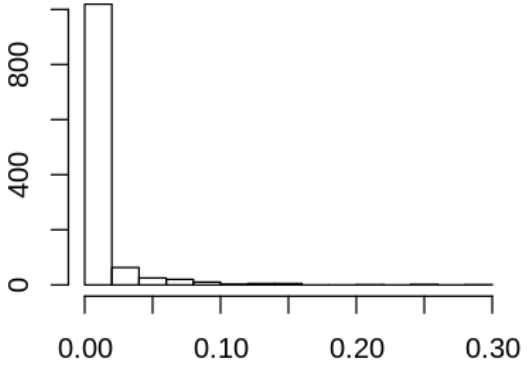
Materials & Methods

- Data
  - Illumina BovineHD data (770K, n=315)
  - WGS data (~10X, n=50)
- Methods
  - Genome assembly ARS-UCD1.2
  - PennCNV software, Smoove, and FreeBayes
  - Shapeit & PLINK software

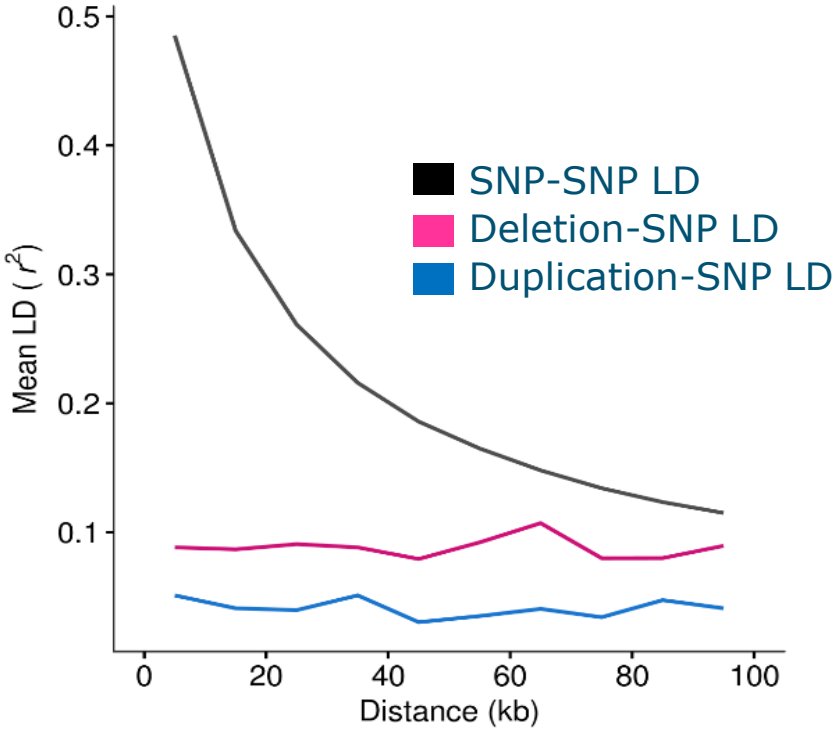
CNV discovery

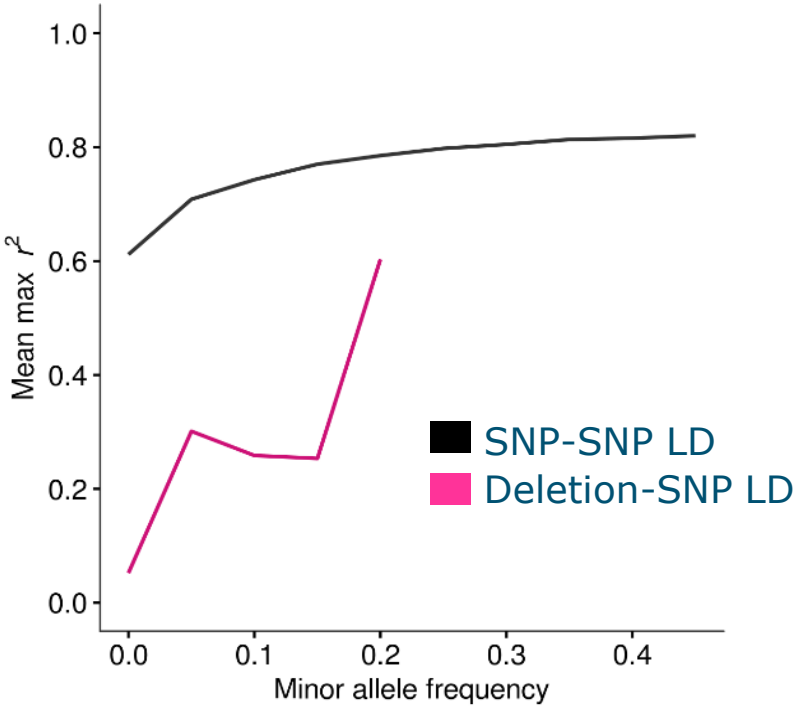
- 8,881 CNVs
- 1,154 CNV regions (CNVRs)





CNV Linkage Disequilibrium (LD)





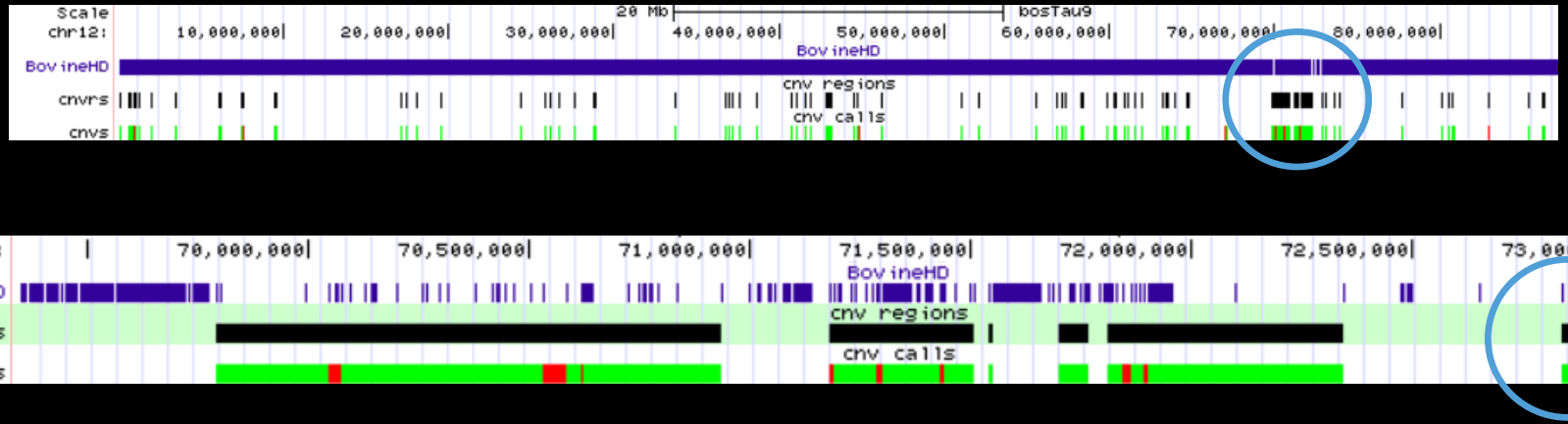
WAGENINGEN
UNIVERSITY & RESEARCH

The dark side of CNV discovery in commercial arrays

Two folds bias in commercial SNP arrays

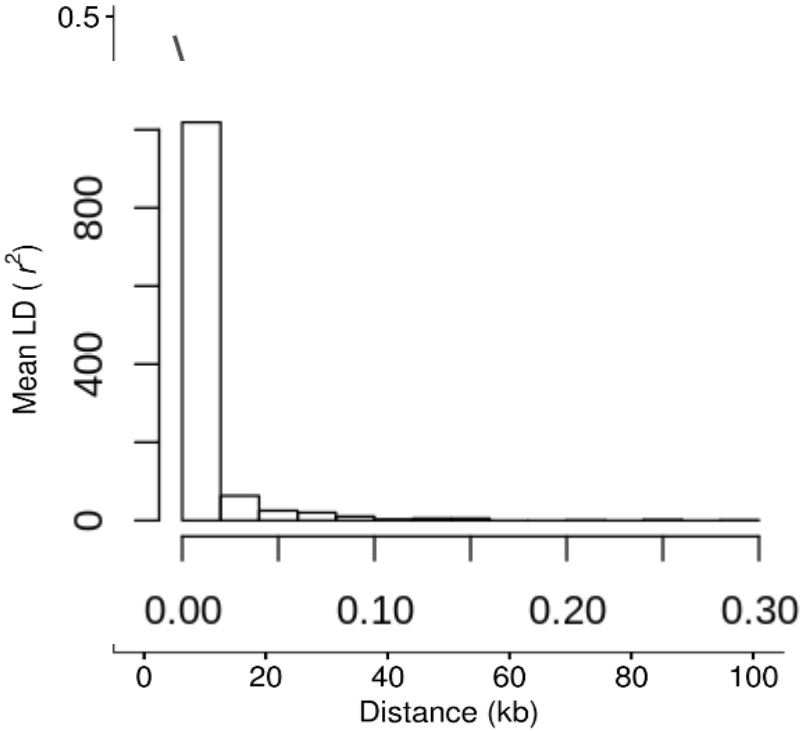
1. SNP density determines size of the detectable CNVs

2. Ascertainment bias determines CNV “detectable” regions



Follow-up on CNV LD

- Why low CNV LD ?
  - Deletions vs. Duplications
  - Low CNV MAF
  - Low local SNP density (ascertainment bias)



Low density (50K)

High density (770K)

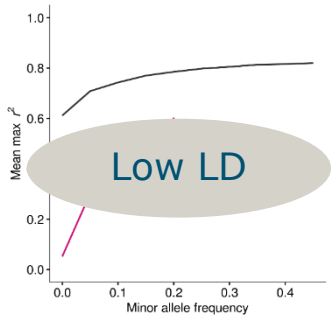
770K+ additional SNPs

Local SNP density

Rare

Minor Allele Frequency

Common



Low LD

High LD

WAGENINGEN
UNIVERSITY & RESEARCH

100years
1918 — 2018

CNV LD using WGS data

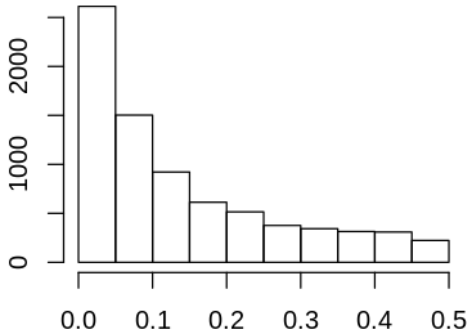
WGS data (n=50)

Calling deletions

7,483 deletions
Uniform MAF

Calling SNPs (15M)

- Bovine 50K
- Bovine 770K
- Bovine 770K+2M



Low density (50K)

High density (770K)

770K+ **additional 2.1M**

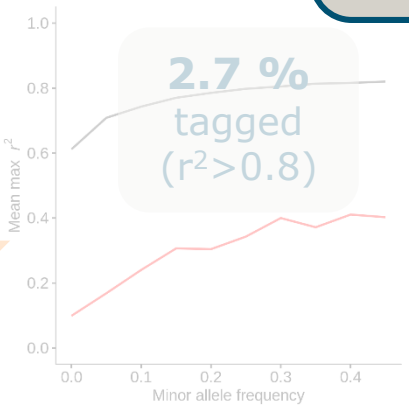
Local SNP density

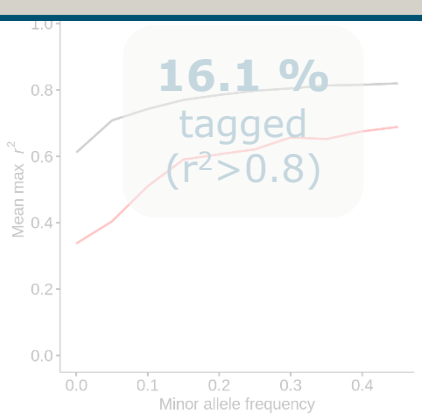
Minor Allele Freq

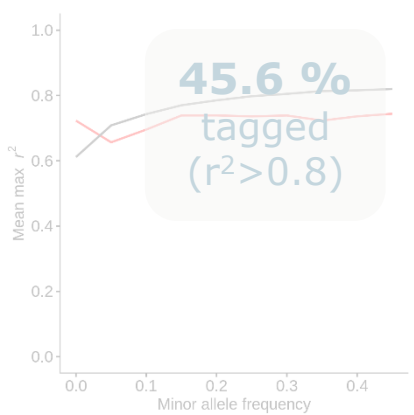
Rare

Common

~47K SNPs from WGS variants tag deletions







WAGENINGEN UNIVERSITY & RESEARCH

Future plan & Take home messages

Validating **~47K deletion tagging SNPs** in genomic prediction

1. Commercial SNP arrays do not effectively tag CNVs
2. Deletions can be tagged using dense SNP set