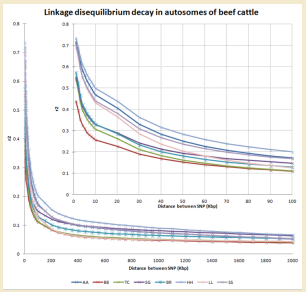
What do SNP chips say about sequence variation?

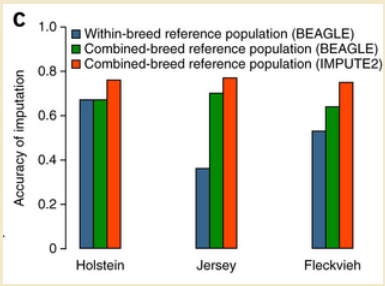
BovineHD linkage disequilibrium



Porto-Neto et al., 2014

- Strong correlations among close SNP
  - correlations drop with increased separation between SNP
- Expect SNP genotypes to indicate nearby sequence variant genotypes?

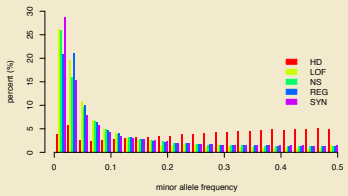
Imputing BovineHD to sequence



Daetwyler et al., 2014

- HD-sequence accuracy ~0.75
- accuracy >0.95 typical for 50K-HD

GPE minor allele frequency

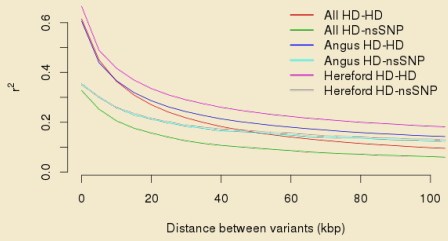


- Differences in minor allele frequencies (MAF) limit correlations between BovineHD and sequence variants
- 0.35 max r^2 between mean HD and mean SV
- Expect SNP genotypes to indicate nearby sequence variant genotypes?

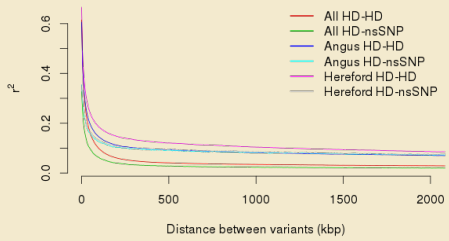
Linkage disequilibrium in GPE bulls

- 176 sires with HD genotypes, low-coverage genome and high-coverage exome sequence
  - 123 purebred bulls (15-19 bulls/breed)
  - 53 F_1 bulls
  - HD SNP with MAF>0.05 in sequenced GPE bulls
  - non-synonymous SNP with MAF>0.05 in sequenced GPE bulls
  - BovineHD chip genotypes, non-synonymous SNP (nsSNP) genotypes called from sequence
  - HD-HD, HD-nsSNP r^2 from all bulls, Angus, Hereford

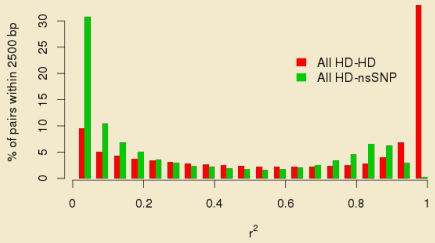
LD decay observed in GPE bulls



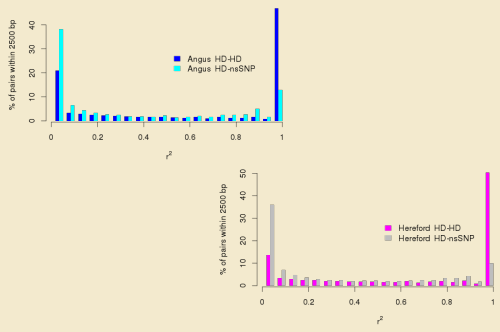
LD decay observed in GPE bulls



Distribution of LD among close variants



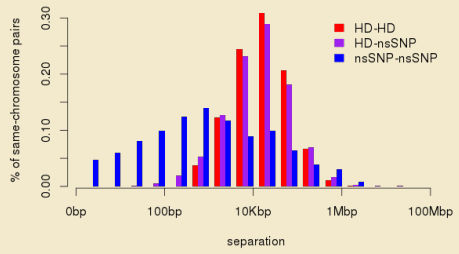
Distribution of LD among close variants



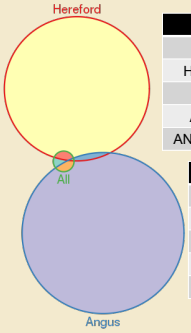
high LD pairs ($r^2 > 0.81$, all bulls)

- HD-HD
  - 75% HD SNP in high LD with other HD SNP on same chromosome
    - 25% 0 other SNP
    - 44% 1-5 other SNP
    - 32% 6-207 other SNP
- HD-nsSNP
  - 20% nsSNP in high LD with HD SNP (< 0.1% HD on other chromosomes)
    - 80% 0 HD SNP
    - 14% 1-5 HD SNP
    - 6% 6-180 HD SNP
- nsSNP-nsSNP
  - 32% nsSNP in high LD with other nsSNP (~0.2% nsSNP on other chromosomes)
    - 68% 0 nsSNP
    - 27% 1-5 nsSNP
    - 5% 6-57 nsSNP

separation between high LD pairs



overlap among high LD pairs



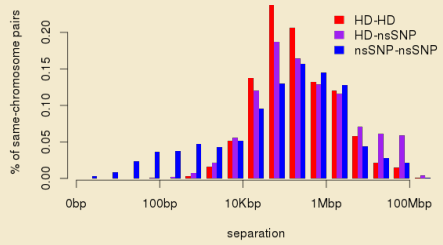
Instances of nsSNP-HD $r^2 > 0.81$			
	Hereford	Angus	All bulls
Hereford	6,633,783	48,438	30,709
Angus		8,293,905	35,133
All bulls			50,006
AN-HH-All			22,322

nsSNP in high LD pairs			
	Hereford	Angus	All bulls
Hereford	19,732	5,906	4,775
Angus		24,429	5,816
All bulls			8,513
AN-HH-All			3,524

moderate LD pairs ($0.25 < r^2 < 0.81$)

- HD-HD
  - 98% HD SNP have moderate LD with other HD SNP on same chromosome
    - 2% 0 other SNP
    - 31% 1-10 other SNP
    - 67% 11-1347 other SNP
- HD-nsSNP
  - 79% nsSNP have moderate LD with HD SNP (36% HD on other chromosomes)
    - 21% 0 HD SNP
    - 30% 1-10 HD SNP (9% HD on other chromosomes)
    - 49% 11-18179 HD SNP (28% HD on other chromosomes)
- nsSNP-nsSNP
  - 90% nsSNP have moderate LD with other nsSNP (83% nsSNP on other chromosomes)
    - 10% 0 nsSNP
    - 54% 1-10 nsSNP (48% nsSNP on other chromosomes)
    - 36% 11-338 nsSNP (35% nsSNP on other chromosomes)

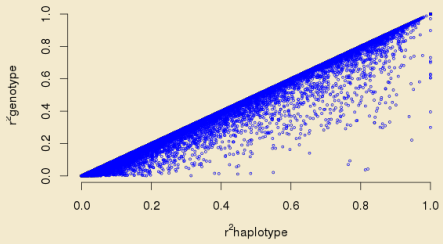
separation between moderate LD pairs



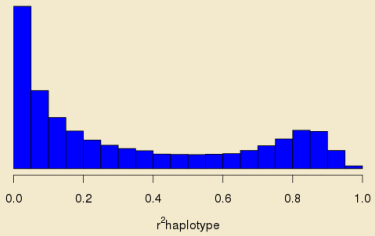
LD with flanking HD SNP

- r^2 from two regression models
  - genotype model
    - $g_{nsSNP} = g_{left\ HD} + g_{right\ HD} + e$
      - g_{nsSNP} – nsSNP genotype called from sequence
        - 0,1,2 copies of alternate allele
      - $g_{left\ HD}$, $g_{right\ HD}$ – BovineHD genotypes of SNP on either side of nsSNP
        - 0,1,2 copies of B allele
  - haplotype model
    - $g_{nsSNP} = AA + AB + BA + BB + e$
      - AA, AB, BA, BB - possible 2-SNP haplotypes
        - 0,1,2 copies of each possibility extracted from findhap imputation

LD with flanking HD SNP



LD with flanking HD SNP



What if nsSNP-HD LD is typical of unknown QTL-chip SNP LD?

- limited high QTL-SNP LD
  - few QTL may have strong correlations with any genotyped SNP
- much low-moderate QTL-SNP LD
  - most QTL may have weak to moderate correlations with many genotyped SNP

What if nsSNP-HD LD is typical of unknown QTL-chip SNP LD?

- genomic selection / GWAS
  - combination of many QTL correlated to SNP genotypes may contribute to SNP effects
    - SNP effects accurately predict variation as long as SNP-QTL correlations are consistent between training and target populations
    - GWAS misleading?
      - signal from correlations with many QTL?
      - signal from strong correlation with distant QTL?
      - missed signal from QTL not strongly correlated with SNP?

Avoiding chip SNP - QTL LD?

- use sequence variants predicted to affect genes instead of standard chip SNP?
  - variation in coding sequence more likely to be QTL?
    - slight accuracy increase for imputed sequence variants over BovineHD (Hayes et al., 2014)
  - variation in coding sequence will not eliminate LD
    - estimated variant effects will be influenced by effects of correlated variants
    - effects more portable than chip SNP effects?

Avoiding chip SNP - QTL LD?

- genotyping coding sequence variation
  - imputation - low accuracy, especially for MAF < 0.20 (Hayes et al., 2014)
    - dependent on LD
  - direct genotypes
    - coding variant assay
    - custom assays

Questions?