Genomic Signatures of Selection in the Horse

Jessica Petersen, Jim Mickelson, Stephanie Valberg, Molly McCue
University of Minnesota, USA
and the Equine Genetic Diversity Consortium

Selective Breeding in the Horse

- Since domestication selective pressures on the horse genome have been directed toward use in agriculture, transportation, and warfare.
- More recently breed registries, and continued breed specialization, have focused more upon improving traits related to aesthetics, performance, and the ability to do work.
- The result is wide variation in phenotypes across breeds, and the fixation or near- fixation of some of the desired traits within many breeds.

Detection of Loci Under Selection

- Genomic segments and the underlying functional alleles also become fixed.
- We have used Illumina 54,000 SNP genotype data collected from 33 breeds to begin to identify putative genomic regions under selection in the modern horse.
- Once regions targeted by selection are identified the variants and processes that have contributed to desired phenotypes can more readily be defined.

Equine Genetic Diversity Consortium

Lisa Andersson, Jeanette Axelsson, Gabriella Lindgren, Sofia Mikko
Ernie Bailey, Kathryn Graves
Danika Bannasch, Cecilia Penedo
Matthew Binns
Alexandre Borges
Katia Cappelli, Stefano Capomaccio, Michela Felicetti, Maurizio Silvestrelli
E Gus Cothran
Artur da Câmara Machado, Maria Susana Lopes
Ottmar Distl
Gérard Guérin
Telhisa Hasegawa
Bianca Haase, Claire Wade
Tosso Leeb, Stefan Rieder
Hannes Lohi, Marja Raekallio, Karin Hemmann
Emmeline W Hill, Nicholas Orr, Pieter Brama, Beatrice McGivney
Richard Piercy
Knut Røed
Oliver Ryder
June Swinburne, Mark Vaudin
Teruaki Tozaki

744 Horses from 33 Breeds

Breed	N
Akhal Teke	19
Andalusian	18
Arabian	24
Belgian	30
Caspian Pony	18
Clydesdale	24
Exmoor	24
Fell Pony	21
Finnhorse	27
Franches-Montagnes	19
French Trotter	17
Hanoverian	15
Icelandic	25
Mangalarga Paulista	15
Miniature	21
Mongolian	19
Morgan	40
New Forest Pony	15
North Swedish Horse	19
Norwegian Fjord	21
Paint	25
Percheron	23
Peruvian Paso	21
Puerto Rican Paso Fino	20
Quarter Horse	40
Saddlebred	25
Shetland	27
Shire	23
Standardbred	25
Swiss Warmblood	14
Tennessee Walking Horse	19
Thoroughbred - United Kingdom	19
Thoroughbred - United States	17
Tuva	15

Analysis

- An average of 22.5 horses/breed were genotyped.
- 500 kb windows of the genome were chosen for analysis.
- A minimal density of 4 SNPs per window was required.
- In total 23,401 within 3,229 windows SNPs were evaluated .
  - Most SNPs not included in analyses were in windows that did not meet the minimal SNP density.
  - The average SNP density was 7.25 SNPs per window (range 4-20).
  - Coverage of the autosomes was 68.7%.

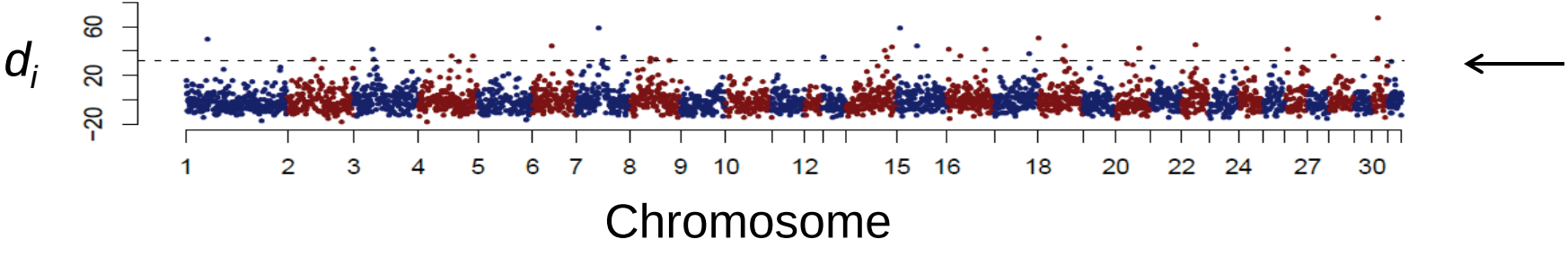
Analysis

$$d_i = \sum_{j \neq i} \frac{F_{ST}^{ij} - E[F_{ST}^{ij}]}{sd[F_{ST}^{ij}]}$$

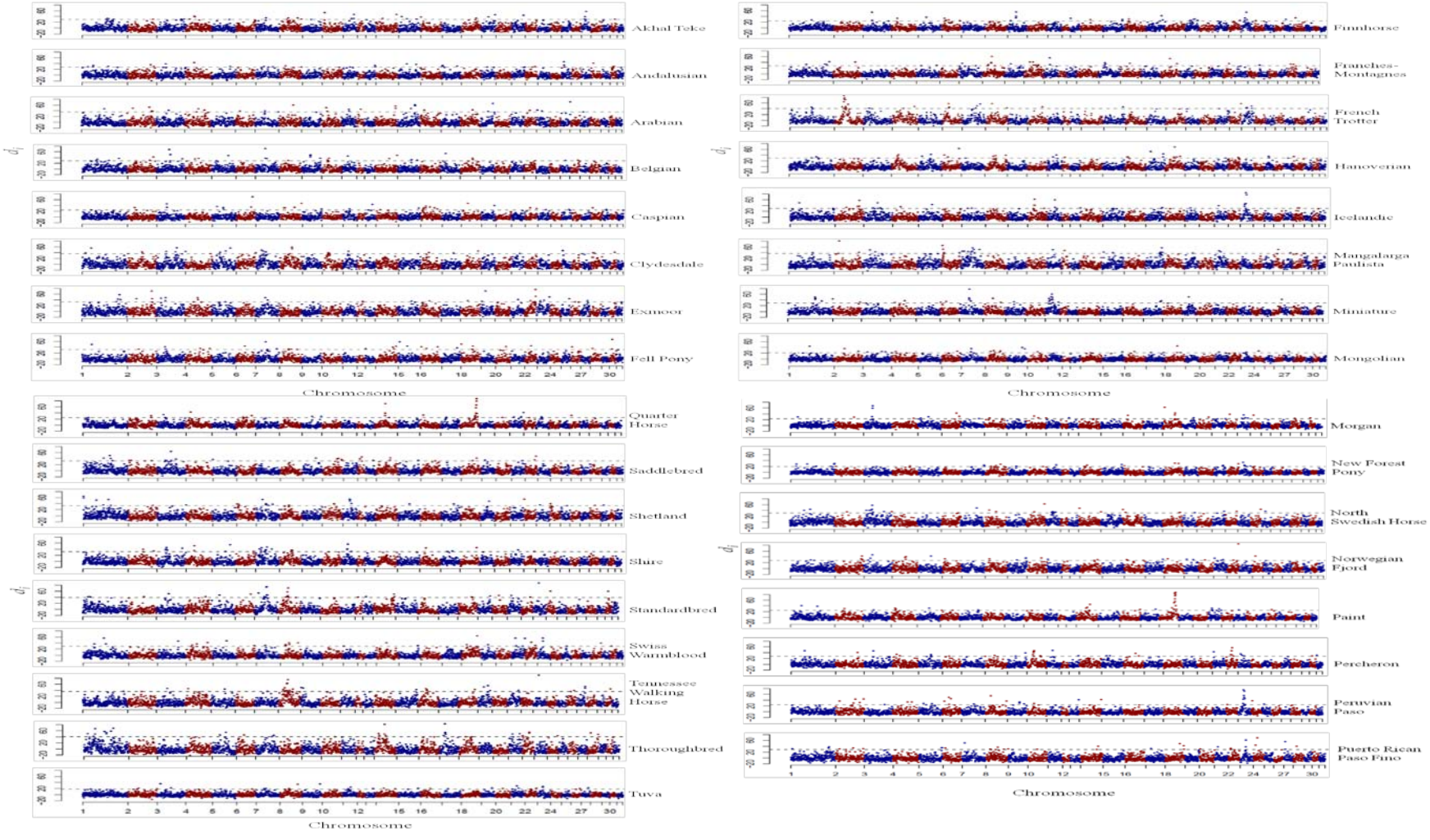
Akey et al. 2010

d_i detects locus specific deviation in allele frequencies for each breed relative to the genome-wide average of pair-wise F_{ST} summed across breeds.

- A large value of d_i indicates greater divergence at that window than observed across the genome as a whole.
- 33 windows within each breed fall into the upper 99th percentile of the empirical distribution and were considered putative signatures of selection.

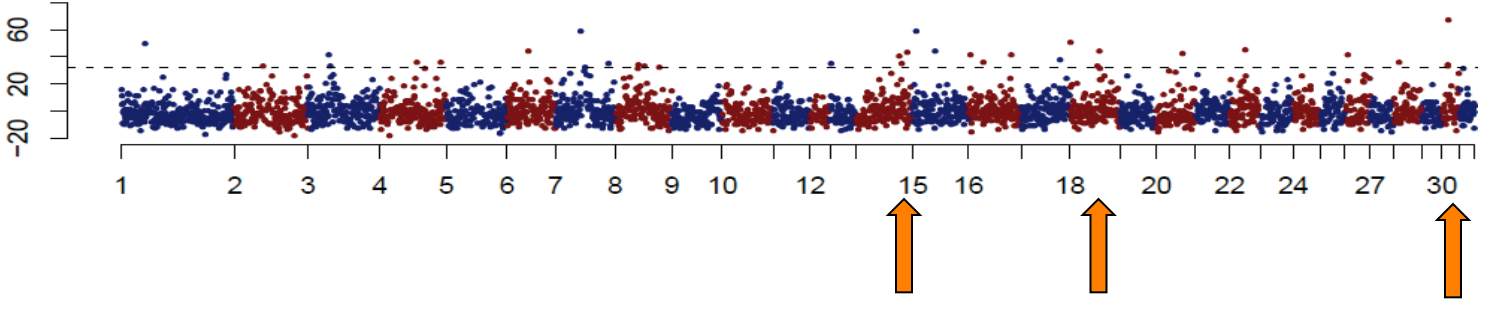


d_i Plots for All 33 Breeds

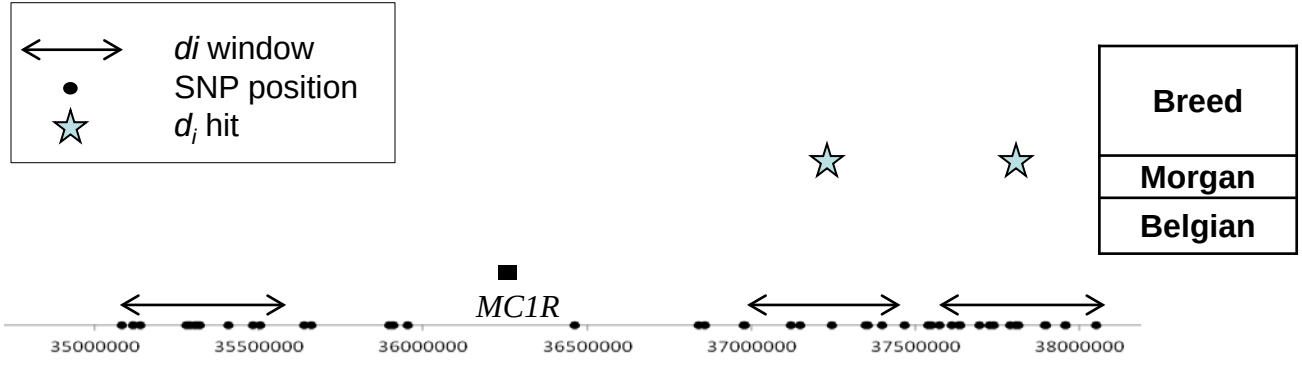


Prioritizing Loci for Follow-up

- Windows containing the highest d_i value within a breed.
- Windows that contain consecutive segments of significant d_i values within a breed.
- Windows that are shared across breeds experiencing selective pressure for similar phenotypes.
- Windows that are near candidate genes with known functional significance.

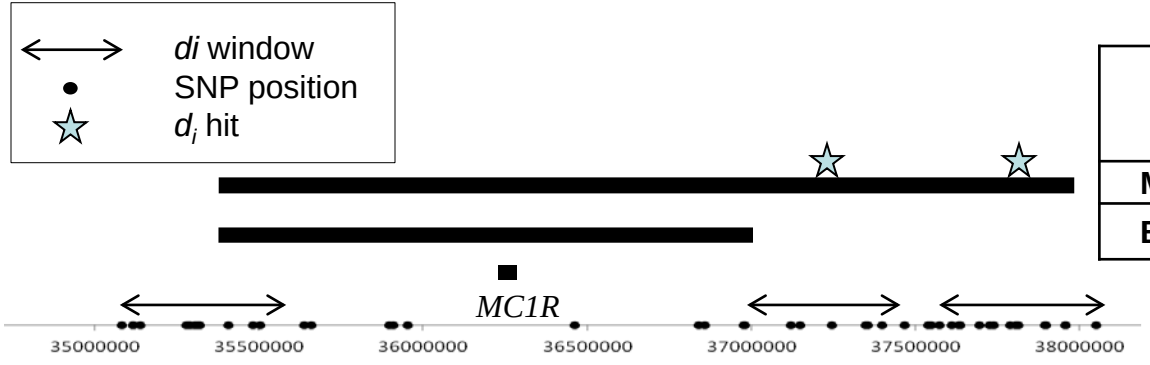


Proof of Principle and Caveats: the *MC1R* locus



- The *MC1R* chestnut allele is selected for and was nearly fixed in our Morgan and Belgian cohorts.
- The highest d_i hit in Morgan horses was on ECA3 in the vicinity of the *MC1R* gene.
- However, Belgians did not have a d_i hit over *MC1R*.

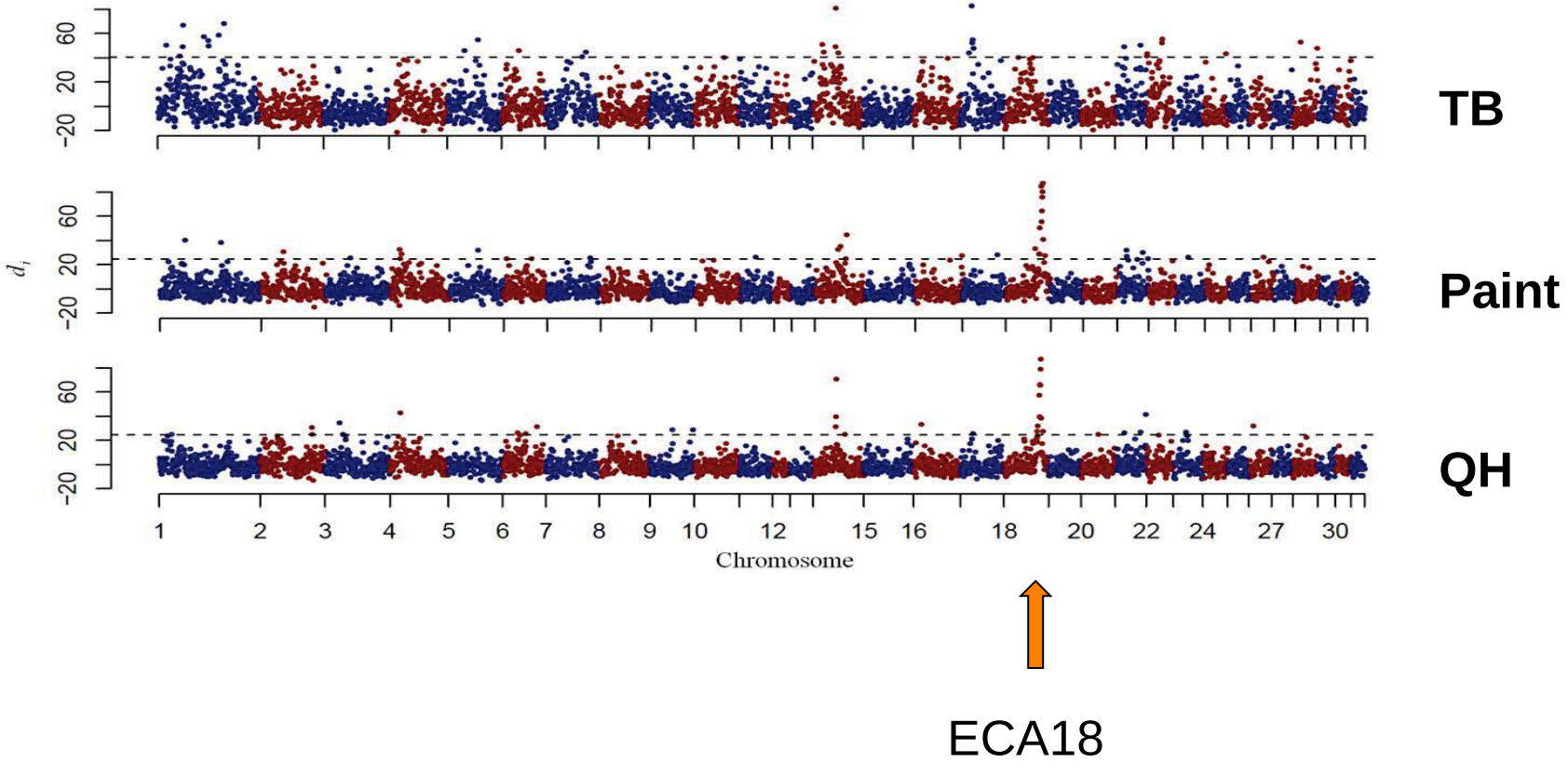
Proof of Principle and Caveats: the *MC1R* locus



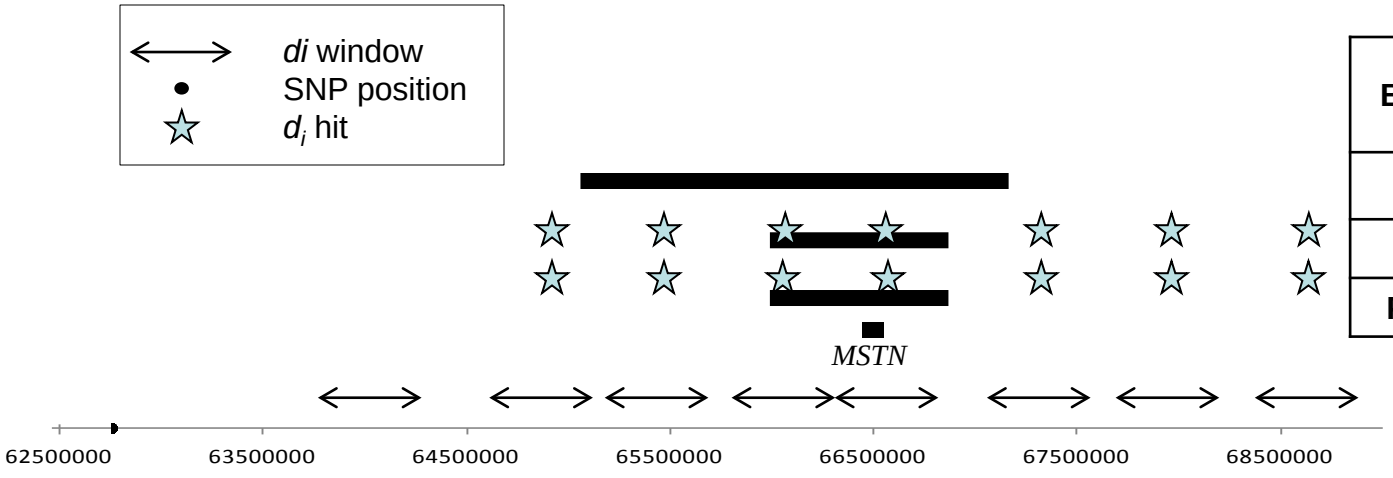
Breed	Haplotype Length (Mb)	Haplotype Frequency
Morgan	2.55	0.95
Belgian	1.57	1.00

- Phasing the SNP data and building haplotypes over the region reveals an extended conserved haplotype that covers the *MC1R* locus in Morgans.
- An identical, but shorter haplotype is then found in the Belgians (and a number of other breeds).

d_i Plots for the Thoroughbred, Paint, and Quarter Horse



ECA18 Haplotypes



Breed	Haplotype Length (Mb)	Haplotype Frequency
TB	2.00	0.53
QH	0.78	0.91
Paint	0.78	1.00

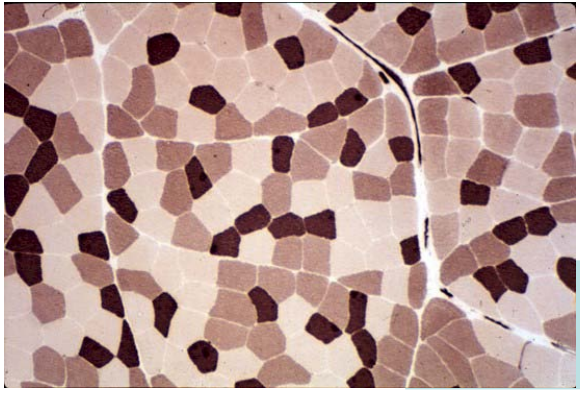
- A minimal shared haplotype within QH and Paints is 0.78 Mb long and occur at a frequency of 0.91 – 1.00.
- The identical haplotype is within a 2 Mb segment in TB and occurs at a frequency of 0.53 in TB.
- 12 genes are in this region including *MSTN.*

Myostatin *(MSTN)* and Racing Performance

- Polymorphisms in equine *MSTN* have been studied by several groups and variants found to be associated with performance in Thoroughbreds.

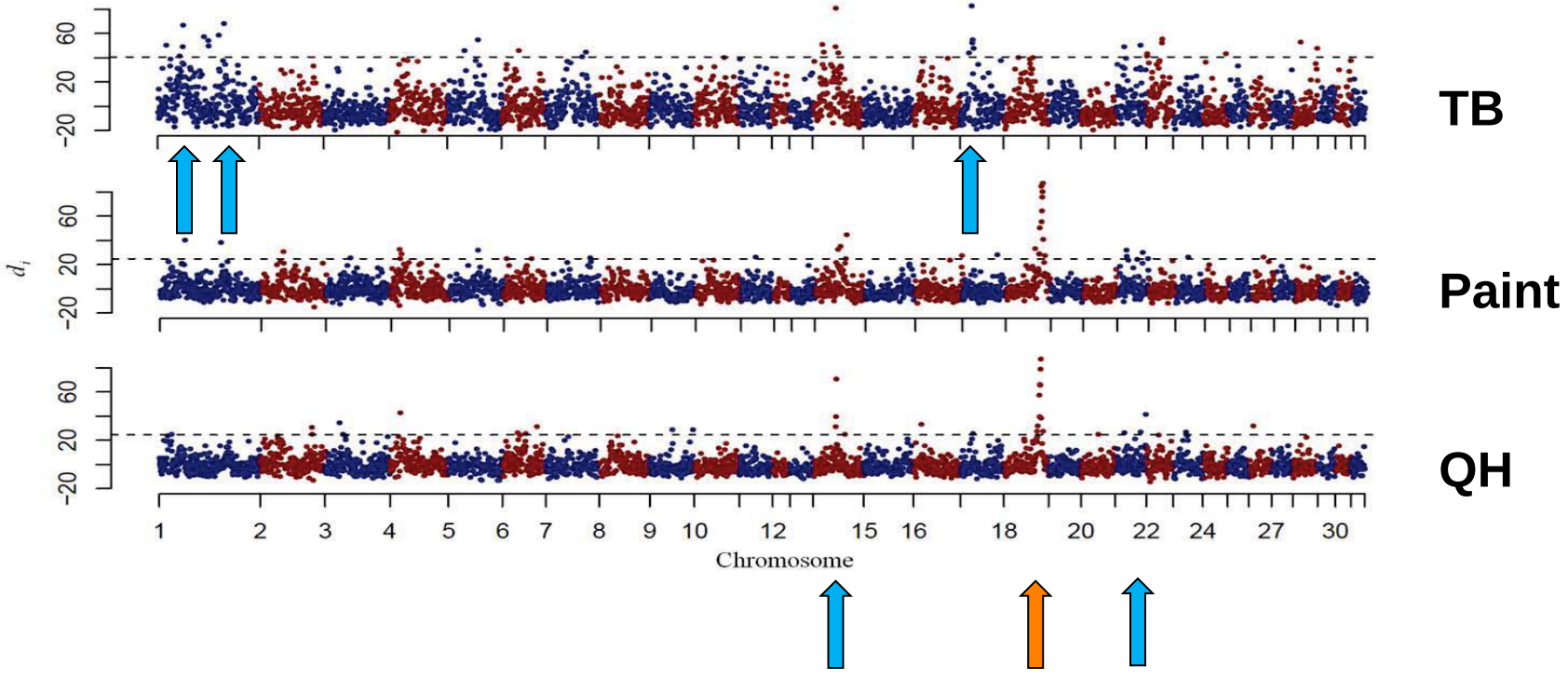
- We further investigated *MSTN* in the Quarter Horse and Paint breeds.
- A SINE insertion in the promoter was present as well as a SNP in intron 1.
- Both variants are correlated at > 0.95 in this selected haplotype.

Effect of *MSTN* Polymorphisms on Gluteal Muscle Fiber Type Proportions



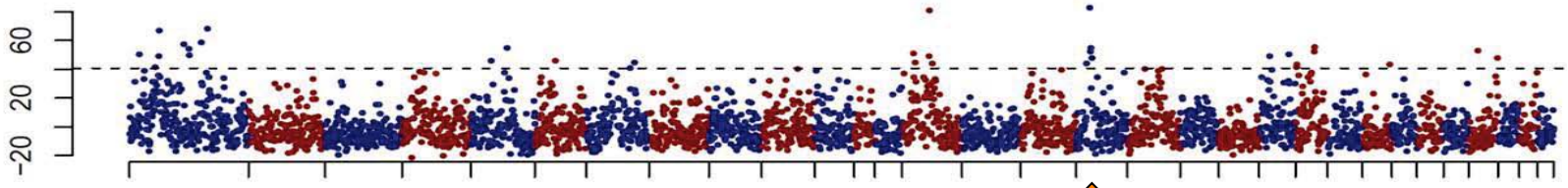
		Slow twitch	Fast twitch	
	Genotype	**Type 1 %**	**Type 2A %**	**Type 2B %**
Intron 1 SNP	TT	21.4[a]	27.0[a]	51.6[a]
	TC	18.2[a]	26.7[a]	55.1[a]
	TT	14.7[b]	24.8[a]	60.8[b]
SINE	NN	20.4[a]	27.2[a]	52.5[a]
	NS	15.9[b]	26.3[a]	57.8[ab]
	SS	15.7[b]	24.0[a]	60.3[b]

d_i Plots for the Thoroughbred, Paint, and Quarter Horse

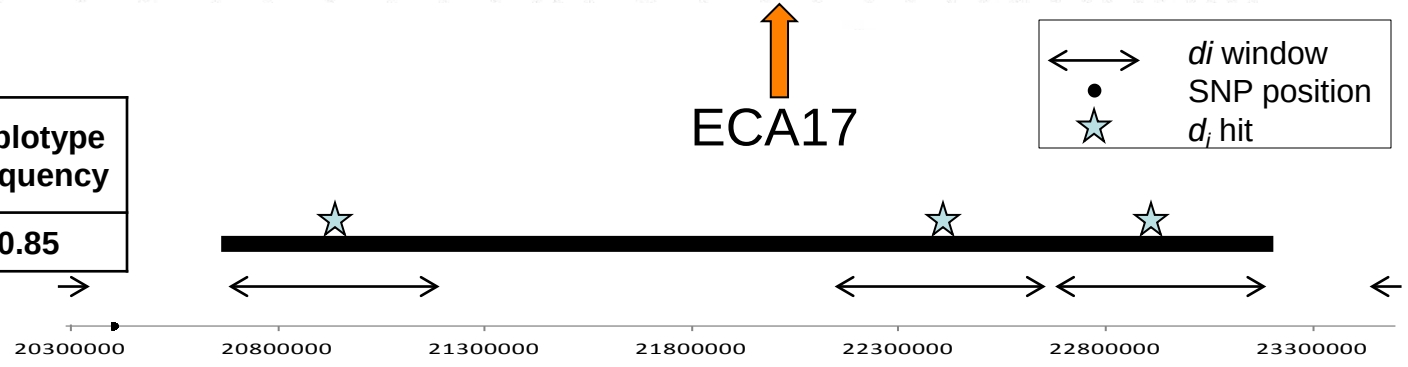


- There are clearly more loci to investigate.

ECA17 in the Thoroughbred

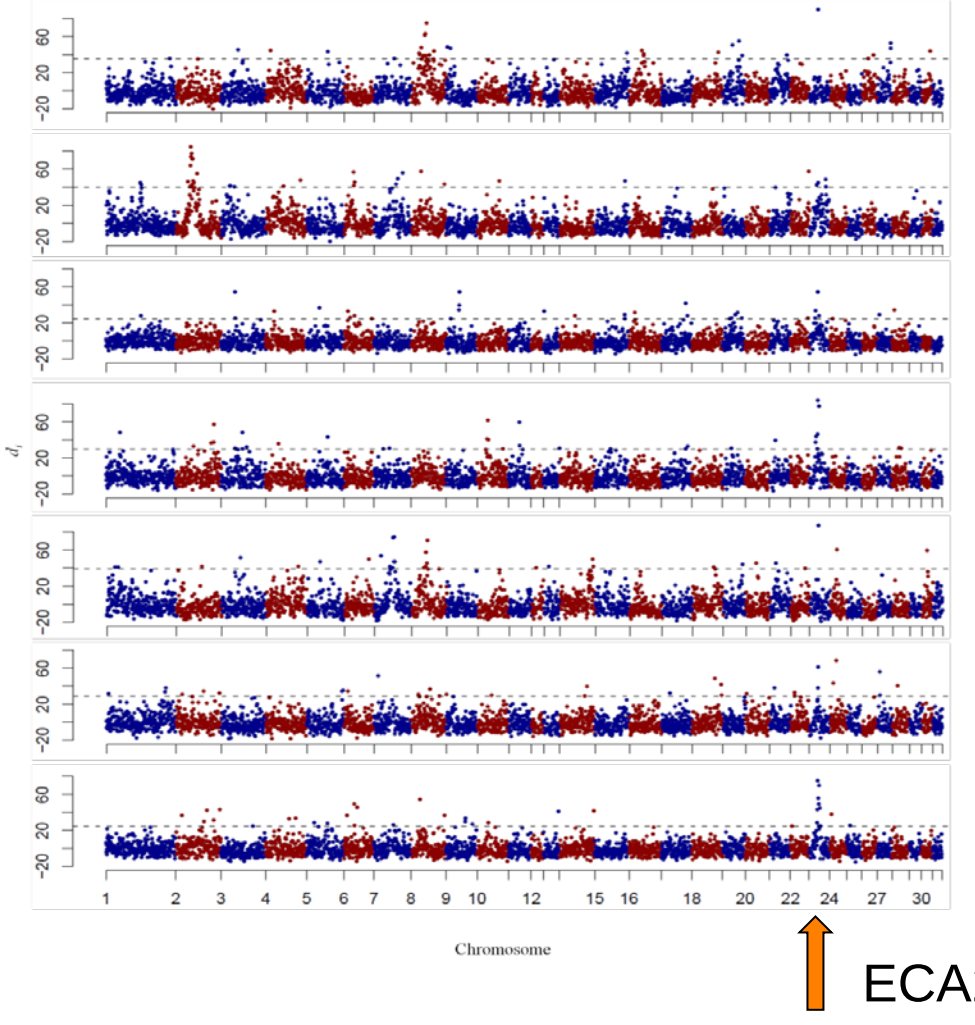


Breed	Haplotype Length (Mb)	Haplotype Frequency
TB	2.49	0.85



- The haplotype was found in 85% of TB chromosomes.
- This haplotype was also observed in Hannoverian, Swiss Warmblood, Quarter Horse, and Paint chromosomes at frequencies < 50%.
- 23 genes are in the region.

d_i Plots for Gaited Breeds and Trotters



Tennessee WH

French Trotter

Finn Horse

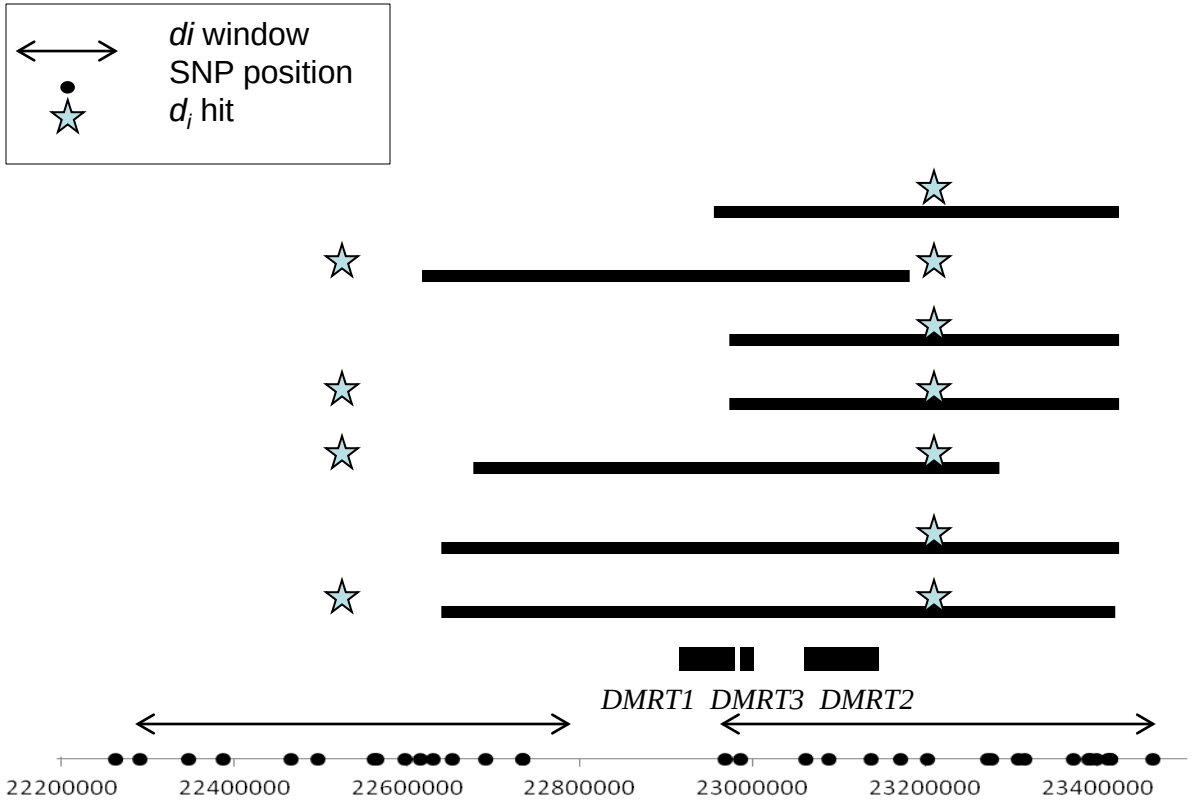
Icelandic

Standardbred

PR Paso Fino

Peruvian Paso

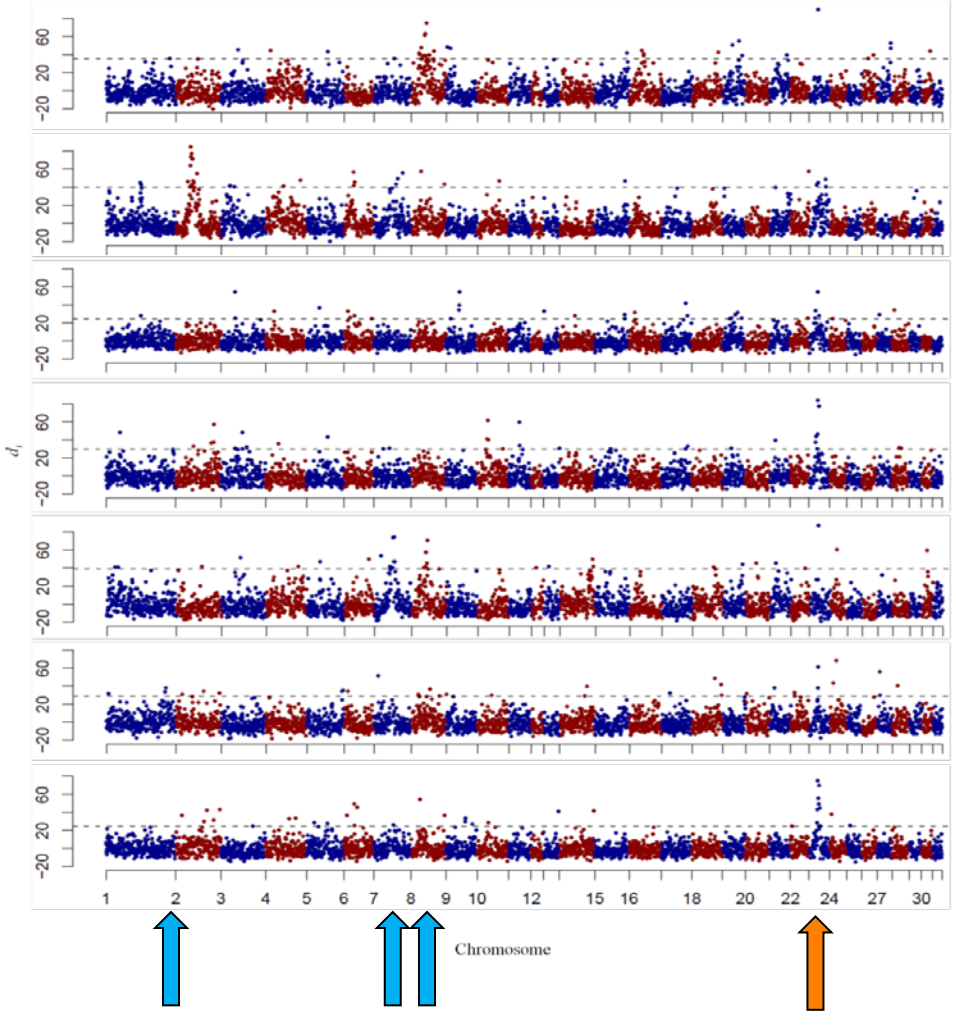
ECA23 Haplotypes



Breed	Haplotype Length (kb)	Haplotype Frequency
STBD	446	1.00
PR Paso Fino	542	0.98
Tenn WH	429	0.95
Icelandic	429	0.91
Peruvian Paso	585	0.90
French Trotter	762	0.74
Finnhorse	759	0.54

- Shared haplotypes within a breed are 429 – 759 kb long and occur at a frequency of 0.54 – 1.00.

d_i Plots for Gaited Breeds and Trotters



Tennessee WH

French Trotter

Finn Horse

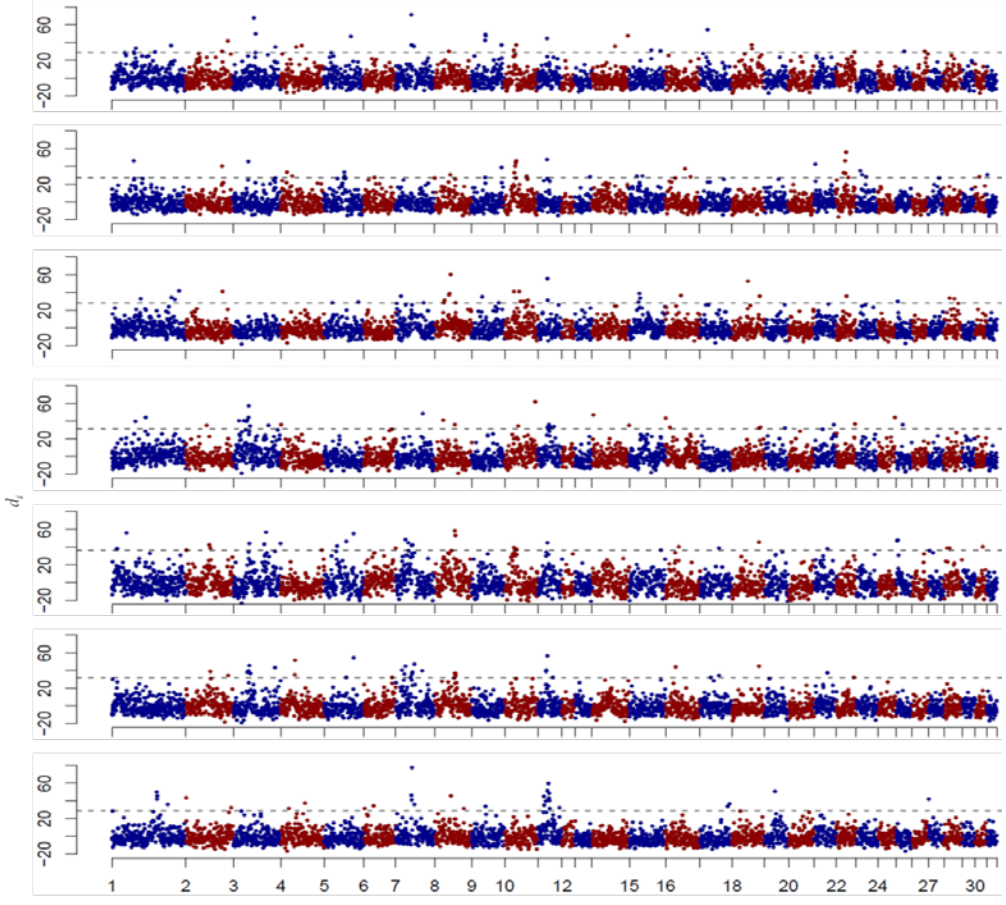
Icelandic

Standardbred

PR Paso Fino

Peruvian Paso

d_i Plots for Draft Breeds and the Miniature



Belgian

Percheron

F-Montagnes

North Swedish

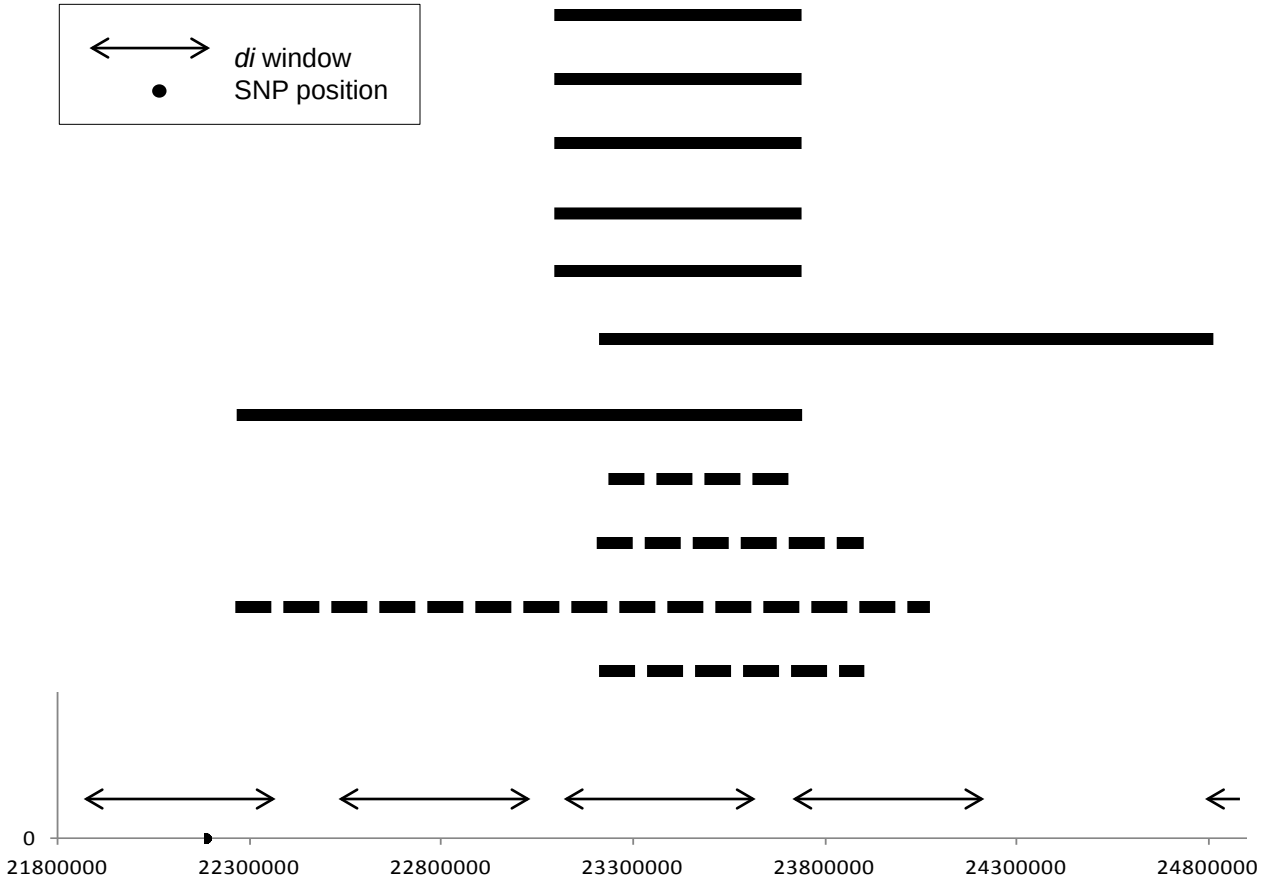
Clydesdale

Shire

Miniature

ECA11

ECA11 Haplotypes

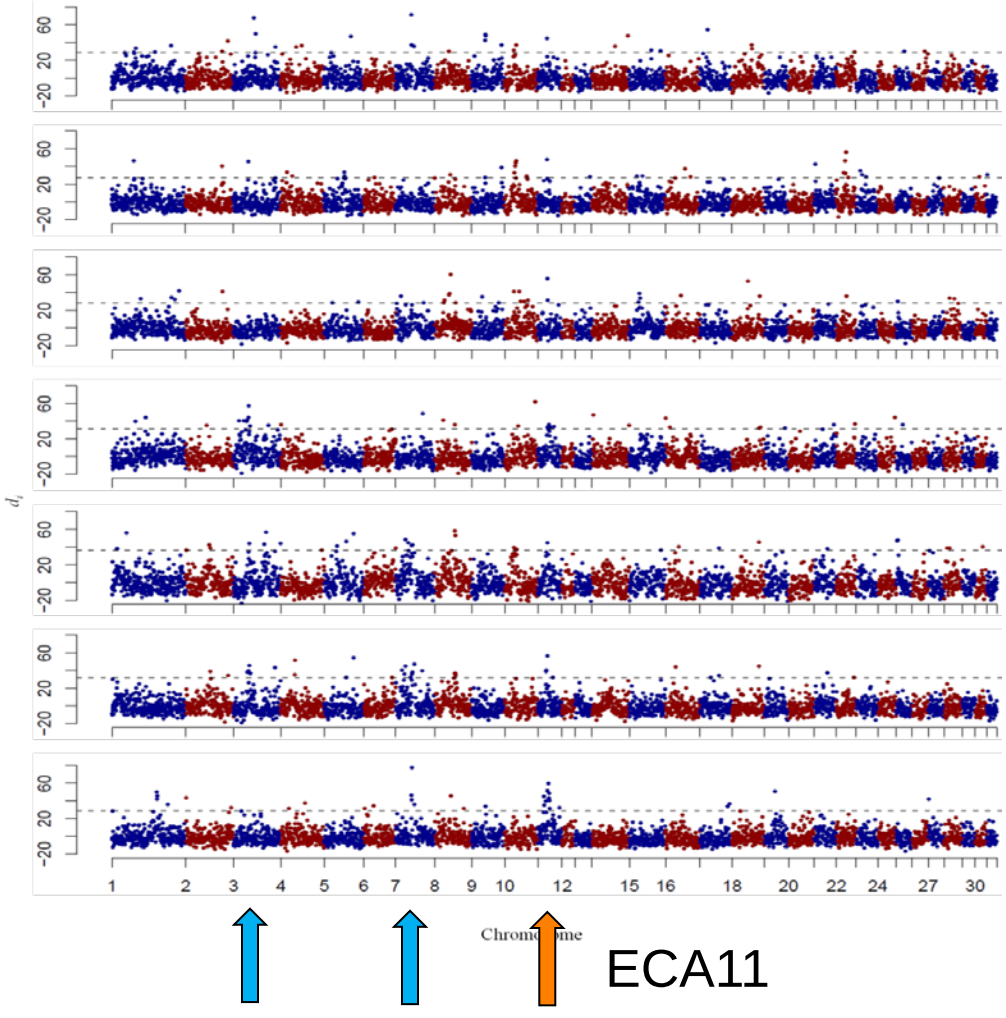


Breed	Haplotype Length (kb)	Haplotype Frequency
Belgian	594	0.82
Clydesdale	594	0.92
Percheron	594	0.74
Shire	594	0.85
Fr Montagnes	594	0.74
N Swedish Draft	1548	0.74
Hannoverian	1424	0.40
Miniature	453	0.95
Shetland	645	0.48
Tennesee WH	1755	0.61
Caspian	638	0.50

ECA11 Haplotypes

- Minimal shared haplotypes in the draft breeds are 0.59 – 1.55 Mb long and occur at a frequency of 0.74 – 0.92.
- In Miniature horses an alternative haplotype is 0.45 Mb long and occurs at a frequency of 0.95.
- 13 genes are in this region and none have been identified previously as being associated with size in mammals.

d_i Plots for Draft Breeds and the Miniature



Belgian

Percheron

F-Montagnes

North Swedish

Clydesdale

Shire

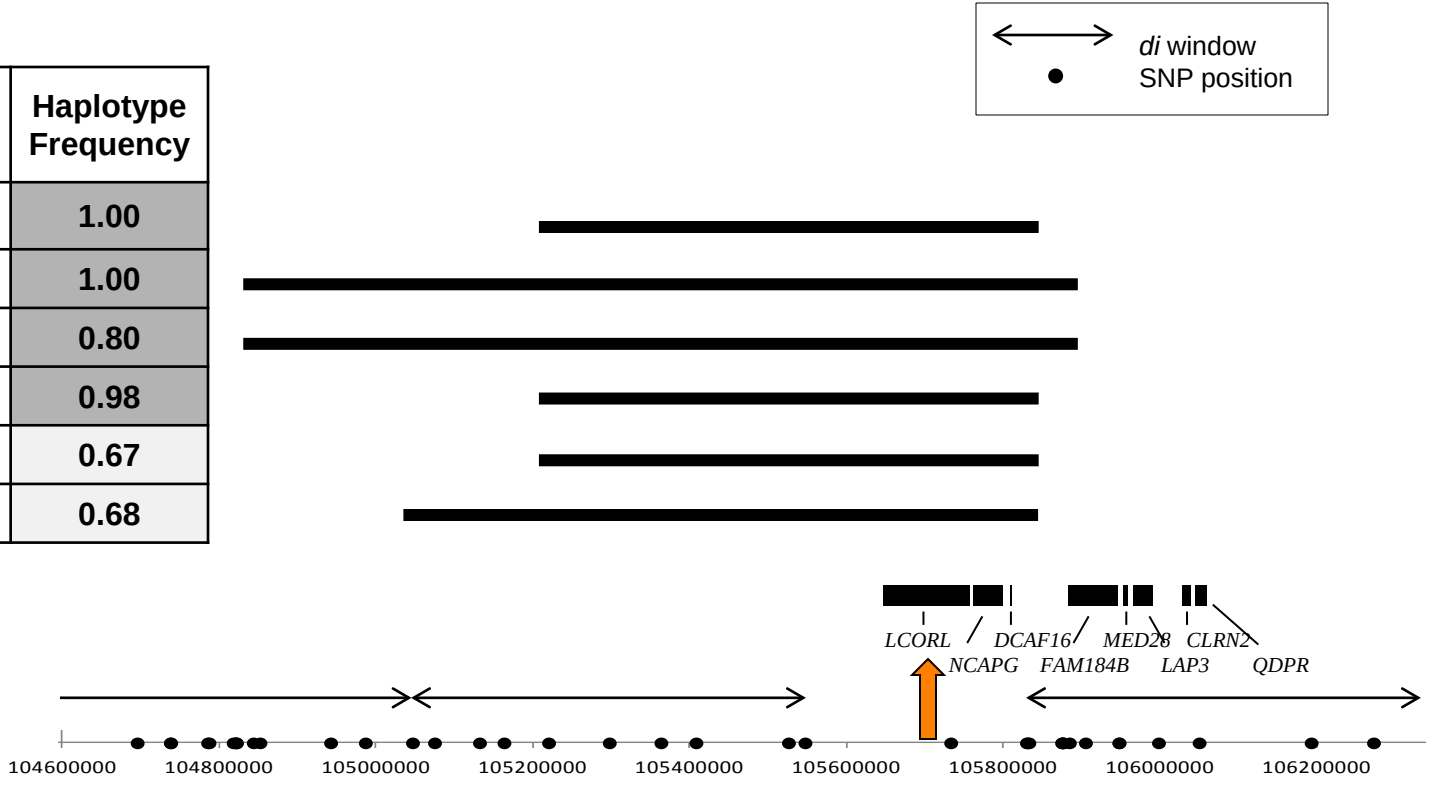
Miniature

Examining Regions that do not have a Significant d_i Hit

- As with the *MC1R* locus in Belgians, there are likely many selected loci in many breeds that were not detected by the d_i analysis alone.
- However, the SNP50 genotypes can be used in “candidate” gene studies.
- Haplotypes can be constructed across any region of interest, and length, frequency and sharing across breeds can be determined.

Candidate Genes (*LCORL/NCAPG*)

Breed	Haplotype Length (Mb)	Haplotype Frequency
Belgian	0.61	1.00
Clydesdale	1.04	1.00
Shire	1.04	0.80
Percheron	0.61	0.98
Finnhorse	0.61	0.67
Swiss WB	0.79	0.68



Identification of Functional Alleles

- We have begun using Agilent arrays to capture ~ 6 Mb from 4 different loci and that sequence is being analyzed.
  - Pooled samples were based on the “selected” haplotype (n=12) vs alternative haplotypes (n=12) at each locus.
  - As a first screen we are looking for alleles in genic regions that are at high frequency in the selected haplotype pool vs the alternate haplotype pool.
  - Move on to more complex genomic alterations.

Caveats

- The number of loci potentially worthy of follow-up investigation is huge!
  - 695 (2.7%) of the 3,229 windows were significant in at least one breed.
- Important loci can be in regions that are not included in the current analysis due to low SNP density.
- Important loci can be in regions of short LD.
- The same window may have a hit in different breeds for different reasons.
- The approach is blinded to phenotype.
- Identification of functional alleles may be challenging.

Conclusions

- This consideration of ~20 horses from 33 breeds has demonstrated the utility of a whole-genome SNP approach to identify genes important in the creation of modern horse breeds.
- Genotype data can be analyzed by the F_{ST}-based d_i statistic across the entire genome, followed by haplotype analysis, or by haplotype analysis around candidate genes.
- Loci apparently being selected for coat color, performance, muscling, gait, and size have been identified.

Conclusions

- Segments investigated thus far are from 0.5 Mb to 2.5 Mb long (1 – 5 windows) and have frequencies from 0.75 – 1.0.
- Loci identified by a high d_i value and high haplotype frequency in some breeds can be present at lower frequency and segregating in other breeds.
- We would be delighted to discuss collaborations to pursue specific loci of interest.