Genomic Signatures of Selection in the Horse
Jessica Petersen, Jim Mickelson, Stephanie Valberg, Molly McCue
University of Minnesota, USA
and the Equine Genetic Diversity 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

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

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

\( d_i \) 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.

Akey et al. 2010

A key diagram showing a scatter plot of \( d_i \) values against chromosome number.
$d_i$: 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*. 
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

TB

Paint

QH

ECA18
ECA18 Haplotypes

- 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

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Type 1 %</th>
<th>Type 2A %</th>
<th>Type 2B %</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intron 1 SNP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT</td>
<td>21.4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>27.0&lt;sup&gt;a&lt;/sup&gt;</td>
<td>51.6&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>TC</td>
<td>18.2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>26.7&lt;sup&gt;a&lt;/sup&gt;</td>
<td>55.1&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>TT</td>
<td>14.7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>24.8&lt;sup&gt;a&lt;/sup&gt;</td>
<td>60.8&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>SINE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NN</td>
<td>20.4&lt;sup&gt;a&lt;/sup&gt;</td>
<td>27.2&lt;sup&gt;a&lt;/sup&gt;</td>
<td>52.5&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>NS</td>
<td>15.9&lt;sup&gt;b&lt;/sup&gt;</td>
<td>26.3&lt;sup&gt;a&lt;/sup&gt;</td>
<td>57.8&lt;sup&gt;ab&lt;/sup&gt;</td>
</tr>
<tr>
<td>SS</td>
<td>15.7&lt;sup&gt;b&lt;/sup&gt;</td>
<td>24.0&lt;sup&gt;a&lt;/sup&gt;</td>
<td>60.3&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Plots for the Thoroughbred, Paint, and Quarter Horse

- There are clearly more loci to investigate.
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
• 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 Haplotypes

<table>
<thead>
<tr>
<th>Breed</th>
<th>Haplotype Length (kb)</th>
<th>Haplotype Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgian</td>
<td>594</td>
<td>0.82</td>
</tr>
<tr>
<td>Clydesdale</td>
<td>594</td>
<td>0.92</td>
</tr>
<tr>
<td>Percheron</td>
<td>594</td>
<td>0.74</td>
</tr>
<tr>
<td>Shire</td>
<td>594</td>
<td>0.85</td>
</tr>
<tr>
<td>Fr Montagnes</td>
<td>594</td>
<td>0.74</td>
</tr>
<tr>
<td>N Swedish Draft</td>
<td>1548</td>
<td>0.74</td>
</tr>
<tr>
<td>Hannoverian</td>
<td>1424</td>
<td>0.40</td>
</tr>
<tr>
<td>Miniature</td>
<td>453</td>
<td>0.95</td>
</tr>
<tr>
<td>Shetland</td>
<td>645</td>
<td>0.48</td>
</tr>
<tr>
<td>Tennesse WH</td>
<td>1755</td>
<td>0.61</td>
</tr>
<tr>
<td>Caspian</td>
<td>638</td>
<td>0.50</td>
</tr>
</tbody>
</table>

*di* window

SNP position

![Haplotype Diagram](image-url)
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

ECA11
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)

<table>
<thead>
<tr>
<th>Breed</th>
<th>Haplotype Length (Mb)</th>
<th>Haplotype Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgian</td>
<td>0.61</td>
<td>1.00</td>
</tr>
<tr>
<td>Clydesdale</td>
<td>1.04</td>
<td>1.00</td>
</tr>
<tr>
<td>Shire</td>
<td>1.04</td>
<td>0.80</td>
</tr>
<tr>
<td>Percheron</td>
<td>0.61</td>
<td>0.98</td>
</tr>
<tr>
<td>Finnhorse</td>
<td>0.61</td>
<td>0.67</td>
</tr>
<tr>
<td>Swiss WB</td>
<td>0.79</td>
<td>0.68</td>
</tr>
</tbody>
</table>

---

### Diagram

- **ECA3** chromosomal region
- **di window** SNP position
- **LCORL** / **NCAPG**
- **DCAF16** / **FAM184B**
- **MED28**
- **CLRN2**
- **QDPR**
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.

![Horses running](image)
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.