

Genetic Variation in Horse Breeds Derived from Whole Genome SNP Data

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



Genetic Diversity in the Modern Horse

- Some breeds are many hundreds of years old while others have been developed relatively recently.
- Some breed registries allow outcrossing and admixture.
- Some breeds are defined by phenotype alone.
- Some populations are classified by their region of origin without active maintenance of a registry.
- Some breeds have experienced severe population bottlenecks.

Previous Studies on Genetic Diversity in the Horse

- Previous studies have examined mitochondrial DNA variation and nuclear DNA (microsatellite) variation.
- Typically a single population, sets of historically related breeds, or breeds from specific geographic regions are examined.
- We have used the Illumina 54,000 SNP genotype data to calculate measures of inter-breed and intra-breed diversity.

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

Telhisia 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



Norges veterinærhøgskole



UNIVERSITÄT
BERN



Animal Health Trust



Institute for
Conservation Research™



Swedish University of
Agricultural Sciences



UNIVERSITY OF HELSINKI



The University of Sydney



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 included.
- Removed SNPs with MAF < 0.05 , as well as ECAX SNPs.
- Removed SNPs that were in LD ($r^2 < 0.1$).
- Resulted in 6,028 SNPs for most analyses.
- On a case by case basis, individuals were removed from the analyses to reduce genome sharing as measured by pi hat values to less than 0.3.

Effective population size (N_e), individual inbreeding estimates (f), inbreeding coefficient (F_{IS}), and expected heterozygosity (H_e)

Breed	N	N_e	f	H_e	Breed	N	N_e	f	H_e
Akhal Teke	19	302	0.088	0.281	New Forest Pony	15	474	0.018	0.300
Andalusian	18	329	0.092	0.293	Norwegian Fjord	21	335	0.092	0.274
Arabian	24	346	0.106	0.280	North Swedish Horse	19	369	0.100	0.276
Belgian	30	431	0.083	0.276	Percheron	23	451	0.068	0.284
Caspian	18	351	0.015	0.292	Peruvian Paso	21	433	0.040	0.293
Clydesdale	24	194	0.263	0.225	Puerto Rican Paso Fino	20	321	0.078	0.278
Exmoor	24	216	0.227	0.242	Paint	25	399	0.048	0.289
Fell Pony	21	289	0.105	0.272	Quarter Horse	40	426	0.050	0.290
Finnhorse	27	575	0.020	0.296	Saddlebred	25	297	0.107	0.268
Florida Cracker	7	171	0.157	0.263	Shetland	27	365	0.148	0.268
Franches-Montagnes	19	316	0.079	0.279	Shire	23	357	0.183	0.252
French Trotter	17	233	0.123	0.262	Standardbred - Norway	25	232	0.152	0.255
Hanoverian	15	269	0.070	0.280	Standardbred - US	15	179	0.166	0.262
Icelandic	25	555	0.062	0.288	Standardbred - all	40	290	0.157	0.260
Lusitano	24	391	0.076	0.292	Swiss Warmblood	15	271	0.079	0.281
Maremmano	24	341	0.040	0.287	Thoroughbred - UK/Ire	19	143	0.168	0.245
Miniature	21	521	0.043	0.292	Thoroughbred - US	17	163	0.159	0.250
Mangalarga Paulista	15	155	0.237	0.228	Thoroughbred - all	36	190	0.163	0.248
Mongolian	19	751	-0.012	0.308	Tuva	15	533	0.006	0.309
Morgan	40	448	0.088	0.287	Tennessee Walking Horse	19	230	0.158	0.256

Within Breed Diversity (N_e)

- The mean effective population size (N_e) across breeds ranged from 190 to 743.
- The mean N_e in this sample collection was 341.
- N_e was lowest in the Thoroughbred (190) , but was also low in the Standardbreds and French Trotters, as well as the Clydesdale (194).
- The highest values of N_e were observed in the Mongolian (743) and Tuva (533), and also in the Icelandic (555), Finnhorse (575), and Miniature (521).

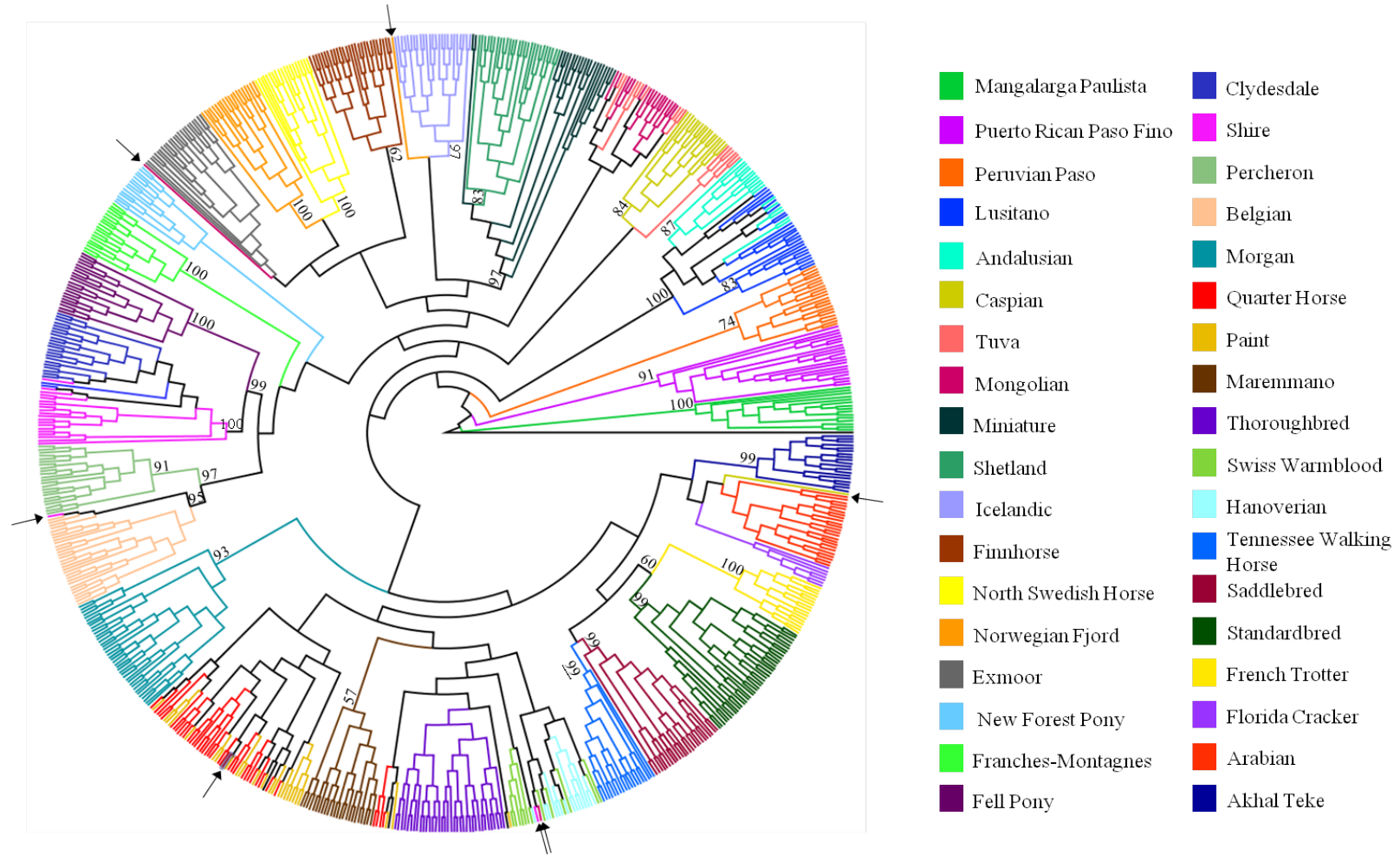
Within Breed Diversity (f)

- Inbreeding coefficients (f) were calculated for each individual based upon observed and expected heterozygosity.
- 17 individuals with estimated $f > 0.3$ included 4 Exmoor, 6 Clydesdale, 2 Mangalarga Paulista, and 1 Akhal Teke, Florida Cracker, Morgan, Shetland, and Tennessee Walking Horse.
- Consequently, average individual estimates of f were greatest in the Clydesdale, Mangalarga Paulista, and Exmoor while the lowest breed means were found in the less-actively managed populations, Warmbloods, Quarter Horse, as well as the Caspian.

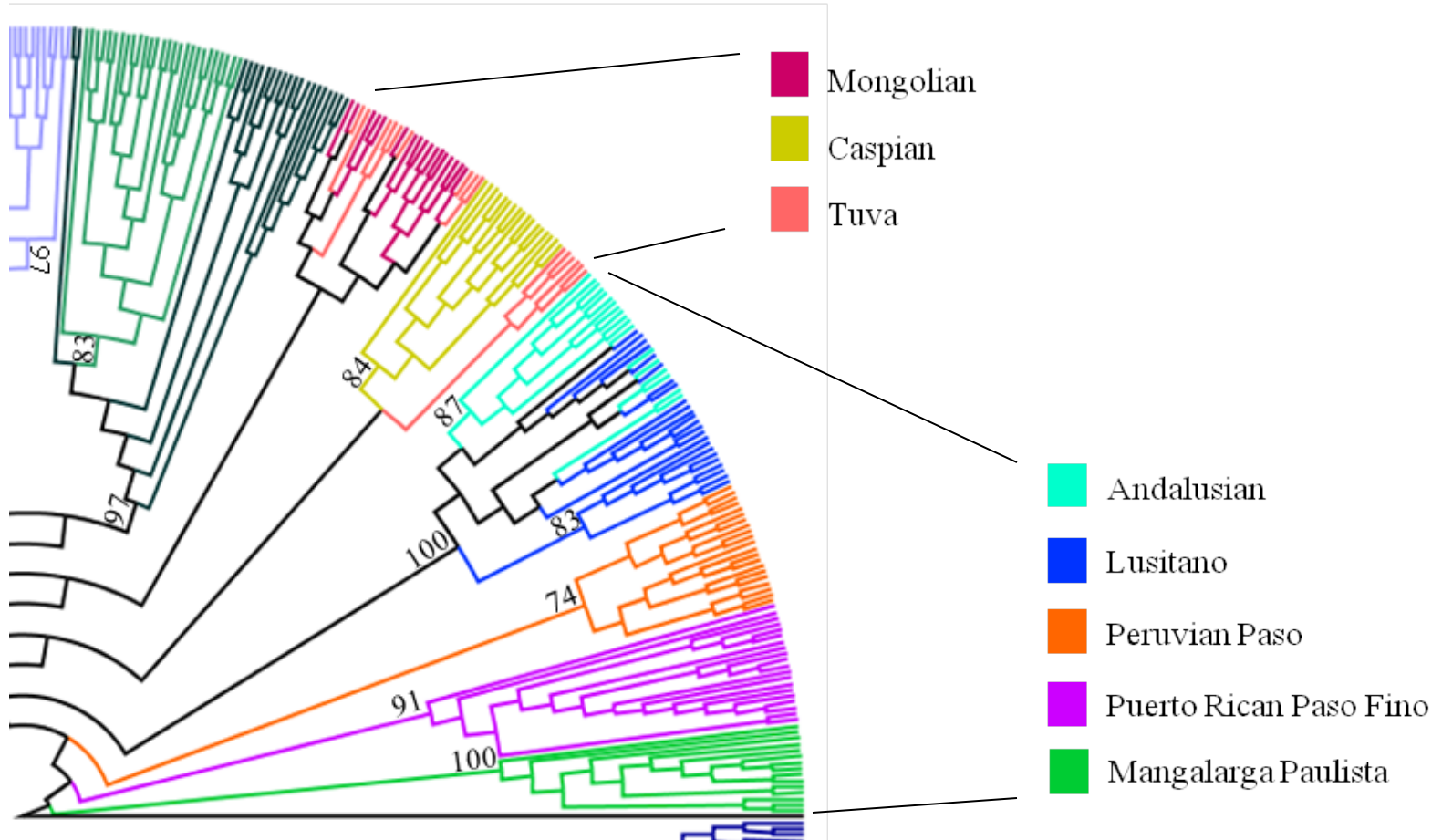
Within Breed Diversity (H_e)

- Diversity as measured by expected heterozygosity (H_e) had a mean across breeds of 0.275.
- H_e ranged from 0.309 and 0.308 in the Mongolian and Tuva, respectively, to 0.225 in the Clydesdale.
- A relatively low H_e of 0.248 was found in the Thoroughbred and in the other racing breeds (Standardbreds and French Trotters).
- A relatively high H_e was present in breeds with continuing admixture and/or large population size (Warmbloods, Quarter Horse).

Parsimony Tree Rooted by the Domestic Ass



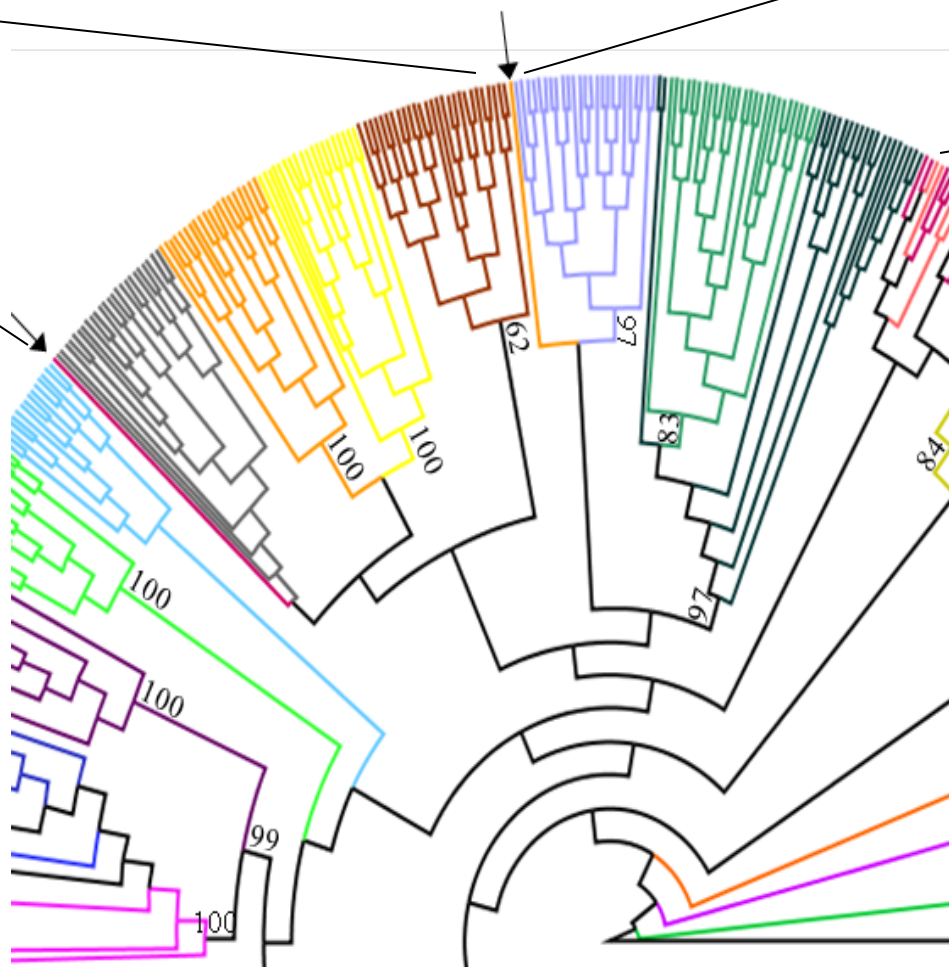
Parsimony Tree



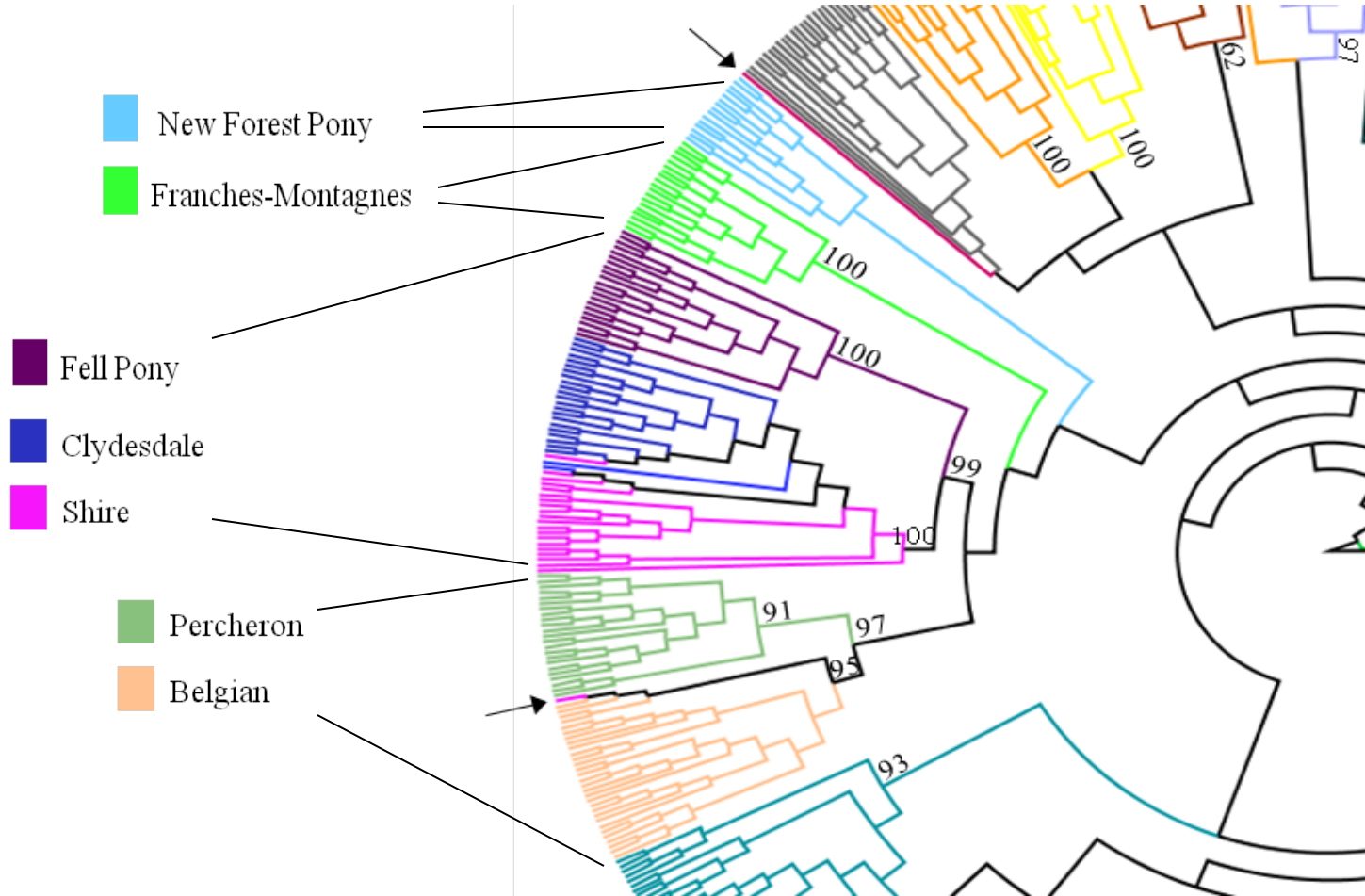
Parsimony Tree

- Finnhorse
- North Swedish Horse
- Norwegian Fjord
- Exmoor

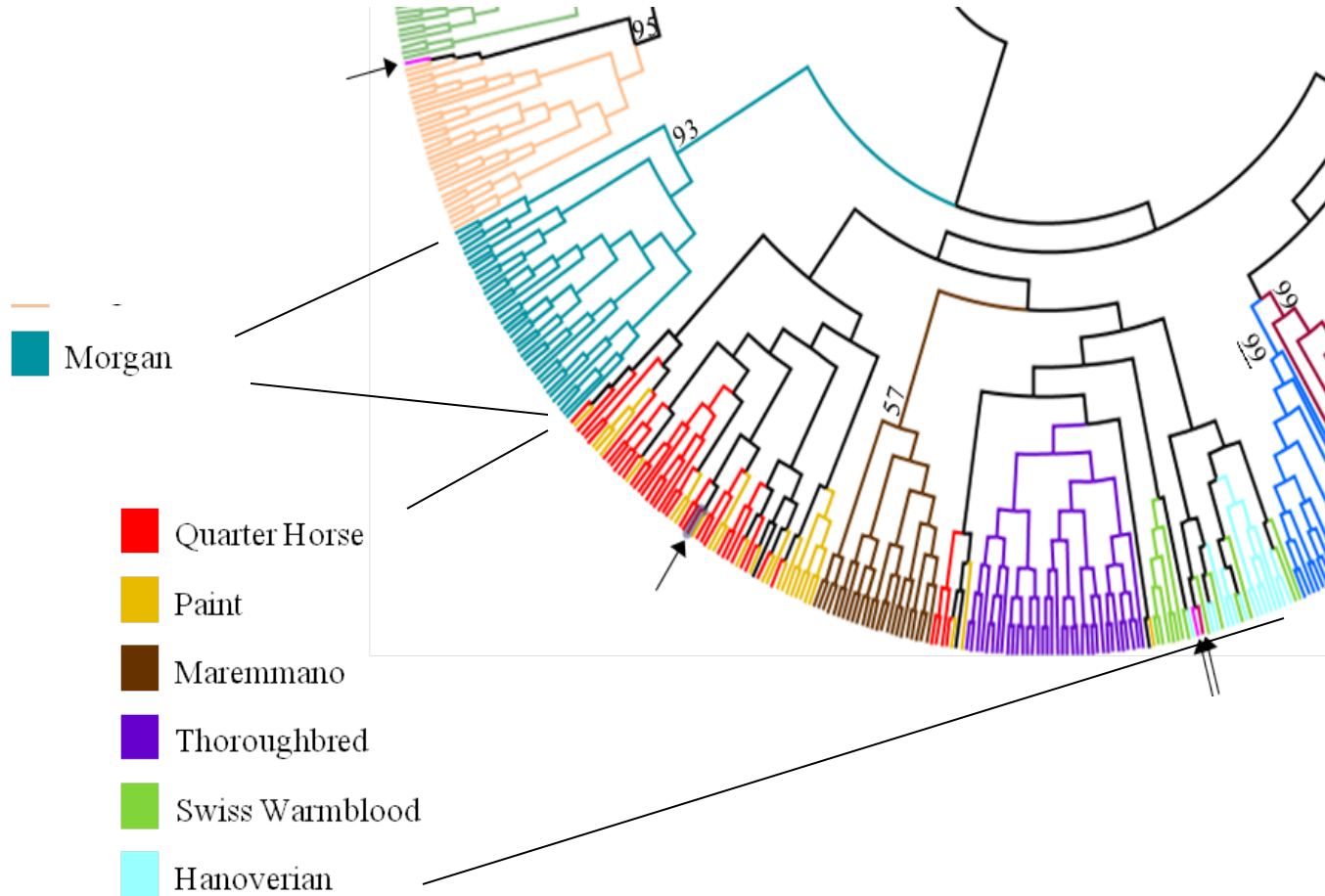
- Icelandic
- Shetland
- Miniature



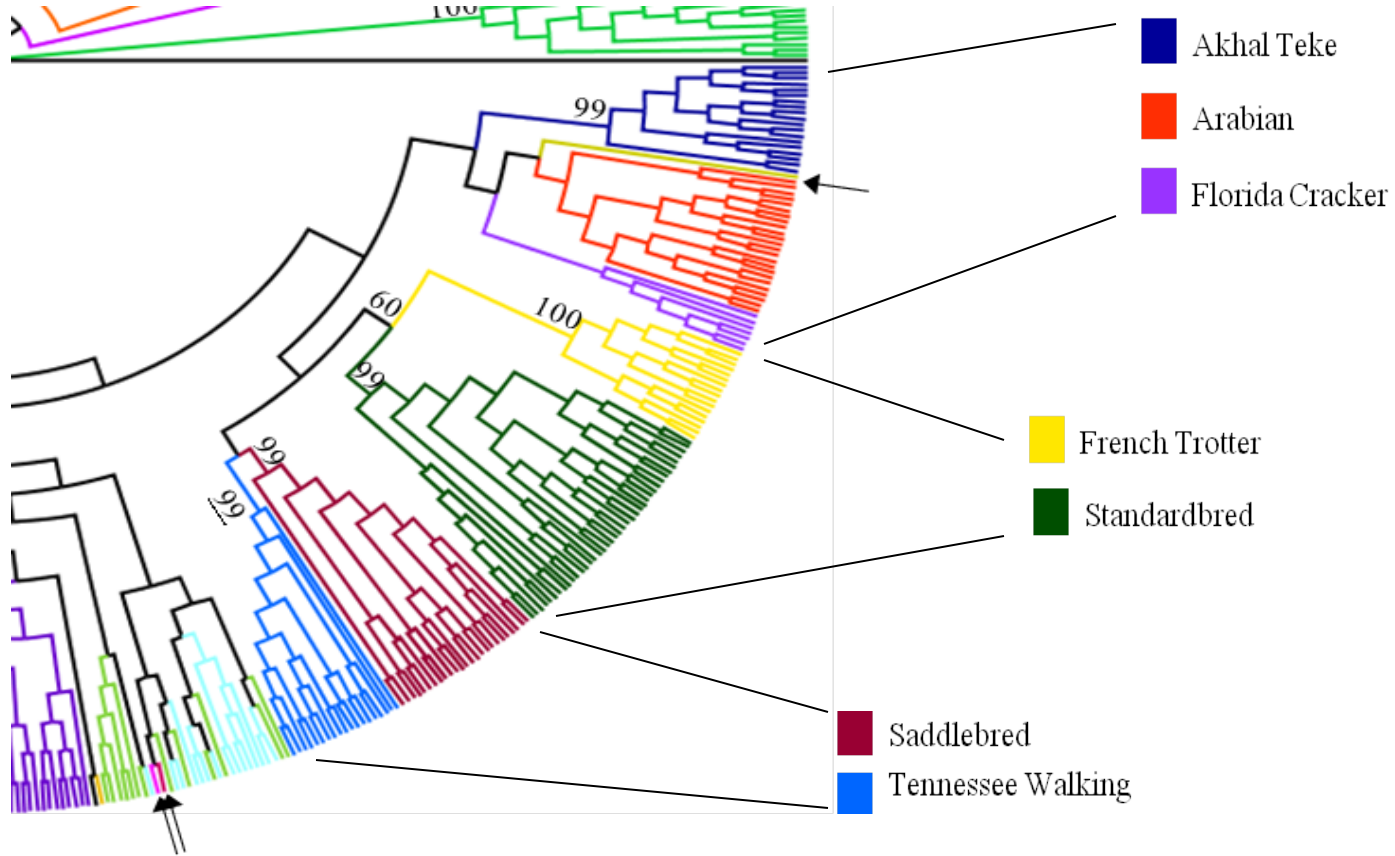
Parsimony Tree



Parsimony Tree



Parsimony Tree



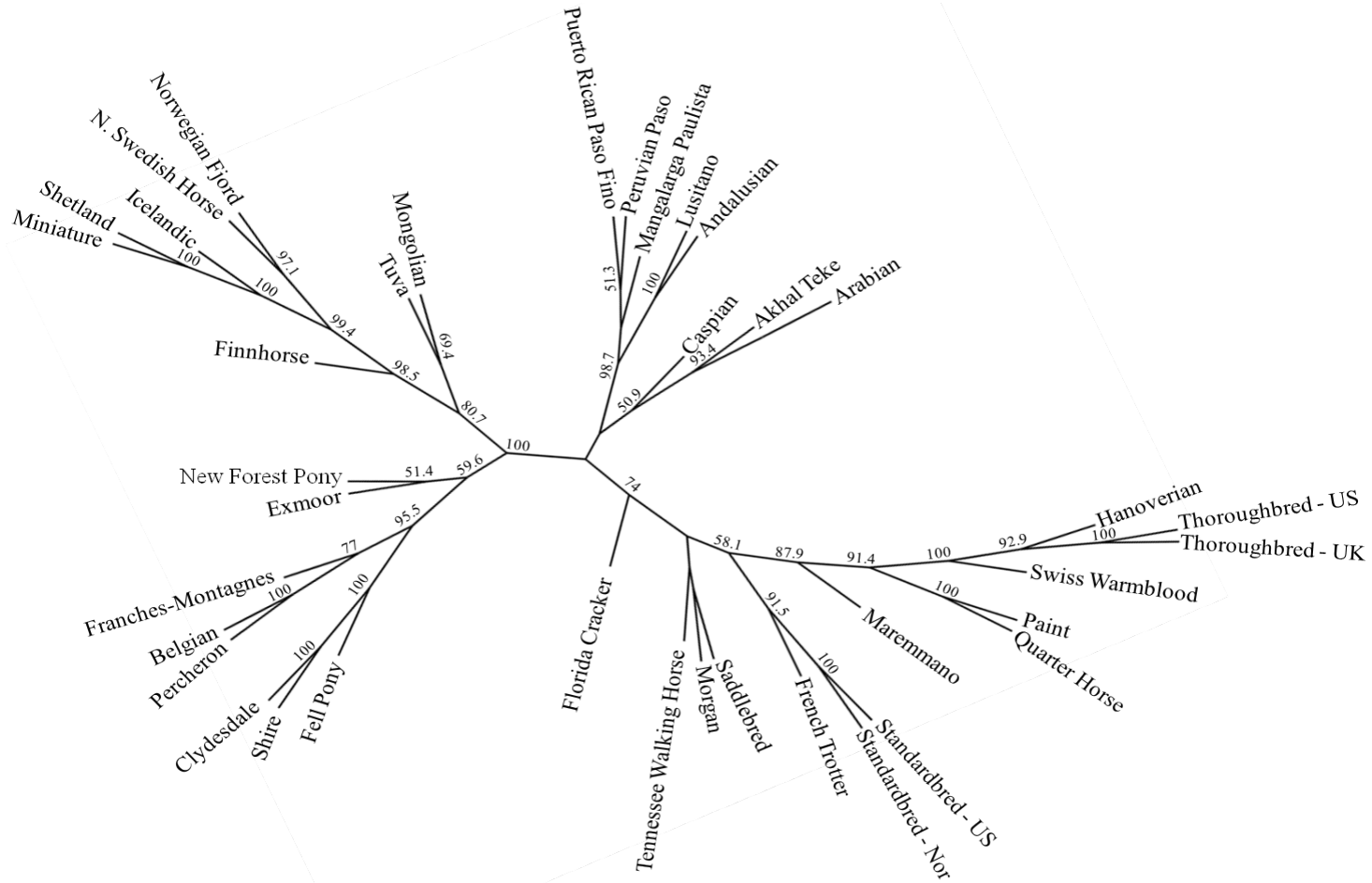
Parsimony Summary

- There is generally tight clustering and monophyly of samples within breeds.
- Major clades of the tree include:
 - Iberian breeds
 - Scandinavian breeds
 - Ponies
 - Heavy draft breeds
 - TB and related admixed breeds
 - Modern US breeds
 - Trotting breeds
 - Middle Eastern breeds

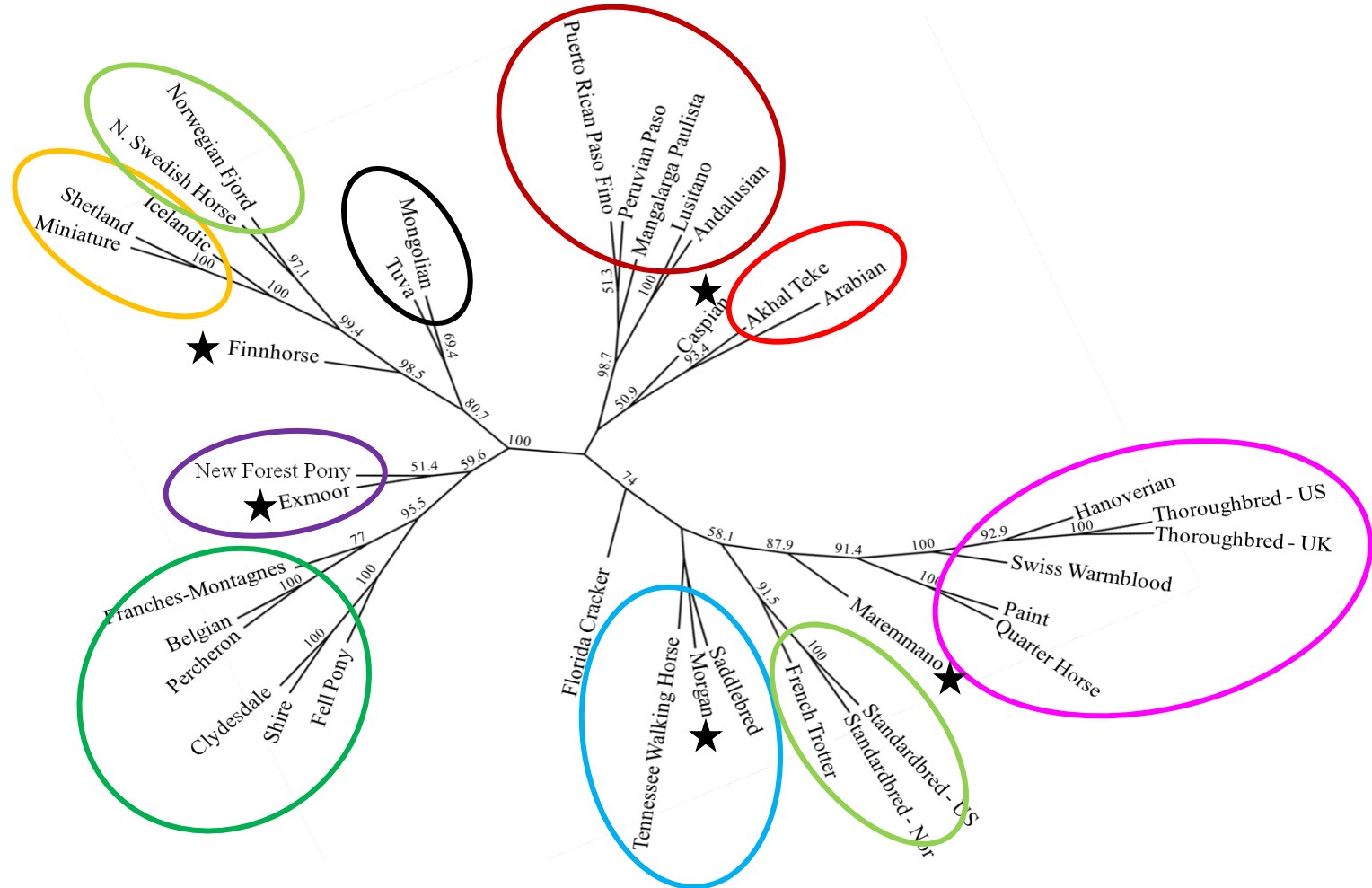
Parsimony Summary

- Some clades are not monophyletic.
 - Paint and Quarter Horse
 - Hannoverian and Swiss Warmblood
- Some individuals are not in clades with most other representatives of their breed.

Unrooted Neighbor Joining Tree



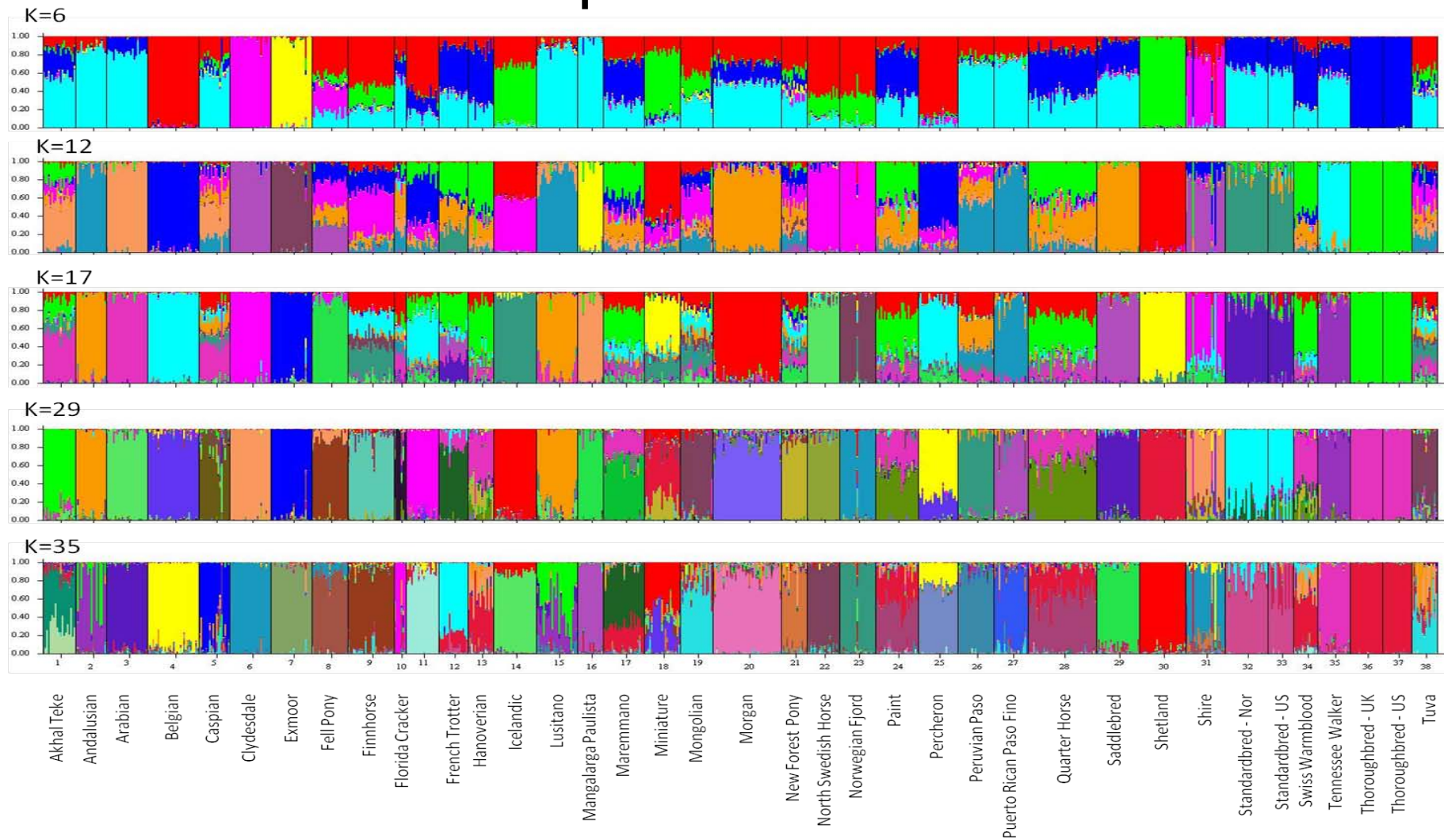
Distance Analysis



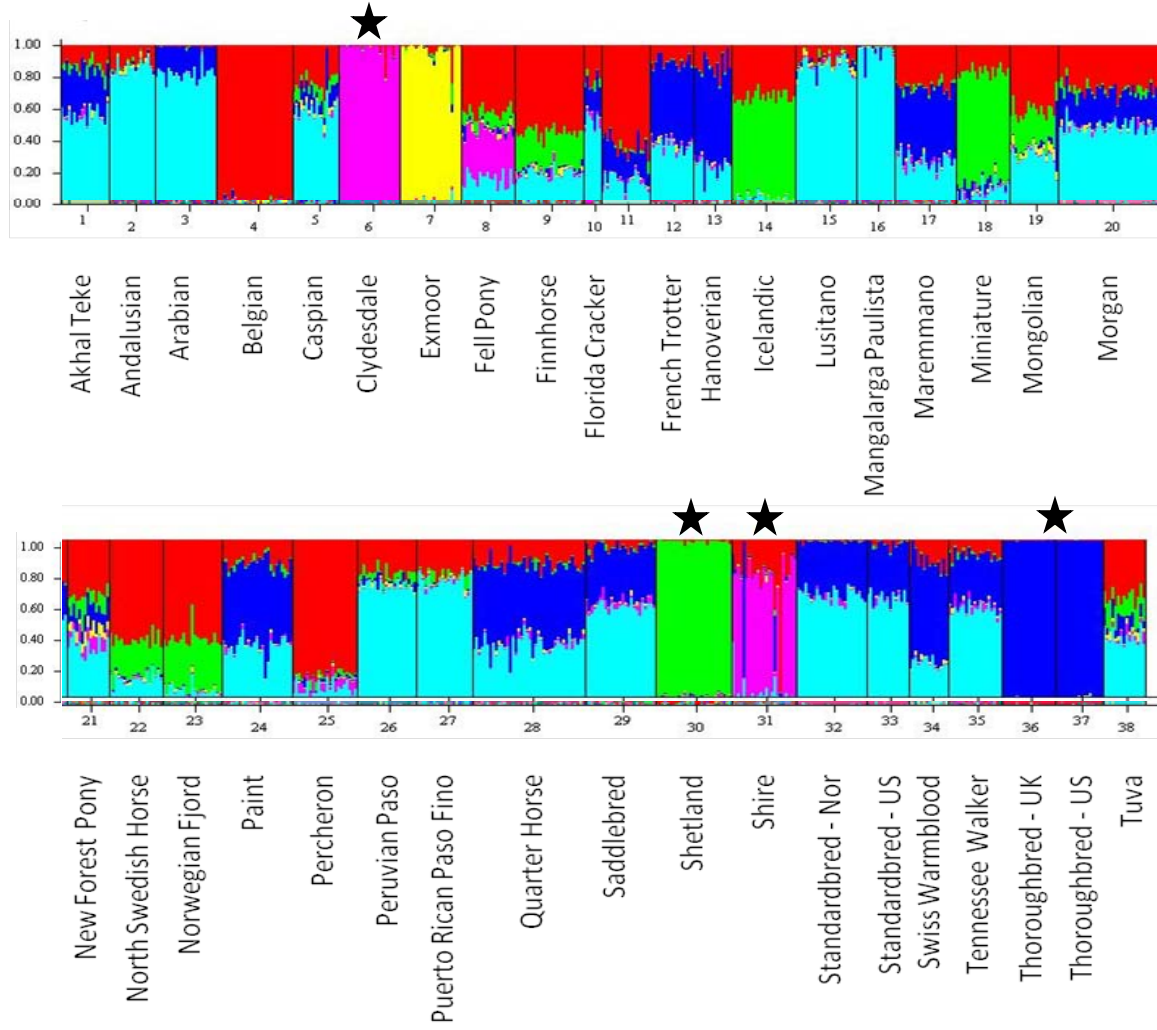
Distance Analysis Summary

- There is quite good agreement with the parsimony analysis with a few differences.
 - Caspian is now with Middle Eastern breed.
 - Finnhorse is not with Scandinavian breeds.
 - Exmoor pony is now with New Forrest pony
 - Morgan is now with Saddlebred and Tennessee Walking Horse.
 - Marremmano is not as close to the TB group.
- Strong bootstrap support for all clades except the US breeds and the Paso, Paso Fino and Paulista.

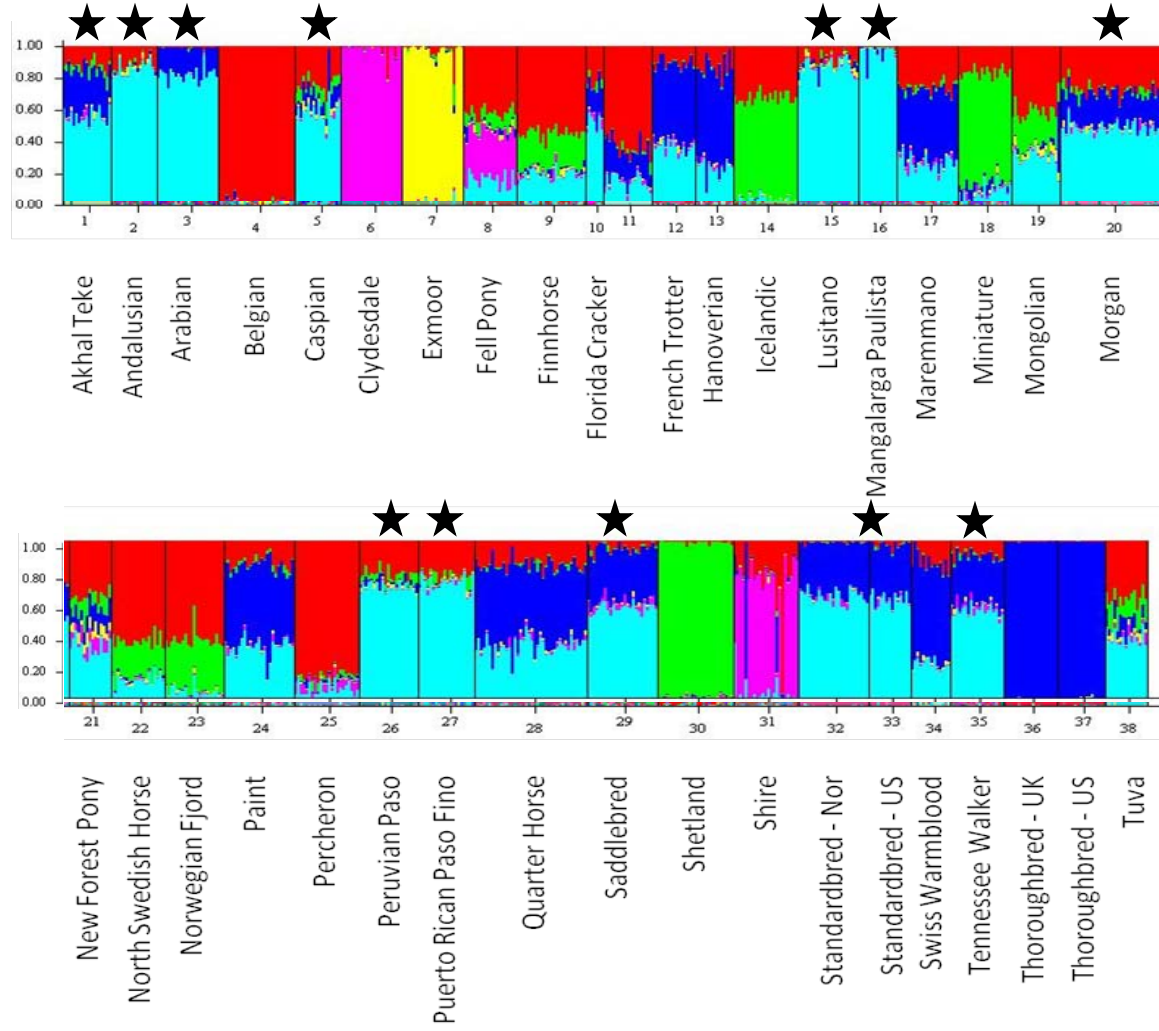
Structure Output for Five Values of K



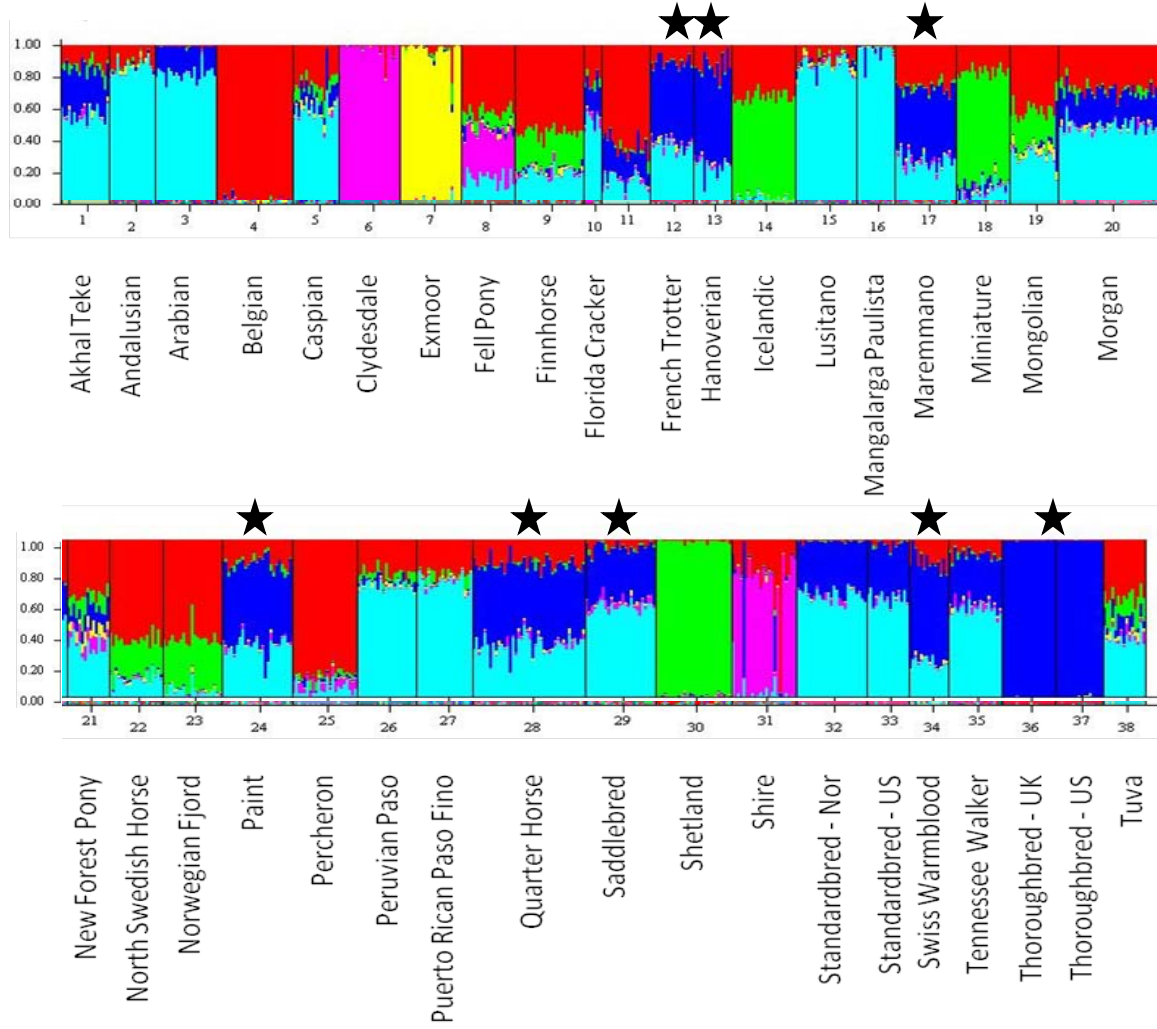
Structure for $K \leq 6$



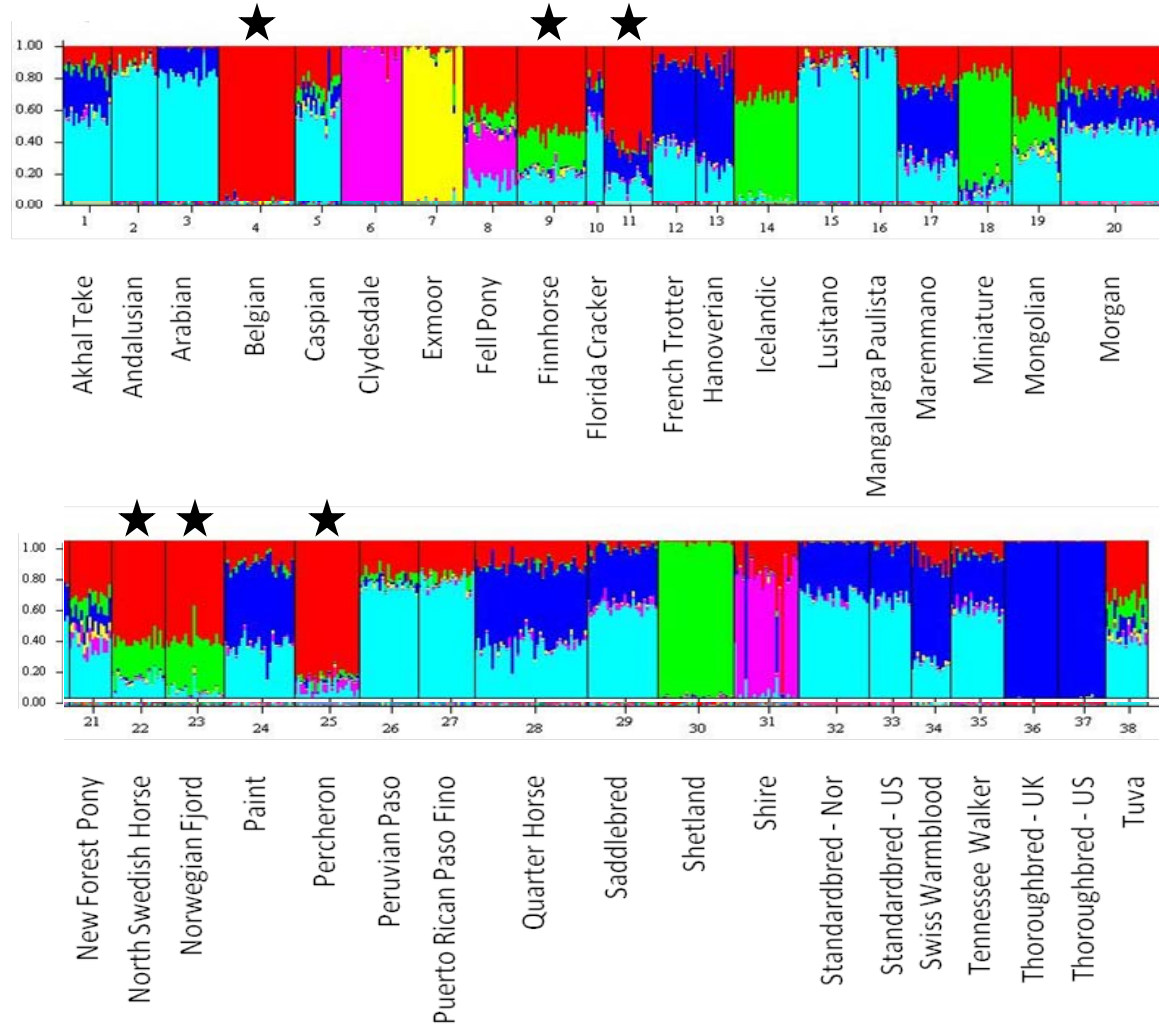
Structure for $K = 6$ (Middle East and Iberian)



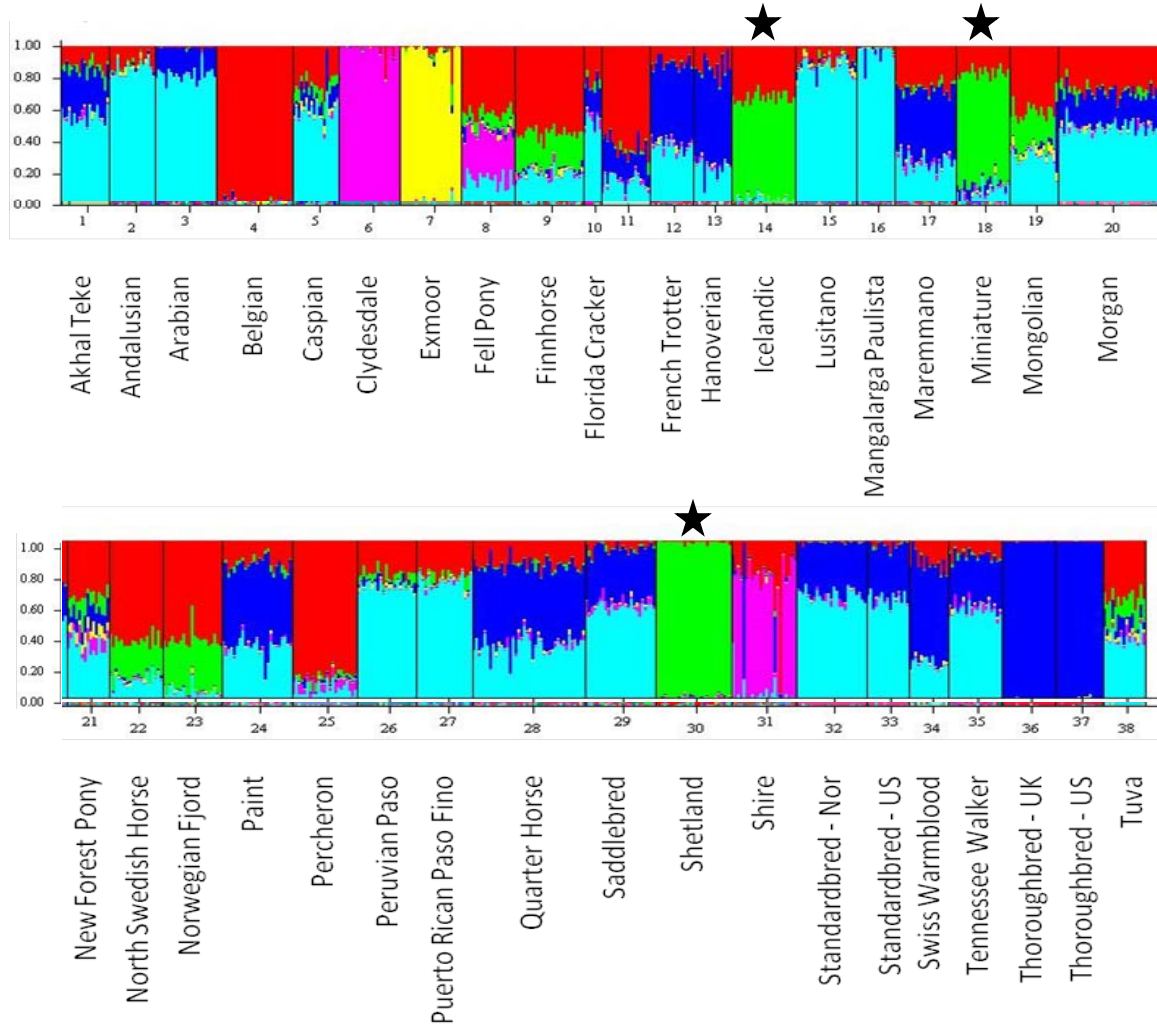
Structure for K = 6 (Thoroughbred)



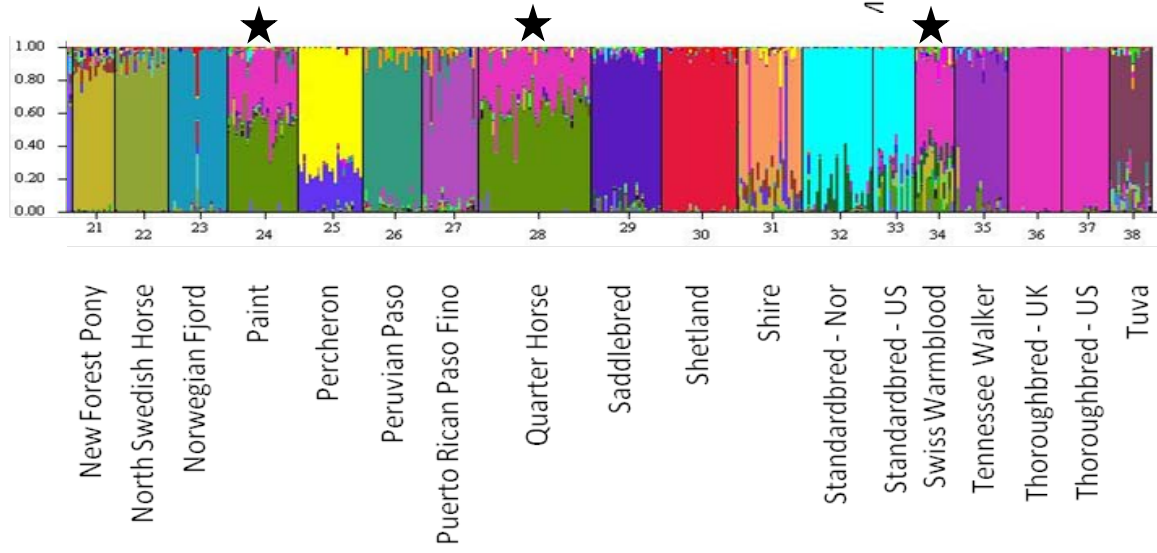
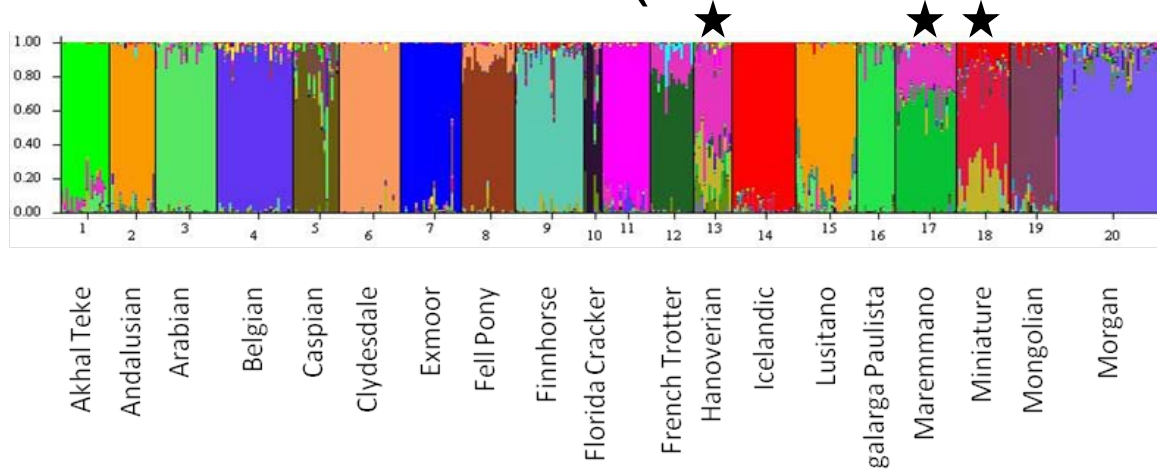
Structure for K = 6 (Scandinavian and N Europe)



Structure for K = 6 (Shetland, Icelandic)



Structure for K = 29 (Admixed Breeds)

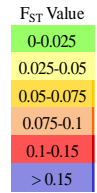
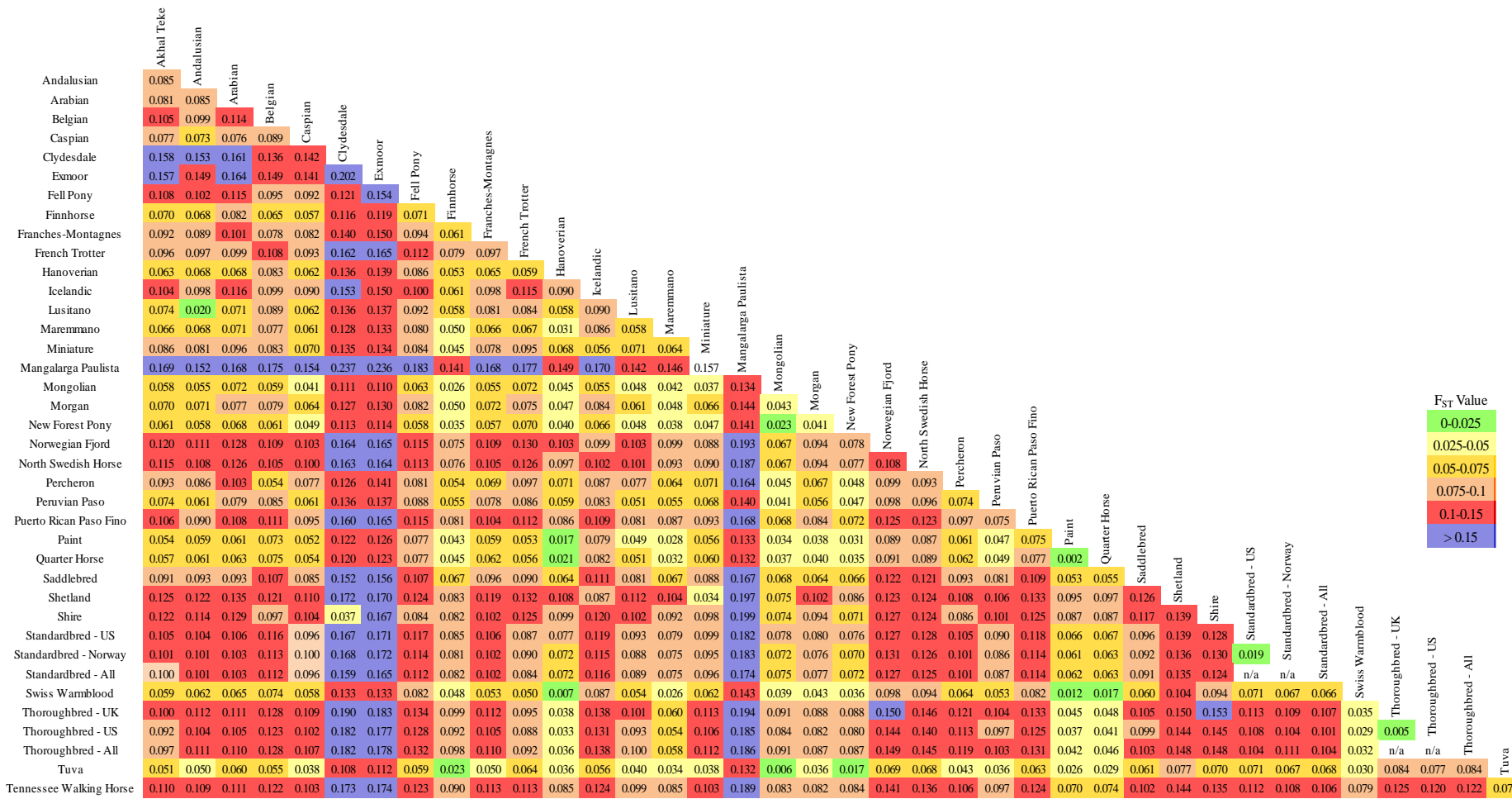


Structure Summary

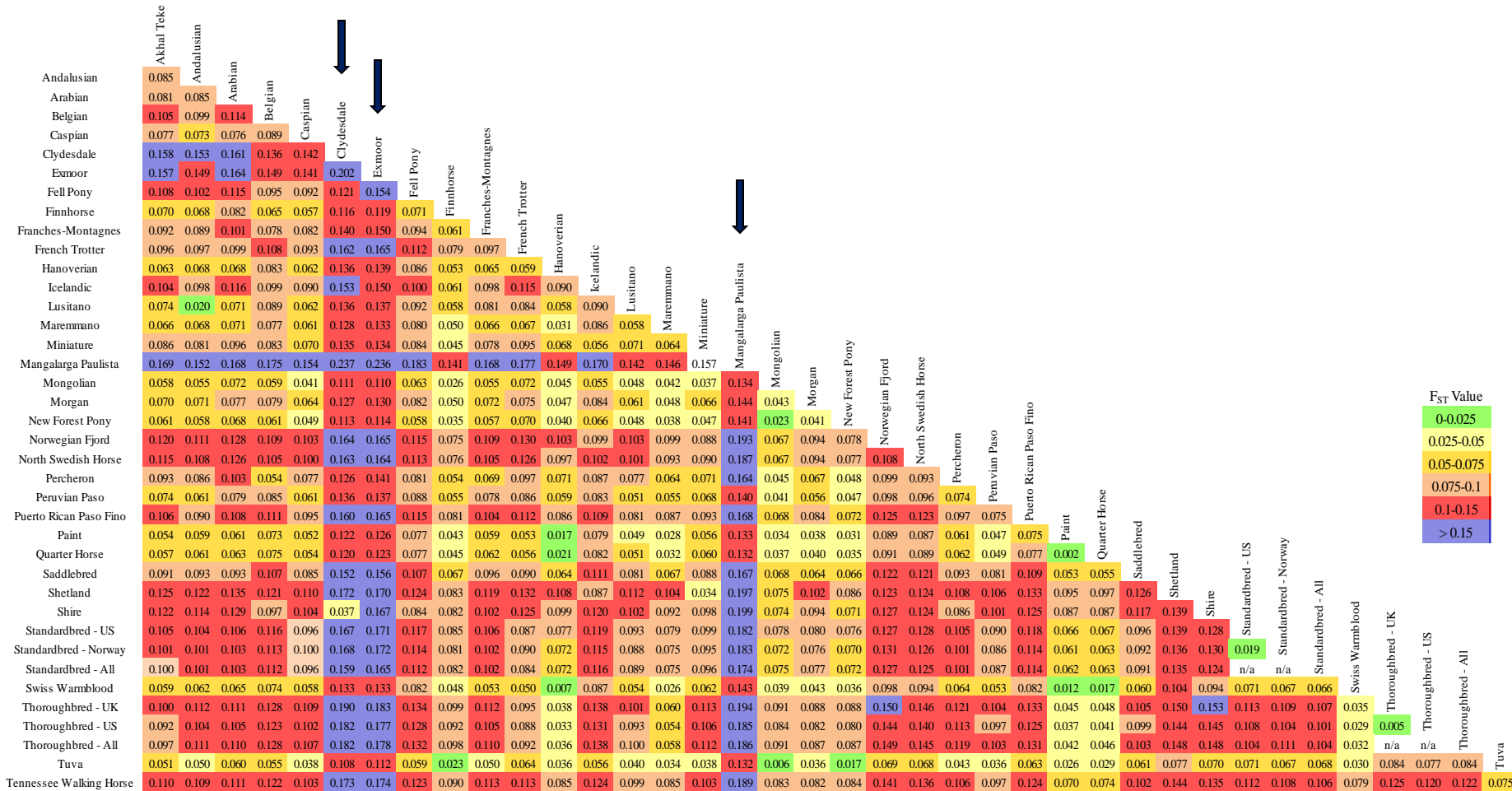
- Several breeds fall out as distinct populations with $K < 3$.
- Geographic groups are evident at $K = 6$.
- Most breeds fall out as distinct populations with $K = 29$.
- Some breeds remain clearly admixed at $K = 29$.



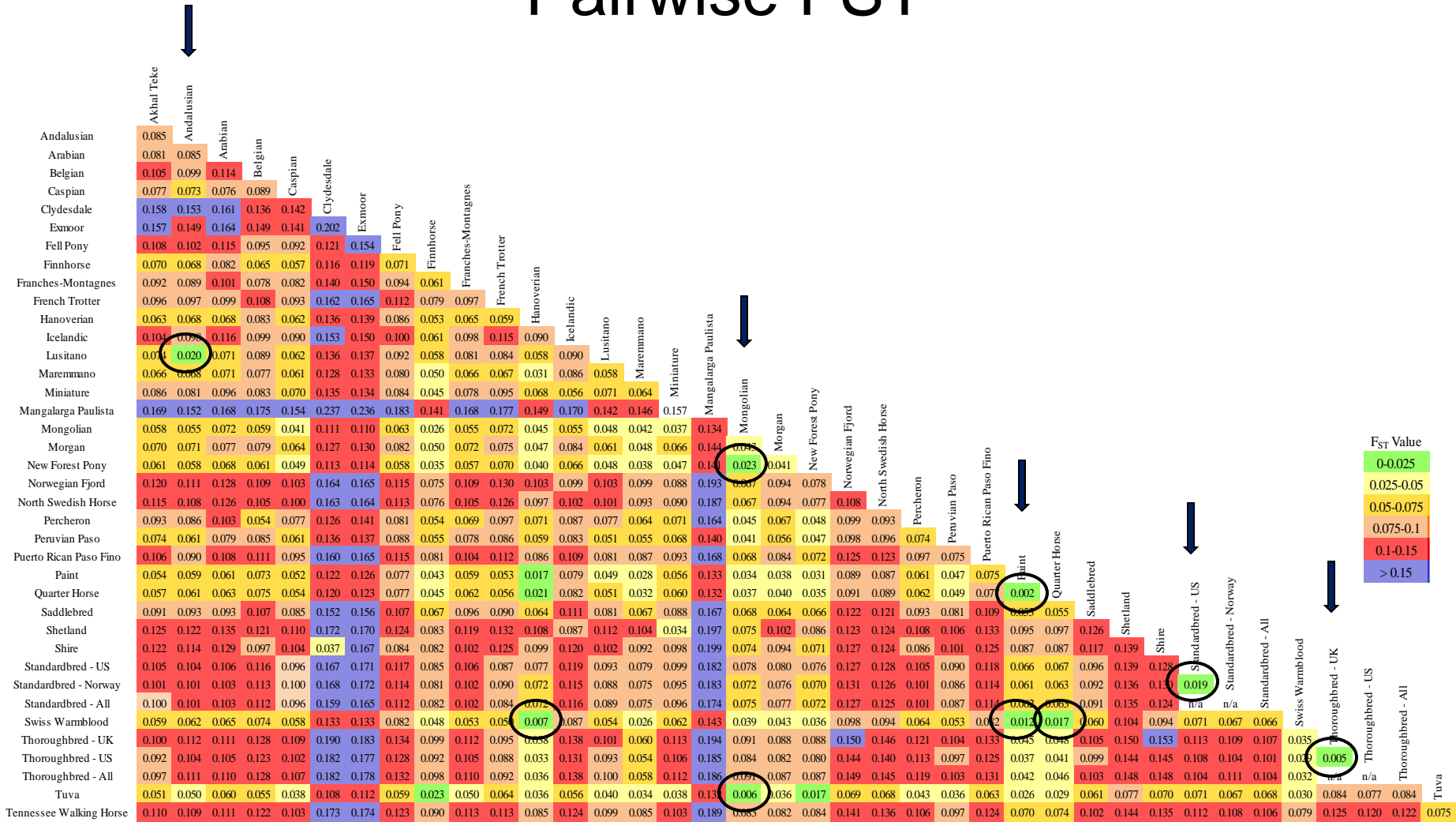
Pairwise FST Values Calculated Using 6,028 Autosomal SNPs



Pairwise FST



Pairwise FST



Conclusions

- Low levels of between population divergence were observed in breeds that are relatively early on in the process of breed development, and in those with high levels of diversity, whether due to large population size, ongoing outcrossing, or large within-breed phenotypic diversity.
- Populations with low within-breed diversity include those which have experienced population bottlenecks, have been under intense selective pressure, or are closed populations with long breed histories.

Conclusions

- F_{ST} calculations, parsimony and distance analysis demonstrated relationships among the breeds that largely reflect geographic origins and known breed histories.
- The potential of whole genome SNP analysis to define relationships between and within breeds has been demonstrated.



Mean of estimated $\ln P(X|K)$ for each of the three runs for $K=1-45$ in Structure

