



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Eidgenössisches Departement für
Wirtschaft, Bildung und Forschung WBF
Agroscope

harasnational.ch



PCA-Informative Individuals (PCA-IIs)

for Structure Identification within populations

Markus Neuditschko
Agroscope – National Stud Farm

Livestock Genomic Resources in a Changing World Conference
Cardiff, 17th of June



THE UNIVERSITY OF
SYDNEY



Key notes of the Franches-Montagnes (FM) breed



- The only indigenous Swiss horse breed
- Established in the 19th century of crossing mares with Warmblood stallions from England and France.
(Due to the demand for a decent military horse)
- Additional cross-breeding with Warmblood stallions from Sweden and France to breed a horse for family and leisure activities.
- Since January 1998 the herdbook is closed and cross-bred horses are not registered.

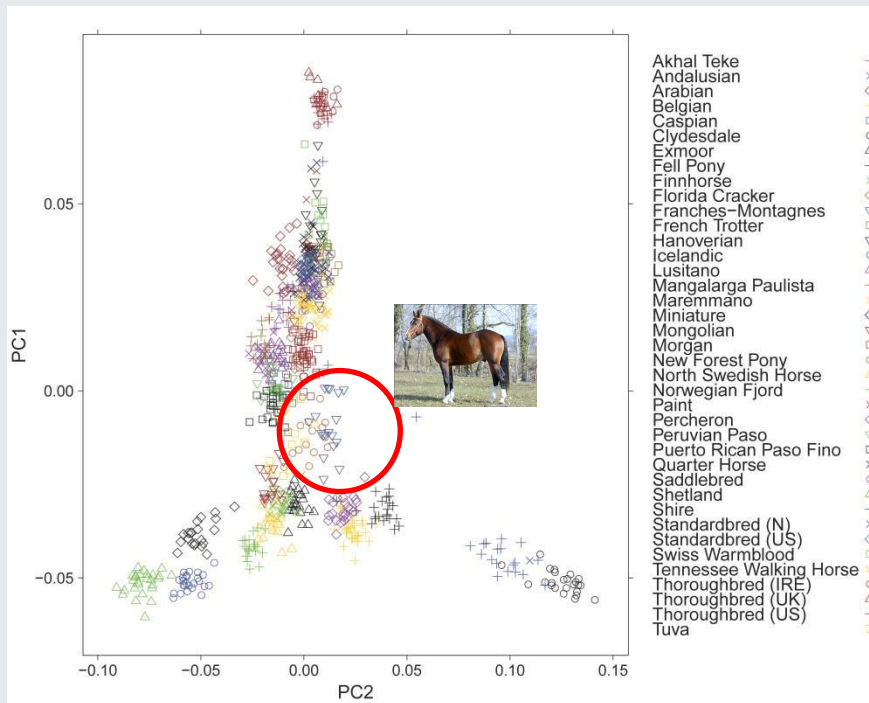




FM in the global world of horse breeds

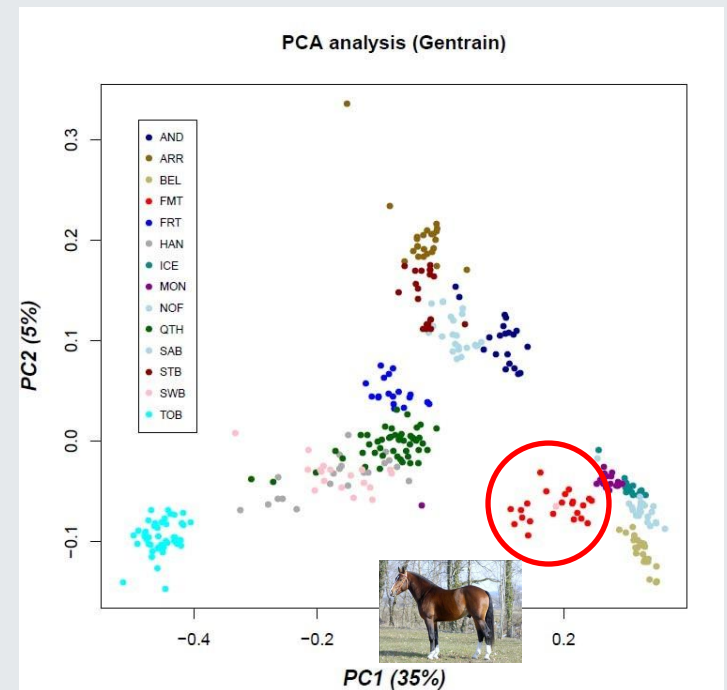
Prinicipal Component Analysis (PCA)

38 horse breeds



Genetic Diversity of the Horse

14 horse breeds



Development and Use of Equine SNP Array

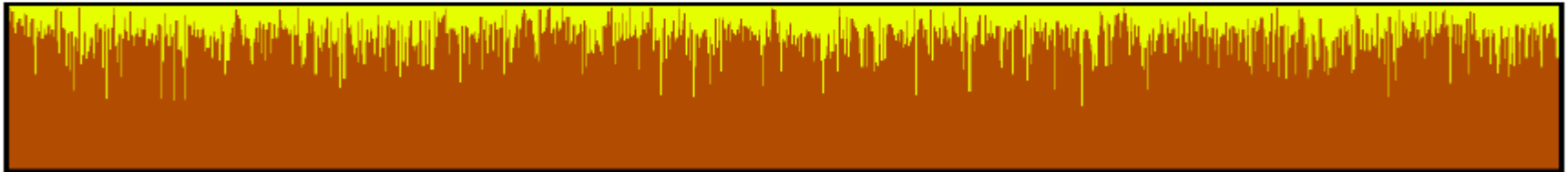


Level of admixture with Warmblood horses

■ Warmblutanteil in %

∅ Warmblutanteil in der gesamten Population = 18,20 %

Als Grundlage für diese Berechnungen wurden 1'077 Freiberger Pferde, 666 Warmblut Pferde und 31'754 SNP Genotypen berücksichtigt.



Die Berechnungen wurden mit der Software ADMIXTURE 1.22 durchgeführt.
Die Visualisierung der Ergebnisse erfolgte mit der Software DISTRUCT 1.1.



PCA-IIs to reveal fine-scale population structures

- Classical **PCA** and model based clustering (e.g. **Admixture**) can be successfully applied to investigate global population structures and the level of admixtures between populations.
- However, these methods do not provide any information of individual's contribution to population structure.
- In order to evaluate the genetic contribution of each individual within populations we developed a novel approach based on the principal definition of Paschou et al. (2007).
- **Note:** Methods outlined by Boichard (**Pedig**) and Goddard and Hayes (**REL**) solely focus on the identification of key ancestors.



The basic principles for Selecting PCA-IIs

- The method presented in this study is based on the **Singular Value Decomposition** (SVD) and requires a symmetric relationship matrix between n individuals.
- PCA-IIs are identified by calculating the correlation of all individuals with the number of significant k principal components (PCs); so called **PCA score**.
- Here, we have used **IBD** and pedigree derived relationship matrices and the empirical method **Horn's parallel analysis** to determine the number of significant PCs.





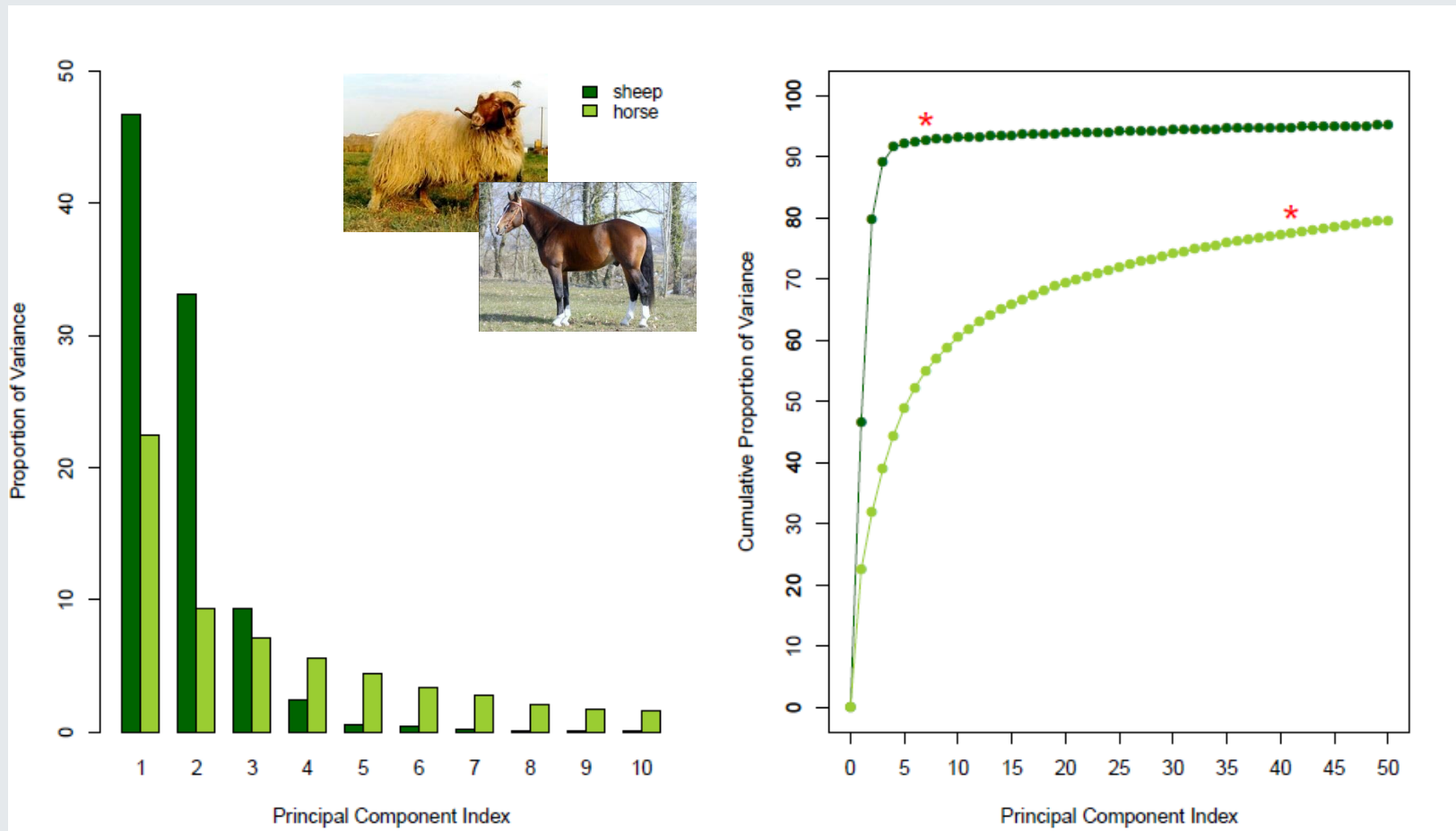
Applied datasets



- Besides the previous described horse dataset (FM) we additionally applied the method to a sheep sample.
- The sheep breed we have used describes a designed 2 breed intercross sheep resource folk, where the formation of population has been staged in three phases:
 - (1) Heterozygous F1 males and females were created by crossing **4 Awassi** founder sires with **30 wool Merino** ewes.
 - (2) **F1 sires** representing each of the founder (F0) were selected and back-crossed to Merino ewes. In total **400, 150, 150** and **150** progeny were born for each for each of the F1 sires.
 - (3) Back cross ewes were being mated to F1 sires and F1 ewes and sires were inter-crossed to produce F2 progeny.

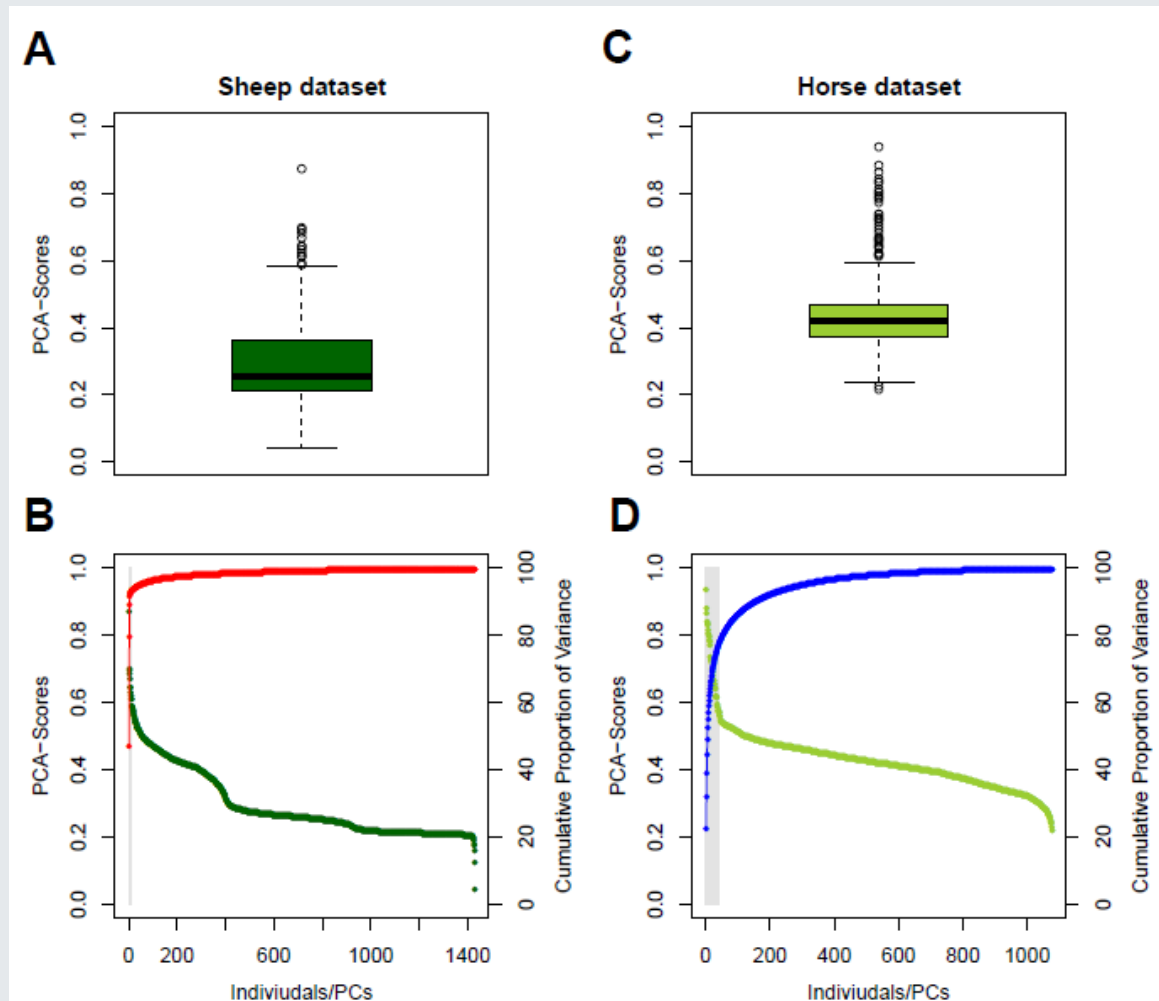


Significant Principal Components





PCA Scores of the individuals





Results (Screening PCA-IIs)

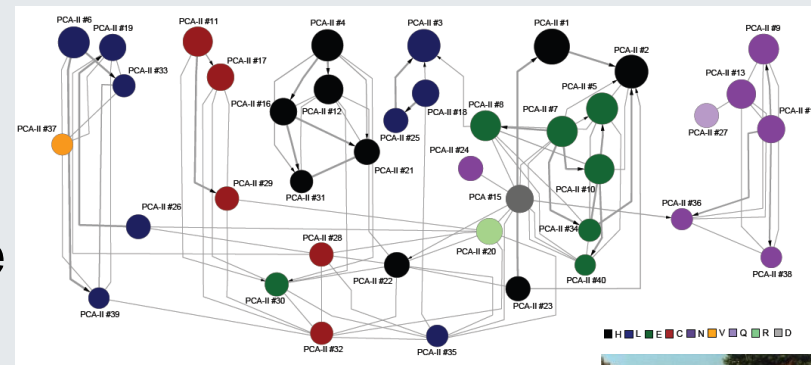
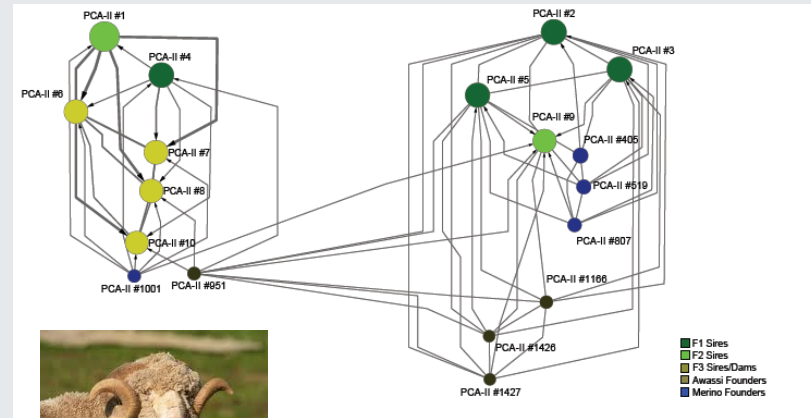
- Within sheep the most informative individual is an inbred ram descending from the F1 foundation sire with the most progeny in the population (400) followed by the 4 F1 foundation sires.
- Screening the top 20 PCA-IIs within the horse population we noticed that:
 - (1) Only stallions are in the top ranking.
 - (2) Top PCA-IIs descending from 4 different lineages (H,E,L,N).
 - (3) No stallions from three lineages (Don, Q and V).
 - (4) Many father son relationships, especially between stallions descending from lineages (H and E).
- Considering the top 40 PCA-IIs stallions descending from all lineages are in the top ranking as well as important breeding mares.





Visualization of PCA-IIs within populations


- To visualize PCA-IIs within populations we have applied recently published approach NETVIEW (Neuditschko *et al.* 2012).
- Using NETVIEW population structures are presented in terms of nodes, edges between nodes and thickness of edges.
- In the final network presentation the **node size** is associated with the **information score** and the **direction** of edges with **ancestry information**.





Conclusions

- The method allows a successful identification of influential individuals using any kind of relationship matrices.
- The combination of PCA-IIs with high definition network analysis allows the accurate identification of key ancestors without the need of individual ancestry.
- Especially useful to investigate population structures in indigenous breeds and wild species, where ancestry information is often lacking or not available.
- Useful for assembling resource populations to facilitate accurate genotype imputation across and within populations.



Thank you for
your attention!