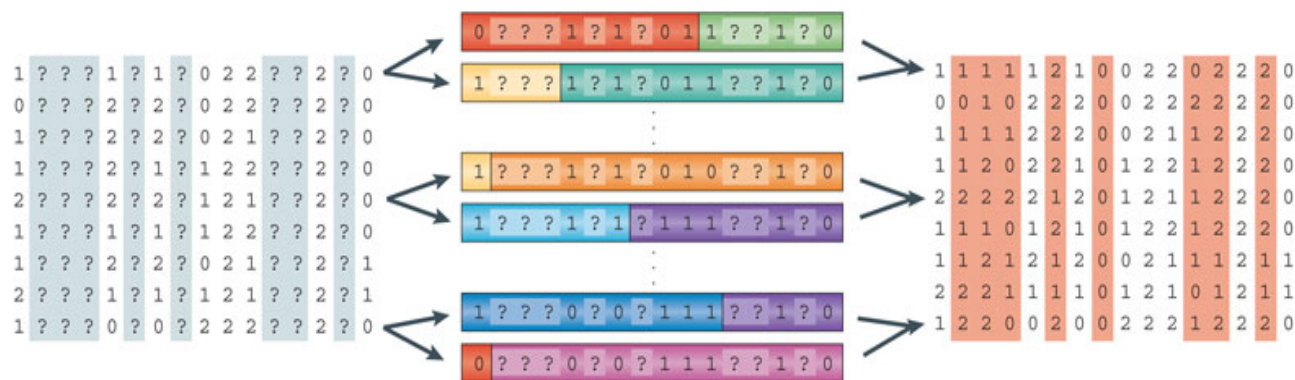




Update zum Imputieren von SNPs



Markus Neuditschko

6. Juni 2018, SABRE-TP Meeting, Zürich



Rückblick

- Aktuelle Methoden für die Selektion von informativen Tieren für die Re-Sequenzierung beschränken sich ausschließlich auf einflussreiche Ahnen (“key ancestors”).
- Diese Selektionsstrategie kann dazu führen, dass die Genotypen Phasen der Referenzpopulation nicht genau abgeleitet werden kann.
- Um die Genotypen Phasen in der Referenzpopulation zu verbessern, haben wir eine neue Methode entwickelt, welche neben den wichtigsten Ahnen auch einflussreiche Nachkommen berücksichtigt.



Graphic





RESEARCH ARTICLE

Identification of key contributors in complex population structures

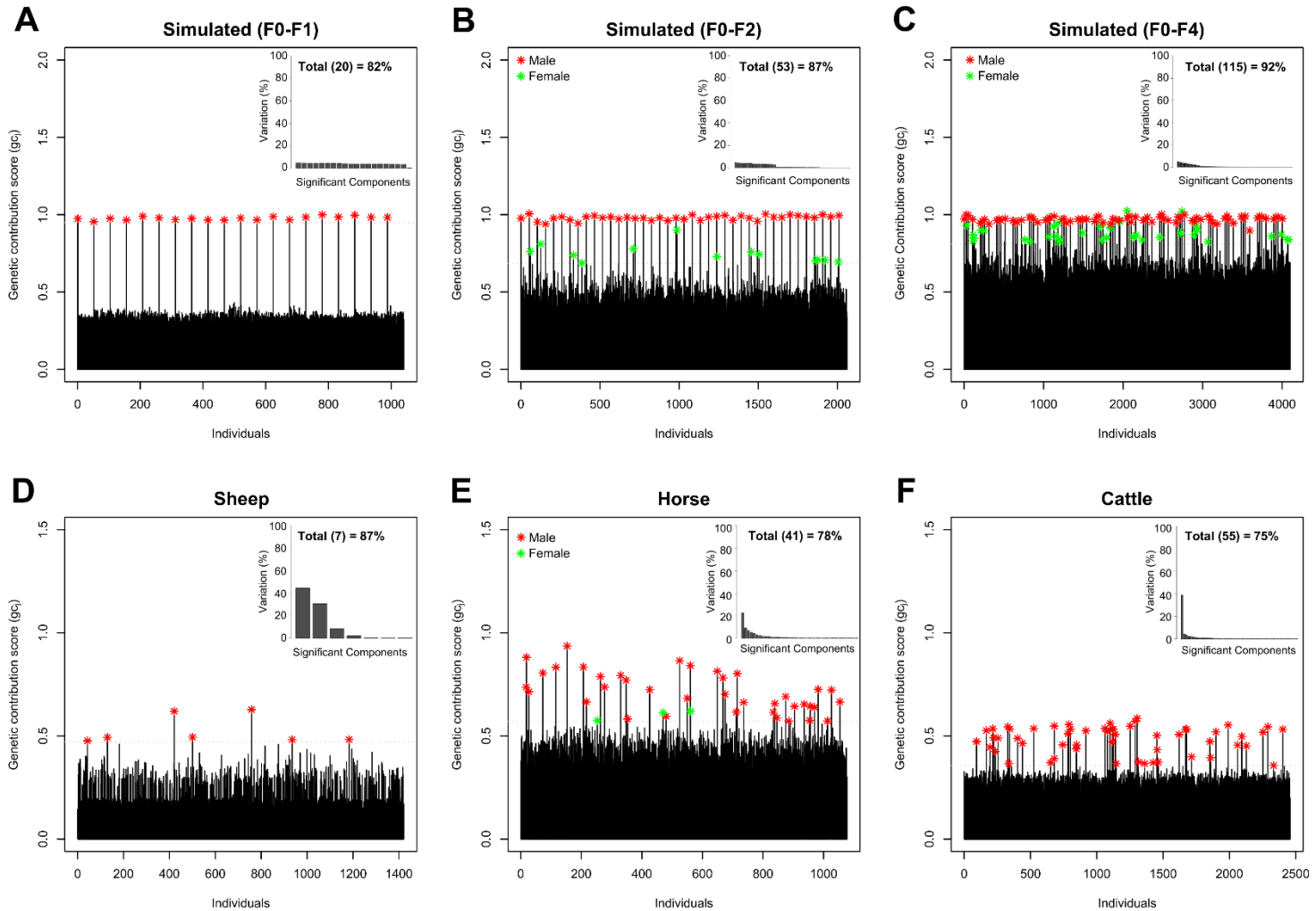
Markus Neuditschko^{1,2*}, Herman W. Raadsma², Mehar S. Khatkar², Elisabeth Jonas^{2,3}, Eike J. Steinig⁴, Christine Flury⁵, Heidi Signer-Hasler⁵, Mirjam Frischknecht^{1,6}, Ruedi von Niederhäusern¹, Tosso Leeb⁶, Stefan Rieder¹

1 Agroscope, Swiss National Stud Farm, Avenches, Switzerland, **2** Reprogen – Animal Bioscience Group, Faculty of Veterinary Science, University of Sydney, Camden, Australia, **3** SLU, Department of Animal Breeding and Genetics, Uppsala, Sweden, **4** College of Marine and Environmental Sciences, James Cook University, Townsville, Australia, **5** School of Agricultural Forest and Food Sciences, Bern University of Applied Sciences, Zollikofen, Switzerland, **6** Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland

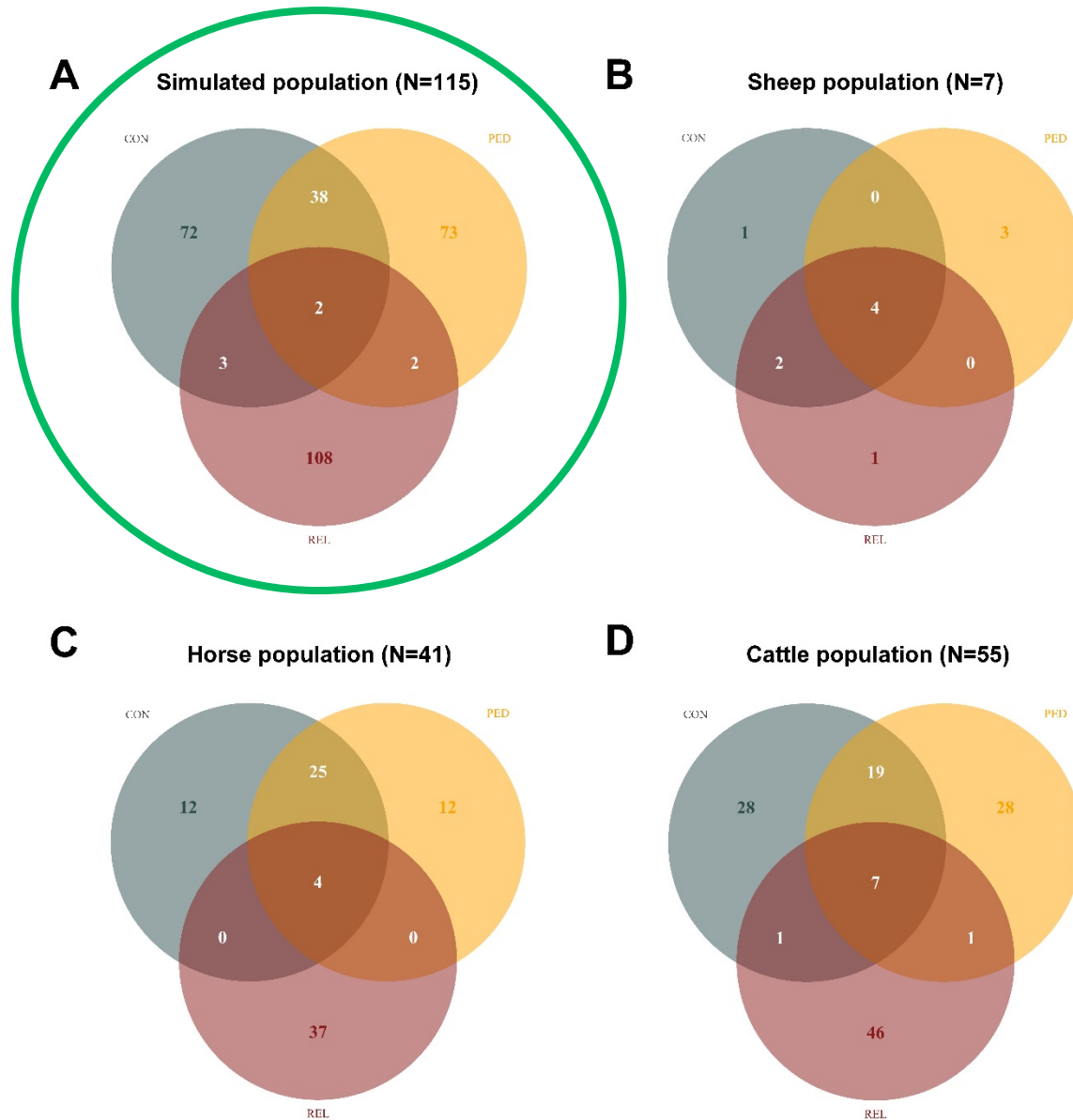
* markus.neuditschko@agroscope.admin.ch

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0177638>

Identifikation einflussreicher Zuchttiere



Identifikation einflussreicher Zuchttiere



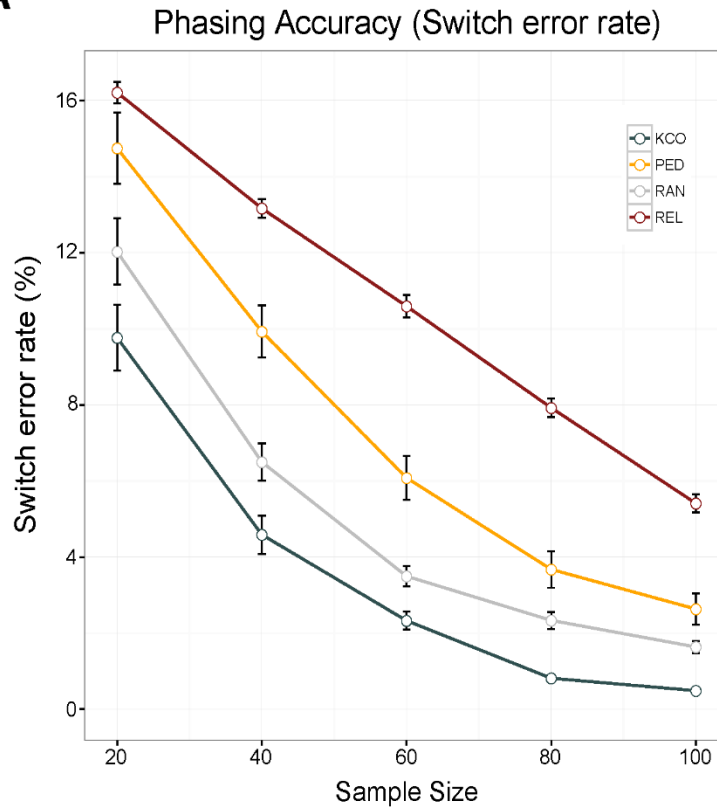
Phasen und Imputations Genauigkeit

- Nach der Auswahl der Referenzpoulation (20 – 100), wurden die abgeleiteten Haplotyp Phasen mit den tatsächlichen Phasen verglichen.
- Die Phasen Genauigkeit wurde anhand der Switch-error-rate ermittelt.
- Für die Berechnung der Imputations Genauigkeit wurden 20% der SNP Genotypen (2'000 SNPs) auf fehlend („missing“) zurückgesetzt, wobei die fehlenden SNPs gleichmäßig über 5 MAF Klassen verteilt waren.
- Die Berechnung der Imputations Genauigkeit erfolgte anhand der Genotype Concordance Rate und der Korrelation zwischen wahren und imputierten SNP Genotyp.

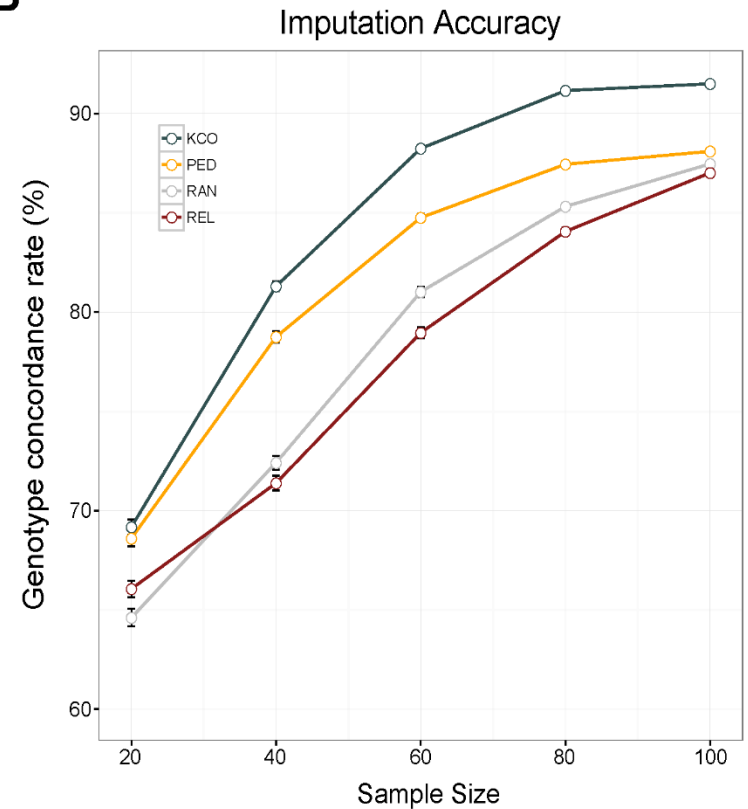


Phasen und Imputations Genauigkeit

A



B



Imputations Genauigkeit MAF Klassen

MAF Klassen	CON	PED	REL	RAN
MAF $0 < 0.1$	84.19%	73.54%	66.04%	72.16%
MAF $0.1 < 0.2$	85.40%	77.56%	73.77%	77.44%
MAF $0.2 < 0.3$	86.84%	79.74%	77.71%	78.81%
MAF $0.3 < 0.4$	87.87%	81.93%	80.28%	80.81%
MAF $0.4 < 0.5$	87.70%	82.46%	80.85%	80.68%



Schlussfolgerung und Ausblick

- Neben der Genotypen Phasen Genauigkeit der Referenzpopulation auch die die Imputations Genauigkeit der Zielpopulation wesentlich verbessert werden kann.
- Vor allem die Imputations Genauigkeit von „rare variants“ ($MAF < 0.1$) signifikant (10%) erhöht werden kann.
- Weitere Analysen sind geplant, welche folgende Aspekte berücksichtigen sollen:
 - wahre Populationsstrukturen (Selle Francais Daten)
 - weitere Selektionsstrategien (Hickey et al. 2017)
 - weitere Imputationsalgorithmen (Beagle 4.2 und MACH)



Danke für Eure Aufmerksamkeit!



Agroscope Swiss national Stud Farm