Published in final edited form as:
Anim Genet. 2014 February ; 45(1): 153. doi:10.1111/age.12093.

Validation of imputation between equine genotyping arrays

A. M. McCoy and **M. E. McCue**
Veterinary Population Medicine Department, University of Minnesota, St. Paul, MN, 55108, USA

Background

Two genotyping arrays are available for the horse, containing ~54 000 and ~65 000 markers, of which only ~45 000 are shared. This leads to a loss of information when combining datasets generated on separate arrays. Genotype imputation offers a potential solution to this problem. Our objective was to assess the accuracy of genotype imputation for the two equine genotyping arrays across scenarios constructed to examine factors previously reported to affect imputation success in domestic animals and humans, including imputed population size, reference population size, reference population makeup (similar to or different from the imputed population) and length of shared haplotype blocks (linkage disequilibrium, LD).[1,2]

Methods

Genotypes from 248 horses of three breeds [Quarter Horse (QH), *n* = 143; Standardbred (STB), *n* = 72; Thoroughbred (TB), *n* = 33] genotyped on the Illumina Equine SNP70 BeadChip were 'masked' down to the 45 703 markers shared by the SNP70 and SNP50 chips and subsequently imputed back to the complete marker set for five chromosomes (ECA 1, 6, 15, 26 and X) using BEAGLE[3] with default settings (Appendix S1, Fig. S1). Additionally, 30 QH genotyped on the SNP50 had their genotypes masked and imputed, using a reference population of 280 horses from 13 diverse breeds.

Results/Conclusions

Results for 20 SNP70 scenarios are summarized in Table S1. Overall, mean imputation success was 94.8% (individual horse range 82.2–100%). Generally, ECA 1, 15 and 26 performed better than did ECA 6 and X. For ECA 6, this may be partly due to the fact that a

Correspondence: A. M. McCoy (mccoy134@umn.edu).

Supporting information
Additional supporting information may be found in the online version of this article.
Appendix S1 Supplemental text.
Figure S1 Complete pipeline for imputation of equine genotyping data.
Figure S2 Mean imputation success with an imputed population *n* = 10 across a range of reference population sizes (*n* = 20–100) for each of three breeds [Quarter Horse (QH), red squares; Standardbred (STB), blue circles; Thoroughbred (TB), green triangles].
Figure S3 Venn diagram of marker overlap between the Illumina Equine SNP50 and SNP70 beadchips.
Table S1 Summary of SNP70 validation scenario results.
Table S2 Summary of preliminary imputation data.
Table S3 Summary of SNP50 validation scenario results.

large block of imputed markers are located at the end of the chromosome and thus do not have an ideal haplotype context for imputation. Contrary to previous reports,[2] size of the imputed population did not impact imputation success. Imputation success increased with larger reference population sizes (Fig. S2) and when imputed and reference populations were breed-matched. However, large mixed breed reference populations resulted in more accurate imputation than did small breed-matched reference populations. Breeds with longer LD had higher imputation success than did those with shorter LD (TB > STB > QH; Fig. S2). These results reflect findings reported in humans.[1,4] Allelic R^2, the estimated squared correlation between the imputed allele dosage and the true allele dosage for a marker, was used as a measure of confidence for imputed genotype calls. The overall mean R^2 was 0.771 (range, 0.582–0.981). Imputation success and R^2 were highly linearly correlated ($r^2 = 0.79$). Results for the SNP50 were comparable to the SNP70 (Appendix S1). The total number of markers available for analysis after imputation was 73 200, an increase of ~27 500 markers from the set shared by the two chips. In conclusion, imputation between the two arrays was highly accurate.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Thanks to Dr. James MacLeod for TB data and Robert Schaefer for custom shell script. AMM was supported by a NIH institutional training grant (T32OD10993).

References

1. Huang L, et al. Am J Hum Genet. 2009; 84:235–250. [PubMed: 19215730]
2. Ventura R, et al. Proc PAG XXI. 2013 P0529. Available at: www.intlpag.org.
3. Browning SR, Browning GL. Am J Hum Genet. 2007; 81:1084–1097. [PubMed: 17924348]
4. Howie B, et al. G3. 2011; 1:457–470. [PubMed: 22384356]