SUPPLEMENTARY INFORMATION:

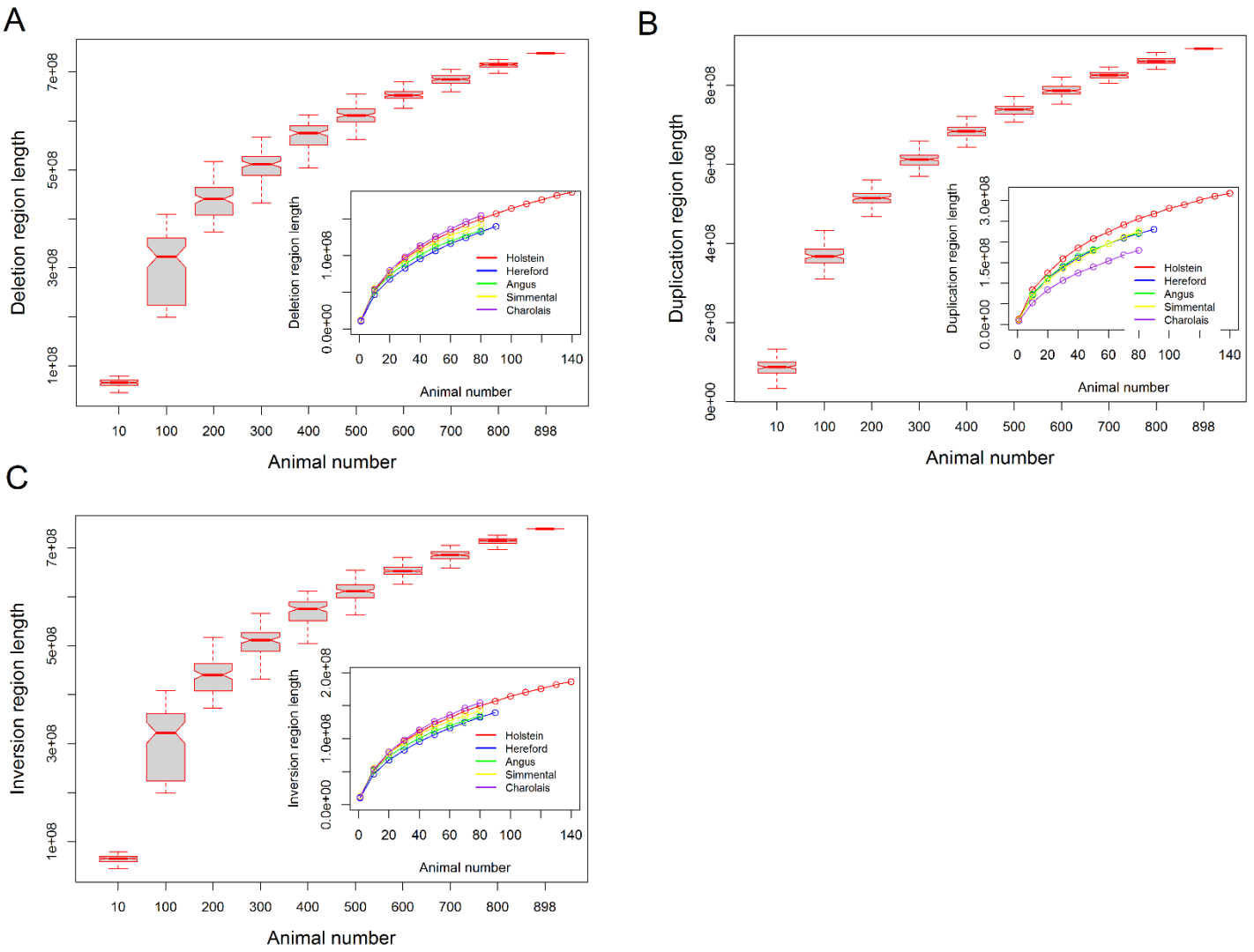
Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolution history

Yang Zhou[1*†], Lv Yang[1*], Xiaotao Han[1], Jiazheng Han[1], Yan Hu[1], Fan Li[1], Han Xia[1], Lingwei Peng[1], Clarissa Boschiero[3], Benjamin D. Rosen[3], Derek M. Bickhart[4], Shujun Zhang[1], Aizhen Guo[2], Curtis P. Van Tassell[3], Timothy P.L. Smith[5], Liguo Yang[1,†], George E. Liu[3,†]

[1]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China
[2]The State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, 430070, China
[3]Animal Genomics and Improvement Laboratory, BARC, USDA-ARS, Beltsville, Maryland 20705, USA
[4]Dairy Forage Research Center, ARS USDA, Madison, Wisconsin 53706 USA
[5]US Meat Animal Research Center, ARS USDA, Clay Center, Nebraska 68933 USA

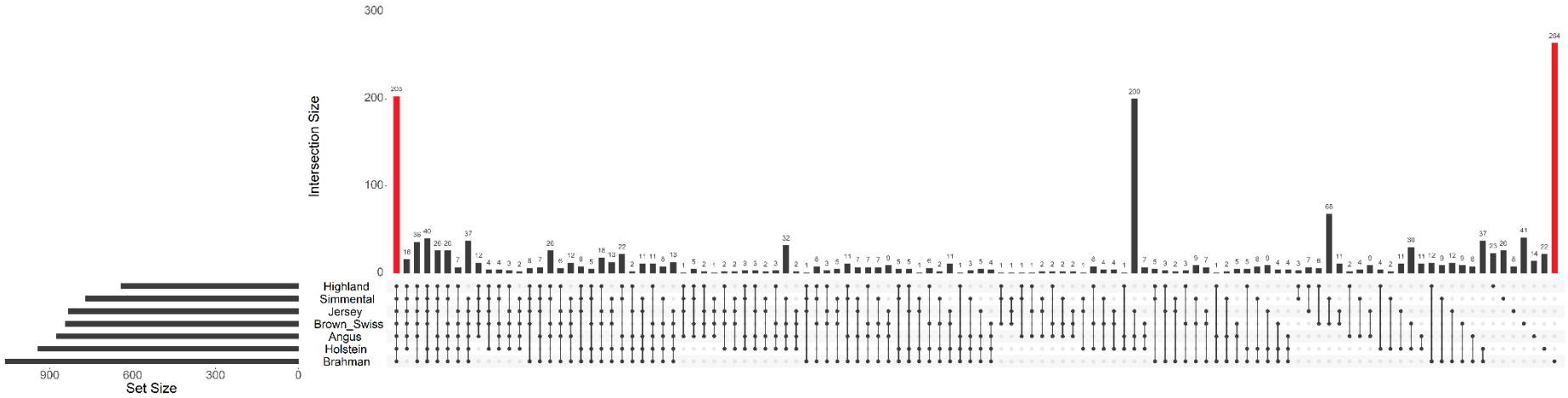
*These authors contributed equally to this work.

E-mail: George.Liu@usda.gov; yangzhou@mail.hzau.edu.cn; ylg@mail.hzau.edu.cn

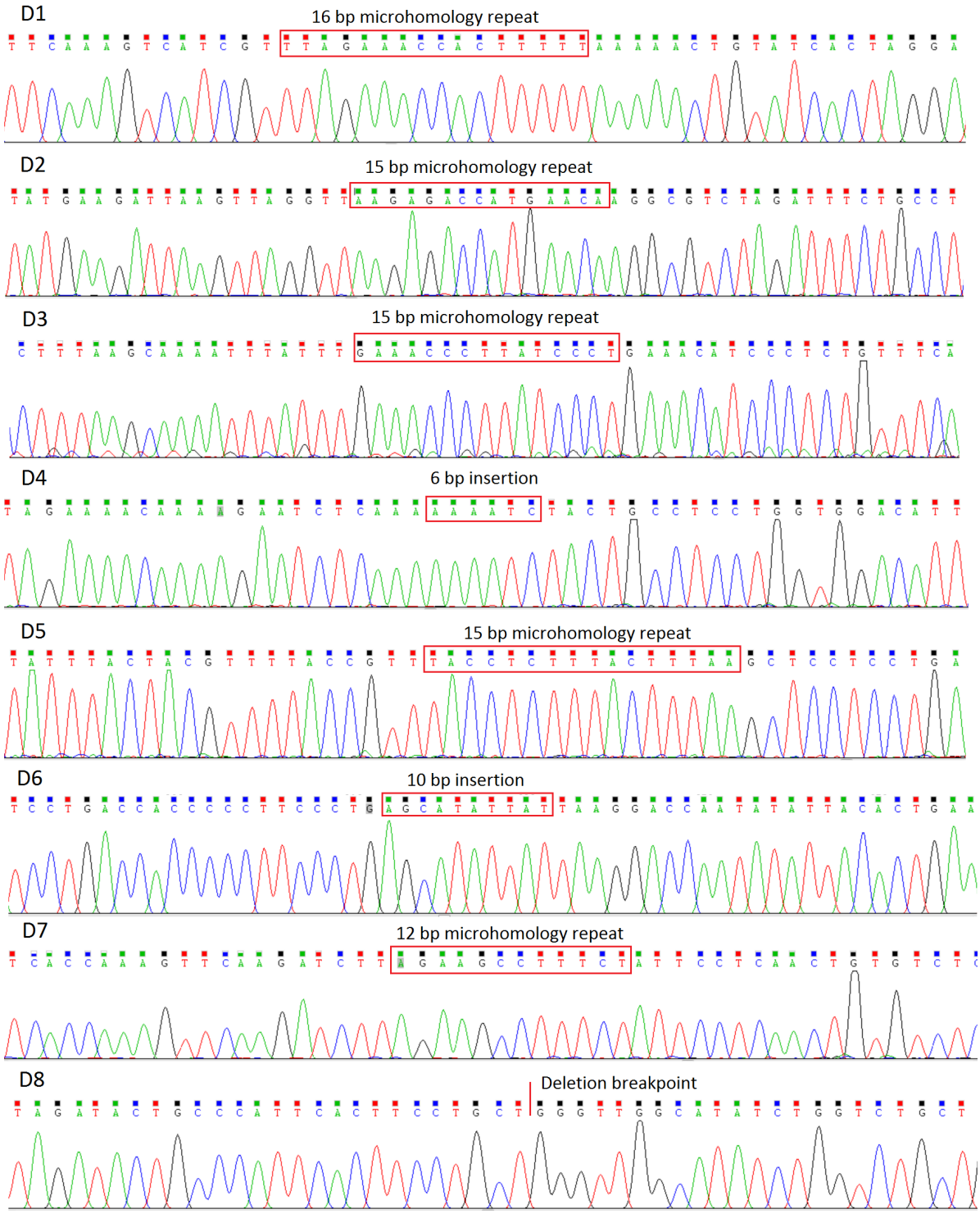


Supplemental Fig S1. Simulations of the increase in SV length detected with the increase of animal number for different SVs and cattle breeds. The Y-axis represents the length of the merged SV detected from the animals randomly sampled. The values used in the boxplot were the SV region length detected in 100 times of random selections from the 898 animals.

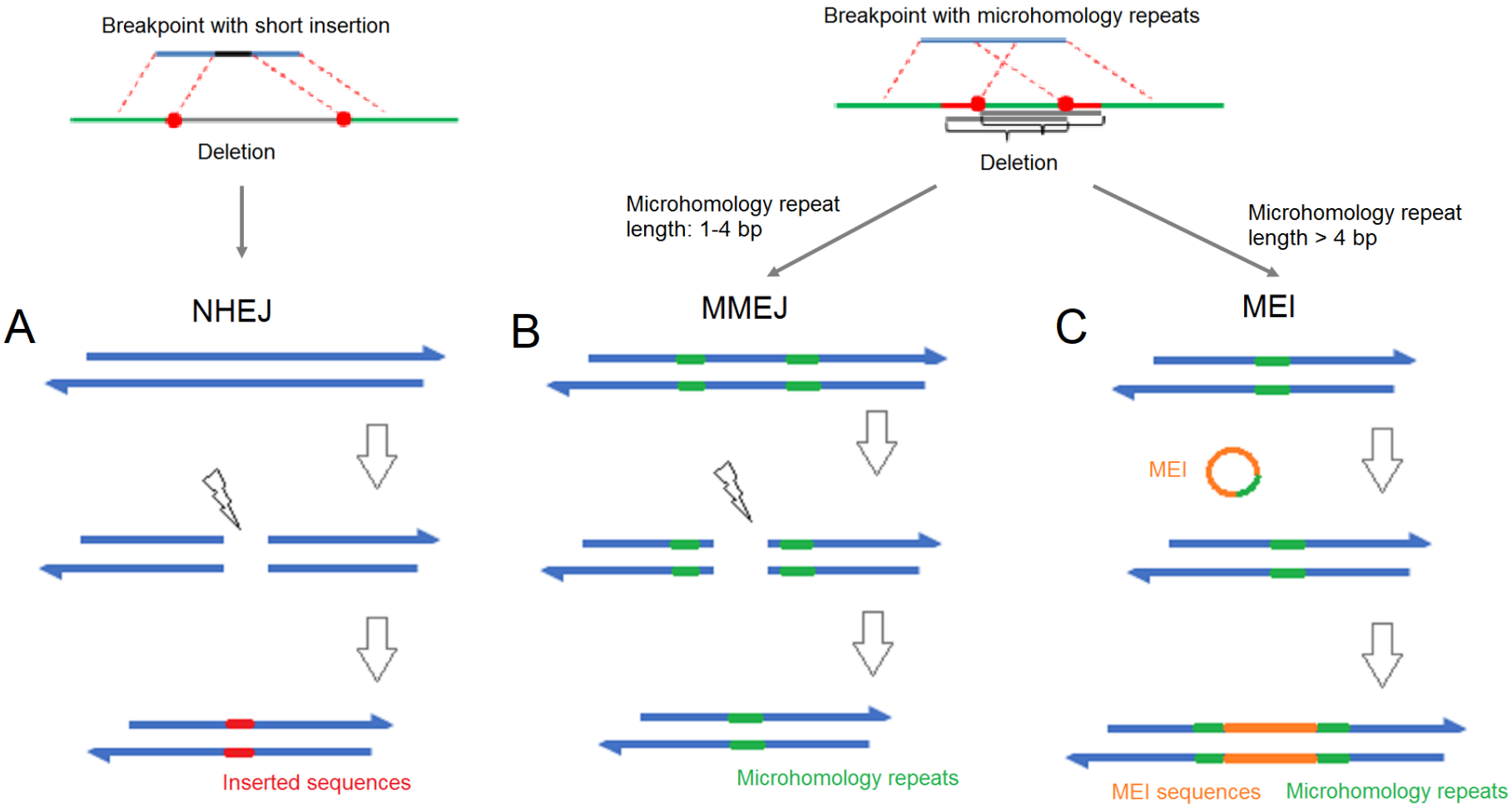
Supplemental Fig S2. The landscape of structural variations on cattle autosomes. SV Desert Gene: Genes located in the SV Desert Region; SV Cluster Gene: Genes located in the SV Cluster Region.



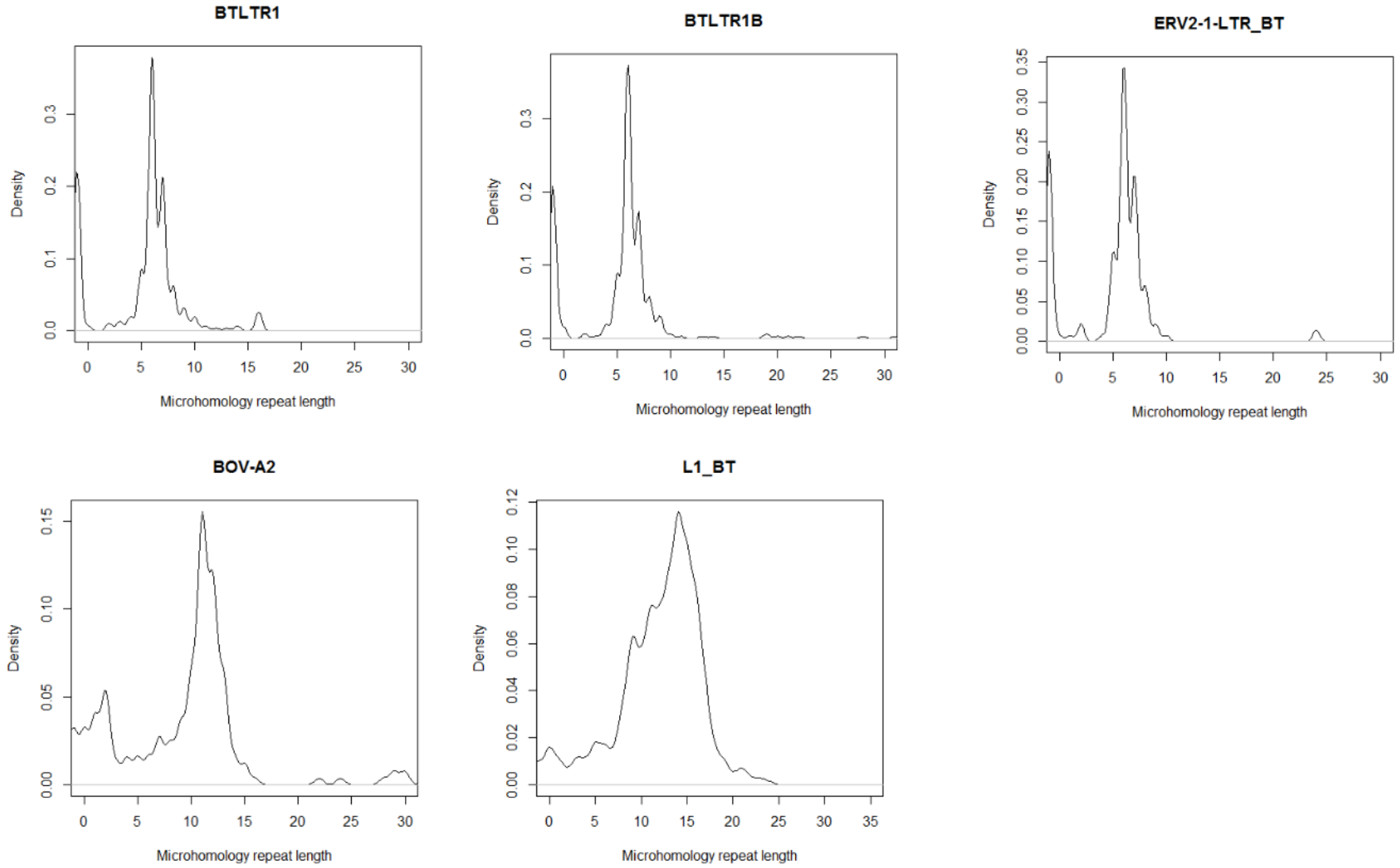
Supplemental Fig S3. UpSet plot for the numbers of successfully matched contig counts shared among seven assemblies of different cattle breeds.



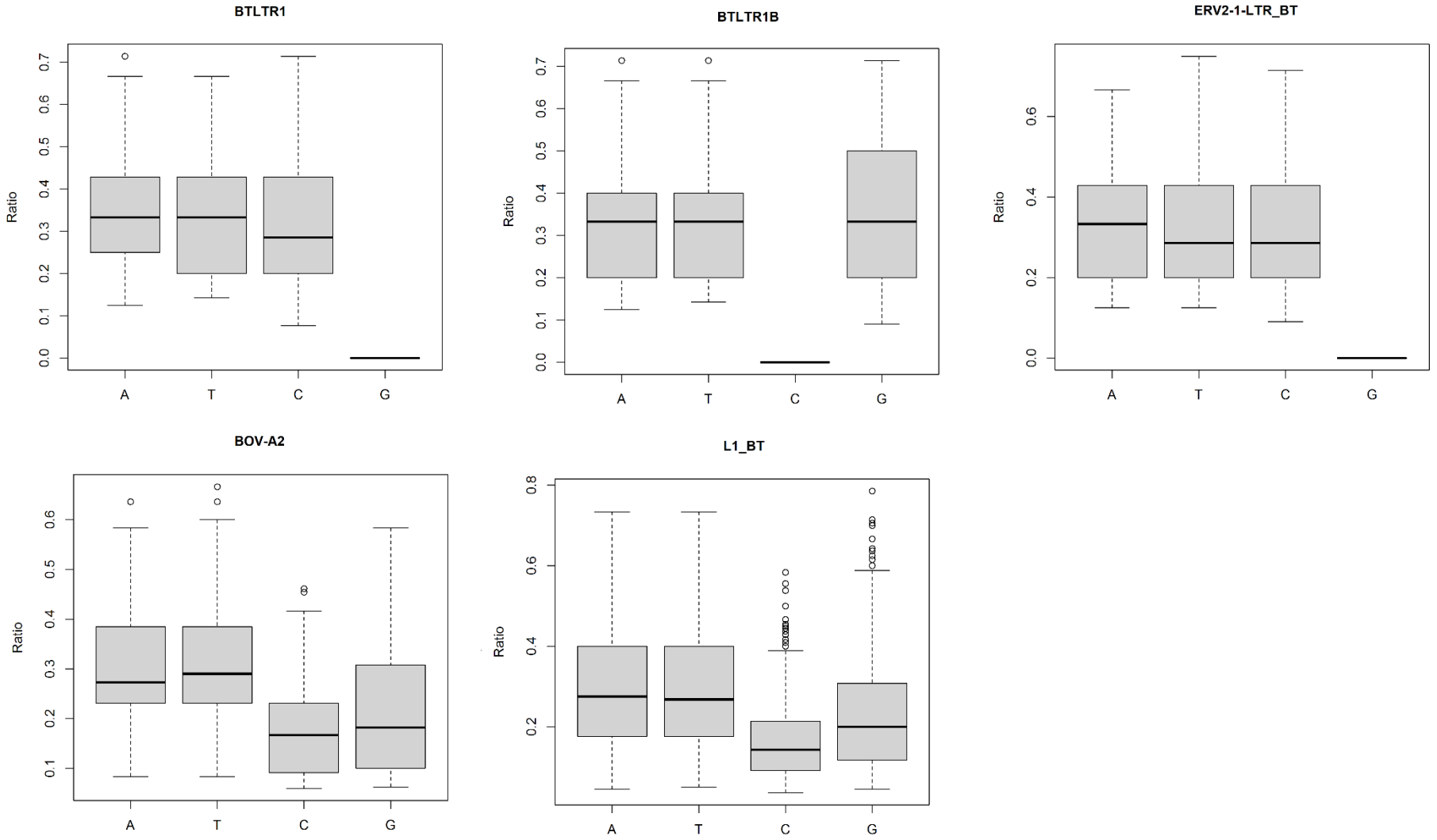
Supplemental Fig S4. Sanger sequencing results for validation of 8 deletions with different signatures.



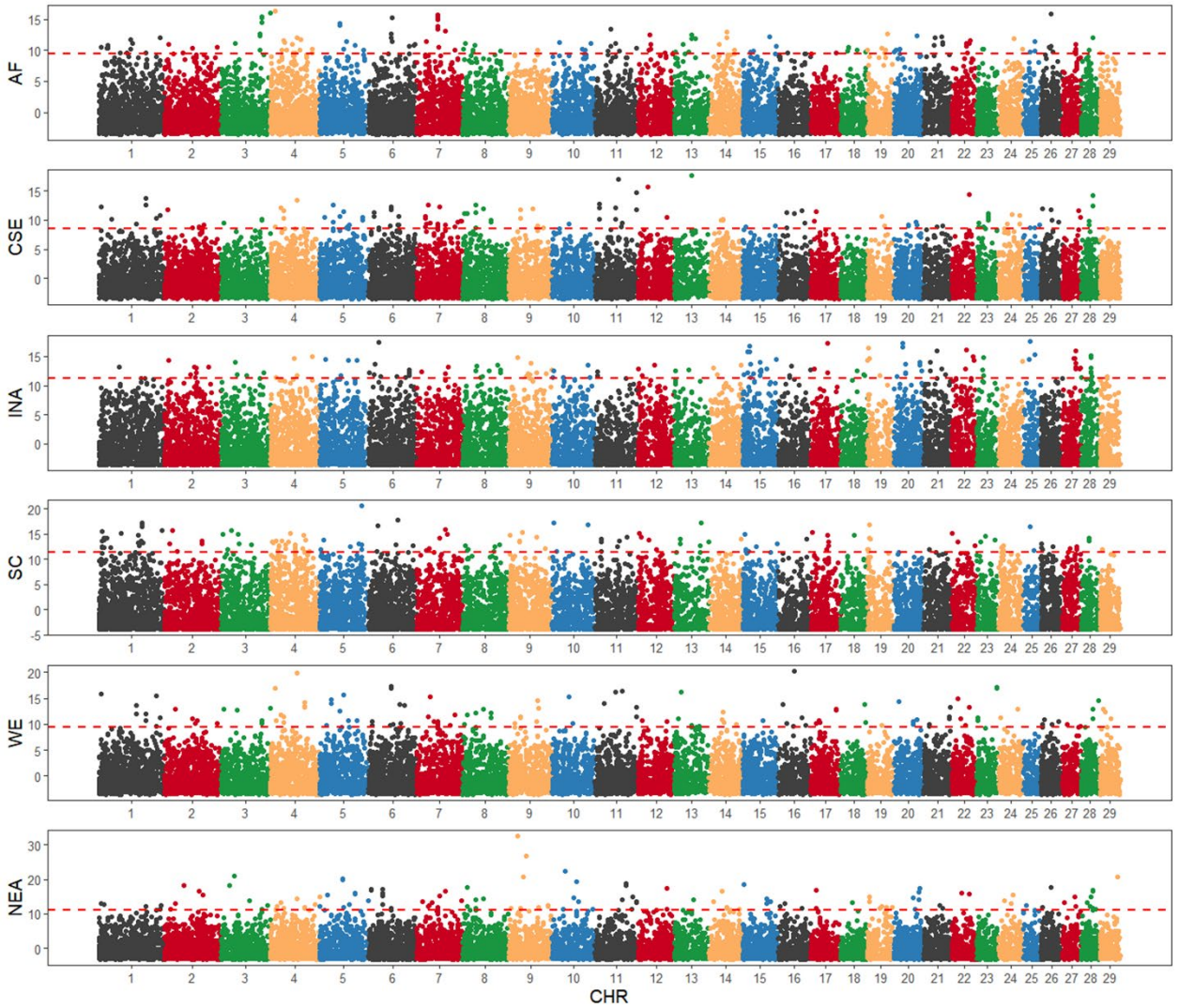
Supplemental Fig S5. Possible mechanisms for the deletion types.



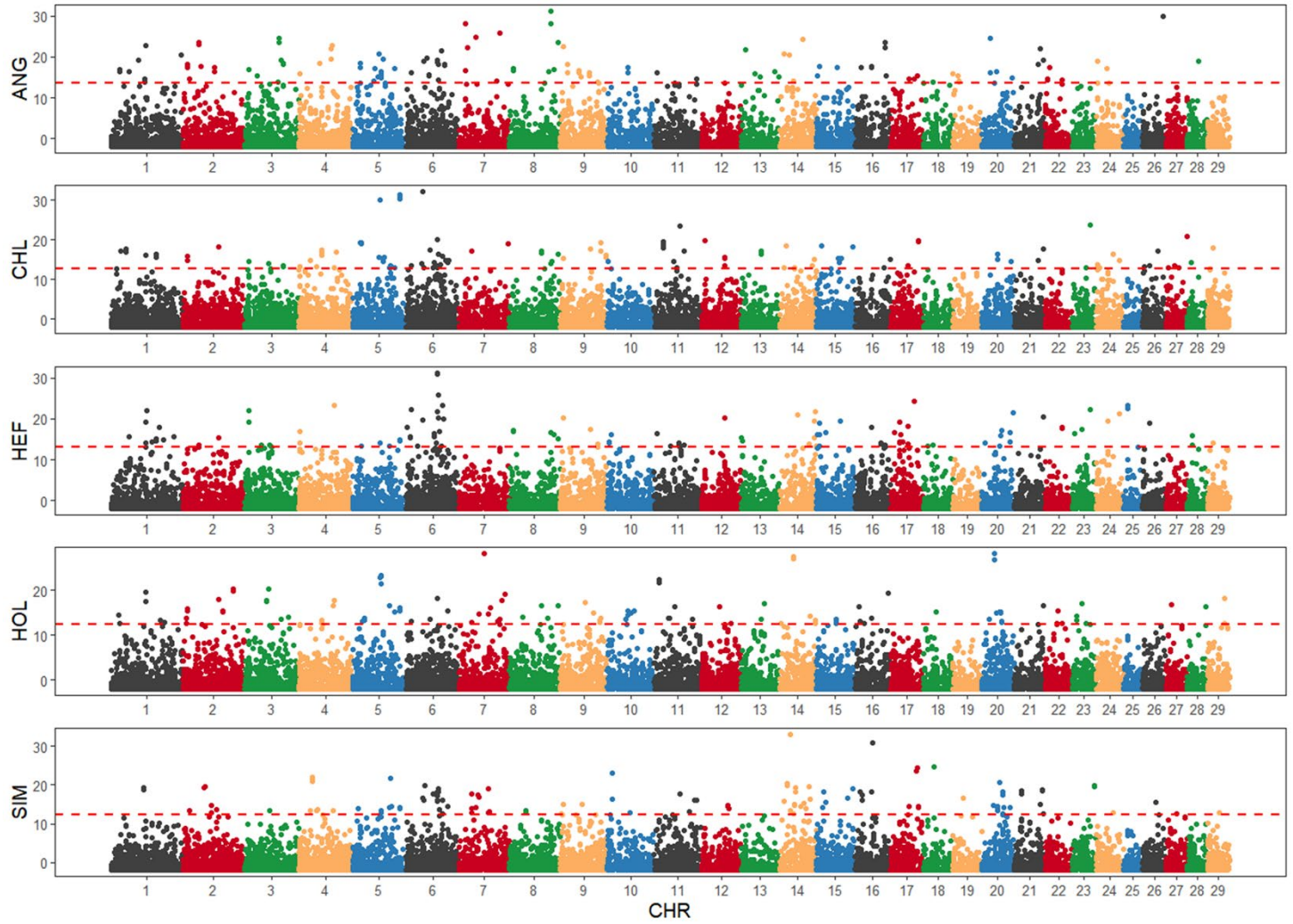
Supplemental Fig S6. Distributions of microhomology repeat lengths for different MEI families. X-axis: the negative values of the represent the lengths of the insertions, and the positive values correspond to the lengths of the microhomology repeats.



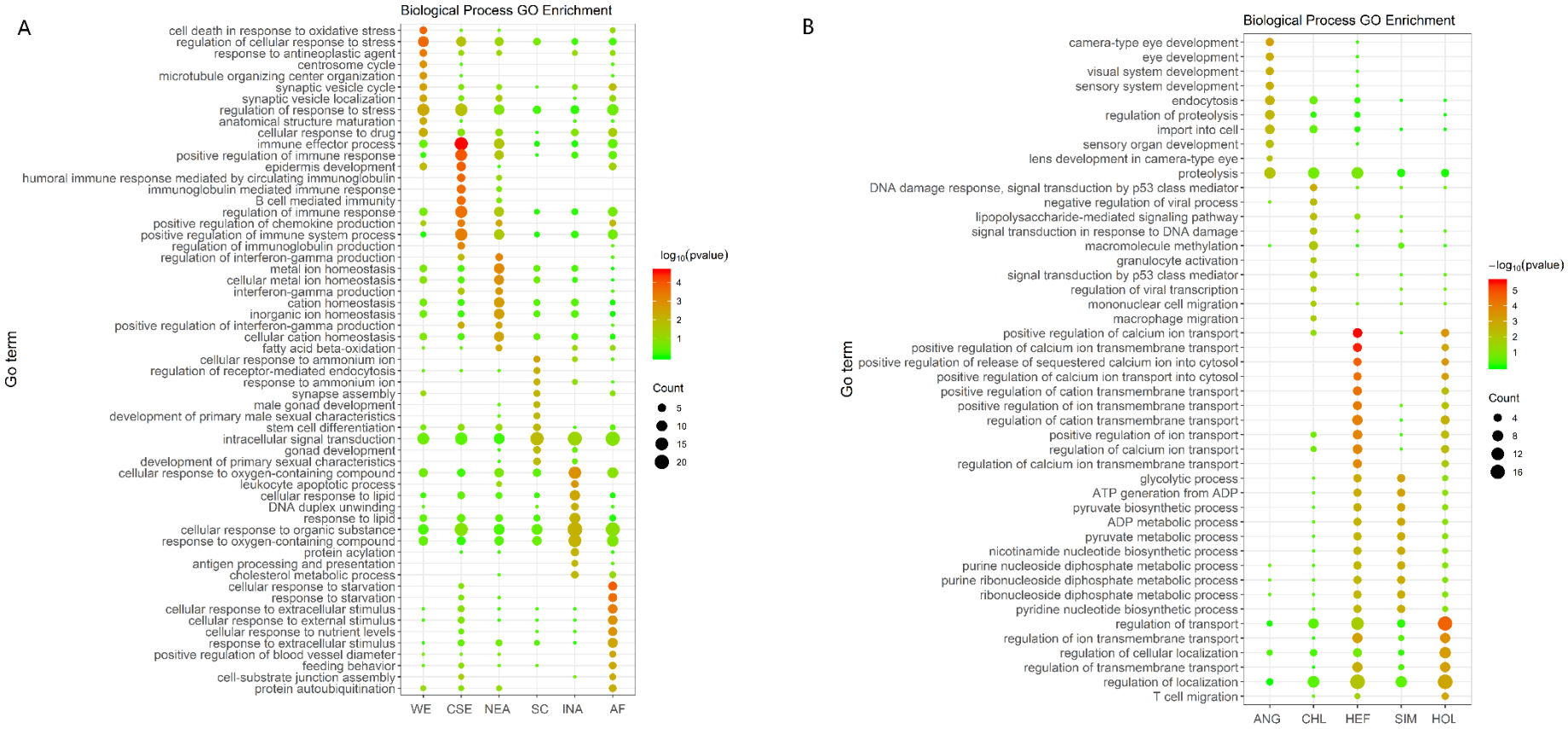
Supplemental Fig S7. Different base compositions of microhomology repeats for different MEI families.



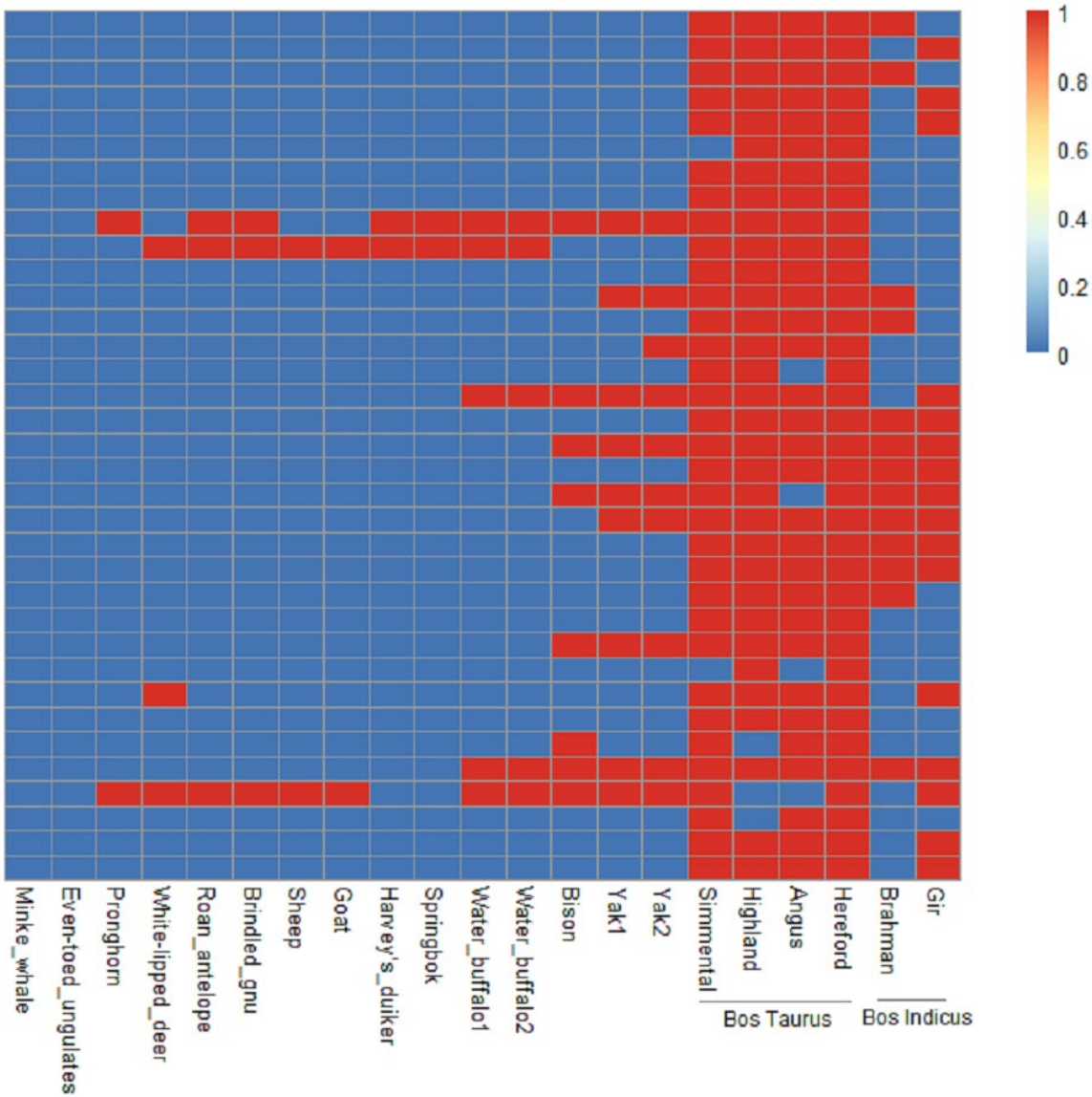
Supplemental Fig S8. Genomic distribution of selection regions in six cattle populations of different geographic origins. WE: West Europe, CSE: Central-South Europe, NEA: Northeast Asia, SC: South China, INA: India, AF: Africa.



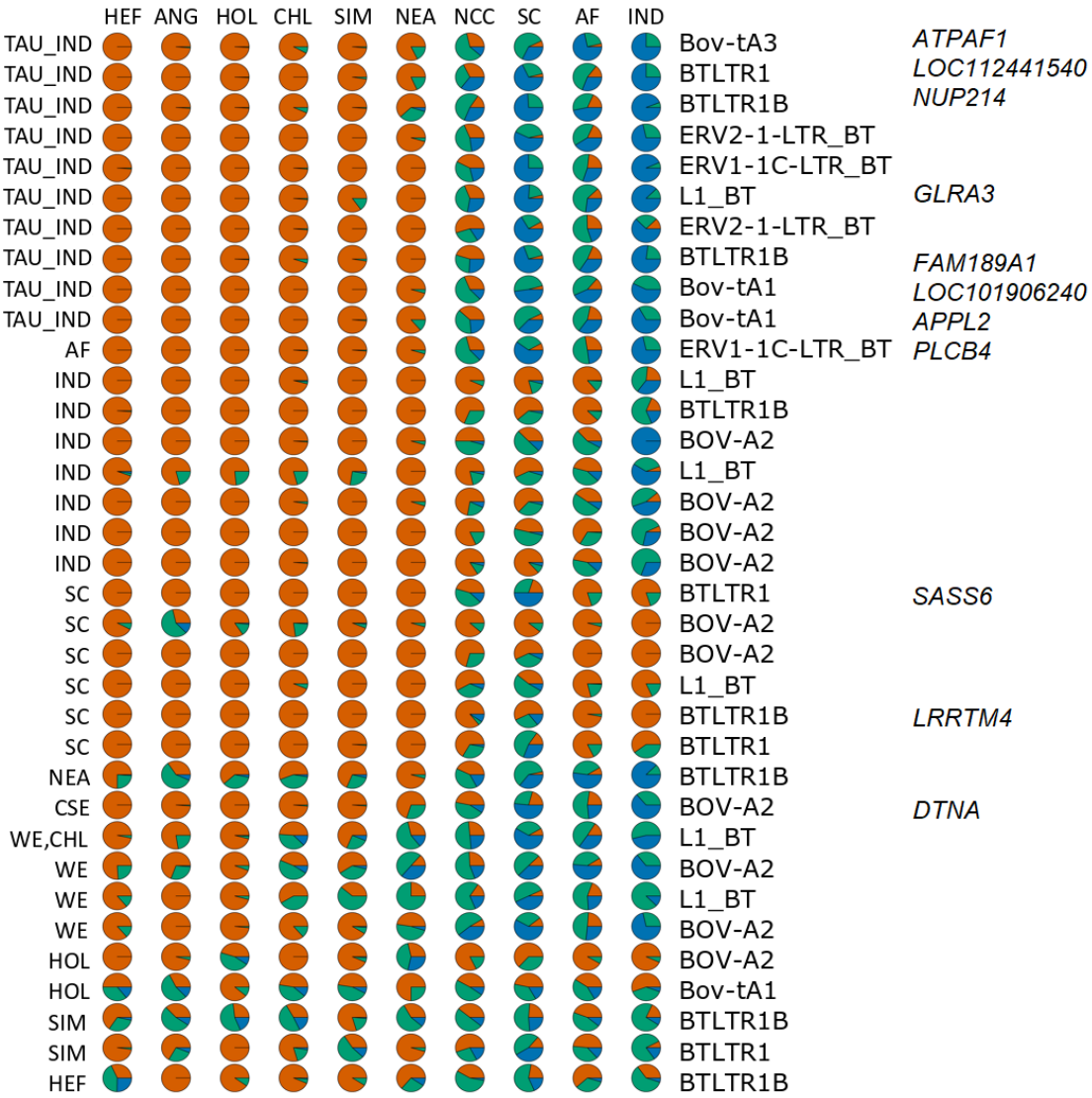
Supplemental Fig S9. Genomic distribution of selection regions in five widely used commercial cattle breeds. ANG: Angus, CHL: Charolais, HEF: Hereford, SIM: Simmental, HOL: Holstein.



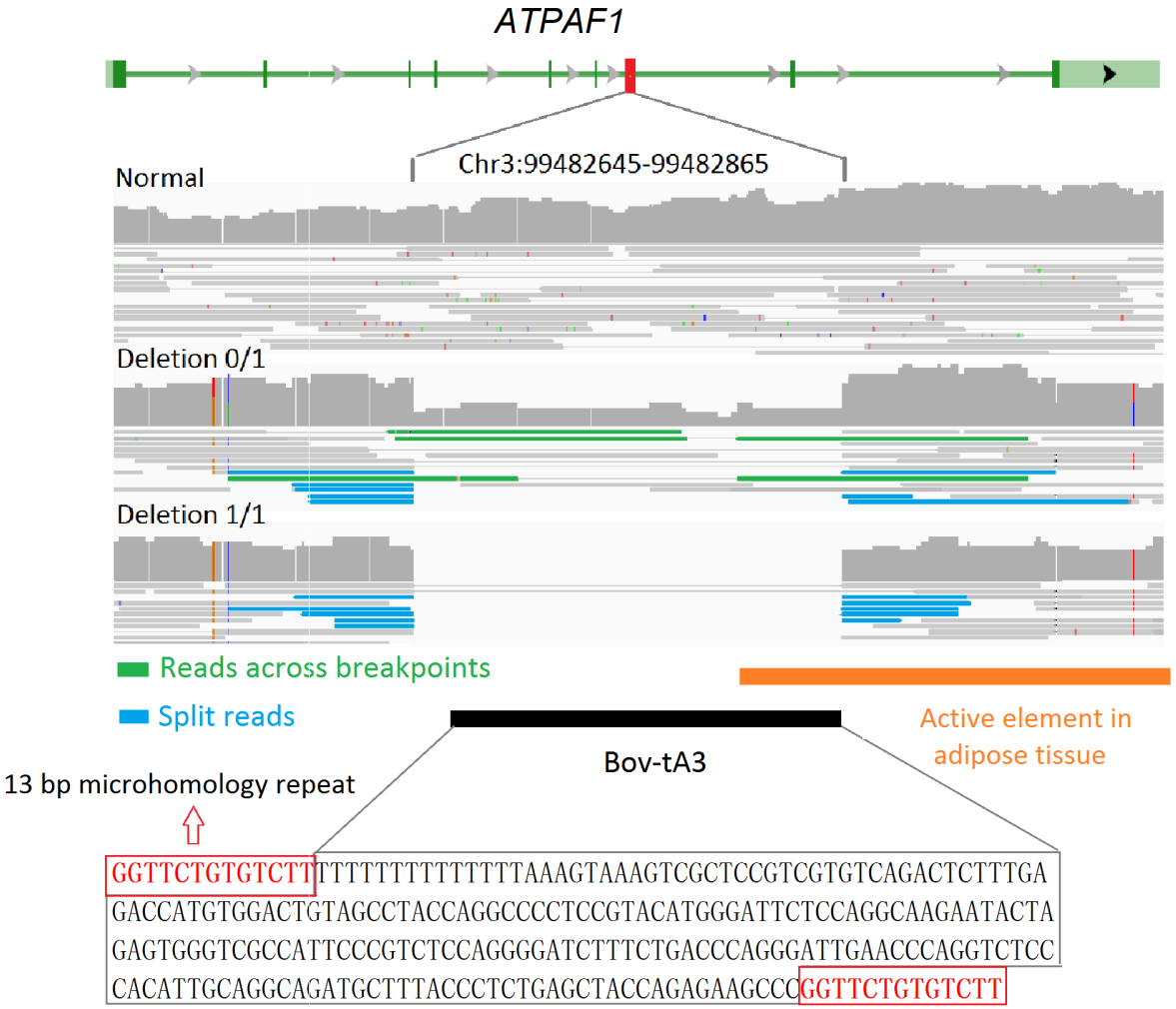
Supplemental Fig S10. Gene Ontology (GO) analysis for the genes under selection. A. Gene Ontology (GO) analysis for the genes under selection in six cattle populations of different geographic origins; B. Gene Ontology (GO) analysis for the genes under selection in five widely used European commercial cattle breeds.



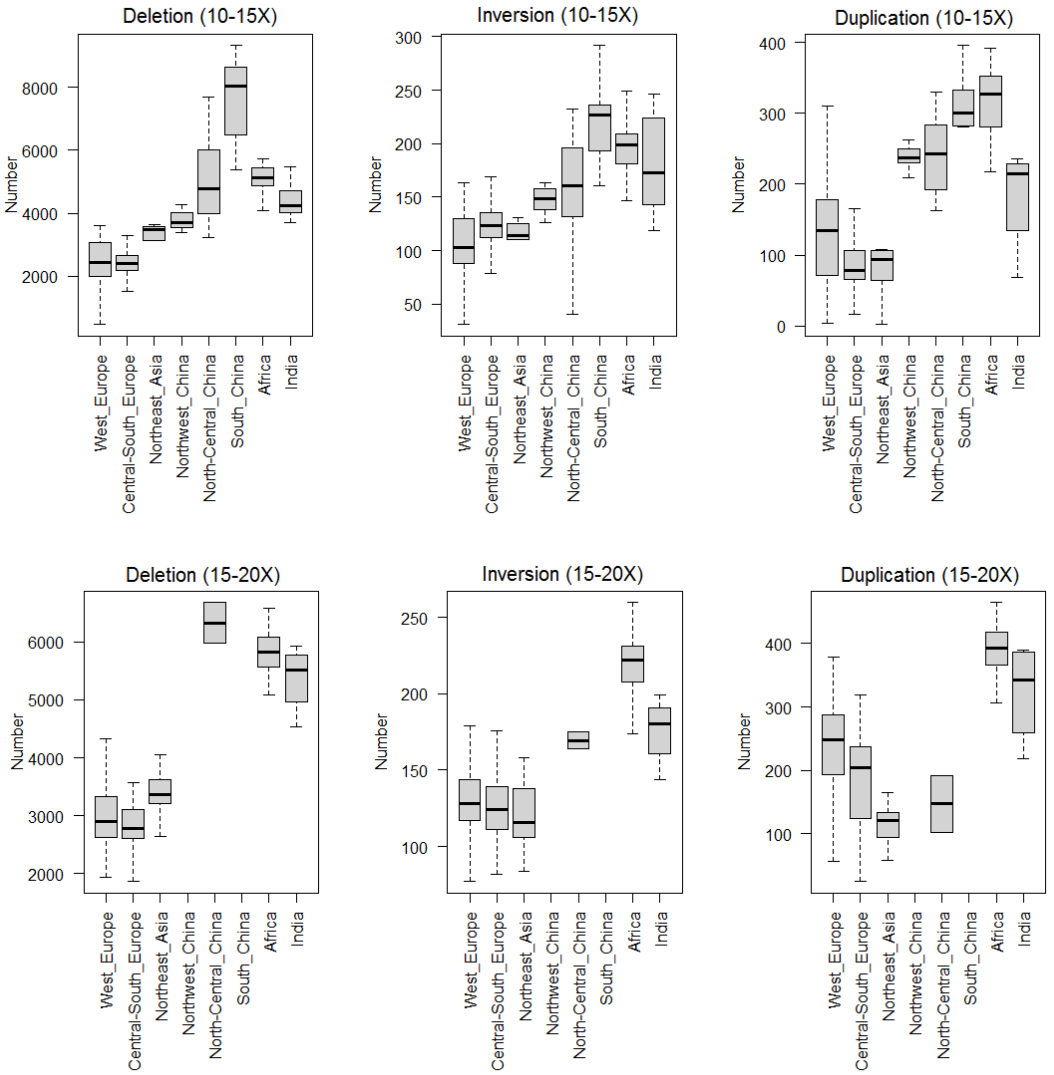
Supplemental Fig S11. Existence of MEIs located in the significantly selected deletions among different ruminants.



Supplemental Fig S12. Different genotypes' frequency distribution of deletions with MEIs in different cattle populations. ANG: Angus, CHL: Charolais, HEF: Hereford, SIM: Simmental, HOL: Holstein; NEA: Northeast Asia, NCC: North-Central China, SC: South China, INA: India, AF: Africa



Supplemental Fig S13. Sequence analysis for the deletion located in *ATPAF1*.



Supplemental Fig S14. Comparison of SV numbers among different cattle populations.