

Original Research

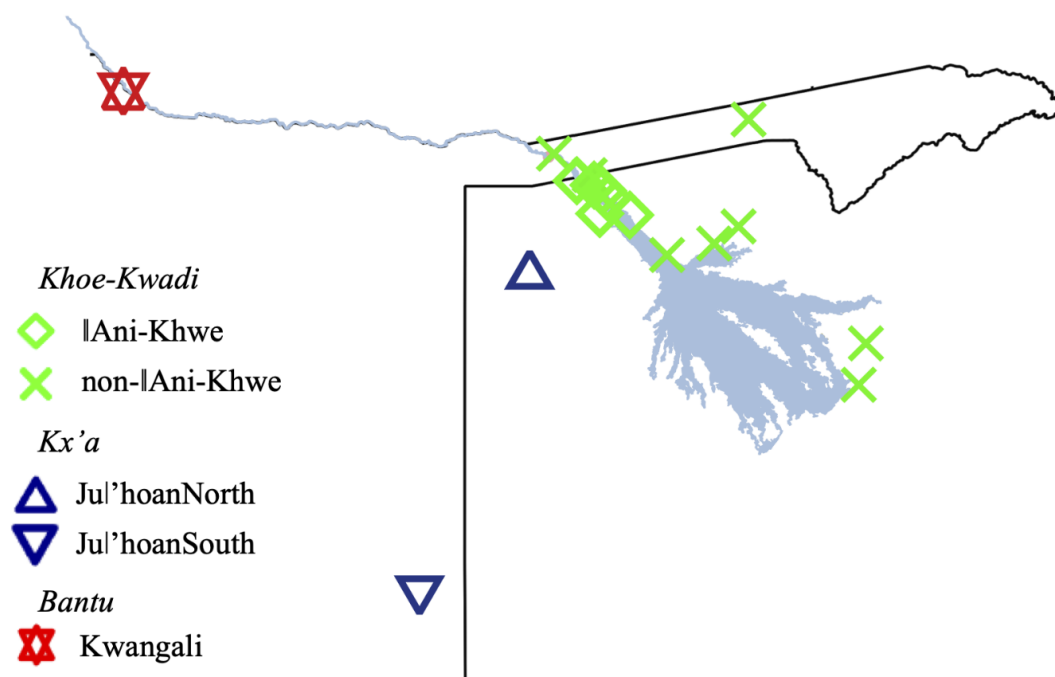
# Post-admixture selection favours Duffy negativity in the Lower Okavango Basin

Artem Kasianov <sup>1,2,\*</sup>, Anne-Maria Fehn <sup>1,2</sup>, Maitseo Bolaane <sup>3</sup>, Gaseitsiwe Masunga <sup>4</sup>,  
 Ezequiel Chimbioputo Fabiano <sup>1,2,5</sup>, Jorge Rocha <sup>1,2,6</sup>

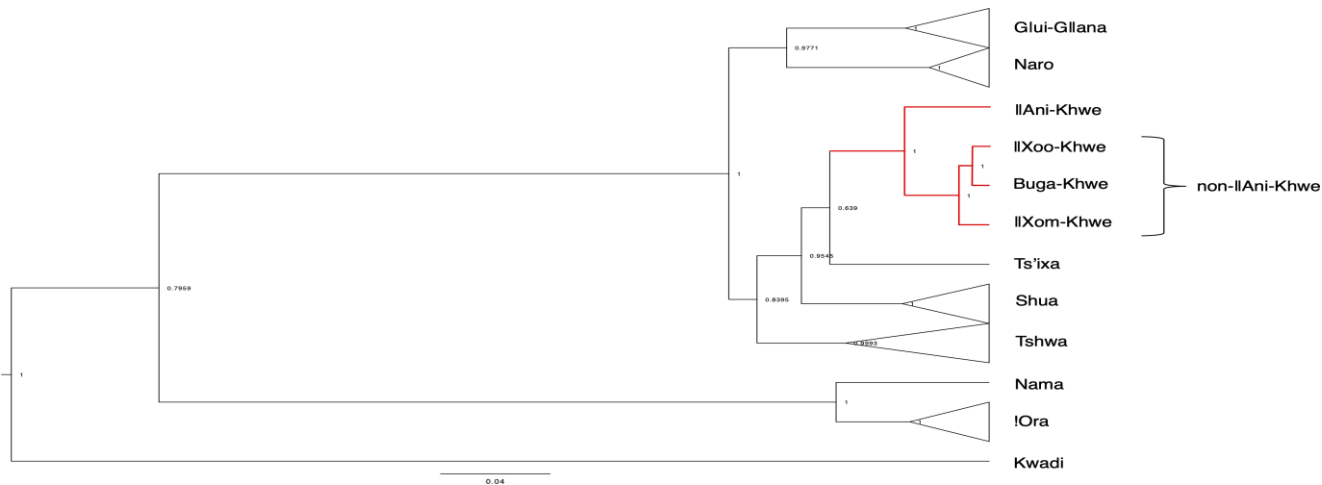
1. CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO Laboratório Associado, Campus de Vairão, Universidade do Porto, 4485-661 Vairão, Portugal; Email: [afehn@cibio.up.pt](mailto:afehn@cibio.up.pt) (A-M.F.); [fabianoezequiel@gmail.com](mailto:fabianoezequiel@gmail.com) (E.C.F.); [jrocha@cibio.up.pt](mailto:jrocha@cibio.up.pt) (J.R.)
2. Biopolis Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, 4485-661 Vairão, Portugal
3. San Research Centre, University of Botswana, P.O. Box 70375, Gaborone, Botswana; Email: [bolaanem@ub.ac.bw](mailto:bolaanem@ub.ac.bw) (M.B.)
4. Okavango Research Institute, University of Botswana, Private Bag 285, Maun, Botswana; Email: [gsmasunga@ub.ac.bw](mailto:gsmasunga@ub.ac.bw) (G.M.)
5. Department of Wildlife Management and Tourism Studies, University of Namibia, Katima Mulilo Campus, Private Bag 13301, Katima Mulilo, Namibia
6. Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, 4099-002 Porto, Portugal

\* **Correspondence:** Artem Kasianov; Email: [artem.kasianov@cibio.up.pt](mailto:artem.kasianov@cibio.up.pt)

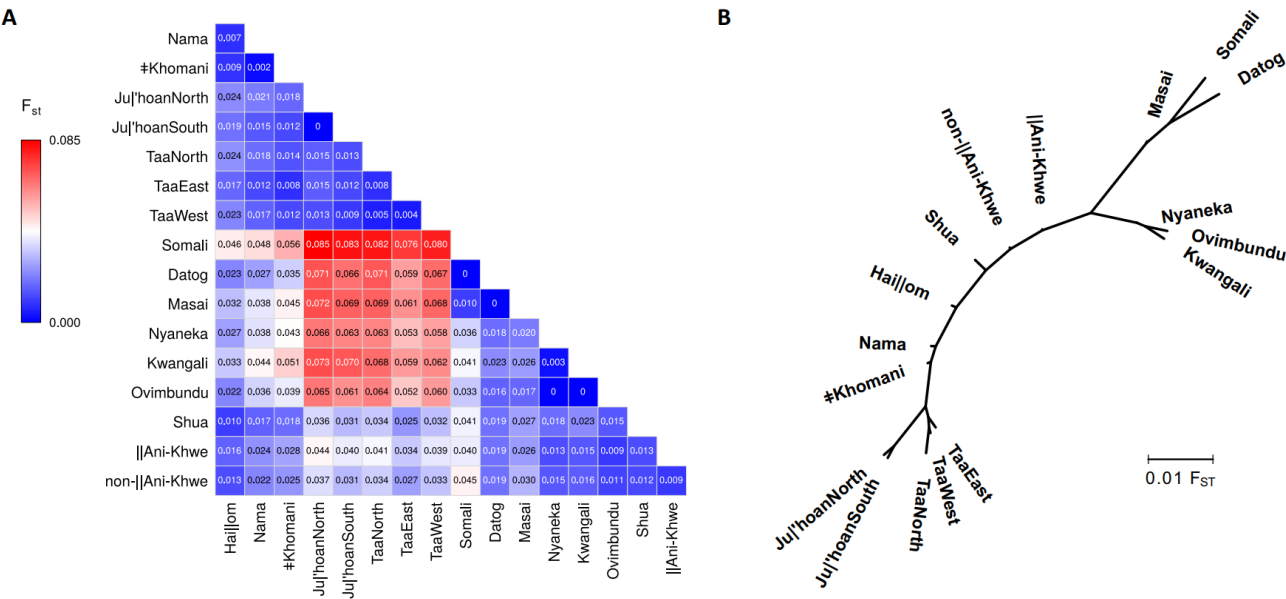
## Supplementary Materials



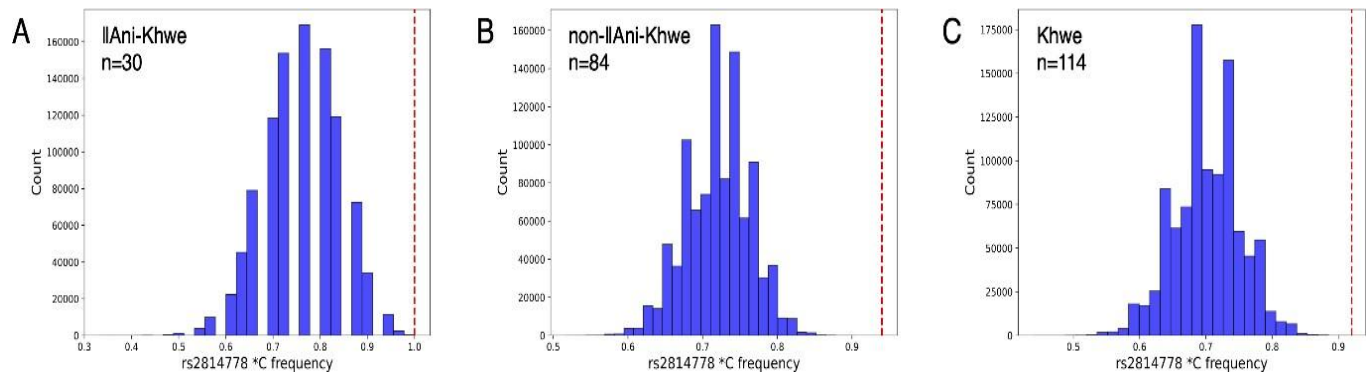
**Figure S1. Geographic locations of sampled Khwe-speaking individuals, together with Kwangali (this study) and Ju|'hoan-speaking groups [1,2], which were used as proxies for parental populations.**



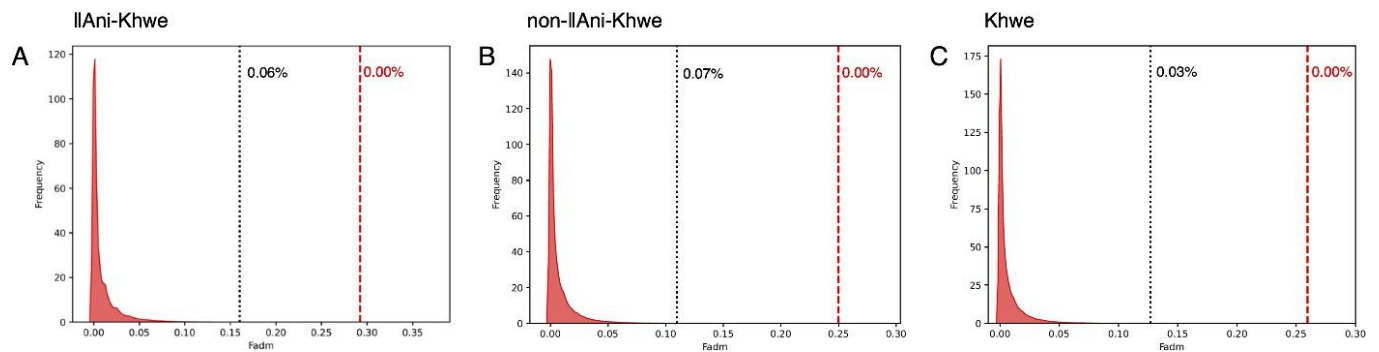
**Figure S2. Consensus tree of a Bayesian phylogenetic analysis of the Khoe-Kwadi language family under the Continuous Markov Chain Model.** The numbers indicate posterior probabilities for each split. Branches corresponding to different Khwe dialects are marked in red (modified from [3]).



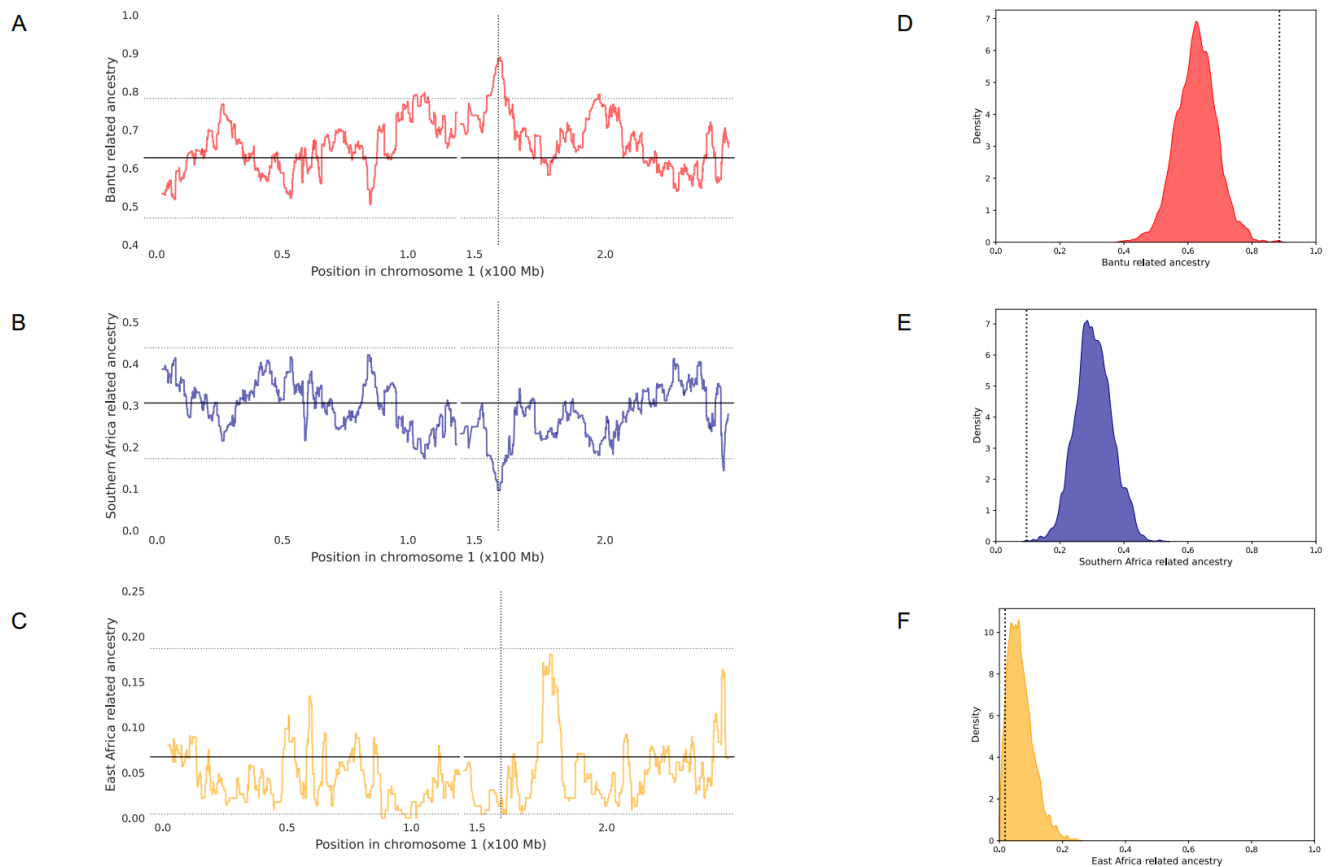
**Figure S3. Pairwise  $F_{ST}$  relationships among newly sampled and contextual populations.** Pairwise  $F_{ST}$  values among newly sampled and contextual populations shown in Figure 1 (A). Neighbour-joining network based on the pairwise  $F_{ST}$  matrix (B).



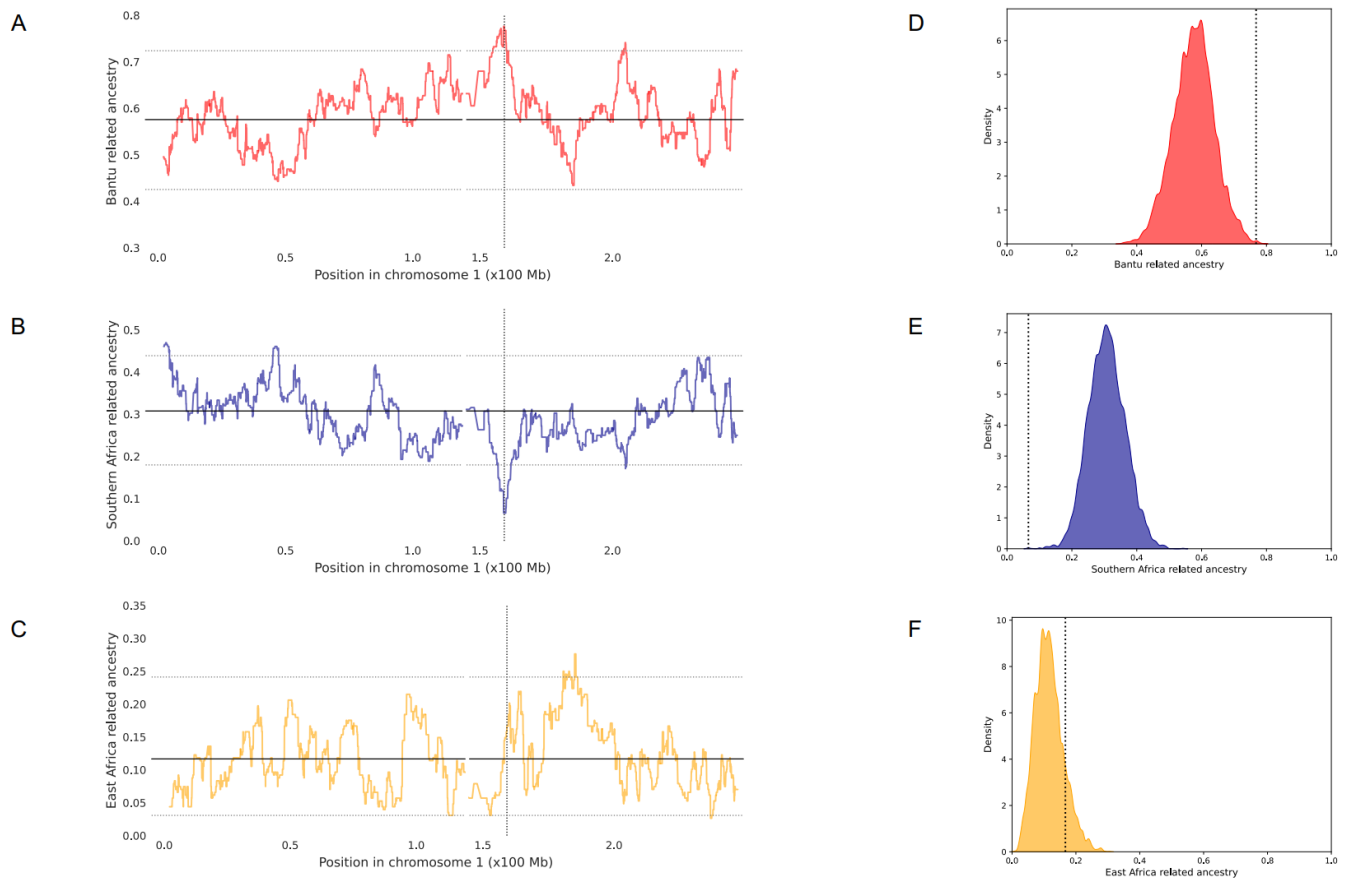
**Figure S4. Distributions of expected  $FY*B^{ES}$  (rs2814558\*C) allele frequencies obtained by binomial resampling with  $10^6$  replicates.** The dotted lines indicate the observed allele frequencies in each Khwe subgroup (A, B) and in the total Khwe sample (C).



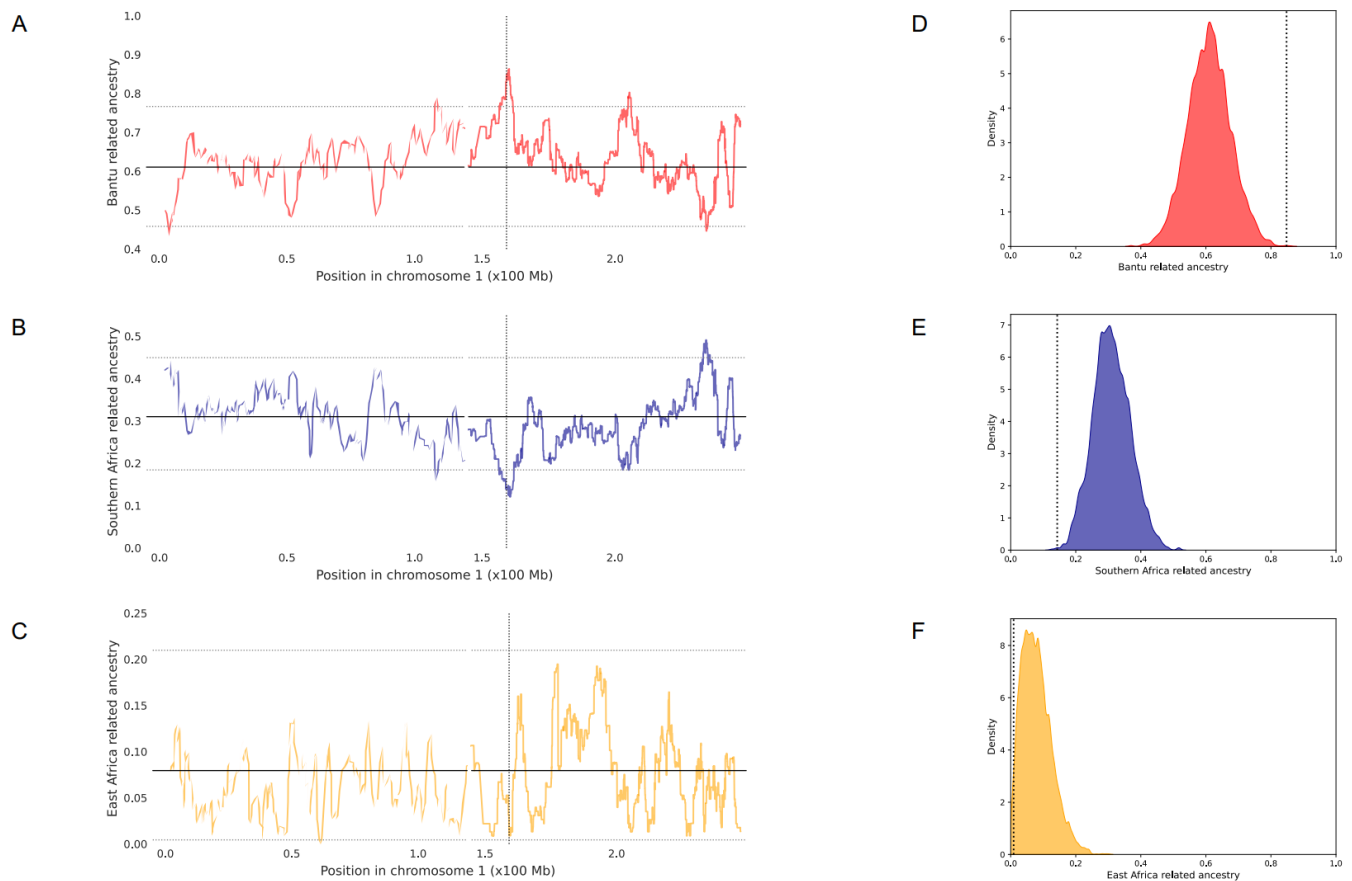
**Figure S5. Genome-wide distributions of the  $F_{adm}$  statistic.** The vertical lines indicate the  $F_{adm}$  values for the SNP rs2814778 in each Khwe subgroup (A, B) and in the total Khwe sample (C). Dotted black lines indicate  $F_{adm}$  values calculated with the most conservative array of parental rs2814778\*C ( $FY*B^{ES}$ ) allele frequencies: Kwangali 100%; Ju|'hoan 11%; Somali 92%. Dashed red lines indicate  $F_{adm}$  values assuming that the  $FY*B^{ES}$  allele was absent in the pre-Bantu hybrid population. The numbers refer to the percentages of all SNPs exhibiting values equal or greater than that of the rs2814778 SNP under each assumption.



**Figure S6. Local ancestry proportions along chromosome 1 and genome-wide distributions of local ancestry proportions in the Khwe, using Kwangali, Ju|'hoan, and Hadza as proxies for Bantu, southern African, and eastern African-related ancestries, respectively.** (A-C) Average local ancestry proportions in 114 Khwe individuals along chromosome 1 for Bantu (A), southern African (B), and eastern African-related (C) ancestries. Solid horizontal lines indicate genome-wide average ancestry proportions, while dashed horizontal lines correspond to the 99th and 1st percentiles of local ancestry distributions. Dashed vertical lines mark the position of the Duffy locus. (D-F) Genome-wide distributions of average Bantu (D), southern African (E), and eastern African-related (F) ancestry proportions. Dashed vertical lines indicate the local ancestry estimate for the RFMix window containing the Duffy locus. The Chromosome 1 region from position 119429575 to position 144929697 was excluded, due to the absence of markers on the Affymetrix Axiom Genome-Wide Human Origins Array. Coordinates are based on the GRCh37/hg19 version of the *H. sapiens* genome.



**Figure S7. Local ancestry proportions along chromosome 1 and genome-wide distributions of local ancestry proportions in the Khwe, using Nyaneka, Ju|'hoan, and Somali as proxies for Bantu, southern African, and eastern African-related ancestries, respectively.** (A-C) Average local ancestry proportions in 114 Khwe individuals along chromosome 1 for Bantu (A), southern African (B), and eastern African-related (C) ancestries. Solid horizontal lines indicate genome-wide average ancestry proportions, while dashed horizontal lines correspond to the 99th and 1st percentiles of local ancestry distributions. Dashed vertical lines mark the position of the Duffy locus. (D-F) Genome-wide distributions of average Bantu (D), southern African (E), and eastern African-related (F) ancestry proportions. Dashed vertical lines indicate the local ancestry estimate for the RFMix window containing the Duffy locus. The Chromosome 1 region from position 119429575 to position 144929697 was excluded, due to the absence of markers on the Affymetrix Axiom Genome-Wide Human Origins Array. Coordinates are based on the GRCh37/hg19 version of the *H. sapiens* genome.



**Figure S8. Local ancestry proportions along chromosome 1 and genome-wide distributions of local ancestry proportions in the Khwe, using Nyaneka, Ju|'hoan, and Hadza as proxies for Bantu, southern African, and eastern African-related ancestries, respectively.** (A-C) Average local ancestry proportions in 114 Khwe individuals along chromosome 1 for Bantu (A), southern African (B), and eastern African-related (C) ancestries. Solid horizontal lines indicate genome-wide average ancestry proportions, while dashed horizontal lines correspond to the 99th and 1st percentiles of local ancestry distributions. Dashed vertical lines mark the position of the Duffy locus. (D-F) Genome-wide distributions of average Bantu (D), southern African (E), and eastern African-related (F) ancestry proportions. Dashed vertical lines indicate the local ancestry estimate for the RFMix window containing the Duffy locus. The Chromosome 1 region from position 119429575 to position 144929697 was excluded, due to the absence of markers on the Affymetrix Axiom Genome-Wide Human Origins Array. Coordinates are based on the GRCh37/hg19 version of the *H. sapiens* genome.

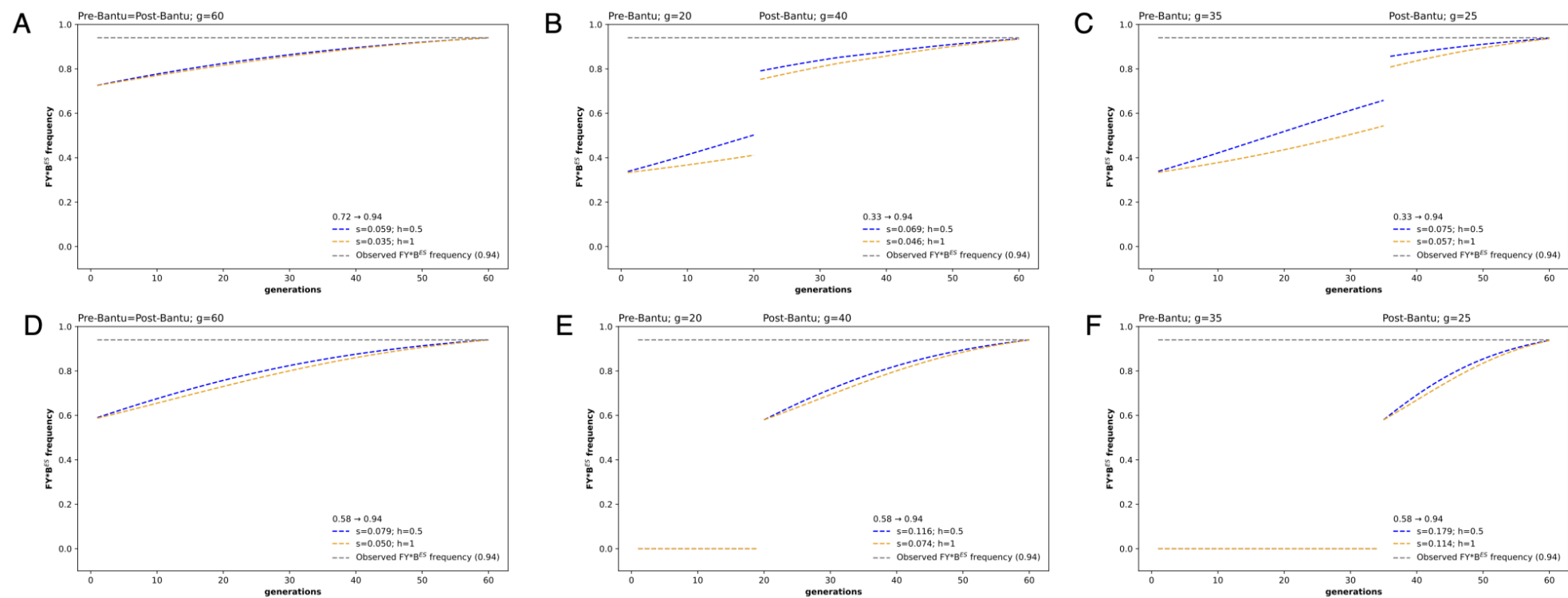


Figure S9. Trajectories of  $FY*B^{ES}$  allele frequencies in the Khwe under the admixture scenarios described in Table S9, using the best-fitting selection coefficients ( $s$ ), dominance coefficients ( $h$ ) of 0.5 and 1, and assuming an infinite population size.

## References

1. Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T, et al. The genetic prehistory of southern Africa. *Nat Commun.* 2012;3:1143. [DOI](#)
2. Lazaridis I, Patterson N, Mitnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature.* 2014;513(7518):409-413. [DOI](#)
3. Fehn AM, Sands B, Phiri A, Bolaane M, Masunga G, Rocha J. Tracing contact and migration in pre-Bantu southern Africa through lexical borrowing. *Evol Hum Sci.* 2025;7:e25. [DOI](#)