

Supporting Information for

Gaps in archaeological metadata reporting: a meta-analysis of human paleogenomic studies in Western Eurasia

Robert Staniuk*^{1,6}, Victor Yan Kin Lee*^{2,6}, Adrian Timpson³, Kristian Kristiansen⁴, Stephen Shennan⁵, Fernando Racimo², Mark G. Thomas*³

¹ Faculty of Archaeology, Adam Mickiewicz University, ul. Uniwersytetu Poznańskiego 7/2.69, 61-614 Poznań, Poland

² Section for Molecular Ecology and Evolution, Globe Institute, University of Copenhagen, 1350 Copenhagen, Denmark

³ Research Department of Genetics, Evolution and Environment, University College London, Gower Street, London, WC1E 6BT, United Kingdom

⁴ Department of Historical Studies, University of Gothenburg, Renströmsgatan 6, 412 55 Göteborg, Sweden

⁵ Institute of Archaeology, University College London, WC1H 0PY, London, United Kingdom

⁶ These authors contributed equally

* Corresponding authors

Email: robert.staniuk@amu.edu.pl

This PDF file includes:

Figures S1 to S21

Legends for Datasets S1 to S7

Other supporting materials for this manuscript include the following:

Dataset S1 to S7

Figures

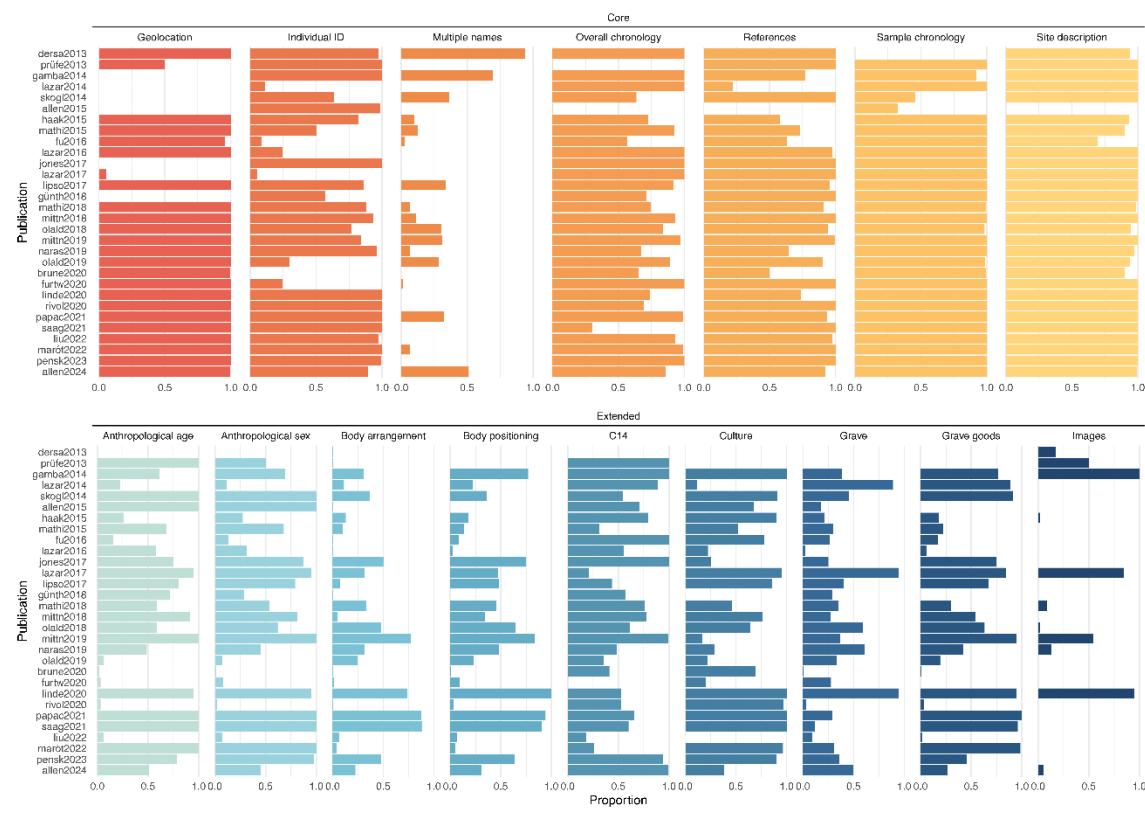


Fig. S1. An overview of archaeological metadata reporting in proportions of the 30 peer-reviewed publications from 2013 to 2024.

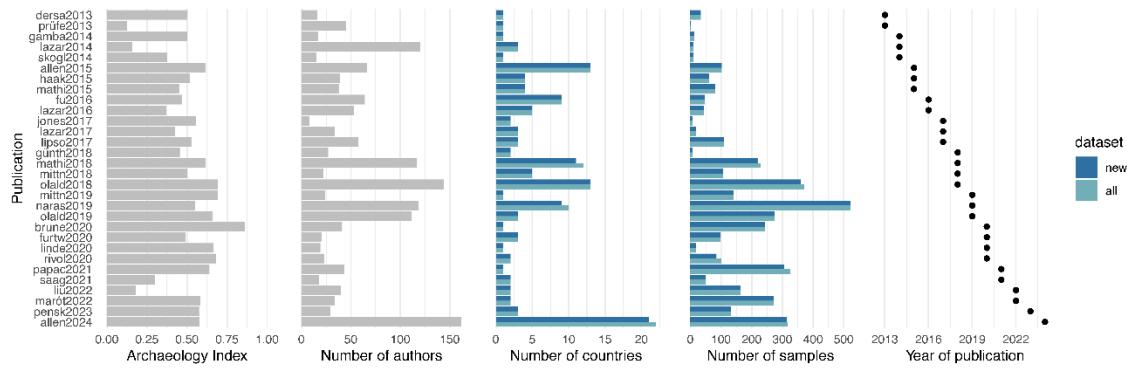


Fig. S2. An overview of the candidate explanatory variable including both newly and resequenced sampled individuals for each publication. “New” refers to only newly sampled individuals; “all” refers to the total dataset.

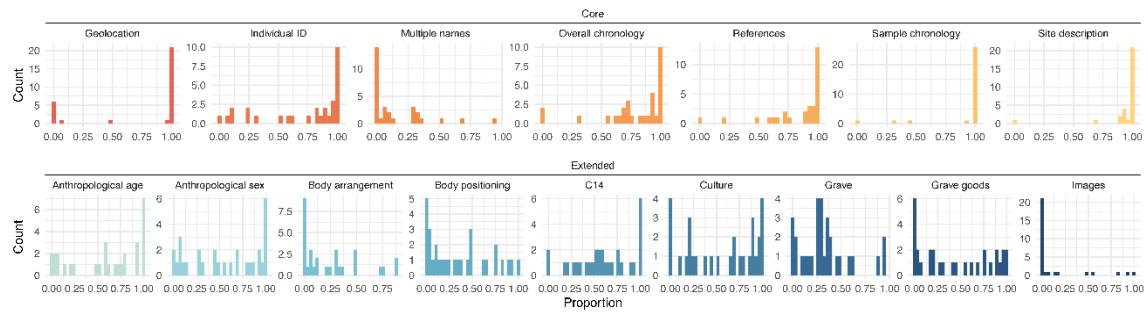


Fig. S3. The distributions of metadata completeness.

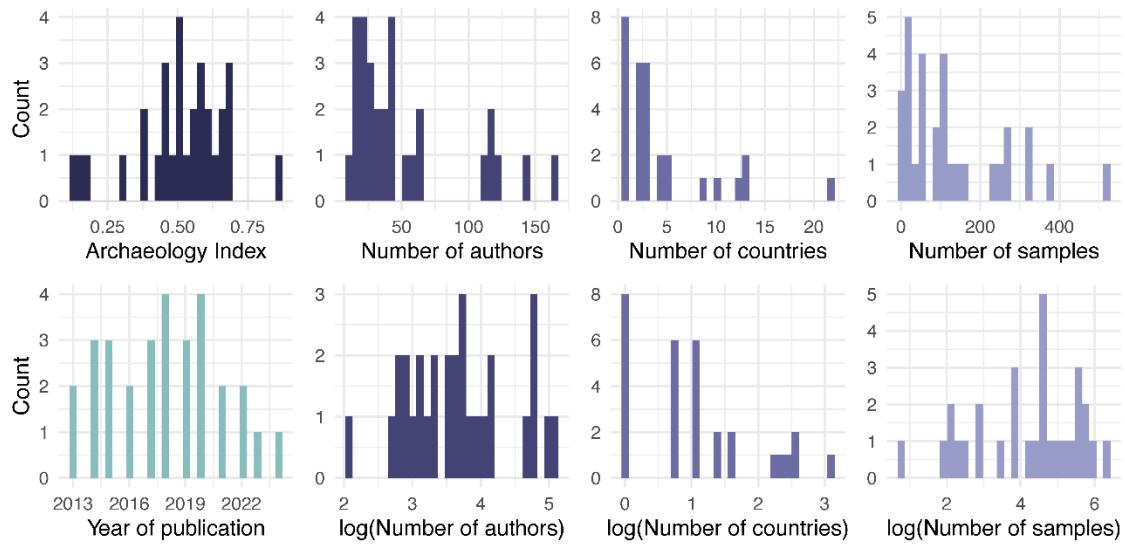


Fig. S4. The distributions of candidate explanatory variables.

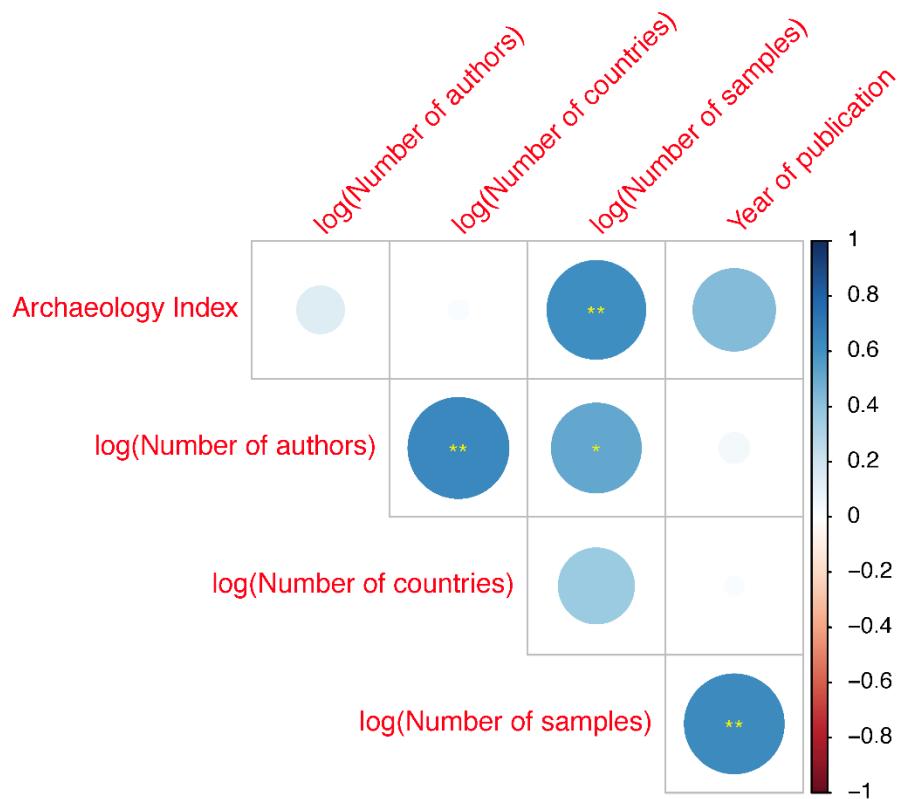


Fig. S5. Spearman correlations between candidate explanatory variables. Spearman's ρ is indicated by the color and the size of the circle whereas the significance level after Bonferroni correction is indicated by the asterisks ("*": $p_{adj} < 0.05$; "**": $p_{adj} < 0.01$; "***": $p_{adj} < 0.001$).

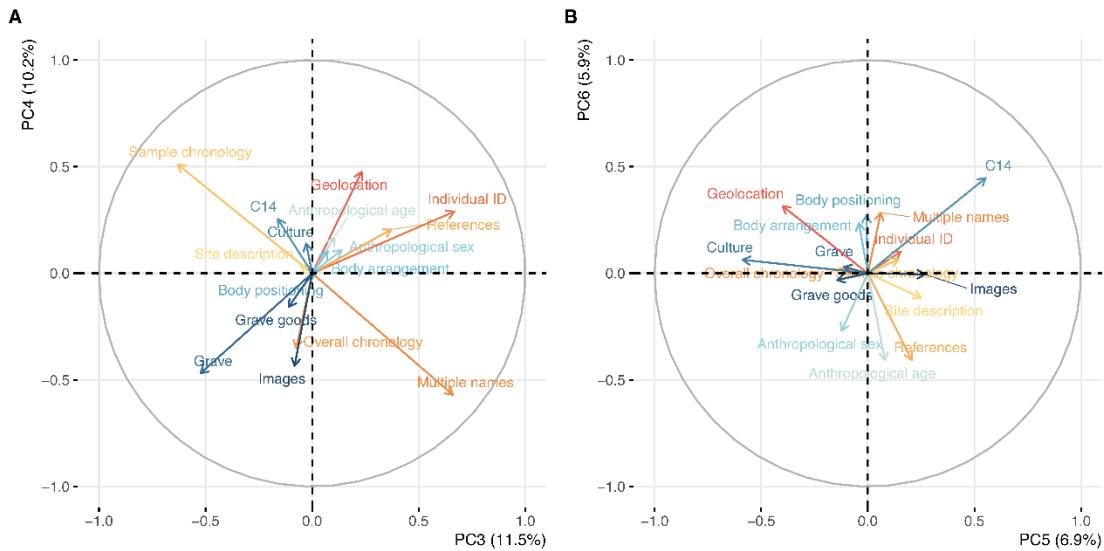


Fig. S6. The loading plots of PCA. The projected vectors correspond to the contributions of metadata variables to the PCs as measured by correlations. (A) Projections on PC3 and PC4. (B) Projections on PC5 and PC6.

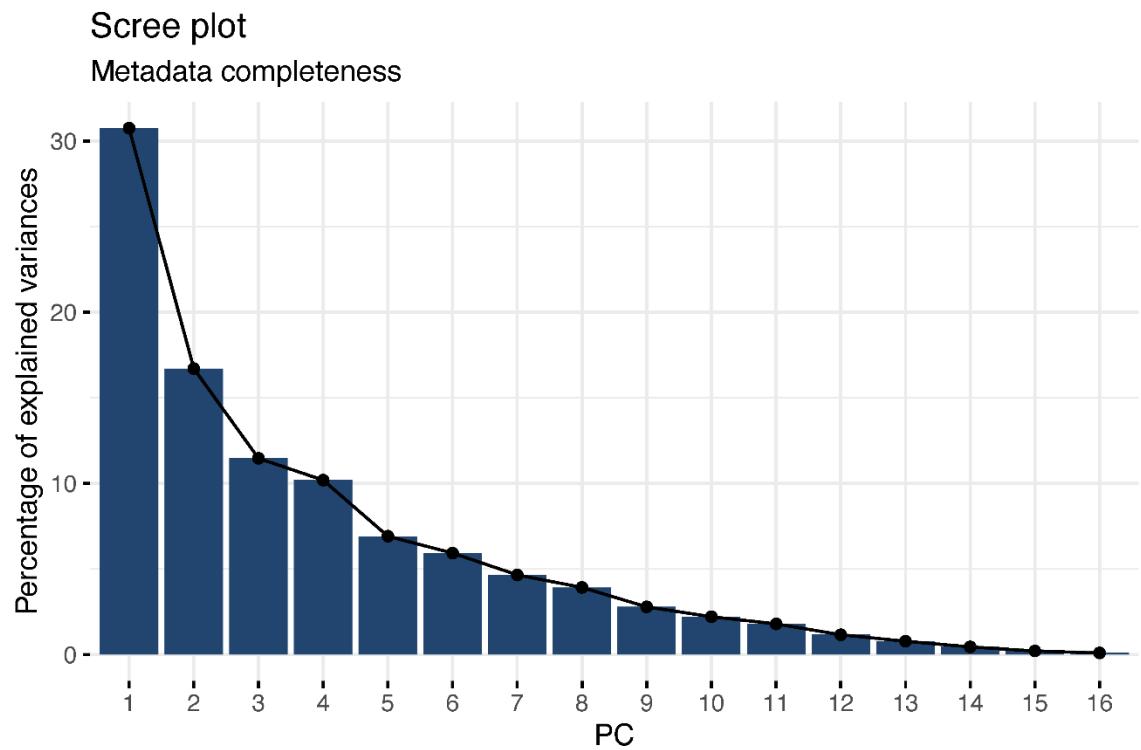


Fig. S7. The scree plot of PCA of metadata completeness.

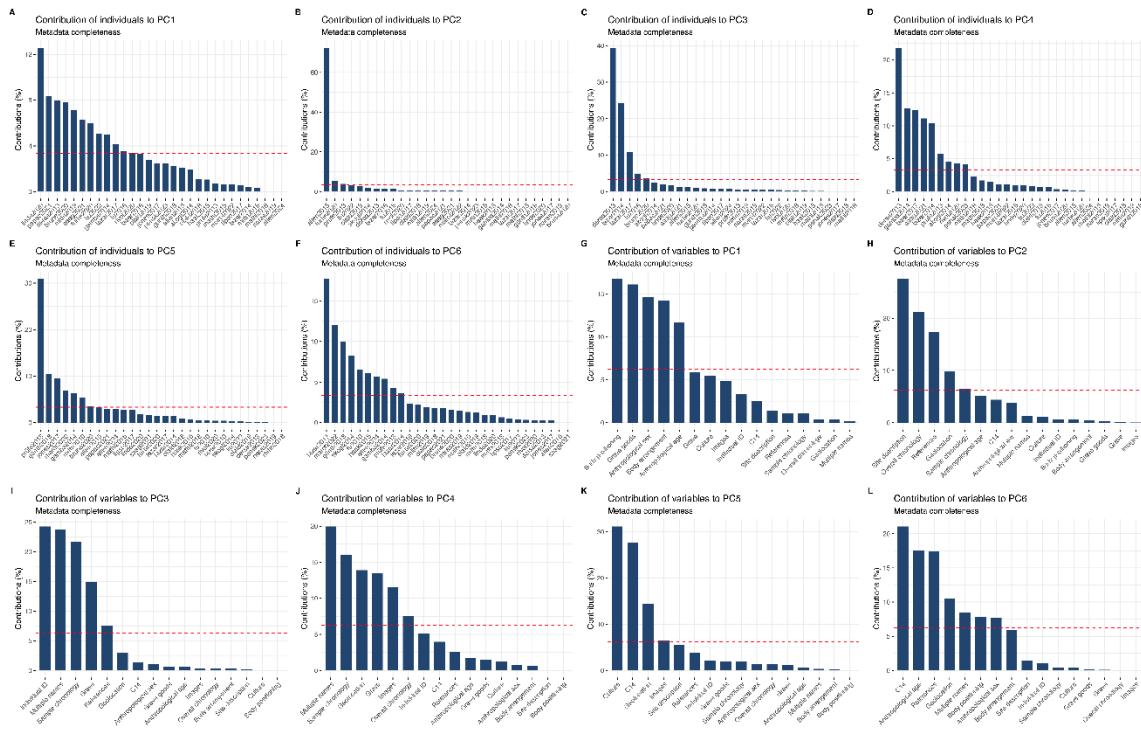


Fig. S8. Contributions of variables to the first 6 PCs of metadata completeness. (A-F) Contributions of publications. (G-L) Contributions of metadata variables.

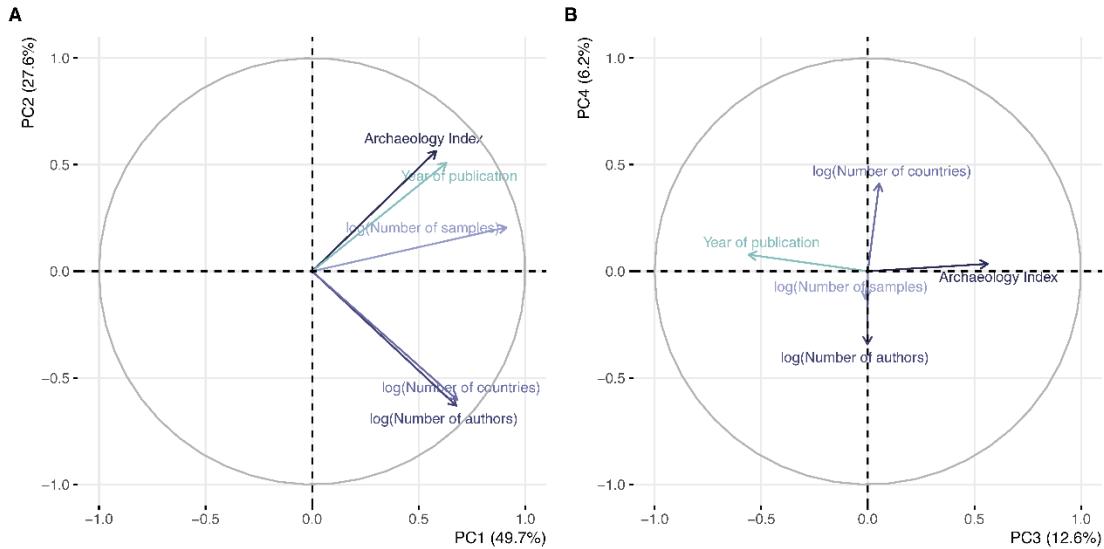


Fig. S9. The loading plots of PCA. The projected vectors correspond to the contributions of candidate explanatory variables to the PCs as measured by correlations. (A) Projections on PC1 and PC2. (B) Projections on PC3 and PC4.

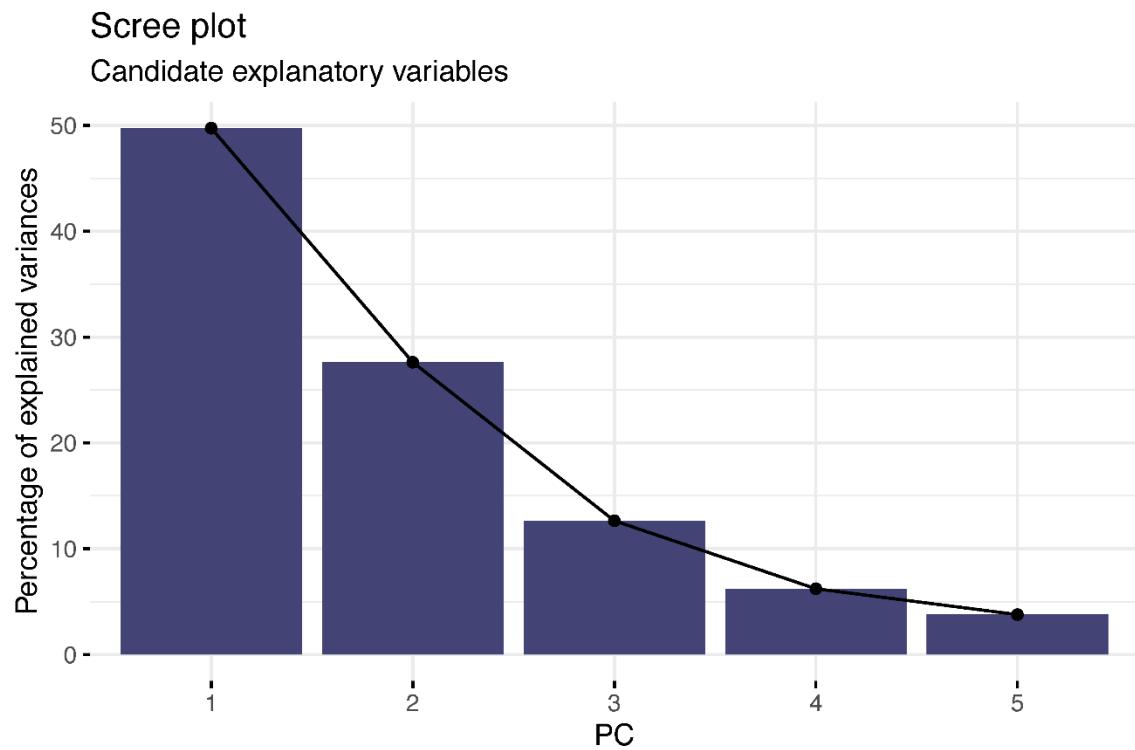


Fig. S10. The scree plot of PCA of metadata completeness.

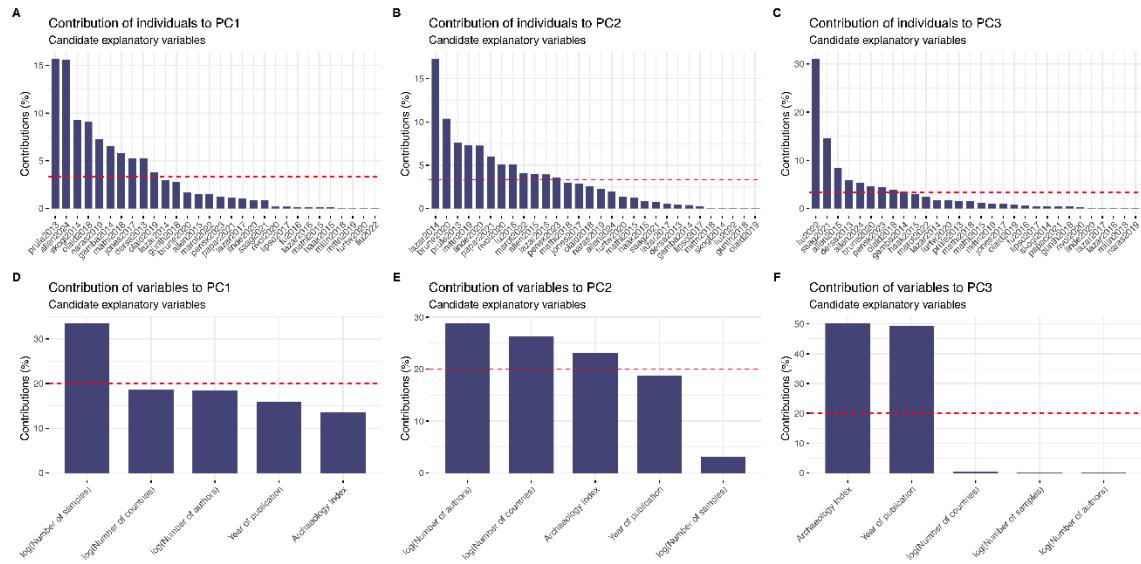


Fig. S11. Contributions of variables to the first 3 PCs of candidate explanatory variables. (A-C) Contributions of publications. (D-F) Contributions of metadata variables.

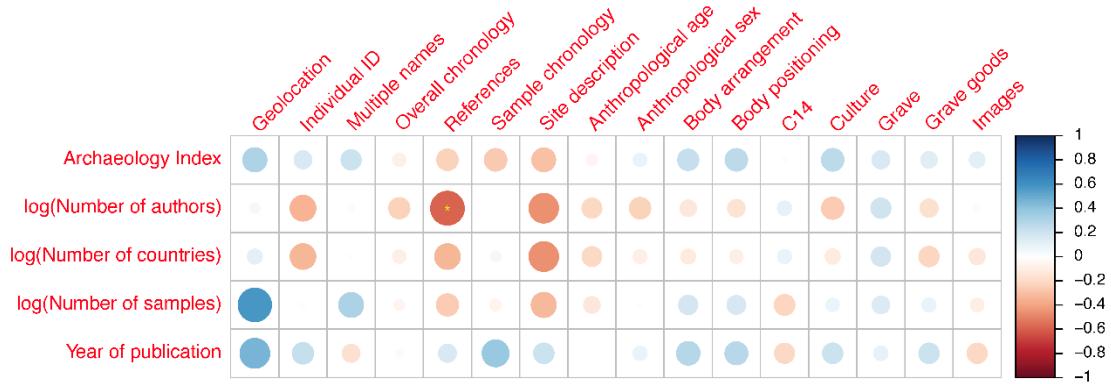


Fig. S12. Spearman correlations between metadata and candidate explanatory variables. Spearman's ρ is indicated by the color and the size of the circle and the significance level after Bonferroni correction is indicated by the asterisks ("*": $\text{padj} < 0.05$; "**": $\text{padj} < 0.01$; "***": $\text{padj} < 0.001$).

Fractional logistic regression

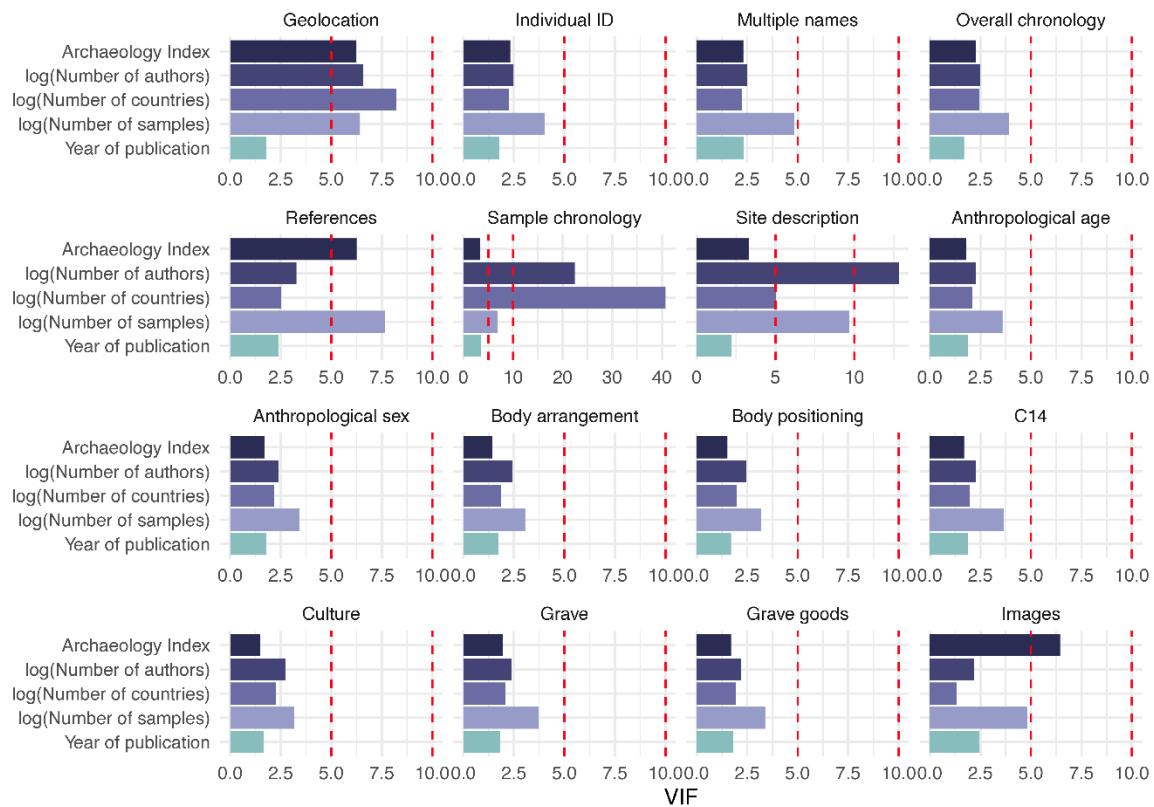


Fig. S13. VIFs of fractional logistic regression models.

Extended–support beta regression

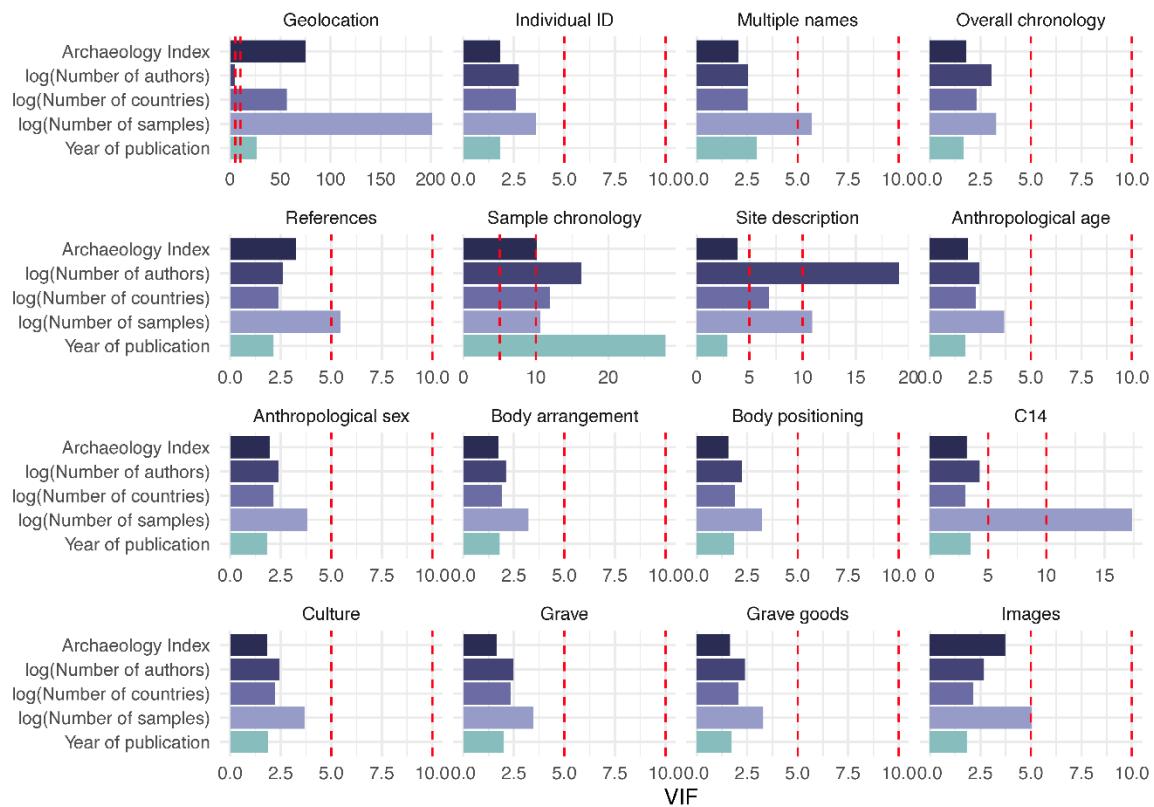


Fig. S14. VIFs of XBX regression model.

Fractional logistic regression

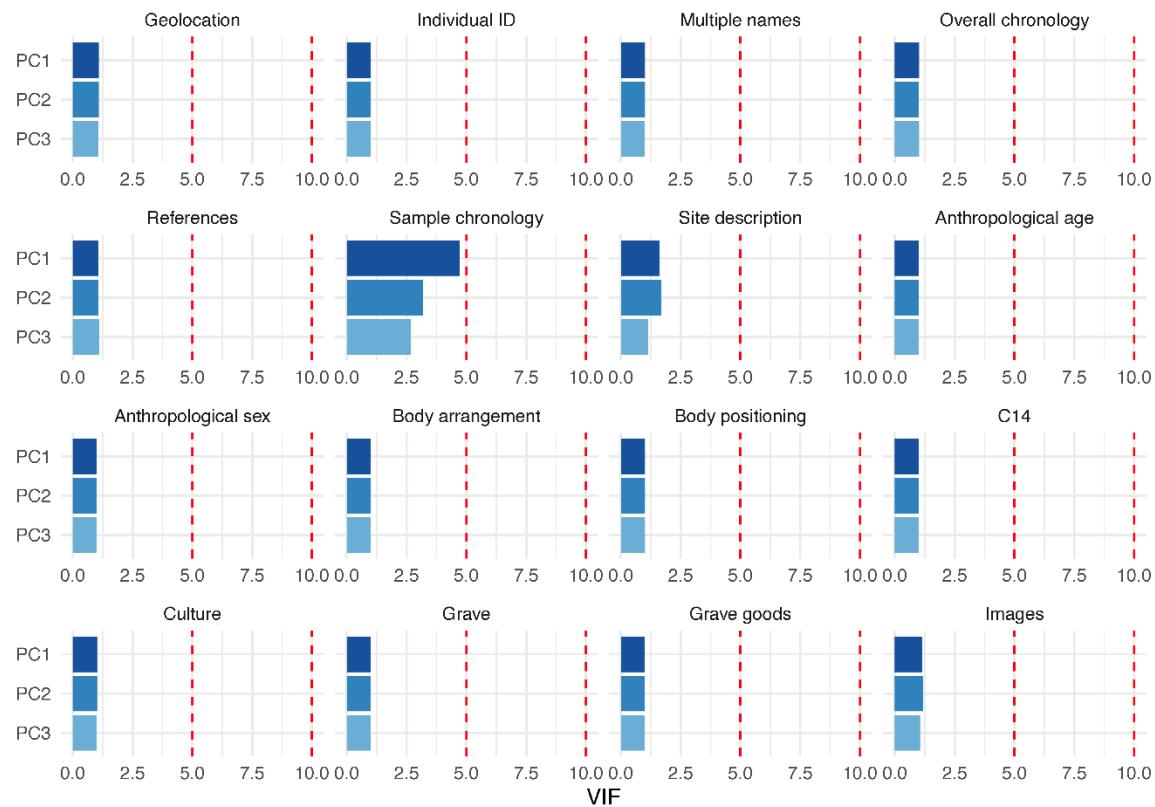


Fig. S15. VIFs of fractional logistic regression models with the first three PCs of the original variables as regressors.

Extended-support beta regression

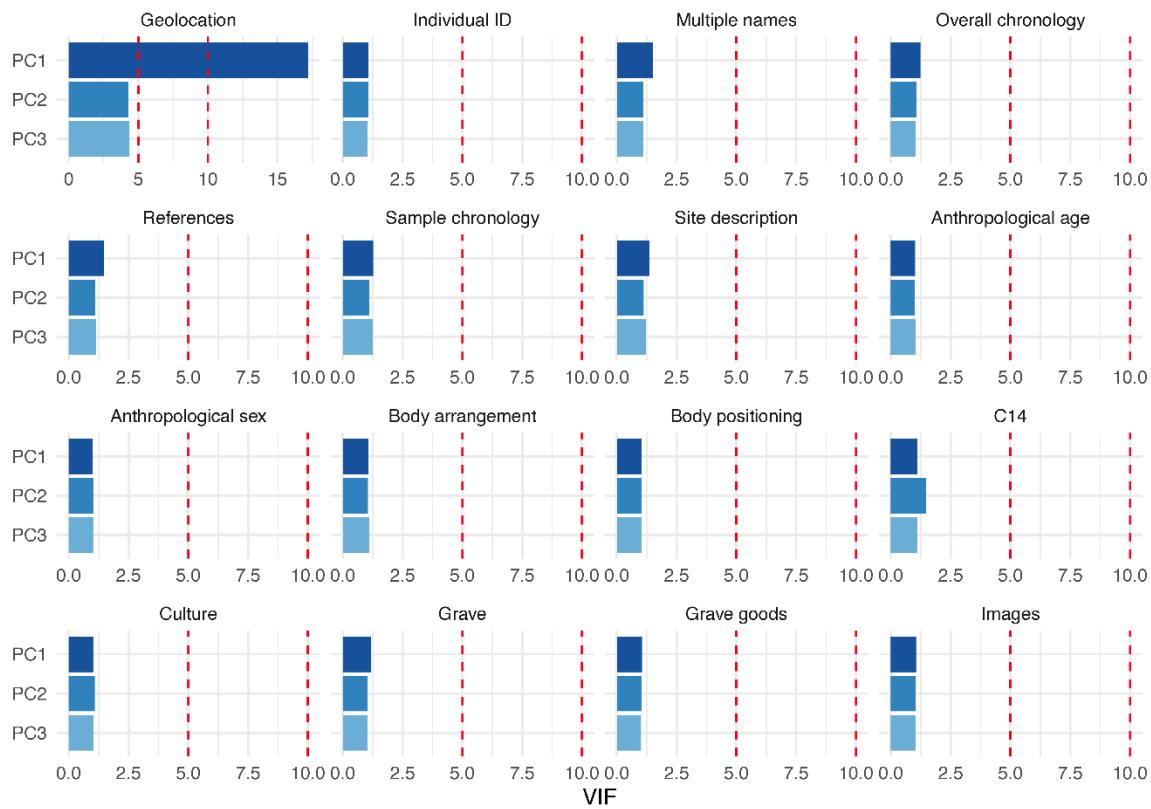


Fig. S16. VIFs of XBX regression models with the first three PCs of the original variables as regressors.

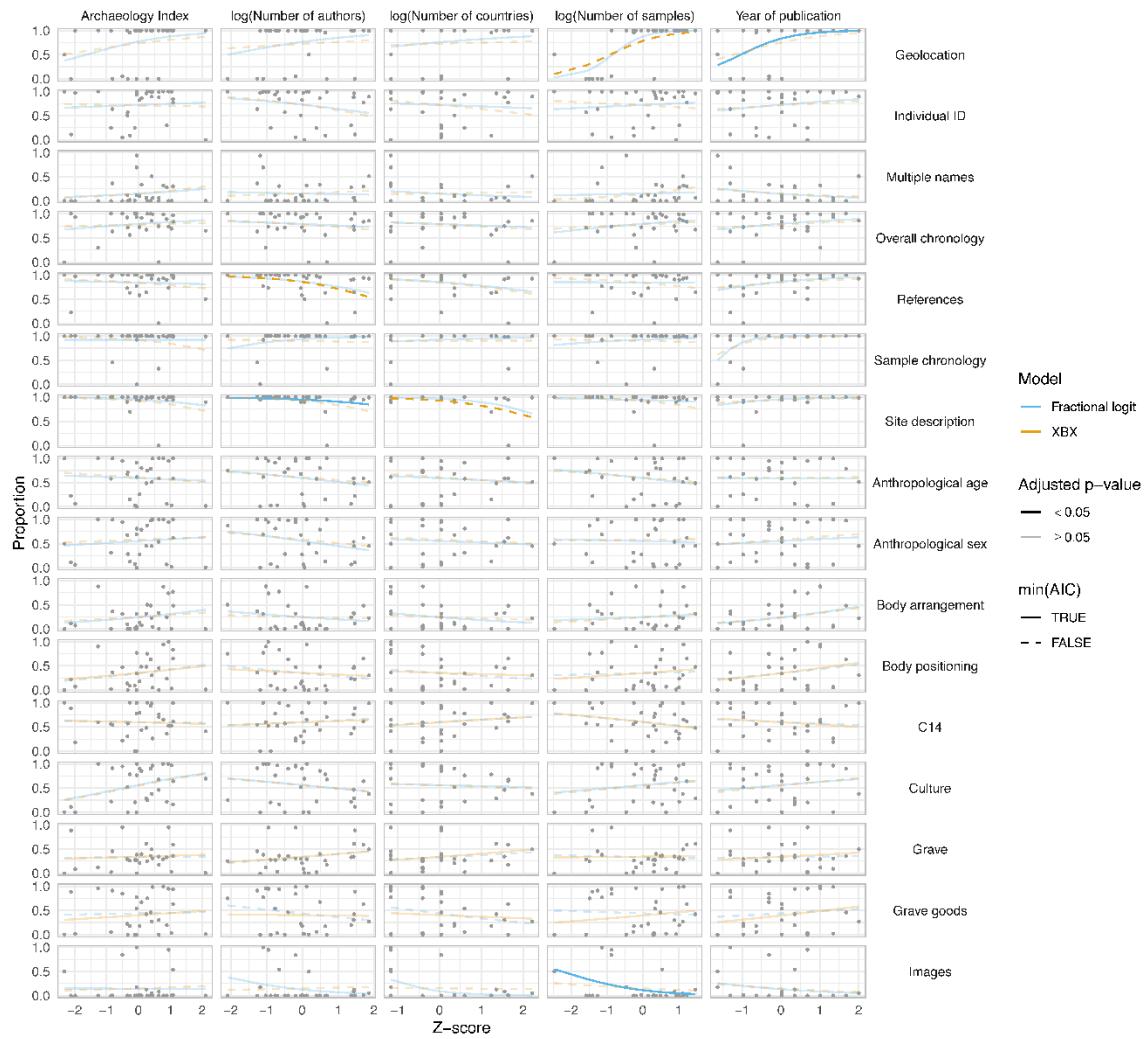


Fig. S17. Model fit of fractional logistic and XBX regression models. An opaque line indicates the coefficient estimate ($\hat{\beta}$) is significant after Bonferroni correction; a solid line indicates the model with the lower AIC.

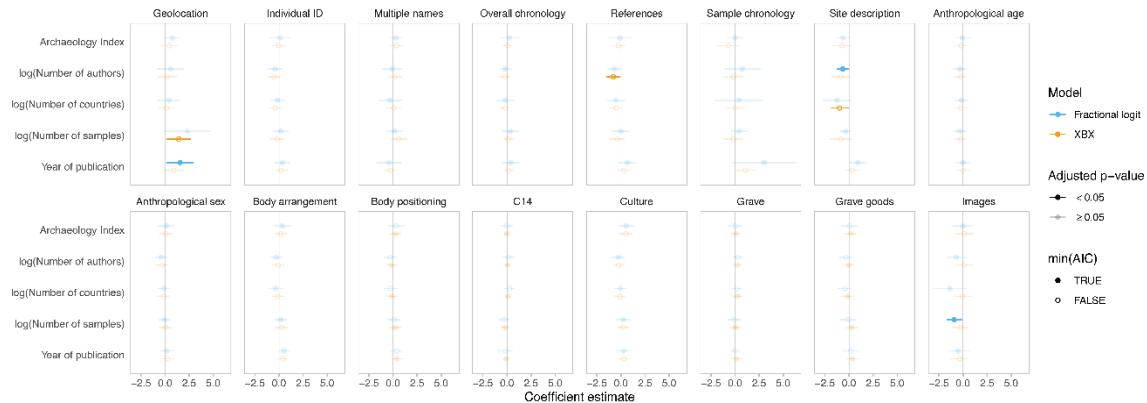


Fig. S18. The effects of candidate explanatory variables on metadata completeness. The line represents the 95% CI around the coefficient estimate ($\hat{\beta}$). An opaque line indicates the coefficient estimate is significant after Bonferroni correction; a solid circle indicates the model with the lower AIC.

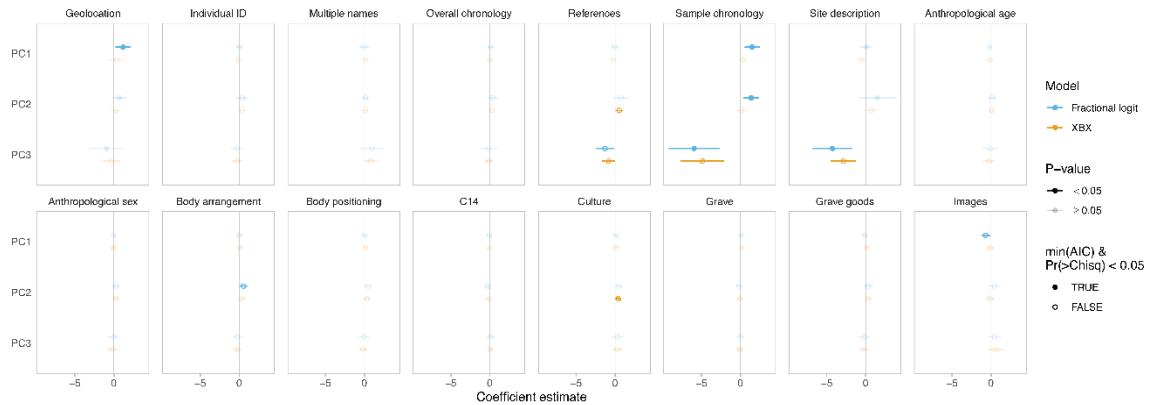


Fig. S19. The effects of the first three PCs of the original variables on metadata completeness. The line represents the 95% CI around the coefficient estimate. An opaque line indicates the coefficient estimate is significant; a solid circle indicates the model with the lower AIC that passes the Wald test for model fit.

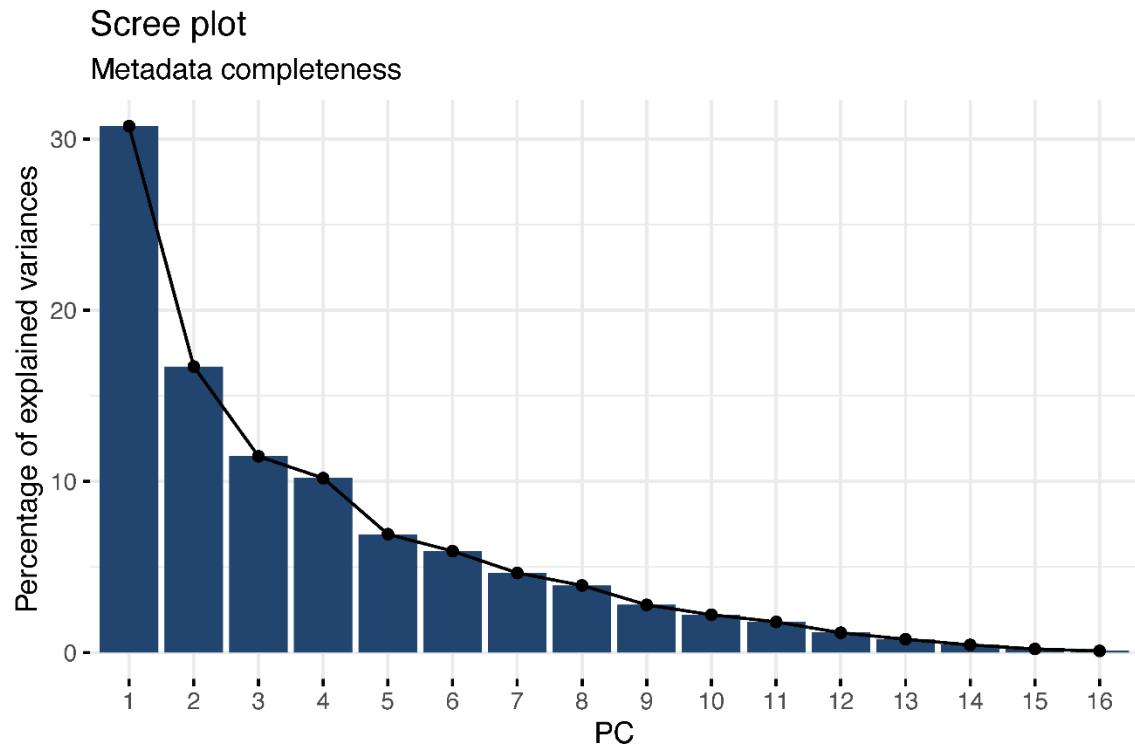


Fig. S20. Model comparisons between fractional logistic and XBX regression models in simple regressions. (A) Log-(pseudo*)likelihood. (B) AIC. (C) BIC. (D) Pseudo R^2 , computed as the squared correlation of the predicted and observed values. *Log-pseudolikelihood applies to the fractional logistic regression model.

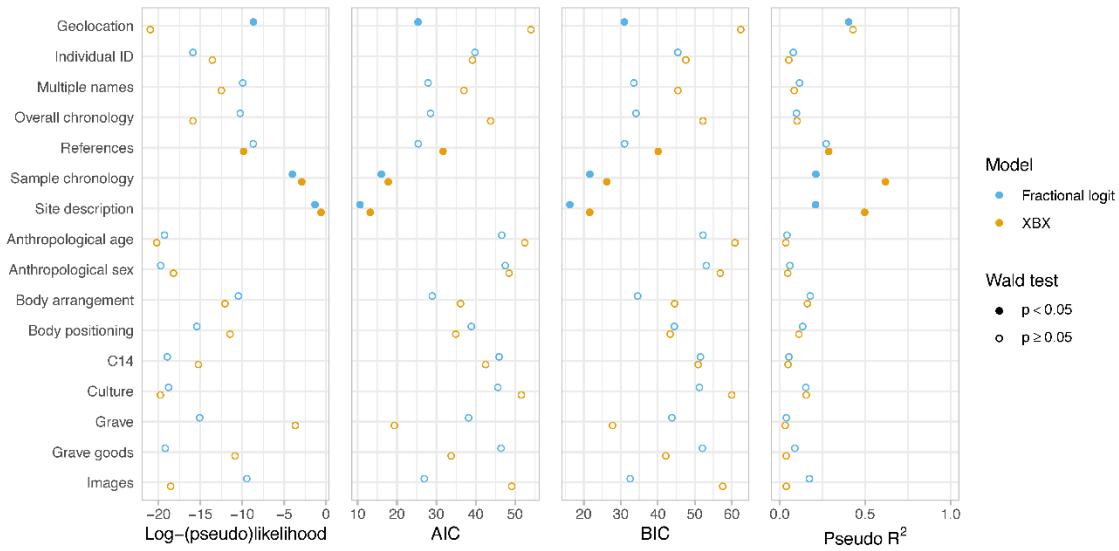


Fig. S21. Model comparisons between fractional logistic and XBX regression models in multiple regressions. Log-pseudolikelihood applies to the fractional logistic regression model. Pseudo R² is computed as the squared correlation of the predicted and observed values.

Dataset S1 (separate file). Spreadsheet in Microsoft Excel named as "2025-aDNAArch-Table-S1-S6".

Table S1: Included in the file as spreadsheet "Table S1". Summary of bibliographic data of papers used for the study.

Table S2: Included in the file as spreadsheet "Table S2". Summary of institutional affiliations reported in the sampled studies and their classification along the genetic-archaeological spectrum.

Table S3: Included in the file as spreadsheet "Table S3". Summary of cited author's institutional affiliations reported in the sampled studies.

Table S4: Included in the file as spreadsheet "Table S4". Summary of metadata survey per reported individual.

Table S5: Included in the file as spreadsheet "Table S5". Encoding table.

Table S6: Included in the file as spreadsheet "Table S6". Dunn (1964) Kruskal-Wallis multiple comparison p-values adjusted with the Bonferroni method.

Table S7: Included in the files as spreadsheet "Table S7". An example of how standardized archaeological metadata can be reported alongside paleogeonomic data.