BREADR: An R Package for the Bayesian Estimation of Genetic Relatedness from Low-coverage Genotype Data

Adam B. Rohrlach [1,2], **Jonathan Tuke** [3], **Divya Ratan Popli** [1], **and Wolfgang Haak** [1]

1 Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
2 School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia **3** School of Computer and Mathematical Sciences, University of Adelaide, Adelaide, South Australia, Australia

Summary

Robust and reliable estimates of how individuals are biologically related to each other are a key source of information when reconstructing pedigrees. In combination with contextual data, reconstructed pedigrees can be used to infer possible kinship practices in prehistoric populations. However, standard methods to estimate biological relatedness from genome sequence data cannot be applied to low coverage sequence data, such as are common in ancient DNA (aDNA) studies. Critically, a statistically robust method for assessing and quantifying the confidence of a classification of a specific degree of relatedness for a pair of individuals, using low coverage genome data, is lacking.

In this paper we present the R-package BREADR (**B**iological **RE**latedness from **A**ncient **D**NA in **R**), which leverages the so-called pairwise mismatch rate, calculated on optimally-filtered genome-wide pseudo-haploid sequence data, to estimate genetic relatedness up to the second degree, assuming an underlying binomial distribution. BREADR also returns a posterior probability for each degree of relatedness, from identical twins/same individual, first-degree, second-degree or "unrelated" pairs, allowing researchers to quantify and report the uncertainty, even for very low-coverage data. We show that this method accurately recovers degrees of relatedness for sequence data with coverage as low as 0.04X using simulated data (produced as in Popli et al.(Popli et al., 2023)).

Statement of need

BREADR is an R package for use in ancient DNA studies, and any studies estimating genetic kinship with extremely low-coverage DNA sequence data, such as in the forensic sciences. BREADR is designed to be a user-friendly software package, using as input pseudo-haploid genotype data in the Eigenstrat format, common in aDNA studies, and allowing for pre-processing quality control by researchers.

State of the Field

Most software packages for estimating genetic relatedness from aDNA sequence data leverage some measure of pairwise genetic dissimilarity, often calculated on pseudo-haploid data, where a random call is sampled for each site, to avoid the issue of phasing.

lcMLkin uses genotype likelihoods to estimate identity-by-descent coefficients (k_0, k_1 and k_2),

from which the co-ancestry coefficient can be calculated, such that $theta = k_1/4 + k_2/2$ (Lipatov et al., 2015). This method requires a reference panel to estimate the genotype likelihoods, and for sites to be unlinked. This method is exploratory in nature, and does not return hard classifications of degrees of relatedness, and is (currently) not peer-reviewed.

TKGWV2 instead estimates relatedness via the method of moments estimator from Queller and Goodnight, denoted $\hat{r}_{ij}$ (Fernandes et al., 2021). This asymmetric estimator assumes no inbreeding in the population and that all sites are unlinked (Queller & Goodnight, 1989). Hence, TKGWV2 cannot be applied to SNP capture data, or data from populations where inbreeding has occurred. This method returns hard classifications of the degree of relatedness based on which expected value of rij the observed value is closest to.

KIN (Popli et al., 2023) and ancIBD (Ringbauer et al., 2024) instead use the spatial distribution of identical-by-descent segments (IBDs). Like lcMLkin, KIN estimates the identity-by-descent coefficients, using IBDs instead of weighted genotype likelihoods, and can identify degrees of relatedness up to the third degree. Conversely, ancIBD uses imputation to increase the resolution of the method up to the seventh-degree. However, this additional resolution comes at the cost of requiring relatively high-coverage sequence data, approximately 0.25X for whole genome sequencing, compared to 0.05X, 0.026X or 0.04X for KIN, TKGV2 and BREADR, respectively.

READ and READv2 use the so-called pairwise-mismatch rate (PMR) to estimate the proportion of overlapping sites for which two individuals have non-matching genotype calls (Alaçamlı et al., 2024; Monroy Kuhn et al., 2018). The PMR was first introduced to overcome the limitations of low-coverage aDNA, however, the publication from Kennett *et al.* did not include a hard-classification method (Kennett et al., 2017). Where Kennett *et al.* used all of the available overlapping sites to estimate the PMR, READ uses a 1MB windowed approach to sample from the distribution of the PMR, but also to account for the effects of linkage disequilibrium (LD), allowing for the automated hard classification of degrees of relatedness. READv2, however, uses a genome-wide estimate in place of the windowed approach of READ.

Like READ and READv2, BREARD estimates the PMR, but instead thins the data to reduce the effects of LD. The PMR on this thinned data can now be reasonably expected to follow a binomial distribution, and theoretical expectations of the PMR are derived and compared to the observed value. From this, statistically rigorous measures of uncertainty are derived, and informative diagnostic plots of the degree of related for pairs of individuals can be produced. Finally, although BREADR only classifies degrees of related up to the second-degree, we also allow for tests of any degree up to the tenth-degree to be formally tested when considering potential pedigree reconstructions.

Current Applications

BREADR produces ready-for-publication results and diagnostic plots, and has already seen use in a number of publications (Ghalichi et al., 2024; Gretzinger et al., 2024; Penske et al., 2024; Pilli et al., 2024; Rivollat et al., 2023).

Estimating Degrees of Relatedness

Consider two individuals, i and j, with N_{ij} overlapping sites (without missingness), where sites are thinned (default value 1×10^5) such that the effects of LD are reduced. We then look at the number of pseudo-haploid genotype calls that do not match, denoted X_{ij}. Hence we have that $X_{ij} \sim \text{Bin}(N_{ij}, p_{ij})$, and that the maximum likelihood estimator for p_{ij} is $\hat{p}_{ij} = X_{ij}/N_{ij}$.

Once all PMRs are calculated, we must account for background relatedness, which can be thought of as the expected PMR for a pair of unrelated individuals, which will differ between populations and sequencing types (such as capture versus shotgun sequencing). Many choices

exist for this value, but assuming that a sampled dataset is made up of mostly unrelated pairs (to the second degree), then the median PMR, denoted $\bar{p}$ will be a reliable estimate (Monroy Kuhn et al., 2018). However, we also allow $\bar{p}$ to be a user-supplied parameter, which may be known from previous studies, or can be varied for sensitivity analyses.

Borrowing from the insights of READ, we now define the expected mean PMR for a relationship of degree $k = 0, 1, 2$ to be

$$p_k = \bar{p} \left(1 - \frac{1}{k+1} \right).$$

Hence, if we assume that the degree of relatedness for individuals i and j is truly of the kth degree, then $X_{ij} \sim \text{Bin}(N_{ij}, p_k)$.

Hence, the likelihood function for relatedness degree k, for individuals i and j, is

$$L\left(X_{ij}|K=k, N_{ij}\right) = \binom{N_{ij}}{X_{ij}} p_k^{X_{ij}} (1-p_k)^{N_{ij}-X_{ij}}.$$

If we let $k = \infty$ be the case that two individuals are "unrelated" (i.e. more than second-degree related), but then we also have that

$$P\left(X_{ij}|K \geq 3, N_{ij}\right) = \sum_{k=3}^{\infty} P\left(K=k|X_{ij}, N_{ij}\right) P\left(K=k\right)$$
$$= \binom{N_{ij}}{X_{ij}} p_k^{X_{ij}} (1-p_k)^{N_{ij}-X_{ij}} f_\lambda(k),$$

where $f_\lambda(k)$ is the 3-truncated Poisson distribution of the form

$$f_\lambda(k) = \frac{1}{\sum_{\ell=0}^{2} \frac{\lambda^\ell e^{-\lambda}}{\ell!}} \left(\frac{\lambda^k e^{-\lambda}}{k!} \right).$$

We choose the 3-truncated Poisson distribution with $\lambda = 10$ as it represents well the unlikelihood of individuals always being closely related once they are more than second-degree related, but also captures the diminishing probabilities of being too distantly-related due to the finite size of populations.

It is then possible to calculate the normalised posterior probabilities of individuals i and j being kth degree related as

$$P\left(K=k|X_{ij}, N_{ij}\right) = \frac{L\left(X_i j|K=k, N_{ij}\right) P\left(K=k\right)}{\left[\sum_{\ell=0}^{2} L\left(X_{ij}|K=\ell, N_{ij}\right) \right] + L\left(X_{ij}|K \geq 3, N_{ij}\right)}$$

for which the denominator equals one, by construction. Hence,

$$P\left(K=k|X_{ij}, N_{ij}\right) = L\left(X_i j|K=k, N_{ij}\right) P\left(K=k\right).$$

However, the prior probabilities for the degrees of relatedness are unclear, and will vary from study to study, and so these are given as an uninformative uniform distribution by default. Since these are user inputs, they should be thought about carefully, and can be explored via sensitivity analyses to explore if the prior probabilities are driving the relatedness classifications.

Finally, while we do not include third-degree or higher relationships in the possible classifications, we do include a method for a statistical test for retaining or rejecting the possibility of a kth-degree relationship, for $0 \leq k \leq 10$. For this, we perform a binomial test for the observed number of mismatches X_{ij}, and the expected probability of pairwise mismatches, as defined in Equation 1, returning a two-sided p-value and the estimated (non-integer) degree of relatedness, found by setting $p_k = p_{ij}$ and $k = \hat{k}$ into Equation 2, and solving for k, i.e.,

$$\hat{k} = \frac{ln(1 - p_{ij}/\hat{p})}{ln(1/2)},$$

if $p_{ij} < \hat{p}$, and set $\hat{k} = \infty$ otherwise, with $k = \infty$ denoting "unrelated".

Example Usage

To show an example of how one might use BREADR to analyse some data, we simulated genotype data from a pedigree of individuals as shown in Figure 1. This pedigree has 2 genetically identical pairs of individuals, 19 first-degree relatives, 12 second-degree relatives (2 of which are half-siblings) and 87 unrelated pairs of individuals. The pedigree also contains a further 9, 6 and 1 third-, fourth- and fifth-degree relatives, respectively, although our method does not attempt to directly identify these deeper relationships. The simulated data resulted in genotypes being called for 29,767,814 segregating sites, at an average of 0.04X coverage.
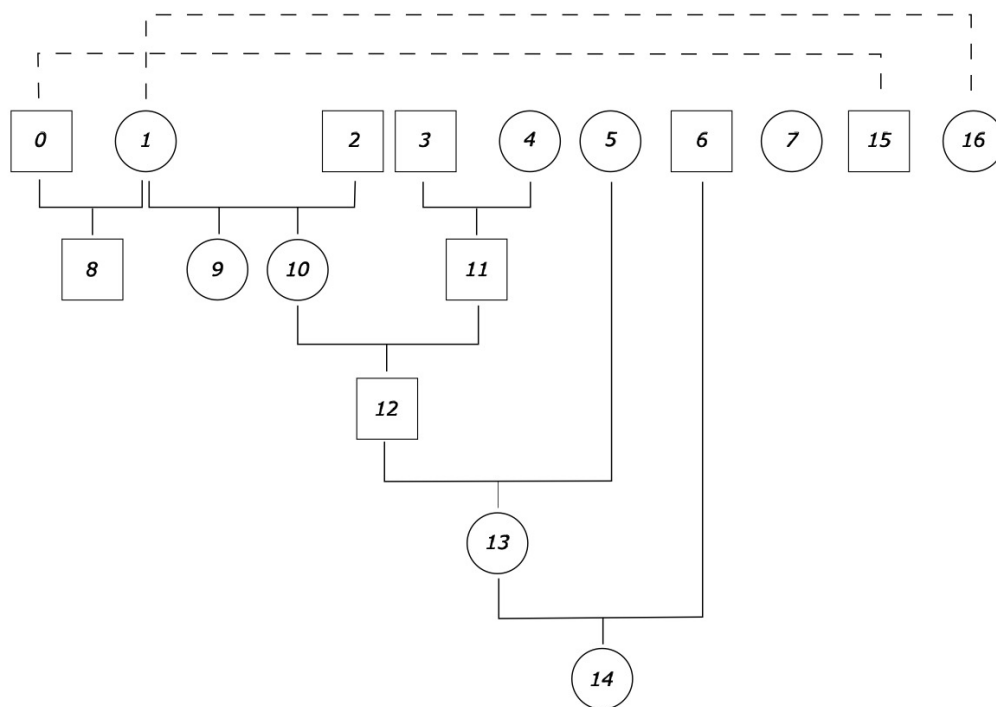

Figure 1: Pedigree for the simulated data. Individuals 0 and 15, and 1 and 16, are genetically identical, respectively. Females are represented by circles, males are represented by squares. Dashed lines indicate identical individuals (such as twins or samples taken from the same individual).

The package BREADR is available under the MIT license from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=BREADR. The package can also be installed and loaded using the following commands:

```
> devtools::install_github("jonotuke/BREADR")
> require(BREADR)
```

The package contains the functions for a full analysis of the genetic relatedness for a large number of individuals, and the visualisation of the overall genetic relatedness, as well as pairwise diagnostic plots to make informed pedigree reconstructions. The raw input data is genotype data called in the Eigenstrat format, and this allows for user pre-processing for data characteristics such as: removing ancient DNA damage, minimum allele frequency cut-offs, random versus majority genotype calls, read and mapping quality scores, etc.

The analysis begins with a processing step where genotype data is converted into count data for each pair of individuals. As this is the most computationally intensive step in the analysis, all downstream analyses are performed on the output of this first step. From this data we calculate the genetic relatedness statistics, calculate posterior probabilities, and make highest posterior probability assignments, for the degrees of genetic relatedness. Lastly, we provide functions for visualising the genetic relatedness estimates and classifications, and for performing ad hoc tests for user defined degrees of relatedness via a binomial test.

Preprocessing the Eigenstrat data for analysis

All downstream functions in the BREADR package require preprocessing of the three standard Eigenstrat files: so-called ind, snp and geno files. During this first step, we take the list of sites on the genome of interest, with the chromosome name and (integer) site position, and the pseudo-haploid genotype calls for all individuals (of interest). Then, for each pair, we take only sites which have overlapping, non-missing calls, and are at least some user-defined number (by default 5×10^5) of positions apart (within chromosomes). Using these site positions, we record the number of filtered, overlapping sites, as well as the number of mismatches per pair.

To generate these required genetic pairwise comparisons for all individuals, users apply the `processEigenstrat()` function. This function requires three input string parameters: the paths to the three Eigenstrat files. At this stage of pre-processing, users have four additional parameters that can be set, and cannot be reset downstream.

First, the *filter_length* parameter (default 1×10^5) allows the user to define the minimum number of positions between sites that can be considered. By removing sites which are close to one another, the assumption of independence for each site-wise comparison is best attained by lowering or removing the effects of linkage disequilibrium.

Second, the *pop_pattern* parameter (default exclude nothing) allows the user to define a set of population names so that only a subset of the individuals is compared. Third, the *filter_deam* parameter (default NO) filters C to T or G to A SNPs from the possible list of sites if the potential effects of post-mortem deamination have not been accounted for in the genotyping process.

Last, since pre-processing is the most computationally expensive step of the analysis, the *outfile* parameter (default NO) is a path-string that allows the user to automatically save the post-processed data as a TSV file.

```
# A tibble: 136 × 4
    pair          nsnps mismatch  pmr
    <chr>         <dbl>    <dbl> <dbl>
1 Ind0 - Ind1   11663     1561 0.134
2 Ind0 - Ind2   11578     1552 0.134
3 Ind0 - Ind3   11600     1531 0.132
4 Ind0 - Ind4   11676     1560 0.134
5 Ind0 - Ind5   11658     1558 0.134
6 Ind0 - Ind6   11587     1592 0.137
```

Figure 2: The first 6 rows of the tibble produced by running `processEigenstrat()` on the simulated data set.

The resulting (see Figure 2) tibble has four columns: the names of the samples/individuals which were compared (pair), the number of overlapping SNPs (nsnps) per pair, the number of overlapping sites for which the pair did not match (mismatch) and the pairwise-mismatch rate (pmr).

Analysing the post-processed data

Following the pre-processing step, the `callRelatedness()` function calls a relationship of Same_Twins, First_Degree , Second_Degree or Unrelated, with additional information, from the output of the `processEigenstrat()` function. Additional parameters can be set to allow the user to customise the analysis.

The *class_prior* parameter (default Uniform) defines the prior probabilities of each relatedness class. The *average_relatedness* parameter defines what the background relatedness should be estimated via (default is the median from the filtered estimates). This can be either: (a) estimated directly from the median PMR from the data, (b) a single value (which is useful for sensitivity analyses when the user is uncertain), or (c) a vector of values equal to the number of rows in the input tibble (which can be useful if different populations have different background relatedness levels). The *median_co* parameter (default 500) defines the minimum number of overlapping SNPs a pair of individuals must share for their PMR to be used in estimating the median PMR, if the user has elected to use the median PMR. Finally, the *filter_n* parameter (default 1) is used to simply remove any pairs of individuals from the analysis if they share fewer overlapping SNPs than this value.

```
# A tibble: 136 × 12
  row pair        relationship   pmr      sd mismatch nsnps ave_rel Same_Twins First_Degree Second_Degree Unrelated
<int> <chr>       <fct>        <dbl>   <dbl>    <dbl> <dbl>   <dbl>      <dbl>        <dbl>         <dbl>     <dbl>
1    1 Ind0 - Ind1 Unrelated    0.134 0.00315     1561 11663   0.131 9.25e-151     6.94e-33      4.84e- 9  1.00e+ 0
2    2 Ind0 - Ind2 Unrelated    0.134 0.00317     1552 11578   0.131 1.87e-150     5.53e-33      3.90e- 9  1.00e+ 0
3    3 Ind0 - Ind3 Unrelated    0.132 0.00314     1531 11600   0.131 8.57e-143     1.09e-29      1.37e- 7  1.00e+ 0
4    4 Ind0 - Ind4 Unrelated    0.134 0.00315     1560 11676   0.131 5.11e-150     1.55e-32      7.16e- 9  1.00e+ 0
5    5 Ind0 - Ind5 Unrelated    0.134 0.00315     1558 11658   0.131 6.37e-150     1.52e-32      6.94e- 9  1.00e+ 0
6    6 Ind0 - Ind6 Unrelated    0.137 0.00320     1592 11587   0.131 1.89e-163     1.87e-38      1.11e-11  1.00e+ 0
```

Figure 3: The first 6 rows of the tibble produced by running `callRelatedness()` on the simulated data set as in Figure 2.

The resulting tibble (see Figure Figure 3) has an additional 9 columns: the row number (pair which is useful for additional functions, the highest-posterior genetic relationship (relationship),

the standard error of the estimate of the PMR (sd), the median used in the calculations (med), and the normalised posterior probabilities for each of the four relatedness categories (Same_Twins, First_Degree, Second_Degree and Unrelated)).

Visually interpreting results

Once posterior probabilities of degrees of relatedness have been calculated, the results can visualised in two ways.

First, an overall picture of the relatedness can be obtained using the `plotLOAF()` function which plots the first N (default $N = 50$) pairs of individuals, sorted by ascending PMR value (see Figure 4). The shape and colour of the point estimates of the PMR indicate the highest posterior degree of relatedness assigned for each pair, with an associated 95% confidence interval represented by the vertical error bars. The dashed coloured lines indicated the expected PMR for each degree of relatedness (given the background relatedness). Note that if we have n individuals, then we will have $\frac{n(n-1)}{2}$ pairs of individuals, which grows factorially as n gets larger. Hence, plotting all pairs of individuals will be infeasible in many cases, and since the majority of pairs will likely be unrelated, users may only wish to plot the closely related pairs of individuals for ease of visualisation. From Figure 4 we can observe two genetically identical individuals, and we also get a scale of the number of first- and second-degree related individuals.
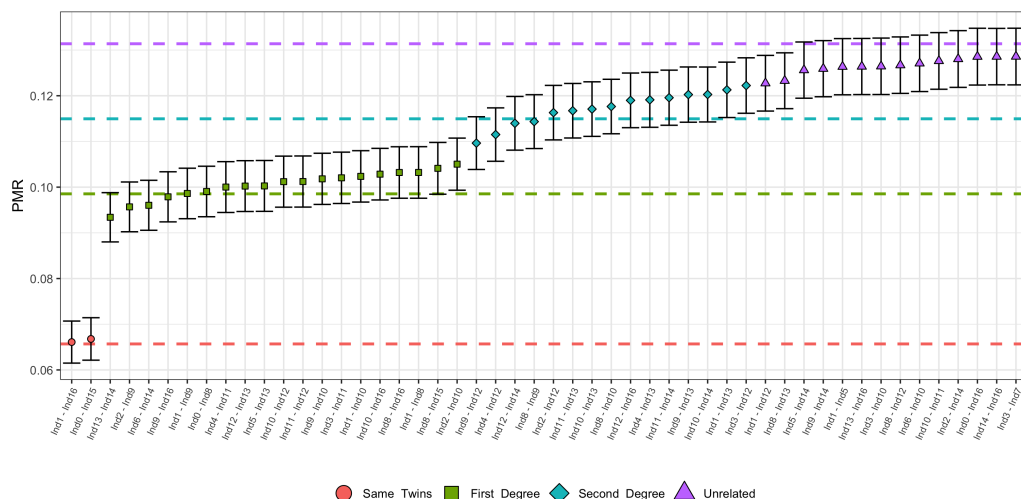

Figure 4: The first 50 ordered pairwise-mismatch rates (PMRs) from the simulated data. Colours indicate classification, dashed lines indicate the expected PMR for each degree of relatedness, and the error bars indicate two standard errors.

Second, a diagnostic plot for the analysis of a single pair of individuals can be obtained using the `callSLICE()` function (see Figure 5). By default, this function produces a two panel plot. Panel A displays the distribution of the PMR for each of the degrees of relatedness (given the number of overlapping sites), as well as the observed PMR (and 95% confidence interval) plotted below these densities. Panel B displays the normalised posterior probabilities for each possible degree of relatedness, both visually and numerically. The user can choose to return just one of these plots, or both. In Figure 5, we misclassify the true second-degree relationship as unrelated, however, the posterior probabilities should lead a researcher to consider a second-degree relationship plausible.
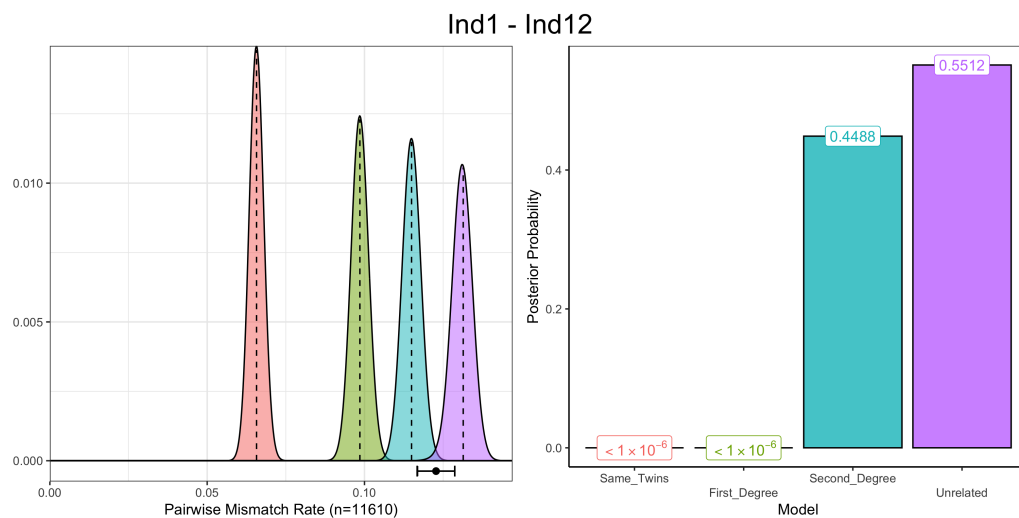
Figure 5: The diagnostic plot for the analysis of the genetic relationship Ind1 and Ind12. Panel A indicates the observed PMR falls between the expected distributions for a second- degree relationship (blue) and an unrelated pair (purple). This is uncertainty is quantified in panel B where the normalised posterior probability assigned to a second-degree relationship is 0.4488, and 0.5512 for unrelated.

Finally, the `savesSLICES()` function allows the user to save all possible pairwise diagnosticplots (as produced by `plotSLICE()`) as PDFs to an output path.

Testing for departure from refined degrees of relatedness

The resolution of the `BREADR` method only allows degrees of relatedness of up to the second degree to be assigned. However, once can in principle test to see if the observed PMR is consistent with any degree of relatedness using a simple binomial test.

```
Testing H0       : "Ind1 - Ind12" are 3rd-degree relatives.
Expected PMR     : 0.1232
Observed PMR     : 0.1227
Estimated degree : 2.9268
p-value          : 0.9025
Decision         : Retain H0
```

Figure 6: The output from the binomial test for whether the PMR observed for Ind1 and Ind12 is consistent with a third-degree relationship. We correctly classified the relationship as "unrelated", as we only report degrees of relatedness to the resolution of the second-degree. However, the binomial test reveals that the observed data is not inconsistent with a third- degree relationship (p = 0.9025), with an estimated degree of $\hat{k} = 2.9268$.

The `test_degree()` function allows the user to test any degree of relatedness up to the tenth degree. If the *printResults* option is set to TRUE, then all of the information about the binomial test is displayed (see Figure 6). From this the null hypothesis, the expected and observed PMR, the estimate of the degree of relatedness, the associated p-value and the decision (at significance level $\alpha = 0.05$) are given. Figure [Figure 6] shows a case where we are able to correctly identify that a third-degree relationship is reasonable for Ind1 and Ind12.

Conclusion

This article describes the usage and functionality of the R package BREADR. Based on the well-understood metric of the PMR, BREADR can infer pairwise genetic relatedness between individuals up to the second degree, and use binomial tests for consistency with PMRs of higher degrees of relatedness. BREADR can calculate Bayesian posterior probabilities for the classification of genetic relationships on even very low-coverage, pseudohaploid sequence data, in the Eigenstrat data format, common in aDNA studies. Additionally, BREADR provides functionality for plotting the overall genetic relatedness, as well as single plots for specific pairs when a need for exploring the strength of evidence for alternative potential genetic relationships is required. Finally, by allowing for researchers to explore the possibility of other potential levels of genetic relatedness in a statistically rigorous framework, quality control analyses and pedigree reconstructions can be performed with more flexibility.

We concede that BREADR has the same limitations of any method that relies on estimating the pairwise-mismatch: namely that estimating the average pairwise-mismatch rate for two unrelated individuals is both difficult to do (and to assess), and is not likely to be constant over large pedigrees, or across populations. We also identify a need to consider the empirical variability of the expected pairwise-mismatch rate for k-degree relatives that can be observed in large modern datasets. However, we encourage researchers to consider these limitations when using BREADR, as well as to use multiple methods for estimating relatedness.

The usefulness for researchers to be able assign statistical confidence to each and every estimated degree of relatedness, even in cases when coverage is extremely low, is critical when reconstructing a pedigree. We have also shown that when BREADR misclassifies genetic relationships, it is due to a true closer genetic relationship (third-degree genetic relationships were misclassified as unrelated). The easy-to-use R-implementation of our method makes BREADR an attractive, and statistically rigorous tool for researchers in archaeogenetics to employ.

Data Description

The genetic data is real, but anonymised, ancient DNA data produced in house at the Max Planck Institute for Evolutionary Anthropology, in Leipzig, Germany. The data comes from a site dated to 2350 BCE. The data was produced using the Illumina sequencing platform, and after adapter trimming, reads were aligned to the human reference genome GRCh37 using BWA(Li & Durbin, 2009) as implemented in EAGER(Peltzer et al., 2016). Sequence reads were filtered for a minimum sequence length of 35 and a mapping quality of at least 25. Data was genotyped using pileupCaller[https://github.com/stschiff/sequenceTools/tree/master/src/SequenceTools] using a random pseudohaploid call for each position on the 1240k capture assay(Mathieson et al., 2015). The processed genotype data can be found at: https://github.com/jonotuke/BREADR/tree/master/data-raw/.

Author Contributions

Adam B. Rohrlach conceived, designed and implemented the method, wrote the paper, prepared the figures, performed the computation work, reviewed drafts of the paper and edited submissions.

Jonathan Tuke conceived the method, created and submitted the package to CRAN, reviewed drafts of the paper and edited submissions.

Divya Ratan Popli created and implemented the computations for benchmarking the accuracy and performance of the method, reviewed drafts of the paper and edited submissions.

Wolfgang Haak conceived the method, reviewed drafts of the paper and edited submissions.

Acknowledgements

Funding

References

Alaçamlı, E., Naidoo, T., Güler, M. N., Sağlıcan, E., Aktürk, Ş., Mapelli, I., Vural, K. B., Somel, M., Malmström, H., & Günther, T. (2024). READv2: Advanced and user-friendly detection of biological relatedness in archaeogenomics. *Genome Biology*, *25*(1), 216. https://doi.org/10.1186/s13059-024-03350-3

Fernandes, D. M., Cheronet, O., Gelabert, P., & Pinhasi, R. (2021). TKGWV2: An ancient DNA relatedness pipeline for ultra-low coverage whole genome shotgun data. *Scientific Reports*, *11*(1), 21262. https://doi.org/10.1038/s41598-021-00581-3

Ghalichi, A., Reinhold, S., Rohrlach, A. B., Kalmykov, A. A., Childebayeva, A., Yu, H., Aron, F., Semerau, L., Bastert-Lamprichs, K., Belinskiy, A. B., & others. (2024). The rise and transformation of bronze age pastoralists in the caucasus. *Nature*, 1–9. https://doi.org/10.1038/s41586-024-08113-5

Gretzinger, J., Schmitt, F., Mötsch, A., Carlhoff, S., Lamnidis, T. C., Huang, Y., Ringbauer, H., Knipper, C., Francken, M., Mandt, F., & others. (2024). Evidence for dynastic succession among early Celtic elites in Central Europe. *Nature Human Behaviour*, 1–14. https://doi.org/10.1038/s41562-024-01888-7

Kennett, D. J., Plog, S., George, R. J., Culleton, B. J., Watson, A. S., Skoglund, P., Rohland, N., Mallick, S., Stewardson, K., Kistler, L., & others. (2017). Archaeogenomic evidence reveals prehistoric matrilineal dynasty. *Nature Communications*, *8*(1), 14115. https://doi.org/10.1038/ncomms14115

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lipatov, M., Sanjeev, K., Patro, R., & Veeramah, K. R. (2015). Maximum likelihood estimation of biological relatedness from low coverage sequencing data. *BioRxiv*, 023374. https://doi.org/10.1101/023374

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S. A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., & others. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, *528*(7583), 499–503. https://doi.org/10.1038/nature16152

Monroy Kuhn, J. M., Jakobsson, M., & Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PloS One*, *13*(4), e0195491. https://doi.org/10.1371/journal.pone.0195491

Peltzer, A., Jäger, G., Herbig, A., Seitz, A., Kniep, C., Krause, J., & Nieselt, K. (2016). EAGER: Efficient ancient genome reconstruction. *Genome Biology*, *17*, 1–14. https://doi.org/10.1186/s13059-016-0918-z

Penske, S., Küßner, M., Rohrlach, A. B., Knipper, C., Nováček, J., Childebayeva, A., Krause, J., & Haak, W. (2024). Kinship practices at the early bronze age site of leubingen in Central Germany. *Scientific Reports*, *14*(1), 3871. https://doi.org/10.1038/s41598-024-54462-6

Pilli, E., Vai, S., Moses, V. C., Morelli, S., Lari, M., Modi, A., Diroma, M. A., Amoretti, V., Zuchtriegel, G., Osanna, M., & others. (2024). Ancient DNA challenges prevailing interpretations of the Pompeii plaster casts. *Current Biology*, *34*(22), 5307–5318. https://doi.org/10.1016/j.cub.2024.10.007

Popli, D., Peyrégne, S., & Peter, B. M. (2023). KIN: A method to infer relatedness from low-coverage ancient DNA. *Genome Biology*, *24*(1), 10. https://doi.org/10.1186/s13059-023-02847-7

Queller, D. C., & Goodnight, K. F. (1989). Estimating relatedness using genetic markers. *Evolution*, *43*(2), 258–275. https://doi.org/10.1111/j.1558-5646.1989.tb04226.x

Ringbauer, H., Huang, Y., Akbari, A., Mallick, S., Olalde, I., Patterson, N., & Reich, D. (2024). Accurate detection of identity-by-descent segments in human ancient DNA. *Nature Genetics*, *56*(1), 143–151. https://doi.org/10.1038/s41588-023-01582-w

Rivollat, M., Rohrlach, A. B., Ringbauer, H., Childebayeva, A., Mendisco, F., Barquera, R., Szolek, A., Le Roy, M., Colleran, H., Tuke, J., & others. (2023). Extensive pedigrees reveal the social organization of a neolithic community. *Nature*, *620*(7974), 600–606. https://doi.org/10.1038/s41586-023-06350-8