Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. doi: 10.18637/jss.v000.i00

Bayesian estimation of Genetic Relatedness from Low-coverage Genotype Data using the BREAD R Package

Adam B Rohrlach
Max Planck Institute for Evolutionary Anthropology,
School of Mathematical Sciences
University of Adelaide

Jonathan Tuke
School of Mathematical Sciences
University of Adelaide

Wolfgang Haak
Max Planck Institute for Evolutionary Anthropology

Divyaratan Popli
Max Planck Institute for Evolutionary Anthropology

Abstract

Robust and reliable estimates of biological relatedness a key source of information when reconstructing pedigrees, which in combination with contextual data 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. Moreover, even current methods which attempt to reconstruct genetic relatedness from aDNA data can still require relatively high genome coverage. Critically, a statistically robust method for assessing the confidence of a classification of a specific degree of relatedness from low coverage genome data is lacking.

In this paper we present the R-package **BREAD** (Biological RElatedness from Ancient DNA), which leverages the so-called pairwise mismatch rate, calculated on optimally-thinned genome-wide pseudo-haploid sequence data, to estimate genetic relatedness up to the second degree, assuming an underlying binomial distribution. **BREAD** 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 particularly 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, and then compare the performance of **BREAD** on empirical data from Bronze Age Iberian human sequence data. The **BREAD** package is designed for pseudo-haploid genotype data, common in aDNA studies.

Keywords: biological kinship, ancient DNA.

1. Introduction

An important quality control step in ancient DNA (aDNA) studies is to use estimates of biological relatedness to see if sequence data may come from different skeletal elements or tissues from the same individual, and can hence be merged. Secondly, it is often used to exclude all but one of a group of closely-related individuals for reducing bias in downstream genetic analyses which rely on independent samples of the allele frequency distribution, and thus could lead to biased results. Recently, due to methodological improvements in sampling, DNA extraction, library preparation and capture techniques, archaeogenetics studies have also begun to explore focused regional studies, aimed at investigating kinship and social organisation in groups of burials or even entire graveyards/burial grounds (Mittnik, Massy, Knipper, Wittenborn, Friedrich, Pfrengle, Burri, Carlichi-Witjes, Deeg, Furtwängler *et al.* (2019); Schroeder, Margaryan, Szmyt, Theulot, Włodarczak, Rasmussen, Gopalakrishnan, Szczepanek, Konopka, Jensen *et al.* (2019); Cassidy, Maoldúin, Kador, Lynch, Jones, Woodman, Murphy, Ramsey, Dowd, Noonan *et al.* (2020); Žegarac, Winkelbach, Blöcher, Diekmann, Krečković Gavrilović, Porčić, Stojković, Milašinović, Schreiber, Wegmann *et al.* (2021); Villalba-Mouco, Oliart, Rihuete-Herrada, Childebayeva, Rohrlach, Fregeiro, Celdrán Beltrán, Velasco-Felipe, Aron, Himmel *et al.* (2021); Rivollat, Thomas, Ghesquière, Rohrlach, Späth, Pemonge, Haak, Chambon, and Deguilloux (2022); Villalba-Mouco, Oliart, Rihuete-Herrada, Rohrlach, Fregeiro, Childebayeva, Ringbauer, Olalde, Celdrán Beltrán, Puello-Mora *et al.* (2022)). Robust estimates of biological relatedness are a key source of information in concert with archaeological and anthropological lines of evidence (among others) for inferring kinship practices in past societies.

Diploid genomes from sequenced individuals are made up of segments of inherited DNA from parent generations, and when inspected to uncover regions that are identical by descent, can be used to infer close genetic relationships. Modern methods for investigating the degrees of genetic relatedness usually require diploid, and even phased, sequence data (Goudet, Kay, and Weir (2018)). However, due to post-mortem DNA damage and degradation, the coverage of human aDNA sequence data can be extremely low and thus impossible to phase (Orlando, Allaby, Skoglund, Der Sarkissian, Stockhammer, Ávila-Arcos, Fu, Krause, Willerslev, Stone *et al.* (2021)). Exciting new methods that employ imputation as a pre-processing step (Ringbauer, Coop, and Barton (2017)) also require relatively high coverage data (for aDNA), a situation which in, for example, whole-cemetery analyses, is extremely unlikely to be the case for all individuals.

Researchers should always use all available methods, where possible, as a best strategy. While methods for estimating the degree of relatedness for low-coverage pseudo-haploid genotype data (Lipatov, Sanjeev, Patro, and Veeramah (2015), Monroy Kuhn, Jakobsson, and Günther (2018), Popli, Peyrégne, and Peter (2023)), we identified that only KIN includes a statistically rigorous measure of classification uncertainty, and that no method includes easy-to-interpret diagnostic plots for interpretation. Here we present **BREAD**: a simple-to-use package, for the R-statistical software. We compare the performance of **BREAD** to READ (Relationship Estimation from Ancient DNA (Monroy Kuhn *et al.* (2018))), a popular, field-standard, peer-reviewed method for estimating biological relatedness up to the second-degree from low-coverage aDNA data.

While READ produces a Z-score for whether or not a pair of individuals can be "more or less" related (Z_{lower} and Z_{upper} respectively), this is calculated by subtracting the midpoint

between the the expected pairwise-mismatch rate (PMR) for the lower and upper classes of relatedness (divided by the jackknife standard deviation). This approach treats the density of the PMR for each related class as uniform, and non-overlapping, with the expected PMR lying precisely at the midpoint between these boundaries. **BREAD** instead assumes a binomial distribution for the distribution of the PMR values, and using the associated probability density functions, returns not just a hard classification (the most "likely" genetic relationship between each pair of individuals) but also the posterior probabilities of all possible genetic relationships in detailed diagnostic plots. With this complete information, researchers are able to better make decisions on pairwise relatedness, informed by rigorous statistical calculations, with measures of uncertainty when considering alternative genetic relationships.

2. The La Almoloya data set

In this paper we focus on a pseudo-haploid Eigenstrat data set (Patterson, Moorjani, Luo, Mallick, Rohland, Zhan, Genschoreck, Webster, and Reich (2012)) generated from prehistoric individuals who were buried at the Early Bronze Age site La Almoloya in today's Spain (Villalba-Mouco *et al.* (2021)). This data was generated using pileupcaller [1] with mapping and sequence quality filtering parameters -q 30 and -Q30, and the –randomHaploid flag. The data consists of 68 individuals, and included 13 pairs of first-degree individuals, 10 pairs of second-degree related individuals, and 2255 unrelated pairs of individuals. The authors identified these relationships using a suite of information including: genetic relatedness estimated by READ, lcMLkin, shared runs of homozygosity, uniparentally-inherited markers from the mitochondrial genome and the Y chromosome, as well as archaeological context information such as age at death and stratigraphy.

2.1. Preprocessing the Eigenstrat data for analysis

All downstream functions in the **BREAD** package require preprocessing of an Eigenstrat "trio": 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 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 *ind*, *geno* and *snp* files that comprise an Eigenstrat trio. 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 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

[1] https://github.com/stschiff/sequenceTools

C->T or G->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.

The resulting 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*) (see Figure 1).

```
          pair nsnps mismatch       pmr
1  Ind1 - Ind2  1518      310 0.2042161
2  Ind1 - Ind3  9435     2093 0.2218336
3  Ind1 - Ind4  8283     1854 0.2238319
4  Ind1 - Ind5  1336      314 0.2350299
5  Ind1 - Ind6  2242      481 0.2145406
6  Ind2 - Ind3  9119     1988 0.2180064
```

Figure 1: The first 6 rows of the tibble produced by running processEigenstrat() on the example Eigenstrat files.

2.2. 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 *median* parameter defines what the background relatedness should be (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 less overlapping SNPs than this value.

The resulting tibble 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*), the normalised posterior probabilities for each of the four relatedness categories (*Same_Twins, First_Degree, Second_Degree* and *Unrelated*), and a strength of evidence statement for the Bayes Factor for the comparison of the genetic relationship which had the two highest-posterior probability values (*BF*) (see Figure 2).

	row	pair	relat…[1]	pmr	sd	misma…[2]	nsnps	ave_rel	Same_Tw…[3]	First_D…[4]
	<int>	<chr>	<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	Ind1 –…	Unrela…	0.204	0.0103	310	1518	0.218	6.71e- 26	2.00e- 4
2	2	Ind1 –…	Unrela…	0.222	0.00428	2093	9435	0.218	1.22e-214	2.41e- 47
3	3	Ind1 –…	Unrela…	0.224	0.00458	1854	8283	0.218	2.00e-194	3.44e- 44
4	4	Ind1 –…	Unrela…	0.235	0.0116	314	1336	0.218	2.68e- 37	5.07e- 10
5	5	Ind1 –…	Unrela…	0.215	0.00867	481	2242	0.218	9.82e- 46	2.62e- 9
6	6	Ind2 –…	Unrela…	0.218	0.00432	1988	9119	0.218	5.06e-195	1.97e- 40

Figure 2: The first 6 rows of the tibble produced by running callRelatedness() on the example PMR data.

2.3. 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 callLOAF() function which plots the first N (by default $N = 50$) pairs of individuals, sorted by PMR (ascending). 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 05% 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 $\frac{n}{2(n-2)}$ 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 brevity.

Second, a diagnostic plot for the analysis of a single pair of individuals can be obtained using the callSLICE() function. By default, this function produces a tow panel plot. The left panel 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. The right panel 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.

Finally, the savesSLICES() function allows the user to save all possible pairwise diagnostic plots (as produced by plotSLICE()) as PDFs to an output folder.

2.4. Testing for departure from refined degrees of relatedness

The resolution of the BREAD 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.

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 3). 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.

```
Testing H0       : "Ind1 - Ind2" are 3rd-degree relatives.
Expected PMR     : 0.2044
Observed PMR     : 0.2042
Estimated degree : 2.9826
p-value          : 0.9823
Decision         : Retain H0
[1] 0.9823165
```

Figure 3: The output from the binomial test for whether the PMR observed for Ind1 and Ind2 is consistent with a third-degree relationship.

3. The statistical model

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 linkage disequilibrium are reduced. We then look at the number of pseudo-haploid genotype calls that *do not* match, denoted X_{ij}. Hence we have that $X_{ij} \sim B(N_{ij}, p_{ij})$, 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. Many choices exist for this value, but assuming that a sample is made up of mostly unrelated pairs (to the second degree), then the median PMR, denoted $\bar{p}$ will be a reliable estimate. 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

$$p_k = \bar{p}\left(1 - \frac{1}{2^{k+1}}\right). \tag{1}$$

Hence, if we assume that the degree of relatedness for individuals i and j is truly of the k^{th} degree, then

$$X_{ij} \sim B(N_{ij}, p_k).$$

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

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

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

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

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 k-degree related as

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

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

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

It is not clear what the prior probabilities for the degrees of relatedness, from same/twin to second-degree related, should be, and so these are given as uniform by default. However, since these are user inputs, these can be explored via sensitivity analyses to test if the prior probabilities are driving the relatedness classifications.

Once posterior probabilities have been calculated, we then find the Bayes factor for individuals i and j for the two degrees of relatedness, denoted $K = s, t$, which had the first and second highest posterior probabilities, *i.e.*

$$B_{ij} = \frac{P(K = s|X_{ij}, N_{ij})}{P(K = t|X_{ij}, N_{ij})}.$$

We then report the strength of evidence, an interpretation of the Bayes factor, for the hard classification of the degree of relatedness between individuals i and j, compared to the next most likely classification. We use the interpretations as defined by Jeffreys (Jeffreys (1998)) (see Table 3).

$\mathbf{B_{ij}}$	**Strength of Evidence**
10^0 to $10^{1/2}$	Weak Evidence
$10^{1/2}$ to 10^1	Substantial
10^1 to $10^{3/2}$	Strong
$10^{3/2}$ to 10^2	Very Strong
$> 10^2$	Decisive

Finally, while we do not include third-degree relationships in the possible classifications, we do include a method for statistical test for retaining or rejecting the possibility of a k^{th}-degree relationship, for $0 \leq k \leq 10$. We simply perform a binomial test for the observed number

of mismatches X_{ij}, and the expected probability of 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 1, and solving for k, *i.e.*,

$$\hat{k} = \begin{cases} \frac{ln(1-p_{ij}/\bar{p})}{ln(1/2)} - 1, & p_{ij} < \bar{p}, \\ \infty, & p_{ij} \geq \bar{p}, \end{cases}$$

where ∞ indicates "unrelated".

4. Results

To test the performance of **BREAD**, we performed analyses on simulated data as described in Popli *et al.* (2023), and on published empirical data from a population genetic and relatedness study (Villalba-Mouco *et al.* (2021)), and compared our results to those obtained from READ.

4.1. Simulated Data

We simulated genotype data from a pedigree of individuals as shown in Figure 4. 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.

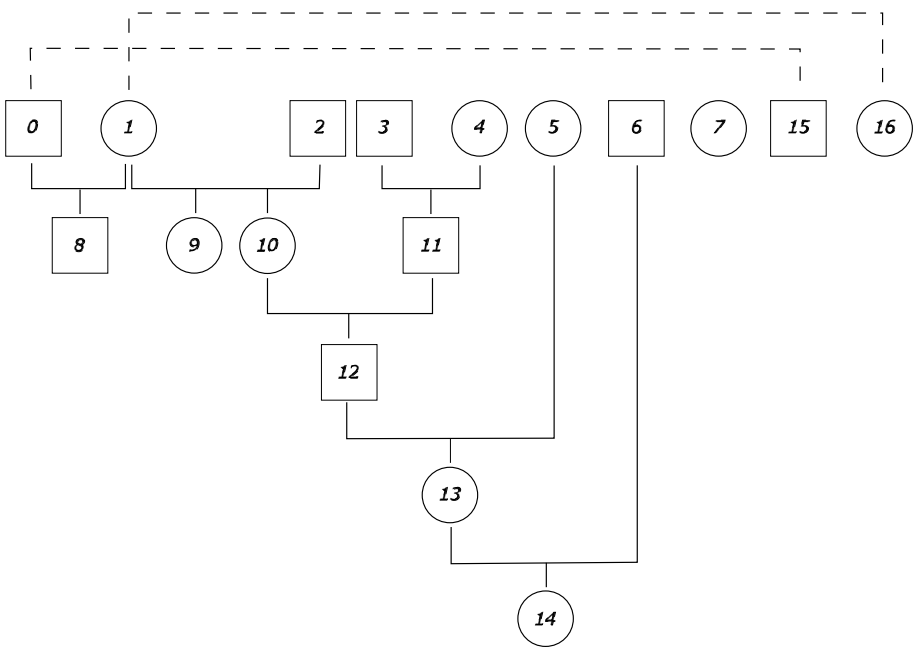

Figure 4: Pedigree for the simulated data. Individuals 0 and 8, and 1 and 9, are genetically identical, respectively. Females are represented by circles, males are represented by squares.

We ran the algorithm with the default thinning parameter of 1×10^5, minimum number of overlapping sites of 500, a uniform prior for the class classifications, and the median used to

estimate the background relatedness. Overall, we correctly assigned 129/136 (94.85%) of the pairwise genetic relationships, indicating a relatively high degree of accuracy.

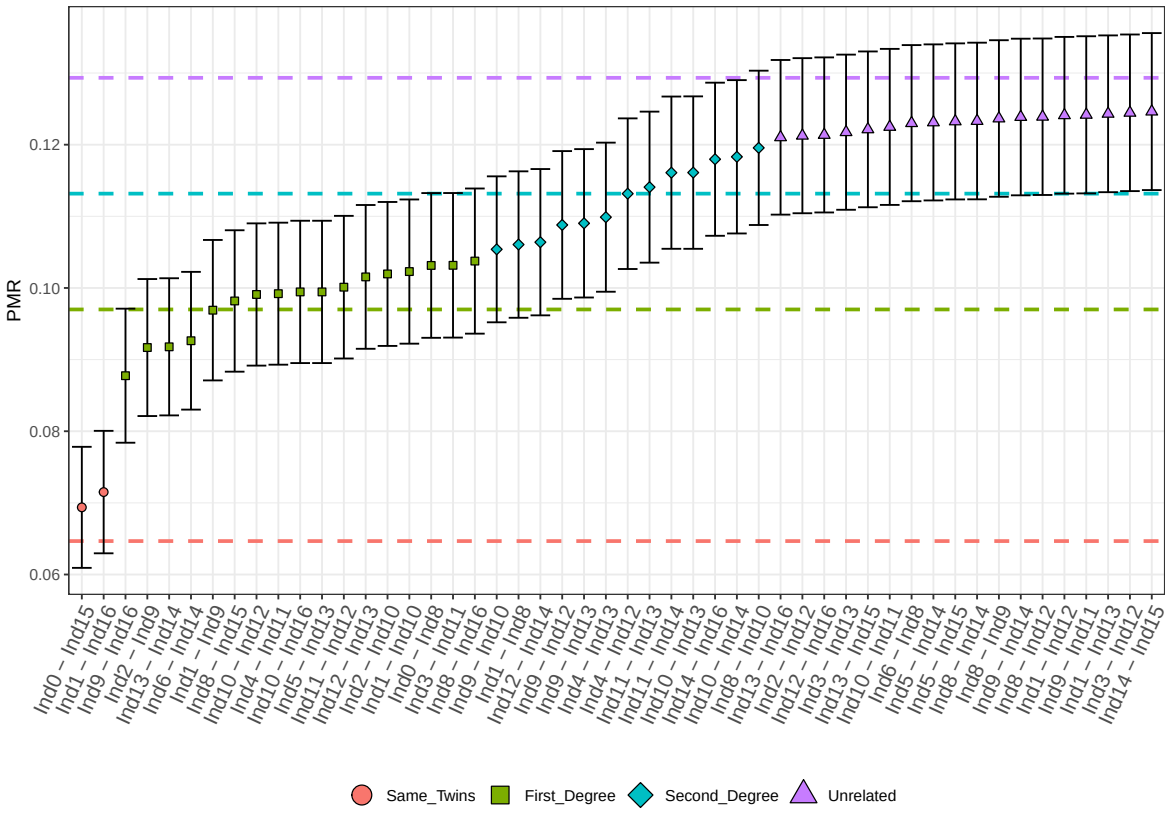

Figure 5: The first 50 ordered pairwise-mismatch rates (PMRs) from the simulated data. Colours indicate classification, dashed lines indicate the expected PMR for each degree of relatedness, and the error bars indicate 2 standard errors.

BREAD correctly classified 2/2 of the genetically identical pairs with high posterior probability (one to machine precision), and 19/19 of the first-degree related pairs (minimum posterior probability 0.993). BREAD correctly identified 10/12 of the second degree pairs, and in the cases where we misclassified the relationships (both times as unrelated) READ also misclassified the second-degree pairs as unrelated. Furthermore, READ misclassified an additional two first-degree related pairs as unrelated.

Finally, BREAD correctly classified 98/103 of the unrelated and third-degree or higher related individuals as "unrelated". Interestingly, the individuals in the remaining five cases were actually related in the third-degree in the pedigree. This was reflected in the fact that, even though there was strong evidence for classifying these individuals as second-degree related compared to unrelated, the binomial tests for third-degree were all non-significant. We note that READ did not misclassify these five relationships, and returned an absolute Z-score less than 2 in 4/5 cases, indicating that a closer relationship than unrelated was also possible.

4.2. Empirical data

We next tested our method on the empirical data from the Early Bronze Age site of La Almoloya, in Spain (Villalba-Mouco *et al.* (2021)) using the same parameters as for the simulated data. When we compared our results to READ, we found that we only disagreed on 2/2278 pairs of individuals: ALM016/ALM017 and ALM057/ALM079. In both cases, READ classified these relationships as unrelated, whereas our method classified them as second-degree. However, we found that the posterior probability for unrelated was 0.21 for ALM057/ALM079 (see Figure 6), and a binomial test indicated that a third-degree relationship could not be rejected (p=0.266). However, these two individuals had no common first- or second-degree relatives in the study, and so a third-dgree relationship between could not definitively be shown using additional relatives. For ALM016/ALM017, the observed PMR of 0.2356 was in between what would be expected for second-degree (0.2271) and third-degree (0.2433), but was shown to be third-degree using additional contextual information in the published study.

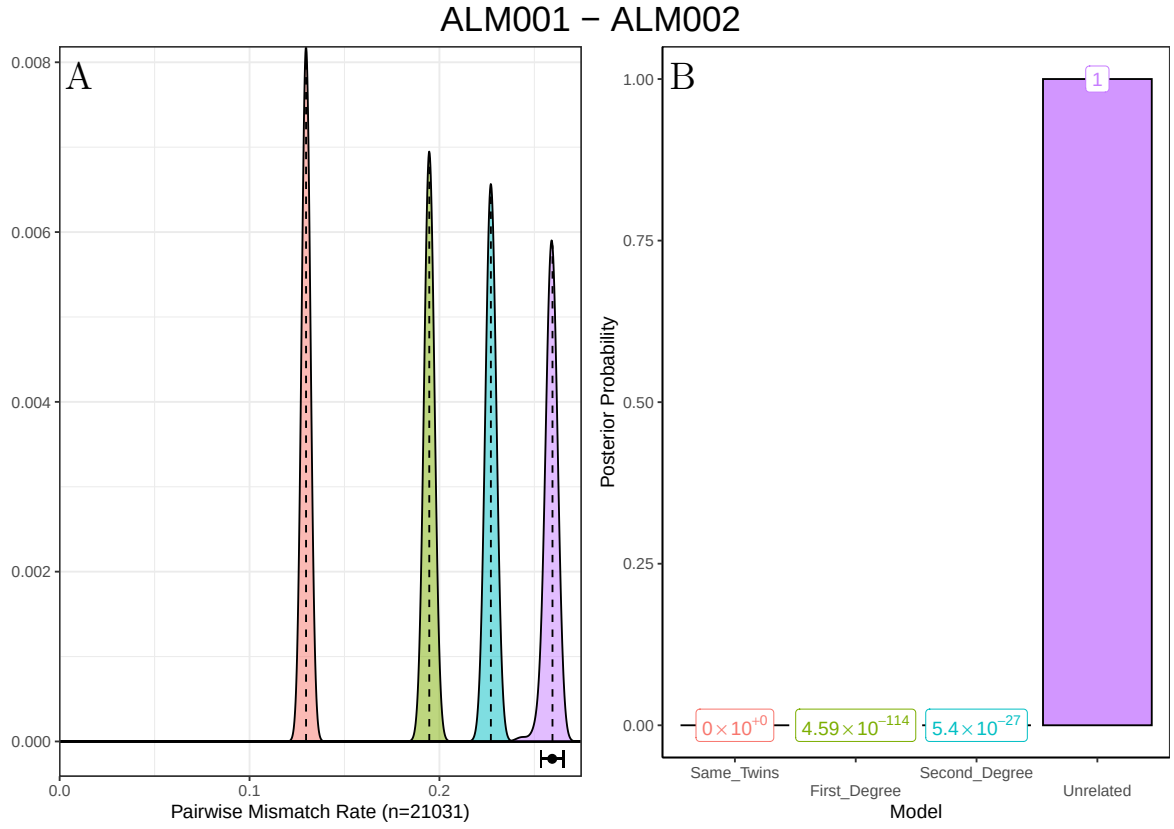

Figure 6: A diagnostic plot for the classification of ALM057 and ALM079, with colours indicating classes. Panel A shows the distributions of the pairwise-mismatch rates (PMR) for the same/twins, first-, second-degree, and unrelated classes, with the observed PMR, with 2 standard errors, indicated below. Dashed lines indicated the mode for each class. Panel B shows the normalised posterior probabilities, with values, for all classes.

Figure 6 shows the additional diagnostic information available using the **BREAD** package. ALM057 and ALM079 were identified as unrelated by READ, and second-degree by **BREAD**. However, the it is also clear from Panel A that the PMR falls somewhere between

the distributions for second-degree related and unrelated pairs of individuals.

5. Summary and Conclusions

This article describes the usage and functionality of the R package **BREAD**. This package is used to infer pairwise genetic relatedness between individuals, using the well-understood PMR, up to the second-degree. **BREAD** 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, **BREAD** 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 have shown that **BREAD** performs equally well on both simulated and empirical data, when compared to the peer-reviewed, field-standard software READ. We have also shown that when **BREAD** misclassifies genetic relationships, it is due to expected statistical variation (as READ also misclassified the genetic relationship), or it was due to a true closer genetic relationship (third-degree genetic relationships were misclassified as unrelated in each case). The easy-to-use R-implementation of our method make **BREAD** an attractive, and statistically rigorous analysis for researchers in archaeogenetics to employ.

6. Bibliography

References

Cassidy LM, Maoldúin RÓ, Kador T, Lynch A, Jones C, Woodman PC, Murphy E, Ramsey G, Dowd M, Noonan A, *et al.* (2020). "A dynastic elite in monumental Neolithic society." *Nature*, **582**(7812), 384–388.

Goudet J, Kay T, Weir BS (2018). "How to estimate kinship." *Molecular ecology*, **27**(20), 4121–4135.

Jeffreys H (1998). *The theory of probability*. OUP Oxford.

Lipatov M, Sanjeev K, Patro R, Veeramah KR (2015). "Maximum likelihood estimation of biological relatedness from low coverage sequencing data." *BioRxiv*, p. 023374.

Mittnik A, Massy K, Knipper C, Wittenborn F, Friedrich R, Pfrengle S, Burri M, Carlichi-Witjes N, Deeg H, Furtwängler A, *et al.* (2019). "Kinship-based social inequality in Bronze Age Europe." *Science*, **366**(6466), 731–734.

Monroy Kuhn JM, Jakobsson M, Günther T (2018). "Estimating genetic kin relationships in prehistoric populations." *PloS one*, **13**(4), e0195491.

Orlando L, Allaby R, Skoglund P, Der Sarkissian C, Stockhammer PW, Ávila-Arcos MC, Fu Q, Krause J, Willerslev E, Stone AC, *et al.* (2021). "Ancient DNA analysis." *Nature Reviews Methods Primers*, **1**(1), 1–26.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D (2012). "Ancient admixture in human history." *Genetics*, **192**(3), 1065–1093.

Popli D, Peyrégne S, Peter BM (2023). "KIN: a method to infer relatedness from low-coverage ancient DNA." *Genome Biology*, **24**(1), 1–22.

Ringbauer H, Coop G, Barton NH (2017). "Inferring recent demography from isolation by distance of long shared sequence blocks." *Genetics*, **205**(3), 1335–1351.

Rivollat M, Thomas A, Ghesquière E, Rohrlach AB, Späth E, Pemonge MH, Haak W, Chambon P, Deguilloux MF (2022). "Ancient DNA gives new insights into a Norman Neolithic monumental cemetery dedicated to male elites." *Proceedings of the National Academy of Sciences*, **119**(18), e2120786119.

Schroeder H, Margaryan A, Szmyt M, Theulot B, Włodarczak P, Rasmussen S, Gopalakrishnan S, Szczepanek A, Konopka T, Jensen TZ, *et al.* (2019). "Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave." *Proceedings of the National Academy of Sciences*, **116**(22), 10705–10710.

Villalba-Mouco V, Oliart C, Rihuete-Herrada C, Childebayeva A, Rohrlach AB, Fregeiro MI, Celdrán Beltrán E, Velasco-Felipe C, Aron F, Himmel M, *et al.* (2021). "Genomic transformation and social organization during the Copper Age–Bronze Age transition in southern Iberia." *Science advances*, **7**(47), eabi7038.

Villalba-Mouco V, Oliart C, Rihuete-Herrada C, Rohrlach AB, Fregeiro MI, Childebayeva A, Ringbauer H, Olalde I, Celdrán Beltrán E, Puello-Mora C, *et al.* (2022). "Kinship practices in the early state El Argar society from Bronze Age Iberia." *Scientific Reports*, **12**(1), 1–15.

Žegarac A, Winkelbach L, Blöcher J, Diekmann Y, Krečković Gavrilović M, Porčić M, Stojković B, Milašinović L, Schreiber M, Wegmann D, *et al.* (2021). "Ancient genomes provide insights into family structure and the heredity of social status in the early Bronze Age of southeastern Europe." *Scientific reports*, **11**(1), 1–11.

Affiliation:

Adam B Rohrlach
Department of Archaeogenetics
Max Planck Institute for Evolutionary Anthropology
Leipzig 04299, Germany
E-mail: adam_ben_rohrlach@eva.mpg.de

Journal of Statistical Software http://www.jstatsoft.org/
published by the Foundation for Open Access Statistics http://www.foastat.org/
MMMMMM YYYY, Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:10.18637/jss.v000.i00 *Accepted:* yyyy-mm-dd