Supplementary Materials:

BADGER: evaluating the performance of ancient DNA genetic relatedness estimation methods using high-fidelity pedigree simulations.

Contents

1 Supplementary Figures

1.1 Figure S1: Alternate template pedigree including inbred individuals

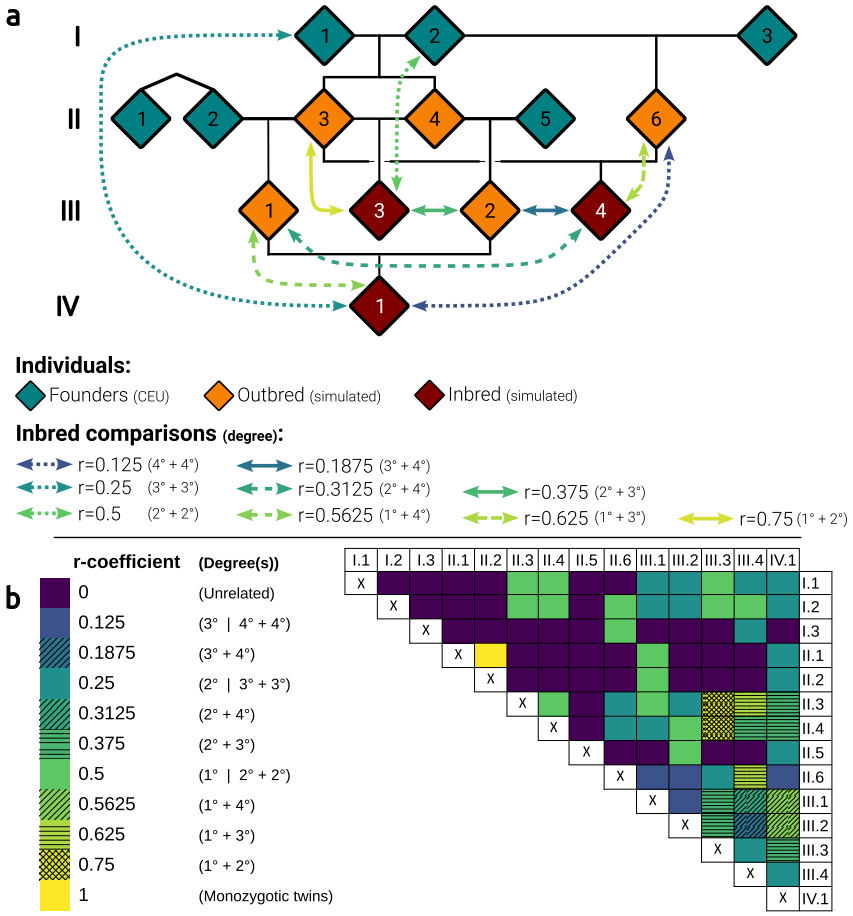

Fig. S1: Alternate template pedigree including three inbred individuals. **a**: Diagram of the alternative template pedigree, provided to BADGER when evaluating the impact of close inbreeding. Coloured arrows denote the pairwise relationships investigated during the inbreeding benchmark. Additional inbred individuals are coloured in dark red, with individuals III.3, III.4 and IV.1 being the result of a mating respectively involving two siblings, two half-siblings, and two first-cousins. **b**: Pairwise matrix of the relationships defined by the simulated pedigree shown in (a). A complete description of the pairwise relationships contained within the input pedigrees used throughout this study is described in Supplementary Table S2

1.2 Figure S2: Alternative input pedigrees used to evaluate the impact of inbreeding on the normalisation procedure of correctKin, KIN and READv2.

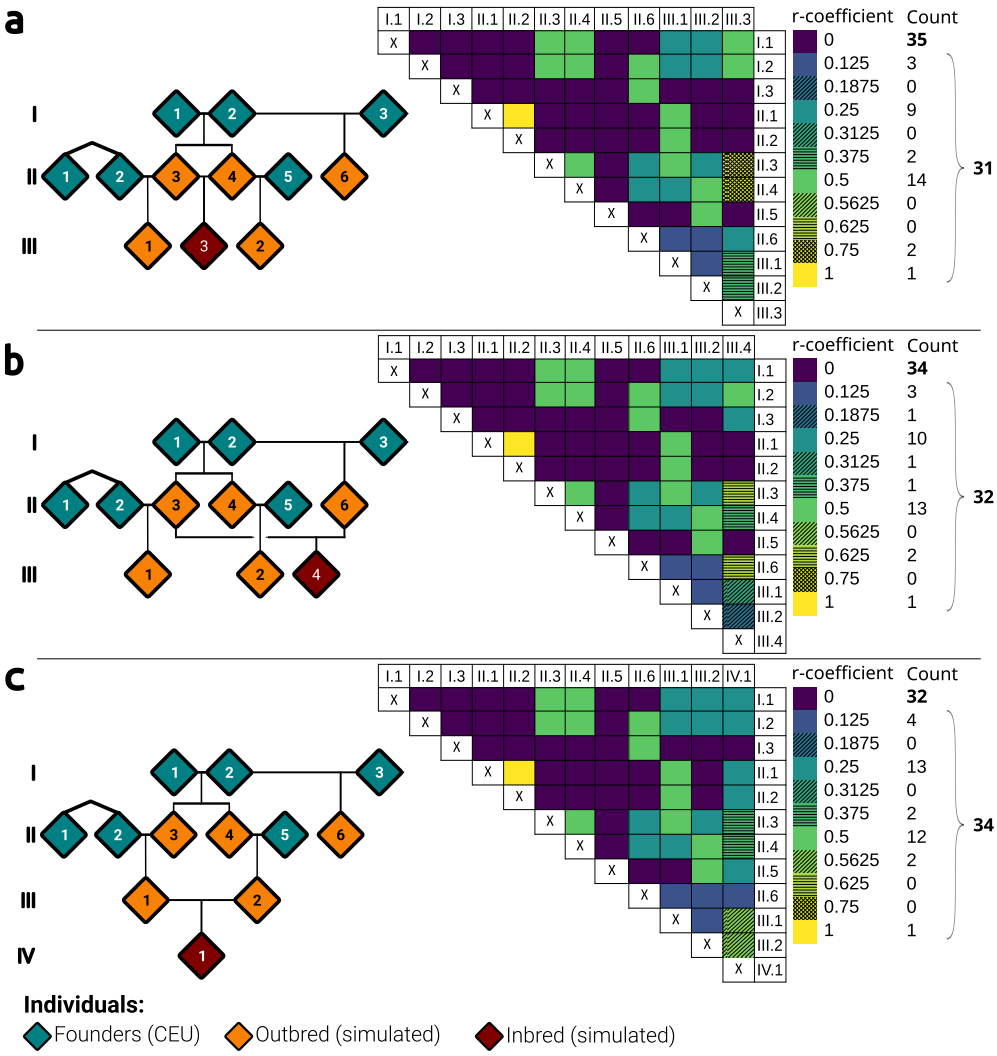

Fig. S2: Diagram and pairwise relationship matrix of three alternative pedigree topologies, given as input to correctKin, KIN and READv2 methods, where only one out of the three inbred individuals (III.3, III.4 and IV.1) is included in the tested cohort. **a**: Full-siblings scenario. **b**: Half-siblings scenario. **c**. First-cousins scenario.

1.3 Figure S3: Comparing the performance of READv2 against its predecessor, READ

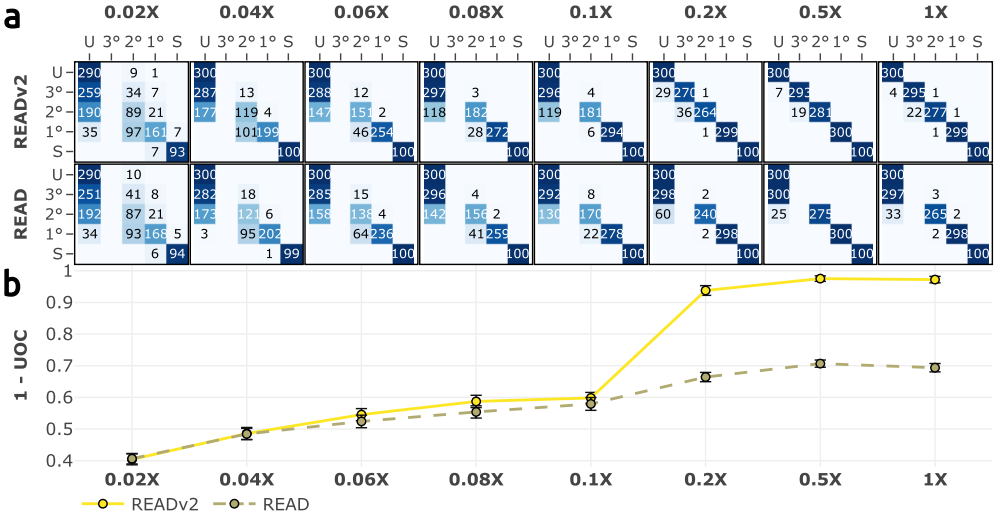

Fig. S3: Benchmark results across increasing values of sequencing depth for the READ (dashed tan line) method and its updated version, READv2 (solid yellow line). **a**: Confusion matrices of the READ and READv2 methods, confronting expected and predicted relationships. Expected and predicted values are displayed in rows and columns, respectively. 1°, 2°, 3° correspond to first-, second-, and third-degree relationships, respectively, U corresponds to "unrelated individuals", and S to "self" (monozygotic twins). **b**: UOC values summarizing the classification performance of each method for the considered sequencing depths. Higher values of $1 - UOC$ indicate higher performance.

1.4 Figure S4: Accuracy and bias of r-coefficients as a function of sequencing depth (pmd-mask).

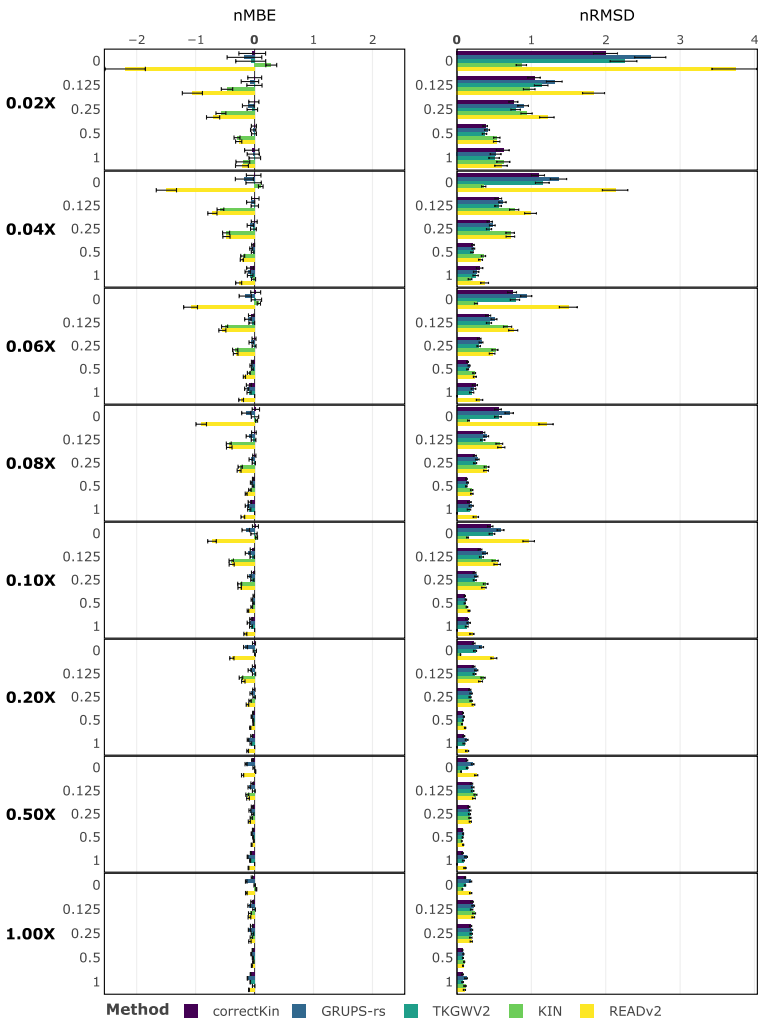

Fig. S4: Normalized estimates of MBE (left column) and $RMSD$ (right column), across all evaluated methods (bar colours), sequencing depths (rows), and expected relatedness coefficients (y-axis ticks), using sample alignment files processed through `pmd-mask`. Increasing values of $nRMSD$ indicate lower accuracy when estimating relatedness coefficients. $nMBE$ values that deviate furthest from zero indicate higher bias, with positive and negative values highlighting a tendency towards over- or under-estimating r-coefficients, respectively. Error bars represent $CI_{95\%}$ for the given estimate.

1.5 Figure S5: Accuracy and bias of r-coefficients as a function of sequencing depth (mapDamage2)

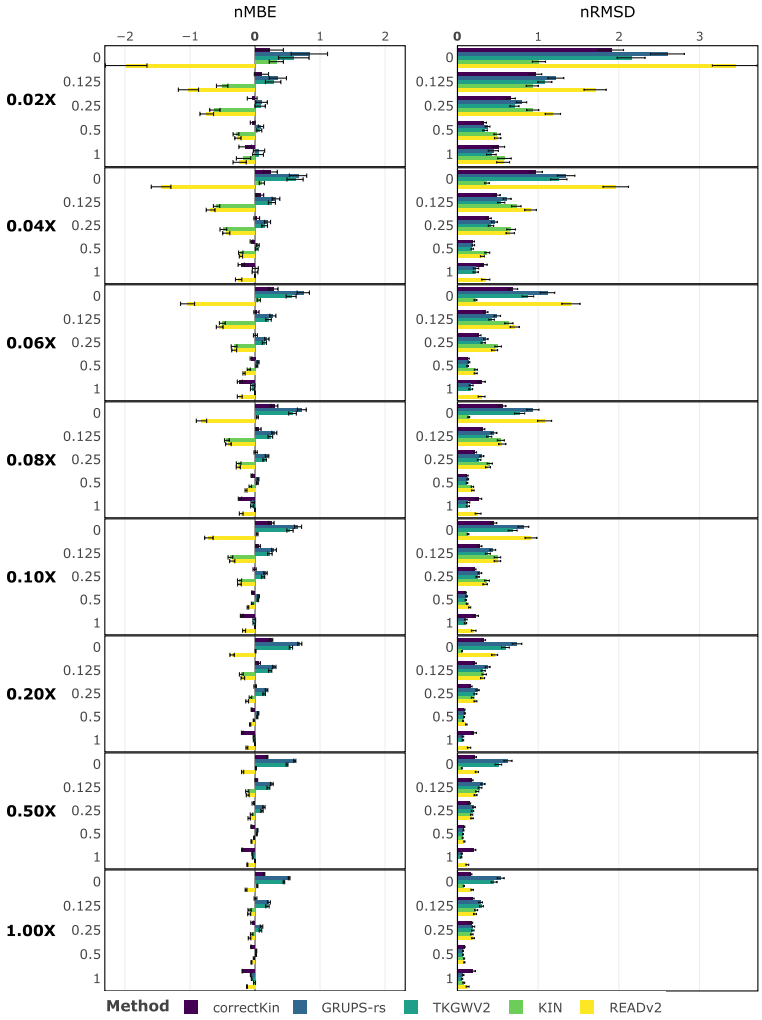

Fig. S5: $nMBE$ (left column) and $nRMSD$ (right column) estimates, across all evaluated methods (bar colours), sequencing depths (rows), and expected relatedness coefficients (y-axis ticks), using sample alignment files processed through `mapDamage2` . Increasing values of $nRMSD$ indicate lower accuracy when estimating relatedness coefficients. $nMBE$ values that deviate furthest from zero indicate higher bias, with positive and negative values highlighting a tendency towards over- or under-estimating r-coefficients, respectively. Error bars represent $CI_{95\%}$ for the given estimate.

1.6 Figure S6: Full grid of confusion matrices and UOC values across increasing values of sequencing depth (mapDamage2)

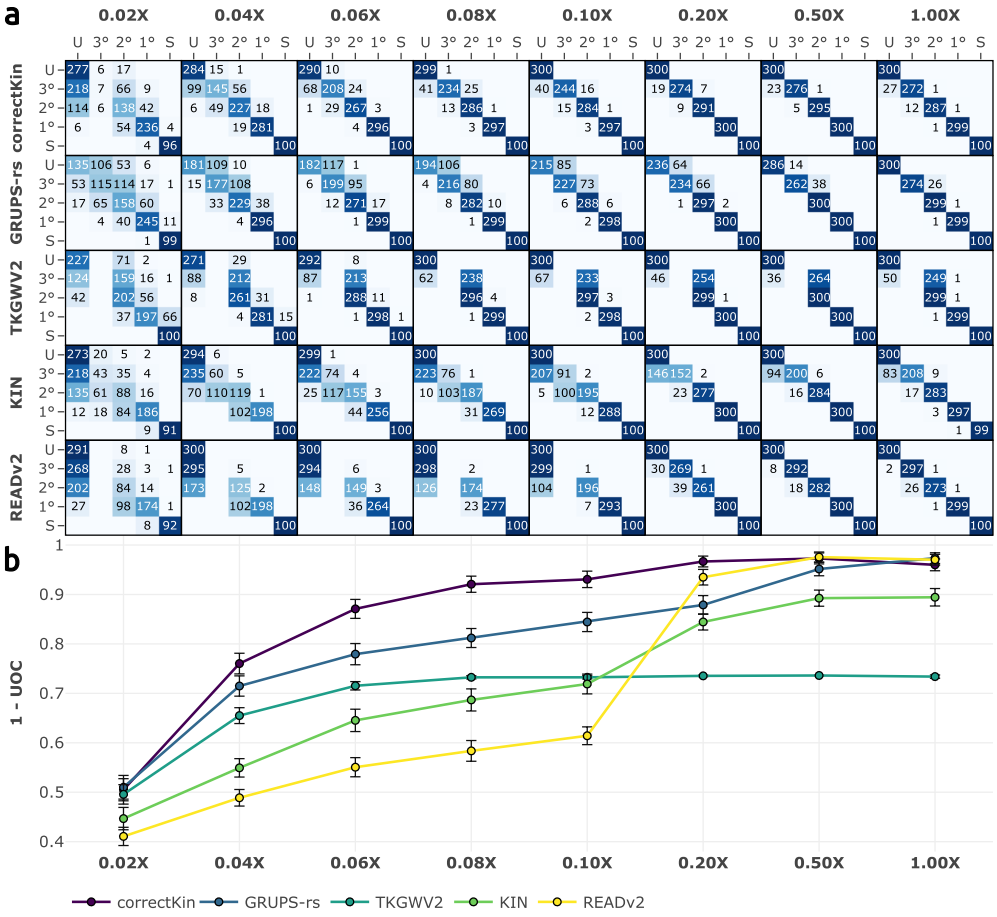

Fig. S6: Benchmark results across increasing values of sequencing depth, using sample alignment files processed through mapDamage2 *post-mortem* damage rescaling software. **a**: Confusion matrices of the five tested methods confronting expected and predicted relationships. Expected and predicted values are displayed in rows and columns, respectively. 1°, 2°, 3° correspond to first-, second-, and third-degree relationships, respectively, U corresponds to "unrelated individuals", and S to "self" (monozygotic twins). **b**: UOC values summarizing the classification performance of each method for the considered sequencing depths. Higher values of $1 - UOC$ indicate higher performance.

1.7 Figure S7: Accuracy and bias of r-coefficients as a function of contamination rate (AFR)

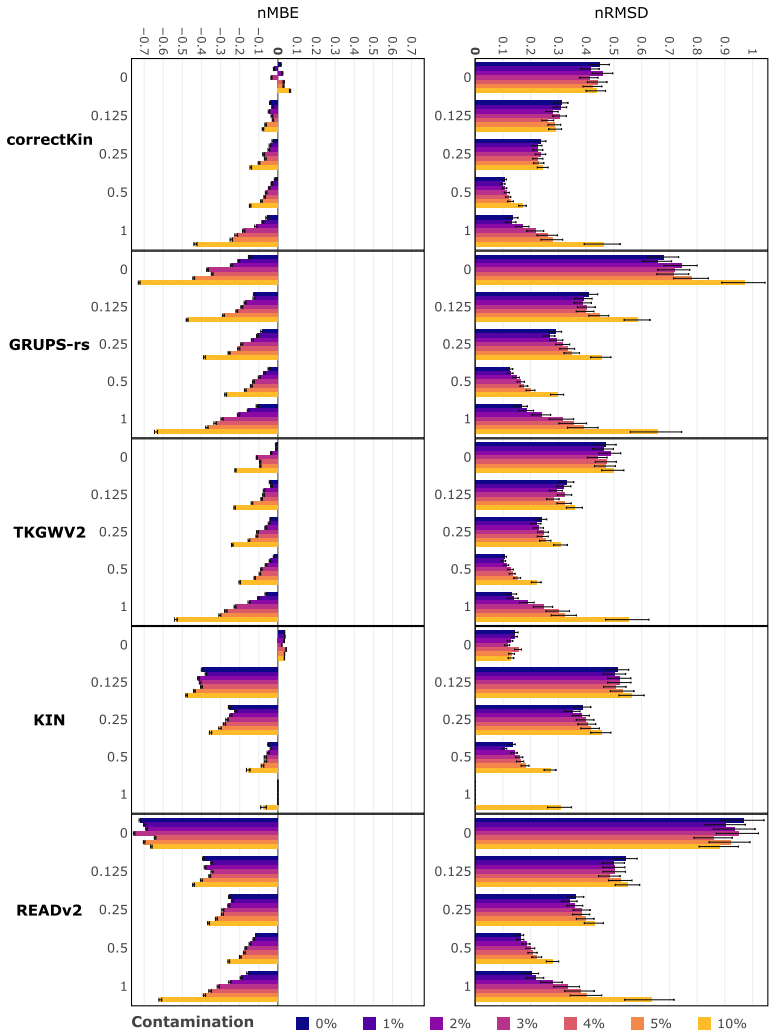

Fig. S7: $nMBE$ (left column) and $nRMSD$ (right column) estimates, across all evaluated methods (rows), rates of contamination from an AFR individual (bar colours [1], and expected relatedness coefficients (y-axis ticks). Increasing values of $nRMSD$ indicate lower accuracy when estimating relatedness coefficients. $nMBE$ values that deviate furthest from zero indicate higher bias, with positive and negative values highlighting a tendency towards over- or under-estimating r-coefficients, respectively. Error bars represent $CI_{95\%}$ for the given estimate.

1.8 Figure S8: Accuracy and bias of r-coefficients as a function of contamination (GBR)

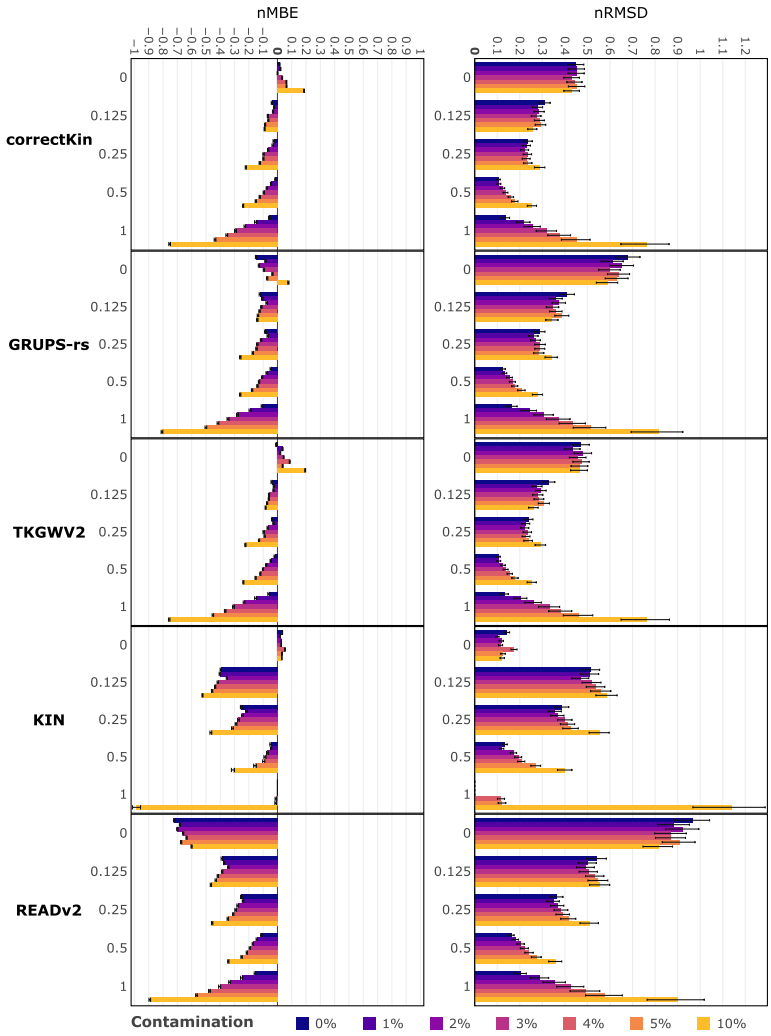

Fig. S8: $nMBE$ (left column) and $nRMSD$ (right column) estimates, across all evaluated methods (rows), rates of contamination from a GBR individual (bar colours)[1], and expected relatedness coefficients (y-axis ticks). Increasing values of $nRMSD$ indicate lower accuracy when estimating relatedness coefficients. $nMBE$ values that deviate furthest from zero indicate higher bias, with positive and negative values highlighting a tendency towards over- or under-estimating r-coefficients, respectively. Error bars represent $CI_{95\%}$ for the given estimate.

1.9 Figure S9: Full grid of confusion matrices and UOC values across increasing values of contamination (AFR)

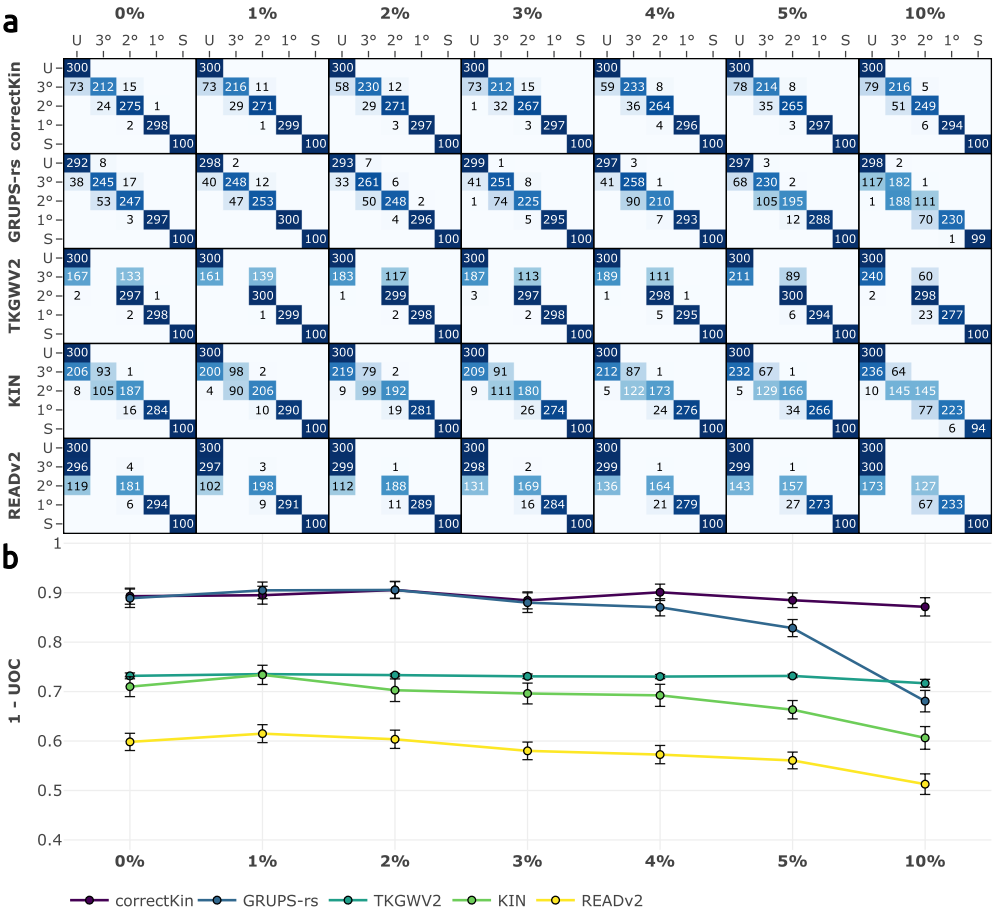

Fig. S9: Benchmark results across increasing values of modern human contamination, using the *AFR* population [1] as a source of contaminating individuals. **a**: Confusion matrices of the five tested methods, confronting expected and predicted relationships. Expected and predicted values are displayed in rows and columns, respectively. 1°, 2°, 3° correspond to first-, second-, and third-degree relationships, respectively, U corresponds to "unrelated individuals", and S to "self" (monozygotic twins). **b**: UOC values summarizing the classification performance of each method for the considered contamination rate. Higher values of $1 - UOC$ indicate higher performance.

1.10 Figure S10: Full grid of confusion matrices and UOC values across increasing values of contamination (GBR)

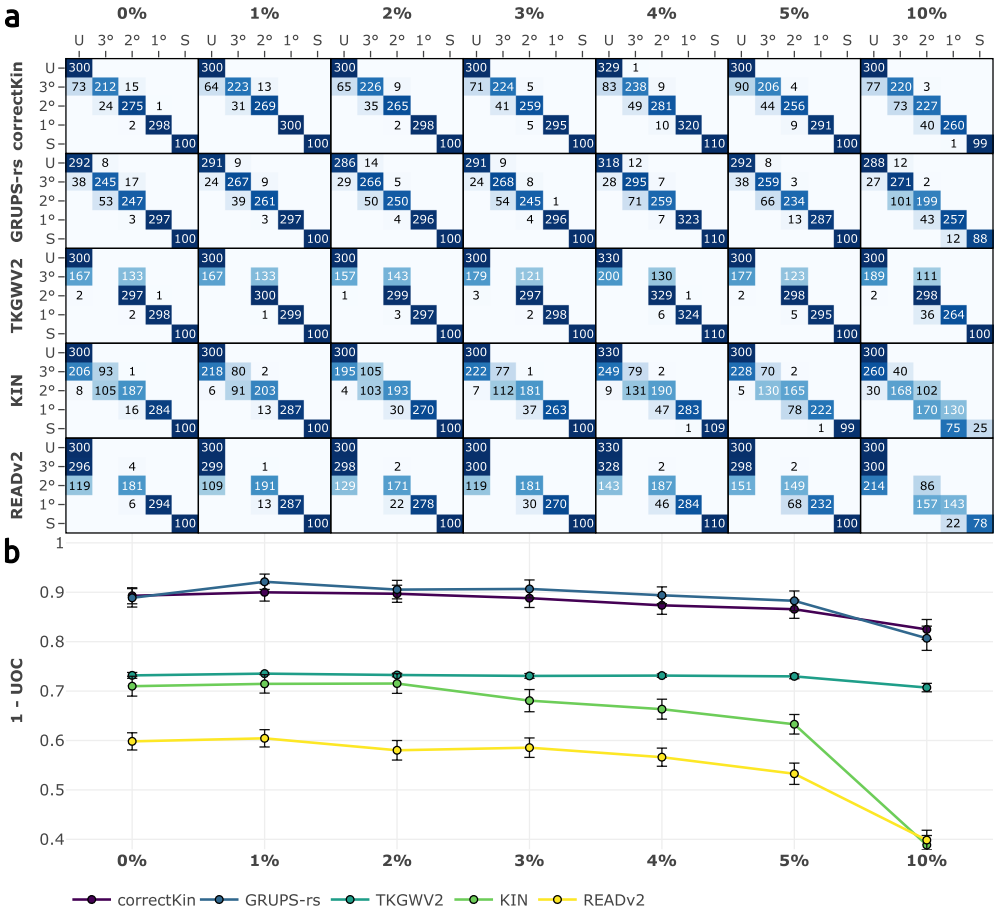

Fig. S10: Benchmark results across increasing values of modern human contamination, using the *GBR* population [1] as a source of contaminating individuals. **a**: Confusion matrices of the five tested methods, confronting expected and predicted relationships. Expected and predicted values are displayed in rows and columns, respectively. 1°, 2°, 3° correspond to first-, second-, and third-degree relationships, respectively, U corresponds to "unrelated individuals", and S to "self" (monozygotic twins). **b**: UOC values summarizing the classification performance of each method for the considered contamination rate. Higher values of $1 - UOC$ indicate higher performance.

1.11 Figure S11: Impact of admixture on the accuracy and bias of r-coefficients

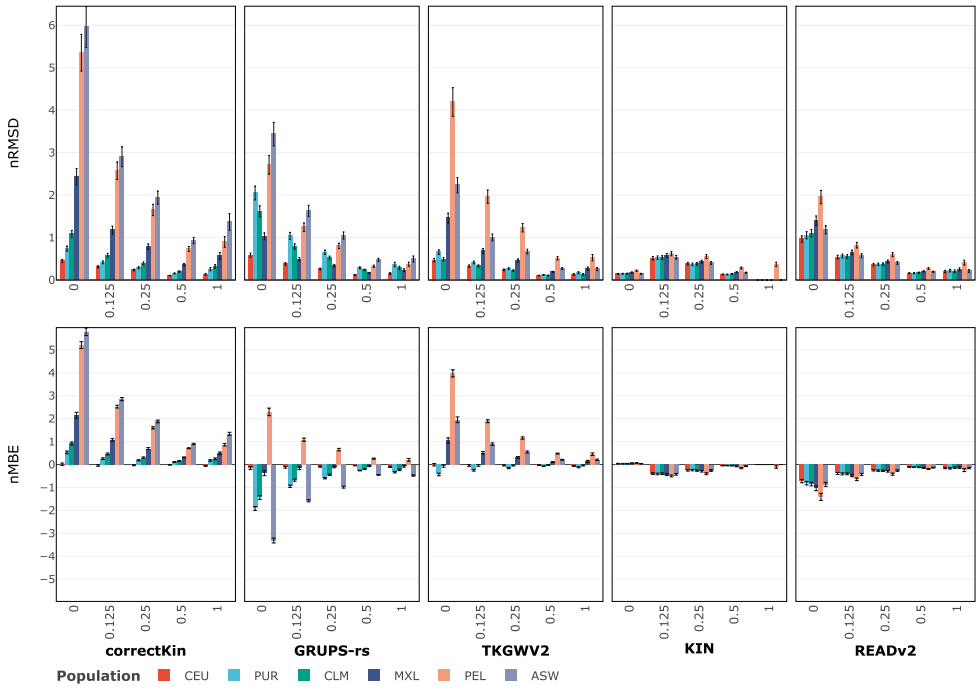

Fig. S11: $nRMSD$ (top row) and $nMBE$ (bottom row) estimates, across all evaluated methods (columns), source populations (bar colours), and expected relatedness coefficients (x-axis ticks), at a simulated sequencing depth of 0.1X. Increasing values of $nRMSD$ indicate lower accuracy when estimating relatedness coefficients. $nMBE$ values that deviate furthest from zero indicate higher bias, with positive and negative values highlighting a tendency towards over- or under-estimating r-coefficients, respectively. Error bars represent $CI_{95\%}$ for the given estimate.

1.12 Figure S12: Confusion matrices and UOC values across increasing values of sequencing depth (ASW)

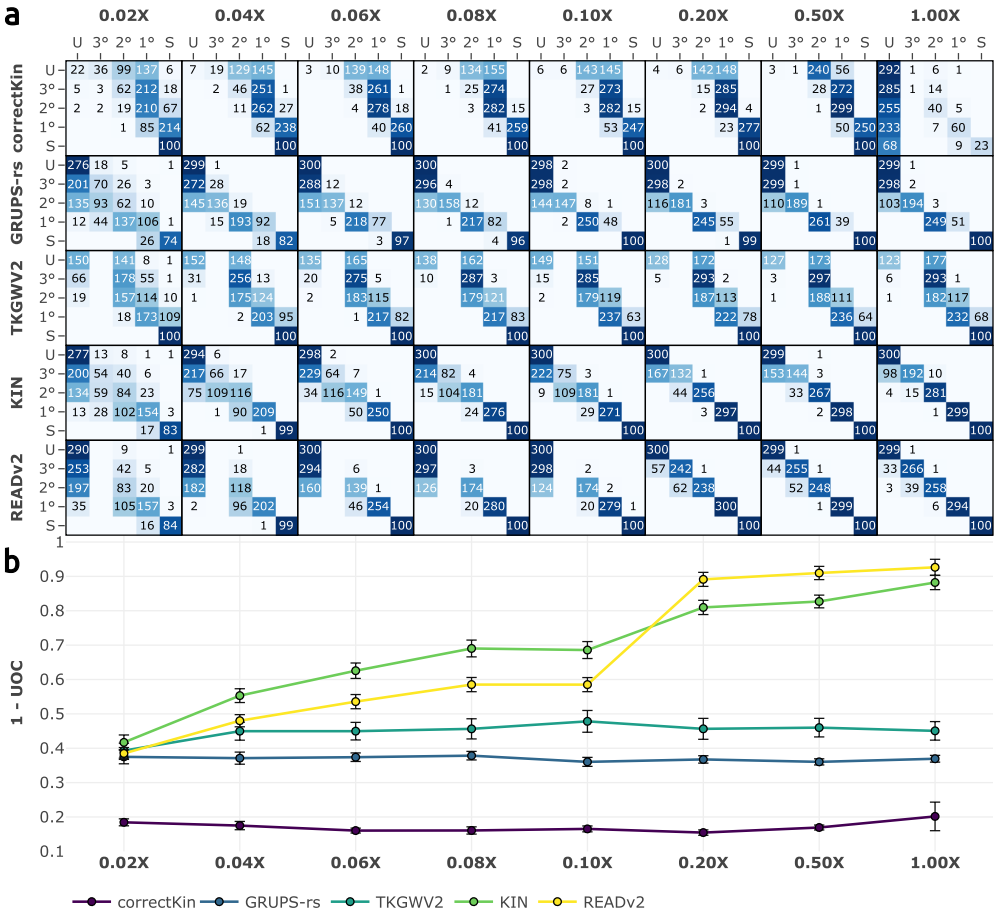

Fig. S12: Benchmark results across increasing values of sequencing depth, using admixed *ASW* individuals as a source population for pedigree individuals [1]. **a**: Confusion matrices of the five tested methods confronting expected and predicted relationships. Expected and predicted values are displayed in rows and columns, respectively. 1°, 2°, 3° correspond to first-, second-, and third-degree relationships, respectively, *U* corresponds to "unrelated individuals", and *S* to "self" (monozygotic twins). **b**: UOC values summarizing the classification performance of each method for the considered sequencing depths. Higher values of $1 - UOC$ indicate higher performance.

1.13 Figure S13: Ancestry proportions of admixed American populations.

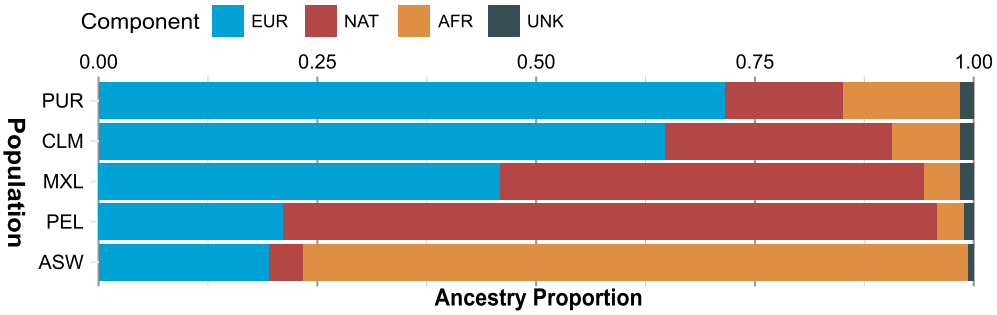

Fig. S13: Ancestry proportions of admixed American populations used during this study. The proportions of this plot were generated using the local ancestry inference results of (Martin et al. 2017) [2] (https://personal.broadinstitute.org/armartin/tgp_admixture). **AFR**: African; **EUR**: European; **NAT**: Native American; **UNK**: Unknown

1.14 Figure S14: Average heterozygosity rate of the European CEU population, and admixed American populations.

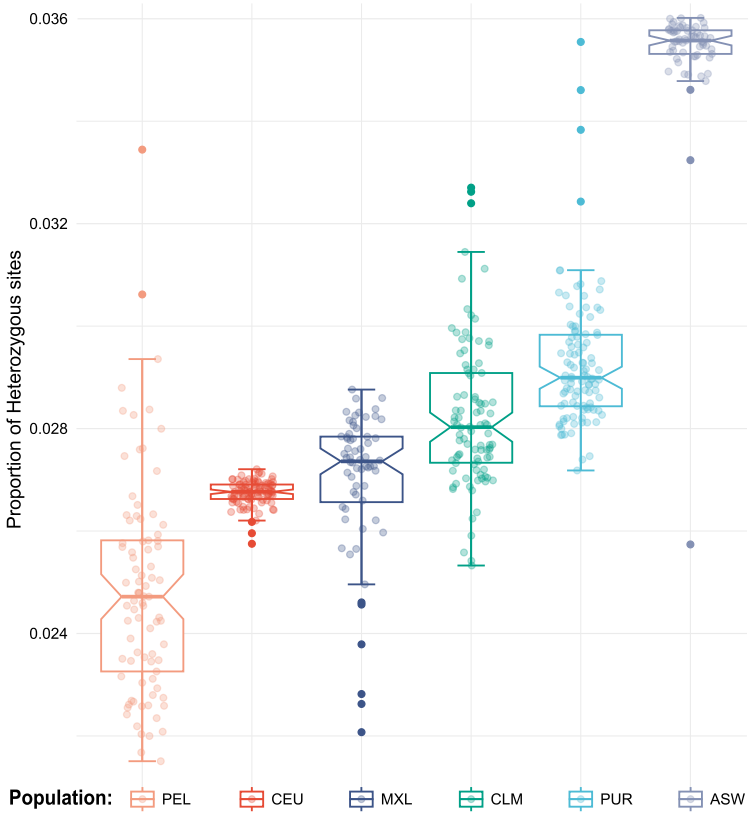

Fig. S14: Sample-wise distribution of the proportions of heterozygous sites of every 1000g-phase3 sample, according to their assigned population. Per-sample counts and proportions of heterozygous sites were directly calculated from the base dataset of the 1000g-phase3 project, using `bcftools stats` [3]. Note that these counts only take SNPs into account. Boxplot notches represent the 95% confidence interval of the median. Whiskers of each boxplot extend from the maximum to the minimum value found within the range $[Q_1 - 1.5 \cdot IQR; Q_3 + 1.5 \cdot IQR]$

1.15 Figure S15: Accuracy and bias of r-coefficients estimates for pairwise comparisons involving inbred individuals

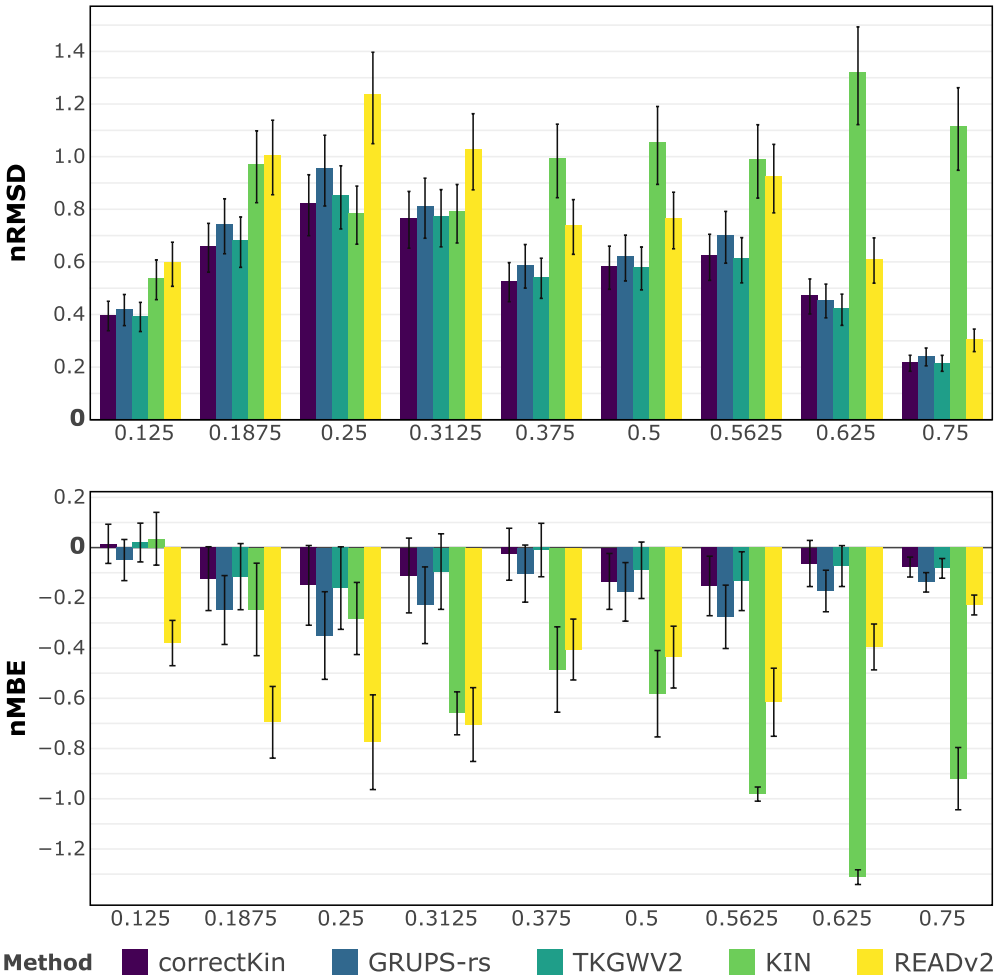

Fig. S15: $nRMSD$ (top row) and $nMBE$ (bottom column) estimates, obtained when simulating inbreeding, across all evaluated methods (bar colours) and expected relatedness coefficients (x-axis ticks). Increasing values of $nRMSD$ indicate lower accuracy when estimating relatedness coefficients. $nMBE$ values that deviate furthest from zero indicate higher bias, with positive and negative values highlighting a tendency towards over- or under-estimating r-coefficients, respectively. Error bars represent $CI_{95\%}$ for the given estimate.

1.16 Figure S16: Impact of inbreeding on the accuracy and bias of r-coefficients estimates for pairwise comparisons involving outbred individuals

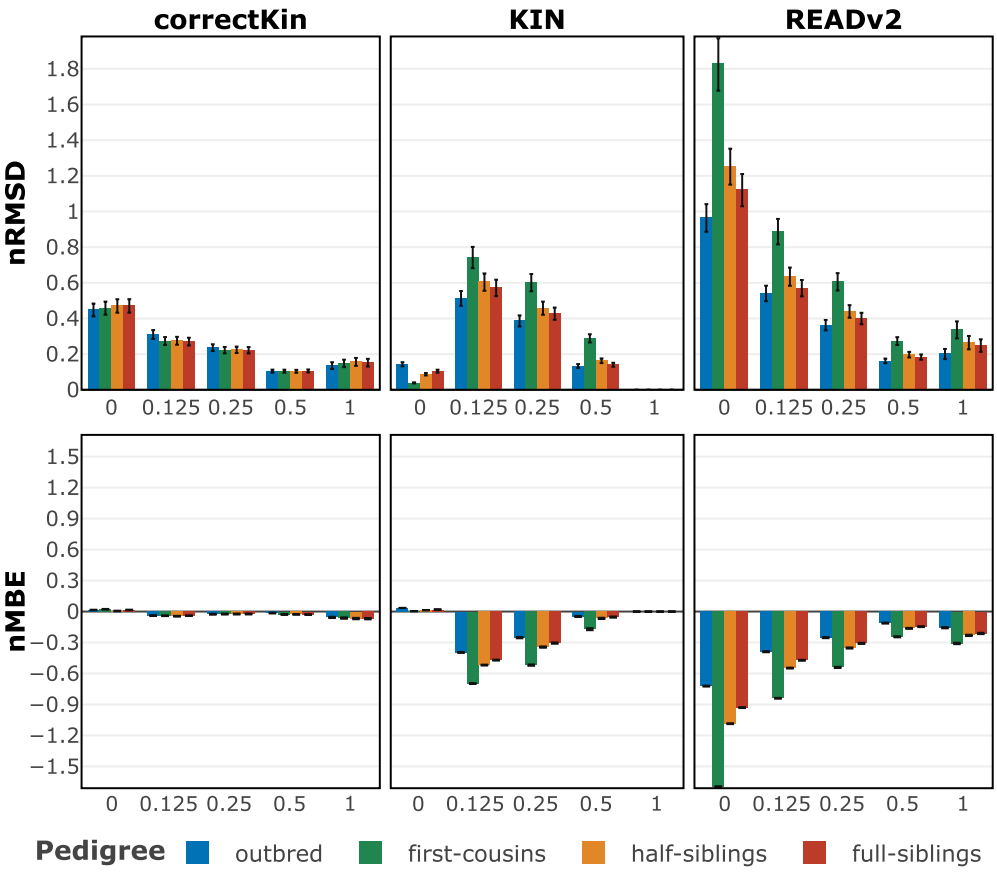

Fig. S16: $nRMSD$ (top row) and $nMBE$ (bottom column) estimates, obtained when simulating inbreeding, across all evaluated methods (columns), expected relatedness coefficients (x-axis ticks) and pedigree scenarii (bar colours). Increasing values of $nRMSD$ indicate lower accuracy when estimating relatedness coefficients. $nMBE$ values that deviate furthest from zero indicate higher bias, with positive and negative values highlighting a tendency towards over- or under-estimating r-coefficients, respectively. Error bars represent $CI_{95\%}$ for the given estimate.

2 Material and Methods

2.1 Description of BADGER's simulation pipeline

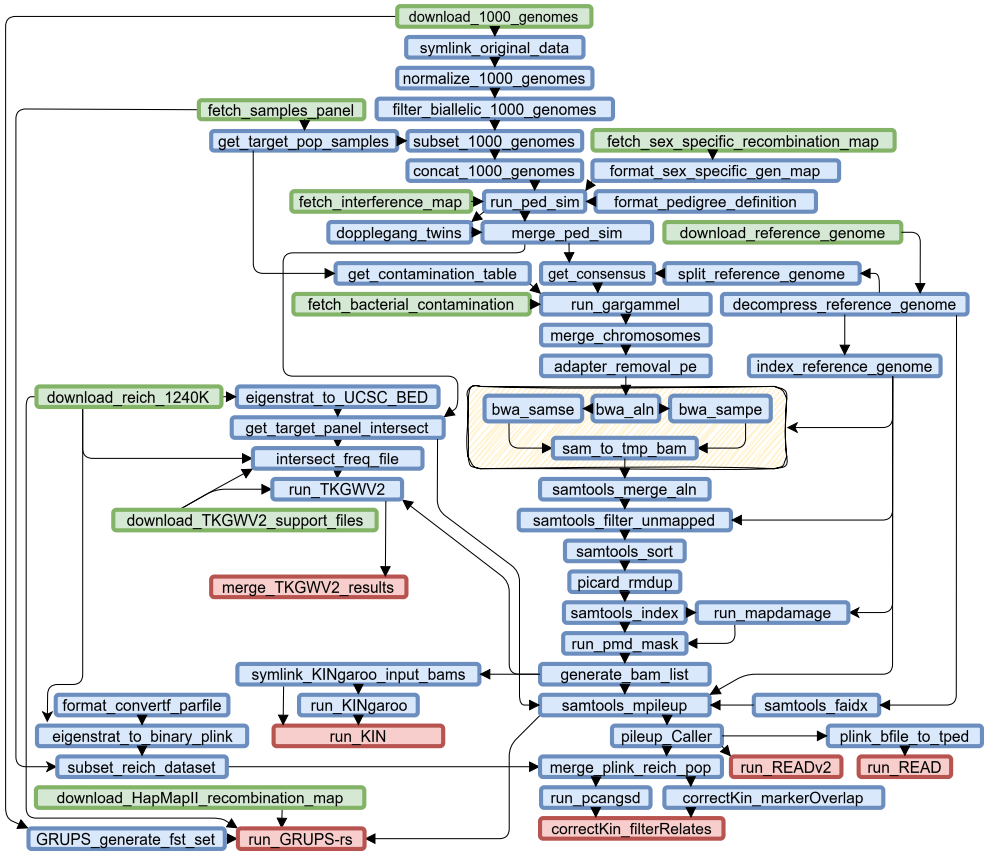

Fig. S17: Complete directed acyclic graph of the BADGER workflow, given the input and parameters provided throughout this study. Node names follow those of the snakemake rules found in BADGER's source code. Green nodes represent data entry points which are automatically downloaded by BADGER. Red nodes represent snakemake rules that are targeted by BADGER by default. The yellow cluster denotes a grouped data input for the reference genome.

2.1.1 1000 genomes dataset pre-processing

BADGER first downloads the 1000 genomes phase3-v5b dataset from the EMBL–EBI FTP website [1] (See: Key Resources Table) and proceeds to apply normalization and left-alignment of indels, using `bcftools norm`. Multi-allelic positions and any lingering unphased genotypes are then filtered out using `bcftools view` (`--phased -m2 -M2`). From the processed dataset, BADGER then generates two data subsets according to the population or super-population label of the samples:

- A concatenated VCF file containing all samples belonging to the selected founder population. This file is used as an input to `ped-sim` when simulating pedigrees, as a source of founder individuals. During this study, the selected founder population was either `CEU`, to simulate a genetically homogeneous population, or one of the five populations belonging to the admixed American population (`ASW`, `CLM`, `MXL`, `PEL`, `PUR`) to simulate an admixed population.
- A set of VCF files (one for each autosome), containing all samples belonging to the selected contaminating population is generated. This second dataset is used as input for `gargammel` when simulating ancient DNA fragments to extract a single contaminating individual, and use its genotype as a source of modern human contamination. During this study, we selected either the `AFR` super-population, or the `GBR` population as a source to simulate modern human contamination.

2.1.2 Pedigree simulations

BADGER leverages the software `ped-sim` to simulate pedigrees in multiple replicates, using founder individuals randomly selected from the `founder dataset`. Here, we parametrized `ped-sim` to simulate sex-specific recombination rates, as well as a crossover interference model, using the refined genetic map from [4] and the interference parameter estimates of [5], respectively. To maximize the number of possible combinations, and given that BADGER only simulates autosomes, the original genetic

sex of the individuals selected as founders was not taken into account to select founders within the pedigree replicates.Simulation of genotyping errors, opposite homozygous errors, missingness, and pseudo-haploid rates were all disabled at this step, to prevent any compounding interactions with the error model of `gargammel` (`--err_rate 0 --err_hom_rate 0 --miss_rate 0 --pseudo_hap 0`).

Simulation of monozygotic twins and/or duplicate samples within the pedigree is performed by merely duplicating the genotype of the selected individuals within the output VCF of `ped-sim` (See rule "*dopplegang_twins*", Supplementary Figure S17).

2.1.3 Ancient DNA simulations

Simulation of raw ancient DNA fragments for every pedigree individual is performed using `gargammel`. As this software requires the use of FASTA-format haplotype sequences, BADGER first uses `bcftools consensus` to apply the variants from the output VCF file of `ped-sim` to a reference `.fasta` file, thus generating a consensus sequence for every pedigree individual and, when simulating non-null rates of modern human contamination, a randomly selected contaminating sample from the `contaminant dataset`.

Here, note that copy-number variations, two-sided inversions, and insertions of `ALU`, `LINE1`, `SVA` and Nuclear Mitochondrial elements are filtered out using regular expressions, to comply with the requirements of `bcftools` (`--exclude 'ALT~"<CN [0-9].*>"||ALT~"<INS:.*>" || ALT~"<INV>'`).

For every individual, haplotype sequences are then inserted in the required `endo` input directory of `gargammel`. Likewise, when simulating human contamination, the haplotype sequences of the randomly sampled individual are inserted in the optional `cont` input directory. BADGER then applies `gargammel` on these input directories, using the user-provided misincorporation probability and fragment size frequency profiles. Here, we elected to use the *post-mortem* damage profile of "Chan_Meso": a young adult female individual dated from the Mesolithic period (9137 ± 124 *cal.BP*) and

exhumed from the "Chan do Lindero" karst system of Pedrafita, Lugo, Spain [6]. This choice of reference was motivated by the fact that Chan_Meso was sequenced on an Illumina HiSeq2000 platform – one of the preset sequencing platform model choice for `gargammel`'s – and exhibits "average" *post-mortem* damage patterns (i.e. an approximate misincorporation rate of 0.22, at the $3'$-and $5'$-end of reads, and a mode of approximately 70 base pairs on its fragment size frequency distribution). Note that while BADGER can be parametrized to handle bacterial contamination from publicly available databases, this capacity was not leveraged during the present study.

To optimize the I/O throughput and runtime performance of BADGER, generation of ancient DNA fragments using `gargammel` is applied in parallel for every pedigree individual, on a per-chromosome basis using a simple scatter-gather approach. Hence, per-chromosome FASTQ files are merely concatenated using `zcat` and `gzip` UNIX command-line utilities.

2.1.4 Alignment

The raw paired-end fragments of every individual composing the pedigree are then trimmed of adapter sequences and collapsed, using `AdapterRemovalv2` [7], requiring a minimum adapter overlap of 1, read length of 17 and base quality of 20. (`--minlength 17 --minquality 20 --minadapteroverlap 1`).

Trimmed fragments are then aligned against the GRCh37 reference genome, using `bwa aln` [8], following the best practices described in [9] (`-l 1024 -n 0.01 -k 2 -o 2`). Note that collapsed single-end sequences and non-collapsed paired-end sequences are mapped separately, using `bwa samse` and `bwa sampe` respectively, and then merged using `samtools merge`. Here, a generic read group tag is placed using `samtools addreplacerg` following merging.

2.1.5 Quality filtering and preprocessing of alignment files

Following alignment and merging, a simple quality filtration step is first applied to the raw binary alignment files of every sample using `samtools view`. Hence, the raw files are trimmed of any sequence that is either a) unmapped, b) measuring less than 30 nucleotides, or c) carrying a mapping quality score lower than 20 (PHRED scale) (`-F4 -q20 -e 'length(seq)>30'`).

Alignment files are then sorted using `samtools sort` and removed of any optical PCR duplicates using `picard MarkDuplicates` (`--REMOVE_DUPLICATES true --VALIDATION_STRINGENCY LENIENT --ASSUME_SORT_ORDER coordinate`)

2.1.6 Correction of *post-mortem* deaminations

Patterns of *post-mortem* deamination were estimated on every sample alignment file, using `mapDamage2`. To estimate the performance impact of applying *post-mortem* damage rescaling, two alternative post-processing methods were then applied:

- "Rescaled" versions of the alignment files, wherein the base quality scores of putative nucleotide misincorporation sites are downscaled, were generated by applying the `--rescale` flag of `mapDamage2`.
- "Masked" version of the alignment files were generated using the in-house software `pmd-mask` and the misincorporation probability estimates of `mapDamage2` (`misincorporation.txt` file). Here, potential `C>T` and `G>A` deamination sites are masked all together, along the `5'` and `3'` ends of fragments, respectively, until the misincorporation probability is less than 1%.

2.1.7 Variant calling

Next, BADGER jointly applies random pseudo-haploid variant calling on every post-processed alignment file by first creating a pileup file with `samtools mpileup`. Here, autosomal bi-allelic SNP positions from the AADR "1240K" SNP dataset, version 52.2

were targeted [10], while disabling Base alignment quality (BAQ) recalculation, and filtering out any position with a mapping and/or base quality lower than 20 (`-RB -q20 -Q20`). This pileup file is then directly given as input to the `pileupCaller` module of `sequenceTools` (https://github.com/stschiff/sequenceTools), to generate pseudo-haploid variant calls (`--randomHaploid --minDepth 1`), in binary PLINK format.

2.1.8 Genetic relatedness estimation

Note that the benchmarked genetic relatedness estimation methods may have differing input data. Hence:

- The random pseudo-haploid variant calls of `pileupCaller` were given as input to READv1 and READv2.
- The joint pileup file of `samtools mpileup` was given as input to GRUPS-rs
- The post-processed binary alignment files of every sample composing a pedigree replicate were given as input to correctKin, KIN and TKGWV2.

correctKin

Following the guidelines of [11], BADGER first generates a subset of the AADR "1240K", to provide correctKin with a user-selected set of reference individual genotypes. Here, all samples belonging to the EUR super-population of the 1000g-phase3 dataset and contained within the 1240K dataset were selected as reference individuals for correctKin during this study. However, as BADGER also makes use of 1000-genomes samples as a source of founder individuals during pedigree simulations, the pipeline first excludes any sample previously given as an input to `ped-sim`, from the list of reference samples added to the correctKin input dataset. BADGER then merges the resulting "1240K" data subset with the pseudo-haploid variant callset of the pedigree replicate, using `plink` (`--bmerge --merge-mode 1 --allow-no-sex --keep-allele-order`). Still following guidelines, a covariance matrix and a marker overlap fraction matrix are generated from this merged dataset, using `pcangsd` [12]

and the `markerOverlap` module of correctKin. Unrelated individuals were filtered out using the `filterRelates` module of correctKin. Here, note that, a) all pairs of individuals not found in the resulting output file were considered as unrelated, and b) pairs of individuals classified as "uncertain" are reclassified as unrelated.

GRUPS-rs

Following the guidelines of [13], BADGER first creates an FSA-encoded dataset of reference individuals from the raw 1000-genomes phase3 dataset, using the `grups-rs fst` module. This preprocessing step is merely intended to increase the runtime efficiency of BADGER and the resulting fsa-encoded 1000g-phase3 dataset is used as an input throughout all pedigree replicates. For every pedigree replicate, BADGER then directly applies `grups-rs pedigree-sims` on the pileup file described in 2.1.7, while requesting 1000 simulation replicates, and using samples from the 1000g-phase3 EUR super-population as reference individuals (`--reps 1000 --pedigree-pop EUR --min-depth 1 --seq-error-rate 0.0`).

Since GRUPS-rs requires the use of a user-constructed template pedigree to perform its simulations, we provided the software with the same template throughout this study (Supplementary Figure S18). Note that this simple pre-constructed template is made available as an example within the sofware's documentation, contains a pair of siblings, half-siblings and first-cousins, and uses these comparisons to estimate first-, second- and third-degree relationships, respectively.

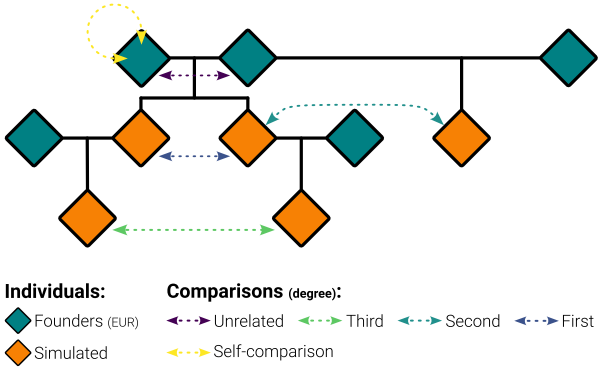

Fig. S18: Diagram of the input template pedigree definition file provided to the GRUPS-rs method throughout this study.

KIN

BADGER first applies the `KINgaroo` module on the post-processed binary alignment files of all individuals composing a pedigree replicate. Bi-allelic autosomal SNPs from the "1240K" dataset are targeted during this preprocessing step (`--bedfile`), while disabling contamination correction (`--contam_parameter 0`). The main `KIN` module is then applied on the output of `KINgaroo` using default parameters. From the final output of `KIN`, a coefficient of relatedness r is deduced for all tested pairs of individuals, using the provided Cotterman coefficient estimates for every given pair, i.e.: $r = k_1/2 + k_2$.

READv1

The pseudo-haploid variant callset of pileupCaller described in section 2.1.7 is given as input to `READ`, with default arguments (normalization statistic: median, sliding window size: 10^6 bp).

From the final output of READv1, the relatedness coefficient r of every tested pair of individuals is derived from the normalized $\overline{P}_0$ estimates of a given pair, using the following equation: $r = 2(1 - \overline{P}_0)$.

READv2

Likewise, for every pedigree replicate, the pseudo-haploid variant dataset produced by `pileupCaller` in section 2.1.7 is provided to `READv2` by BADGER, using the default parameters (`--norm_method median`). Contrary to READv1, the coefficient of relatedness r is directly obtained from the final output results of the software.

TKGWV2

BADGER applies `TKGWV2` on every tested pair of post-processed alignment files, using the support 1000g-phase3 EUR population bed files and allele frequencies provided in [14] (`1000GP3_22M_noFixed_noChr.bed` and `1000GP3_EUR_1240K.frq`, respectively). Here, note that providing `TKGWV2` with a pre-defined dataset of allele frequencies incurs the risk of targeting positions that were not simulated by `ped-sim` during a particular run, as BADGER makes successive use of pedigree simulations using a pre-processed 1000g dataset, followed by the creation of consensus sequences (which will naturally exclusively contain reference alleles). Hence, to alleviate this potential source of reference bias, we first filter the provided `1000GP3_EUR_1240K.frq` file by removing any position that was not found within the raw output VCF file previously emitted by `ped-sim`. Also note that, contrary to all other candidate methods, `TKGWV2` may only be applied on a single pair of individuals. Thus, applying this method on an entire pedigree replicate requires BADGER to employ a scatter-gather approach, by running the software on every tested pair of individuals, and subsequently merging the results.

2.2 Statistical analysis and benchmark using badger.plots

Following the application of BADGER in multiple replicates, the statistical analysis and performance estimation of each method is handled using `badger.plots` : a command line interface, written as a companion software to BADGER. This software thus sequentially performs :

1. The deserialization and consolidation of the results of each genetic relatedness estimation software, across all BADGER simulation replicates and sets of studied biological parameters.
2. The calculation of summary statistics regarding the classification performance of each method, for each biological parameter studied.
3. The estimation of the average accuracy and bias of each method's r-coefficient calculation for each degree of relationship.
4. The generation of interactive plots summarizing these performance statistics.

2.2.1 Estimation of classification performance

For every biological condition and method, we constructed confusion matrices confronting the predicted degrees of relationship for all pairwise comparisons, against the "true" degrees of relationship, defined by the topology of the template pedigree originally given as an input for BADGER. From these confusion matrices – one for every pedigree replicate – we calculated the Uniform Ordinal Classification Index (UOC) as a measure of a method's overall classification performance [15]. Briefly, this performance metric, adapted from the ordinal classification index of Cardoso and Sousa [16], is bound between 0 and 1 (0 implying perfectly accurate classification), insensitive to class-imbalance, and markedly takes into account the inherent relative order, and ranking distance separating two degrees of relationship. As such, the ordinal nature of estimating genetic relatedness is retained when estimating performance (e.g. misclassifying an "Unrelated" pairs of individuals as "First-degree" is more penalized than misclassifying them as "Second-degree"). Here, our implementation of the UOC metric was incorporated into `badger.plots` by adapting the pseudo-code found in [16], and source code provided in [17]. For every method and biological condition, the average UOC of every pedigree replicate is then calculated and plotted as a final aggregate summary statistic. 95% confidence intervals are directly estimated from the

distribution of UOC values across all simulation replicates, using normal approximation. When applicable, estimates for the area under the curve (AUC) of UOC values of every method were obtained through trapezoidal integration, using the R package `pracma` version 2.4.4, and its associated function `trapz`.

2.2.2 Average accuracy and bias of relatedness coefficients

For every simulated degree of relationship, and across all tested methods and biological conditions, we calculate the average Root Mean Square Deviation (RMSD) and (Mean Bias Error) between the *calculated* and the *expected* relatedness coefficients (r), in an effort to gain insight regarding the average accuracy and bias of each method. Note that, as many of the methods tested here do not *directly* compute r-coefficients, `badger.plots` must first derives this metric from the raw output of every method. Hence:

- As KIN estimates the Jacquard genetic identity coefficients of every pairwise relationship [18], we derived an r-coefficient from the provided k_1 and k_2 values, i.e.: $r = \frac{k_1}{2} + k_2$
- for READ and READv2, an r-coefficient can be calculated from the normalized $\overline{P_0}$ of a given pair, using the following formula: $r = 2 \cdot (1 - \overline{P_0})$
- for GRUPS-rs, an r-coefficient can be derived by first calculating normalized estimates of the $PWD_{i,j}^{obs}$ metric of a given pair i, j, which is obtained by dividing this raw estimate by the average expected distribution of unrelated pairs ($\widehat{PWD}_{i,j,unrelated}^{sim}$). It follows that the r-coefficient can be derived using the following equation: $r = 2(1 - \frac{PWD_{i,j}^{obs}}{\widehat{PWD}_{i,j,unrelated}^{sim}})$
- Finally, as both correctKin and TKGWV2 compute a kinship coefficient (ϕ), the r-coefficient is simply obtained by multiplying this estimate by 2: $r = 2 \cdot \phi$

Here, it must be noted that the distance separating the expected r-coefficient of a given degree of relationship from neighbouring distributions varies with the degree of

relatedness, and is effectively halved for each additional degree separating two individuals. This implies that a given deviation from the expected average can have a greatly differing impact on the general accuracy, depending on the degree for which it is observed (e.g. a standard deviation of 0.1 for the r-coefficient between two individuals is insignificant when considering first-degree relationships, but would consistently cause misclassifications in the case of third-degree relationships).

Thus, to properly compare the accuracy and bias, both across a given method and degree of relatedness, we propose to first normalize the RMSD and MBE of a given relationship by the range of its theoretical distribution. This can be done by dividing the RMSD of MBE value by the distance separating the two midpoints found between the mean of a given relationship k and its neighbouring ones ($k-1$ and $k+1$).

$$nRMSD_{m,k} = \frac{\sqrt{\frac{\sum_{i=1}^{i=N}(\widehat{r}_k - r_{m,k,i})^2}{N}}}{\frac{\widehat{r}_{k+1} - \widehat{r}_{k-1}}{2}} \tag{1}$$

$$nMBE_{m,k} = \frac{\frac{\sum_{i=1}^{i=N}(r_{m,k,i} - \widehat{r}_k)}{N}}{\frac{\widehat{r}_{k+1} - \widehat{r}_{k-1}}{2}} \tag{2}$$

where m and k represent a given method and degree of relatedness, respectively. $\widehat{r}_k$ is the expected relatedness coefficient of relationship k ; $r_{m,k,i}$, the calculated relatedness coefficient for the i^{th} pair of individuals, and N, the total amount of observations of a given relationship k ($k = 0, 0.125, 0.25, 0.5, 1$). 95% confidence intervals for these metrics were calculated using the principles and methods described in [19, 20], i.e.:

$$nRMSD_{m,k} \in \left[nRMSD_{m,k}\left(1 - \sqrt{1 - \frac{1.96\sqrt{2}}{\sqrt{N-1}}}\right); RMSD_{m,k}\left(\sqrt{1 + \frac{1.96\sqrt{2}}{\sqrt{N-1}}} - 1\right)\right] \tag{3}$$

$$nMBE_{m,k} \in \left[nMBE_{m,k} \pm \frac{1.96 \cdot \sigma_{m,k}}{\sqrt{N}} \right] \quad (4)$$

where $\sigma_{m,k}$ is the population standard deviation of the relatedness coefficients obtained from method m, and belonging to relationship k.

2.3 Key Resources Table

Reagent or Resource	Source	Identifier
Deposited data		
1000g-phase3-v20130502	IGSR [1]	https://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/
HapMapII	IHMP [21]	http://ftp.ncbi.nlm.nih.gov/hapmap/recombination/2011-01_phaseII_B37/
GRCh37-release113	Church et al. [22]	http://ftp.ensembl.org/pub/grch37/release-113/fasta/homo_sapiens/dna/
Cross-over interference model	Campbell et al. [5]	https://raw.githubusercontent.com/williamslab/ped-sim/refs/heads/master/interfere/nu_p_campbell.tsv
Sex-specific genetic maps	Bhérer et al. [4]	https://github.com/cbherer/Bherer_etal_SexualDimorphismRecombination/
AADR dataset v52.2	Mallick et al. [10]	https://reichdata.hms.harvard.edu/pub/datasets/amh_repo/curated_releases/V52/V52.2/SHARE/public.dir/
TKGWV2 support files	Fernandes et al. [14]	https://github.com/danimfernandes/tkgwv2
Softwares and algorithms		
AdapterRemoval-v2.3.3	Schubert et al. [7]	RRID:SCR_011834
ANGSD-v0.939	Korneliussen et al. [23]	RRID:SCR_021865
BADGER-v0.5.1	This study	https://github.com/MaelLefeuvre/badger/tree/v0.5.1
bcftools-1.15	Li [3]	RRID:SCR_005227
conda-23.1.0	Conda contributors [24]	RRID:SCR_018317
correctKin	Nyerki et al. [11]	RRID:SCR_026952
gargammel-1.1.4	Renaud et al. [25]	RRID:SCR_026953

grups-rs-0.3.2	Lefeuvre et al. [13]	RRID:SCR_026954
kin-3.1.3	Popli et al. [26]	RRID:SCR_026955
mapDamage-v2.2.1	Jónsson et al. [27]	RRID:SCR_001240
pcangsd-0.99	Meisner and Albrechtsen [12]	RRID:SCR_026956
ped-sim-v1.4	Caballero et al. [28]	RRID:SCR_026957
picard-v2.27.4	Broad Institute [29]	RRID:SCR_006525
plink-v1.9	Chang et al. [30]	RRID:SCR_001757
pmd-mask-v0.3.2	This study	https://github.com/MaelLefeuvre/pmd-mask/tree/v0.3.2
READ-v1.0	Kuhn et al. [31]	RRID:SCR_026958
READv2-v2.00	Alaçamlı et al. [32]	RRID:SCR_026959
samtools-v1.15	Li [3]	RRID:SCR_002105
sequenceTools-v1.5.2	Schiffels [33]	https://github.com/stschiff/sequenceTools/tree/v1.5.2
snakemake-7.20.0	Mölder et al. [34]	RRID:SCR_003475
TKGWV2	Fernandes et al. [14]	RRID:SCR_026960
python-3.11.0	Python Software Foundation	RRID:SCR_008394
R-v4.1.2	R Development Core Team	RRID:SCR_001905

3 Description of the pmd-mask command line utility

3.1 Rationale, behaviour and workflow description

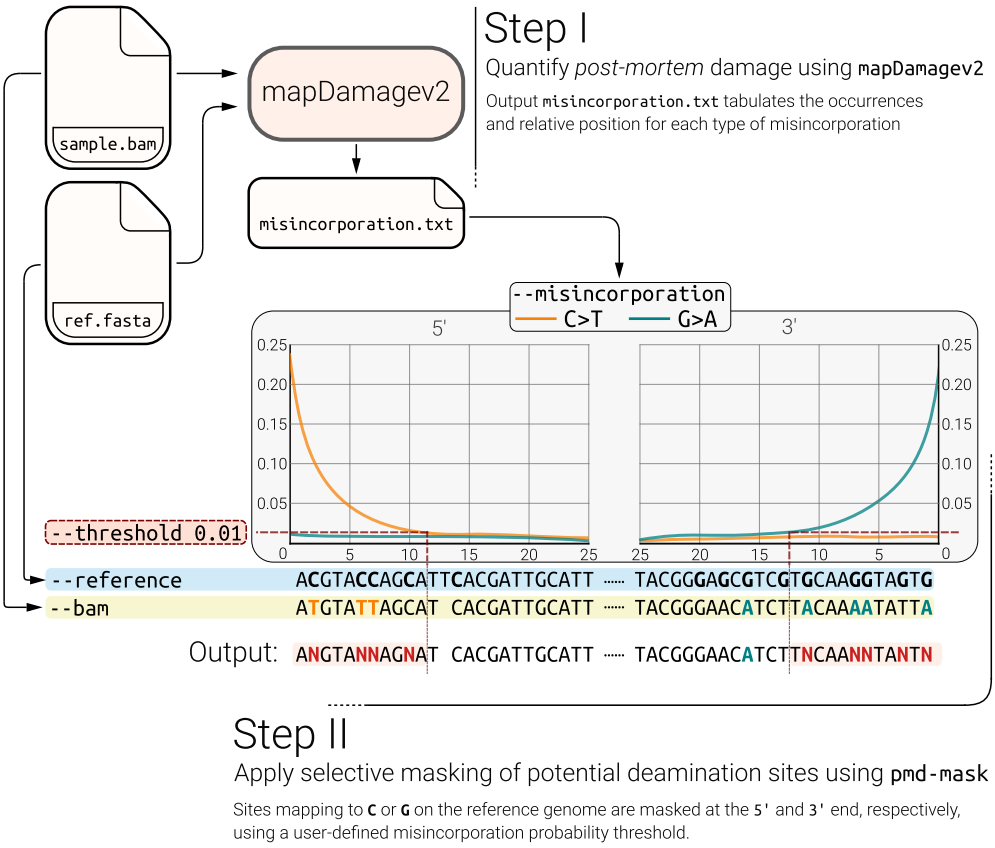

Fig. S19: Summary diagram describing the main process, required inputs, and workflow surrounding the `pmd-mask` command-line utility.

`pmd-mask` is a simple command-line pre-processing tool written in the Rust programming language, which masks the positions from reads that are likely to be impacted by *post-mortem* damage (PMD). Briefly, `pmd-mask` leverages nucleotide– and position-specific misincorporation rate estimates emitted from the mapDamage2 software (typically, in the form of a `misincorporation.txt` output file) to apply soft-trimming on the read extremities, until the PMD rate reaches a user defined threshold

by setting candidate nucleotides to 'N' and their base-quality to 0. Note that the default threshold is here defined as a misincorporation rate of 0.01, but may be modified at leisure by the user, using the `--threshold` argument. Note that, in its current state, the `pmd-mask` algorithm only considers the misincorporation rate found at a given position, and the genotype found in the *reference* genome. Therefore, nucleotides at the extremities of a read are masked, regardless of the actual genotype observed at a given position (Supplementary Figure S19). Pseudo-code snippets, summarizing the main loop of the program can be read in Algorithm-1. Here, the devised approach is one that i) is expected to incur less potential bias than when applying PMD-rescaling through mapDamage2 (i.e. using its provided `--rescale` flag), and ii) carries the benefit of mitigating the loss of information usually displayed when applying hard-clipping, by instead specifically targeting potential `C>T` and `G>A` transition sites on both the $5'$ and $3'$ end of the read, respectively. In other terms, this method may be regarded as a conservative compromise between *post-mortem* damage rescaling methods such as mapDamage2, PMDtools, or ATLAS [27, 35, 36] and hard-clipping methods such as the trimBam module of the bamUtil software [37]. Required inputs for `pmd-mask` are as follows:

- `--bam`: An input `.bam` file (SAM, BAM, and CRAM formats are accepted). pmd-mask can either read from a file (using `-b`|`--bam`) or from the standard input, through shell piping.
- `--misincorporation`: A mapDamage2 `misincorporation.txt` file. This file provides strand-specific PMD frequency estimates, which are used to compute the threshold at which masking should be performed. Evidently, this file must have been generated from the input BAM file to provide a sound estimate.
- `--reference`: A reference genome, in the form of a `.fasta` file . This genome must of course be identical to the one used to align the input BAM file.

Additional instructions regarding the installation and usage of `pmd-mask`, as well as its source code is made publicly available at https://github.com/MaelLefeuvre/pmd-mask, under GPL-v3.0 licencing.

3.2 Pseudo-code describing the main algorithm of pmd-mask

Algorithm 1 Pseudo-code describing the main algorithm of pmd-mask.

Ensure: : $args.threshold \in [0.00, 1.00[$

```
bam ← bam_reader(args.bam_path);
refseq ← fasta_reader(args.reference_path)
misincorporation ← misincorporation_reader(args.misincorporation_path)

if args.threshold is not null then
    threshold ← args.threshold
else
    threshold ← 0.01

function MASK(read, position)
    read.nucleotide[position] ←' N';
    read.quality[position] ← 0

output_bam ← bam.copy_header()
for read ∈ bam do
    ▷ Extract read length, coordinate and strand information ◁
    n ← read.length
    chr ← read.chromosome
    pos ← read.position
    strand ← read.strand
    ▷ Mask 5' Cytosines ◁
    for mask5' : (i = 0; i < n; i++) do
        if misincorporation['C > T'][chr][strand][i] ≤ args.threshold then
            break mask5p'
        else if reference.get(chr, pos + i) ==' C' then
            MASK(read, i)
    ▷ Mask 3' Guanines ◁
    for mask3' : (i = n; i > 0; i − −) do
        if misincorporation['G > A'][chr][strand][i] ≤ args.threshold then
            break mask3'
        else if reference.get(chr, pos + i) ==' G' then
            MASK(read, i)

    ▷ Store masked read ◁
    output_bam+ = read
 return (output_bam)
```

References

[1] The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74. Number: 7571. https://doi.org/10.1038/nature15393.

[2] Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. The American Journal of Human Genetics. 2017 Apr;100(4):635–649. https://doi.org/10.1016/j.ajhg.2017.03.004.

[3] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov;27(21):2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

[4] Bhérer C, Campbell CL, Auton A. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. Nature Communications. 2017 Apr;8(1):14994. Number: 1. https://doi.org/10.1038/ncomms14994.

[5] Campbell CL, Furlotte NA, Eriksson N, Hinds D, Auton A. Escape from crossover interference increases with maternal age. Nature Communications. 2015 Feb;6(1):6260. Number: 1. https://doi.org/10.1038/ncomms7260.

[6] González-Fortes G, Jones ER, Lightfoot E, Bonsall C, Lazar C, Grandal-d'Anglade A, et al. Paleogenomic Evidence for Multi-generational Mixing between Neolithic Farmers and Mesolithic Hunter-Gatherers in the Lower Danube Basin. Current Biology. 2017 Jun;27(12):1801–1810.e10. https://doi.org/10.1016/j.cub.2017.05.023.

[7] Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Research Notes. 2016 Feb;9(1):88. https://doi.org/10.1186/s13104-016-1900-2.

[8] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009 Jul;25(14):1754–1760. Number: 14. https://doi.org/10.1093/bioinformatics/btp324.

[9] Oliva A, Tobler R, Cooper A, Llamas B, Souilmi Y. Systematic benchmark of ancient DNA read mapping. Briefings in Bioinformatics. 2021 Sep;22(5):bbab076. https://doi.org/10.1093/bib/bbab076.

[10] Mallick S, Micco A, Mah M, Ringbauer H, Lazaridis I, Olalde I, et al. The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. Scientific Data. 2024 Feb;11(1):182. https://doi.org/10.1038/s41597-024-03031-7.

[11] Nyerki E, Kalmár T, Schütz O, Lima RM, Neparáczki E, Török T, et al. correctKin: an optimized method to infer relatedness up to the 4th degree from low-coverage ancient human genomes. Genome Biology. 2023 Feb;24(1):38. Number: 1. https://doi.org/10.1186/s13059-023-02882-4.

[12] Meisner J, Albrechtsen A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. Genetics. 2018 Oct;210(2):719–731. https://doi.org/10.1534/genetics.118.301336.

[13] Lefeuvre M, Martin MD, Jay F, Marsolier MC, Bon C. GRUPS-rs, a high-performance ancient DNA genetic relatedness estimation software relying on pedigree simulations. Human Population Genetics and Genomics. 2024 Jan;4(1). https://doi.org/10.47248/hpgg2404010001.

[14] Fernandes DM, Cheronet O, Gelabert P, Pinhasi R. TKGWV2: an ancient DNA relatedness pipeline for ultra-low coverage whole genome shotgun data. Scientific Reports. 2021 Oct;11(1):21262. Number: 1. https://doi.org/10.1038/s41598-021-00581-3.

[15] Silva W, Pinto JR, Cardoso JS. A Uniform Performance Index for Ordinal Classification with Imbalanced Classes. In: 2018 International Joint Conference on Neural Networks (IJCNN); 2018. p. 1–8. ISSN: 2161-4407.

[16] Cardoso JS, Sousa R. Measuring the performance of ordinal classification. International Journal of Pattern Recognition and Artificial Intelligence. 2011 Dec;25(08):1173–1195. https://doi.org/10.1142/S0218001411009093.

[17] Albuquerque T, Cruz R, Cardoso JS. Ordinal losses for classification of cervical cancer risk. PeerJ Computer Science. 2021 Apr;7:e457. https://doi.org/10.7717/peerj-cs.457.

[18] Jacquard A. Genetic Information Given by a Relative. Biometrics. 1972;28(4):1101–1114. https://doi.org/10.2307/2528643.

[19] Nicholls A. Confidence limits, error bars and method comparison in molecular modeling. Part 1: The calculation of confidence intervals. Journal of Computer-Aided Molecular Design. 2014 Sep;28(9):887–918. https://doi.org/10.1007/s10822-014-9753-z.

[20] Nicholls A. Confidence limits, error bars and method comparison in molecular modeling. Part 2: comparing methods. Journal of Computer-Aided Molecular Design. 2016 Feb;30(2):103–126. https://doi.org/10.1007/s10822-016-9904-5.

[21] Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007

Oct;449(7164):851–861. https://doi.org/10.1038/nature06258.

[22] Church DM, Schneider VA, Graves T, Auger K, Cunningham F, Bouk N, et al. Modernizing Reference Genome Assemblies. PLOS Biology. 2011 Jul;9(7):e1001091. https://doi.org/10.1371/journal.pbio.1001091.

[23] Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics. 2014 Nov;15(1):356. https://doi.org/10.1186/s12859-014-0356-4.

[24] Conda contributors.: conda: A system-level, binary package and environment manager running on all major operating systems and platforms. Available from: https://docs.conda.io/projects/conda/.

[25] Renaud G, Hanghøj K, Willerslev E, Orlando L. gargammel: a sequence simulator for ancient DNA. Bioinformatics. 2017 Feb;33(4):577–579. Number: 4. https://doi.org/10.1093/bioinformatics/btw670.

[26] Popli D, Peyrégne S, Peter BM. KIN: a method to infer relatedness from low-coverage ancient DNA. Genome Biology. 2023 Jan;24(1):10. Number: 1. https://doi.org/10.1186/s13059-023-02847-7.

[27] Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 2013 Jul;29(13):1682–1684. Number: 13. https://doi.org/10.1093/bioinformatics/btt193.

[28] Caballero M, Seidman DN, Qiao Y, Sannerud J, Dyer TD, Lehman DM, et al. Crossover interference and sex-specific genetic maps shape identical by descent sharing in close relatives. PLOS Genetics. 2019 Dec;15(12):e1007979. Number: 12. https://doi.org/10.1371/journal.pgen.1007979.

[29] Broad Institute.: Picard: A set of command line tools (in Java) for manipulating high-throughput sequencing (HTS) data and formats such as SAM/BAM/CRAM and VCF. Broad Institute. Available from: http://broadinstitute.github.io/picard.

[30] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015 Dec;4(1):s13742–015–0047–8. https://doi.org/10.1186/s13742-015-0047-8.

[31] Kuhn JMM, Jakobsson M, Günther T. Estimating genetic kin relationships in prehistoric populations. PLOS ONE. 2018 Apr;13(4):e0195491. Number: 4. https://doi.org/10.1371/journal.pone.0195491.

[32] Alaçamlı E, Naidoo T, Güler MN, Sağlıcan E, Aktürk S, Mapelli I, et al. READv2: advanced and user-friendly detection of biological relatedness in archaeogenomics. Genome Biology. 2024 Aug;25(1):216. https://doi.org/10.1186/s13059-024-03350-3.

[33] Schiffels S.: sequenceTools. Available from: https://github.com/stschiff/sequenceTools.

[34] Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, et al.: Sustainable data analysis with Snakemake. F1000Research. Available from: https://f1000research.com/articles/10-33.

[35] Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences. 2014 Feb;111(6):2229–2234. https://doi.org/10.1073/pnas.1318934111.

[36] Link V, Kousathanas A, Veeramah K, Sell C, Scheu A, Wegmann D.: ATLAS: Analysis Tools for Low-depth and Ancient Samples. bioRxiv. Pages: 105346 Section: New Results. Available from: https://www.biorxiv.org/content/10.1101/105346v2.

[37] Jun G, Wing MK, Abecasis GR, Kang HM. An efficient and scalable analysis framework for variant extraction and refinement from population scale DNA sequence data. Genome Research. 2015 Apr;p. gr.176552.114. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab. https://doi.org/10.1101/gr.176552.114.