

Haplotype-based inference of recent effective population size in modern and ancient DNA samples

Received: 3 July 2022

Accepted: 10 November 2023

Published online: 01 December 2023

 Check for updatesRomain Fournier¹, Zoi Tsangalidou¹, David Reich^{2,3,4,5,7} & Pier Francesco Palamara^{1,6,7} 

Individuals sharing recent ancestors are likely to co-inherit large identical-by-descent (IBD) genomic regions. The distribution of these IBD segments in a population may be used to reconstruct past demographic events such as effective population size variation, but accurate IBD detection is difficult in ancient DNA data and in underrepresented populations with limited reference data. In this work, we introduce an accurate method for inferring effective population size variation during the past ~2000 years in both modern and ancient DNA data, called HapNe. HapNe infers recent population size fluctuations using either IBD sharing (HapNe-IBD) or linkage disequilibrium (HapNe-LD), which does not require phasing and can be computed in low coverage data, including data sets with heterogeneous sampling times. HapNe shows improved accuracy in a range of simulated demographic scenarios compared to currently available methods for IBD-based and LD-based inference of recent effective population size, while requiring fewer computational resources. We apply HapNe to several modern populations from the 1,000 Genomes Project, the UK Biobank, the Allen Ancient DNA Resource, and recently published samples from Iron Age Britain, detecting multiple instances of recent effective population size variation across these groups.

The increasing availability of high-quality genomic data for both modern and ancient samples is creating exciting new opportunities for data-driven investigation of key evolutionary parameters. Among these, the effective size of a population plays an essential role in population biology¹. A population's effective size is defined as the number of individuals in an idealized evolutionary model^{2,3}, and the ability to infer it from genomic data has a wide range of applications, including the study of past demographic events^{4,5} and cultural practices⁶, the quantification of the effectiveness of natural selection^{1,7}, and the prediction of viability in conservation biology⁸.

Several statistical tools have been developed to reconstruct the trajectory of effective population size from genomic data⁹, each

leveraging different genomic features and enabling the analysis of different data types. Methods that rely on the site frequency spectrum (SFS) of a sample^{10–13} avoid modeling recombination and are thus scalable, but require high-quality sequencing data to estimate the SFS and have been observed to be statistically inefficient¹⁴. Methods that model both mutation and recombination processes^{15–19}, on the other hand, tend to scale to smaller sample sizes and require high-quality genome sequencing data. Recent approaches enable simultaneous modeling of recombination and allele frequencies in unphased sequencing data¹⁸, or scaling to larger sample sizes for accurately phased sequencing data²⁰ and for unphased low-coverage data^{21,22}. Finally, several methods that focus on capturing the signature of

¹Department of Statistics, University of Oxford, Oxford, UK. ²Department of Genetics, Harvard Medical School, Boston, MA, USA. ³Broad Institute of Harvard and MIT, Cambridge, MA, USA. ⁴Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. ⁵Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA. ⁶Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. ⁷These authors jointly supervised this work: David Reich, Pier Francesco Palamara. ✉ e-mail: palamara@stats.ox.ac.uk

recombination through the sharing of identical-by-descent (IBD) haplotypes^{23–27} or linkage disequilibrium^{28–31} (LD) have been developed.

Inference of recent population size fluctuations is particularly appealing because it provides unique insights into demographic and evolutionary processes that are specific to the analyzed population. IBD-based methods have been used to infer recent demographic history^{23–25,27} in SNP array and sequencing data. A key limitation of these methods is that they rely on accurate detection of IBD regions^{32–35}. The performance of these algorithms depends on accurate long-range computational phasing, which may be hard to obtain, particularly in low-coverage ancient DNA data. While being a less direct measure of the signature of past recombination events, LD-based summary statistics can be computed in unphased samples, including SNP array and ancient DNA data. LD has been extensively modeled^{36–40} and applied to infer effective population size^{28–31,40,41}. The most recent methods for IBD- and LD-based inference, IBDNe²⁷ and GONE³¹, enable inference of population size fluctuations in time, without assuming a strictly parametrized demographic model. This strategy, however, poses additional challenges, due to the need to adequately regularize the inferred models^{25,27} to avoid reporting spurious fluctuations, while preserving manageable computational costs.

Here, we present a new method, called HapNe, that enables flexible inference of recent effective population size fluctuations using IBD or LD summary statistics, and can be used to analyze both phased and unphased SNP array or sequencing data, including low coverage or ancient DNA data with heterogeneous sampling time. Using extensive coalescent simulations, we show that HapNe accurately and efficiently infers recent demographic history, while regularizing the model to control for spurious oscillations in recent generations. We apply HapNe to reconstruct recent demographic history in both modern and ancient data, including populations from the 1000 Genomes Project and different postcodes from the U.K. Biobank data set, where we observed a bottleneck in the Late Middle Ages corresponding to the period of the Black Death. We also analyze ancient individuals from the Caribbean, Scandinavian Vikings, and individuals who lived in England during the Iron Age, observing isolation and expansion events that are consistent with past historical events, such as the transition from the Archaic to the Ceramic periods in the Caribbean.

Results

Overview of the HapNe algorithm

The HapNe algorithm infers recent effective population size using either IBD or LD data (see Methods and Supplementary Note for a detailed description of the algorithm). We refer to these two approaches as HapNe-IBD and HapNe-LD, respectively. HapNe-IBD uses IBD-sharing information to compute summary statistics related to the count of IBD segments of different lengths.

Briefly (see Supplementary Note for a detailed derivation, also see refs. 23–25,27) this approach leverages the relationship between the length and the age of IBD segments to infer the effective number of individuals that lived in a population in the past. More in detail, two individuals sharing a common ancestor may co-inherit IBD regions, which tend to be larger when the shared ancestor is recent because fewer recombination events occur in the genealogical lineages connecting these individuals. The distribution of the lengths of IBD segments thus provides information about the age of shared ancestors in a population as well as the density of common ancestors across time, which is proportional to effective population size. HapNe-IBD models these relationships using a composite likelihood that links the observed distribution of IBD segment lengths with a population's effective size trajectory.

A limitation of IBD-based inference of effective population size trajectories, however, is that it relies on accurate detection of IBD segments. This typically requires phasing information and additional modeling of haplotype sharing to differentiate between identical-by-

state (IBS) and truly IBD regions. Accurate phasing and haplotype modeling may not be possible if the analyzed genomes are not of high quality or not well represented in reference panels. HapNe-LD, on the other hand, leverages summary statistics related to long-range LD (Pearson correlation between sites). These long-range correlations arise because the shared ancestors transmitting long IBD segments are typically more recent than genomic variations that are found at high frequency in the population, so that these ancestors are themselves carriers of these variants. Therefore, long-range LD is driven by the underlying presence of IBD segments, capturing demographic information used in IBD-based inference. HapNe-LD relies on a composite likelihood that models the expected distribution of IBD segment lengths and the genomic correlations they induce. These LD statistics are easy to compute and do not require genotypes to be either phased or of high quality, enabling the analysis of past demographic events in low coverage or aDNA data.

HapNe-IBD and HapNe-LD both optimize a composite likelihood. To ensure that the model is appropriately regularized, HapNe utilizes a prior on the effective population size $N_e(t)$ that favors models with minimal population size fluctuations. When the analyzed IBD or LD data does not contain sufficient signal, this regularization mechanism prevents inferring spurious variation in $N_e(t)$, which may be incorrectly interpreted as past demographic events. The resulting approximate posterior is optimized to compute a maximum-a-posteriori (MAP) estimator of $N_e(t)$ and bootstrap resampling is used to provide estimates of uncertainty through approximate 95% confidence intervals. Both methods automatically exclude genomic regions harboring unusually large amounts of IBD or LD, which may be caused by natural selection or the presence of structural variation rather than past demographic events. In addition, HapNe-LD implements a test to detect the presence of possible biases due to the presence of strong LD caused by population structure and can handle samples originating from different time points. The HapNe program is freely available as an open-source software package (see Code Availability).

Performance on simulated modern data

We used extensive coalescent simulations to benchmark HapNe-IBD and HapNe-LD against other recent methods for haplotype-based inference of recent effective population size. To this end, we considered several demographic scenarios (Fig. 1a, dotted black lines), including: a constant population size of $N_e(t) = 20,000$; an exponentially expanding population with 200,000 haploid individuals at $t = 0$ and 20,000 at $t = 50$ generations; an exponentially collapsing population with 2000 living individuals at $t = 0$ and 20,000 at $t = 100$; and a population undergoing a strong bottleneck, evolving from 200,000 haploid individuals at $t = 0$ to 2000 at $t = 25$, and then growing back to 20,000 at $t = 50$. For each of these populations, we simulated 256 diploid individuals. We generated realistic SNP-array data and used the simulated ancestral recombination graph to extract ground truth IBD segments longer than 1cM (see Methods).

We initially considered the performance of HapNe-IBD and IBDNe³³ in an idealized setting where ground truth IBD sharing information is available (see Supplementary Fig. 1). In this scenario, HapNe-IBD generally produced lower error than IBDNe, measured using the root mean squared log-error (RMSLE) over the past 50 generations (see Methods). HapNe-IBD produced stable estimates of effective population size in the very recent past, whereas IBDNe tended to output spurious oscillations, a caveat that was highlighted by the authors³³. We next inferred and analyzed LD summary statistics from the simulated array data using HapNe-LD. Because the LD signal reflects the presence of underlying IBD segments (see Supplementary Note), analysis of ground truth IBD data may be seen as an upper bound on the accuracy of HapNe-LD. We observed the RMSLE of HapNe-LD applied to SNP array data to be close to that of HapNe-IBD using ground truth IBD data, suggesting that HapNe-LD achieves close

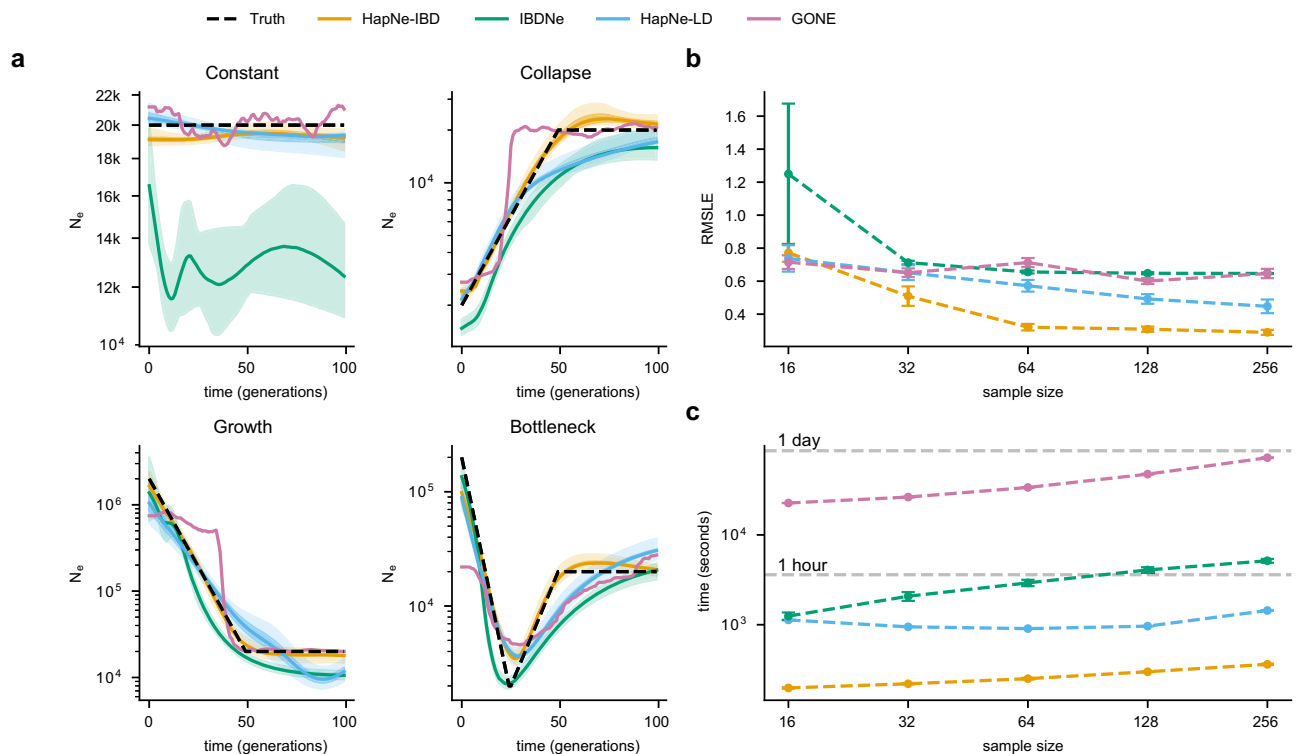


Fig. 1 | Benchmarks in simulated modern populations. **a** N_e estimates obtained from HapNe-IBD, IBDNe, HapNe-LD, and GONE on simulated SNP-array data (256 individuals) for four different demographic scenarios. The light and dark-shaded areas correspond to 95% and 50% confidence intervals estimated using bootstrap quantiles. **b** Accuracy of the different methods on the “Bottleneck” demographic

model as a function of sample size. **c** Total running time for each method (including IBD segment detection and within-chromosome LD estimation, see Methods). In **(b and c)**, we report the mean value across ten independent simulations as well as error bars representing $1.96 \times \text{s.e.m.}$

to optimal performance in these simulations, despite not utilizing phasing information (see Supplementary Fig. 1b). We evaluated the robustness of HapNe-LD under a misspecified recombination map and did not observe strong biases in simulations involving noisy recombination rate estimates (Supplementary Fig. 2), or when using different genetic maps in real analyses (Supplementary Fig. 3). We also tested the performance of GONE³¹, a recent LD-based method, and observed larger RMSLE in the past 50 generations (see Fig. 1b). Due to its regularization procedure, HapNe-LD tended to infer smooth changes in population size, whereas GONE inferred more rapid fluctuations (see Fig. 1a). GONE did not produce bootstrap confidence intervals in these simulations, due to an insufficient number of available SNPs (see Methods).

We next considered a more realistic scenario for the application of IBD-based methods (HapNe-IBD and IBDNe), where we inferred IBD sharing from simulated SNP array data (assuming perfect phasing, see Methods). We detected IBD sharing using the HapIBD software³⁵ (see Methods); similar results were obtained by using other IBD detection methods, as well as input parameters for IBDNe (see Supplementary Fig. 4, Methods). Fig. 1a shows the output of all four methods on a data set of 256 diploid samples and results for other sample sizes are summarized in Fig. 1b (also see Supplementary Figs. 5 and 6, as well as Supplementary Fig. 7 for simulations involving larger sample sizes).

In most cases, the noise introduced by inferring IBD from the data resulted in biases in the inferred effective population sizes; IBDNe tended to underestimate recent effective population size, while HapNe-IBD tended to overestimate ancestral population size (Supplementary Fig. 5). We observed the error in IBD detection to be dependent on several factors, including demographic history, the length of the inferred segments, the software used, as well as the IBD postprocessing strategy. (see Supplementary Fig. 8).

We finally benchmarked the computational speed of these methods and observed HapNe-IBD and HapNe-LD to be more computationally efficient than IBDNe and GONE (see Fig. 1c). Computing LD scales only linearly with the number of analyzed samples, while detecting pairwise IBD sharing requires computation that is quadratic in the number of samples, making LD-based analyses more scalable. Unlike IBDNe, which requires more time to fit larger samples, HapNe-IBD only computes a fixed-size vector of the IBD segment lengths, significantly reducing computational costs for larger samples. The difference in computational time between HapNe-IBD and HapNe-LD is mainly driven by differences in the time required to compute IBD and LD summary statistics.

We next assessed the robustness of HapNe to the presence of haplotype phasing errors, genotyping errors, and population structure, which are often encountered in analyses of real data (see Methods). We observed HapNe-LD, which relies on unphased two-locus statistics, to be more robust than HapNe-IBD to the presence of phasing and genotyping errors. In these simulations, computational phasing reduced the sensitivity of IBD detection, resulting in an upward shift of the demographic models inferred by both HapNe-IBD and IBDNe. HapNe-LD, on the other hand, remained unaffected (Fig. 2a). We observed similar effects when genotyping errors were included (Fig. 2b, c). HapNe-LD was robust to the presence of errors, which however, caused IBD segments to break into smaller regions that led HapNe-IBD to infer spurious oscillations (see Fig. 2b, c and Supplementary Fig. 8). Finally, we investigated the impact of LD-induced by admixture (admixture LD), simulating a scenario where a population originating from a recent admixture event involving two diverged ancestral groups (see Methods, Fig. 2d). HapNe-LD estimates and removes the effects of cross-chromosome LD (see Methods), which partially corrects for the presence of population structure (Supplementary Fig. 9). This, however, does not fully account for

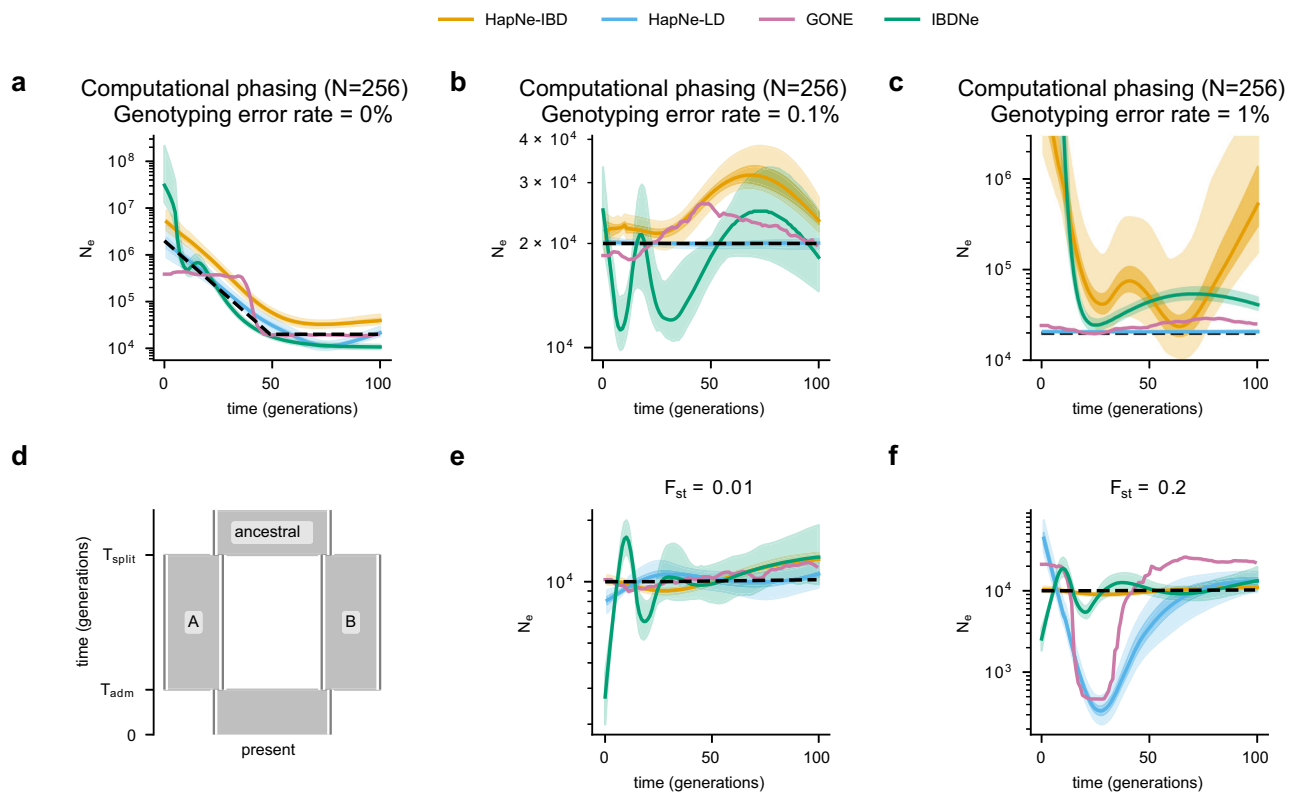


Fig. 2 | Effects of phasing, genotyping error, and population structure on population size inference. These show inference results for computationally phased genotypes from 256 simulated diploid samples N , using genotyping error rates of 0% in (a), 0.1% in (b), and 1% in (c). LD-based inference remained more robust across different demographic models and error rates. Inferred models for IBD-based methods were shifted up compared to simulations with no error (see Fig. 1). **d** This shows a demographic model used in simulations involving recent

admixture. The dataset contains 100 diploid individuals sampled from a population that originates from an admixture event at time T_{adm} between two populations separating at time T_{split} (see Methods). The global effective population size (defined in Supplementary Note, Section 1.1.8) is shown using dashed black lines. Inference results for $T_{adm} = 25$ and different values of the fixation index F_{st} between populations A and B is shown in (e) and (f). The light and dark-shaded areas correspond to 95% and 50% confidence intervals estimated using bootstrap quantiles.

admixture LD, which decays with genetic distance and can lead LD-based methods to infer a spurious bottleneck around the time of the admixture event (Supplementary Fig. 10). We observed the strength of this bottleneck to be proportional to the degree of differentiation between the ancestral populations. The inferred models, however, were not substantially biased for admixture events involving populations with a fixation index (F_{st}) below 0.02 (Supplementary Fig. 11), roughly corresponding to the highest F_{st} observed between European populations (0.023 between Southern Italy and Finland Kuusamo), but lower than F_{st} observed across other groups (e.g., 0.192 between Yoruba and Japan)⁴². We, therefore, caution that HapNe-LD results may be biased in analyses of populations that experienced recent admixture events involving groups for which high F_{st} values are observed.

Overall, HapNe-IBD and HapNe-LD provided improved accuracy and substantially reduced computational times compared to existing methodologies. Although IBD-based inference of effective population sizes is potentially more accurate than LD-based analysis, the need to accurately detect IBD sharing is likely to introduce substantial biases in the inferred population sizes. HapNe-LD's performance was observed to be close to that of IBD-based methods applied to ground truth IBD data and may be applied in the analysis of large sample sizes, providing several practical advantages over IBD-based methods in the analysis of real data sets.

Performance on simulated aDNA data

HapNe-LD does not require phased or high coverage data, making it especially suitable for the analysis of effective population sizes of ancient populations, where phase determination can be poor.

However, analyses of aDNA data suffer from several limitations. First, analyses based on aDNA data sets tend to contain fewer samples sequenced at relatively low coverage compared with modern panels. Furthermore, different sequencing strategies balancing sample size and coverage might lead to different performances in effective population size inferences. Finally, individuals sampled at a site are unlikely to have lived at the same time, with a few notable exceptions^{43,44}. If not modeled, this source of time heterogeneity may lead to biased effective size estimates. We set out to test HapNe-LD's robustness to these sources of confounding. We first created synthetic aDNA samples by generating pseudo-haploid individuals with different levels of missingness m , mimicking the effects of reduced sequencing coverage C , with $m \approx e^{-C}$ (see Methods). We tested the relative impact of the simulated sample size s and coverage on HapNe-LD's inference accuracy (see Fig. 3a and Supplementary Fig. 12 for additional demographic scenarios). As expected, RMSLE decreases when more samples are available and when coverage increases (see Fig. 3b and Supplementary Fig. 13). We then tested whether HapNe-LD would perform better when analyzing a larger number of low-coverage samples rather than a smaller number of high-coverage samples. To this end, we performed simulations where the overall number of sequencing reads is kept approximately constant, while the number of analyzed samples and their coverage are varied (see Fig. 3c and Supplementary Fig. 13). We considered an analysis involving 256 individuals and observed that reducing coverage from 30 \times to 1.4 \times had no significant impact on the performance while requiring only about 5% of the reads. Using an equivalent number of reads to perform high coverage (30 \times) sequencing would only allow sequencing 16 individuals, resulting in

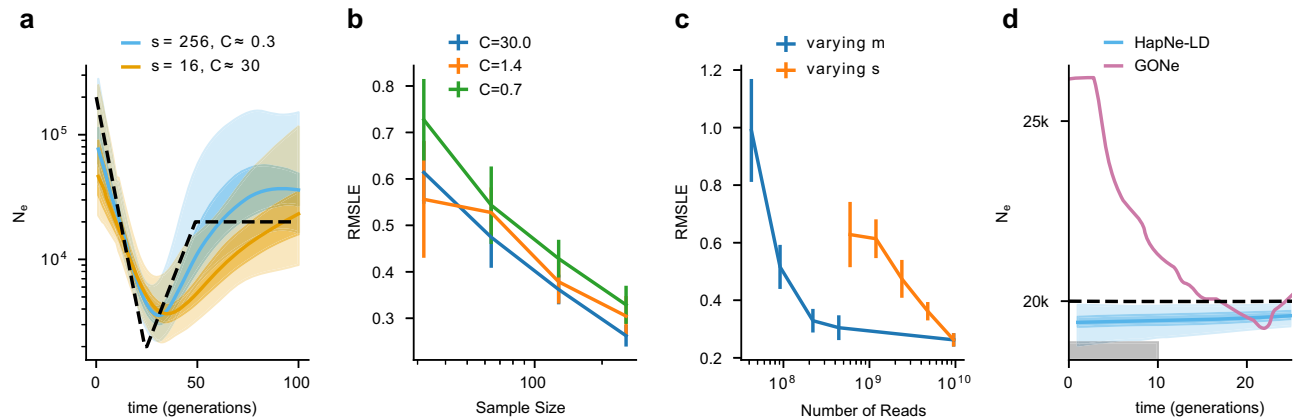


Fig. 3 | Results in simulated aDNA data. **a** HapNe-LD inference results for simulated aDNA-like data under the “Bottleneck” demographic scenario (dashed lines) where the number s of simulated samples and fraction m of missing SNPs, or equivalently the coverage C , are varied (see Methods). **b** RMSLE over the first 50 generations for different coverage levels. **c** Comparison of the accuracy of HapNe-LD based on two sequencing strategies. The orange line reports RMSLE for high coverage data ($m = 0$, $C = 30$) with varying sample size s . The blue line reports

RMSLE for fixed $s = 256$ and varying coverage. **d** HapNe-LD and GONE inference results for a simulation where individuals from a population of constant size of $N_e = 20,000$ are uniformly sampled over an interval $\Delta T = 10$ generations (gray shaded area). In panels **a** and **d**, the light and dark-shaded areas correspond to 95% and 50% confidence intervals estimated using bootstrap quantiles. In (**b**, **c**), we report the mean value across ten independent simulations as well as error bars representing $1.96 \times \text{s.e.m.}$

significantly higher RMSLE. These results suggest that sequencing at a coverage higher than 1–2 \times does not lead to significant improvements in HapNe-LD’s performance, and that HapNe-LD is more accurate when a larger number of individuals is sequenced at lower coverage compared to settings in which a smaller number of high coverage samples is analyzed, because more independent IBD segments are present in the latter case. At low sample sizes ($s < 10$), only demographic histories with a strong contraction in the recent past yielded enough signal for HapNe to infer fluctuations in our experiments (see Supplementary Fig. 14). We also examined the impact of imputing aDNA using a modern reference panel (see Methods). In these experiments, HapNe-LD obtained improved accuracy of the inferred demographic history, despite a relatively high error rate of 2.9% in the imputed aDNA genotypes (see Supplementary Fig. 15). This suggests that genotype imputation of aDNA samples may be a viable route to improve demographic inference, although additional care may be required to consider and address potential biases introduced by imputation. In these experiments, we observed that imputation induces a significant level of cross-chromosome LD, which will be flagged in HapNe’s output.

Lastly, we considered potential biases arising due to heterogeneous sampling times of the analyzed aDNA individuals. We used analytical modeling (see Methods and Supplementary Note) to confirm that, if not accounted for, heterogeneous sampling times lead to biased recent effective population size estimates. We performed simulations of aDNA samples originating from heterogeneous time locations under a constant demographic history, uniformly drawing the time offset of each sample between 0 and ΔT generations in the past (see Methods). In this setting, we observed that using GONE to infer effective population size leads to the spurious inference of a recent population expansion, consistent with analytical predictions under unmodeled time heterogeneity (see Fig. 3d). The HapNe-LD algorithm allows utilizing prior knowledge of sampling times (e.g., from radiocarbon dating or archeological context) in the form of a user-provided time interval for each analyzed individual (see Methods). Using simulations, we verified that this approach effectively removes recent biases due to time heterogeneity.

Inference of recent effective population sizes in the UK Biobank and 1000 Genomes Project data sets

We used HapNe-IBD and HapNe-LD to analyze recent effective population size variation using genotype data from the UK Biobank data set

(see Methods). Accurate inference of recent demographic events requires a combination of large sample sizes and small effective population sizes, which make it possible to estimate recent coalescent rates. In this case, large recent effective population sizes generally present across the UK are balanced by the large sample sizes available in the UK Biobank data set. In order to mitigate the impact of admixture LD, we focused on the larger group of samples with self-reported white British ancestry, and only considered unrelated individuals to avoid biasing demographic inference in recent generations. We grouped individuals based on the postcode of their self-reported birthplace and report analyses for three of these postcodes (see Fig. 4a, Methods). We also used FastSMC to detect IBD segments within each of these postcodes. Regions with unusually high LD or IBD sharing were excluded using HapNe’s filter (Supplementary Figs. 16 and 17).

Effective size trajectories inferred from these regions in the UK all exhibit a bottleneck event during the Late Middle Ages, which roughly corresponds to the period of the Black Death (Fig. 3a, vertical dashed line). The inferred population size for individuals from the Llandudno postcode has a significantly smaller effective population size compared to the ones inferred for Glasgow and Edinburgh. Such a smaller effective size offers a stronger source of recent demographic signal, allowing to perform inference using a smaller sample size ($s = 2089$ for Llandudno, $s = 14,724$ for Glasgow, and $s = 9981$ for Edinburgh). In contrast, detecting the more subtle contraction to a larger minimum bottleneck size in Glasgow required a substantially larger sample size, as highlighted when we downsampled data from this postcode to 2000 individuals (see Supplementary Fig. 18). In this experiment, the bottleneck was only apparent in the output of HapNe-IBD, suggesting that LD-based analysis may lead to comparably lower statistical efficiency in cases where high-quality IBD signal is available. Demographic models inferred by HapNe-IBD and HapNe-LD are broadly consistent, although HapNe-IBD tends to report a larger effective population size, with a significant shift towards more recent times. These observations are compatible with the presence of underlying IBD segments that are undetected or broken into smaller segments, due to the presence of phasing or genotyping errors in the data.

We next applied HapNe-IBD and HapNe-LD to data from the 1000 Genomes Project (1 kGP,⁴⁶). Unlike the UK Biobank, most 1 kGP groups contain a small number of samples, which originate from large populations. Furthermore, several groups represented in the 1 kGP data set

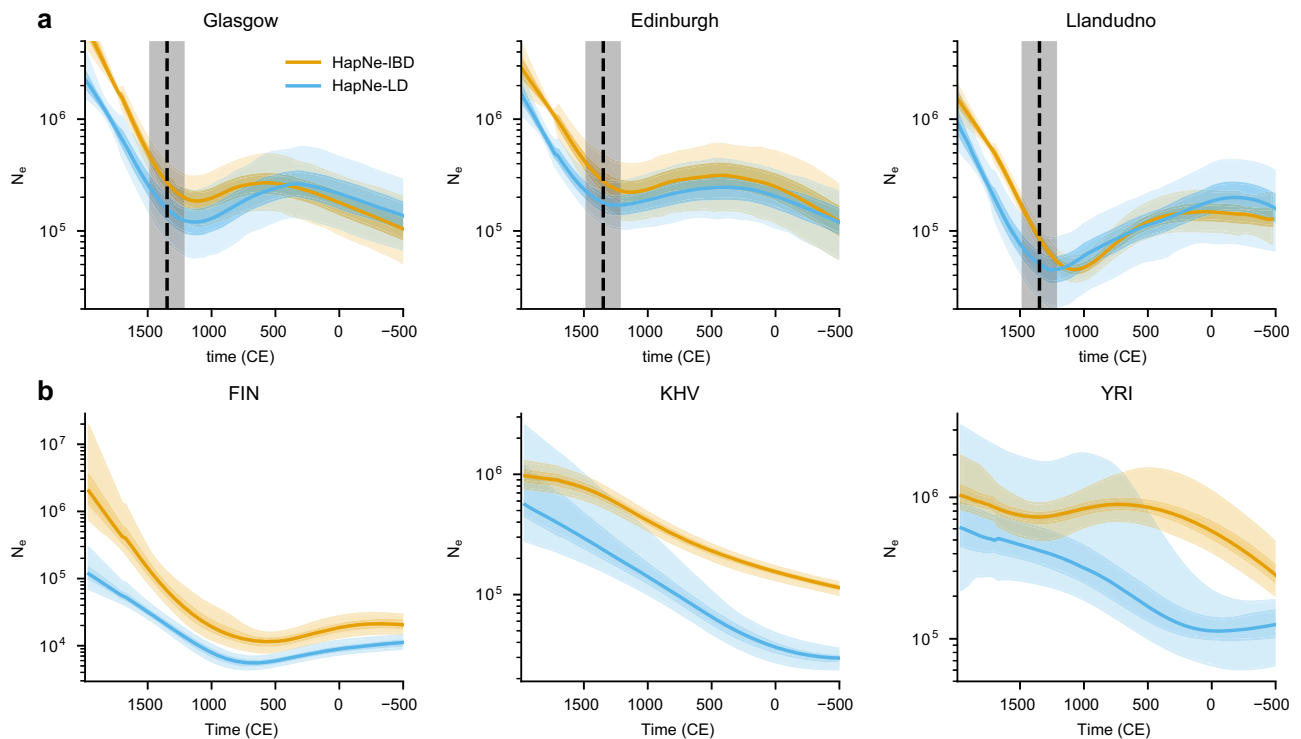


Fig. 4 | HapNe-IBD and HapNe-LD estimates of recent effective population sizes in modern populations. **a** Inference results for three postcodes: Glasgow (G), $s = 14,724$; Edinburgh (EH), $s = 9981$; and Llandudno (LL), $s = 2089$ from the UK Biobank data set. The vertical dashed line corresponds to the estimated date of the Black Death in the UK (1348, ref. 45). HapNe results are converted to years assuming 29 years per generation. The shaded gray area depicts how the placement of the

Black Death would shift with respect to the inferred demographic models if values between 23 and 35 years per generation were assumed. **b** Inference results for three populations (Finnish, European, FIN, $s = 99$; Kinh in Ho Chi Minh City, Vietnam, South Asian, KHV, $s = 99$; Yoruba in Ibadan, Nigeria, African, YRI, $s = 108$) from the 1000 Genomes Project. The light and dark-shaded areas correspond to 95% and 50% confidence intervals estimated using bootstrap quantiles.

are known to have undergone recent admixture⁴⁶. As shown in simulations, HapNe-LD is not suited for the analysis of groups with a history of recent admixture involving groups with large F_{st} and we, therefore, only retained a subset of populations for subsequent analysis (see Methods). We next used HapNe-LD to compute LD for each population and estimated recent IBD sharing using the HapIBD algorithm followed by a post-processing step to merge detected segments (see Methods). We then inferred recent effective population sizes using the HapNe-LD and HapNe-IBD methods and reported results for the five populations with no significant levels of CCLD and for which both methods found sufficient signal to infer demographic variation.

Figure 4b shows results for three populations meeting these criteria. Results for the other populations retained for this analysis are shown in Supplementary Fig. 19, which also includes results obtained using IBDNe. IBD-based inference consistently resulted in larger inferred effective population sizes compared to LD-based inference. As observed in simulations, this is likely caused by reduced sensitivity in the detection of IBD segments in real data, due to the presence of phasing and genotyping errors. This effect was more pronounced in 1kGP compared to the UK Biobank data set, where a larger sample size leads to higher accuracy in phasing and IBD detection. HapNe-LD infers a bottleneck at 1000 CE for the FIN population, consistent with previous reports^{27,31,47}. This demographic event is inferred to have an earlier onset using IBD data, likely also a result of noisy IBD detection. Both HapNe methods suggest a recent expansion for the individuals from the Kinh population in Ho Chi Minh City, Vietnam (KHV) and for the Yoruba population in Ibadan, Nigeria (YRI). Some biases observed in our simulations are also present in the inferred demographic histories of other populations. These include population collapses in the demographic history inferred for populations with significant levels

of CCLD and known recent admixture between diverged ancestral groups (Supplementary Fig. 20).

Overall, these results suggest that HapNe-LD and HapNe-IBD provide similar results when large samples and high-quality IBD data are available and suggest guidelines for the application of these tools. HapNe-LD provides more robust results than HapNe-IBD in data sets where phasing and IBD detection accuracy are reduced, at the cost of slightly reduced statistical efficiency, particularly when the presence of population structure requires the estimating cross-chromosome LD (see Supplementary Note). LD induced by recent admixture involving populations with large F_{st} can lead to the inference of spurious population reductions in the recent past (see Supplementary Fig. 20).

Inference of recent demographic history in ancient populations

We applied the HapNe-LD method to aDNA sampled from four different sites for which large cohorts from similar time strata were available (see Methods and Supplementary Tables 1–8). We excluded individuals for which known close relatives were present in the data set (see Methods), which would otherwise lead to a smaller inferred effective population size. It is still possible that undetected distant relationships remain present in these groups; if present, our analyses interpret these more distant relationships as reflecting demographic phenomena, rather than sample ascertainment. We first analyzed a group of recently published individuals excavated in Pocklington, Yorkshire, UK⁴⁸ (see Fig. 5a). The archaeological context suggests that this group belongs to the Arras culture, which is distinctive relative to other Iron Age cultures in the UK but shows similarities with contemporary cultures in the Paris Basin and Ardennes/Champagne regions of France. These individuals were found to be unusually highly drifted from nearby groups, although their F -statistics do not highlight significantly divergent admixture histories. This suggests that these

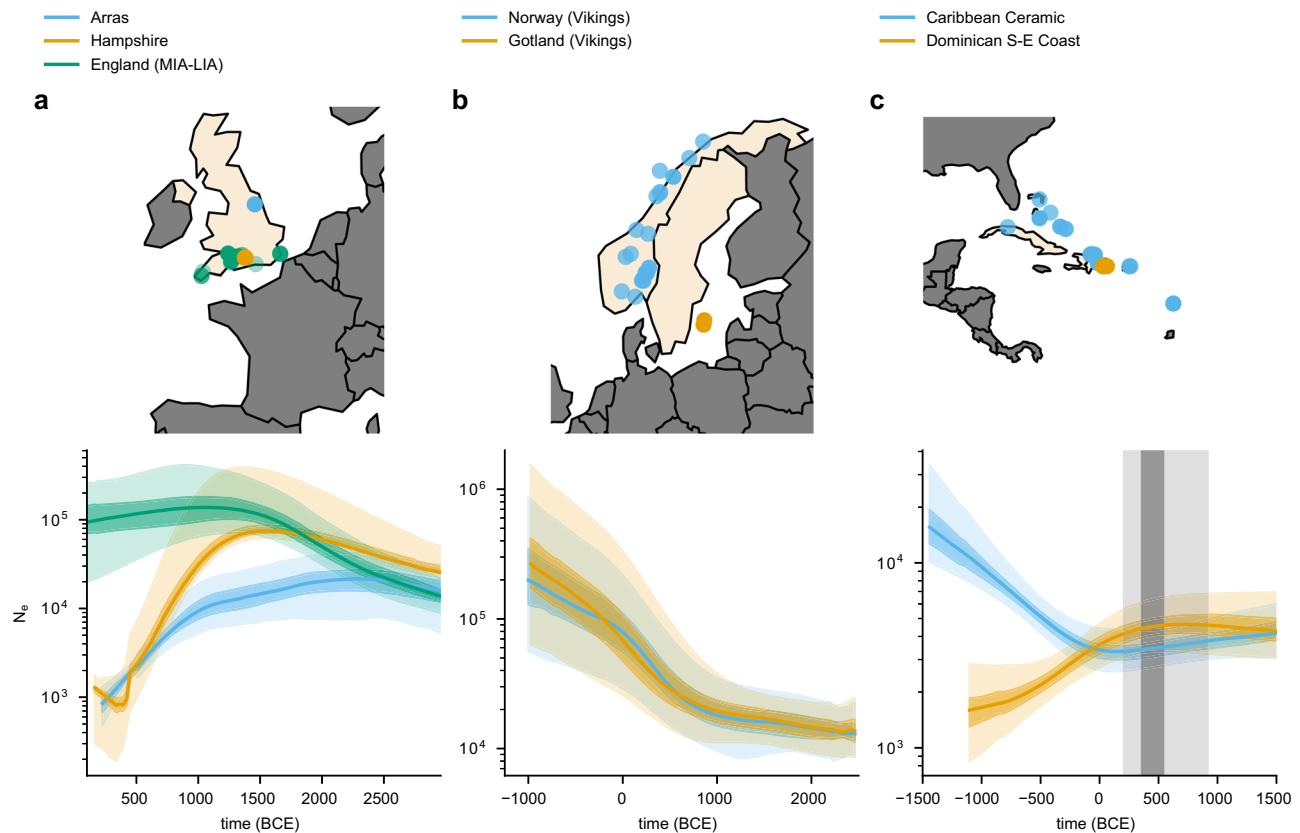


Fig. 5 | HapNe-LD estimates of recent effective population sizes in ancient populations. **a** Analysis of 49 Middle to Late Iron Age individuals from South England, compared to a subset of 14 individuals from Hampshire, and to 24 individuals related to the Arras culture near Yorkshire. **b** Inference based on 22 Viking samples found in modern Norway (blue) and 28 found in Gotland, a Swedish island (red). **c** Effective population size inference based on 71 unrelated individuals from the Caribbean Ceramic clade and 18 from the Dominican South-East coast subclade. The dark-gray shaded area corresponds to the estimated date for the transition

from the Archaic to Ceramic culture in the region. The light and dark-colored shaded areas correspond to 95% and 50% confidence intervals estimated using bootstrap quantiles. The light gray-shaded area depicts how the placement of this transition would shift with respect to the inferred demographic models if values between 25 and 35 years per generation were assumed. The dots on the maps represent the location of the samples. The figure was made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.

groups share common origins but may have been isolated for some time or that they originated from a later migration event⁴⁸. This hypothesis is compatible with the effective population sizes we observed when running HapNe on 24 individuals from the Arras culture, which we compared with 49 individuals from South England (Supplementary Tables 2–4). Because the Arras are sampled from a smaller geographic region compared to the samples from South England, we also considered a subset of 14 individuals from South England sampled from a similarly localized region in Hampshire.

Both the Arras and the Hampshire groups displayed a significant recent population contraction, probably reflecting geographic localization. However, the demographic model inferred for the Hampshire samples grows back to the population size inferred for South England, whereas the size observed for the Arras samples remains smaller. This difference in population sizes is consistent with the recent observation of high genetic drift between the Arras and other groups in these regions, possibly reflecting isolation or distinct origins⁴⁸. Cross-chromosome LD for these groups was found to be negligible, suggesting that the observed demographic signature is not due to population structure (see Supplementary Table 1), although recent admixture might create a similar collapse (see Supplementary Fig. 11). The small population size of the Arras group might also explain why this population was found to be unusually highly drifted from nearby groups. The recent effective population size inferred for individuals in the South of England was compatible with population size estimates obtained for modern UK Biobank individuals, although confidence

intervals were large over the first 1000 years due to a reduced sample size.

We next analyzed 22 genetically similar individuals from the Viking Age buried in Norway, together with 28 individuals from the south-east Swedish island of Gotland⁴³ (Fig. 5b and Supplementary Tables 5 and 6). Norwegian and Swedish Vikings have been observed to have a slightly smaller proportion of ancestry from Neolithic farmers from Anatolia compared to Vikings from Gotland. On the other hand, Vikings from Gotland have a relatively higher estimated fraction of ancestry shared with Bronze Age individuals from the Baltic region. Despite these differences, the demographic histories inferred by HapNe-LD for the recent past of these individuals substantially overlap, and both trajectories show a significant expansion during the iron age (–500 to 800 CE).

Finally, we focused on 71 unrelated individuals from the Caribbean, first analyzed in ref. 49 ($n = 62$) and ref. 50 ($n = 9$) spanning –1149 to –1440 CE (Supplementary Tables 7 and 8). For these samples, HapNe-LD infers a weak sign of a bottleneck occurring around 1 CE, followed by a significant expansion, as shown in Fig. 5c (blue line). This pattern may reflect the transition from the Archaic to Ceramic context about 2500–2300 years ago (Fig. 5a, gray area), which has been associated with migration events in the region⁴⁹. We also extracted and separately analyzed a subgroup of individuals from South-East Dominican sites (Fig. 5c, red). These individuals are part of a subclade previously identified in ref. 49. The population size inferred for this group matches that of the broader Caribbean group in the deep

past, consistent with common origins, but shows a distinctive sign of contraction in the more recent past. LD induced by population structure is detectable in these individuals, which may partially explain the observed contraction (Supplementary Table 1). Nevertheless, the sizes inferred by HapNe-LD in the recent past roughly match those inferred using runs of homozygosity⁵¹, supporting the possibility of a population contraction starting after the transition from the Archaic to the Ceramic period⁴⁹. As in the case of the Arras and Southern England individuals, these demographic patterns may also be due to isolation by distance, where samples originating from different islands result in a larger effective size when considered together.

Discussion

We developed an algorithm, called HapNe, that leverages the count of IBD segments of different lengths (HapNe-IBD) or long-range LD (HapNe-LD) to infer recent effective population size fluctuations in modern or ancient DNA data. HapNe-IBD and HapNe-LD implement a number of preprocessing steps, as well as tests to verify that sufficient recent demographic signal is present in the data and to detect the presence of admixture LD. Both methods minimize a power-likelihood based on an analytic link between observed summary statistics and the effective population size and use regularization to avoid producing spurious oscillations. We used extensive simulation to show that both HapNe methods were more accurate and computationally faster than available algorithms for IBD-based and LD-based inference of recent demographic history, producing lower error and fewer spurious oscillations. These simulations also showed that while HapNe-LD does not require high-quality or phased data and scales better with sample size, its performance can be close to that of IBD-based methods applied to ground truth IBD information. Finally, we applied HapNe to several modern and aDNA data sets, detecting evidence for recent past demographic events across these populations. These include population size contractions corresponding to the period of the Black Death in different regions of the UK, as well as bottleneck and expansion events in 1000 Genome Project populations. In aDNA data, these analyses provided evidence for divergence and isolation events, as well as shared demographic histories in subgroups from several ancient populations with diverse geographic and temporal origins.

Our analyses suggest that LD-based inference of recent demographic variation provides a route to circumventing biases that may arise in IBD-based demographic inference. Although the spectrum of shared IBD haplotypes is an effective source of information for analyses of past demographic events, accurately estimating IBD sharing is complicated in low coverage and aDNA data and may lead to biased results. This may also be the case in modern populations when limited data availability prevents accurate phase estimation. Although summary statistics of LD rely on less direct observation of historical recombination events, they may be effectively computed in unphased and low-coverage data sets. This enables analyzing recent demographic events in samples from poorly represented populations and, coupled with modeling of heterogeneous sampling time, in aDNA data sets. Performing both IBD-based and LD-based analyses may offer validation for an inferred demographic model and allow testing for the presence of biases in either approach. An additional source of potential bias in methods for demographic inference is linked to the need to make assumptions about the type of demographic model being inferred. In this context, approaches that avoid relying on a predefined set of models provide more flexibility, but require further tuning strategies to balance the desired sensitivity to past demographic events with the need to prevent the inference of spurious fluctuations. Our work suggests that the use of self-tuning regularization mechanisms helps mitigate the risk of spurious inferred fluctuations. Finally, our analyses highlight the importance of accurately preprocessing both IBD and LD signals before performing demographic inference, as results may vary significantly if unfiltered data is utilized. Key

preprocessing steps include testing for the presence of admixture LD and systematically filtering out regions of the genome that harbor unusually high IBD sharing or LD (see e.g., Supplementary Fig. 16). These may be due to natural selection or the presence of structural variation and lead to biases in analyses of demographic history and selection if not accounted for.

We outline several limitations and directions of future development for this work. First, HapNe-LD assumes that the LD signal observed in the data is solely due to past population size fluctuations. In some instances, residual admixture LD can be present in the data after filtering, causing a spurious contraction in the recent past and creating the need to carefully interpret models that resemble this type of signature. In general, a contraction in the recent past might be due to a decrease in the census size of the population, isolation, or an artifact of recent admixture. Similarly, HapNe-IBD currently only relies on the observed spectrum of IBD sharing, which may be biased due to inaccurate IBD detection. Future work may allow explicit modeling of type-1 and type-2 errors in IBD detection, mitigating biases in the inferred demographic models. Since the HapNe-LD and HapNe-IBD methods are subject to different sources of bias, their output can be compared to check the validity of their output. Second, while regularization helps prevent the inference of spurious demographic fluctuation, it leads to favoring constant and exponential demographic histories that lack fluctuations if these are not supported by the data. When interpreting demographic models inferred by HapNe, it is important to note that an inferred constant growth rate may reflect insufficient evidence for past demographic variation (see e.g., Supplementary Fig. 18), which causes the model to be more strongly regularized. For the same reason, when the data contain limited demographic information, HapNe may produce tight bootstrap confidence intervals for demographic histories that are closer to constant population size. These should also not be interpreted as strong evidence for a constant size, but as a lack of evidence for past fluctuations. Third, we note that both LD-based and IBD-based analyses rely on preprocessing steps to filter out genomic regions that may otherwise lead to biases. These excluded regions harbor unusually high or low density of LD or IBD signal, which may be caused by non-demographic factors such as natural selection or the underlying presence of structural variation⁵² (see Supplementary Fig. 17). HapNe-LD and HapNe-IBD automatically detect and remove these genomic regions. This is currently achieved by dropping entire chromosome arms but a less conservative and likely more computationally intensive approach may be devised. Fourth, HapNe-LD currently corrects for time heterogeneity by marginalizing out the time difference between the samples included in the analysis. Future extensions may allow HapNe to increase inference accuracy by correcting the LD for each pair of samples separately.

Finally, HapNe-LD makes several model simplifications, including the assumption that the analyzed samples come from a single population. A promising direction for future work is to extend HapNe to allow it to explicitly account for multiple populations and infer coalescent rates across groups, improving the analysis of more complex demographic models such as those involving isolation by distance, divergence, and admixture. Similarly, HapNe-LD is currently focused on the inference of recent demographic history, but may be extended to the analysis of deeper time scales by modeling variation in allele frequencies, which are currently assumed to be constant in time. Despite these limitations, we expect that the HapNe framework developed in this work will offer valuable insights into past demographic events in both modern and ancient DNA data.

Methods

Simulated genetic data

We used the ARGON simulator⁵³ (version 0.1.160415) to generate synthetic genotypes and ground truth IBD data for modern and ancient

populations. Simulations with time heterogeneity were performed using msprime⁵⁴ (version 1.1.1). We simulated genomes of 36.23 Morgans, split into 39 independent regions corresponding to human chromosome arms. We used a mutation rate of $\mu = 1.65 \times 10^{-8}$ and a recombination rate of $\rho = 1 \times 10^{-8}$ per generation per base pair, except in experiments where human genetic maps were used. To simulate SNP data, we then downsampled sequencing data to match the genotype density and allele frequency spectrum observed using Chromosome 2 of the UK Biobank data set, using 50 evenly spaced MAF bins. We generated unphased diploid individuals by randomly pairing simulated haplotypes. Ancient data was generated using a similar procedure, downsampling to 1240k SNP array densities. We also added two additional steps to simulate low-coverage data. We first transformed the data into pseudo-haploid individuals by randomly sampling one haplotype at each site. We then set each site as missing with probability m , related to a simulated coverage parameter C through the relationship $m \approx e^{-C}$, further described below.

In the simulations that incorporate admixture, we employed a demographic history starting with an ancestral population of size $N_e = 10,000$ which splits into two populations, A and B, at time $t = t_{\text{split}}$. Populations A and B then evolve independently, each with an effective population size of $N_e = 5000$. At time $t = t_{\text{adm}}$, the two populations merge into a single population, with an equal contribution. The newly created population has a size of $N_e = 10,000$. It then experiences one of four scenarios (Supplementary Fig. 11). In one scenario, the population remains constant until $t = 0$. In the other scenarios, the population undergoes an exponential expansion with $N_e = 100,000$ at $t = 0$, an exponential collapse with $N_e = 1000$ at $t = 0$, or a bottleneck with N_e exponentially decaying to 1,000 at $t = t_{\text{adm}}/2$ before exponentially growing back to 10,000 at $t = 0$. For convenience, we report the value of F_{st} instead of t_{split} when discussing the results; the link between these two quantities is found described in Supplementary Note Section 1.1.9.

Simulation of missingness and coverage

We simulated low coverage data by discarding a proportion m of the SNPs of each individual, but often report results referring to corresponding sequencing coverage parameters. To this end, we assumed a simple model where a genome of length G is sequenced using N reads of length L . Using this notation, the probability that a randomly selected site along the genome is not spanned by a read is:

$$m = \left(1 - \frac{L}{G}\right)^N = \left(1 - \frac{C}{N}\right)^N \approx e^{-C}, \tag{1}$$

where $C \equiv \frac{NL}{G}$ represents the coverage parameter.

This relation can also be used to obtain a link between m and the number of reads:

$$N = -s \frac{\log(m)}{z}, \tag{2}$$

where $z = -\log(1 - \frac{L}{G}) > 0$ and s is the number of sampled individuals with missingness m .

Imputation of simulated ancient data

To simulate imputed aDNA data, we simulated 32 ancient samples from 30 generations before present as well as 200 modern individuals, using msprime⁵⁴ (v 1.1.1). These individuals were sampled from a demographic model with an ancestral population size of 20,000 haploid individuals, which undergoes an exponential expansion

starting at 80 generations before present. The expansion makes the population reach 2,000,000 individuals at generation 30, after which it maintains a constant size until present day. We downsampled the simulated 32 ancient individuals to match 1240k SNP array mutation and allele frequency densities. We simulated short reads (150bp) at the target coverage on the array SNPs. These reads were written to SAM file format and compressed into BAM file format using samtools (version 1.13). We further simulated a reference fasta file using msprime reference allele at all polymorphic positions and then aligned the reads to the reference sequence and called genotype likelihoods using BCFtools (version 1.15.1). We then used GLIMPSE (version 1.1.1) to impute the ancient data, leveraging the called genotype likelihoods, using the 200 modern simulated samples as a reference panel, assuming perfect phasing and 0 genotyping error. This imputation process results in a genotype call being made for every variant that is polymorphic in the reference panel or the ancient samples. We reported imputation r^2 , the squared correlations between the imputed allele dosage and the ground-truth sequencing data for the ancient samples across the polymorphic variants (Supplementary Fig. 15). Imputation quality at coverage of $1\times$ was above 90% for SNPs with MAF above 25%, used in our analyses.

Computation of LD

We consider a panel of s individuals, M sites and genotypes $\tilde{G}_{i,x} \sim \text{Bin}(2, p_x)$ for individual i at site x with minor allele frequency p_x . We first standardize the genotypes by computing $G_{i,x} = \frac{\tilde{G}_{i,x} - 2\hat{p}_x}{\sqrt{2\hat{p}_x(1-\hat{p}_x)}}$, where \hat{p}_x is the estimated allele frequency. The LD between two sites x and y is computed as the R^2 statistic:

$$R_{x,y}^2 = \frac{\left(\sum_{i=1}^s G_{i,x} G_{i,y}\right)^2 - \left(\sum_{i=1}^s G_{i,x}^2 G_{i,y}^2\right)}{s(s-1)}. \tag{3}$$

The computation of this statistic scales linearly with the number of samples ($O(s)$). Note that this estimator is biased due to the use of \hat{p}_x instead of the unknown allele frequency p_x during the normalization step. We describe a procedure used at runtime to debias these estimates in the Supplementary Note. The LD of pseudo-diploid individuals is computed using the same approach, with $\frac{1}{2}\tilde{G}_{i,x} \sim \text{Bin}(1, p_x)$.

Detection of IBD segments

We used FastSMC³⁴ (version 1.2), HapIBD³⁵ and RefinedIBD³³ (version 17jan20) to detect IBD segments in simulated and real data analyses. HapIBD and RefinedIBD were used with recommended parameters for SNP-array data (default parameters). We ran FastSMC using parameters $\min m = 0.5$ (minimum cM length) and $t = 100$ (IBD time threshold). Decoding quantities were generated based on 30 samples using a European demographic history. FastSMC was run using multiple jobs, so that each job considers at most 100 haploid samples. The IBD segments inferred by HapIBD and RefinedIBD were post-processed using the merge-ibd-segments tool (see URLs), using the default parameters. We observed that this post-processing step improves the accuracy of the inferred IBD segment length distribution when genotyping and sequencing errors are present in the data. In these scenarios, FastSMC may break segments into shorter regions based on their estimated posterior probability. However, a post-processing tool to merge these fragmented regions and improve the accuracy of segment length estimates is not available for FastSMC, so we used HapIBD and RefinedIBD for most analyses, except for the use of FastSMC in the analysis of the UK Biobank data³⁴, which is accurately phased.

HapNe-IBD and HapNe-LD algorithms

We developed two algorithms to infer recent effective population size fluctuations $N_e(t)$ from a set of s samples, called HapNe-IBD and HapNe-LD. Both approaches take summary statistics $\{Y_{i,b}\}$ as input and maximize a pseudo-posterior function for $N_e(t)$. The input data set $\{Y_{i,b}\}$ is split into 39 genomic regions corresponding to chromosome arms indexed by i , using 0.5 cM long bins indexed by b .

HapNe-IBD takes as input a list of IBD segments of length $N \sim O(s^2)$. Input data $\{Y_{i,b}\}$ corresponds to the count of IBD segments in region i whose length lies in bin b . Bins start at 2cM and end at the largest detected IBD segment. We assume that each of these counts is the realization of a Poisson random variable, with demographic-dependent mean parameter $\mu_b(N_e(t))L_i$, where L_i is the length of the i th region ($\mu_b(N_e(t))$ is described in the Supplementary Note). To handle overdispersion, we used a quasi-likelihood approach to compute a weight parameter ϕ_b^2 that multiplies the variance in each bin.

HapNe-LD uses average R^2 statistics as input data $\{Y_{i,b}\}$. This input is computed in $O(sm)$, where m is the total number of loci. We assumed that these observations are realizations of a Normal random variable, with a distance-dependent mean parameter $\mu_b(N_e(t))$ (see Supplementary Note for a detailed description of $\mu_b(N_e(t))$). The variance parameters ϕ_b^2 were estimated using the usual variance estimator within each bin.

Give a set of IBD or LD observations $\{Y_{i,b}\}$ for the i th genomic region and b th bin, HapNe aims to maximize $P(N_e(t)|\{Y_{i,b}\})$ under the following assumptions. First, $N_e(t)$ is a piece-wise exponential function from $t=0$ to $t=t_{max}$ generations, and remains constant afterwards. In all our analyses, we used $t_{max}=125$ generations. The lengths of the time intervals are iteratively tuned so that each time interval contains the same number of expected ancestors of IBD segments (see Supplementary Note). Second, we assume that there exists a prior on the effective population size $p_{N_e}(\theta)$, where θ represents the set of parameters defining $N_e(t)$. A discussion about the choice of this prior can be found in the Supplementary Note. Third, we assume that the total likelihood $P(\{Y_{i,b}\}|N_e)$ can be approximated by a power likelihood^{25,56} and be written as $P(\{Y_{i,b}\}) = \prod_{i,b} P(Y_{i,b})^c$. If we assume that bins on different chromosomes are not correlated, the exponent c captures the correlations between the bins of a region. When there is no correlation in the data, $c=1$ recovers the true likelihood, whereas if all bins are fully correlated, setting $c = \frac{1}{n_{bins}}$ leads to the likelihood of a single observation. We discuss how this hyperparameter is automatically tuned using a heuristic model selection rule in the Supplementary Note.

Once the time intervals and the value of the regularization parameter are fixed, HapNe assesses the uncertainty of the prediction by performing 100 bootstrap iterations. For each iteration, HapNe samples chromosome arms with replacement to create new input data, and estimates the effective population size. The 2.5th, 25th, 75th, and 97.5th percentiles are reported at each generation to obtain 50% and 95% confidence intervals.

Comparisons to other methods

To perform method comparisons, we simulated genotypes based on the demographic models shown in Fig. 1 and used the methodology described above to compute summary statistics. We ran HapNe-IBD, HapNe-LD, IBDNe (version 23Apr20.ae9), and GONE (retrieved Jun 22, 2021). We used default parameters for all methods, except for IBDNe where we set $gmin=1$ in simulated data, as recently recommended⁵⁷. The simulated SNP array data did not contain enough sites to perform the SNP bootstrapping strategy used by GONE to produce confidence intervals in sequencing data. All computations were run on an Intel Skylake 2.6 GHz architecture on the Oxford Biomedical Research Computing cluster.

We reported the root mean squared log-error (RMSLE) over the first 50 generations as a measure of accuracy. If $N_e(t)$ and $\hat{N}_e(t)$ denote

the true and predicted demographic models, the accuracy is defined as:

$$\text{RMSLE} = \sqrt{\frac{1}{50} \sum_{t_i=1}^{50} \left(\log(\hat{N}_e(t_i)) - \log(N_e(t_i)) \right)^2} \quad (4)$$

We performed ten independent sets of simulations and computed error bars reported in each plot as $1.96 \times \text{s.e.m.}$

Filtering of high IBD and LD regions

To mitigate the impact of natural selection and structural variation, HapNe applies a filtering algorithm to exclude chromosome arms with unusual amounts of IBD sharing or LD. For LD data, parameters of a normal distribution are computed for each bin using the median and quantiles of the observed data. We used this quantile-based approach instead of moment-based estimators so that the inference is robust in the presence of the outlier regions we aim to filter out. Then, each genomic region is discarded using the following two heuristic rules. First, the deviation between the observed LD in the region and the median must be within 6 standard deviations. Second, the observed values must cross the median at least once, i.e., a region cannot have all its observations above or below the median. The IBD data is filtered using a similar approach. For each region, the mean of the Poisson distribution and the dispersion factors are computed for each bin using all others regions. The region is discarded if the sum of its squared deviance residuals is in the upper or lower α -quantile of the underlying χ^2 distribution, with $\alpha = 10^{-12}$. The procedure is performed a second time, without considering the discarded regions, to prevent outliers to impact the final result.

LD-based test for population structure

Population structure creates long-range LD between unlinked pair of sites. HapNe allows testing for LD due to population structure by computing cross-chromosome LD (CCLD). In the absence of CCLD, we expect the correlations between two sites x and y located on different chromosomes to be only due to finite sample sizes (see Supplementary Note):

$$\mathbb{E} \left[G_{i,x} G_{i,y} G_{j,x} G_{j,y} - \frac{4}{(N_x - 1)(N_y - 1)} \right] = 0, \quad (5)$$

where N_x and N_y are the number of observed haplotypes on sites x and y , respectively. Because the LD is only computed between pairs of sites containing at least two overlapping observations, N_x and N_y are not independent variables. HapNe-LD computes the empirical mean of Eq. (5) for each pair of chromosomes and then performs a t -test to check for deviation from the 0-mean hypothesis.

Time heterogeneity in the set of analyzed samples

Most aDNA data sets contain samples originating from different time points, with an estimated date range spanning many generations when the archeological context is used to date the samples. We thus extended HapNe-LD to account for time heterogeneity and uncertainty. The user can provide a date range for each sample. This information is used by HapNe to compute the density of the ages of a randomly selected pair of individuals. This density is then used to marginalize out the age of the oldest sample and the generation gap between the two individuals under the SMC approximation, resulting in an unbiased estimator of the effective population size (see Supplementary Note).

Inference of demographic history in the UK Biobank

We analyzed the subset of 305,784 unrelated samples with self-reported White British ancestry, corresponding to the individuals

reported in ref. 58 that did not withdraw from the study and whose birth location can be assigned to a postcode in the U.K. (13,995 were removed because of this last condition). We focused on 727,103 genotyped autosomal variants, which we phased using Beagle 5.1⁵⁹. We then grouped the individuals based on their self-reported birth location, labeling each of them with the first 1 or 2 letters of their corresponding postcode. We randomly picked postcodes with different sample sizes to infer population sizes. LD computations and IBD detection steps were performed using the procedure described above, using genetic maps corresponding to the GRCh37 genome build.

Inference of Demographic history in the 1000 Genomes Project

We downloaded the $N = 2504$ high-coverage sequenced samples from the 1000 Genomes Project data set⁶⁰. The samples were grouped according to 26 population labels. For the IBD detection step, we downsampled the sequenced data to match the density of SNP array variants found in the UK Biobank, using the procedure described above. IBD segments were inferred using HapIBD and merged using the post-processing tool described above. HapNe-LD only utilized variants with $MAF > 0.25$, as in previous analyses (see Supplementary Note). Populations flagged as having unusual cross-chromosome LD or no signal were excluded from the analysis (Supplementary Fig. 20). Genetic maps corresponding to the GRCh38 genome build were used for this analysis.

Inference of demographic history in ancient data

We downloaded version 50.0 of the Allen Ancient DNA Resource (AADR) dataset^{61,62}. For each analysis, we started by removing related individuals reported in the annotation files present in the dataset. For each family, the individual with the highest coverage was kept. Information about sample ages was also extracted from the annotation file and used as input for HapNe-LD. To ensure high data quality, we filtered the datasets to include only variants and individuals with a missing data rate m of less than 80%, corresponding to a coverage of approximately 0.22 \times . Specific information about each population is present in Supplementary Tables 1–8.

Software

We used ARGON⁵³ (version 0.1.160415) and msprime⁵⁴ (version 1.1.1) to simulate synthetic data. We used Plink 1.9⁶³ and Plink 2.0⁶⁴ to preprocess genetic files. We used FastSMC³⁴ (version 1.2), HapIBD³⁵ (version 1.0, 23Apr20.f1a), and RefinedIBD³³ (version 17jan20) to detect IBD segments in simulated and real data analyses. We used IBDNe²⁷ (version 23Apr20.ae9), and GONE³¹ (retrieved on Jun 22, 2021) to infer effective population sizes. Data processing and plotting were performed using Numpy (1.23.4)⁶⁵, Pandas (1.5.1)⁶⁶, SciPy (1.9.3)⁶⁷, Numba (0.56.3)⁶⁸, Matplotlib (3.4.3)⁶⁹, Seaborn (0.12.2)⁷⁰, and Geopandas (0.12.2)⁷¹.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data sets used for this study can be accessed using the following links. The data sets simulated for this study are available at <https://doi.org/10.5281/zenodo.10024899>. UK Biobank data can be accessed by approved researchers through <http://www.ukbiobank.ac.uk/>. Other data sets can be downloaded from the following URLs: genetic maps https://ftp-trace.ncbi.nih.gov/1000genomes/ftp/technical/working/20110106_recombination_hotspots/, 1000 Genomes Project phase three⁶⁰<https://www.internationalgenome.org/data/> and the Allen Ancient DNA Resource⁶¹<https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data> (Supplementary Tables 1–8).

Code availability

The HapNe software package is freely available at <https://palamaralab.github.io/software/hapne>.

References

- Charlesworth, B. Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **10**, 195–205 (2009).
- Wright, S. Evolution in Mendelian populations. *Genetics* **16**, 97–159 (1931).
- Wright, S. Inbreeding and homozygosity. *Proc. Natl Acad. Sci.* **19**, 411–420 (1933).
- Pickrell, J. K. & Reich, D. Toward a new history and geography of human genes informed by ancient DNA. *Trends Genet.* **30**, 377–389 (2014).
- Nielsen, R. et al. Tracing the people of the world through genomics. *Nature* **541**, 302–310 (2017).
- Sikora, M. et al. Ancient genomes show social and reproductive behavior of early upper paleolithic foragers. *Science* **358**, 659–662 (2017).
- Kondrashov, A. S. Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *J. Theoret. Biol.* **175**, 583–594 (1995).
- Franklin, I. R. & Frankham, R. How large must populations be to retain evolutionary potential? *Anim. Conserv.* **1**, 69–70 (1998).
- Schraiber, J. G. & Akey, J. M. Methods and models for unravelling human evolutionary history. *Nat. Rev. Genet.* **16**, 727–740 (2015).
- Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional snp frequency data. *PLoS Genet.* **5**, e1000695 (2009).
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C. & Foll, M. Robust demographic inference from genomic and SNP data. *PLoS Genet.* **9**, e1003905 (2013).
- Bhaskar, A., Wang, Y. R. & Song, Y. S. Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Res.* **25**, 268–279 (2015).
- Kamm, J., Terhorst, J., Durbin, R. & Song, Y. S. Efficiently inferring the demographic history of many populations with allele count data. *J. Am. Stat. Assoc.* **115**, 1472–1487 (2020).
- Terhorst, J. & Song, Y. S. Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proc. Natl Acad. Sci.* **112**, 7677–7682 (2015).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Sheehan, S., Harris, K. & Song, Y. S. Estimating variable effective population sizes from multiple genomes: a sequentially Markov conditional sampling distribution approach. *Genetics* **194**, 647–662 (2013).
- Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genome sequences. *Nat. Genet.* **46**, 919–925 (2014).
- Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
- Steinrucken, M., Kamm, J., Spence, J. P. & Song, Y. S. Inference of complex population histories using whole-genome sequences from multiple populations. *Proc. Natl Acad. Sci.* **116**, 17115–17120 (2019).
- Speidel, L., Forest, M., Shi, S. & Myers, S. R. A method for genome-wide genealogy estimation for thousands of samples. *Nat. Genet.* **51**, 1321–1329 (2019).
- Speidel, L. et al. Inferring population histories for ancient genomes using genome-wide genealogies. *Mol. Biol. Evol.* **38**, 3497–3511 (2021).

22. Upadhyaya, G. & Steinrücken, M. Robust inference of population size histories from genomic sequencing data. *PLoS Comput. Biol.* **18**, e1010419 (2022).
23. Palamara, P. F., Lencz, T., Darvasi, A. & Pe'er, I. Length distributions of identity by descent reveal fine-scale demographic history. *Am. J. Hum. Genet.* **91**, 809–822 (2012).
24. Palamara, P. F. & Pe'er, I. Inference of historical migration rates via haplotype sharing. *Bioinformatics* **29**, i180–i188(2013).
25. Ralph, P. & Coop, G. The geography of recent genetic ancestry across Europe. *PLoS Biol.* **11**, 1001555 (2013).
26. Harris, K. & Nielsen, R. Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet.* **9**, e1003521 (2013).
27. Browning, S. R. & Browning, B. L. Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *Am. J. Hum. Genet.* **97**, 404–418 (2015).
28. Sved, J. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoret. Popul. Biol.* **2**, 125–141 (1971).
29. Tenesa, A. et al. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* **17**, 520–526 (2007).
30. McEvoy, B. P., Powell, J. E., Goddard, M. E. & Visscher, P. M. Human population dispersal “out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome Res.* **21**, 821–829 (2011).
31. Santiago, E. et al. Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Mol. Biol. Evol.* **37**, 3642–3653 (2020).
32. Gusev, A. et al. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* **19**, 318–326 (2008).
33. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
34. Saada, J. N. et al. Identity-by-descent detection across 487,409 British samples reveals fine-scale population structure and ultra-rare variant associations. *Nat. Commun.* **11**, 6130 (2020).
35. Zhou, Y., Browning, S. R. & Browning, B. L. A fast and simple method for detecting identity-by-descent segments in large-scale data. *Am. J. Hum. Genet.* **106**, 426–437 (2020).
36. Hill, W. G. Estimation of linkage disequilibrium in randomly mating populations. *Heredity* **33**, 229–239 (1974).
37. Weir, B. S. Inferences about linkage disequilibrium. *Biometrics* **35**, 235–254 (1979).
38. Excoffier, L. & Slatkin, M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol. Biol. Evol.* **12**, 921–927 (1995).
39. Waples, R. S. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conserv. Genet.* **7**, 167–184 (2006).
40. Ragsdale, A. P. & Gravel, S. Models of archaic admixture and recent history from two-locus statistics. *PLoS Genet.* **15**, e1008204 (2019).
41. Mezzavilla, M. Neon: An R package to estimate human effective population size and divergence time from patterns of linkage disequilibrium between SNPs. *J. Comput. Sci. Syst. Biol.* **8**, 037–044 (2015).
42. Nelis, M. et al. Genetic structure of Europeans: a view from the north-east. *PLoS One* **4**, e5472 (2009).
43. Margaryan, A. et al. Population genomics of the Viking world. *Nature* **585**, 390–396 (2020).
44. Novak, M. et al. Genome-wide analysis of nearly all the victims of a 6200 year old massacre. *PLoS One* **16**, e0247332 (2021).
45. Aberth, J. The black death 1348–1350: a brief history with documents. The Bedford Series in History and Culture (St Martin’s Press, New York, NY, 2005).
46. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
47. Kere, J. Human population genetics: lessons from Finland. *Annu. Rev. Genom. Hum. Genet.* **2**, 103–128 (2001).
48. Patterson, N. et al. Large-scale migration into Britain during the middle to late bronze age. *Nature* (2021).
49. Fernandes, D. M. et al. A genetic history of the pre-contact Caribbean. *Nature* **590**, 103–110 (2021).
50. Nägele, K. et al. Genomic insights into the early peopling of the Caribbean. *Science* **369**, 456–460 (2020).
51. Ringbauer, H., Novembre, J. & Steinrücken, M. Human parental relatedness through time—detecting runs of homozygosity in ancient DNA. *Nat. Commun.* **12**, 5425 (2020).
52. Price, A. L. et al. Long-range LD can confound genome scans in admixed populations. *Am. J. Hum. Genet.* **83**, 132–135 (2008).
53. Palamara, P. F. Argon: fast, whole-genome simulation of the discrete time wright-fisher process. *Bioinformatics* **32**, 3032–3034 (2016).
54. Kelleher, J., Etheridge, A. M. & McVean, G. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLoS Comput. Biol.* **12**, e1004842 (2016).
55. Bissiri, P. G., Holmes, C. C. & Walker, S. G. A general framework for updating belief distributions. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **78**, 1103–1130 (2016).
56. Holmes, C. C. & Walker, S. G. Assigning a value to a power likelihood in a general Bayesian model. *Biometrika* **104**, 497–503 (2017).
57. Cai, R., Browning, B. L. & Browning, S. R. Identity-by-descent-based estimation of the X chromosome effective population size with application to sex-specific demographic history. *G3 Genes|Genomes|Genetics* **13**, jkad165 (2023).
58. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
59. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
60. Byrka-Bishop, M. et al. High-coverage whole-genome sequencing of the expanded 1000 genomes project cohort including 602 trios. *Cell* **185**, 3426–3440.e19 (2022).
61. Mallick, S. & Reich, D. The Allen ancient DNA resource (AADR): a curated compendium of ancient human genomes, Harvard dataverse, v50.0 data release [10 Oct 2021] (2021).
62. Mallick, S. et al. The Allen Ancient DNA Resource (AADR): a curated compendium of ancient human genomes. Preprint at <https://www.biorxiv.org/content/10.1101/2023.04.06.535797v1> (2023).
63. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
64. Chang, C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, s13742–015–0047–8 (2015).
65. Harris, C. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
66. Team, T. pandas-dev/pandas: Pandas. (Zenodo, 2020,2), <https://doi.org/10.5281/zenodo.3509134>.
67. Virtanen, P. et al. SciPy 1.0 contributors SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
68. Lam, S. et al. Numba: a llvm-based python jit compiler. In: *Proc. Second Workshop on the LLVM Compiler Infrastructure in HPC*. pp. 1–6 (2015).
69. Hunter, J. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
70. Waskom, M. seaborn: statistical data visualization. *J. Open Source Softw.* **6**, 3021 (2021).
71. Jordahl, K. et al. geopandas/geopandas: v0.8.1. (Zenodo,2020,7).

Acknowledgements

We thank Simon Myers, Geoff Nicholls, Juba Nait Saada, and Fergus Cooper for helpful discussions and suggestions; Arjun Biddanda and Shai Carmi for comments on an early version of the manuscript; and Brian Zhang and Arjun Biddanda for sharing code used for various parts of the analysis. This work was conducted using the UK Biobank resource (Application #43206). We thank the participants of the UK Biobank project. Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. This work was supported by the Angus McLeod Scholarship (to R.F.); EPSRC grant EP/SO23151/1 (to Z.T.); NIH grant R21-HG010748-01 (to P.F.P.); Wellcome Trust ISSF grant no. 204826/Z/16/Z (to P.F.P.); and ERC Starting Grant ARGPHENO 850869 (to P.F.P. and R.F.). D.R. is an investigator of the Howard Hughes Medical Institute and this work was also supported by grants from the National Institutes of Health (GM100233 and HG012287), and the John Templeton Foundation (grant 61220). This research was funded in whole, or in part, by the Wellcome Trust (204826/Z/16/Z and 203141/Z/16/Z). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the author-accepted manuscript of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Author contributions

P.F.P. designed the study; R.F. and P.F.P. developed and implemented the HapNe model; R.F., Z.T. and P.F.P. generated and analyzed simulated data; R.F., D.R., and P.F.P. analyzed real data; R.F. developed software tools; R.F. and P.F.P. wrote the manuscript with input and edits from D.R. and Z.T.; P.F.P. and D.R. supervised analyses.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-43522-6>.

Correspondence and requests for materials should be addressed to Pier Francesco Palamara.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023