Bioinformatics, 39(1), 2023, btac738
https://doi.org/10.1093/bioinformatics/btac738
Advance Access Publication Date: 15 November 2022
Original Paper

Genetics and population analysis

Achieving improved accuracy for imputation of ancient DNA

Kristiina Ausmees * and **Carl Nettelblad** *

Department of Information Technology, Uppsala University, Uppsala 751 05, Sweden

*To whom correspondence should be addressed.

Associate Editor: Russell Schwartz

Received on June 2, 2022; revised on November 7, 2022; editorial decision on November 9, 2022; accepted on November 14, 2022

Abstract

Motivation: Genotype imputation has the potential to increase the amount of information that can be gained from the often limited biological material available in ancient samples. As many widely used tools have been developed with modern data in mind, their design is not necessarily reflective of the requirements in studies of ancient DNA. Here, we investigate if an imputation method based on the full probabilistic Li and Stephens model of haplotype frequencies might be beneficial for the particular challenges posed by ancient data.

Results: We present an implementation called prophaser and compare imputation performance to two alternative pipelines that have been used in the ancient DNA community based on the Beagle software. Considering empirical ancient data downsampled to lower coverages as well as present-day samples with artificially thinned genotypes, we show that the proposed method is advantageous at lower coverages, where it yields improved accuracy and ability to capture rare variation. The software prophaser is optimized for running in a massively parallel manner and achieved reasonable runtimes on the experiments performed when executed on a GPU.

Availability and implementation: The C++ code for prophaser is available in the GitHub repository https://github.com/scicompuu/prophaser.

Contact: kristiina.ausmees@it.uu.se or carl.nettelblad@it.uu.se

Supplementary information: Supplementary information is available at *Bioinformatics* online.

1 Introduction

The possibility of sequencing ancient human remains has allowed genetic analyses to be incorporated into studies of the history and evolution of our species. While methodological advances have led to an increase in the availability of ancient human genetic data, sample quality, quantity and cost still pose limitations for its analysis. Consequently, a significant amount of the available ancient DNA (aDNA) is sequenced at relatively low coverages of 1× or less, which leads some standard genomic and population genetic tools to be infeasible or unreliable to use (Marciniak and Perry, 2017).

One approach to extending the information content and making available a wider range of methods for such data is to impute genotypes at sites that are unobserved or otherwise not possible to call confidently. Imputation has been successfully applied to present-day data in a wide range of studies, including for conforming samples that have been genotyped on different arrays (Li *et al.*, 2009) as well as calling dense diploid genotypes from low- and medium-coverage data (Auton *et al.*, 2015).

Genotype imputation is commonly performed by statistical models that take into account patterns of genetic variation in a panel of reference individuals with known haplotypes. The haplotype frequency model of Li and Stephens (2003) is the foundation for many commonly used imputation tools such as IMPUTE2 (Howie *et al.*, 2009) and Beagle versions 4.1 and above (Browning and Browning, 2016). Beagle 4.0 (Browning and Browning, 2007) uses a slightly different underlying model based on local haplotype clusters. Among more recent methods is GLIMPSE (Rubinacci *et al.*, 2021), which is designed for imputing low-coverage data and is based on a modified version of the Li and Stephens framework.

In the aDNA community, imputation has mainly been performed using the algorithm of Beagle 4.0 with genotype likelihoods as input (Antonio *et al.*, 2019; Ausmees *et al.*, 2022; Cassidy *et al.*, 2020; Gamba *et al.*, 2014; Jones *et al.*, 2015; Martiniano *et al.*, 2017). In a recent study, an evaluation of different imputation pipelines on an empirical ancient sample was conducted (Hui *et al.*, 2020). They compared imputation performed in a single step based on genotype likelihoods, as implemented in Beagle 4.0 and GLIMPSE, to a two-step methodology in which either GLIMPSE or Beagle is used to determine sites that can be confidently called based on the likelihoods, after which imputation is performed by Beagle 5.0 based on the resulting genotypes. They found that the two-step methodology resulted in more refined posterior genotype probabilities (GP) that allowed for an improved balance between accuracy and the number of retained sites when performing post-imputation filtering.

In general, methods for genotype imputation have been designed to target present-day data, with algorithm simplifications made in order to prioritize computational efficiency and applicability to increasingly large study samples and reference panels. While Beagle versions up to 4.0 allowed for imputation to be performed based on genotype likelihoods, later versions require genotypes as input, which may introduce errors for low-coverage data in which genotypes cannot be confidently called. Many methods also reduce the number of possible states induced by the phased reference, e.g. by the haplotype cluster model of Beagle 4.0, or by considering a subset of reference haplotypes based on some measure of similarity to the sample, as in IMPUTE2 and GLIMPSE. The latter two are also examples of models that decouple the phasing and imputation steps, iterating haplotype estimation based on sample data and subsequent haploid imputation. This is a strategy that can reduce the computational requirements of imputation and has laid the foundation for pre-phasing approaches in which the two steps are completely separated and imputation requires phased input data. Also supported by IMPUTE2 and Beagle since versions 4.1, pre-phasing improves runtime for imputation with an often small trade-off in accuracy, although larger errors may result for populations where the haplotype phase is more difficult to estimate (Das *et al.*, 2018; Howie *et al.*, 2012).

These trends in priorities for the development of imputation methods do not necessarily reflect the requirements of aDNA studies, where the goal is often to extract as much information as possible from a few precious samples in order to perform population genetic analyses, rather than e.g. preparing tens or hundreds of thousands of samples for genome-wide association studies. Further, as low coverage and sequencing errors are prevalent in ancient data, and assumptions of the available phased reference panels being representative of the study samples may not hold, algorithm simplifications may incur larger trade-offs in accuracy than for present-day data.

In this article, we present a tool for performing genotype imputation that has been designed with ancient and other low-coverage data in mind. Our approach is different from other implementations based on the Li and Stephens model outlined above, in that it is based on genotype likelihood input and implements the full probabilistic model for the entire state space of possible haplotype configurations, rather than a subset. It also integrates over the unknown phase of the target data while performing imputation. Unlike these previous implementations, in which algorithmic simplifications have been made in order to increase computational efficiency, our contribution is intended for scenarios in which particularly challenging data motivates the use of the more comprehensive and expensive algorithm.

In order to maintain practical usability on large problem sizes, we present an implementation, made available as the software prophaser that extends previous work on reducing the memory requirements of the algorithm. This allows the tool to be executed efficiently on GPUs for datasets of relevant sizes. We can show that this yields wall-clock runtimes comparable to current best practice CPU-based workflows for a reference set of realistic size. We evaluate prophaser on simulated low-coverage data based on present-day samples from the 1000 Genomes project as well as downsampled data from five empirical ancient individuals and compare performance to both a single- and a two-step imputation pipeline based on Beagle from Hui *et al.* (2020).

2 Materials and methods

2.1 Empirical data

Five ancient samples for which high-coverage genomes ranging from 19× to 57× are available were used for evaluation of imputation performance. These were sf12 from Günther *et al.* (2018), Loschbour and LBK from Lazaridis *et al.* (2014), ne1 from Gamba *et al.* (2014) and ans17 from Fraser and Sánchez-Quinto *et al.* (in preparation). Gold standard genotypes to which imputed data were compared were obtained by the same pipeline as described in their respective publications, and further filtered to keep only sites that were confidently called. Low-coverage data for the evaluation samples were generated by downsampling reads to coverages ranging from 0.1× to 2×, after which genotype likelihoods were generated using a pipeline based on the Genome Analysis Toolkit (GATK) v3.5.0 (McKenna *et al.*, 2010) tool UnifiedGenotyper. Unlike most previous studies of the imputation of aDNA, we did not filter the imputation input to remove genotypes that could be affected by post-mortem damage, which was motivated by the findings in Hui *et al.* (2020). More information about the samples and data preparation procedure can be found in Ausmees *et al.* (2022).

2.2 Simulated data

Simulated low-coverage data were generated for 10 individuals from the CEU population of the 1000 Genomes Phase 3 dataset (Auton *et al.*, 2015), filtered to keep only biallelic SNPs, by the following procedure.

For every site j with a biallelic SNP with called genotype G_j:

- Decide whether or not to sample a read with probability p.
- If a read is not sampled, set the likelihood to 0.3333 for all three possible genotypes.
- If a read is sampled:
  a. Sample one read from the distribution defined in Equation 1, which, for the biallelic SNP at location j that is sampled randomly N_j times, describes the probability of sampling A_j copies of the alternative allele A and R_j copies of the reference allele R (so that $A_j + R_j = N_j$). This assumes independently observed reads, and a constant sequencing error rate per base δ.

$$P(A_j|G_j) = \begin{Bmatrix} Bin(A_j, N_j, 1-\delta) & \text{if } G_j = AA \\ Bin(A_j, N_j, 0.5) & \text{if } G_j = RA \\ Bin(A_j, N_j, \delta) & \text{if } G_j = RR \end{Bmatrix}. \quad (1)$$

  b. Let X denote the allele that was sampled, and Y the allele that was not sampled. Calculate likelihoods for the three genotypes according to Equation 2, which is based on the formula for genotype likelihoods from the GATK framework (McKenna *et al.*, 2010).

$$L(G) = \begin{Bmatrix} 0.5(1-\delta) + 0.5(1-\delta) & \text{for } G = XX \\ 0.5(\delta/3) + 0.5(1-\delta) & \text{for } G = YX \\ 0.5(\delta/3) + 0.5(\delta/3) & \text{for } G = YY \end{Bmatrix}. \quad (2)$$

For different simulated coverage levels $c \in [0, 1]$, the sampling probability p was selected so that the fraction of sampled sites would be approximately c. The called genotypes of the original 1000 Genomes data, G_j, were used as gold standard genotypes to which imputed data were compared to assess accuracy. See the Supplementary Material for more information about the simulated data.

2.3 Model and implementation

The method implemented in the prophaser is based on a diploid version of the Hidden Markov Model (HMM) presented in Li and Stephens (2003). We first outline this model, in which observed data correspond to called genotypes, and later describe the implemented modification which instead considers genotype likelihoods.

For a set of H phased reference haplotypes at genomic positions $j \in \{1 \dots M\}$, the observed states G_j of the model correspond to the sample genotypes, and the hidden states $X_j = (x^1, x^2)$ consist of a pair of values that indicate which reference haplotype is used as a template for each chromosome at the position. The observation probabilities, shown in Table 1, are based on the similarity between the observation and the genotype implied by the template haplotypes, and depend on an error parameter ϵ_j. The transition

Table 1. The observation probability $P(G_j|X_j)$ of genotype G_j at position j given the hidden state, with $G(X_j)$ denoting the genotype implied by the hidden state X_j

	$G(X_j)$		
G_j	0	1	2
0	$(1-\varepsilon_j)^2$	$(1-\varepsilon_j)\varepsilon_j$	ε_j^2
1	$2\varepsilon_j(1-\varepsilon_j)$	$(1-\varepsilon_j)^2+\varepsilon_j^2$	$2\varepsilon_j(1-\varepsilon_j)$
2	ε_j^2	$(1-\varepsilon_j)\varepsilon_j$	$(1-\varepsilon_j)^2$

probabilities of the HMM are shown in Equation 3. These depend on the number of switches in the template haplotype for each chromosome and are parameterized by H and θ_j, reflecting the probability of historical recombination events between the two genomic positions. We refer to the Supplementary Material for more information about the model and its implementation.

$$P(X_j|X_{j-1}) = \begin{cases} \dfrac{\theta_j^2}{H^2} & \text{if 2 switch} \\ \dfrac{(1-\theta_j)\theta_j}{H} + \dfrac{\theta_j^2}{H^2} & \text{if 1 switch} \\ (1-\theta_j)^2 + \dfrac{2(\theta_j-\theta_j^2)}{H} + \dfrac{\theta_j^2}{H^2} & \text{if 0 switch} \end{cases} \quad (3)$$

The described model was used to define an HMM that incorporates the uncertainty of genotypes by having as observed data a vector of genotype likelihoods $L(g)$ for each possible genotype g. For the case of biallelic SNPs, $g \in \{0,1,2\}$ and the observation at position j is thus $Y_j = [L(0), L(1), L(2)]$. Uniform likelihoods were assumed for unobserved sample sites. Integration over the unknown genotype yields the joint probability of hidden and observed states shown in Equation 4.

$$P(Y,X) = P(X_1)\prod_{j=2}^{M} P(X_j|X_{j-1}) \prod_{j=1}^{M}\sum_{g=0}^{2} P(g|X_j)L(g) \quad (4)$$

The forward–backward algorithm (Rabiner, 1989) was used to calculate posterior probabilities for the hidden states given the sequence of observed data $P(X_j|Y_1,\ldots,Y_M)$ for all positions j. Phased haplotypes could then be obtained by selecting the state with the highest probability, and posterior GP by integration over the hidden states.

Our implementation reuses the collapsing of transition probabilities involving switches between haplotypes outlined in Li *et al.* (2010). This brings down the effective computational complexity to linear in the number of states (quadratic in the number of haplotypes). Li *et al.* (2010) also employed a checkpointing scheme, only keeping every k elements in the permanent forward-backward tables, reducing memory usage to be proportional to the square root of the number of variant positions. In our implementation, we extend the checkpointing scheme to two levels, rather than one. This makes memory usage proportional to the cubic root of the number of SNPs, with a trebling in the amount of computations relative to a non-checkpointing counterpart. The reductions in memory usage make the resulting algorithm more cache-friendly and the total workset small enough to fit in the local memory of a conventional GPU. Using OpenMP offloading pragmas, the same source code can be compiled for multicore CPUs and compliant GPUs.

2.4 Experimental setup

Imputation performance was evaluated for three different methodologies: the proposed implementation prophaser as well as a single-step and a two-step pipeline based on Beagle. The single-step approach is the one that has mainly been used in aDNA studies previously and consisted of using Beagle 4.0 to impute data using genotype likelihoods as input. The two-step methodology corresponds to the one that was recommended in Hui *et al.* (2020) based on a comparison of single- and two-step pipelines based on Beagle and GLIMPSE. In this pipeline, Beagle 4.1 was used to estimate posterior GP from the genotype likelihoods, to which a pre-imputation filter of maximum GP ≥ 0.99 was applied. The filtered genotypes were used as input for imputation, which was performed using Beagle 5.2. See the Supplementary Material for more details on the experimental setup as well as additional results.

Imputation was performed on chromosome 20 using a panel of phased reference haplotypes from the 1000 Genomes Phase 3 dataset and a genetic map derived from the HapMap2 (Frazer *et al.*, 2007) project. The phased reference data contained 2504 individuals and were filtered to include only biallelic SNPs. Experiments were performed on the simulated and empirical datasets separately. When imputing the simulated data, the 10 CEU individuals whose genotypes the sample data were derived from were removed from the reference set.

Imputation performance was assessed by genotype concordance, defined as the fraction of evaluated sites where the imputed genotype is the same as the gold standard genotype. We also report the number of SNPs evaluated in terms of raw numbers or the fraction they make up of the total amount of sites that are possible to compare, i.e. the number of sites that overlap the gold standard data and the non-filtered output from imputation. Where results are reported for different allele frequencies, the minor allele frequency (MAF) within the phased reference panel was used.

The Beagle pipelines were run on CPUs, with Beagle 4.0 allocated two cores of a compute server consisting of two 8-core Xeon E5-2660 processors, 218GB memory and Beagle versions 4.1 and 5.2 running on compute servers consisting of two 10-core Xeon E5-2630 V4 processors, 128 GB memory using four and five cores, respectively. The prophaser runs were performed on an NVIDIA A100 Tensor Core GPU with 40 GB on-board HBM2 VRAM.

3 Results

For the results in this section, we only considered sites at which the gold standard genotype was heterozygous, as these are more challenging to impute and less subject to chance agreement to the reference majority allele. We also note that we did not see any tendencies of the methods to systematically mispredict homozygotes as heterozygote, and that the patterns in relative performance between the methods considered for different coverages that are visible in the results below were also observed when considering all genotypes. A post-imputation filter requiring the GP value of the most probable imputed genotype to reach a certain threshold was also applied, with different thresholds used for the three imputation methodologies in order to obtain comparable numbers of evaluated SNPs. We found that the two-step Beagle methodology resulted in a less nuanced posterior GP, with many genotypes having the value 1.0, and consequently a relatively large amount of sites being retained. We therefore used a high threshold of 0.999 for this pipeline and adjusted the other two accordingly with 0.9 for single-step Beagle and 0.8 for prophaser. We prioritized obtaining similar numbers of evaluated sites for the lower coverages in particular, as these are of main interest in this study and also the cases with the largest differences in performance between the methods. We refer to the Supplementary Material for additional results considering all genotypes as well as different GP thresholds for the post-imputation filter.

For runtimes, we report the execution time per sample and imputation window, see the Supplementary Material for more information about how the different methodologies were executed. Using the computational resources described in the previous section, prophaser execution took 40 min. The runtimes of the Beagle software depended on coverage and ranged between 2.2 and 7.12 h for the single-step pipeline, and between 1 min and 2.5 h for the two-step pipeline.

3.1 Empirical data

Figure 1 summarizes the performance of the three imputation methodologies for different coverages of the empirical dataset. The single-step Beagle pipeline showed more variable outcomes for the different evaluation individuals and was outperformed by the other

approaches. The largest difference between the methods was seen for coverages below 0.5×, where prophaser had higher concordance while retaining a larger share of SNPs. Figure 2 shows the imputation performance over the allele frequency spectrum for coverages 0.1×, 0.5× and 2×. Again, the differences between methodologies were more pronounced for lower coverages, with the prophaser outperforming the others consistently across the MAF spectrum at 0.1× coverage but only showing discernible advantages at the lowest MAF values for 2× data.

3.2 Simulated data

For the simulated data, the range of coverages considered contained lower values than that of the empirical data. Figure 3 shows that for coverages below 0.2×, the single-step Beagle methodology outperformed the two-step one and had the highest concordance overall for the lowest coverage considered of 0.01×. Inspection of the corresponding plot in Figure 4 shows that differences between prophaser and single-step Beagle varied over allele frequencies, where the latter had fewer sites retained at the lower end of the spectrum and a corresponding higher concordance, whereas the opposite holds at the higher end of the MAF range. Their relative performance may be subject to the particular post-imputation GP thresholds used, whereas the two-step Beagle pipeline had consistently lower concordance than the single-step, even when less sites were retained.

For the higher coverages, the trends were largely similar to those observed for the empirical dataset, with the main difference being increased performance for the single-step pipeline over the two-step one for coverage 0.1×, although this could also be driven by differences in number of sites retained by the filter. Another possible explanation for the different patterns seen for the two datasets is that the haplotypes of the simulated samples are expected to have a higher degree of similarity to those in the reference panel as they are derived from the same population, in contrast to the ancient individuals which are temporally distinct from the present-day individuals in the panel.

4 Discussion and conclusion

With the majority of ancient genomes being sequenced at coverages for which reliable diploid variant calls are typically not feasible (Günther and Jakobsson, 2019; Marciniak and Perry, 2017; Parks and Lambert, 2015), the use of imputation methods that are based on probabilistic genotypes has become standard in the aDNA community.

In this article, we propose an imputation tool that accepts genotype likelihood input and implements the full probabilistic Li and Stephens framework for haplotype frequencies. We evaluate performance on empirical ancient samples and simulated sparse genotypes, and show that the proposed tool achieves higher heterozygote concordance than two alternative methodologies adopted in the aDNA community that are based on the Beagle software. We further show that it provides reliable genotype calls over a larger part of the allele frequency spectrum and that it is particularly advantageous on low-coverage data, and we therefore believe that it may be of use in

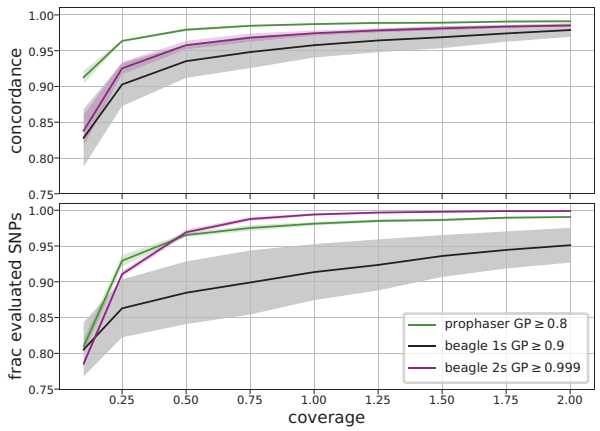

Fig. 1. Genotype concordance (top) and fraction of evaluated SNPs (bottom) for heterozygote sites per coverage level, averaged over the five empirical evaluation individuals, with shaded regions displaying standard deviation. The legend indicates the GP thresholds used in the post-imputation filter for the different methodologies

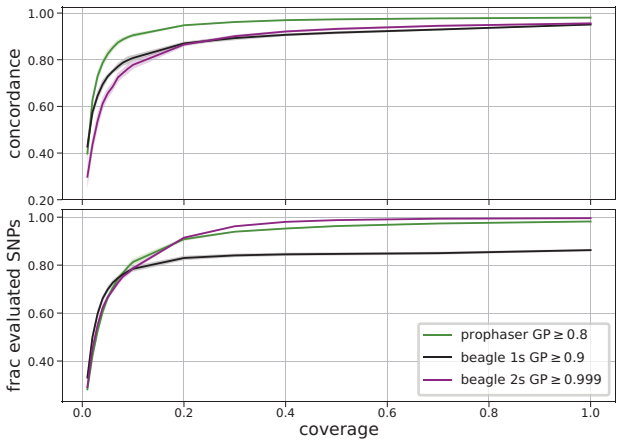

Fig. 3. Genotype concordance (top) and fraction of evaluated SNPs (bottom) for heterozygote sites per coverage level, averaged over the 10 simulated evaluation individuals, with shaded regions displaying standard deviation. The legend indicates the GP thresholds used in the post-imputation filter for the different methodologies

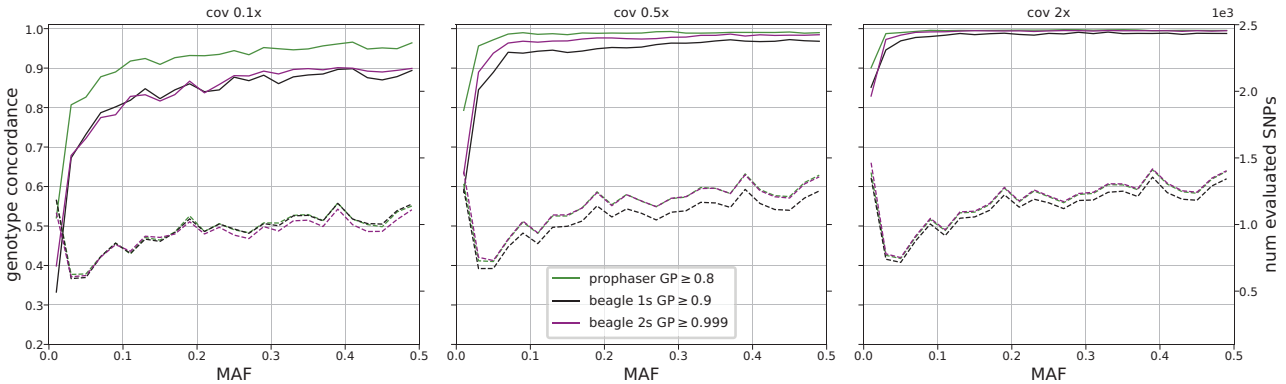

Fig. 2. Genotype concordance (solid lines) and number of evaluated SNPs (dashed lines) for heterozygote sites, displayed over 25 MAF bins, averaged over the five empirical evaluation individuals, for three coverage levels. The three plots share both left and right y-axes, with axis ranges indicated on the respective figure edge. The legend indicates the GP thresholds used in the post-imputation filter

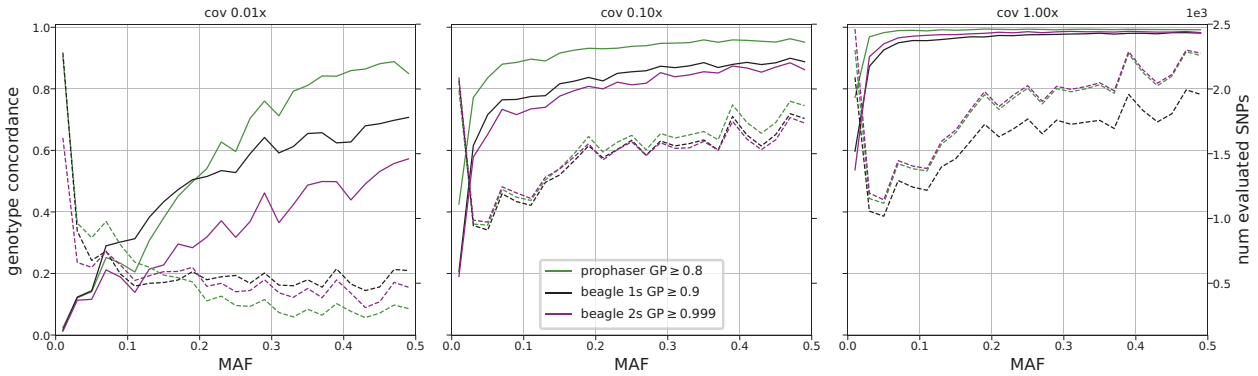

Fig. 4. Genotype concordance (solid lines) and number of evaluated SNPs (dashed lines) for heterozygote sites, displayed over 25 MAF bins, averaged over the 10 simulated evaluation individuals, for three coverage levels. The three plots share both left and right y-axes, with axis ranges indicated on the respective figure edge. The legend indicates the GP thresholds used in the post-imputation filter

studies of aDNA as a means to increase information content where high-coverage data are not available.

In addition to aDNA imputation, the tool would also be useful for other cases of challenging low-coverage imputation. Our simulated results on downsampling of 1000 Genomes individuals show that the tool is applicable in such scenarios. However, we believe that the benefits are the highest for ancient data, where the low coverage can also be unevenly distributed across the genome, individual calls can be less certain, and the divergence between the haplotypes found in reference individuals, and those in the study population, can be greater.

The presented implementation is optimized to run in a highly parallel manner, and we show that it achieves feasible runtimes on a realistic problem size when executed on a GPU. This design is in line with recent trends in the field of machine learning, where such processing architectures are seeing increased use due to applications like the training of deep artificial neural networks.

The underlying model lends itself to further customization for aDNA, and future work may focus on investigating the possibilities to include sample- or position-dependent adjustments to the HMM. This could include the incorporation of known patterns of errors in aDNA into the observation probabilities, or using estimates of divergence between sample and reference to adjust transition probabilities.

Acknowledgements

The authors acknowledge the use of computational resources provided by Swedish National Infrastructure for Computing (SNIC) and associated centers under projects SNIC 2021/22-94, UPPMAX 2020/2-5 and SNIC 2017/7-162, as well as the use of the Berzelius resource funded by the Knut och Alice Wallenbergs Stiftelse (KAW) under project Berzelius-2021-30.

Funding

This work was supported by Formas grant [2017-00453].

Conflict of Interest: none declared.

Data availability

The authors state that all data necessary for confirming the conclusions presented in the article are represented fully within the article and in its online supplementary material.

References

Antonio,M.L. *et al.* (2019) Ancient Rome: a genetic crossroads of Europe and the Mediterranean. *Science*, **366**, 708–714.

Ausmees,K. *et al.* (2022) An empirical evaluation of genotype imputation of ancient DNA. *G3 Genes—Genomes—Genetics*, **12**, jkac089.

Auton,A. *et al.*; 1000 Genomes Project Consortium. (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.

Browning,B. and Browning,S. (2016) Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.*, **98**, 116–126.

Browning,S.R. and Browning,B.L. (2007) 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.

Cassidy,L.M. *et al.* (2020) A dynastic elite in monumental Neolithic society. *Nature*, **582**, 384–388.

Das,S. *et al.* (2018) Genotype imputation from large reference panels. *Annu. Rev. Genomics Hum. Genet.*, **19**, 73–96.

Frazer,K.A. *et al.*; International HapMap Consortium. (2007) A second generation human haplotype map of over 3.1 million SNPS. *Nature*, **449**, 851–861.

Gamba,C. *et al.* (2014) Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.*, **5**, 5257.

Günther,T. and Jakobsson,M. (2019) Population genomic analyses of DNA from ancient remains. In: Balding,D.J. *et al.* (eds) *Handbook of Statistical Genomics, Chapter 10*. Wiley, Hoboken, NJ, pp. 295–324.

Günther,T. *et al.* (2018) Population genomics of mesolithic scandinavia: investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol.*, **16**, e2003703.

Howie,B. *et al.* (2012) Fast and accurate genotype i in genome-wide association studies through pre-phasing. *Nat. Genet.*, **44**, 955–959.

Howie,B.N. *et al.* (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.*, **5**, e1000529.

Hui,R. *et al.* (2020) Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Sci. Rep.*, **10**, 18542.

Jones,E.R. *et al.* (2015) Upper palaeolithic genomes reveal deep roots of modern Eurasians. *Nat. Commun.*, **6**, 8912.

Lazaridis,I. *et al.* (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*, **513**, 409–413.

Li,N. and Stephens,M. (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, **165**, 2213–2233.

Li,Y. *et al.* (2009) Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, **10**, 387–406.

Li,Y. *et al.* (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.

Marciniak,S. and Perry,G.H. (2017) Harnessing ancient genomes to study the history of human adaptation. *Nat. Rev. Genet.*, **18**, 659–674.

Martiniano,R. *et al.* (2017) The population genomics of archaeological transition in west Iberia: investigation of ancient substructure using imputation and haplotype-based methods. *PLoS Genet.*, **13**, e1006852.

McKenna,A. *et al.* (2010) The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**, 1297–1303.

Parks,M. and Lambert,D. (2015) Impacts of low coverage depths and post-mortem DNA damage on variant calling: a simulation study. *BMC Genomics*, **16**, 19.

Rabiner,L.R. (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE*, **77**, 257–286.

Rubinacci,S. *et al.* (2021) Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.*, **53**, 120–126.