The spatiotemporal patterns of major human admixture events during the European Holocene

Manjusha Chintalapati[1#], Nick Patterson[2#], Priya Moorjani[1,3#]
[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94620 USA
[2]Broad Institute of Harvard and MIT, Cambridge, MA, 02142, USA
[3]Center for Computational Biology, University of California, Berkeley, CA 94620 USA

[#]To whom correspondence should be addressed: Manjusha Chintalapati <m_chintalapati@berkeley.edu>, Nick Patterson <nickp@broadinstitute.org> and Priya Moorjani <moorjani@berkeley.edu>

Abstract

Recent studies have shown that admixture has been pervasive throughout human history. While several methods exist for dating admixture in contemporary populations, they are not suitable for sparse, low coverage ancient genomic data. Thus, we developed *DATES* that leverages ancestry covariance patterns across the genome of a single individual to infer the timing of admixture. *DATES* provides reliable estimates under various demographic scenarios and outperforms available methods for ancient DNA applications. Using *DATES* on ~1,100 ancient genomes, we reconstruct major gene flow events during European Holocene. By studying the genetic formation of Anatolian farmers, we infer that gene flow related to Iranian Neolithic farmers occurred before 9,600 BCE, predating the advent of agriculture in Anatolia. Contrary to the archaeological evidence, we estimate that early Steppe pastoralist groups (Yamnaya and Afanasievo) were genetically formed more than a millennium before the start of steppe pastoralism. Using time transect samples across sixteen regions, we provide a fine-scale chronology of the Neolithization of Europe and the rapid spread of Steppe pastoralist ancestry across Europe. Our analyses provide new insights on the origins and spread of farming and Indo-European languages, highlighting the power of genomic dating methods to elucidate the legacy of human migrations.

Introduction

Recent studies have shown that population mixture (or "admixture") is pervasive throughout human history, including mixture between the ancestors of modern humans and archaic hominins (i.e., Neanderthals and Denisovans), as well as in the history of many contemporary human groups such as African Americans, South Asians and Europeans (Hellenthal et al., 2014; Moorjani et al., 2016). Many admixed groups are formed due to population movements involving ancient migrations that pre-date historical records. The recent availability of genomic data for a large number of present-day and ancient genomes provides an unprecedented opportunity to reconstruct population events using genetic data, providing evidence complementary to linguistics and archaeology. Understanding the timing and signatures of admixture offers insights into the historical context in which the mixture occurred and enables the characterization of the evolutionary and functional impact of the gene flow.

To characterize patterns of admixture, genetic methods use the insight that the genome of an admixed individual is a mosaic of chromosomal segments inherited from distinct ancestral populations (Chakraborty and Weiss, 1988). Due to recombination, these ancestral segments get shuffled in each generation and become smaller and smaller over time. The length of the segments is inversely proportional to the time elapsed since the mixture (Chakraborty and Weiss, 1988; Moorjani et al., 2011). Several genetic approaches—ROLLOFF (Moorjani et al., 2011), ALDER (Loh et al., 2013), Globetrotter (Hellenthal et al., 2014), and Tracts (Gravel, 2012)— have been developed that use this insight by characterizing patterns of admixture linkage disequilibrium (LD) or haplotype lengths across the genome to infer the timing of mixture. Haplotype-based methods perform chromosome painting or local ancestry inference at each locus in the genome and characterize the distribution of ancestry tract lengths to estimate the time of mixture (Gravel, 2012; Hellenthal et al., 2014). This requires accurate phasing and inference of local ancestry, which is often difficult when the admixture events are old (as ancestry blocks become smaller over time) or when reference data from ancestral populations is unavailable. Admixture LD-based methods, on the other hand, measure the extent of the allelic correlation across markers to infer the time of admixture (Loh et al., 2013; Moorjani et al., 2011). They do not require phased data from the target or reference populations and work reliably for dating older admixture events (>100 generations). However, they tend to be less efficient in characterizing admixture events between closely related ancestral groups.

While highly accurate for dating admixture events using data from present-day samples, current methods do not work reliably for dating admixture events using ancient genomes. Ancient DNA samples often have high rates of DNA degradation, contamination (from human and other sources) and low sequencing depth, leading to a large proportion of missing variants and uneven coverage across the genome. Additionally, most studies generate pseudo-haploid genotype calls––consisting of a single allele call at each diploid site—that can lead to some issues in the inference. In such sparse datasets, estimating admixture LD can be noisy and biased (see Simulations below). Moreover, haplotype-based methods require phased data from both admixed and reference populations which remains challenging for ancient DNA specimens.

An extension of admixture LD-based methods, recently introduced by Moorjani et al. (2016), leverages ancestry covariance patterns that can be measured in a single sample using low coverage data. This approach measures the allelic correlation across neighboring sites, but instead of measuring admixture LD across multiple samples, it integrates data across markers within a single diploid genome. Using a set of ascertained markers that are informative for Neanderthal ancestry (where sub-Saharan Africans are fixed for the ancestral alleles and Neanderthals have a derived allele), Moorjani et al. (2016) inferred the timing of Neanderthal gene flow in Upper Paleolithic Eurasian samples and showed the approach works accurately in ancient DNA samples (Moorjani et al., 2016). However, this approach is inapplicable for dating admixture events within modern human populations, as there are very few fixed differences across populations (Auton et al., 2015).

Motivated by the single sample statistic in Moorjani et al. (2016), we developed *DATES (Distribution of Ancestry Tracts of Evolutionary Signals*) that measures the ancestry covariance across the genome in a single admixed individual, weighted by the allele frequency difference between two ancestral populations. This method was first introduced in Narasimhan et al. (2019), where it was used to infer the date of gene flow between groups related to Ancient Ancestral South Indians, Iranian farmers, and Steppe pastoralists in ancient South and Central Asian populations (Narasimhan et al., 2019). In this study, we evaluate the performance of *DATES* by performing extensive simulations for a range of demographic scenarios and compare the approach to other published genomic dating methods. We then apply *DATES* to infer the chronology of the genetic formation of the ancestral populations of Europeans and the spatiotemporal patterns of admixture during the European Holocene using data from ~1,100 ancient DNA specimens spanning ~8,000–350 BCE.

Results

Overview of *DATES*: Model and simulations

DATES estimates the time of admixture by measuring the weighted ancestry covariance across the genome using data from a single diploid genome and two reference populations (representing the ancestral source populations). *DATES* works like haplotype-based methods as it is applicable to dating admixture in a single genome and not like admixture LD-based methods, which by definition require multiple genomes to be co-analyzed; but unlike haplotype-based methods, it is more flexible as it does not require local ancestry inference. There are three main steps in *DATES*: we start by first learning the genome-wide ancestry proportions by performing a simple regression analysis to model the observed genotypes in an admixed individual as a linear mix of allele frequencies from the two reference populations. For each marker, we then compute the likelihood of the observed genotype in the admixed individual using the estimated ancestry proportions and allele frequencies in each reference population (this is similar in spirit to local ancestry inference). This information is, in turn, used to compute the joint likelihood for two neighboring markers to test if they derive ancestry from the same ancestral group, accounting for

the probability of recombination between the two markers. Finally, we compute the covariance across pairs of markers located at a particular genetic distance, weighted by the allele frequency differences in the reference populations (Methods).

Following (Moorjani et al., 2016), we bin the markers that occur at a similar genetic distance across the genome, rather than estimating admixture LD for each pair of markers, and compute the covariance across increasing genetic distance between markers. The estimated covariance is expected to decay exponentially with genetic distance, and the rate of decay is informative of the time of the mixture (Moorjani et al., 2011). Assuming the gene flow occurred instantaneously, we infer the average date of gene flow by fitting an exponential distribution to the decay pattern (Methods). In cases where data for multiple individuals is available, we compute the likelihood by summing over all individuals. To make *DATES* computationally tractable, we implemented the fast Fourier transform (FFT) for calculating ancestry covariance as described in ALDER (Loh et al., 2013). This provides a speedup from $O(n^2)$ to $O(n \log n)$ that reduces the typical runtimes from hours to seconds with minimal loss in accuracy (Appendix 1—figure 2).

To assess the reliability of *DATES*, we performed simulations where we constructed ten admixed diploid genomes by randomly sampling haplotypes from two source populations (Methods). Briefly, we simulated individual genomes with 20% European and 80% African ancestry by using phased haplotypes of Northern Europeans (Utah European Americans, CEU) and west Africans (Yoruba from Nigeria, YRI) from the 1000 Genomes Project respectively (Auton et al., 2015). As reference populations in *DATES*, we used closely related surrogate populations of French and Yoruba respectively, from the Human Genome Diversity Panel (Rosenberg et al., 2002). We first investigated the accuracy of *DATES* by varying the time of admixture between 10–300 generations. For comparison, we also applied ALDER (Loh et al., 2013) to these simulations. Both methods reliably recovered the time of admixture up to 200 generations or ~5,600 years ago, assuming a generation time of 28 years (Moorjani et al., 2016), though *DATES* was more precise than ALDER for older admixture events (>100 generations) (Figure 1—figure supplement 1, Appendix 1—table 4). Further, *DATES* shows accurate results even for single samples (Figure 1A, Figure 1—figure supplement 2A) and even when few reference individuals are available for dating (Figure 1A, Figure 1—figure supplement 2B). Moreover, we find use of large numbers of reference samples, which are often easily available, can help to improve the inference. In *DATES*, allele frequencies of the reference populations are used in computing the likelihood as well as the weighted pairwise ancestry covariance across the genome (Methods). Thus, reducing the sampling noise in inferred allele frequencies can, in turn, improve the precision of inferred dates.

Next, we tested *DATES* for features such as varying admixture proportions and use of surrogate populations as reference groups. By varying of European ancestry proportion between ~1–50% (the rest derived from west Africans), we observed *DATES* accurately estimated the timing in all cases (Figure 1—figure supplement 3A). However, the inferred admixture proportion was overestimated for lower admixture proportions (<10%) (Figure 1—figure supplement 3B). Thus, we caution against using *DATES* for estimating ancestry proportions and recommend other methods based on *f*-statistics (Patterson et al., 2012). *DATES* works reliably for dating admixtures

between related groups such as Europeans and Mexicans (F_{ST} ~ 0.03), though it was unable to distinguish mixtures of Southern and Northern Europeans (F_{ST} < 0.005) (Figure 1—figure supplement 5).

We found *DATES* is robust to the use of highly divergent surrogates as reference populations. For example, the use of Khomani San as the reference population instead of the true ancestral population of Yoruba (F_{ST} ~ 0.1) provides unbiased dates of admixture (Figure 1—figure supplement 4). In this regard, for ancient DNA where sometimes only sparse data is available, one can also use present-day samples as reference populations to increase the quality and sample size of the ancestral groups. In principle, as long as the allele frequencies in the reference samples are correlated to the ancestral allele frequencies, the inference of admixture dates should remain unbiased (Methods). In practice, however, recent demographic events (e.g., strong founder events or admixture from additional sources, etc.) in the history of the present-day samples could lead to significant deviation from the ancestral allele frequencies. Thus, the reference populations should be carefully chosen.

Another idea is to use the admixed populations themselves as one of the reference populations as demonstrated by the single reference setup in ALDER (Loh et al., 2013). Admixed individuals have intermediate allele frequencies to the ancestral populations and thus weighted LD or ancestry covariance can be computed with only one reference population (albeit, with reduced power). Loh et al., 2013 showed that the use of admixed populations as one of the references does not bias the rate of decay of the weighted LD (i.e., time of admixture), though the amplitude of the decay curve (not used in *DATES*) can be biased under some scenarios. To verify *DATES* provides reliable results under this setup, we performed simulations and applied *DATES* with a single reference population and used the admixed population as the other reference. Like ALDER, our inferred dates of admixture were unbiased (Figure 1—figure supplement 6).

An important feature of *DATES* is that it does not require phased data and is applicable to datasets with small sample sizes, making it in principle useful for ancient DNA applications. To test the reliability of *DATES* for ancient genomes, we simulated data mimicking the relevant features of ancient genomes, namely small sample sizes (n = 1–20), large proportions of missing genotypes (between 10–60%), and pseudo-haploid genotype calls (instead of diploid genotype calls) in reference and / or target samples. *DATES* showed reliable results in both cases, even only a single admixed individual was available (Figure 1B, Figure 1—figure supplements 1–9). In contrast, admixture LD-based methods require more than one sample and do not work reliably with missing data. For example, ALDER estimates were very unstable for simulations with >40% missing data. For older dates (>100 generations), there was a slight bias even with >10% missing genotypes (Appendix 1—figure 5). As LD calculations leverage shared patterns across samples, variable missingness of genotypes across individuals leads to substantial loss of data leading to unstable and noisy inference. We also generated data for combinations of features including small sample sizes, pseudo-haploid genotypes with large proportions of missing genotypes in both target and reference samples and use of highly divergent reference samples. We found *DATES* yielded reliable results with large amounts (~40-60%) of missing data, either in the target or references, even when highly divergent reference populations were used (Figure 1—figure supplement 8).

This was also true for when a single target sample was available but as expected, the inference becomes noisier for older dates and large fractions of missing data (Figure 1—figure supplement 9). This highlights a major advantage of *DATES* for ancient DNA applications as it provides robust results even in sparse datasets.

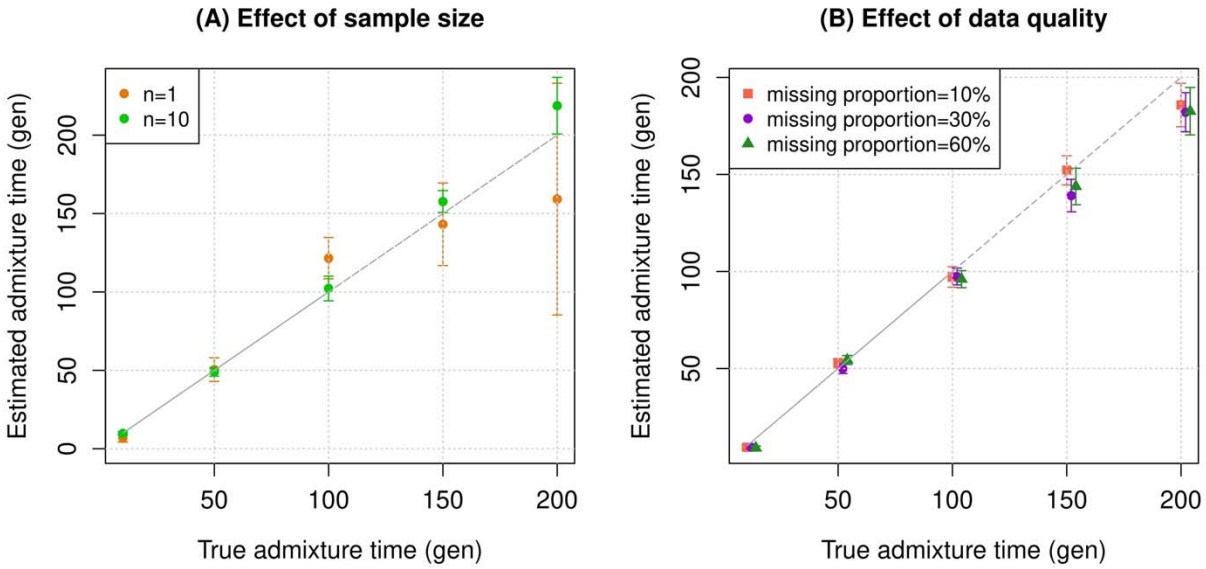


Figure 1: Simulation results. We constructed *n* admixed individuals with 20% European (CEU) and 80% African (YRI) ancestry using ~380,000 genome-wide SNPs for admixture dates ranging between 10–200 generations. To minimize any issues with overfitting, we used French and Yoruba from the Human Genome Diversity Panel as reference populations in *DATES*. We show the true time of admixture (X-axis, in generations) and the estimated time of admixture (± 1 SE) (Y-axis, in generations). Standard errors were calculated using a weighted block jackknife approach by removing one chromosome in each run (Methods). (**A)** Effect of sample size: We varied the sample size (*n*) of target group between 1–10 individuals; (**B)** Effect of data quality: To mimic the features of ancient genomes, we generated *n*=10 target individuals with pseudo-haploid genotypes and missing genotype rate as 10% (orange), 30% (purple) and 60% (green). R code to replicate this figure is available at:
https://github.com/manjushachintalapati/DATES_EuropeanHolocene/blob/main/Figure1.R

See Figure 1—figure supplements **1-9** for additional simulations to test the performance of *DATES*. R code to replicate supplement figures:
https://github.com/manjushachintalapati/DATES_EuropeanHolocene/blob/main/Figure1_supplementaryFigures.R

DATES assumes a model of instantaneous gene flow with a single pulse of mixture between two source populations. However, many human populations have a history of multiple pulses of gene flow. To test the performance of *DATES* for multi-way admixture events, we generated admixed individuals with ancestry from three sources (East Asians, Africans, and Europeans)

where the gene flow occurred at two distinct time points (Appendix 2—figure 1). By applying *DATES* with pairs of reference populations at a time and fitting a single exponential to the ancestry covariance patterns, we observed that *DATES* recovered both admixture times in case of equal ancestry proportion from the three ancestral groups when the associated reference groups were used for dating (Appendix 2—figure 2). In the case of unequal admixture proportions from three ancestral groups, *DATES* inferred the timing of the recent admixture event in most cases, though some confounding was observed, especially when the ancestry proportion of the recent event was low (Appendix 2—figure 3). However, if the reference populations were set up to match the model of gene flow, we observed that we could reliably recover the time of the recent gene flow event. For example, there is limited confounding if the two references used in *DATES* include: (i) the source population for the recent event and (ii) both ancestral populations that contributed to the first event or the intermediate admixed group formed after the first event (Appendix 2—table 1). This highlights how the choice of reference populations can help to tune the method to infer the timing of specific admixture events reliably.

Finally, we explored the impact of more complex demographic events, including continuous admixture and founder events using coalescent simulations (Appendix 2). In the case of continuous admixture, *DATES* inferred an intermediate timing between the start and the end of the gene flow period, similar to other methods like ALDER and Globetrotter (Hellenthal et al., 2014; Loh et al., 2013) (Appendix 2—table 2). In the case of populations with founder events, we inferred unbiased dates of admixture in most cases except when the founder event was extreme ($N_e \sim 10$) or the population had maintained a low population size ($N_e < 100$) until the present (i.e., no recovery bottleneck) (Appendix 2—figure 4, Appendix 2—table3). In humans, few populations have such extreme founder events, and thus, in most other cases, our inferred admixture dates should be robust to founder events (Tournebize et al., 2020). We note that while *DATES* is not a formal test of admixture, in simulations, we find that in the absence of gene flow, the method does not infer significant dates of admixture even when the target has a complex demographic history (Appendix 2—figure 6, Appendix 2—figure 7).

Comparison to other methods

We assessed the reliability of *DATES* in real data by comparing our results with published methods: Globetrotter, ALDER, and ROLLOFF. These methods are designed for the analysis of present-day samples that typically have high-quality data with limited missing variants. In addition, Globetrotter uses phased data which is challenging for ancient DNA samples. Thus, instead of rerunning other methods, we took advantage of the published results for contemporary samples presented in Hellenthal et al., (2014) (Hellenthal et al., 2014). Following (Hellenthal et al., 2014), we created a merged dataset including individuals from Human Genome Diversity Panel (Rosenberg et al., 2002), Behar et al. (2010) (Behar et al., 2010), and Henn et al. (2012) (Henn et al., 2012) (Methods). We applied *DATES* and ALDER to 29 target groups using the reference populations reported in Table S12 in Hellenthal et al. 2014, excluding one group where the population label was unclear. Interestingly, the majority of these groups (25/29) failed ALDER's

formal test of admixture; either because the results of the single reference and two reference analyses yielded inconsistent estimates or because the target had long-range shared LD with one of the reference populations (Appendix 1—table 5). Using *DATES*, we inferred significant dates of admixture in 20 groups, and 14 of those were consistent with estimates based on Globetrotter. In the case of the six populations that disagreed across the two methods, most of the populations appear to have a history of multiple pulses of gene flow either involving more than two populations (e.g., Brahui (Pagani et al., 2017)) or multiple instances of contact between the same two reference groups (e.g., Mandenka (Price et al., 2009)) or the model of admixture for incorrect (e.g., present-day Bulgarians that have ancestry from western hunter-gatherers, Near Eastern farmers, and Steppe pastoralists from Eurasia but were modeled as a mixture of Polish and Cypriots in Globetrotter). In such cases, the inconsistencies across the methods are hard to interpret as in complex admixture cases, Globetrotter and *DATES* could be capturing different events or the weighting of both events could differ. Finally, the estimated admixture timing based on *DATES*, ROLLOFF, and ALDER (assuming two-way admixture regardless of the formal test results) were found to be highly concordant (Appendix 1—table 5).

Fine-scale patterns of population mixtures in ancient Europe

Recent ancient DNA studies have shown that present-day Europeans derive ancestry from three distinct sources: (a) hunter-gatherer-related ancestry that is closely related to Mesolithic hunter-gatherers (HG) from Europe; (b) Anatolian farmer-related ancestry related to Neolithic farmers from the Near East and associated to the spread of farming to Europe; and (c) Steppe pastoralist-related ancestry that is related to the Yamnaya pastoralists from Russia and Ukraine (Allentoft et al., 2015; Haak et al., 2015; Mathieson et al., 2018; Skoglund et al., 2012). Many open questions remain about the timing and dynamics of these population interactions, in particular related to the formation of the ancestral groups (which were themselves admixed) and their expansion across Europe. To characterize the spatial and temporal patterns of mixtures in Europe in the past 10,000 years, we used 1,096 ancient European samples from 152 groups from the publicly available Allen Ancient DNA Resource (AADR) spanning a time range of ~8,000–350 BCE (Methods, Supplementary file 1A). Using *DATES*, we characterized the timing of the various gene flow events, and below, we describe the key events in chronological order focusing on three main periods.

***Holocene to Mesolithic*:** Pre-Neolithic Europe was inhabited by hunter-gatherers until the arrival of the first farmers from the Near East (Haak et al., 2010; Hofmanová et al., 2016). There was large diversity among hunter-gatherers with four main groups— western hunter-gatherers (WHG) that were related to the Villabruna cluster in central Europe, eastern hunter-gatherers (EHG) from Russia and Ukraine related to the Upper Paleolithic group of Ancestral North Eurasians (ANE) ancestry, Caucasus hunter-gatherers (CHG) from Georgia associated to the first farmers from Iran, and the GoyetQ2-cluster associated to the Magdalenian culture in Spain and Portugal (Fernandes et al., 2018; Fu et al., 2016; Jones et al., 2015; Rivollat et al., 2020; Skoglund et al., 2012). Most

Mesolithic HGs fall on two main clines of relatedness: one cline that extends from Scandinavia to central Europe showing variable WHG–EHG ancestry, and the other in southern Europe with WHG–GoyetQ2 ancestry (Rivollat et al., 2020). This ancestry is already present in the 17,000 BCE *El Mirón* individual from Spain, suggesting that the GoyetQ2-related gene flow occurred well before the Holocene. However, the WHG–EHG cline was formed more recently during the Mesolithic period, though the precise timing of the spread of EHG ancestry remains less well understood.

To characterize the formation of the WHG–EHG cline, we used genomic data from 16 ancient HG groups (n=101) with estimated ages of ~7,500–3,600 BCE. We first verified the ancestry of each HG group using *qpAdm* that compares the allele frequency correlations between the target and a set of source populations to formally test the model of admixture and then infer the ancestry proportions for the best-fitted model (Haak et al., 2015). For each target population, we chose the most parsimonious model, i.e., fitting the data with the minimum number of source populations. Consistent with previous studies, our *qpAdm* analysis showed that most HGs from Scandinavia, the Baltic Sea region, and central Europe could be modeled as a two-way mixture of WHG and EHG-related ancestry (Supplementary file 2A). To confirm that the target populations do not harbor Anatolian farmer-related ancestry (that could lead to some confounding in estimated admixture dates), we applied D-statistics of the form D(Mbuti, *target*, WHG, Anatolian farmers) where *target* = Mesolithic HGs. We observed that none of the target groups had a stronger affinity to Anatolian farmers than WHG (Supplementary file 2B). Together, these results suggest that the mixtures we date below reflect pre-Neolithic contacts between the HGs.

To infer the timing of the mixtures in the history of Mesolithic European HGs, we applied *DATES* to hunter-gatherers from Scandinavia, the Baltic regions, and central Europe. *DATES* infers the time of admixture in generations before the sample lived. Accounting for the average sampling age of the specimens and the mean human generation time of 28 years (Moorjani et al., 2016), we inferred the admixture time in BCE using 1950 conversion (Methods). We report the average dates (or median, where specified) in BCE in the main text and provide additional details in Figure 3 and Supplementary file 1B including sample size, dates in generations and BCE. Among HGs, we inferred that the earliest admixture occurred in Scandinavian HGs from Norway and Sweden with a range of average dates of ~80–113 generations before the samples lived (Figure 3–figure supplement 1). This translates to admixture dates of ~10,200–8,000 BCE, with the most recent dates inferred in Motala HG's from Sweden suggesting substantial substructure in hunter-gatherers (Figure 3). In the Baltic region, we inferred the range of admixture dates of ~8,700–6,000 BCE in Latvia and Lithuania HGs, postdating the mixture in Scandinavia. In southeast Europe, the Iron Gates region of the Danube Basin shows widespread evidence of mixtures between hunter-gatherer groups and, in the case of some outliers, the mixture of hunter-gatherers and Anatolian farmer-related ancestry as early as the Mesolithic period (Feldman et al., 2019). Further, these groups showed a strong affinity to the WHG-related ancestry in Anatolian populations, suggesting ancient interactions with Near Eastern populations (Feldman et al., 2019). We applied *qpAdm* to test the model of admixture in Iron Gates HG and found that the parsimonious model with WHG and EHG provides a good fit to the data. Further, when we tested

the model with Anatolian-related ancestry using Anatolian HG (AHG) as an additional source population, AHG was not required as the AHG ancestry proportion was not significant (Supplementary file 2A). Applying *DATES* to Iron Gates HG with WHG and EHG as reference populations, we inferred this group was genetically formed in on average ~9,200 (95% confidence interval: 10,000–8400 BCE). Our samples of the Iron Gates HGs include a wide range of C14 dates between 8,800–5,700 BCE. We confirmed our dates were robust to the sampling age of the individuals as we obtained statistically consistent dates when all samples were combined as one group or when subsets of samples were grouped in bins of 500 years (Figure 3 —figure supplement 2). The most recent dates of ~7,500 BCE were inferred in eastern Europe in Ukraine HGs, highlighting how the WHG-EHG cline was formed over a period ~2000–3000 years (Figure 3, Supplementary file 1B).

***Early to middle Neolithic**:* Neolithic farming began in the Near East—the Levant, Anatolia, and Iran—and spread to Europe and other parts of the world (Haak et al., 2010; Kılınç et al., 2016; Skoglund et al., 2012). The first farmers of Europe were related to Anatolian farmers, whose origin remains unclear. The early Neolithic Anatolian farmers (Aceramic Anatolian farmers) had majority ancestry from AHG with some gene flow from the first farmers from Iran (Feldman et al., 2019). AHG, in turn, had ancestry from Levant HG (Natufians) and some mysterious hunter-gatherer group related to the ancestors of WHG individuals from central Europe— a gene flow event that likely occurred in the late Pleistocene (Feldman et al., 2019). Using *qpAdm*, we confirmed that early Anatolian farmers could be modeled as a mixture of AHG and Iran Neolithic farmer-related groups (Supplementary file 2C). To learn about the timing of the genetic formation of early Anatolian farmers, we applied *DATES* using one reference group as a set of pooled individuals of WHG-related and Levant Neolithic farmers-related individuals as a proxy of AHG ancestry and the second reference group containing pooled Iran Neolithic farmer-related individuals. We note that the application of *DATES* to three-way admixed groups can lead to intermediate dates between the first and second pulse of gene flow unless the reference populations are chosen carefully (Appendix 2—table 1). Our setup for early Anatolian farmers should have limited confounding and should recover the timing of the most recent event (in this case, the gene flow from CHG or Iran Neolithic-related groups) reliably. We infer the Iran Neolithic farmer-related gene flow occurred ~10,900 BCE (12,200–9,600 BCE) (Figure 2), predating the origin of farming in Anatolia (Bramanti et al., 2009). During the subsequent millennia, these early farmers further admixed with Levant Neolithic groups to form Anatolian Neolithic farmers who spread towards the west to Europe and in the east to mix with Iran Neolithic farmers, forming the Chalcolithic groups of Seh Gabi and Hajji Firuz (Supplementary file 2C). Using *DATES*, we inferred that these Chalcolithic groups were genetically formed in ~7,600–5,700 BCE (Supplementary file 1B).

In Europe, the Anatolian Neolithic farmers mixed with the local indigenous hunter-gatherers contributing between ~40–98% of ancestry to the Neolithic Europeans. To elucidate the fine-scale patterns and regional dynamics of these mixtures, we applied *DATES* to time transect samples from 94 groups (*n*=657) sampled from sixteen regions in Europe, ranging from ~6,000-

1,900 BCE and encompassing individuals from the early Neolithic to Chalcolithic periods (Supplementary file 1A). Using *qpAdm*, we first confirmed that the Neolithic Europeans could be modeled as a mixture of European hunter-gatherer-related ancestry and Anatolian farmer-related ancestry and inferred their ancestry proportions (Supplementary file 2D). For most target populations (~80%), we found the model of gene flow between Anatolian farmer-related and WHG-related ancestry provided a good fit to the data (p-value > 0.05). In some populations, we found variation in the source of the HG-related ancestry and including either EHG or GoyetQ2 improved the fit of the model. In five groups, none of the models fit, despite excluding outlier individuals whose ancestry profile differed from the majority of the individuals in the group (Supplementary file 2E). To confirm that the target populations do not harbor Steppe pastoralist-related ancestry, we applied D-statistics of the form D(Mbuti, *target*, Anatolian farmers, Steppe pastoralists) where *target* = Neolithic European groups. We observed that four groups had a stronger affinity to Steppe pastoralists compared to Anatolian farmers, and hence we excluded these from further analysis (Supplementary file 2F). After filtering, we applied *DATES* to 86 European Neolithic groups using WHG-related individuals and Anatolian farmers as reference populations.

Earlier analysis has suggested that farming spread along two main routes in Europe, from southeast to central Europe ('continental route') and along the Mediterranean coastline to Iberia ('coastal route') (D. Gronenborn, 2014; J. Guilaine, 2003; Rivollat et al., 2020). Consistent with this, we inferred one of the earliest timings of gene flow was in the Balkans around 6,400 BCE. Using the most comprehensive time-transect in Hungary with 19 groups (n=63) spanning from middle Neolithic to late Chalcolithic, we inferred the admixture dates ranged from ~6,100–4,500 BCE. Under a model of a single shared gene flow event in the common ancestors of all individuals, we would expect to obtain similar dates of admixture (before present) after accounting for the age of the ancient specimens. Similar to Lipson et al. (2017), we observed that the estimated dates in middle Neolithic individuals were substantially older than those inferred in late Neolithic or Chalcolithic individuals (Figure 3). This would be expected if the underlying model of gene flow involved multiple pulses of gene flow, such that the timing in the middle Neolithic samples reflects the initial two-way mixture and the timing in the Chalcolithic samples captures both recent and older events. Interestingly, Lipson et al. (2017) and other recent studies have documented increasing HG ancestry from ~3-15% from the Neolithic to Chalcolithic period (Haak et al., 2015; Lipson et al., 2017; Rivollat et al., 2020), suggesting that there was additional HG gene flow after the initial mixture. This highlights that the interactions between local hunter-gatherers and incoming Anatolian farmers were complex with multiple gene flow events between these two groups, which explains the increasing HG ancestry and more recent dates in Chalcolithic individuals (Supplementary file 2D).

Mirroring the pattern in Hungary, we documented the resurgence of HG ancestry in the Czech Republic, France, Germany, and southern Europe. In central Europe, we inferred that the Anatolian farmer-related gene flow ranged between ~5,600-5,000 BCE across Germany and Czech Republic, with some exceptions. For instance, in the Blätterhöhle site from Germany, the inferred dates were more recent (~4,000 BCE), consistent with the occupation of both hunter-gatherers and

farmers in this region until the late Neolithic (Lipson et al., 2017). In eastern Europe, using samples related to the Funnel Beaker culture (TRB; from German *Trichterbecher*) from Poland, we dated the Anatolian farmer-related gene flow occurred on average ~4,700 BCE (5,300–4,200 BCE). Following the TRB decline, the Baden culture and the Globular Amphora culture appeared in many areas of Poland and Ukraine (Fernandes et al., 2018). These cultures had close contact with the Corded Ware complex and Steppe pastoralists' societies, though we found a parsimonious model without Steppe pastoralist-related ancestry provides a good fit to GAC individuals (Supplementary file 2D). Applying *DATES*, we inferred the Anatolian farmer-related and HG mixture in GAC ranged between ~4,700–3,900 BCE, predating the spread of Steppe pastoralists to eastern Europe (Allentoft et al., 2015; Haak et al., 2015).

Along the Mediterranean route, we characterized Anatolian farmer-related gene flow in Italy, Iberia, France, and the British Isles. Using samples from five groups in Italy, we inferred the earliest dates of Anatolian farmer-related gene flow of ~6,100 BCE, and within the millennium, the ancestry spread from Sardinia to Sicily (Figure 3). In Iberia, the Anatolian farmer-related mixture ranged from ~5,700–4,300 BCE and showed evidence for an increase in HG ancestry from ~9–20% after the initial gene flow. In France, previous studies have shown that Anatolian farmer-related ancestry came from both routes, along the Danubian in the north and along the Mediterranean in the south (Rivollat et al., 2020). This is reflected in the source of the HG ancestry, which is predominantly EHG and WHG-related in the north and includes WHG and Goyet-Q2 ancestry in the south (Rivollat et al., 2020). Consistently, we also observed that the admixture dates in France were structured along these routes, with the median estimate of ~5,100 BCE in the east and much older ~5,500 BCE in the south (Supplementary file 1B). In Scandinavia, we inferred markedly more recent dates of admixture of ~4,300 BCE using samples from Sweden associated with the TRB culture and Ansarve Megalithic tombs, consistent with a late introduction of farming to Scandinavia (Mittnik et al., 2018).

Finally, we inferred recent dates of admixture in Neolithic samples from the British Isles (England, Scotland, and Ireland) with the median timing of ~5,000 BCE across the three regions. Interestingly, unlike in western and southern Europe, there was no resurgence in HG ancestry during the Neolithic in Britain (Brace et al., 2019). This suggests our dates can be interpreted as the time of the main mixture of HGs and Anatolian farmers in this region, implying that the farmer-related ancestry reached Britain a millennium after its arrival in continental Europe. By 4,300 BCE, we find that Anatolian farmer-related ancestry is present in nearly all regions in Europe.

***Late Neolithic to Bronze Age*:** The beginning of the Bronze Age was a period of major cultural and demographic change in Eurasia, accompanied by the spread of Yamnaya Steppe pastoralist-related ancestry from Pontic-Caspian steppes across Europe and South Asia (Haak et al., 2015). The archaeological record documents that the early Steppe pastoralists cultures of Yamnaya and Afanasievo, with characteristic burial styles and pottery, appeared around ~3,300 to 2,600 BCE (Morgunova and Khokhlova, 2013). These groups were formed as a mixture of EHG-related groups and CHG-related groups associated with the first farmers from Iran (Jones et al., 2015; Narasimhan et al., 2019; Wang et al., 2019). Using *qpAdm*, we first tested how well this model fits

470 the data from 8 early Steppe pastoralist groups, including seven groups associated with Yamnaya
471 culture and one group related to the Afanasievo culture (Methods). For all but two Yamnaya groups
472 (from Hungary Baden and Russia Kalmykia), we found this model provides a good fit to the data
473 (Supplementary file 2G). We note that the samples from Kalmykia in our dataset were shotgun
474 sequenced, and in the *qpAdm* analysis, we are mixing shotgun and capture data that could
475 potentially lead to technical issues. To understand the timing of the formation of the early Steppe
476 pastoralist-related groups, we applied *DATES* using pooled EHG and pooled Iranian Neolithic
477 farmers. Focusing on the groups with the largest sample sizes, Yamnaya Samara (n=10) and
478 Afanasievo (n=19), we inferred the admixture occurred between 40–45 generations before the
479 individuals lived, translating to an admixture timing of ~4,100 BCE (Supplementary file 1B). We
480 obtained qualitatively similar dates across four Yamnaya and one Afanasievo groups, consistent
481 with the findings that these groups descend from a recent common ancestor (for Ozera samples
482 from Ukraine, the dates were not significant). This is also further supported by the insight that the
483 genetic differentiation across early Steppe pastoralist groups is very low (F_{ST} ~ 0.000-0.006)
484 (Supplementary file 2H). Thus, we combined all early Steppe pastoralist individuals in one group
485 to obtain a more precise estimate for the genetic formation of proto-Yamnaya of ~4,400 to 4,000
486 BCE (Figure 2). These dates are noteworthy as they pre-date the archaeological evidence by more
487 than a millennium (ANTHONY, 2007) and have important implications for understanding the
488 origin of proto-Pontic Caspian cultures and their spread to Europe and South Asia.
489

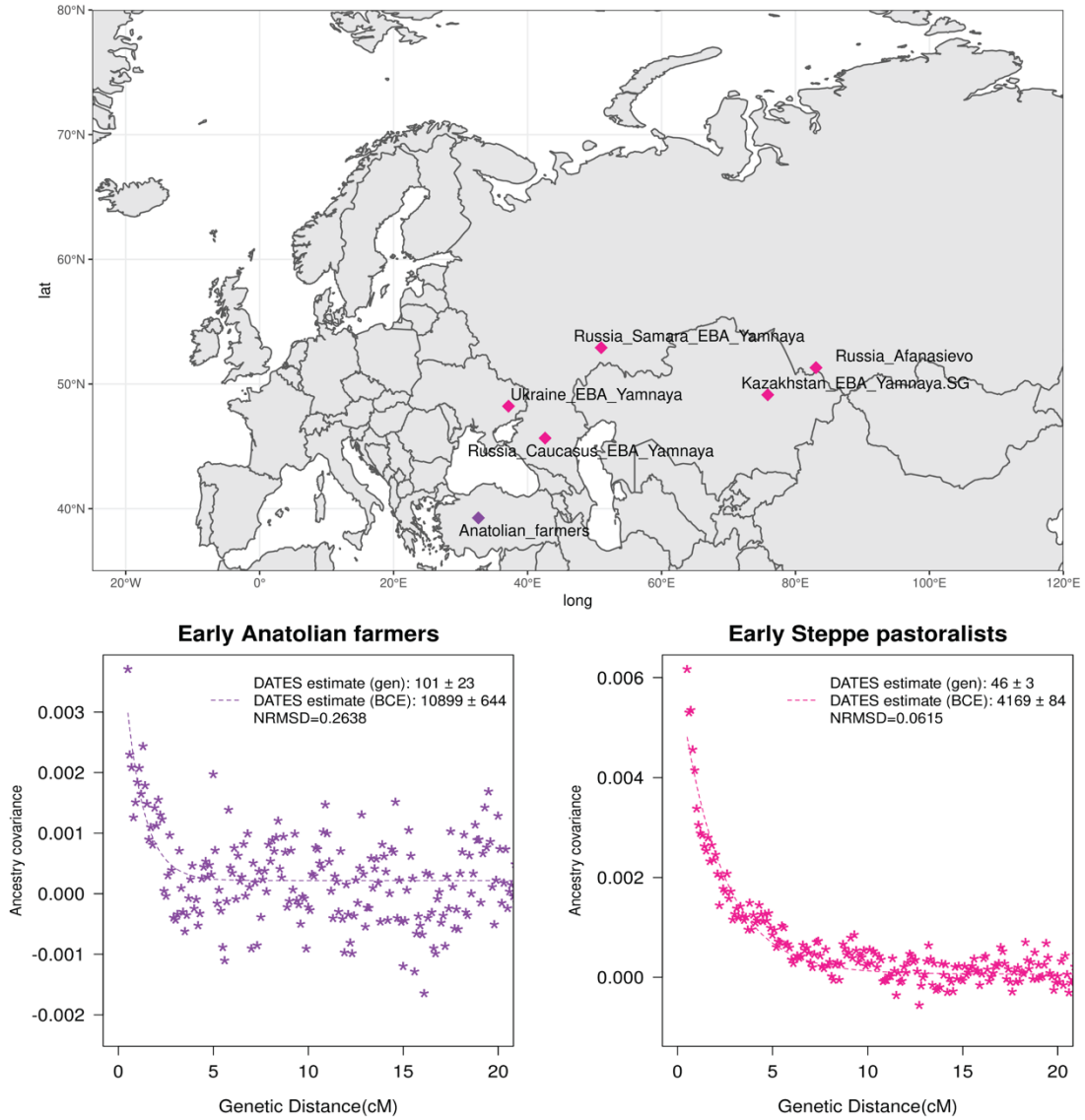


Figure 2: Genetic formation of early Anatolian farmers and early Bronze Age Steppe pastoralists. The top panel shows a map with sampling locations of the target groups analyzed for admixture dating. The bottom panels show the inferred times of admixture for each target using *DATES* by fitting an exponential function with an affine term $y = Ae^{-\lambda d} + c$, where d is the genetic distance in Morgans and $\lambda = (t+1)$ is the number of generations since admixture (t) (Methods). We start the fit at a genetic distance (d) > 0.5cM to minimize confounding with background LD and estimate a standard error by performing a weighted block jackknife removing one chromosome in each run. For each target, in the legend, we show the inferred average dates of admixture (± 1 SE) in generations before the individual lived, in BCE accounting for the average age of all the individuals and the mean human generation time, and the NRMSD values to assess the fit of the exponential curve (Methods). The bottom left shows the ancestry covariance decay curve for early Anatolian farmers inferred using one reference group as a set of pooled individuals of WHG-related and Levant Neolithic farmers-related individuals as a proxy of AHG ancestry and the second reference group containing Iran Neolithic farmer-related individuals. The bottom right shows the ancestry covariance decay curve for early Steppe pastoralists groups, including all

Yamnaya and Afanasievo individuals as the target group and EHG-related and Iran Neolithic farmer-related groups as reference populations. R code to replicate this figure is available at: https://github.com/manjushachintalapati/DATES_EuropeanHolocene/blob/main/Figure2.R

Over the following millennium, the Yamnaya-derived ancestry spread across Europe through Corded Ware Complex (CWC) and Bell Beaker complex (BBC) cultures. Present-day Europeans derive between ~10-60% Steppe pastoralist-related ancestry, which was not seen in Neolithic samples. To obtain a precise chronology of the spread of Steppe pastoralist-related ancestry across Europe, we analyzed 109 late Neolithic, Chalcolithic, and BA samples dated between 3,000–750 CE from 18 regions, including samples associated with the CWC and BBC cultures. We first confirmed that most target samples had Steppe pastoralist-related ancestry, in addition to European HG-related and Anatolian farmer-related ancestry using *qpAdm*. We excluded 20 groups that could not be parsimoniously modeled as a three-way mixture even after removing individual outliers. After filtering, we retained 79 groups for dating Steppe pastoralist-related gene flow across Europe (Supplementary file 2I, Supplementary file 2J). As Bronze Age Europeans have ancestry from three distinct groups, we applied *DATES* using the following two reference populations, one group including early Steppe pastoralists (Yamnaya and Afanasievo) and the other group with pooled samples of WHG-related and Anatolian farmer-related individuals, which is the proxy for the ancestral Neolithic Europe population.

To learn about the spread of CWC culture across Europe, we used seven late Neolithic and Bronze age groups, including five associated with CWC artifacts. Using *DATES*, we inferred that the oldest date of Steppe pastoralists gene flow in Europe was ~3,200 BCE in Scandinavia in samples associated with Battle Axe Culture in Sweden and Single Grave Culture in Denmark that were both contemporary to CWC. The samples from Scandinavia showed large heterogeneity in ancestry, including some individuals with majority Steppe pastoralist-related ancestry (and negligible amounts of Anatolian farmer-related ancestry), consistent with patterns expected from recent gene flow (Malmström et al., 2019). Strikingly, we inferred the timing of admixture in central Europe (Germany and the Czech Republic) and eastern Europe (Estonia and Poland) to be remarkably similar. These dates fall within a narrow range of ~3,000–2,900 BCE across diverse regions, suggesting that the mixed population associated with the Corded Ware culture formed over a short time and spread across Europe rapidly with very little further mixture (Supplementary file 1B).

Following the Corded Ware culture, from around 2,800 to 2,300 BCE, Bell Beaker pottery became widespread across Europe (Fokkens, H. and Nicolis, F, 2012). Using 19 Chalcolithic and Bronze Age samples, including ten associated with Beaker-complex artifacts, we inferred the dynamics of the spread of the Beaker complex across Europe. We inferred the oldest date of Steppe pastoralist-related admixture was ~3,200 BCE (3600–2800 BCE) in EBA Mallorca samples from Iberia. We note the EBA Mallorca sample is not directly associated with Beaker culture, but *qpAdm* modeling suggests that this individual is clade with the small subset of Iberian Beaker-complex-associated individuals who carried Steppe pastoralist-related ancestry (Fernandes et al., 2020). Most individuals from Iberia, however, had negligible Steppe pastoralist-related ancestry

547 suggesting the Beaker culture was not accompanied by major gene flow in Iberia despite the
548 earliest dates (Supplementary file 2I). In central and western Europe, where Steppe pastoralist gene
549 flow was more pervasive, we inferred the median date of the mixture was ~2,700 BCE with the
550 oldest dates in the Netherlands, followed by Germany and France (Figure 3). There was, however,
551 large heterogeneity in the dates across Europe and even within the same region. For example,
552 comparing two BA groups from the Netherlands suggests a wide range of dates ~3,000 BCE and
553 2,500 BCE, and four groups from Germany indicate a range of ~2,900–2,700 BCE. From central
554 Europe, the Steppe pastoralist-related ancestry spread quickly to the British Isles, where people
555 with Steppe pastoralist ancestry replaced 90% of the genetic ancestry of individuals from Britain.
556 Our estimates for the time of gene flow in Bell Beakers samples from England suggest that the
557 gene flow occurred ~2,700 BCE (2770-2550 BCE). Our estimated dates of admixture are older
558 than the dates of arrival of this ancestry in Britain (Olalde et al., 2018) and, interestingly, overlap
559 the dates in central Europe. Given that a significant fraction of the Beaker individuals were recent
560 migrants from central Europe, we interpret our dates reflect the admixture into ancestors of the
561 British Beaker people, occurring in mainland Europe (Olalde et al., 2018).
562 The middle to late Bronze age led to the final integration of Steppe pastoralist-related
563 ancestry in Europe. In southern Europe, early BA samples had limited Steppe pastoralist-related
564 ancestry, though present-day individuals harbor between ~5–30% of this ancestry (Haak et al.,
565 2015). Using pooled samples of middle to late BA from Spain, we inferred major mixture occurred
566 ~2,500 BCE in Iberia. We inferred a similar timing in Italy using individuals associated with the
567 Bell Beaker culture and early BA samples from Sicily (Supplementary file 1B). In Sardinia, a
568 majority of the BA samples do not have Steppe pastoralist-related ancestry. In a few individuals,
569 we found evidence for Steppe pastoralist-related ancestry, though in most cases, this ancestry
570 proportion overlapped 0 and the inferred dates of admixture were very noisy (Supplementary file
571 2I). Using Iron Age samples from Sardinia, we inferred the gene flow occurred ~2,600 BCE,
572 though there is a large uncertainty associated with this estimate (3,700–1,490 BCE). In other parts
573 of continental Europe and the British Isles, the Steppe pastoralist-related ancestry got diluted over
574 time, as evidenced by more recent dates in LBA than EBA or MBA samples in Germany, England,
575 and Scotland, and an increase in Neolithic farmer ancestry during this period (Olalde et al., 2019)
576 (Supplementary file 1B).
577 Finally, the Corded Ware Complex expanded to the east to form the archaeological
578 complexes of Sintashta, Srubnaya, Andronovo, and the Bronze Age cultures of Kazakhstan.
579 Samples associated with these cultures harbor mixed ancestry from the Yamnaya Steppe
580 pastoralist-related groups (CWC, in some cases) and Neolithic individuals from central Europe
581 (Supplementary file 2K) (Narasimhan et al., 2019). Applying *DATES* to 8 Middle to late Bronze
582 Age (MLBA) Steppe pastoralist groups, we inferred the precise timing for the formation of these
583 groups beginning in the third millennium BCE. These groups were formed chronologically, with
584 the date of genetic formation of ~3,200 BCE for Sintashta culture, followed by ~2,900 BCE for
585 Srubnaya and Andronovo cultures. In the central Steppe region (present-day Kazakhstan), we
586 obtained median dates of ~2,800 BCE for the expansion of Steppe pastoralist-related ancestry in
587 four Kazakh cultures of Maitan Alakul, Aktogai, and Kairan. By ~2,700 BCE, most of these

cultures had almost 60-70% Yamnaya Steppe pastoralist-related ancestry (Supplementary file 1B). These groups, in turn, expanded eastwards, transforming the genetic composition of populations in South Asia.

Figure 3: Timeline of admixture events in ancient Europe. We applied *DATES* to ancient samples from Europe. In the right panel, we show the sampling locations of the ancient specimens, and in the left panel, we show the admixture dates for each target group listed on the X-axis. The inferred dates in generations were converted to dates in BCE by assuming a mean generation time

of 28 years (Moorjani et al., 2011) and accounting for the average sampling age (shown as grey dots) of all ancient individuals in the target group (Methods). The top panel shows the formation of WHG–EHG cline (in blue) using Mesolithic hunter-gatherers as the target and EHG and WHG as reference populations. The middle panel shows admixture dates of local HGs and Anatolian farmers (in orange) using Neolithic European groups as targets and Anatolian farmers-related groups and WHG-related groups as reference populations. The bottom panel shows the spread of Steppe pastoralist-related ancestry (in green) estimated using middle and late Neolithic, Chalcolithic, and Bronze Age samples from Europe as target populations and early Steppe pastoralist-related groups (Afanasievo and Yamnaya Samara) and a set of Anatolian farmers and WHG-related groups as reference populations. For the middle to late Bronze Age samples from Eurasia, we used the early Steppe pastoralist-related groups and the Neolithic European groups as reference populations. The cultural affiliation (CWC, BBC, or Steppe MLBA cultures) of the individuals is shown in the legend. R code to replicate this figure is available at:
https://github.com/manjushachintalapati/DATES_EuropeanHolocene/blob/main/Figure3.R

See Figure 3—figure supplements **1** and **2** for decay curves for all samples and stratified dates for Iron Gates hunter-gatherers.
R code to replicate supplement figures:
https://github.com/manjushachintalapati/DATES_EuropeanHolocene/blob/main/Figure3_supplementaryFigures.R

Discussion

We developed *DATES* that measure ancestry covariance in a single diploid individual genome to estimate the time of admixture. Using extensive simulations, we show that *DATES* provide accurate estimates of the timing of admixture across a range of demographic scenarios. Application of *DATES* to present-day samples shows that the results are concordant with published methods—Rolloff, ALDER, and Globetrotter. For sparse datasets, *DATES* outperforms published methods as it does not require phased data and works reliably with limited samples, large proportions of missing variants as well as pseudo-haploid genotypes. This makes *DATES* ideally suited for the analysis of ancient DNA samples. We illustrate the application of *DATES* by reconstructing population movements and admixtures during the European Holocene. We confirm and extend signals that were previously identified such as the resurgence of hunter-gatherer ancestry during the Neolithic and provide new details about the genetic formation of the ancestral populations of Europeans and the spread of Corded Ware and Bell Beaker Complex cultures across Europe. Together, our analysis provides a detailed timeline and insights into the dynamics of the Neolithization of Europe and the spread of Steppe pastoralist-related ancestry across Europe.

First, we document that the Mesolithic hunter-gatherers formed as a mixture of WHG and EHG ancestry ~10,200 to 7400 BCE. These dates are consistent with the archeological evidence for the appearance of lithic technology associated with eastern HGs in Scandinavia and the Baltic regions (Günther et al., 2018; Kashuba et al., 2019; Mathieson et al., 2018). Next, we studied the timing of the genetic formation of Anatolian farmers. The earliest evidence of agriculture comes from the Fertile Crescent, the southern Levant, and the Zagros Mountains of Iran and dateed to around 10,000 BCE. In central Anatolia, farming has been documented c. 8,300 BCE (Baird et al.,

2018; Bellwood, 2005). It has been long debated if Neolithic farming groups from Iran and the Levant introduced agriculture to Anatolia or hunter-gatherers in the region locally adopted agricultural practices. The early Anatolian farmers can be modeled as a mixture of local hunter-gatherers people related to Caucasus hunter-gatherers or the first farmers from Iran (Feldman et al., 2019). By applying *DATES* (assuming a single instantaneous admixture), we inferred that the Iran Neolithic gene flow occurred around 10,900 BCE (~12,200–9,600 BCE). An alternate possibility is that there was a long period of gradual gene flow between the two groups and our dates reflect intermediate dates between the start and end of the gene flow. An upper bound for such a mixture comes from the lack of Iran Neolithic ancestry in Anatolian HGs at 13,000 BCE, and a lower bound comes from the C14 dates of early Anatolian farmers, one of which is directly dated at 8269–8210 BCE (Feldman et al., 2019). In either case (instantaneous admixture or gradual gene flow), the genetic mixture that formed Anatolian farmers predates the advent of agriculture in this region. This supports the model that Anatolian hunter-gatherers locally transitioned to agricultural subsistence, and most probably, there was cultural diffusion from other regions in Near East (Iran and Levant) (Feldman et al., 2019). Future studies with more dense temporal sampling will shed light on the demographic processes that led to the transition from foraging to farming in the Near East, and in turn, elucidate the relative roles of demic and cultural diffusion in the dispersal of technologies like agriculture across populations.

Using data from sixteen regions in Europe, we reconstruct a detailed chronology and dynamics of the expansion and admixture of Anatolian farmers during the Neolithic period. We infer that starting in ~6,400 BCE, gene flow from Anatolian farmers became widespread across Europe, and by ~4,300 BCE, it was present in almost all parts of continental Europe and the British Isles. These dates are significantly more recent than the estimates of farming based on archaeological evidence in some parts of Europe, suggesting that the local hunter-gatherers and farmers co-existed for more than a millennium before the mixture occurred (Haak et al., 2015; Lipson et al., 2017). In many regions, after the initial mixture, there was a resurgence of HG ancestry, highlighting the complexities of these ancient interactions. We note that our results are consistent with two previous genetic studies, Lipson et al. (2017) and Rivollat et al. (2020), that applied genetic dating methods to a subset of samples we used in our analysis. Lipson et al. (2017) used a modified version of ALDER to infer the timing of admixture in three regions (n=151), and we obtained statistically consistent results for all overlapping samples (within two standard errors) (Appendix 1—table 6). An advantage of our approach over the modified ALDER approach is that we do not rely on helper samples (higher coverage individuals combined with the target group) for dating; unless these have a similar ancestry profile, they could bias the inferred dates. Our results are concordant with Rivollat et al. (2020) that used a previous version of *DATES* to infer the timing of Neolithic gene flow in 32 groups (vs. 86 groups in our study). We find the performance of both versions of *DATES* is similar, though some implementation details have improved (Appendix 1 - table 1).

The second major migration occurred when populations associated with the Yamnaya culture in the Pontic-Caspian steppes expanded across Europe. Our analysis reveals the precise timing of the genetic formation of these early Steppe pastoralist groups–Yamnaya and

Afanasievo–occurred ~4,400–4,000 BCE. This estimate predates the archaeological evidence by more than a millennium (ANTHONY, 2007) and suggests the presence of an ancient "ghost" population of proto-Yamnaya around this time. Understanding the source and location of this ghost population will provide deep insights into the history of Pontic-Caspian cultures and the origin of Indo-European languages that have been associated to have spread with Steppe pastoralists ancestry to Europe and South Asia (Haak et al., 2015; Kassian et al., 2021). Starting in ~3,200 BCE, the Yamnaya-derived cultures of Corded Ware Complex and Bell Beaker complex spread westwards, bringing Steppe pastoralist-related ancestry to Europe. Our analysis reveals striking differences in the spread of these three cultures: the Yamnaya were genetically formed a millennium before the evidence for pastoralism, while CWC formation is coincident with the archaeological dates and similar across diverse regions separated by thousands of kilometers, suggesting a rapid spread after the initial formation of this group. In contrast, the formation and expansion of people with Steppe pastoralist-related ancestry associated with Bell Beakers cultural artifacts are much more complex and heterogeneous across regions. We find the earliest evidence of Steppe pastoralist-related ancestry in Iberia around 3200 BCE, though this ancestry only becomes widespread after 2,500 BCE. In central Europe, the gene flow occurred simultaneously with archaeological evidence and was coexisting with the Corded Ware complex in some parts (Müller, 2001; Olalde et al., 2018). Finally, in the British Isles, the Bell Beaker culture spreads rapidly from central Europe and replaces almost 90% of the ancestry of individuals in this region (Olalde et al., 2018).

Recent analysis has shown remarkable parallels in the history of Europe and South Asia; with both groups deriving ancestry from local indigenous HGs, Near Eastern farmers, and Steppe pastoralist-related groups (Narasimhan et al., 2019). Interestingly, however, the timing of the two major migrations events differs across the two subcontinents. Both mixtures occurred in Europe almost a millennium before they occurred in South Asia. In Europe, the Neolithic migrations primarily involved Anatolian farmers, while the source of Neolithic ancestry is closer to Iran Neolithic farmers in South Asia. The Steppe pastoralist-related gene flow occurred in the context of the spread of CWC and BBC cultures in Europe around 3,200-2,500 BCE; in South Asia, this ancestry arrived with Steppe MLBA cultures in 1,800-1,500 BCE (Narasimhan et al., 2019). The Steppe MLBA groups were genetically formed as an admixture of Steppe pastoralist derived groups and European Neolithic farmers following the eastward expansion of CWC groups between ~3,200–2,700 BCE. Understanding the origin and migration paths of the ancestral groups thus helps to illuminate the differences in the timeline of the spread of Steppe pastoralists across the two subcontinents of Eurasia.

Genomic dating methods like *DATES* provide an independent and complementary approach for reconstructing population history. By focusing on the genetic clock based on recombination rate, we provide an independent estimate of the timing of evolutionary events up to several thousands of years. Our analysis also has advantages over the temporal sampling of ancient DNA, in that we can obtain direct estimates of when a population was formed, rather than inferring putative bounds for the timing based on the absence/presence of a particular ancestry signature (which may be sensitive to sampling choice or density). Genetic approaches provide

complementary evidence to archaeology and linguistics as they date the time of admixture and not migration. Both dates are similar in many contemporary populations like African Americans and Latinos, though this may not be generally true (Hellenthal et al., 2014). This is underscored by our dates for the Neolithic farmer mixture, which postdates evidence of material culture related to agriculture by almost two millennia in some regions. This suggests that European HGs and farmers resided side by side for several thousand years before mixing (Bollongino et al., 2013; Skoglund et al., 2014). This underscores how genetic dates can provide complementary evidence to archaeology and help to build a comprehensive picture of population origins and movements.

Methods and Materials

Dataset

We analyzed 1,096 ancient European samples from 152 groups restricting to data from 1,233,013 autosomal SNP positions that were genotyped using the Affymetrix Human Origins array (the V44.3 release of the Allen Ancient DNA Resource (AADR); https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data). We filtered this dataset to remove samples that were marked as contaminated, low coverage, outliers, duplicates or first- or second-degree relatives. We grouped individuals together from a particular culture or region. Details of sample affiliation and grouping used is described in Supplementary file 1A.

Modeling admixture history

We applied *qpAdm* from ADMIXTOOLS to identify the best fitting model and estimate the ancestry proportions in a target population modeled as a mixture of *n* "reference" populations using a set of "Outgroup" populations (Haak et al., 2015). We set the details: YES parameter, which reports a normally distributed Z-score to evaluate the goodness of fit of the model (estimated with a Block Jackknife). For each target population, we chose the most parsimonious model, i.e., fitting the data with the minimum number of source populations. We excluded models where the *p*-value < 0.05 indicating a poor fit to the data. Details of the *qpAdm* analysis for each group are reported in Supplementary file 2. We also applied *D*-statistics in some cases using *qpDstat* in ADMIXTOOLS with default parameters.

DATES: model and implementation

DATES leverages the weighted ancestry covariance patterns across the genome of an admixed individual to infer the time of admixture. This method extends the idea introduced in ROLLOFF and ALDER and ref. (Moorjani et al., 2016) to be applicable to dating admixture events between modern human populations using a single genome (Moorjani et al., 2016).

Basic model and notation

Assume we have an admixed individual *C* with ancestry from source populations *A* and *B*, with ancestry proportion of α and $\beta = (1-\alpha)$ respectively. This mixture occurred t generations ago. First, we model the genotypes of *C* as a linear mix of allele frequencies of populations *A* and *B*. For any SNP *i*, let the genotype of *C* be g_i and allele frequency in *A* and *B* be $p_A(i)$ and $p_B(i)$. We can then infer the mixing fraction α from population *A* by solving the simple linear regression by minimizing the residuals.

$$R = \sum_i (g_i - (\alpha p_A(i) + (1-\alpha) p_B(i)))^2 \quad (1)$$

Let a_i be the probability of observing g_i in *C* given the observed genotype in *A,* and b_i be the probability of observing g_i in *C* given the observed genotype in *B*

$$a_i = P(g_i|A)$$
$$b_i = P(g_i|B)$$

We can then compute the likelihood L_i of observing a genotype g_i in the admixed individual

$$L_i = \alpha a_i + \beta b_i \quad (2)$$

For a pair of neighboring markers S_1, S_2 located at a genetic distance of *d* Morgans, the probability of no recombination between the two markers is given by $\theta = e^{-td}$. Accounting for recombination, the log likelihood that the two markers have the same ancestry is then given by:

$$\mathcal{L} = \log\left[(1-\theta)L_1 L_2 + \theta(\alpha a_1 a_2 + \beta b_1 b_2)\right] \quad (3)$$

Let K_i represent the ancestry at marker S_i. Expanding as a power series in θ, the coefficient of θ is QK_1K_2, where

$$Q = \alpha\beta$$
$$K_i = \frac{(a_i - b_i)}{L_i} \quad (4)$$

We can compute the ancestry covariance, *A(d)*, across pairs of markers S_1, S_2 separated by distance *d* as

$$A(d) = \frac{\sum_{S(d)} (K_1 - \overline{K_1})(K_2 - \overline{K_2})}{|S(d)|}$$

where *S(d)* is a set of markers S_1, S_2 located *d* Morgans apart.

The ancestry covariance $A(d)$ is expected to follow an exponential decay with *d* with the rate of decay depending on the time since admixture (t)

$$A(d) \sim e^{-(t+1)d}$$

The factor of (t+1) is because, in the first-generation following admixture, the admixed population derives one chromosome from each ancestral group. The mixing of chromosomes only begins in

799 the following generations as the chromosomes recombine. This means that if we fit t generations,
800 we are likely to underestimate the time of admixture. We note that previous methods like ALDER
801 and ROLLOFF, however, incorrectly fit t generations to infer the time of mixture. In practice,
802 however, this has little effect on the inference except for every recent admixture dates. We infer
803 the time of the mixture by fitting an exponential distribution with affine term using least squares.
804 *DATES* is applicable for dating admixture in a single individual. When multiple individuals from
805 an admixed population are available, *DATES* computes the log-likelihood by summing over all
806 individuals.
807
808 *Application to real data*
809 We applied *DATES* using genome-wide SNP data from the target population and two reference
810 populations. To infer the allele frequency in the ancestral populations more reliably, where
811 specified, we pooled individuals deriving the majority of their ancestry from the population of
812 interest (Supplementary file 1A). We computed the weighted ancestry covariance between 0.45cM
813 (to minimize the impact of background LD) to 100 cM, with a bin size of 0.1 cM. We plotted the
814 weighted covariance with genetic distance and obtained a date by fitting an exponential function
815 with an affine term $y = Ae^{-\lambda d} + c$, where d is the genetic distance in Morgans and $\lambda = (t+1)$ is
816 the number of generations since admixture (t). We computed standard errors using weighted block
817 jackknife, where one chromosome was removed in each run (Busing et al., 1999). We examined
818 the quality of the exponential fit by computing the normalized root-mean-square deviation
819 (NRMSD) between the empirical ancestry covariance values z and the fitted ones $\hat{z}$, across all the
820 genetic distance bins (Tournebize et al., 2020).
821
822 The estimated dates of admixture were considered <u>significant</u> if the Z-score > 2, λ < 200
823 generations and NRMSD < 0.7. We converted the inferred dates from generations to years by
824 assuming a mean generation time of 28 years (Moorjani et al., 2016). For ancient samples, we
825 added the sampling age of the ancient specimen (Supplementary file 1A). When multiple
826 individuals were available, we used the average sampling ages to offset the admixture dates. We
827 report dates in BCE by assuming the 1950 convention.
828
829 **Comparison of old and new version of *DATES***
830
831 An earlier version of *DATES* (version v753) was released in (Narasimhan et al., 2019). The current
832 method (version 4010) released in this study differs in some key aspects of the implementation as
833 described below:
834 <u>a) Use of regression model vs. likelihood approach</u>: In v753, we used a regression model to infer
835 the residuals at each site in the genotype by conditioning on the allele frequency in the reference
836 population and the genome-wide estimate of the admixture proportion (Narasimhan et al., 2019).
837 In contrast, in the current version (v4010), we use a more rigorous likelihood framework where
838 we infer the probability of ancestry from each reference population at each site in the genome
839 (equation (3).

b) Rate of decay of exponential fit: In v753, like ALDER and Rolloff, we fit an exponential decay with the rate of t generations. However, this assumes that mosaic chromosomes are formed in the generation when the gene flow occurs. However, in reality, the mixing of ancestry only begins in the following generations as the chromosomes of distinct ancestry recombine. To correctly account for this effect, we fit an exponential with the rate of (t+1) in *DATES* v4010. In practice, this has a minor effect on the dates reported earlier, as in most cases the uncertainty is much larger than one generation.

c) Goodness of fit test: In v4010, we implemented the NRMSD to assess the fit of the exponential curve. NRMSD computes the deviation between the empirical estimate and fitted data in order to provide a statistical way to characterize the noisiness of the fitted curve. Lower values of NRMSD suggest a better fit, however, there is no clear interpretation of the absolute value of NRMSD. Based on the empirical distribution of NRMSD values in our study samples (Appendix 1—figure 3), we infer a conservative threshold of 0.7 to define a "good" fit. We caution that users should adjust this threshold based on their application and always visually inspect their exponential fits to ensure reliable results.

(d) Support for arbitrary number of chromosomes: Unlike v753 that was optimized for parameters in humans, the new version supports an arbitrary number of chromosomes (inputted by the user) so *DATES* can be used in any species

A comparison of the two version of *DATES* using simulated data (Appendix 1—table 1) and empirical data (Appendix 1—table 2, Appendix 1—table 3) yields qualitatively similar results

Simulations

We constructed admixed genomes following the approach described in (Moorjani et al., 2011). This method requires phased haplotypes from two source populations and uses two key parameters to simulate data from admixed individuals, (a) the mixture proportion (α) that represents the probability that a particular sampled haplotype comes from one of the reference panels, namely $source_1$ and $source_2$, and (b) the time of mixture (t) which is the number of generations since mixture. To simulate an admixed individual, we begin at the start of the chromosome and sample a haplotype from either $source_1$ with a probability (α) and $source_2$ with a probability $(1-\alpha)$. At each subsequent marker, we check if there was a recombination event between the two neighboring markers. A recombination event occurs with a probability of $(1-e^{-\lambda g})$, where g is the genetic distance in Morgans. We use the time of $\lambda = (t+1)$ generations to account for the fact that in the first-generation following admixture, the offspring inherits one chromosome of each ancestry. In the next generation, the crossovers lead to a mixing of ancestry. Thus, when a recombination event occurs, we resample the ancestry between $source_1$ or $source_2$, otherwise, we copy the haplotype from the same source population (Note, a recombination event can lead to a switch to a haplotype

of the same ancestry). Once the ancestry is chosen, we randomly pick a haplotype from the ancestral pool (without replacement) and copy its sequence to the genome of the admixed individual. This process is continued until we reach the end of the chromosome. Using this approach, we generate the genomes of n admixed individuals. The simulated haploid chromosomes are merged at random to construct diploid admixed individuals. This algorithm requires more than $2n$ ancestral haplotypes for generating data for n diploid admixed individuals (Moorjani et al., 2011). For more than two reference populations, the same algorithm is repeated iteratively. We used 111 CEU and 112 YRI phased 1000 genomes phase 3 dataset (Auton et al., 2015) for generating 10 admixed genomes (unless otherwise stated) for ~380,000 SNPs. For the inference, we used French and Yoruba from Human Genome Diversity Panel (HGDP) (Rosenberg et al., 2002). We generated data for various demographic scenarios, where we varied the time the admixture (t), proportion of mixture (α), sample size in the reference and target populations, divergence between the ancestral and reference populations used and studied their impact on the estimated dates. We also characterized the impact of features of ancient DNA such as missing data, pseudo-haploid genotypes and limited sample size. In order to simulate pseudo-haploid genotypes, we randomly sampled an allele at each heterozygous site and assigned it as the homozygous genotype at that site (Harney et al., 2021). To generate missing data, we set the genotype call at a site as "missing" or "unknown" (in eigenstrat format as *9*) where the proportion of missing genotypes ranged between 5–60% in our simulations. We also evaluated the impact of the reference populations used in *DATES* in case of simple and multiple pulses of admixture.

To study the impact of complex scenarios of admixture involving founder events and continuous gene flow, we used a coalescent simulator, MaCs (Chen et al., 2008). We simulated 100Mb of three populations with an effective population size of 12,500, mutation rate of 1.2×10^{-8} and recombination rate 1×10^{-8} per base pair per generation respectively (Halldorsson et al., 2019; Jónsson et al., 2017). We assumed the admixture occurred continuously over a period of time or was followed by the bottleneck. In case of the latter, the duration of the bottleneck was 1-10 generations with reduction in effective population size from 12,500 to 10–1,000 and the population recovered to its original size after the bottleneck or maintained a small size until present (no recovery founder event). For each simulation, we generated data for two haploid chromosomes and combined these to generate one diploid chromosome.

Software availability

The executable and source code for *DATES* will be available on GitHub:
https://github.com/MoorjaniLab/DATES_v4010

Data Availability

The following previously published datasets were used in the study

The Allen Ancient DNA Resource (AADR) version V44.3
https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data

1000 genomes project, 1000 Genomes phase 3 release (Auton et al., 2015).

Human Genome Diversity Project (HGDP) (Rosenberg et al., 2002).

Funding

The Burroughs Wellcome Fund
Priya Moorjani

The Sloan Research fellowship
Priya Moorjani

National Institutes of Health (NIH R35GM142978)
Priya Moorjani

Fellowship through the Radcliffe Institute for Advanced Study at Harvard University.
Nick Patterson

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgments

We thank Monty Slatkin, Ziyue Gao, David Reich, Iosif Lazaridis, and Vagheesh Narasimhan for their comments on the manuscript. We thank Iosif Lazaridis for helpful discussions about population models in the Near East, Remi Tournebize for suggestions for evaluating the fit of exponential decay curves, and Neel Alex for suggestions for implementation of FFT for an earlier version of DATES.

Competing interests

The authors declare that no competing interests exist.

Additional files

Figure 1—figure supplements.

Figure 1—figure supplement 1. Varying time of admixture up to 300 generations. We simulated data for 10 admixed individuals with 20% European (CEU) and 80% African (YRI) ancestry and varied the time of admixture between 10 to 300 generations. The X-axis shows the true time of admixture, and the Y-axis shows the estimated time of admixture (± 1 SE) inferred using *DATES*.

Figure 1—figure supplement 2. Impact of sample size of the target (admixed) and reference populations. We simulated n admixed individuals with 20% European (CEU) and 80% African (YRI) ancestry and applied *DATES* with *m* reference samples of French and Yoruba ancestry. (A) Effect of sample size of target population. Each panel shows the results of simulations with *n* target individuals shown in the legend and *m* = 28 French and *m* = 21 Yoruba reference samples from each source group. (B) Effect of sample size (*m*) of reference populations. Each panel shows the results of simulations with *n*=10 target individuals and *m* reference samples from each source group shown in the legend. The true admixture time is shown on X-axis, and the estimated time of admixture (± 1 SE) is shown on Y-axis.

Figure 1—figure supplement 3. Impact of admixture proportion. We simulated data for 10 admixed individuals with European (CEU) ancestry (α) in the range of 1–40% (the rest derived from Africans). We ran *DATES* to infer the time of admixture and ancestry proportion. (A) Impact on the estimated time of admixture: Each panel shows the estimated date of admixture for a different value of α shown in the legend. The true admixture time is shown on X-axis, and the estimated time of admixture (± 1 SE) is shown on Y-axis. (B) Impact on estimated ancestry proportion: Each panel shows the estimated proportion of admixture for a different value of α shown in the legend. The red dashed horizontal line further indicates the value of α used. The true time of admixture is shown on X-axis with the inferred proportion of admixture on Y-axis.

Figure 1—figure supplement 4. Impact of divergence between the ancestral population and reference populations used in *DATES*. We simulated 10 admixed individuals with 80% European (CEU) and 20% African (YRI) ancestry. We applied *DATES* to infer the timing of admixture using reference populations. In each panel, we show the estimated dates of admixture using French and a group that is increasingly divergent from Yoruba (shown in the legend as the F_{ST} with Yoruba). The true admixture time is shown on X-axis, and the estimated time of admixture (± 1 SE) is shown on Y-axis.

Figure 1—figure supplement 5. Impact of divergence between the two source populations. We simulated *n* admixed individuals with 20% European (CEU) and 80% ancestry from a range of populations with increasing relatedness to Europeans (shown in the legend as the F_{ST} to Europeans). Specifically, the other reference population we used was either West Africans (YRI), East Asians (CHB), South Americans (MXL) or Southern Europeans (TSI). We used the following reference populations for the inference: French (for all simulations) with one of the other

1001 references as either Yoruba, Tujia, Maya or Italian respectively. We show results for varying target
1002 sample sizes of (A) n = 10 and (B) n = 1. The true admixture time is shown on X-axis, and the
1003 estimated time of admixture (± 1 SE) is shown on Y-axis. We note the inferred dates for CEU/TSI
1004 mixtures were not significant for older timescales and hence not shown.
1005
1006 **Figure 1—figure supplement 6. Impact of using the admixed individuals themselves as one**
1007 **of the reference groups in *DATES*.** We simulated data for 10 admixed individuals with European
1008 ancestry (α) in the range of 20–80% (the rest derived from Africans) using CEU and YRI as
1009 reference populations. Using a non-overlapping set of CEU and YRI individuals, we generated 10
1010 additional individuals that we used as reference samples in *DATES*. For each simulation, we ran
1011 *DATES* with Europeans (French) and a non-overlapping set of simulated admixed individuals as
1012 the reference populations (shown in blue), or Yoruba and simulated admixed individuals (shown
1013 in orange). The true admixture time is shown on X-axis, and the estimated time of admixture (± 1
1014 SE) is shown on Y-axis.
1015
1016 **Figure 1—figure supplement 7. Impact of sample size and data quality of target samples.** We
1017 simulated data for n admixed individuals with 20% European (CEU) and 80% African (YRI)
1018 ancestry. In each panel, we varied three key features of the data from the target population, notably
1019 sample size (n = 1 or 10), type of genotypes (diploid or pseudo-haploid) and missing genotype rate
1020 (between 10–60%). (A) Diploid genotypes with missing data for n=10 admixed individuals. Each
1021 panel shows the results of x% of missing diploid genotypes (shown in the legend). (B) Pseudo-
1022 haploid genotypes with missing data for n=10 admixed individuals. Each panel shows the results
1023 of x% of missing pseudo-haploid genotypes (shown in the legend) (C) Pseudo-haploid genotypes
1024 with missing data for n=1 admixed individuals. Each panel shows the results of x% of missing
1025 pseudo-haploid genotypes (shown in the legend). The true admixture time is shown on X-axis, and
1026 the estimated time of admixture (± 1 SE) is shown on Y-axis.
1027
1028 **Figure 1—figure supplement 8. Impact of data quality of target and reference populations**
1029 **as a function of divergence between true and reference populations used in *DATES*.** We
1030 simulated data for $n=10$ admixed individuals with 20% European (CEU) and 80% African (YRI)
1031 ancestry with pseudo-haploid genotypes. The reference populations used also had pseudo-haploid
1032 genotypes. We further varied three key features of the data, missing genotype rate in reference
1033 populations, missing genotype rate in target populations and divergence between true source
1034 populations and reference population used for the analysis. In each row, we show the admixture
1035 dates using reference populations with increasing divergence to true source population (F_{ST} shown
1036 in the row title). In each column, we varied the missing genotype rate in the target population
1037 (shown in the column title). Further, each panel shows results of missing data in the reference
1038 genomes (shown in the legend). (a) Reference populations of French and Yoruba (F_{ST}(true,
1039 reference) ~ 0). (b) Reference populations of French and Bantu Kenya (F_{ST}(true, reference) ~
1040 0.009). (c) Reference populations of French and San (F_{ST}(true, reference) ~ 0.103). The true

admixture time is shown on X-axis, and the estimated time of admixture (± 1 SE) is shown on Y-axis.

Figure 1—figure supplement 9. Impact of small sample size and data quality of target and reference populations as a function of divergence between true and reference populations used in *DATES*. We simulated data for *n=1* admixed individuals with 20% European (CEU) and 80% African (YRI) ancestry with pseudo-haploid genotypes. The reference populations used also had pseudo-haploid genotypes. We further varied three key features of the data, missing genotype rate in reference populations, missing genotype rate in target populations and divergence between true source populations and reference population used for the analysis. In each row, we show the admixture dates using reference populations with increasing divergence to true source population (F_{ST} shown in the row title). In each column, we varied the missing genotype rate in the target population (shown in the column title). Further, each panel shows results of missing data in the reference genomes (shown in the legend). (a) Reference populations of French and Yoruba (F_{ST}(true, reference) ~ 0). (b) Reference populations of French and Bantu Kenya (F_{ST}(true, reference) ~ 0.009). (c) Reference populations of French and San (F_{ST}(true, reference) ~ 0.103). The true admixture time is shown on X-axis, and the estimated time of admixture (± 1 SE) is shown on Y-axis.

Figure 3 —figure supplements.

Figure 3—figure supplement 1. *DATES* ancestry covariance decay curves. We show the weighted ancestry covariance decay curves generated using *DATES* for all the target groups analyzed in the study. Each subplot shows the decay curve for one target population with the associated reference groups shown in the title. For each target, in the legend, we show the inferred average dates of admixture (± 1 SE) in generations before the individual lived, in BCE that accounts for the average age of all the individuals in the target and the mean generation time of human populations (see Methods). We also show the NRMSD values for all fitted curves and the plots with NRMSD > 0.7 are shown in grey. For consistency, we use the same colors as Figure 3.

Figure 3—figure supplement 2. Timing of WHG and EHG admixture in Iron Gates HG samples. The time of admixture in Iron Gates HG samples. grouped in bins of C14 age of 500 years. The C14 age is shown on X-axis and the admixture time in BCE for corresponding samples is shown on the Y-axis.

Supplementary file 1- Data and admixture dates inferred using *DATES* for European groups during the Holocene (Excel sheet)

A) Information on ancient samples used in our study
B) Estimated dates of admixture for population mixture events during the European Holocene.

Supplementary file 2 -Formal tests of admixture for populations in Europe using *qpAdm* and D-statistics with default parameters in ADMIXTOOLS (Excel sheet)

A) Modeling population admixture of HG groups using *qpAdm* in ADMIXTOOLS.
B) D-statistics to assess the affinity of Mesolithic HG groups to WHG or Anatolian farmers
C) Modeling population admixture of Near Eastern farmers using *qpAdm* in ADMIXTOOLS.
D) Modeling population admixture of Neolithic European groups using *qpAdm* in ADMIXTOOLS.
E) Modeling population admixture of Neolithic European groups per individual using *qpAdm* in ADMIXTOOLS.
F) D-statistics to explore the affinity of the target groups to Steppe pastoralists or Anatolian Farmers.
G) Modeling population admixture of Early Steppe pastoralists groups using *qpAdm* in ADMIXTOOLS.
H) Genetic distance (FST) in early Steppe pastoralists groups.
I) Modeling population admixture of Bronze Age groups using *qpAdm* in ADMIXTOOLS
J) Modeling population admixture of Bronze Age groups per individual using *qpAdm* in ADMIXTOOLS.
K) Modeling population admixture of Middle Late Bronze Age (MLBA) Steppe pastoralists groups Age groups using *qpAdm* in ADMIXTOOLS.

References

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas A-S, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Casa PD, Dąbrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolář J, Kriiska A, Lasak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtchyan R, Moiseyev V, Paja L, Pálfi G, Pokutta D, Pospieszny Ł, Price TD, Saag L, Sablin M, Shishlina N, Smrčka V, Soenov VI, Szeverényi V, Tóth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Pontén T, Brunak S, Nielsen R, Kristiansen K, Willerslev E. 2015. Population genomics of Bronze Age Eurasia. *Nature* **522**:167–172. doi:10.1038/nature14507

ANTHONY DW. 2007. The Horse, the Wheel, and Language. Princeton University Press. doi:10.2307/j.ctt7sjpn

Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, Gabriel SB, Gibbs RA, Green ED, Hurles ME, Knoppers BM, Korbel JO, Lander ES, Lee C, Lehrach H, Mardis ER, Marth GT,

1124 McVean GA, Nickerson DA, Schmidt JP, Sherry ST, Wang J, Wilson RK, Gibbs RA,
1125 Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, Lee S, Muzny D, Reid JG,
1126 Zhu Y, Wang J, Chang Y, Feng Q, Fang X, Guo X, Jian M, Jiang H, Jin X, Lan T, Li G,
1127 Li J, Li Y, Liu S, Liu X, Lu Y, Ma X, Tang M, Wang B, Wang G, Wu H, Wu R, Xu X,
1128 Yin Y, Zhang D, Zhang W, Zhao J, Zhao M, Zheng X, Lander ES, Altshuler DM, Gabriel
1129 SB, Gupta N, Gharani N, Toji LH, Gerry NP, Resch AM, Flicek P, Barker J, Clarke L,
1130 Gil L, Hunt SE, Kelman G, Kulesha E, Leinonen R, McLaren WM, Radhakrishnan R,
1131 Roa A, Smirnov D, Smith RE, Streeter I, Thormann A, Toneva I, Vaughan B, Zheng-
1132 Bradley X, Bentley DR, Grocock R, Humphray S, James T, Kingsbury Z, Lehrach H,
1133 Sudbrak R, Albrecht MW, Amstislavskiy VS, Borodina TA, Lienhard M, Mertes F,
1134 Sultan M, Timmermann B, Yaspo M-L, Mardis ER, Wilson RK, Fulton L, Fulton R,
1135 Sherry ST, Ananiev V, Belaia Z, Beloslyudtsev D, Bouk N, Chen C, Church D, Cohen R,
1136 Cook C, Garner J, Hefferon T, Kimelman M, Liu C, Lopez J, Meric P, O'Sullivan C,
1137 Ostapchuk Y, Phan L, Ponomarov S, Schneider V, Shekhtman E, Sirotkin K, Slotta D,
1138 Zhang H, McVean GA, Durbin RM, Balasubramaniam S, Burton J, Danecek P, Keane
1139 TM, Kolb-Kokocinski A, McCarthy S, Stalker J, Quail M, Schmidt JP, Davies CJ, Gollub
1140 J, Webster T, Wong B, Zhan Y, Auton A, Campbell CL, Kong Y, Marcketta A, Gibbs
1141 RA, Yu F, Antunes L, Bainbridge M, Muzny D, Sabo A, Huang Z, Wang J, Coin LJM,
1142 Fang L, Guo X, Jin X, Li G, Li Q, Li Y, Li Z, Lin H, Liu B, Luo R, Shao H, Xie Y, Ye C,
1143 Yu C, Zhang F, Zheng H, Zhu H, Alkan C, Dal E, Kahveci F, Marth GT, Garrison EP,
1144 Kural D, Lee W-P, Fung Leong W, Stromberg M, Ward AN, Wu J, Zhang M, Daly MJ,
1145 DePristo MA, Handsaker RE, Altshuler DM, Banks E, Bhatia G, del Angel G, Gabriel
1146 SB, Genovese G, Gupta N, Li H, Kashin S, Lander ES, McCarroll SA, Nemesh JC,
1147 Poplin RE, Yoon SC, Lihm J, Makarov V, Clark AG, Gottipati S, Keinan A, Rodriguez-
1148 Flores JL, Korbel JO, Rausch T, Fritz MH, Stütz AM, Flicek P, Beal K, Clarke L, Datta
1149 A, Herrero J, McLaren WM, Ritchie GRS, Smith RE, Zerbino D, Zheng-Bradley X,
1150 Sabeti PC, Shlyakhter I, Schaffner SF, Vitti J, Cooper DN, Ball EV, Stenson PD, Bentley
1151 DR, Barnes B, Bauer M, Keira Cheetham R, Cox A, Eberle M, Humphray S, Kahn S,
1152 Murray L, Peden J, Shaw R, Kenny EE, Batzer MA, Konkel MK, Walker JA, MacArthur
1153 DG, Lek M, Sudbrak R, Amstislavskiy VS, Herwig R, Mardis ER, Ding L, Koboldt DC,
1154 Larson D, Ye K, Gravel S, The 1000 Genomes Project Consortium, Corresponding
1155 authors, Steering committee, Production group, Baylor College of Medicine, BGI-
1156 Shenzhen, Broad Institute of MIT and Harvard, Coriell Institute for Medical Research,
1157 European Molecular Biology Laboratory EBI, Illumina, Max Planck Institute for
1158 Molecular Genetics, McDonnell Genome Institute at Washington University, US
1159 National Institutes of Health, University of Oxford, Wellcome Trust Sanger Institute,
1160 Analysis group, Affymetrix, Albert Einstein College of Medicine, Bilkent University,
1161 Boston College, Cold Spring Harbor Laboratory, Cornell University, European Molecular
1162 Biology Laboratory, Harvard University, Human Gene Mutation Database, Icahn School
1163 of Medicine at Mount Sinai, Louisiana State University, Massachusetts General Hospital,
1164 McGill University, National Eye Institute N. 2015. A global reference for human genetic
1165 variation. *Nature* **526**:68–74. doi:10.1038/nature15393
1166 Baird D, Fairbairn A, Jenkins E, Martin L, Middleton C, Pearson J, Asouti E, Edwards Y,
1167 Kabukcu C, Mustafaoğlu G, Russell N, Bar-Yosef O, Jacobsen G, Wu X, Baker A, Elliott
1168 S. 2018. Agricultural origins on the Anatolian plateau. *Proc Natl Acad Sci* **115**:E3077–
1169 E3086. doi:10.1073/pnas.1800163115

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, Chaubey G, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parfitt T, Hammer MF, Skorecki K, Villems R. 2010. The genome-wide structure of the Jewish people. *Nature* **466**:238–242. doi:10.1038/nature09103

Bellwood P. 2005. First Farmers. The Origins of Agricultural Societies, (Wiley-Blackwell, Chichester. ed.

Bollongino R, Nehlich O, Richards MP, Orschiedt J, Thomas MG, Sell C, Fajkosová Z, Powell A, Burger J. 2013. 2000 years of parallel societies in Stone Age Central Europe. *Science* **342**:479–481. doi:10.1126/science.1245049

Brace S, Diekmann Y, Booth TJ, van Dorp L, Faltyskova Z, Rohland N, Mallick S, Olalde I, Ferry M, Michel M, Oppenheimer J, Broomandkhoshbacht N, Stewardson K, Martiniano R, Walsh S, Kayser M, Charlton S, Hellenthal G, Armit I, Schulting R, Craig OE, Sheridan A, Parker Pearson M, Stringer C, Reich D, Thomas MG, Barnes I. 2019. Ancient genomes indicate population replacement in Early Neolithic Britain. *Nat Ecol Evol* **3**:765–771. doi:10.1038/s41559-019-0871-9

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind C-J, Lueth F, Terberger T, Hiller J, Matsumura S, Forster P, Burger J. 2009. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**:137–140. doi:10.1126/science.1176869

Busing FMTA, Meijer E, Leeden RVD. 1999. Delete-m Jackknife for Unequal m. *Stat Comput* **9**:3–8.

Chakraborty R, Weiss KM. 1988. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci* **85**:9119–9123. doi:10.1073/pnas.85.23.9119

Chen GK, Marjoram P, Wall JD. 2008. Fast and flexible simulation of DNA sequence data. *Genome Res* **19**:136–142. doi:10.1101/gr.083634.108

D. Gronenborn,. 2014. The persistence of hunting and gathering: Neolithic western temperate and Central Europe, in The Oxford Handbook of the Archaeology and Anthropology of Hunter-Gatherers, V. Cummings, P. Jordan, M. Zvelebil, Eds. Oxford University Press.

Feldman M, Fernández-Domínguez E, Reynolds L, Baird D, Pearson J, Hershkovitz I, May H, Goring-Morris N, Benz M, Gresky J, Bianco RA, Fairbairn A, Mustafaoğlu G, Stockhammer PW, Posth C, Haak W, Jeong C, Krause J. 2019. Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat Commun* **10**:1218. doi:10.1038/s41467-019-09209-7

Fernandes DM, Mittnik A, Olalde I, Lazaridis I, Cheronet O, Rohland N, Mallick S, Bernardos R, Broomandkhoshbacht N, Carlsson J, Culleton BJ, Ferry M, Gamarra B, Lari M, Mah M, Michel M, Modi A, Novak M, Oppenheimer J, Sirak KA, Stewardson K, Mandl K, Schattke C, Özdoğan KT, Lucci M, Gasperetti G, Candilio F, Salis G, Vai S, Camarós E, Calò C, Catalano G, Cueto M, Forgia V, Lozano M, Marini E, Micheletti M, Miccichè RM, Palombo MR, Ramis D, Schimmenti V, Sureda P, Teira L, Teschler-Nicola M, Kennett DJ, Lalueza-Fox C, Patterson N, Sineo L, Coppa A, Caramelli D, Pinhasi R, Reich D. 2020. The spread of steppe and Iranian-related ancestry in the islands of the western Mediterranean. *Nat Ecol Evol* **4**:334–345. doi:10.1038/s41559-020-1102-0

Fernandes DM, Strapagiel D, Borówka P, Marciniak B, Żądzińska E, Sirak K, Siska V, Grygiel R, Carlsson J, Manica A, Lorkiewicz W, Pinhasi R. 2018. A genomic Neolithic time

transect of hunter-farmer admixture in central Poland. *Sci Rep* **8**:14879. doi:10.1038/s41598-018-33067-w

Fokkens, H. and Nicolis, F. 2012. Background to Beakers. Inquiries into regional cultural backgrounds of the Bell Beaker complex. Leiden: Sidestone Press.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, Nickel B, Peltzer A, Rohland N, Slon V, Talamo S, Lazaridis I, Lipson M, Mathieson I, Schiffels S, Skoglund P, Derevianko AP, Drozdov N, Slavinsky V, Tsybankov A, Cremonesi RG, Mallegni F, Gély B, Vacca E, Morales MRG, Straus LG, Neugebauer-Maresch C, Teschler-Nicola M, Constantin S, Moldovan OT, Benazzi S, Peresani M, Coppola D, Lari M, Ricci S, Ronchitelli A, Valentin F, Thevenet C, Wehrberger K, Grigorescu D, Rougier H, Crevecoeur I, Flas D, Semal P, Mannino MA, Cupillard C, Bocherens H, Conard NJ, Harvati K, Moiseyev V, Drucker DG, Svoboda J, Richards MP, Caramelli D, Pinhasi R, Kelso J, Patterson N, Krause J, Pääbo S, Reich D. 2016. The genetic history of Ice Age Europe. *Nature* **534**:200–205. doi:10.1038/nature17993

Gravel S. 2012. Population Genetics Models of Local Ancestry. *Genetics* **191**:607–619. doi:10.1534/genetics.112.139808

Günther T, Malmström H, Svensson EM, Omrak A, Sánchez-Quinto F, Kılınç GM, Krzewińska M, Eriksson G, Fraser M, Edlund H, Munters AR, Coutinho A, Simões LG, Vicente M, Sjölander A, Jansen Sellevold B, Jørgensen R, Claes P, Shriver MD, Valdiosera C, Netea MG, Apel J, Lidén K, Skar B, Storå J, Götherström A, Jakobsson M. 2018. Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLOS Biol* **16**:e2003703. doi:10.1371/journal.pbio.2003703

Haak W, Balanovsky O, Sanchez JJ, Koshel S, Zaporozhchenko V, Adler CJ, Der Sarkissian CSI, Brandt G, Schwarz C, Nicklisch N, Dresely V, Fritsch B, Balanovska E, Villems R, Meller H, Alt KW, Cooper A, the Genographic Consortium. 2010. Ancient DNA from European Early Neolithic Farmers Reveals Their Near Eastern Affinities. *PLoS Biol* **8**:e1000536. doi:10.1371/journal.pbio.1000536

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**:207–211. doi:10.1038/nature14317

Halldorsson BV, Palsson G, Stefansson OA, Jonsson H, Hardarson MT, Eggertsson HP, Gunnarsson B, Oddsson A, Halldorsson GH, Zink F, Gudjonsson SA, Frigge ML, Thorleifsson G, Sigurdsson A, Stacey SN, Sulem P, Masson G, Helgason A, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K. 2019. Characterizing mutagenic effects of recombination through a sequence-level genetic map. *Science* **363**:eaau1043. doi:10.1126/science.aau1043

Harney É, Patterson N, Reich D, Wakeley J. 2021. Assessing the performance of qpAdm: a statistical tool for studying population admixture. *Genetics* **217**:iyaa045. doi:10.1093/genetics/iyaa045

Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, Myers S. 2014. A Genetic Atlas of Human Admixture History. *Science* **343**:747–751. doi:10.1126/science.1243518

Henn BM, Botigué LR, Gravel S, Wang W, Brisbin A, Byrnes JK, Fadhlaoui-Zid K, Zalloua PA, Moreno-Estrada A, Bertranpetit J, Bustamante CD, Comas D. 2012. Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genet* **8**:e1002397. doi:10.1371/journal.pgen.1002397

Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-del-Molino D, van Dorp L, López S, Kousathanas A, Link V, Kirsanow K, Cassidy LM, Martiniano R, Strobel M, Scheu A, Kotsakis K, Halstead P, Triantaphyllou S, Kyparissi-Apostolika N, Urem-Kotsou D, Ziota C, Adaktylou F, Gopalan S, Bobo DM, Winkelbach L, Blöcher J, Unterländer M, Leuenberger C, Çilingiroğlu Ç, Horejs B, Gerritsen F, Shennan SJ, Bradley DG, Currat M, Veeramah KR, Wegmann D, Thomas MG, Papageorgopoulou C, Burger J. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proc Natl Acad Sci* **113**:6886–6891. doi:10.1073/pnas.1523951113

J. Guilaine,. 2003. Aspects de la néolithisation en Méditerranée et en France, in The Widening Harvest. The Neolithic Transition in Europe: Looking Back, Looking Forward, A. Ammerman, P. Biagi, Eds. Archaeological Institute of America.

Jones ER, Gonzalez-Fortes G, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego Llorente M, Cassidy LM, Gamba C, Meshveliani T, Bar-Yosef O, Müller W, Belfer-Cohen A, Matskevich Z, Jakeli N, Higham TFG, Currat M, Lordkipanidze D, Hofreiter M, Manica A, Pinhasi R, Bradley DG. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun* **6**:8912. doi:10.1038/ncomms9912

Jónsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, Ward LD, Arnadottir GA, Helgason EA, Helgason H, Gylfason A, Jonasdottir Adalbjorg, Jonasdottir Aslaug, Rafnar T, Frigge M, Stacey SN, Th. Magnusson O, Thorsteinsdottir U, Masson G, Kong A, Halldorsson BV, Helgason A, Gudbjartsson DF, Stefansson K. 2017. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**:519–522. doi:10.1038/nature24018

Kashuba N, Kırdök E, Damlien H, Manninen MA, Nordqvist B, Persson P, Götherström A. 2019. Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter–gatherers in Scandinavia. *Commun Biol* **2**:185. doi:10.1038/s42003-019-0399-1

Kassian AS, Zhivlov M, Starostin G, Trofimov AA, Kocharov PA, Kuritsyna A, Saenko MN. 2021. Rapid radiation of the inner Indo-European languages: an advanced approach to Indo-European lexicostatistics. *Linguistics* **59**:949–979. doi:10.1515/ling-2020-0060

Kılınç GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Bıçakçı E, Baird D, Dönertaş HM, Ghalichi A, Yaka R, Koptekin D, Açan SC, Parvizi P, Krzewińska M, Daskalaki EA, Yüncü E, Dağtaş ND, Fairbairn A, Pearson J, Mustafaoğlu G, Erdal YS, Çakan YG, Togan İ, Somel M, Storå J, Jakobsson M, Götherström A. 2016. The Demographic Development of the First Farmers in Anatolia. *Curr Biol* **26**:2659–2666. doi:10.1016/j.cub.2016.07.057

Lipson M, Szécsényi-Nagy A, Mallick S, Pósa A, Stégmár B, Keerl V, Rohland N, Stewardson K, Ferry M, Michel M, Oppenheimer J, Broomandkhoshbacht N, Harney E, Nordenfelt S, Llamas B, Gusztáv Mende B, Köhler K, Oross K, Bondár M, Marton T, Osztás A, Jakucs J, Paluch T, Horváth F, Csengeri P, Koós J, Sebők K, Anders A, Raczky P, Regenye J, Barna JP, Fábián S, Serlegi G, Toldi Z, Gyöngyvér Nagy E, Dani J, Molnár E, Pálfi G, Márk L, Melegh B, Bánfai Z, Domboróczki L, Fernández-Eraso J, Antonio Mujika-Alustiza J, Alonso Fernández C, Jiménez Echevarría J, Bollongino R, Orschiedt

J, Schierhold K, Meller H, Cooper A, Burger J, Bánffy E, Alt KW, Lalueza-Fox C, Haak W, Reich D. 2017. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**:368–372. doi:10.1038/nature24476

Loh P-R, Lipson M, Patterson N, Moorjani P, Pickrell JK, Reich D, Berger B. 2013. Inferring Admixture Histories of Human Populations Using Linkage Disequilibrium. *Genetics* **193**:1233–1254. doi:10.1534/genetics.112.147330

Malmström H, Günther T, Svensson EM, Juras A, Fraser M, Munters AR, Pospieszny Ł, Tõrv M, Lindström J, Götherström A, Storå J, Jakobsson M. 2019. The genomic ancestry of the Scandinavian Battle Axe Culture people and their relation to the broader Corded Ware horizon. *Proc R Soc B Biol Sci* **286**:20191528. doi:10.1098/rspb.2019.1528

Marija Gimbutas. 2001. The Living Goddesses. University of California Press.

Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, Broomandkhoshbacht N, Candilio F, Cheronet O, Fernandes D, Ferry M, Gamarra B, Fortes GG, Haak W, Harney E, Jones E, Keating D, Krause-Kyora B, Kucukkalipci I, Michel M, Mittnik A, Nägele K, Novak M, Oppenheimer J, Patterson N, Pfrengle S, Sirak K, Stewardson K, Vai S, Alexandrov S, Alt KW, Andreescu R, Antonović D, Ash A, Atanassova N, Bacvarov K, Gusztáv MB, Bocherens H, Bolus M, Boroneanţ A, Boyadzhiev Y, Budnik A, Burmaz J, Chohadzhiev S, Conard NJ, Cottiaux R, Čuka M, Cupillard C, Drucker DG, Elenski N, Francken M, Galabova B, Ganetsovski G, Gély B, Hajdu T, Handzhyiska V, Harvati K, Higham T, Iliev S, Janković I, Karavanić I, Kennett DJ, Komšo D, Kozak A, Labuda D, Lari M, Lazar C, Leppek M, Leshtakov K, Vetro DL, Los D, Lozanov I, Malina M, Martini F, McSweeney K, Meller H, Menđušić M, Mirea P, Moiseyev V, Petrova V, Price TD, Simalcsik A, Sineo L, Šlaus M, Slavchev V, Stanev P, Starović A, Szeniczey T, Talamo S, Teschler-Nicola M, Thevenet C, Valchev I, Valentin F, Vasilyev S, Veljanovska F, Venelinova S, Veselovskaya E, Viola B, Virag C, Zaninović J, Zäuner S, Stockhammer PW, Catalano G, Krauß R, Caramelli D, Zariņa G, Gaydarska B, Lillie M, Nikitin AG, Potekhina I, Papathanasiou A, Borić D, Bonsall C, Krause J, Pinhasi R, Reich D. 2018. The genomic history of southeastern Europe. *Nature* **555**:197–203. doi:10.1038/nature25778

Mittnik A, Wang C-C, Pfrengle S, Daubaras M, Zariņa G, Hallgren F, Allmäe R, Khartanovich V, Moiseyev V, Tõrv M, Furtwängler A, Andrades Valtueña A, Feldman M, Economou C, Oinonen M, Vasks A, Balanovska E, Reich D, Jankauskas R, Haak W, Schiffels S, Krause J. 2018. The genetic prehistory of the Baltic Sea region. *Nat Commun* **9**:442. doi:10.1038/s41467-018-02825-9

Moorjani P, Patterson N, Hirschhorn JN, Keinan A, Hao L, Atzmon G, Burns E, Ostrer H, Price AL, Reich D. 2011. The History of African Gene Flow into Southern Europeans, Levantines, and Jews. *PLoS Genet* **7**:e1001373. doi:10.1371/journal.pgen.1001373

Moorjani P, Sankararaman S, Fu Q, Przeworski M, Patterson N, Reich D. 2016. A genetic method for dating ancient genomes provides a direct estimate of human generation interval in the last 45,000 years. *Proc Natl Acad Sci* **113**:5652. doi:10.1073/pnas.1514696113

Morgunova NL, Khokhlova OS. 2013. Chronology and Periodization of the Pit-Grave Culture in the Region Between the Volga and Ural Rivers Based on Radiocarbon Dating and Paleopedological Research. *Radiocarbon* **55**:1286–1296. doi:10.1017/S0033822200048190

Müller Johannes Willigen, Samuel van,. 2001. New radiocarbon evidence for european Bell Beakers and the consequences for the diffusion of the Bell Beaker Phenomenon. *Bell Beakers Today* **1**.

Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, Lazaridis I, Nakatsuka N, Olalde I, Lipson M, Kim AM, Olivieri LM, Coppa A, Vidale M, Mallory J, Moiseyev V, Kitov E, Monge J, Adamski N, Alex N, Broomandkhoshbacht N, Candilio F, Callan K, Cheronet O, Culleton BJ, Ferry M, Fernandes D, Freilich S, Gamarra B, Gaudio D, Hajdinjak M, Harney É, Harper TK, Keating D, Lawson AM, Mah M, Mandl K, Michel M, Novak M, Oppenheimer J, Rai N, Sirak K, Slon V, Stewardson K, Zalzala F, Zhang Z, Akhatov G, Bagashev AN, Bagnera A, Baitanayev B, Bendezu-Sarmiento J, Bissembaev AA, Bonora GL, Chargynov TT, Chikisheva T, Dashkovskiy PK, Derevianko A, Dobeš M, Douka K, Dubova N, Duisengali MN, Enshin D, Epimakhov A, Fribus AV, Fuller D, Goryachev A, Gromov A, Grushin SP, Hanks B, Judd M, Kazizov E, Khokhlov A, Krygin AP, Kupriyanova E, Kuznetsov P, Luiselli D, Maksudov F, Mamedov AM, Mamirov TB, Meiklejohn C, Merrett DC, Micheli R, Mochalov O, Mustafokulov S, Nayak A, Pettener D, Potts R, Razhev D, Rykun M, Sarno S, Savenkova TM, Sikhymbaeva K, Slepchenko SM, Soltobaev OA, Stepanova N, Svyatko S, Tabaldiev K, Teschler-Nicola M, Tishkin AA, Tkachev VV, Vasilyev S, Velemínský P, Voyakin D, Yermolayeva A, Zahir M, Zubkov VS, Zubova A, Shinde VS, Lalueza-Fox C, Meyer M, Anthony D, Boivin N, Thangaraj K, Kennett DJ, Frachetti M, Pinhasi R, Reich D. 2019. The formation of human populations in South and Central Asia. *Science* **365**:eaat7487. doi:10.1126/science.aat7487

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, Rohland N, Mallick S, Szécsényi-Nagy A, Mittnik A, Altena E, Lipson M, Lazaridis I, Harper TK, Patterson N, Broomandkhoshbacht N, Diekmann Y, Faltyskova Z, Fernandes D, Ferry M, Harney E, de Knijff P, Michel M, Oppenheimer J, Stewardson K, Barclay A, Alt KW, Liesau C, Ríos P, Blasco C, Miguel JV, García RM, Fernández AA, Bánffy E, Bernabò-Brea M, Billoin D, Bonsall C, Bonsall L, Allen T, Büster L, Carver S, Navarro LC, Craig OE, Cook GT, Cunliffe B, Denaire A, Dinwiddy KE, Dodwell N, Ernée M, Evans C, Kuchařík M, Farré JF, Fowler C, Gazenbeek M, Pena RG, Haber-Uriarte M, Haduch E, Hey G, Jowett N, Knowles T, Massy K, Pfrengle S, Lefranc P, Lemercier O, Lefebvre A, Martínez CH, Olmo VG, Ramírez AB, Maurandi JL, Majó T, McKinley JI, McSweeney K, Mende BG, Modi A, Kulcsár G, Kiss V, Czene A, Patay R, Endrődi A, Köhler K, Hajdu T, Szeniczey T, Dani J, Bernert Z, Hoole M, Cheronet O, Keating D, Velemínský P, Dobeš M, Candilio F, Brown F, Fernández RF, Herrero-Corral A-M, Tusa S, Carnieri E, Lentini L, Valenti A, Zanini A, Waddington C, Delibes G, Guerra-Doce E, Neil B, Brittain M, Luke M, Mortimer R, Desideri J, Besse M, Brücken G, Furmanek M, Hałuszko A, Mackiewicz M, Rapiński A, Leach S, Soriano I, Lillios KT, Cardoso JL, Pearson MP, Włodarczak P, Price TD, Prieto P, Rey P-J, Risch R, Rojo Guerra MA, Schmitt A, Serralongue J, Silva AM, Smrčka V, Vergnaud L, Zilhão J, Caramelli D, Higham T, Thomas MG, Kennett DJ, Fokkens H, Heyd V, Sheridan A, Sjögren K-G, Stockhammer PW, Krause J, Pinhasi R, Haak W, Barnes I, Lalueza-Fox C, Reich D. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**:190–196. doi:10.1038/nature25738

Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, Dulias K, Edwards CJ, Gandini F, Pala M, Soares P, Ferrando-Bernal M, Adamski N, Broomandkhoshbacht N, Cheronet O, Culleton BJ, Fernandes D, Lawson AM, Mah M, Oppenheimer J,

Stewardson K, Zhang Z, Jiménez Arenas JM, Toro Moyano IJ, Salazar-García DC, Castanyer P, Santos M, Tremoleda J, Lozano M, García Borja P, Fernández-Eraso J, Mujika-Alustiza JA, Barroso C, Bermúdez FJ, Viguera Mínguez E, Burch J, Coromina N, Vivó D, Cebrià A, Fullola JM, García-Puchol O, Morales JI, Oms FX, Majó T, Vergès JM, Díaz-Carvajal A, Ollich-Castanyer I, López-Cachero FJ, Silva AM, Alonso-Fernández C, Delibes de Castro G, Jiménez Echevarría J, Moreno-Márquez A, Pascual Berlanga G, Ramos-García P, Ramos-Muñoz J, Vijande Vila E, Aguilella Arzo G, Esparza Arroyo Á, Lillios KT, Mack J, Velasco-Vázquez J, Waterman A, Benítez de Lugo Enrich L, Benito Sánchez M, Agustí B, Codina F, de Prado G, Estalrrich A, Fernández Flores Á, Finlayson C, Finlayson G, Finlayson S, Giles-Guzmán F, Rosas A, Barciela González V, García Atiénzar G, Hernández Pérez MS, Llanos A, Carrión Marco Y, Collado Beneyto I, López-Serrano D, Sanz Tormo M, Valera AC, Blasco C, Liesau C, Ríos P, Daura J, de Pedro Michó MJ, Diez-Castillo AA, Flores Fernández R, Francès Farré J, Garrido-Pena R, Gonçalves VS, Guerra-Doce E, Herrero-Corral AM, Juan-Cabanilles J, López-Reyes D, McClure SB, Merino Pérez M, Oliver Foix A, Sanz Borràs M, Sousa AC, Vidal Encinas JM, Kennett DJ, Richards MB, Werner Alt K, Haak W, Pinhasi R, Lalueza-Fox C, Reich D. 2019. The genomic history of the Iberian Peninsula over the past 8000 years. *Science* **363**:1230–1234. doi:10.1126/science.aav4040

Pagani L, Colonna V, Tyler-Smith C, Ayub Q. 2017. An Ethnolinguistic and Genetic Perspective on the Origins of the Dravidian-Speaking Brahui in Pakistan 13.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. *Genetics* **192**:1065–1093. doi:10.1534/genetics.112.145037

Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S. 2009. Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. *PLoS Genet* **5**:e1000519. doi:10.1371/journal.pgen.1000519

Rivollat M, Jeong C, Schiffels S, Küçükkalıpçı İ, Pemonge M-H, Rohrlach AB, Alt KW, Binder D, Friederich S, Ghesquière E, Gronenborn D, Laporte L, Lefranc P, Meller H, Réveillas H, Rosenstock E, Rottier S, Scarre C, Soler L, Wahl J, Krause J, Deguilloux M-F, Haak W. 2020. Ancient genome-wide DNA from France highlights the complexity of interactions between Mesolithic hunter-gatherers and Neolithic farmers. *Sci Adv* **6**:eaaz5344. doi:10.1126/sciadv.aaz5344

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic Structure of Human Populations. *Science* **298**:2381. doi:10.1126/science.1078311

Skoglund P, Malmstrom H, Omrak A, Raghavan M, Valdiosera C, Gunther T, Hall P, Tambets K, Parik J, Sjogren K-G, Apel J, Willerslev E, Stora J, Gotherstrom A, Jakobsson M. 2014. Genomic Diversity and Admixture Differs for Stone-Age Scandinavian Foragers and Farmers. *Science* **344**:747–750. doi:10.1126/science.1253448

Skoglund P, Malmstrom H, Raghavan M, Stora J, Hall P, Willerslev E, Gilbert MTP, Gotherstrom A, Jakobsson M. 2012. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. *Science* **336**:466–469. doi:10.1126/science.1216304

Osborne, M. R., & Smyth, G. K. (1986). An Algorithm for Exponential Fitting Revisited. *Journal of Applied Probability*, *23*, 419–430. JSTOR. https://doi.org/10.2307/3214370

Tournebize R, Chu G, Moorjani P. 2020. Reconstructing the history of founder events using genome-wide patterns of allele sharing across individuals. *bioRxiv* 2020.09.07.286450. doi:10.1101/2020.09.07.286450

Wang C-C, Reinhold S, Kalmykov A, Wissgott A, Brandt G, Jeong C, Cheronet O, Ferry M, Harney E, Keating D, Mallick S, Rohland N, Stewardson K, Kantorovich AR, Maslov VE, Petrenko VG, Erlikh VR, Atabiev BCh, Magomedov RG, Kohl PL, Alt KW, Pichler SL, Gerling C, Meller H, Vardanyan B, Yeganyan L, Rezepkin AD, Mariaschk D, Berezina N, Gresky J, Fuchs K, Knipper C, Schiffels S, Balanovska E, Balanovsky O, Mathieson I, Higham T, Berezin YB, Buzhilova A, Trifonov V, Pinhasi R, Belinskij AB, Reich D, Hansen S, Krause J, Haak W. 2019. Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. *Nat Commun* **10**:590. doi:10.1038/s41467-018-08220-8

DATES: Implementation, versions, and comparison with other published methods

(A) *DATES* FFT implementation

In order to make *DATES* computationally tractable, we implemented the fast Fourier transform (FFT) for computing ancestry covariance as described in ALDER (Loh et al., 2013). Briefly, we perform an algebraic transformation of the ancestry covariance statistic and compute the FFT convolution in discrete equally sized bins (referred to as mesh points). This provides a speedup from $O(n^2)$ to $O(n \log n)$, which reduces the typical runtimes from hours to seconds.

In *DATES*, we compute the ancestry covariance $A(d)$ (Methods) by expanding the numerator as below. $X_j(d)$, (j = 0, 1, 2) where

$$X_2(d) = \sum_{s(d)} K_1 K_2$$

$$X_1(d) = 2\left(\sum_{s(d)} \overline{K_1} K_2 + \overline{K_2} K_1\right)$$

$$X_0(d) = \sum_{s(d)} \overline{K_1}\,\overline{K_2}$$

Where d is genetic distance in Morgans between pair of neighboring markers S_1, S_2 are and K_i represents the ancestry at marker S_i. We discuss an approximate calculation of X_2. The calculations of X_1 and X_0 are similar.

Like ALDER, we divide the genome in windows based on the position in the genetic map (instead of genetic distance). We set a mesh on the genetic map (default mesh size is 0.01 centiMorgans (cM)), mapping every SNP to the nearest mesh point. For a mesh point, u define T_u to be the set of SNPs mapping to u and

$$K(u) = \sum_{i \in T(u)} K_i$$

We now set

$$X_2'(d) = \sum_{u,v:|u-v|=d} K(u)K(v)$$

Where $|u - v|$ is the genetic distance of u, v. X_2' can be computed by FFT. We note that the use of the mesh is the only source of approximation in the FFT implementation to compute X_2. The

mesh discretization parameter, *qbin*, provides a trade-off between runtime and accuracy, smaller mesh size leads to higher accuracy and longer run time.

To explore the impact of *qbin* on the estimated accuracy, we performed simulations for varying sample sizes (n=1 and n=20) and ran *DATES* using varying *qbin* between 1–100. We find the method works reliably for all *qbin* values (Appendix 1—figure 1). Moreover, there is almost a 5–10-fold speedup in a run between *qbin* values of 10 vs. 100 (Appendix 1—figure 2). The run time is invariant to the proportion of time of admixture. The default value of *qbin* in *DATES* is 10 but we advise the user to perform simulations for their dataset size and population model to set this parameter reliably.

Assessing the exponential fit

Following Tournebize et al. (2020), we examined the quality of the exponential fit by computing the normalized root-mean-square deviation (NRMSD) between the empirical ancestry covariance values z and the fitted ones $\hat{z}$, across all the genetic distance bins (where, N the number of bins) (Tournebize et al., 2020).

$$NRMSD = \frac{1}{max(\hat{z}) - min(\hat{z})}\sqrt{\frac{\sum^{D}(z-\hat{z})^2}{N}}$$

We calculated NRMSD for all ancient DNA populations in our study and the distribution of these values is shown in (Appendix 1—figure 3). Focusing on the most extreme values of NRMSD, we show that the statistic is useful in identifying poor fits where the fitted line deviates from the data or the fit is highly dispersed (Appendix 1—figure 4). However, the absolute value of this statistic does not have any statistical meaning. Based on the empirical distribution of NRMSD values in our study samples (Appendix 1—figure 3), we use a threshold of 0.7 to flag poor fits. We caution that users should adjust this threshold based on their application and always visually inspect their exponential fits to ensure reliable results.

(B) Comparison of versions of *DATES*

An earlier version of *DATES* (version v753) was released in (Narasimhan et al., 2019). The current method (version 4010) released in this study differs in some key aspects of the implementation (Methods). To compare the two versions of *DATES*, we performed simulations and generated 10 admixed individuals with 20% European and 80% African ancestry where the time of admixture varied between 10–300 generations (similar to Appendix 2—figure 1). We also varied the sample sizes of the admixed population between 1-20 in increments of 5. Our estimated admixture using v753 and v4010 are highly concordant suggesting although the implementation has changed, the results are similar (Appendix 1—table 1A-B). Further, we

compared the dates of admixture that were reported using the earlier version. To this end, we repeated the analysis for Narasimhan et al. (2019) for ancient South Asians and Rivollat et al. (2020) for ancient Neolithic samples in Europe (Rivollat et al., 2020). In both cases, we obtained consistent results as reported earlier (Appendix 1—table 2, Appendix 1—table 3).

Appendix 1—table 1: Comparison of results using *DATES* v753 and v4010 using simulated data.

A) Simulated data with the target sample size (*n*) of 10 individuals

True time of admixture (generations)	*DATES* (V753) (mean ± SE)	*DATES* (v4010) (mean ± SE)
10	10.0 ± 0.5	11.0 ± 0.6
20	19.1 ± 1.5	19.6 ± 1.5
30	28.0 ± 1.5	28.9 ± 1.3
40	46.1 ± 2.1	45.7 ± 1.8
50	55.5 ± 2.9	55.6 ± 2.8
60	59.4 ± 2.2	60.5 ± 2.4
70	69.3 ± 4.2	69.8 ± 4.0
80	84.2 ± 4.4	84.0 ± 3.9
90	97.7 ± 3.7	93.7 ± 3.6
100	107.4 ± 5.4	106.7 ± 4.5
110	113.6 ± 5.5	112.9 ± 4.7
120	122.7 ± 5.4	124.3 ± 5.5
130	138.6 ± 7.7	134.1 ± 6.2
140	153.2 ± 9.2	152.0 ± 8.4
150	147.5 ± 9.0	146.6 ± 8.2
160	181.4 ± 9.8	176.6 ± 8.2
170	178.0 ± 8.1	175.9 ± 7.4
180	180.7 ± 10.6	182.3 ± 9.3
190	172.9 ± 15.3	174.8 ± 12.7
200	204.7 ± 17.1	208.8 ± 13.4
210	194.5 ± 13.3	196.3 ± 11.9
220	255.8 ± 17.0	250.7 ± 13.8
230	251.9 ± 18.6	237.0 ± 13.0
240	234.7 ± 18.3	241.5 ± 14.2
250	228.3 ± 13.8	233.2 ± 11.9
260	254.2 ± 21.7	253.0 ± 16.1
270	291.4 ± 22.4	292.1 ± 20.0
280	252.4 ± 25.2	248.1 ± 22.4
290	277.9 ± 22.4	285.4 ± 20.5
300	318.6 ± 23.3	315.1 ± 20.3

118

B) Simulated data with sample size (*n*) ranging between 10-20 individuals. 119 120

True time of admixture (generations)	Sample size	*DATES* (V753) (mean ± SE)	*DATES* (v4010) (mean ± SE)
10	1	6.9 ± 2.5	7.8 ± 2.6
10	5	8.8 ± 0.8	9.9 ± 0.8
10	10	9.9 ± 1.2	10.8 ± 1.2
10	15	10.9 ± 0.7	11.8 ± 0.7
10	20	10.3 ± 0.6	11.3 ± 0.6
50	1	51.7 ± 7.1	51.5 ± 7.5
50	5	59.7 ± 3.5	58.6 ± 3.2
50	10	48.7 ± 2.7	50.1 ± 2.7
50	15	54.2 ± 2.1	54.7 ± 2.1
50	20	52.9 ± 1.9	53.1 ± 1.9
100	1	124.5 ± 17.6	122.5 ± 13.2
100	5	107.2 ± 7.6	108.2 ± 7.5
100	10	103.3 ± 7.7	103.2 ± 7.8
100	15	99.4 ± 4.5	100.1 ± 3.8
100	20	105.5 ± 3.9	103.4 ± 3.3
150	1	136.4 ± 29.2	144.2 ± 26.3
150	5	142.6 ± 11.4	143 ± 11.7
150	10	156.9 ± 9	158.6 ± 7
150	15	142.9 ± 7.8	146.1 ± 6.7
150	20	156.5 ± 5.4	152.9 ± 4.3
200	1	195.4 ± 88.4	160.2 ± 73.9
200	5	210.9 ± 20.7	206.7 ± 18.7
200	10	225 ± 18.7	219.8 ± 18
200	15	200 ± 10.6	197.7 ± 9
200	20	189.4 ± 11	190.3 ± 9

Appendix 1—table 2: Comparison of results with Narasimhan, Patterson et al. 2019.

Pop	**Reference populations***	***DATES* (v753)**	***DATES* (v4010)**
		(mean ± SE; in generations)	**(mean ± SE; in generations)**
Indus_Periphery_Pool	AASI and Iranian-farmer-related	71 ± 15	62 ± 7
SPGT	AASI and Steppe-pastoralist-related	26 ± 3	28 ± 3

Note:
* We used the reference populations of AASI ancestry that includes South Asians from the 1000 Genomes Project (Phase 3) including Sri Lankan Tamil from the UK (*STU.SG*) and Indian Telugu from the UK (*ITU.SG*), as well as *BIR.SG* and Iranian farmer-related ancestry including Aigyrzhal_BA, Sarazm_EN, Geoksyur_EN, Parkhai_Anau_EN, and Steppe-pastoralist-related including Central_Steppe_MLBA.

Appendix 1—table 3: Comparison of dates of the spread of Neolithic farming from Rivollat et al. 2020.

Population	***n***	***DATES* (v753)**	**Population in our study (v44 1240K)**	***n* (v44 1240K)**	***DATES* (v4010)#**
Bulgaria_MP_Neolithic	9	8.4 ± 2.3	Bulgaria_MalakPreslavets_N	3	8.05 ± 3
Serbia_Neolithic	4	--	Serbia_EN	3	22.8 ± 9.8
Romania_EN	2	32.1 ± 10.4	Romania_EN*	2	29.7 ± 7.1
Croatia_Impressa	2	--	Croatia_EN_Impressa	2	--
Hungary_ALPc_MN	23	21.5 ± 4.7	Hungary_MN_ALPc	21	21.9 ± 1.6
Hungary_LBK_MN	10	12.8 ± 5.2	Hungary_MN_LBK	6	18.6 ± 7.4
Hungary_ALBK_MN	2	14.8 ± 3.2	Hungary_MN_ALBK_Szakalhat	2	19.3 ± 3.3
Hungary_LN	18	21.5 ± 3.7	Hungary_LN	18	28.03 ± 3.8
Austria_LBK_EN	8	15.5 ± 4.6	Austria_EN_LBK	9	17.6 ± 2.3
Czech_MN	5	18.3 ± 7.7	Czech_MN	4	32.9 ± 6.3
France_MN	3	26.5 ± 5.6	France_MN	43	30 ± 1.3
Iberia_EN	10	15.6 ± 2.5	Spain_EN	11	20.6 ± 3.6
Iberia_MN	7	52.4 ± 4.3	Spain_MLN	42	56.3 ± 4
Germany_LBK_EN	27	14.4 ± 2.6	Germany_EN_LBK	54	17.4 ± 2.7
Germany_Blatterhohle_MN	4	12.3 ± 2.5	Germany_Blatterhohle_MN	4	16.2 ± 2.9
Germany_Esperstedt_MN	1	--	Germany_MN_Esperstedt	1	--
England_Neolithic	29	45.5 ± 5.5	England_N.SG	17	--
Wales_Neolithic	6	45.3 ± 7.4	Wales_N	4	50.7 ± 3.3
Scotland_Neolithic	42	50.9 ± 3.8	Scotland_N	30	56.6 ± 2.9
Ireland_Neolithic	13	46.9 ± 7.5	Ireland_MN.SG	26	50.8 ± 2.2

Note:
(blue) indicates samples sizes that differ across both studies
For *DATES*, we used pooled WHG and Anatolian farmers as the reference populations except for samples marked with *.
-- indicates cases where the results were not significant as the 95% CI includes 0

(C) Comparison of *DATES* with published methods

(i) Comparing *DATES* and ALDER in simulations

To compare the *DATES* with ALDER, we simulated data for n (=20 or 100) admixed individuals using our admixture simulator with 20% European and 80% African ancestry using CEU and YRI phased individuals from 1000 Genomes Project (Auton et al., 2015). We used French and Yoruba from the Human Genome Diversity Panel (HGDP) dataset (Rosenberg et al., 2002) as the reference populations to represent European and African source populations respectively. We applied *DATES* and ALDER to the same dataset. We ran ALDER using default settings and allowed ALDER to pick the minimum distance to start the exponential fit. ALDER estimates the date of admixture by fitting an exponential to the weighted covariance statistic with genetic distance and performs a least-squares fit using $y = Ae^{-td} + c$, where d is the genetic distance in Morgans and t is the number of generations since admixture. We note this differs from *DATES* which assumes the exponential decay parameter of $(t + 1)$, though in practice this has little effect on the comparisons. For the timing of admixture between 10-200 generations, we observed that both methods accurately estimated the time of admixture in most cases, though *DATES* provided more precise estimates than ALDER for older admixture dates (Appendix 1—table 4).

Appendix 1—table 4: Comparison of ALDER and *DATES* for varying samples sizes and times of admixture. We simulated data for 20 and 100 admixed individuals using the CEU and YRI from 1000G with the mixture proportion of 20% from European and 80% African ancestry. The dates reported here for *DATES* are using exponential fit to $\lambda - 1$ generations.

	Number of individuals, *n*=20		**Number of individuals, *n*=100**	
Time of admixture (gen)	**ALDER mean ± 1SE (gen)**	***DATES* mean ± 1SE (gen)**	**ALDER mean ± 1SE (gen)**	***DATES* mean ± 1SE (gen)**
10	9.3 ± 0.8	10.7 ± 0.6	10.2 ± 0.3	10 ± 0.3
20	19.4 ± 1.3	19.7 ± 0.8	20.2 ± 0.3	20.3 ± 0.3
30	28.5 ± 1.7	30.8 ± 1.5	30.6 ± 0.9	30.5 ± 0.7
40	40.9 ± 2	40.3 ± 1.5	40.6 ± 0.7	40.6 ± 0.4
50	47.9 ± 3.6	49.6 ± 1.6	50 ± 1.1	50.9 ± 0.7
60	55.7 ± 2.7	60.3 ± 1.5	62 ± 2.2	63.2 ± 1
70	71.4 ± 4	74 ± 2.7	74 ± 2.2	72.4 ± 1.3
80	80.6 ± 4.8	82.5 ± 2.9	85.3 ± 2.3	84.4 ± 1.1
90	87.8 ± 4.2	88.9 ± 3	94.1 ± 2.7	92.9 ± 1.3
100	93.7 ± 4.9	98.1 ± 2.9	101.9 ± 3.9	103.6 ± 1

110	121.4 ± 5.4	118.2 ± 3.7	120.7 ± 4.3	115.5 ± 1.8
120	116.5 ± 8.5	128.4 ± 3.9	121.2 ± 5.1	121.5 ± 1.7
130	138.2 ± 9.2	133.7 ± 4.6	130.2 ± 4.8	132.8 ± 1.7
140	134.5 ± 17.5	142.4 ± 7	144.9 ± 7.3	145.3 ± 3.1
150	144.8 ± 23.8	149.5 ± 7.4	155.1 ± 7	157.5 ± 2.8
160	141.9 ± 11.3	166.7 ± 5.9	154.5 ± 8.5	161.7 ± 2.4
170	173.4 ± 13.7	175.1 ± 6.9	170.3 ± 6.2	173.5 ± 3
180	204.6 ± 17.8	195.5 ± 7.1	174.2 ± 7	180.7 ± 3.3
190	221.3 ± 23.9	210.4 ± 9.4	191.2 ± 16.2	197.2 ± 4.6
200	202.8 ± 11.1	196 ± 6.5	188.5 ± 16.5	202.6 ± 4.7

Next, we generated 10 simulated individuals with missing genotypes varying between 5–60% (in increments of 5%) as described in Methods and applied both *DATES* and ALDER. Using the same setup in both methods, we inferred that *DATES* reliably recovers the time of admixture even with high missing proportions such as 60% (Appendix 1—figure 5). However, ALDER becomes every noisy with large proportions of missing data (>40%). For older dates (>100 generations), we observed biased estimates even with >10% missing genotypes (Appendix 1—figure 5) As the missing sites vary among individuals, admixture-LD based methods such as ALDER that combine information across individuals become noisy as there are few sites without non-missing genotypes remaining for inference. However, *DATES* performs the analysis for single individuals (using all non-missing genotypes for that individual) and then averages the inferred estimates across individuals. This provides substantial robustness to variable missingness across individuals.

180

181 (ii) **Comparison of *DATES* with published admixture dating methods- Rolloff,**
182 **Globetrotter, ALDER.**

183

184 **Appendix 1—table 5: Admixture dates in present day populations inferred using Rolloff,**
185 **Globetrotter, ALDER and *DATES*.**

Population	**nk**	**Source1**	**Source2**	**Rolloff**	**Globetrotter**	**ALDER formal test**	**ALDER_2-ref dates**	***DATES***	**Comments**
Hazara	22	Mongola (10)	Iranian (13)	23 ± 1	22 ± 0.9	Long-range LD	--	24.6 ± 1.0	
Uzbekistani	15	Mongola (10)	Iranian (13)	20 ± 1.4	19 ± 1.1	**SUCCEEDS**	19.18 ± 2.22	21.3 ± 1.4	
Uyghur	10	Mongola (10)	Iranian (13)	23 ± 2.6	22 ± 1.3	**SUCCEEDS**	16.73 ± 1.38	22.2 ± 2.1	
Makrani	22	Bantu Kenya (11)	Balochi (21)	18 ± 1.8	18 ± 1.2	Long-range LD	--	13.2 ± 1.6	
Druze	42	Yoruba (21)	Cypriot (12)	39 ± 7.3	37 ± 1.9	FAILS	44.02 ± 6.37	43.4 ± 6.1	
Mozabite	25	Yoruba (21)	Moroccan (22)	23 ± 1.9	21 ± 1.3	Long-range LD	--	21.6 ± 1.8	
Turkish	17	Mongola (10)	Iranian (13)	28 ± 3.2	24 ± 1.5	FAILS	25.62 ± 2.48	28.5 ± 2.3	
Brahui	23	Bantu Kenya (11)	Balochi (21)	13 ± 3.4	20 ± 1.5	Long-range LD	--	10.4 ± 1.6*	Possibly multi-way admixture (Pagani et al., 2017)
Yemeni	4	Bantu Kenya (11)	Syrian (16)	15 ± 2.3	14 ± 1.8	FAILS	6.29 ± 2.89	12.7 ± 1.6	
Pima	14	Turkish (17)	Mayan (21)	9 ± 3.6	6 ± 0.9	**SUCCEEDS**	6.29 ± 0.89	7.8 ± 1.1	
Bantu South Africa	8	San Khomani (30)	Yoruba (21)	26 ± 2.5	25 ± 2.3	Long-range LD	--	27.9 ± 2.2	
Tu	10	Greek (20)	Han N-China (10)	33 ± 6.3	25 ± 2.3	FAILS	28.83 ± 2.8	31.3 ± 1.96	
West Sicilian	10	Yoruba (21)	East Sicilian (10)	26 ± 7.8	27 ± 3.9	FAILS	42.72 ± 16.34	37.4 ± 16.4	
Cambodian	10	Uyghur (10)	Han (34)	17 ± 4.7	20 ± 2.7	**SUCCEEDS**	24.28 ± 5.36	33.6 ± 3.7*	
Georgian	20	Adygei (17)	Greek (20)	--	30 ± 3.3	FAILS	3.15 ± 1.22	--	
Romanian	13	Lithuanian (10)	East Sicilian (10)	--	31 ± 2.6	FAILS	--	--	
Bulgarian	18	Polish (16)	Cypriot (12)	--	28 ± 3.5	FAILS	40.95 ± 16.42	91.1 ± 24.7*	Possibly multi-way admixtur

									e (see Main text)
Hezhen	8	Tujia (10)	Mongola (10)	--	13 ± 1.3	FAILS	2.92 ± 1.4	--	
Oroqen	9	Yakut (25)	Mongola (10)	--	15 ± 2	Long-range LD	--	--	
Hungarian	18	Cypriot (12)	Polish (16)	65 ± 24	39 ± 3.5	FAILS	54.83 ± 25.27	61.8 ± 19.1	
Han N-China	10	Turkish (17)	Tujia (10)	37 ± 11.1	26 ± 3.8	FAILS	48.17 ± 10.36	44.3 ± 5.1*	
Daur	9	Tujia (10)	Mongola (10)	--	21 ± 1.7	FAILS	--	--	
Greek	20	Polish (16)	Cypriot (12)	69 ± 18.5	36 ± 3.7	FAILS	55.54 ± 8.93	62.6 ± 16.9	
Melanesian	10	Papuan (16)	Cambodian (10)	66 ± 12.1	28 ± 7.6	FAILS	64.91 ± 5.42	68.6 ± 7.1*	
Mandenka	22	Moroccan (22)	Yoruba (21)	22 ± 10.3	19 ± 4.2	FAILS	17.25 ± 6.05	85.8 ± 19.0 *#Ω	Possibly multiple admixture events (Price et al., 2009)
Indian	13	Cambodian (10)	Sindhi (23)	91 ± 41.1	53 ± 8.4	FAILS	n/a	n/a	There are multiple "Indian" groups in the dataset making it unclear which target was used
North Italian	12	Cypriot (12)	French (28)	--	71 ± 11.8	FAILS	12.44 ± 4.32	--	
Polish	16	French (28)	Lithuanian (10)	--	31 ± 5.1	FAILS	--	--	
Tuscan	8	Cypriot (12)	French (28)	--	35 ± 6.1	FAILS	--	--	
San Namibia	5	Sandawe (28)	San Khomani (30)	--	48 ± 8.9	Long-range LD	--	--	

NOTE

- Columns 1-5 include results from Table S12 from Hellenthal et al. 2014. We only show significant dates (|Z| > 2)
- Following Hellenthal et al., we created a merged dataset of the Human Genome Diversity Panel, Henn et al. and Behar et al. containing 1642 individuals and 465543 SNPs. This dataset was used for ALDER and *DATES* analysis.
- Standard errors in *DATES* were estimated using chromosome jackknife (see Methods)

-- indicates results where the inferred results were not significant, either the method failed or the 95% CI included 0

*- indicates *DATES* estimates that significantly differ from Globetrotter estimates (not within two SEs)

#- indicates *DATES* estimates that significantly differ from Rolloff results

Ω- indicates *DATES* estimates that significantly differ from ALDER results

n/a- indicates target population was unclear

(iii) Comparing *DATES* and modified ALDER for admixture times in Neolithic samples

Here we compare the performance of *DATES* with a modified version of ALDER used in (Lipson et al., 2017) where they use a helper high coverage sample in order to infer the timing of admixture in three regions. An advantage of our approach over modified ALDER analysis is that we do not need helper samples. Our results suggest that we obtain statistically consistent results for all overlapping samples (within two standard errors). In some cases, higher coverage samples maybe available, though unless these samples have similar ancestry, they could bias the dates obtained.

Appendix 1—table 6: Admixture dates for Neolithic European groups using DATES and modified ALDER (Lipson et al. 2017)

Region	Population	Modified ALDER* (Lipson et al. 2017)	DATES#
Germany	Blätterhöhle_MN	18.5 ± 4.6	14 ± 3
Germany	Germany_MN	26.2 ± 4.4	36 ± 20
Germany	LBK_EN	14.9 ± 2.4	18 ± 3
Hungary	LBK_EN	17.8 ± 2.0	22 ± 2
Hungary	Baden_CA	27.6 ± 3.8	49 ± 11
Hungary	Lasinja_CA	29.3 ± 5.2	32 ± 6
Hungary	LBKT_MN	30.3 ± 5.8	23 ± 10
Hungary	Protoboleraz_CA	44.3 ± 6.4	46 ± 7
Hungary	Starcevo_EN	4.5 ± 1.9	5 ± 2
Hungary	TDLN	20.9 ± 2.7	29 ± 4
Hungary	Tisza_LN	18.2 ± 6.6	27 ± 9
Spain	Iberia_CA	49.6 ± 5.2	55 ± 6
Spain	Iberia_EN	19.4 ± 2.3	20 ± 3
Spain	Iberia_MN	49.9 ± 7.7	52 ± 8

NOTE:
* Modified ALDER - We report the individual level dates from Extended Data Table 4 based on average of individual level dates calculated using Anatolian farmers and WHG as sources and high coverage Anatolian farmers as helper samples. For details, see Lipson et al. 2017
# DATES - We used pooled WHG groups and Anatolian farmers as references in DATES

Additional files

Appendix 1—figure 1. Impact of the discretization parameter (*qbin*) on accuracy. We show three subplots for a sample size of n=1 (Panel A) and n=20 (Panel B). For each subplot, we simulated data for n admixed individuals with 20% ancestry from Europeans (1000 Genomes, CEU) and 80% ancestry from Africans (1000 Genomes, YRI) with the time of admixture (t)

shown on the X-axis and the estimated admixture time inferred using *DATES* on Y-axis. We ran *DATES* using varying *qbin* values shown in different colors.

Appendix 1—figure 2. Impact of the discretization parameter (*qbin*) on run time. We show three subplots for sample size of *n=1* (top left), *n=20* (top right) and *n=100* (bottom). For each subplot, we simulated data for *n* admixed individuals with 20% ancestry from Europeans (1000 Genomes, CEU) and 80% ancestry from Africans (1000 Genomes, YRI) with the time of admixture (*t*) of 100 generations ago. We show the impact of *qbin* (X-axis) on the runtime measured in seconds. For sample sizes, $n > 1$, r^2 between *qbin* and run time is >0.99.

Appendix 1—figure 3. Histogram of the NRMSD values computed as the normalized residual between the empirical and fitted decay curves, for all the ancient DNA populations reported in Figure 3 —Supplementary file 1. The red vertical line represents the value NRMSD=0.7, which we used as the threshold to exclude populations from our analysis because visual inspection of fitted curves above this threshold suggests the results are too noisy to make a reliable inference (see Figure 3–figure supplement 1 for all fitted decay curves).

Appendix 1—figure 4. Ancestry covariance curves for the lowest (left) and highest (right) values in our ancient DNA populations in our study. For details of all curves and NRMSD, see Figure 3 — figure supplement 1.

Appendix 1—figure 5: Effect of missing genotypes on the performance of *DATES* and ALDER: We simulated data for 10 admixed individuals with varying proportions of missing data (shown in each panel). The estimated admixture times (± 1 SE) from *DATES* (green) and ALDER (pink) are shown on Y-axis and the true time of admixture is shown on X-axis. For a fair comparison with ALDER, the dates reported here for *DATES* are using exponential fit to $\lambda - 1$ generations (instead of the default of λ generations).

Simulations to test the performance of *DATES* under complex admixture scenarios involving founder events, multiple pulses of gene flow or gradual admixture

DATES models admixed individuals as a mix of two source populations. However, in real data, the target could have ancestry from more than two source populations, or the gene flow can occur continuously over a period of time. To test the performance of *DATES* under these more complex scenarios, we performed additional simulations.

A. Multiple pulses of gene flow

Using the simulation model described in Methods, we generated a target population that has ancestry from three groups (*PopA*, *PopB*, *PopC*) that mixed at two distinct times (t_1 and t_2 generations ago) (Appendix 2—figure 1). Specifically, we generated data for three sets of admixed populations each with 10 individuals, where *PopA*, *PopB,* and *PopC* differ across runs. For each simulation, the older pulse of admixture occurred t_2 (=30, 60, 100) generations ago and *PopA* and *PopB* mixed with α_1/α_2 ancestry respectively. This mixture was followed by additional gene flow from *PopC* that contributed α_3 ancestry at t_1 (=10) generation ago. We used CEU, YRI, and CHB as *PopA*, *PopB,* and *PopC* and varied the order of the three ancestral populations to generate multiple sets of simulated individuals. We estimated admixture time for all three sets of simulated individuals using French, Tujia, and Yoruba as reference populations.

(i) Equal proportions of ancestry from three sources.

In this setup, we simulated admixed individuals with ancestry from *PopA*, *PopB,* and *PopC* with $\alpha_1 = 50\%$, $\alpha_2 = 50\%$, $\alpha_3 = 33\%$ thus the effective ancestry proportion in the admixed population would be 33% from each ancestral group. We varied the order of the three ancestrals and applied *DATES* to infer the timing of the mixture. Our results showed that when the references populations used for the inference correspond to the true admixing sources, in most cases, we reliably estimate the timing of both the pulses of admixture (Appendix 2—figure 2).

(ii) Unequal proportions of ancestry from three ancestrals.

In most real-world scenarios, the ancestry proportions of the admixing groups are unlikely to be exactly the same or similar. Thus, we generated data for groups with unequal proportions of ancestry from *PopA*, *PopB,* and *PopC* by setting $\alpha_1 = 20\%$, $\alpha_2 = 80\%$, and $\alpha_3 = 20\%$ or 80%. We varied the order of the three ancestral groups and applied *DATES* to infer the timing of the mixture. We observed that we recovered the timing of the recent pulse of admixture in most cases (Appendix 2—figure 3). In some cases, there was confounding in the timing of the recent event, when the % of ancestry from *PopC* was low (20%). In Appendix 2—table 1, we explore how choosing ancestral populations that are more aligned with the model of admixture (that can be reliably inferred using other methods like *qpAdm*) can alleviate this bias (Appendix 2—table 1).

(iii) Choice of references in a multi-pulse admixture

Using the admixture scenario described above, we examined how the choice of reference populations impacts the inferred dates of admixture. From Appendix 2—figure 3, we find that using PopC as one of the references gives more reliable dates. In most real-world scenarios, the ordering of gene flow events is known. For instance, for present-day Europeans, it's known that Steppe pastoralist gene flow occurs after the gene flow between Anatolian farmers and local European hunter-gatherers. Thus, we fixed one reference group as *PopC* and explored how the choice of the second reference population impacts the recovery of the recent gene flow event. To ensure that our observed dates are not biased, we ran 10 replicates for each run and report the average of the 10 runs below. Specifically, for the target populations with ancestries from *PopA*, *PopB,* and *PopC*, we ran *DATES* with the following reference populations:

(i) Ref1: *PopC* and Ref2: *PopA*

(ii) Ref1: *PopC* and Ref2: *PopB*

(iii) Ref1: *PopC* and Ref1: admixed individuals with *PopA* and *PopB* ancestry (30% *PopA* / 70% *PopB*) ancestry

(iv) Ref1: *PopC* and Ref2: pooled samples from *PopA* and *PopB*

We observed that in all four cases using *PopC* (admixing source in the most recent pulse) as one of the sources allows us to recover the recent pulse of admixture. When the other reference population is either the pooled set of samples of *PopA* and *PopB* (admixing sources in the older pulse) or admixed individuals with ancestry from *PopA* and *PopB*, we reliably infer the younger pulse of admixture in all cases (Appendix 2—table 1). In other cases, we observe some confounding in the inferred dates with the estimated dates falling intermediate of two dates.

Appendix 2—table 1: Impact of reference populations in two-way admixed groups.

Model: We generated target populations with two pulses of gene flow where *PopA* and *PopB* mixed at time t_1 generations ago with ancestry proportion of α_1 and α_2, followed by gene flow from *PopC* at t_2 generations ago with ancestry proportion of α_3							
Target	**t_2/t_1**	**Ref1**	**Ref 2**	**α_1 =20%, α_2 =80%, α_3 =80%**	**α_1 =50%, α_2 =50%, α_3 =50%**	**α_1 =20%, α_2 =80%, α_3 =20%**	**α_1 =20%, α_2 =80%, α_3 =10%**
A. Using reference populations *PopA* and *PopC*							
PopA = CEU *PopB = YRI* *PopC = CHB*	100/10	Han	French	10.3	10.7	12.0	12.6
PopA = CEU *PopB = CHB* *PopC = YRI*	100/10	Yoruba	French	10.3	9.7	10.2	10.6
PopA = CHB *PopB = YRI* *PopC = CEU*	100/10	French	Han	10.8	11.6	20.7	44
PopA = CEU *PopB = YRI* *PopC = CHB*	60/10	Han	French	11.3	10.7	11.8	14.4
PopA = CEU *PopB = CHB* *PopC = YRI*	60/10	Yoruba	French	10.4	10.8	10.3	11.2
PopA = CHB *PopB = YRI* *PopC = CEU*	60/10	French	Han	11.2	12.6	19.9	40.7

B. Using reference populations *PopB* and *PopC*							
PopA = CEU *PopB = YRI* *PopC = CHB*	100/10	Han	Yoruba	10.2	11.5	11.3	11.9
PopA = CEU *PopB = CHB* *PopC = YRI*	100/10	Yoruba	Han	10.1	9.7	10.3	10.5
PopA = CHB *PopB = YRI* *PopC = CEU*	100/10	French	Yoruba	10.1	12.8	12.2	14
PopA = CEU *PopB = YRI* *PopC = CHB*	60/10	Han	Yoruba	11.0	12.6	12.2	14.6
PopA = CEU *PopB = CHB* *PopC = YRI*	60/10	Yoruba	Han	10.3	11.1	10.6	11.4
PopA = CHB *PopB = YRI* *PopC = CEU*	60/10	French	Yoruba	10.4	13.4	12.4	18
C. Using reference populations *PopC* and "admixed" individuals with ancestry from *PopA* and *PopB* *(30% PopA / 70% PopB)* ancestry							
PopA = CEU *PopB = YRI* *PopC = CHB*	100/10	Han	Admixed (30% CEU/ 70% YRI)	10.1	10.9	10.8	10.5
PopA = CEU *PopB = CHB* *PopC = YRI*	100/10	Yoruba	Admixed (30% CEU/ 70% CHB)	10.1	9.5	10.0	10
PopA = CHB *PopB = YRI* *PopC = CEU*	100/10	French	Admixed (30% CHB/ 70% YRI)	10.0	11.2	10.9	10.8
PopA = CEU *PopB = YRI* *PopC = CHB*	60/10	Han	Admixed (30% CEU/ 70% YRI)	11	11.3	10.9	11.8
PopA = CEU *PopB = CHB* *PopC = YRI*	60/10	Yoruba	Admixed (30% CEU/ 70% CHB)	10.3	10.8	10.3	10.5
PopA = CHB *PopB = YRI* *PopC = CEU*	60/10	French	Admixed (30% CHB/ 70% YRI)	10.2	11.3	10.6	12.1
D. Using reference populations *PopC* and pooled individuals of *PopA* and *PopB* ancestry							
PopA = CEU *PopB = YRI* *PopC = CHB*	100/10	Han	French + Yoruba	10.1	10.99	10.4	9.5
PopA = CEU *PopB = CHB* *PopC = YRI*	100/10	Yoruba	French + Han	10.2	9.5	10.03	9.9
PopA = CHB *PopB = YRI* *PopC = CEU*	100/10	French	Han + Yoruba	9.9	10.4	10.6	10
PopA = CEU *PopB = YRI* *PopC = CHB*	60/10	Han	French + Yoruba	10.9	10.8	10.3	10.3
PopA = CEU *PopB = CHB* *PopC = YRI*	60/10	Yoruba	French + Han	10.3	10.7	10.1	10.5
PopA = CHB *PopB = YRI* *PopC = CEU*	60/10	French	Han + Yoruba	10.0	10.3	9.8	11

Note: the estimated dates are shown per scenario are averages of 10 simulations.

B. Coalescent simulations

To evaluate the performance of *DATES* under demographic models involving gradual gene flow or founder events, we performed simulations using the coalescent simulator, *macs* (Chen et al., 2008). For all the simulations described below, we generated data for three populations (*PopA*, *PopB,* and *PopC*) for a region of 100Mb with 22 replicates. The effective population size (N_e) of all three populations was assumed to be 12,500 with the mutation rate and recombination rate was assumed as 1.2 x 10^{-8} and 1 x 10^{-8} per base pair per generation respectively (Halldorsson et al., 2019; Jónsson et al., 2017). The divergence time between population *A* and *B* was assumed to be 1800 generations, which translates to an estimated F_{ST} (*PopA*, *PopB*) of 0.067. *PopC* was formed by admixture between *PopA* to *PopB* that occurred either continuously over a period of λ generations or instantaneously at time *t*. We combined two haploid chromosomes at random to generate one diploid chromosome.

(i) Impact of continuous gene flow

To model continuous gene flow, we simulated a gradual mixture in *PopC* from *PopA*/ *PopB* over a period of $\lambda\,(= 5 - 60)$ generations, leading to 20%/ 80% *PopA*/ *PopB* ancestry. Applying *DATES* to *PopC* with *Pop A* and *PopB* as the reference populations showed that the inferred time was intermediate between the start and end of the period of gene flow (Appendix 2—table 2). This is similar to the results of other admixture dating methods like Globetrotter, ALDER, and ROLLOFF (Hellenthal et al., 2014; Loh et al., 2013; Moorjani et al., 2016) and can be explained by the fact that continuous admixture leads to mixtures of exponential curves, and resolving the timing in such case can be challenging due to the well-known difficulty of fitting a sum of exponentials to data with even a small amount of noise (Osborne & Smyth, 1986).

macs command line for continuous admixture with $\lambda = 5$ generations:
macs 120 1e8 -t 6e-4 -r 5e-4 -I 3 50 20 50 -em 0.0002 2 1 2000 -em 0.0003 2 1 0 -ej 0.00032 2 3 -ej 0.036 1 3

Appendix 2—table 2: Impact of continuous gene flow. The table shows true and inferred times of admixture in *PopC* using *PopA* and *PopB* used as the reference populations.

The true period of continuous admixture, λ generations	Inferred time of admixture (mean ± 1 SE) is shown on Y-axis. generations)
10-15	15 ± 1
20-30	23 ± 2
40-60	53 ± 3
40-100	64 ± 4

(ii) Impact of founder event/ bottleneck post admixture

Many human populations have a history of founder events in their recent evolutionary past (Tournebize et al., 2020). A founder event generates long-range LD in the target population, which could in principle be spuriously inferred as admixture-related ancestry covariance, confounding the dates of admixture. To explore the effect of this scenario, we simulated *PopC* that has ancestry from *PopA* and *PopB* due to gene flow that occurred T_A generations ago. Following admixture, *PopC* experienced a bottleneck that occurred T_B (=10, 80 or 100) generations ago where the effective population size decreased from 12,500 to N_B (=10, 100, 500, and 1000) for a duration of D_B generations (=1, 5 or 10). After T_B, the population recovered to the original population size of 12,500 (Appendix 2—figure 4A). We applied *DATES* to *PopC* using *PopA* and *PopB* as reference populations and found that the estimated dates of admixture were accurate when the bottleneck was not extreme (Appendix 2—figure 4B). In the case of strong bottlenecks where N_B is less than 100, we observed a downward bias in the estimated admixture time.

Macs command line:
macs 120 1e8 -t 6e-4 -r 5e-4 -I 3 50 20 50 -em 0.002 2 1 10000 -em 0.00202 2 1 0 -en 0.0002 2 0.0002 -en 0.0003 2 1 -ej 0.00204 2 3 -ej 0.036 1 3

Another scenario we considered is when a population undergoes a severe bottleneck but does not recover (i.e., maintains a historically low population size to present). To test *DATES* for this demographic scenario, we simulated an admixed population that experienced a bottleneck post-admixture that occurred 100 generations ago. The effective population size was then reduced from 12,500 to N_B (=100–4000). This population maintained a small size until the present. Using *DATES* with the target as *PopC* and *PopA* and *PopB* as reference populations, we observed the inferred admixture times can be biased when $N_B < 1000$; there is no bias when the effective population size is larger (Appendix 2—figure 5, Appendix 2—table 3).

macs command line:
macs 120 1e8 -t 6e-4 -r 5e-4 -I 3 50 20 50 -em 0.002 2 1 10000 -em 0.00202 2 1 0 -en 0 2 0.04 -en 0.00198 2 1 -ej 0.00204 2 3 -ej 0.036 1 3

Appendix 2—table 3: Admixture time estimates from *DATES* for populations with extreme bottlenecks with a historically low population size that does not recover until the present.

Admixture	Ne before admixture	Ne post admixture	Inferred time of admixture
100	12500	4000	96 ± 5
		3500	96 ± 5
		3000	88 ± 4
		2500	99 ± 5

		2000	92 ± 6
		1500	78 ± 6
		1000	86 ± 6
		500	54 ± 9
		100	42 ± 10

(ii) Simulations with no admixture in the target population

To investigate if *DATES* gives spurious results for admixture in the absence of gene flow from the reference populations, we generated data for populations without a history of recent admixture. We simulated individuals for three populations *PopA*, *PopB,* and *PopC*, where the divergence between *PopA* and *PopB* was 1800 generations and divergence between *PopC* and *PopB* was 1000 generations. *PopC* had a bottleneck that occurred T_B (=100 or 10) generations ago where the population size reduced to N_B (=100 or 10). Applying *DATES* with *PopC* as the target with *PopA* and *PopB* as reference populations, we observed no evidence of ancestry decay in *PopC* — the ancestry covariance curves were noisy and the 95% CI for the dates included 0 (Appendix 2—figure 6).

macs command line:
macs 120 1e8 -t 6e-4 -r 5e-4 -I 3 50 20 50 -en 0.0002 2 0.0002 -en 0.0003 2 1 -ej 0.02 2 3 -ej 0.036 1 3

Further, we simulated a target population with a severe bottleneck without recovery to present (without any admixture), where *PopC* undergoes bottleneck at 100 or 500 generations without recovery till present. The effective population size reduces from 12,500 to 1000 or 650 at 100 generations or 500 generations and is maintained at the low size to present (Appendix 2—figure 7). Using *DATES* on *PopC* as a target with *PopA* and *PopB* as sources, we observed that the ancestry covariance curves were noisy and the dates were not significant. This shows that *DATES* does not provide spurious evidence for admixture even for populations that have a complex history with strong founder events.

macs command line:
macs 120 1e8 -t 6e-4 -r 5e-4 -I 3 50 20 50 -en 0 2 0.013 -en 0.00202 2 1 -ej 0.00204 2 3 -ej 0.036 1 3

Additional files

Appendix 2—figure 1: Model for multiple pulses of admixture. The admixed population (*Target)* derives ancestry from three populations, from the two gene flow events that occurred t_2 generations ago (older pulse) between *PopA* and *PopB* with α_1 / α_2 ancestries respectively resulting in an intermediate group *S*, which then mixes with *PopC* with α_3 ancestry *at* t_1 generations ago (younger pulse).

Appendix 2—figure 2: Multiple pulses of admixture with equal proportions of ancestry from sources. We generated a target population that has ancestry from three groups (*PopA*, *PopB*, *PopC*) with ancestry proportions of 33%, 33% and 33% respectively that mixed at two distinct times (t_1 = 30,60,100 and t_2 = 10 generations ago). We used CEU, YRI, and CHB as *PopA*, *PopB,* and *PopC* and varied the order of the three ancestrals, and applied *DATES* with pairs of populations as the reference to infer the timing of the mixture. We show the expected dates (t_1 or t_2 depending on the references used), the orange dashed line corresponds to the older pulse and the blue dashed line corresponds to the younger pulse of admixture. The blue points correspond to *DATES* estimates using Pop*A* and Pop*C* or Pop*B* and Pop*C* as references. The orange points correspond to *DATES* estimates using Pop*A* and Pop*B* as references. Panel (A) shows the admixture scenario with *PopA* = CEU, *PopB* = YRI and *PopC* = CHB, Panel (B) shows the admixture scenario with *PopA* = CEU, *PopB* = CHB and *PopC* = YRI, and Panel (C) shows the admixture scenario with *PopA* = CHB, *PopB* = YRI and *PopC* = CEU.

Appendix 2—figure 3: Two pulses of admixture with unequal proportions of ancestry from reference populations. We generated a target population that has ancestry from three groups with variation in ancestry from three sources. Model A: *PopA*, *PopB*, *PopC* with ancestry proportion of 4%, 16% and 80% respectively that mixed at two distinct times (t_1 = 30,60,100 and t_2 = 10 generations ago). Model B: *PopA*, *PopB*, *PopC* with ancestry proportion of 16%, 64% and 20% respectively that mixed at two distinct times (t_1 = 30,60,100 and t_2 = 10 generations ago). We used CEU, YRI, and CHB as *PopA*, *PopB,* and *PopC* and varied the order of the three ancestral populations, and applied *DATES* with pairs of populations as the reference to infer the timing of the mixture. Figures shows the true time of admixture on the X-axis and inferred time on Y-axis, the orange dashed line corresponds to the older pulse and the blue dashed line corresponds to the younger pulse of admixture. The blue points correspond to *DATES* estimates using Pop*A* and Pop*C* or Pop*B* and Pop*C* as references. The orange points correspond to *DATES* estimates using Pop*A* and Pop*B* as references. Panel (A) shows the admixture scenario with *PopA* = CEU, *PopB* = YRI and *PopC* = CHB, Panel (B) shows the admixture scenario with *PopA* = CEU, *PopB* = CHB and *PopC* = YRI, and Panel (C) shows the admixture scenario with *PopA* = CHB, *PopB* = YRI and *PopC* = CEU.

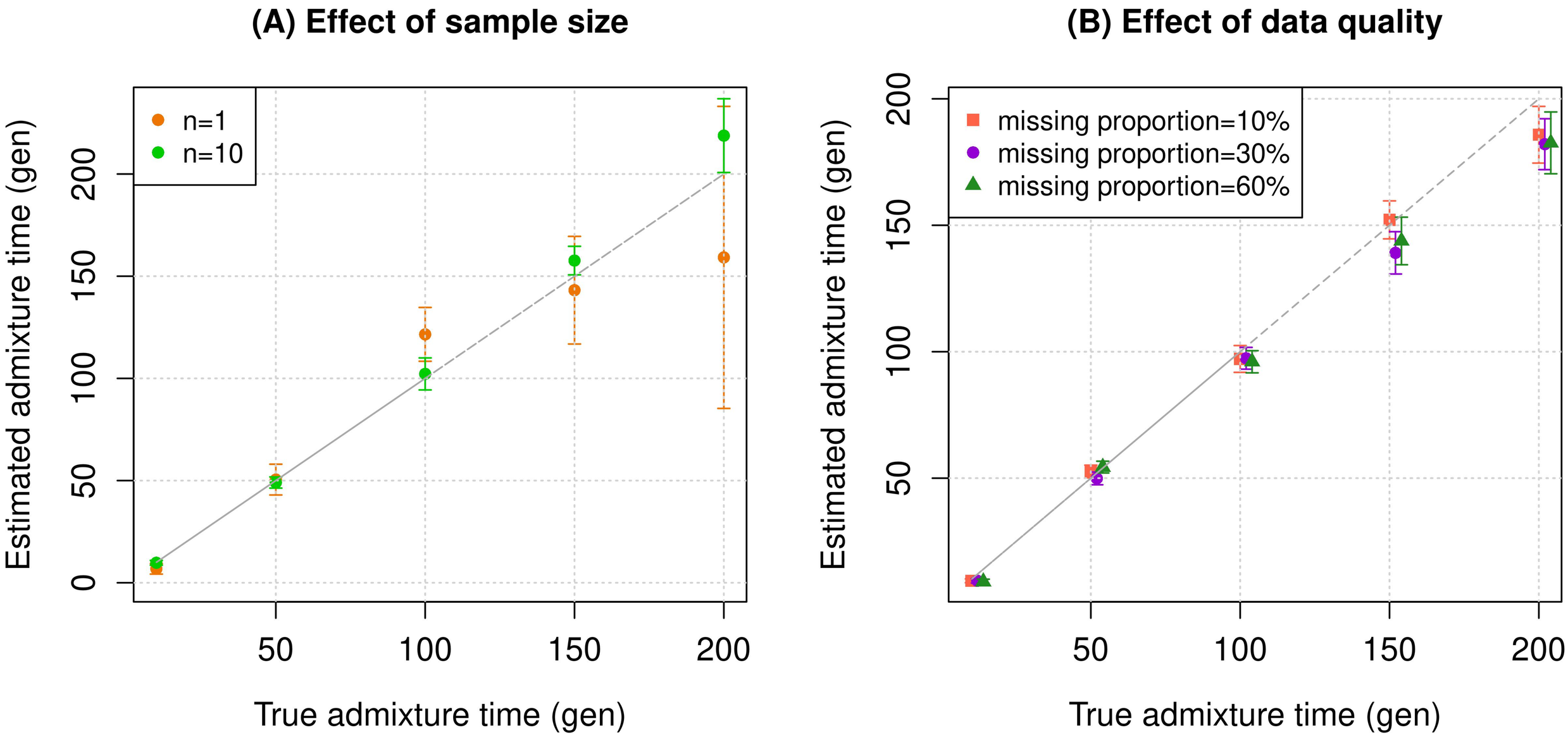
Appendix 2—figure 4: Impact of founder events on inferred dates of admixture: A) Schematic for demographic scenario shows that *PopC* was formed through admixture between *PopA* and *PopB* at time T_A. Following admixture, *PopC* experienced a severe bottleneck that occurred T_B generations ago where the effective population size decreased from 12,500 to N_B for a duration of D_B generations. After T_B, the population recovered to the original population size. B) We simulated data for 10 individuals with admixture occurring at 50, 100, and 200 generations with bottleneck post admixture for a period of D_B =1, 5, or 10 generations (shown in the legend) with the effective population size during bottleneck as N_B =10, 100, 500 or 1000 individuals (shown as four panels). The true admixture time is shown on X-axis, and the estimated time of admixture (± 1 SE) is shown on Y-axis.

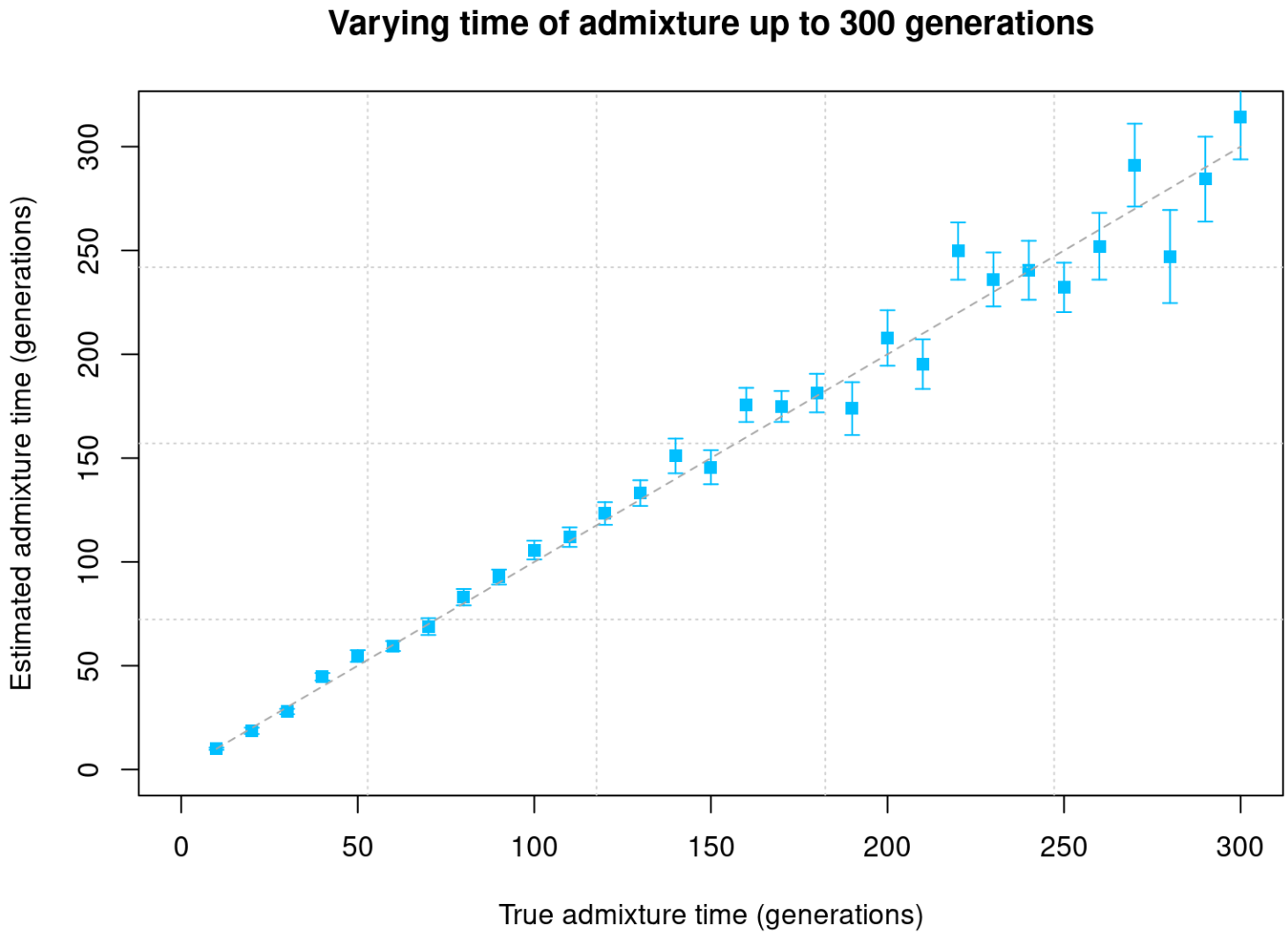
Appendix 2—figure 5: Impact of founder event with no recovery in admixed population: Schematic for the demographic history of the admixed group *PopC* that has ancestry from *PopA* and *PopB* followed by a severe bottleneck post-admixture without recovery to present (i.e.,

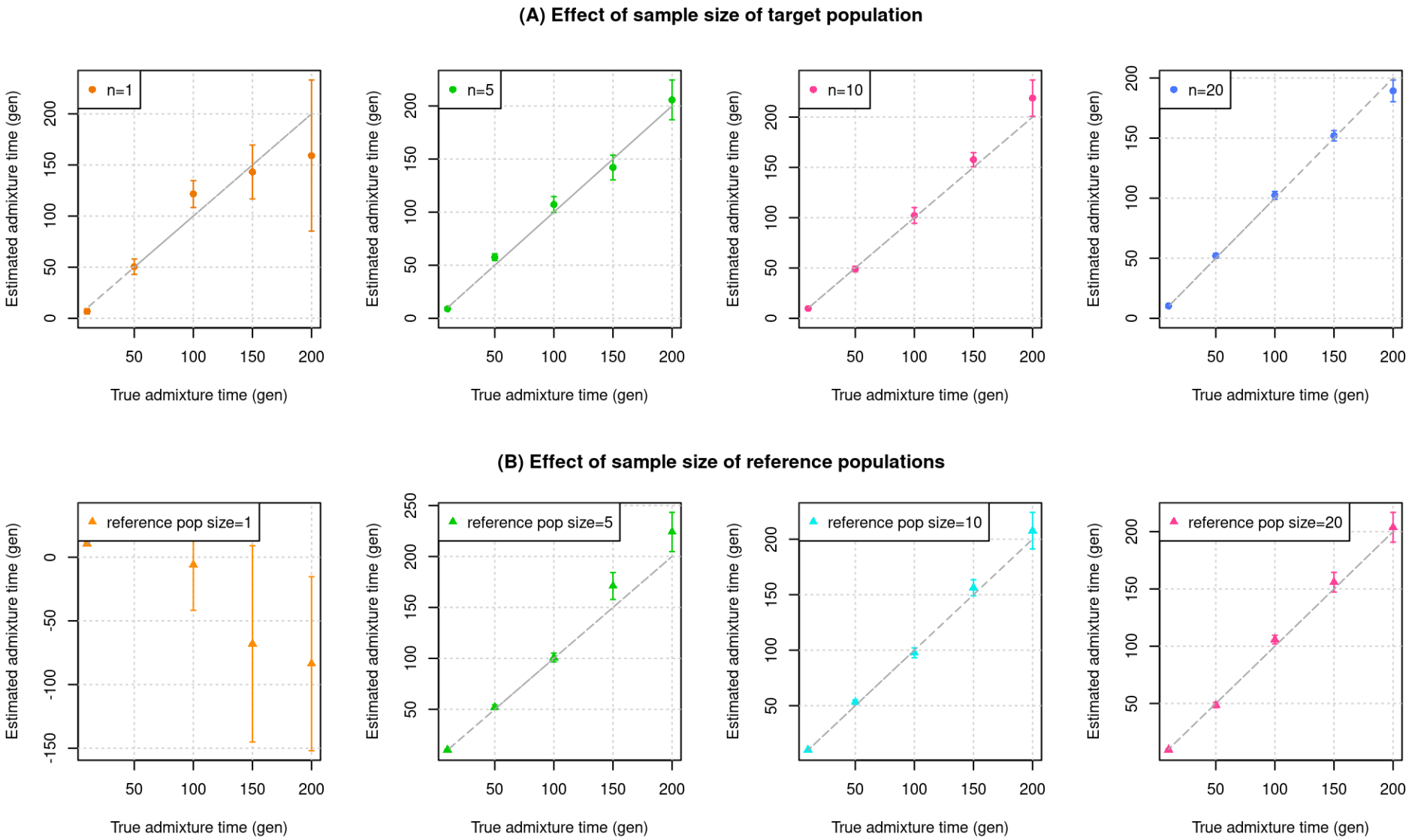
maintenance of historically low population size to present N_B). The F_{ST} (*PopA*, *PopC)* is 0.202 and F_{ST} (*PopB*, *PopC)* is 0.168.

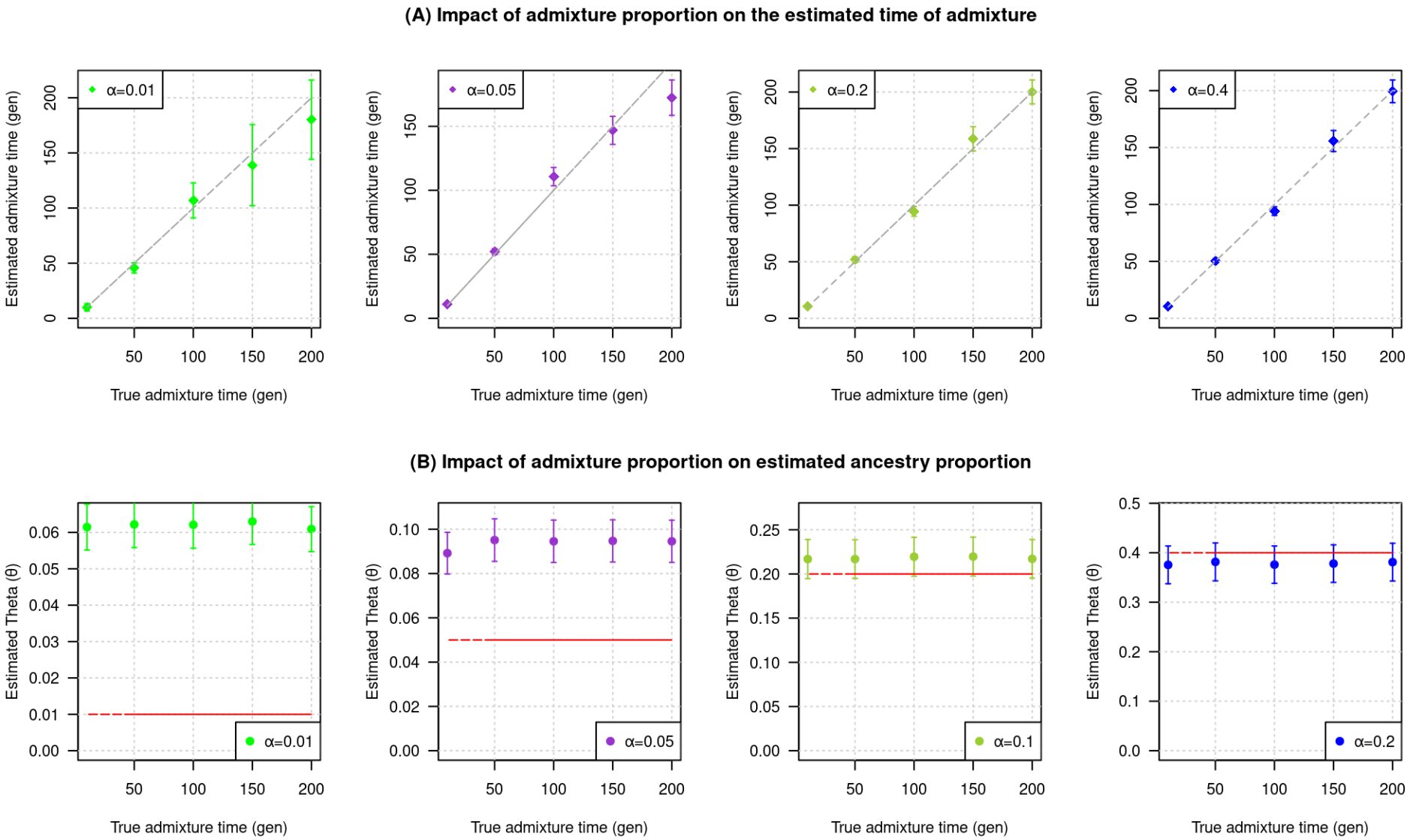
Appendix 2—figure 6: Impact of models with no admixture, severe founder event: A) Demographic scenario with three populations. *PopA* and *PopB* diverged 1800 generations ago and *PopB* and *PopC* diverged 1000 generations ago. *PopC* had a bottleneck at T_B generations ago with population size during bottleneck N_B. B) Ancestry covariance curves for *PopC*. We simulated data for 25 individuals from *PopA* and *PopB* and 10 individuals from *PopC*. We applied *DATES* on *PopC* with *PopA* and *PopB* as sources and show the decay curves for different timing and effective population size of founder events. Note, none of the simulations show significant exponential fits and all dates include 0 in the estimated CI.

Appendix 2—figure 7: Effect of drift, no admixture: Impact of models with no admixture, severe founder event (without recovery): A) Demographic scenario with three populations. *PopA* and *PopB* diverged 1800 generations ago and *PopB* and *PopC* diverged 1000 generations ago. *PopC* had a bottleneck at T_B generations ago with population size during bottleneck N_B. The population size N_B is maintained to present. B) Ancestry covariance curves for *PopC*. We simulated data for 25 individuals from *PopA* and *PopB* and 10 individuals from *PopC*. We applied *DATES* on *PopC* with *PopA* and *PopB* as sources and show the decay curves for different timing and effective population size of founder events.

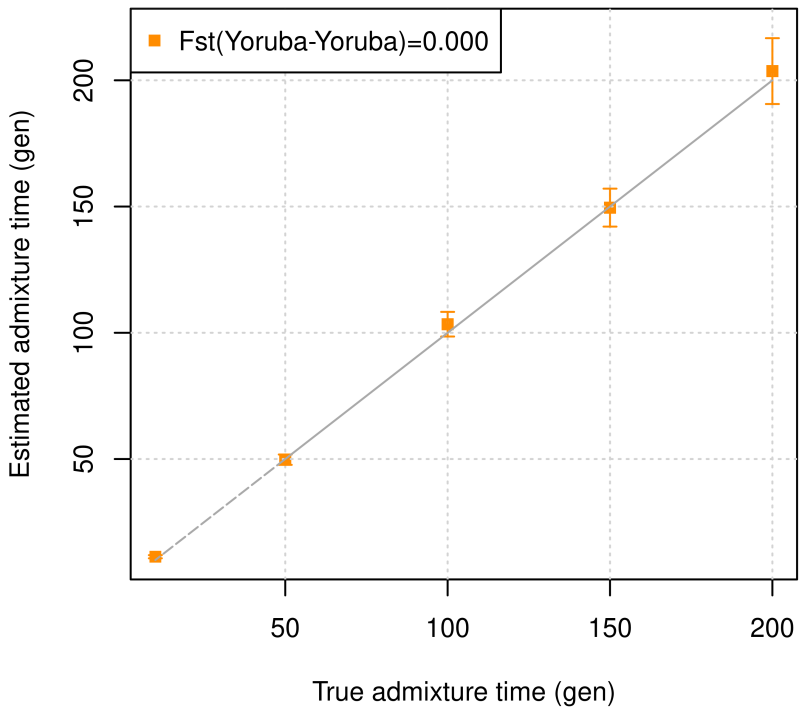

(A) Effect of sample size
n=1
n=10
Estimated admixture time (gen)
True admixture time (gen)
0
50
100
150
200
(B) Effect of data quality
missing proportion=10%
missing proportion=30%
missing proportion=60%
Estimated admixture time (gen)
True admixture time (gen)
50
100
150
200

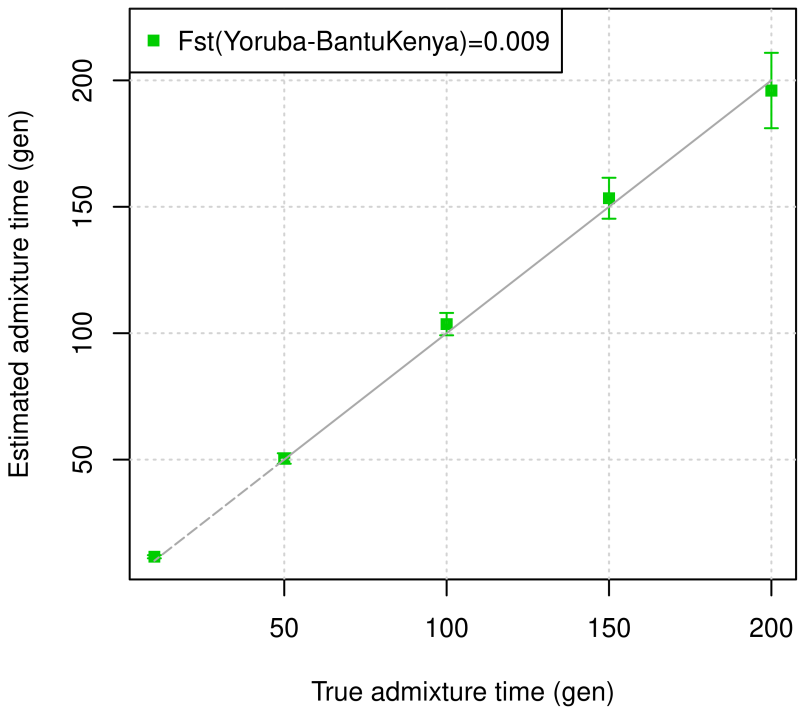

Varying time of admixture up to 300 generations
Estimated admixture time (generations)
True admixture time (generations)
0
50
100
150
200
250
300

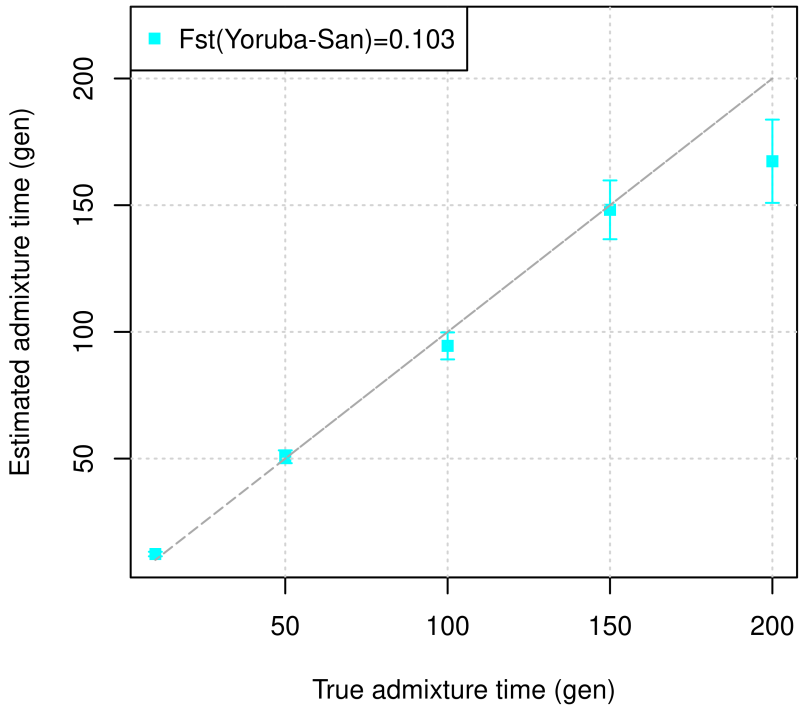

(A) Effect of sample size of target population
n=1
n=5
n=10
n=20
Estimated admixture time (gen)
True admixture time (gen)
(B) Effect of sample size of reference populations
reference pop size=1
reference pop size=5
reference pop size=10
reference pop size=20
Estimated admixture time (gen)
True admixture time (gen)

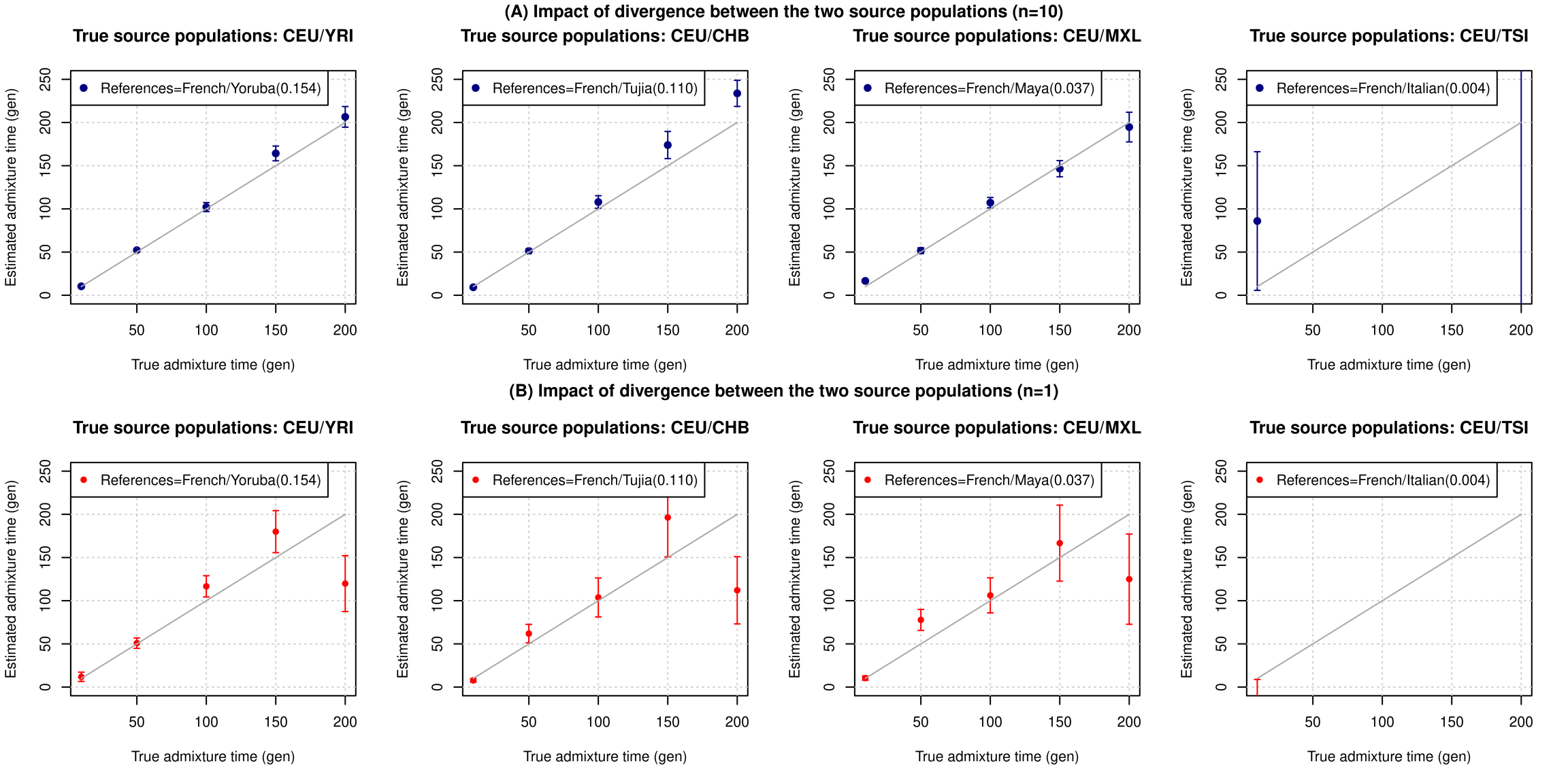

(A) Impact of admixture proportion on the estimated time of admixture
α=0.01
α=0.05
α=0.2
α=0.4
Estimated admixture time (gen)
True admixture time (gen)
(B) Impact of admixture proportion on estimated ancestry proportion
α=0.01
α=0.05
α=0.1
α=0.2
Estimated Theta (θ)
True admixture time (gen)

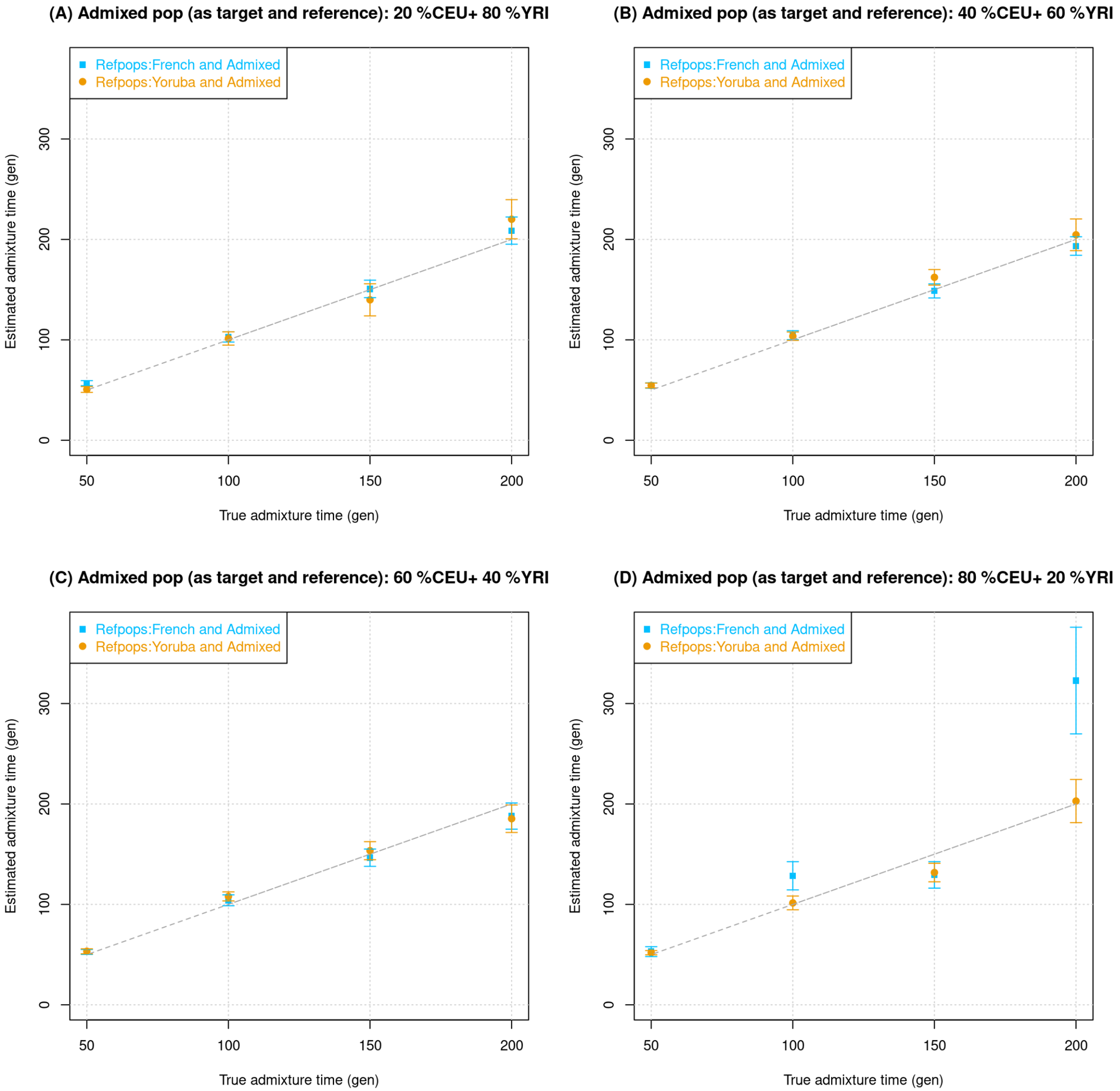
Impact of divergence between the ancestral population and reference populations used in DATES

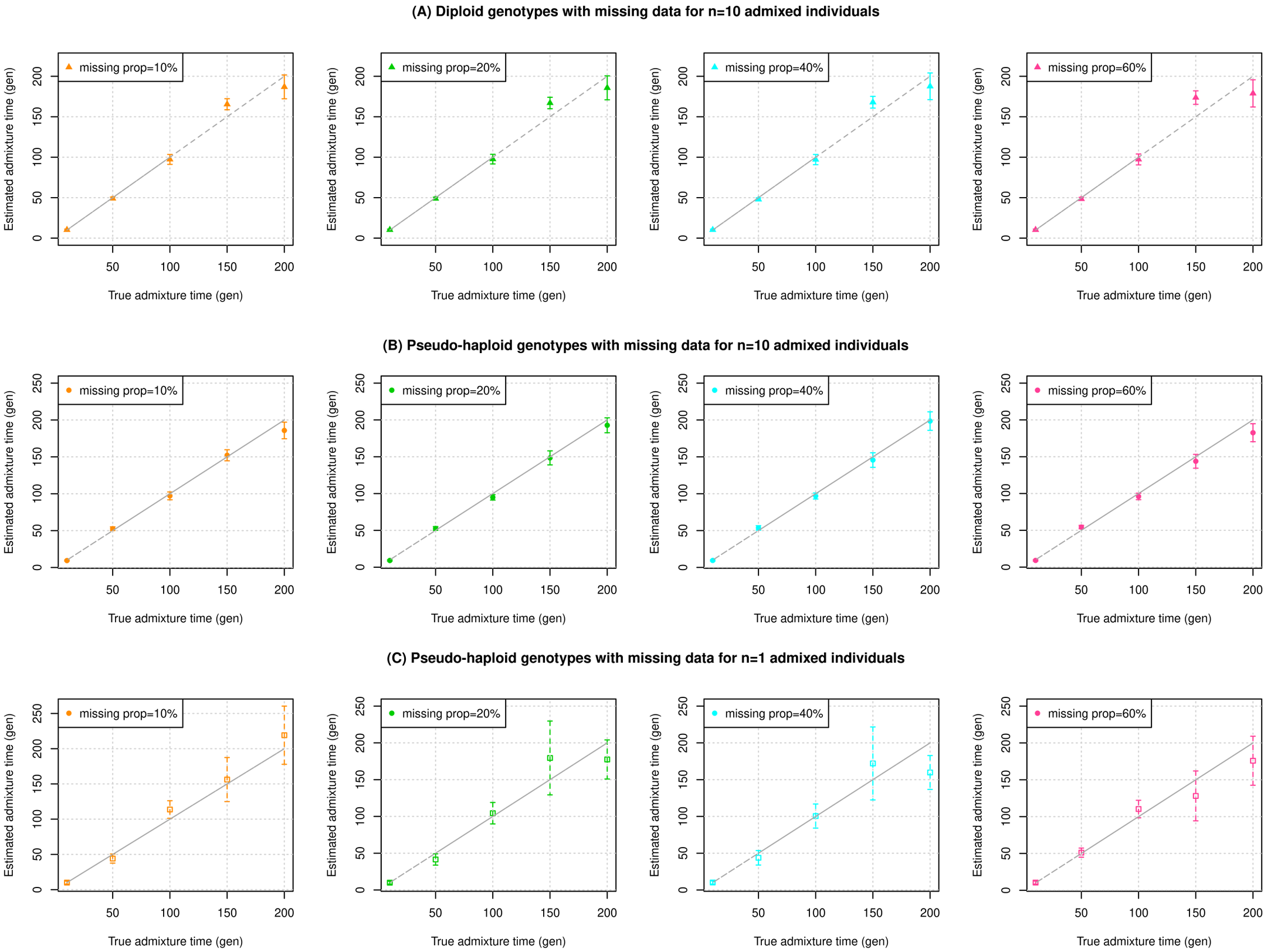


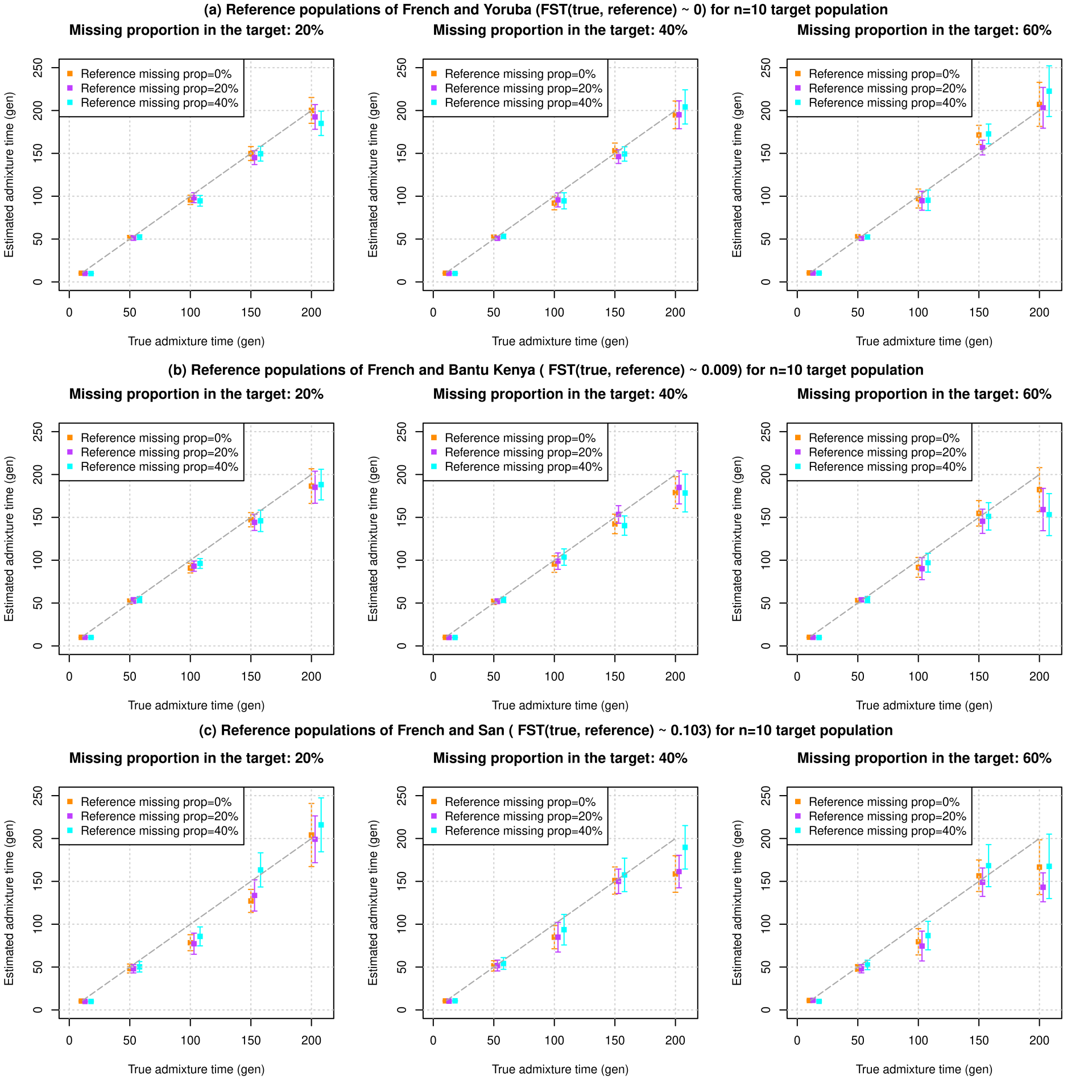


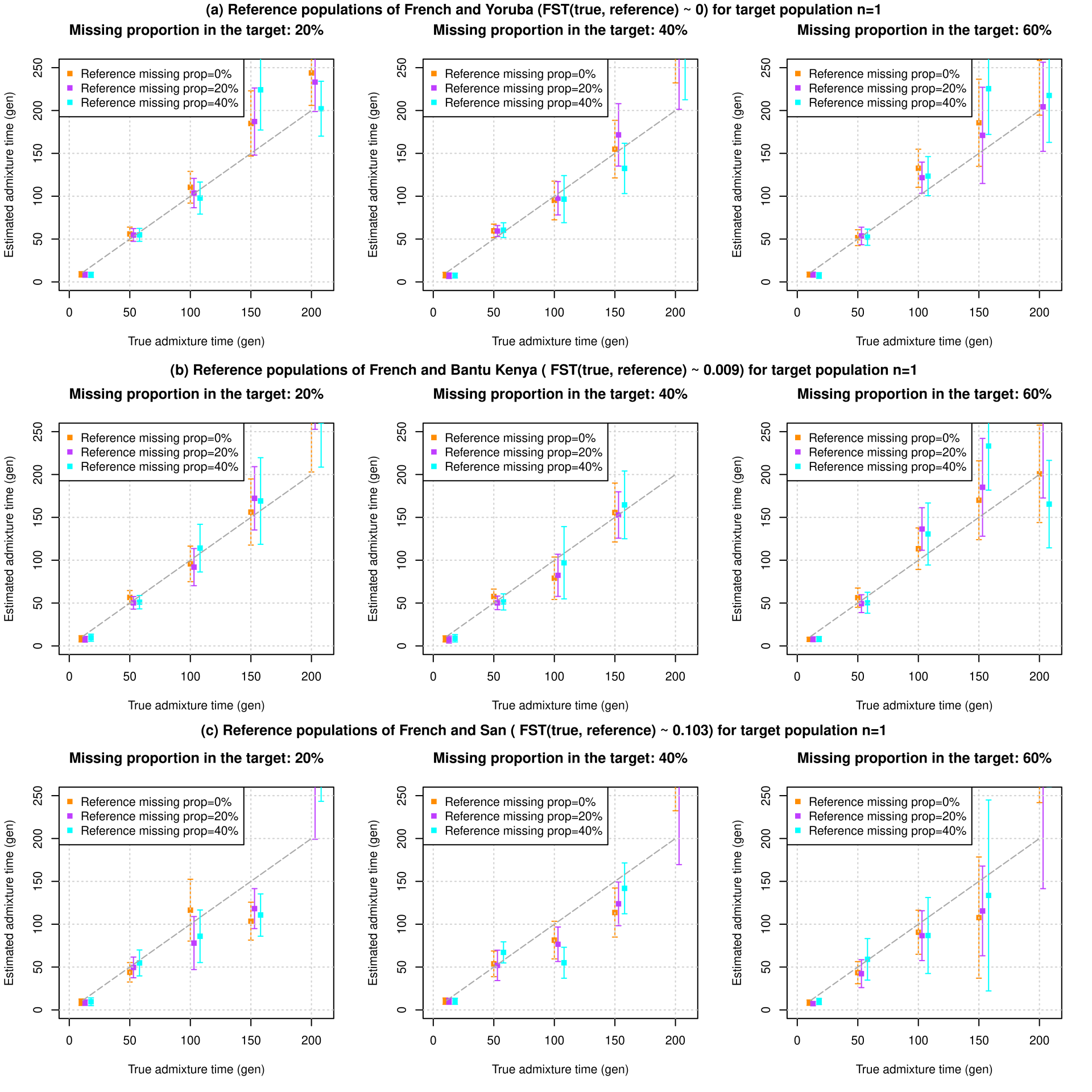


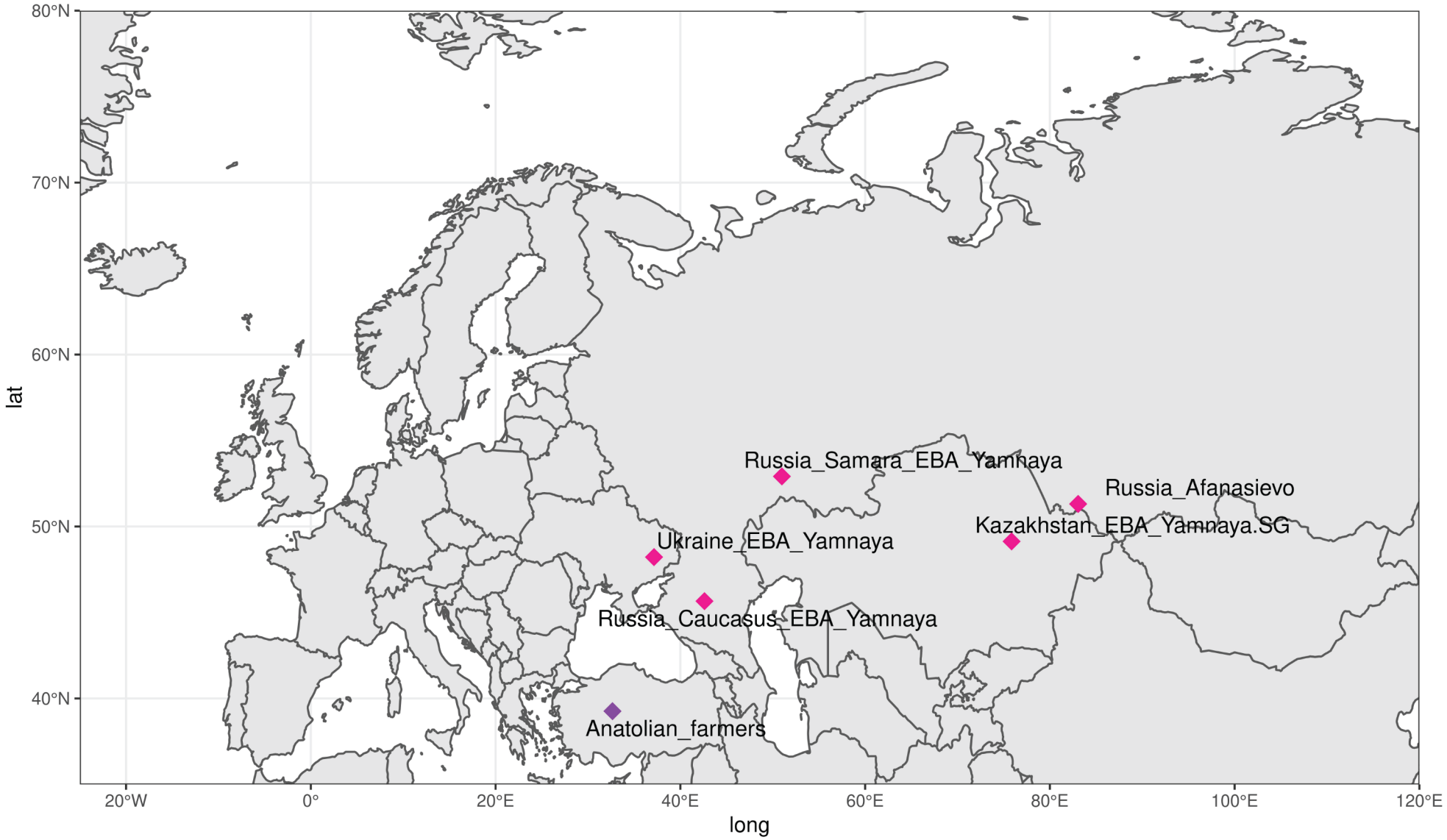

(A) Impact of divergence between the two source populations (n=10)
True source populations: CEU/YRI
References=French/Yoruba(0.154)
True source populations: CEU/CHB
References=French/Tujia(0.110)
True source populations: CEU/MXL
References=French/Maya(0.037)
True source populations: CEU/TSI
References=French/Italian(0.004)
(B) Impact of divergence between the two source populations (n=1)
True source populations: CEU/YRI
References=French/Yoruba(0.154)
True source populations: CEU/CHB
References=French/Tujia(0.110)
True source populations: CEU/MXL
References=French/Maya(0.037)
True source populations: CEU/TSI
References=French/Italian(0.004)
Estimated admixture time (gen)
True admixture time (gen)
0
50
100
150
200
250

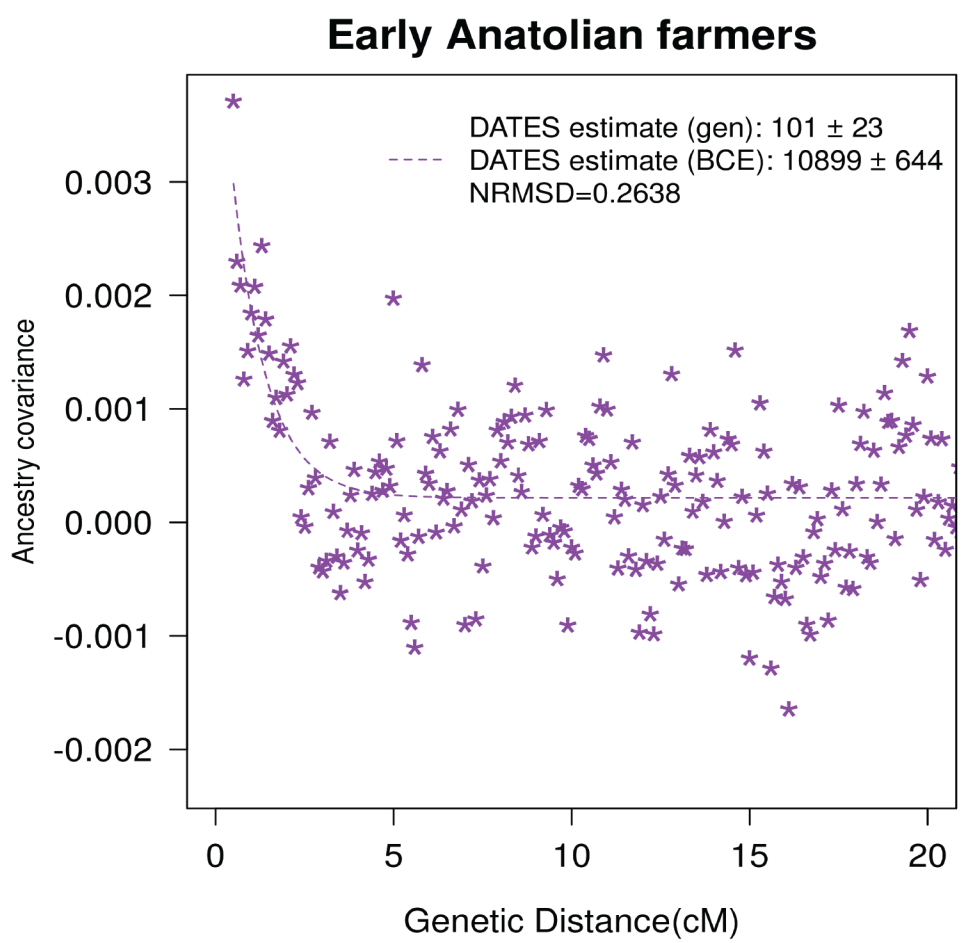

(A) Admixed pop (as target and reference): 20 %CEU+ 80 %YRI
(B) Admixed pop (as target and reference): 40 %CEU+ 60 %YRI
(C) Admixed pop (as target and reference): 60 %CEU+ 40 %YRI
(D) Admixed pop (as target and reference): 80 %CEU+ 20 %YRI
Refpops:French and Admixed
Refpops:Yoruba and Admixed
Estimated admixture time (gen)
True admixture time (gen)
0
100
200
300
50
100
150
200

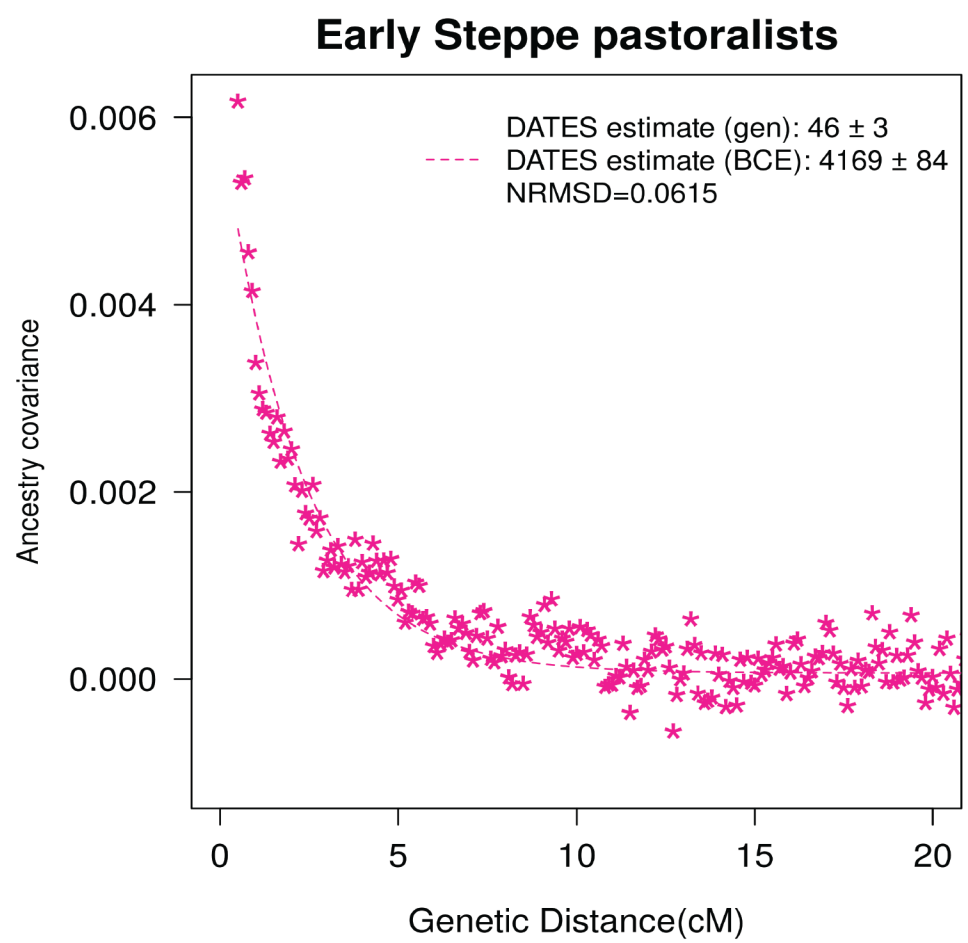

(A) Diploid genotypes with missing data for n=10 admixed individuals
missing prop=10%
missing prop=20%
missing prop=40%
missing prop=60%
(B) Pseudo-haploid genotypes with missing data for n=10 admixed individuals
missing prop=10%
missing prop=20%
missing prop=40%
missing prop=60%
(C) Pseudo-haploid genotypes with missing data for n=1 admixed individuals
missing prop=10%
missing prop=20%
missing prop=40%
missing prop=60%
Estimated admixture time (gen)
True admixture time (gen)
0
50
100
150
200
250

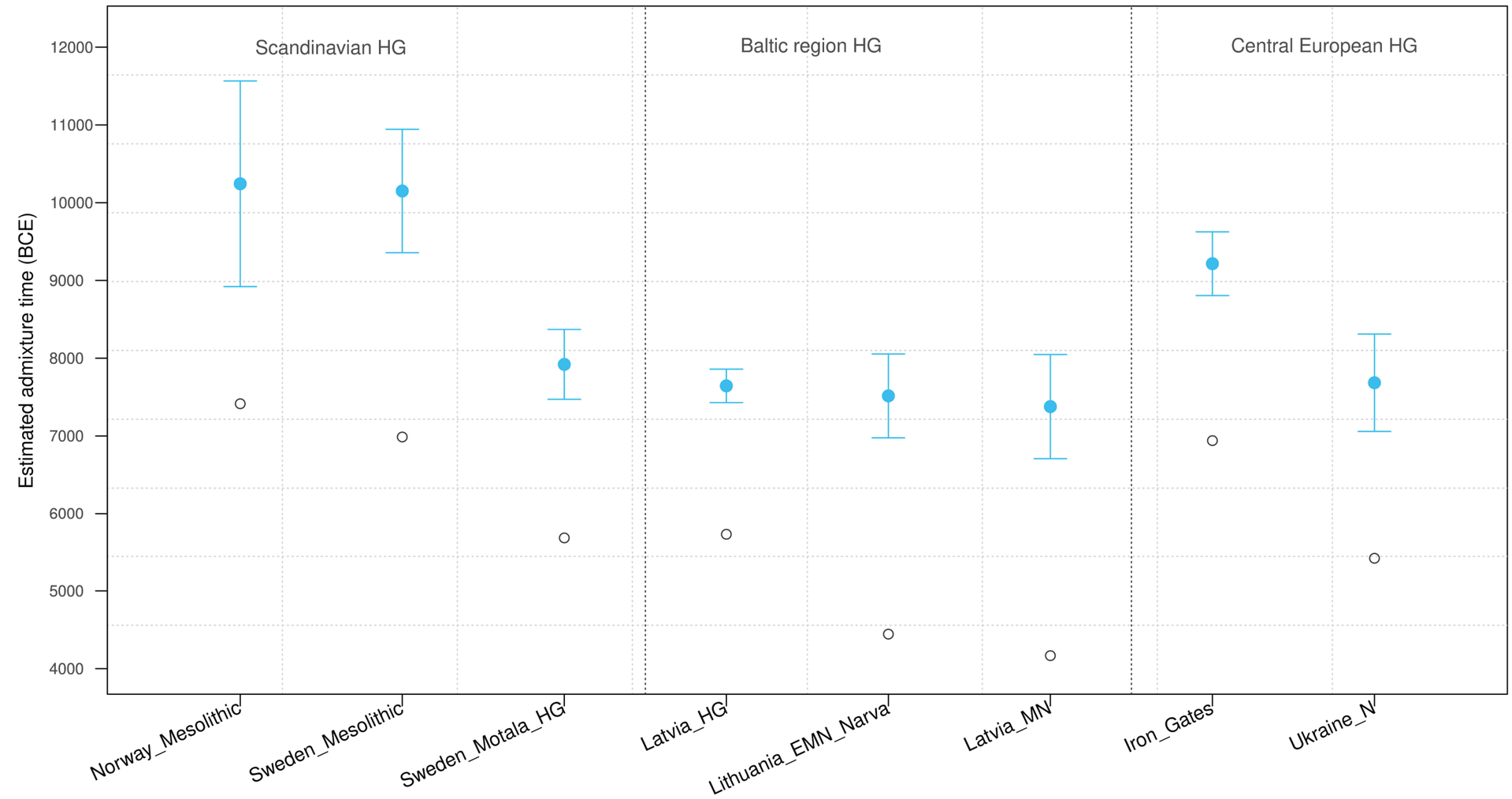

(a) Reference populations of French and Yoruba (FST(true, reference) ~ 0) for n=10 target population
Missing proportion in the target: 20%
Missing proportion in the target: 40%
Missing proportion in the target: 60%
(b) Reference populations of French and Bantu Kenya (FST(true, reference) ~ 0.009) for n=10 target population
Missing proportion in the target: 20%
Missing proportion in the target: 40%
Missing proportion in the target: 60%
(c) Reference populations of French and San (FST(true, reference) ~ 0.103) for n=10 target population
Missing proportion in the target: 20%
Missing proportion in the target: 40%
Missing proportion in the target: 60%
Reference missing prop=0%
Reference missing prop=20%
Reference missing prop=40%
Estimated admixture time (gen)
True admixture time (gen)
0
50
100
150
200
250

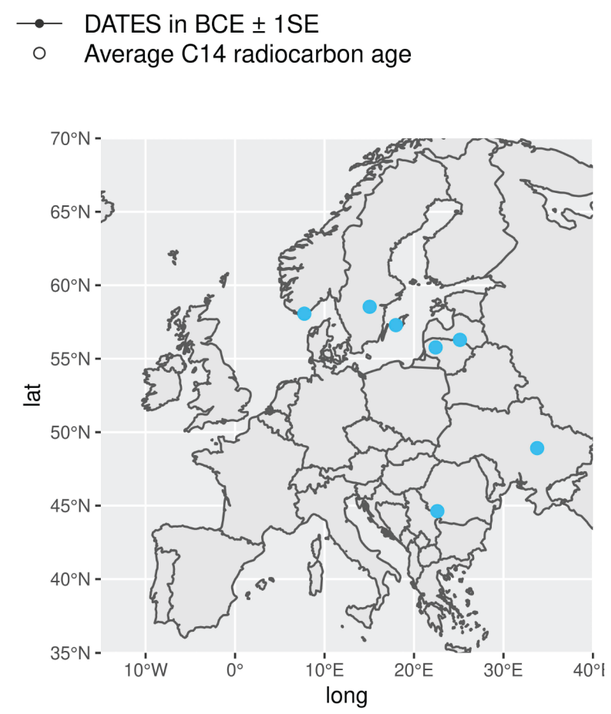

(a) Reference populations of French and Yoruba (FST(true, reference) ~ 0) for target population n=1
Missing proportion in the target: 20%
Missing proportion in the target: 40%
Missing proportion in the target: 60%
(b) Reference populations of French and Bantu Kenya (FST(true, reference) ~ 0.009) for target population n=1
Missing proportion in the target: 20%
Missing proportion in the target: 40%
Missing proportion in the target: 60%
(c) Reference populations of French and San (FST(true, reference) ~ 0.103) for target population n=1
Missing proportion in the target: 20%
Missing proportion in the target: 40%
Missing proportion in the target: 60%
Reference missing prop=0%
Reference missing prop=20%
Reference missing prop=40%
Estimated admixture time (gen)
True admixture time (gen)
0
50
100
150
200
250

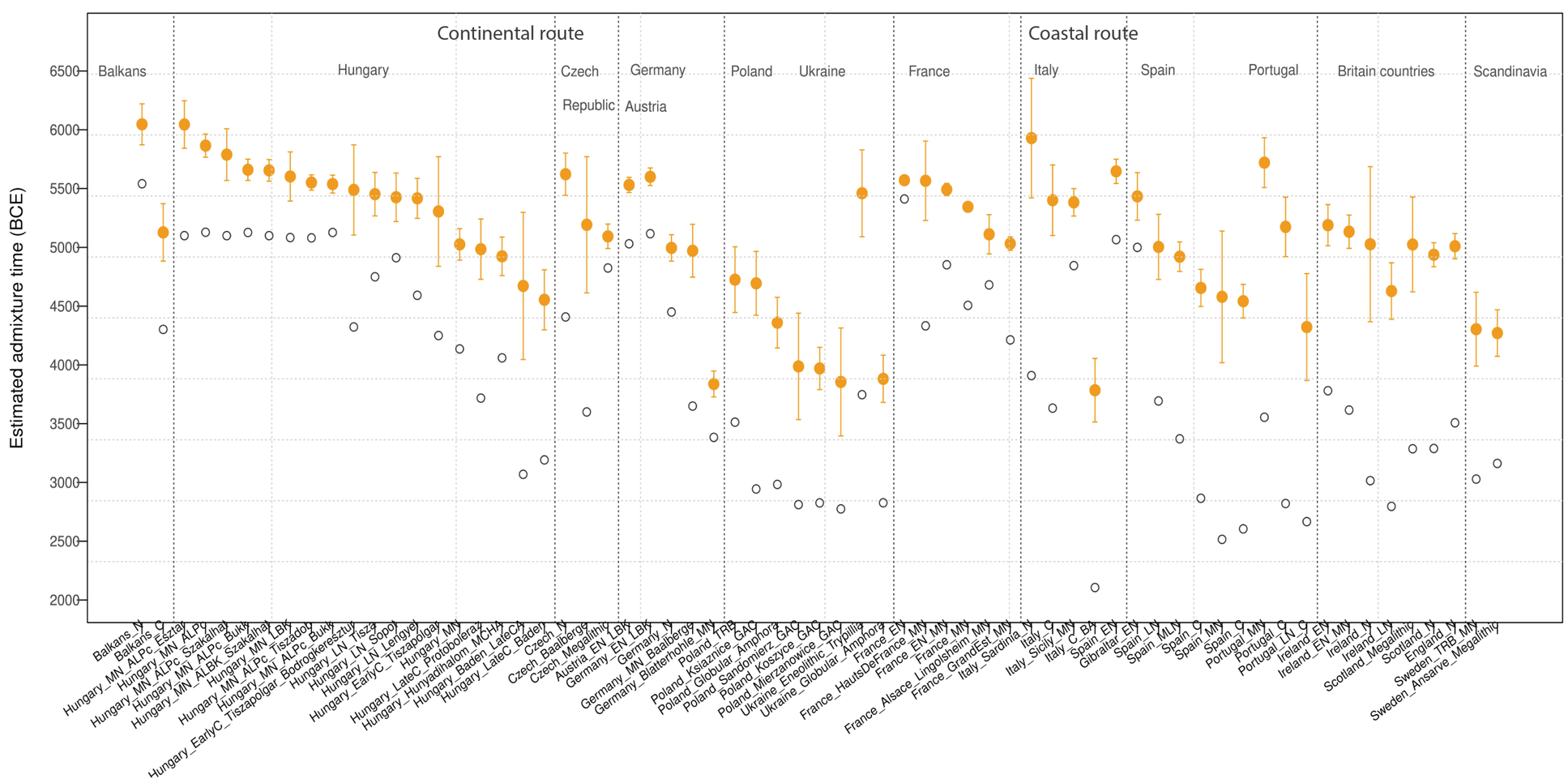

80°N
70°N
60°N
50°N
40°N
lat
20°W
0°
20°E
40°E
60°E
80°E
100°E
120°E
long
Russia_Samara_EBA_Yamnaya
Russia_Afanasievo
Kazakhstan_EBA_Yamnaya.SG
Ukraine_EBA_Yamnaya
Russia_Caucasus_EBA_Yamnaya
Anatolian_farmers

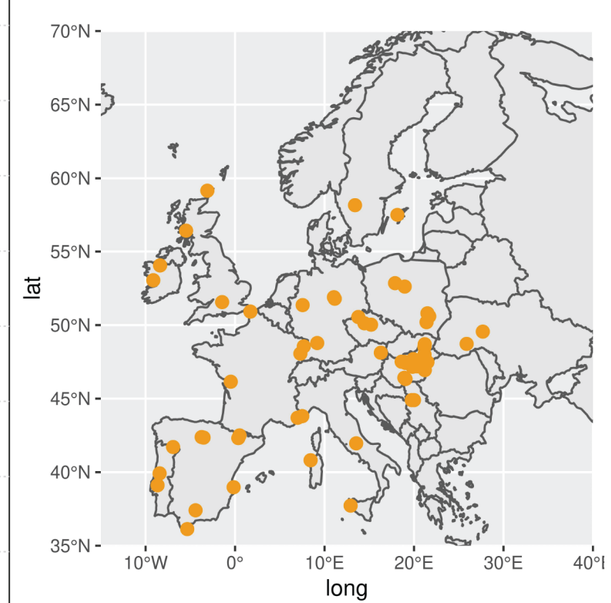

Early Anatolian farmers
DATES estimate (gen): 101 ± 23
DATES estimate (BCE): 10899 ± 644
NRMSD=0.2638
Ancestry covariance
0.003
0.002
0.001
0.000
-0.001
-0.002
0
5
10
15
20
Genetic Distance(cM)

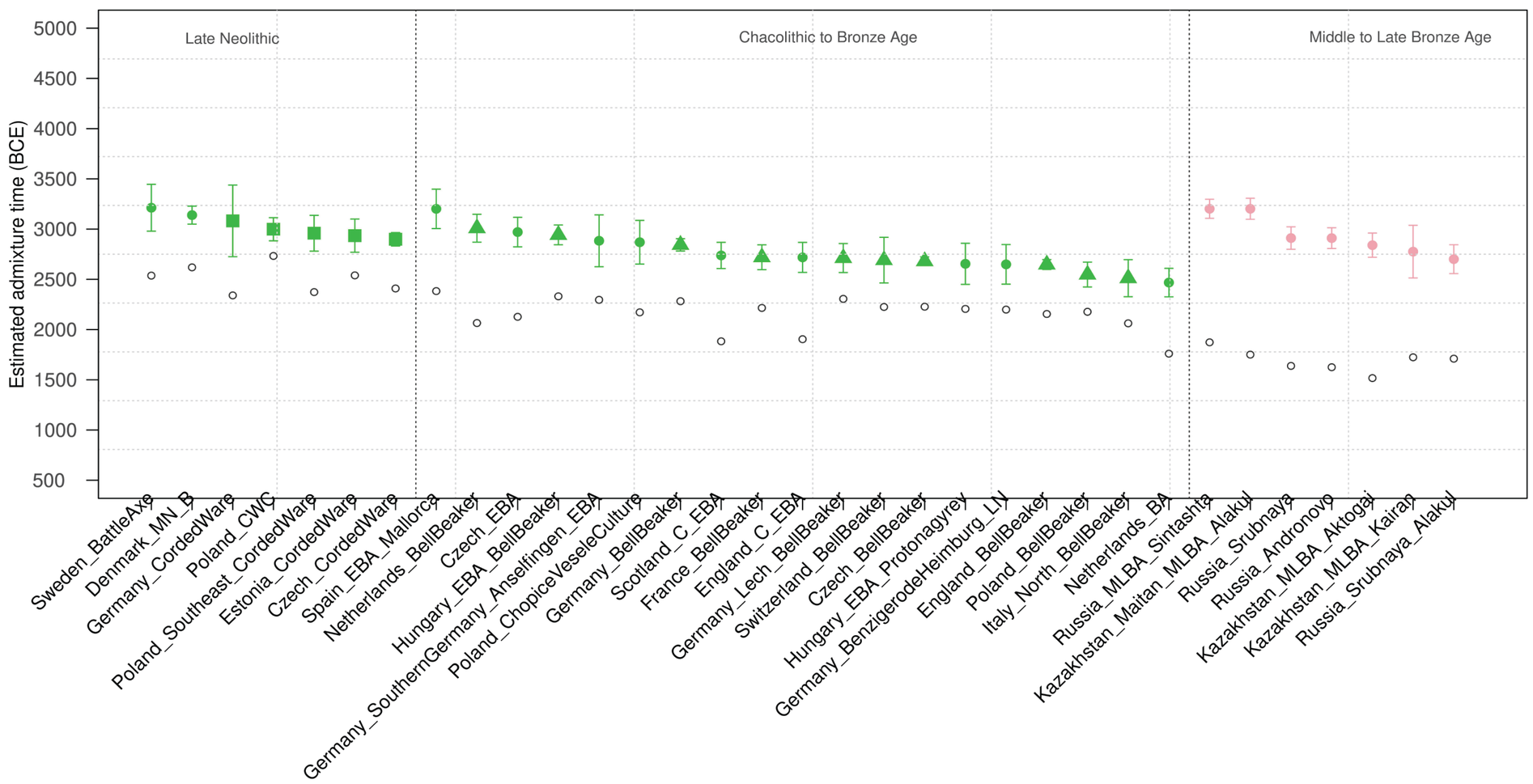

Early Steppe pastoralists
DATES estimate (gen): 46 ± 3
DATES estimate (BCE): 4169 ± 84
NRMSD=0.0615
Ancestry covariance
0.006
0.004
0.002
0.000
0
5
10
15
20
Genetic Distance(cM)

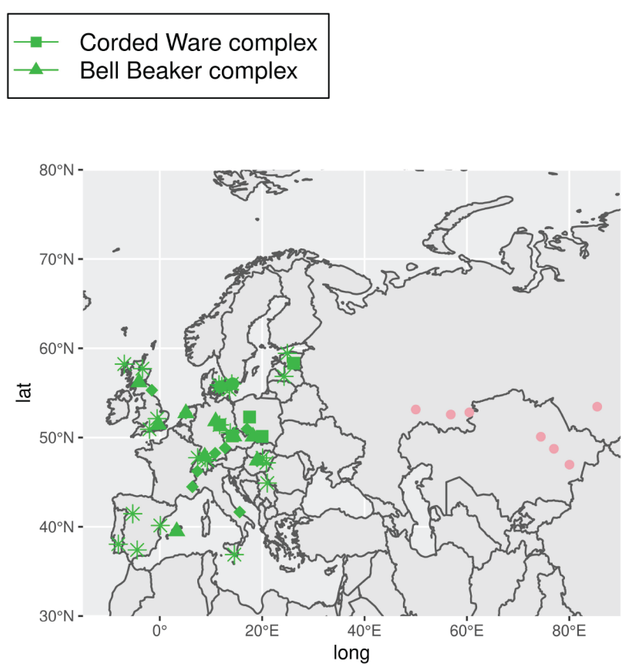

Scandinavian HG
Baltic region HG
Central European HG
Estimated admixture time (BCE)
12000
11000
10000
9000
8000
7000
6000
5000
4000
Norway_Mesolithic
Sweden_Mesolithic
Sweden_Motala_HG
Latvia_HG
Lithuania_EMN_Narva
Latvia_MN
Iron_Gates
Ukraine_N

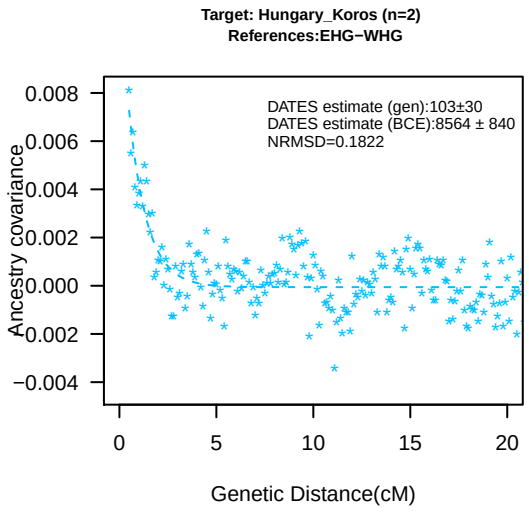

DATES in BCE ± 1SE
Average C14 radiocarbon age
70°N
65°N
60°N
55°N
50°N
45°N
40°N
35°N
lat
10°W
0°
10°E
20°E
30°E
long

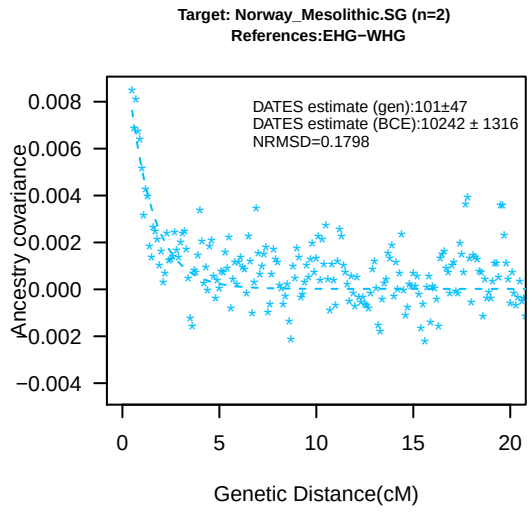

Continental route
Coastal route
Balkans
Hungary
Czech Republic
Germany
Austria
Poland
Ukraine
France
Italy
Spain
Portugal
Britain countries
Scandinavia
Estimated admixture time (BCE)
6500
6000
5500
5000
4500
4000
3500
3000
2500
2000

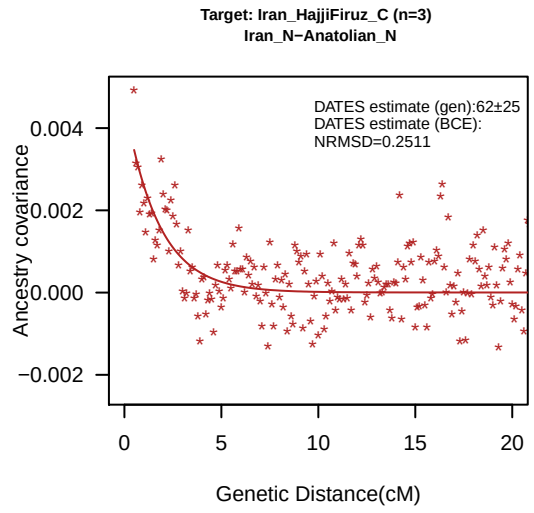

70°N
65°N
60°N
55°N
50°N
45°N
40°N
35°N
lat
10°W
0°
10°E
20°E
30°E
long

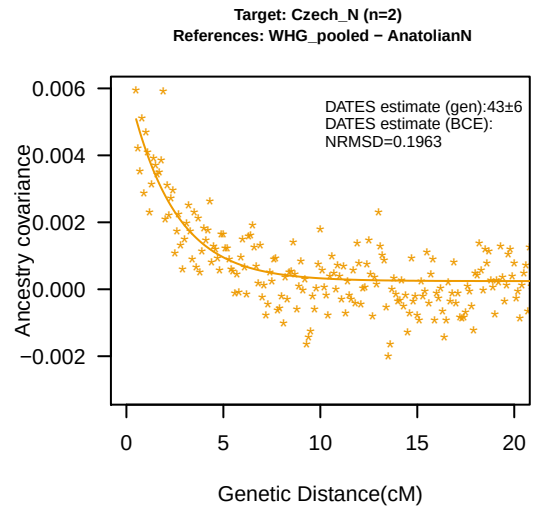

Late Neolithic
Chacolithic to Bronze Age
Middle to Late Bronze Age
Estimated admixture time (BCE)
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
Sweden_BattleAxe
Denmark_MN_B
Germany_CordedWare
Poland_CWC
Poland_Southeast_CordedWare
Estonia_CordedWare
Czech_CordedWare
Spain_EBA_Mallorca
Netherlands_BellBeaker
Czech_EBA
Hungary_EBA_BellBeaker
Germany_SouthernGermany_Anselfingen_EBA
Poland_ChopiceVeseleCulture
Germany_BellBeaker
Scotland_C_EBA
France_BellBeaker
England_C_EBA
Germany_Lech_BellBeaker
Switzerland_BellBeaker
Czech_BellBeaker
Hungary_EBA_Protonagyrev
Germany_BenzigerodeHeimburg_LN
England_BellBeaker
Poland_BellBeaker
Italy_North_BellBeaker
Netherlands_BA
Russia_MLBA_Sintashta
Kazakhstan_Maitan_MLBA_Alakul
Russia_Srubnaya
Russia_Andronovo
Kazakhstan_MLBA_Aktogai
Kazakhstan_MLBA_Kairan
Russia_Srubnaya_Alakul

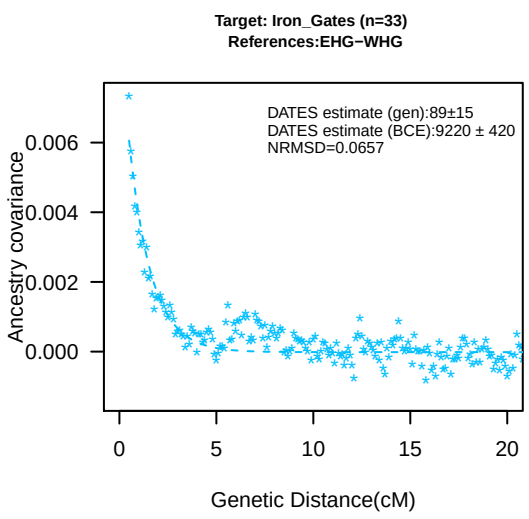

Corded Ware complex
Bell Beaker complex
80°N
70°N
60°N
50°N
40°N
30°N
lat
0°
20°E
40°E
60°E
80°E
long

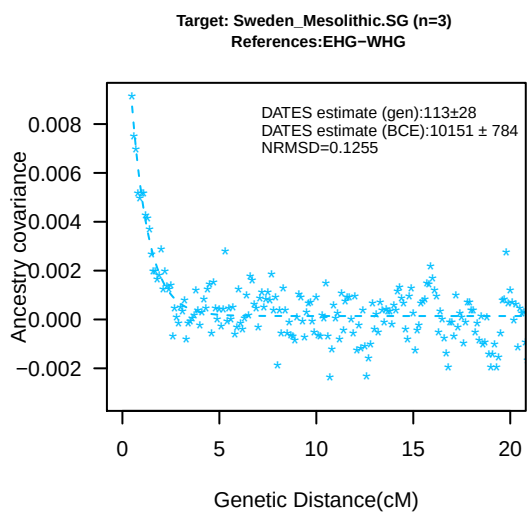

Target: Hungary_Koros (n=2)
References:EHG−WHG
DATES estimate (gen):103±30
DATES estimate (BCE):8564 ± 840
NRMSD=0.1822
Ancestry covariance
Genetic Distance(cM)

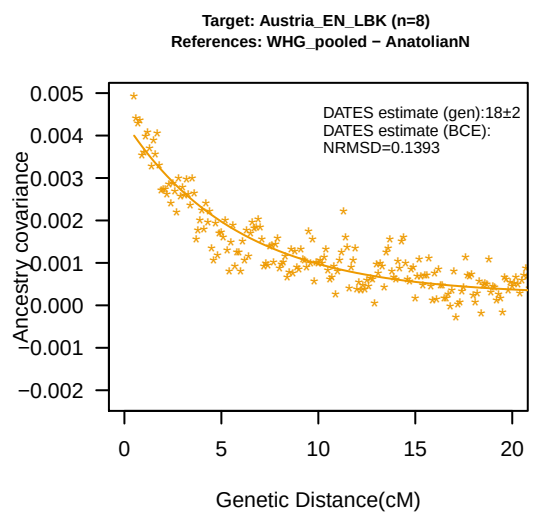

Target: Norway_Mesolithic.SG (n=2)
References:EHG−WHG
DATES estimate (gen):101±47
DATES estimate (BCE):10242 ± 1316
NRMSD=0.1798
Ancestry covariance
Genetic Distance(cM)

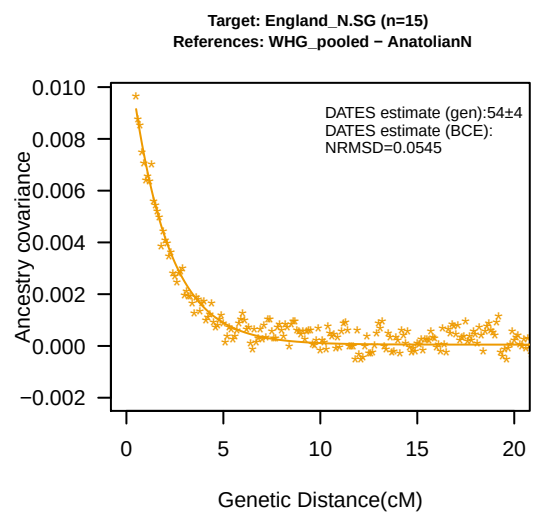

Target: Iran_HajjiFiruz_C (n=3)
Iran_N−Anatolian_N
DATES estimate (gen):62±25
DATES estimate (BCE):
NRMSD=0.2511
Ancestry covariance
Genetic Distance(cM)

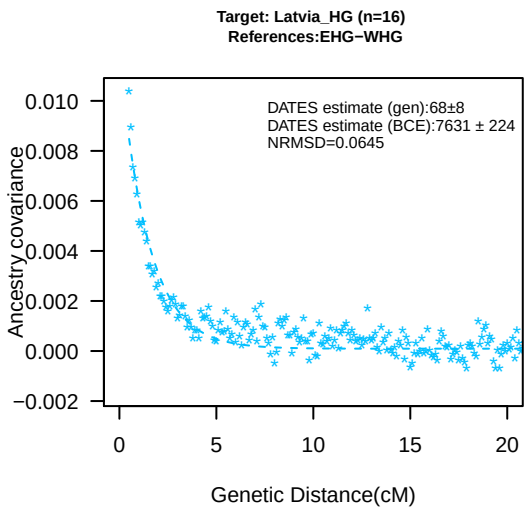

Target: Czech_N (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):43±6
DATES estimate (BCE):
NRMSD=0.1963
Ancestry covariance
Genetic Distance(cM)

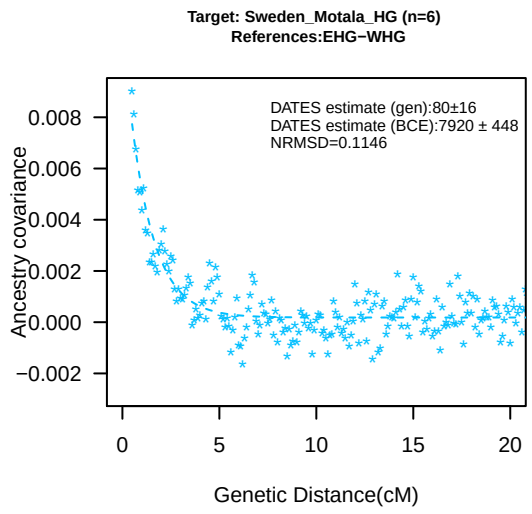

Target: Iron_Gates (n=33)
References:EHG−WHG
DATES estimate (gen):89±15
DATES estimate (BCE):9220 ± 420
NRMSD=0.0657
Ancestry covariance
Genetic Distance(cM)

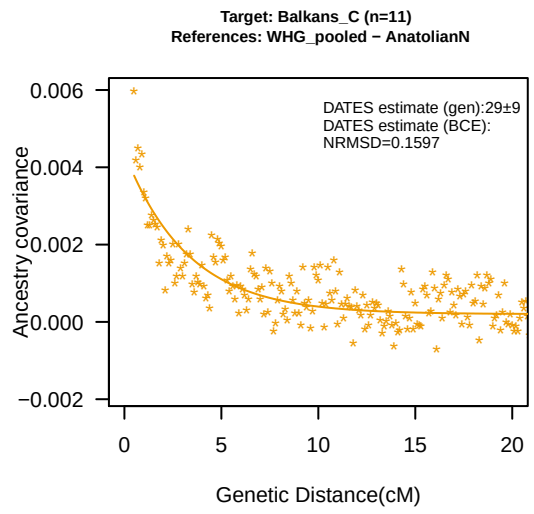

Target: Sweden_Mesolithic.SG (n=3)
References:EHG−WHG
DATES estimate (gen):113±28
DATES estimate (BCE):10151 ± 784
NRMSD=0.1255
Ancestry covariance
Genetic Distance(cM)

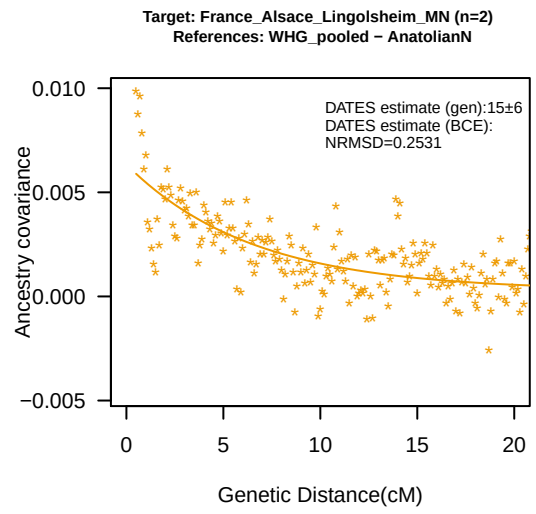

Target: Austria_EN_LBK (n=8)
References: WHG_pooled − AnatolianN
DATES estimate (gen):18±2
DATES estimate (BCE):
NRMSD=0.1393
Ancestry covariance
Genetic Distance(cM)

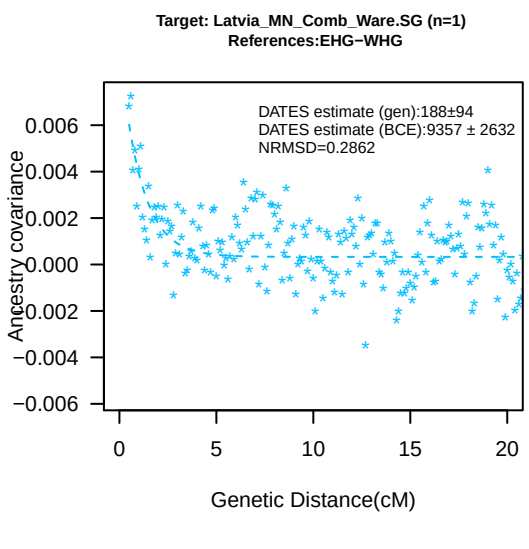

Target: England_N.SG (n=15)
References: WHG_pooled − AnatolianN
DATES estimate (gen):54±4
DATES estimate (BCE):
NRMSD=0.0545
Ancestry covariance
Genetic Distance(cM)

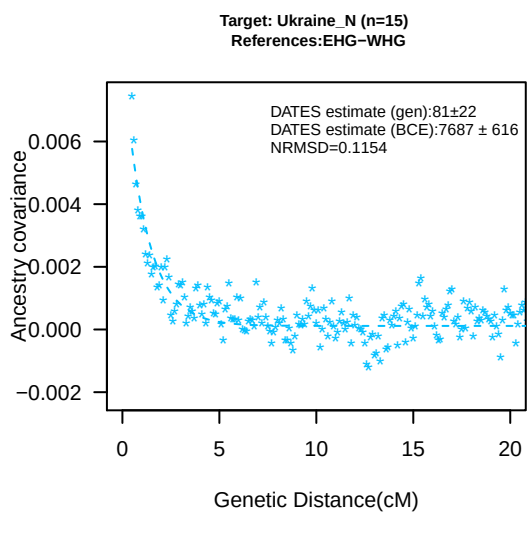

Target: Latvia_HG (n=16)
References:EHG−WHG
DATES estimate (gen):68±8
DATES estimate (BCE):7631 ± 224
NRMSD=0.0645
Ancestry covariance
Genetic Distance(cM)

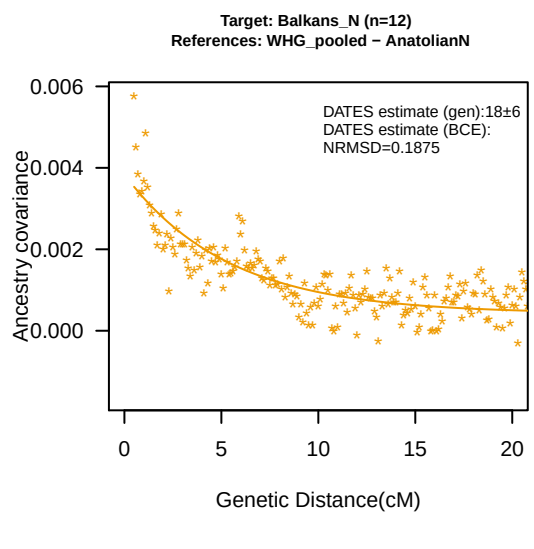

Target: Sweden_Motala_HG (n=6)
References:EHG−WHG
DATES estimate (gen):80±16
DATES estimate (BCE):7920 ± 448
NRMSD=0.1146
Ancestry covariance
Genetic Distance(cM)

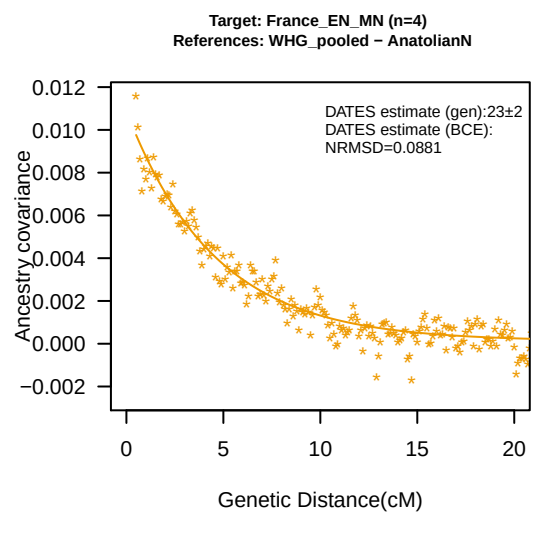

Target: Balkans_C (n=11)
References: WHG_pooled − AnatolianN
DATES estimate (gen):29±9
DATES estimate (BCE):
NRMSD=0.1597
Ancestry covariance
Genetic Distance(cM)

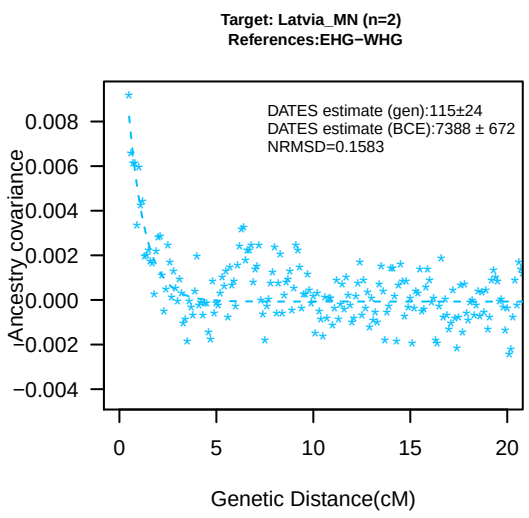

Target: France_Alsace_Lingolsheim_MN (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):15±6
DATES estimate (BCE):
NRMSD=0.2531
Ancestry covariance
Genetic Distance(cM)

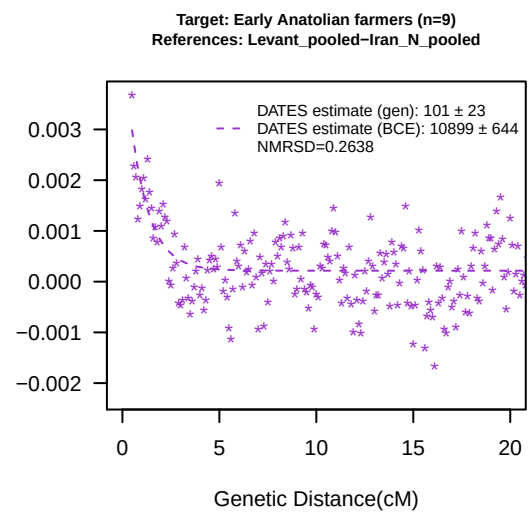

Target: Latvia_MN_Comb_Ware.SG (n=1)
References:EHG−WHG
DATES estimate (gen):188±94
DATES estimate (BCE):9357 ± 2632
NRMSD=0.2862
Ancestry covariance
Genetic Distance(cM)

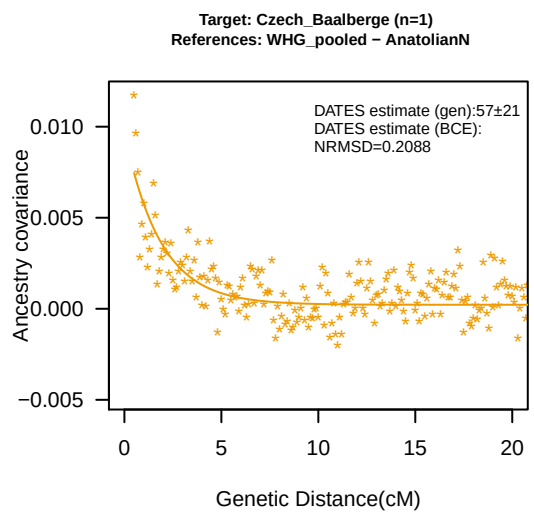

Target: Ukraine_N (n=15)
References:EHG−WHG
DATES estimate (gen):81±22
DATES estimate (BCE):7687 ± 616
NRMSD=0.1154
Ancestry covariance
Genetic Distance(cM)

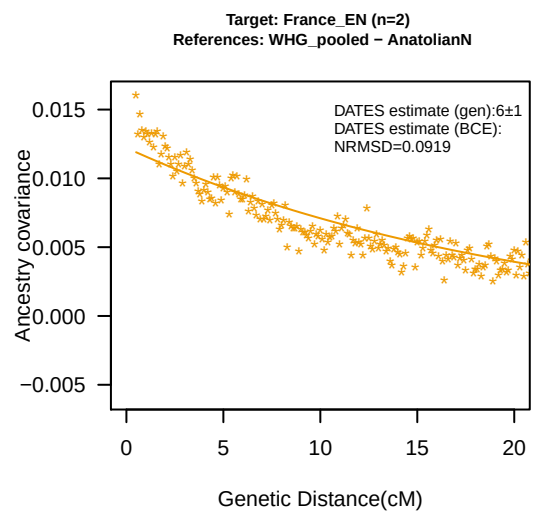

Target: Balkans_N (n=12)
References: WHG_pooled − AnatolianN
DATES estimate (gen):18±6
DATES estimate (BCE):
NRMSD=0.1875
Ancestry covariance
Genetic Distance(cM)

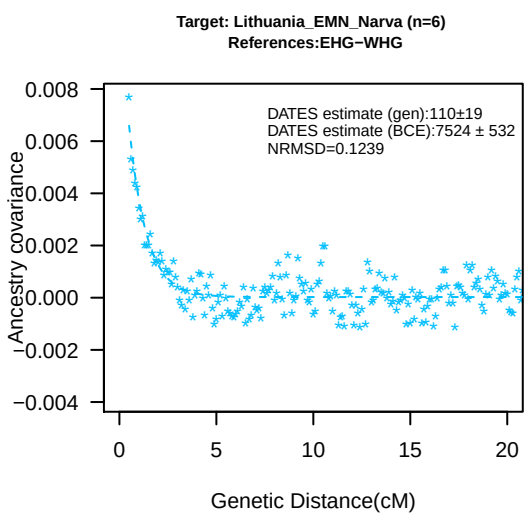

Target: France_EN_MN (n=4)
References: WHG_pooled − AnatolianN
DATES estimate (gen):23±2
DATES estimate (BCE):
NRMSD=0.0881
Ancestry covariance
Genetic Distance(cM)

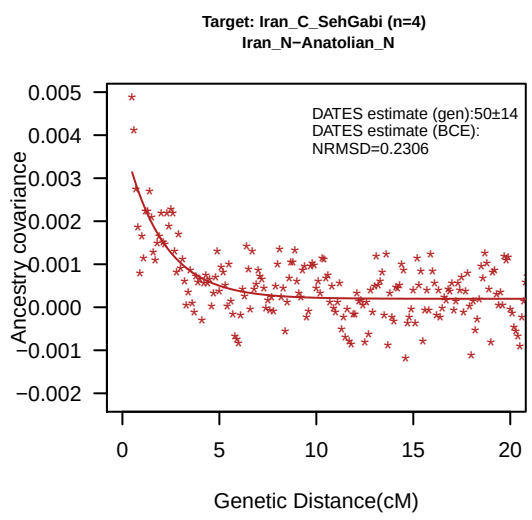

Target: Latvia_MN (n=2)
References:EHG−WHG
DATES estimate (gen):115±24
DATES estimate (BCE):7388 ± 672
NRMSD=0.1583
Ancestry covariance
Genetic Distance(cM)

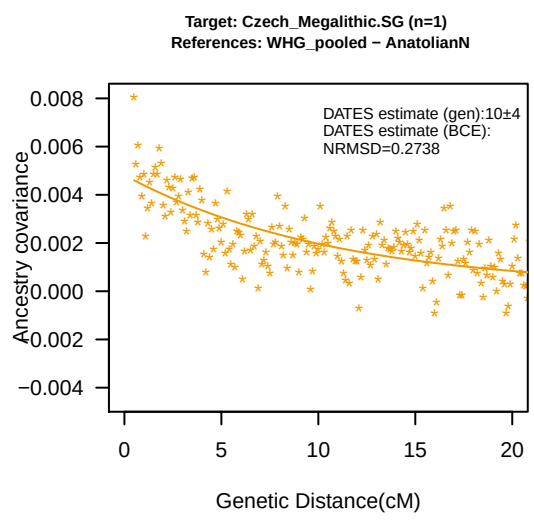

Target: Early Anatolian farmers (n=9)
References: Levant_pooled−Iran_N_pooled
DATES estimate (gen): 101 ± 23
DATES estimate (BCE): 10899 ± 644
NMRSD=0.2638
Ancestry covariance
Genetic Distance(cM)

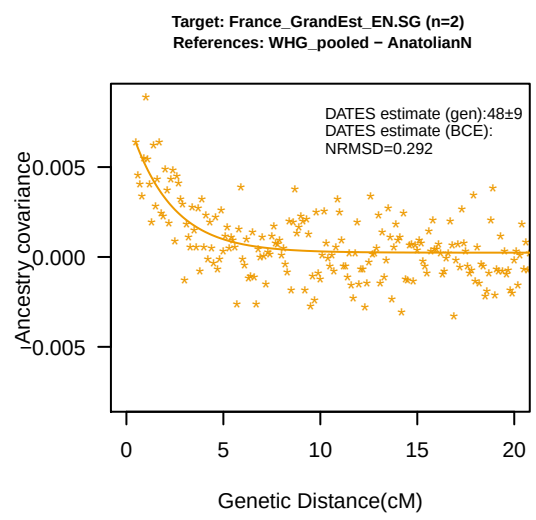

Target: Czech_Baalberge (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):57±21
DATES estimate (BCE):
NRMSD=0.2088
Ancestry covariance
Genetic Distance(cM)

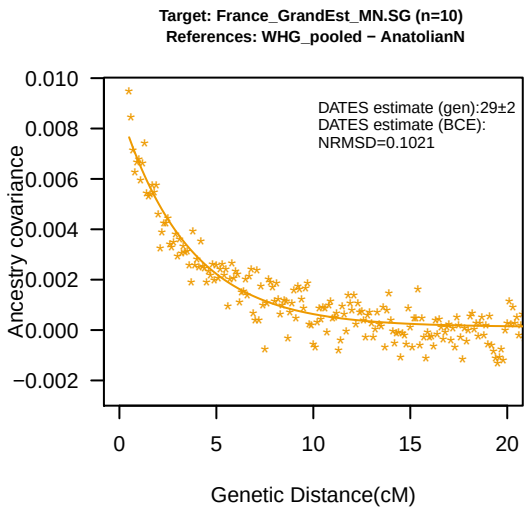

Target: France_EN (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):6±1
DATES estimate (BCE):
NRMSD=0.0919
Ancestry covariance
Genetic Distance(cM)

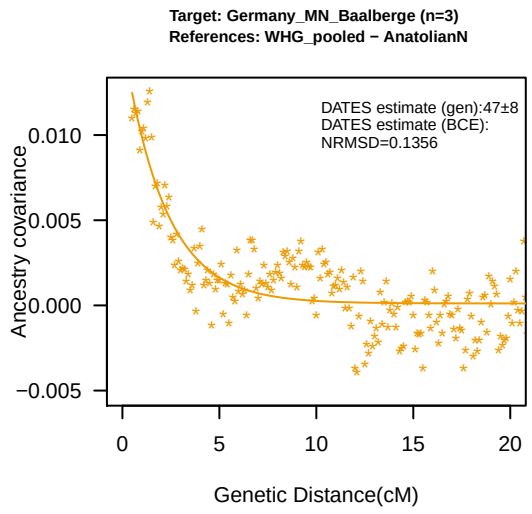

Target: Lithuania_EMN_Narva (n=6)
References:EHG−WHG
DATES estimate (gen):110±19
DATES estimate (BCE):7524 ± 532
NRMSD=0.1239
Ancestry covariance
Genetic Distance(cM)

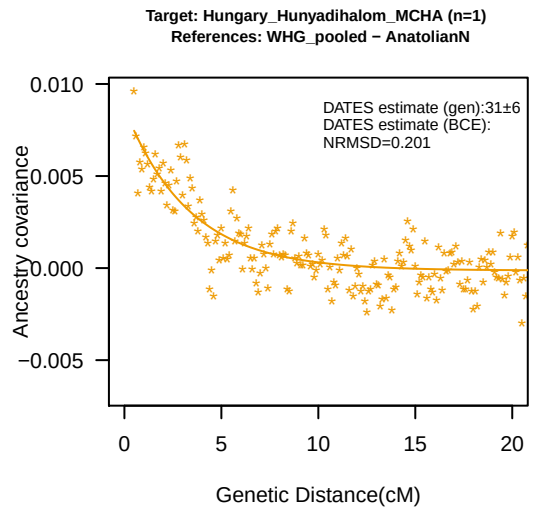

Target: Iran_C_SehGabi (n=4)
Iran_N−Anatolian_N
DATES estimate (gen):50±14
DATES estimate (BCE):
NRMSD=0.2306
Ancestry covariance
Genetic Distance(cM)

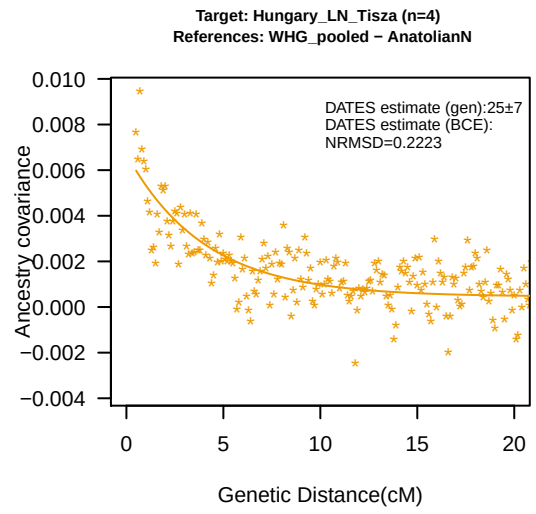

Target: Czech_Megalithic.SG (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):10±4
DATES estimate (BCE):
NRMSD=0.2738
Ancestry covariance
Genetic Distance(cM)

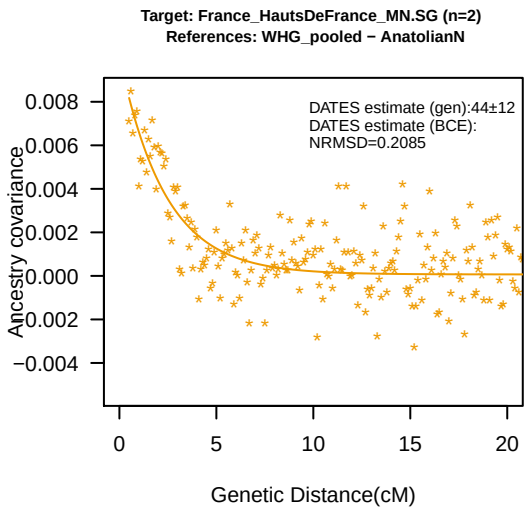

Target: France_GrandEst_EN.SG (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):48±9
DATES estimate (BCE):
NRMSD=0.292
Ancestry covariance
Genetic Distance(cM)

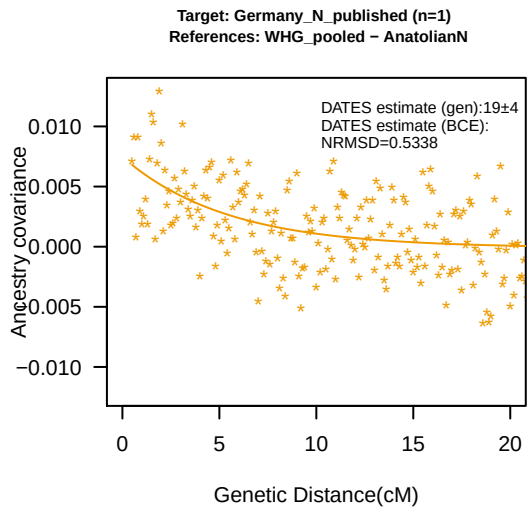

Target: France_GrandEst_MN.SG (n=10)
References: WHG_pooled − AnatolianN
DATES estimate (gen):29±2
DATES estimate (BCE):
NRMSD=0.1021
Ancestry covariance
Genetic Distance(cM)

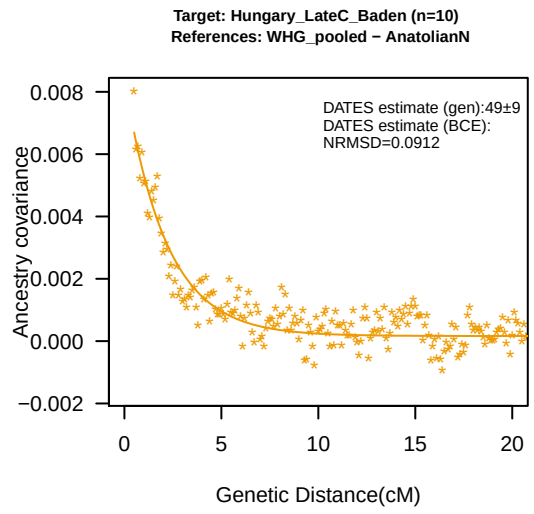

Target: Germany_MN_Baalberge (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):47±8
DATES estimate (BCE):
NRMSD=0.1356
Ancestry covariance
Genetic Distance(cM)

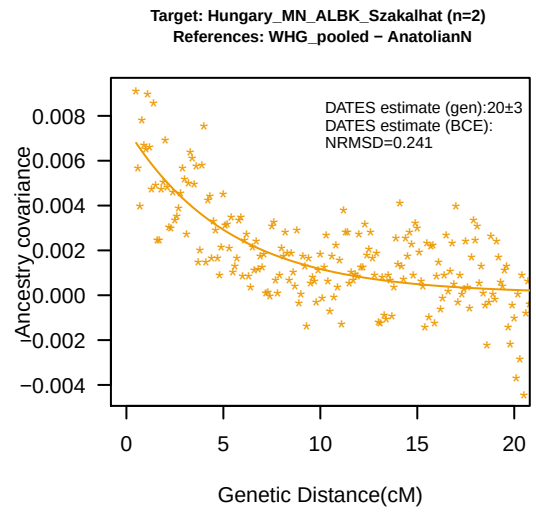

Target: Hungary_Hunyadihalom_MCHA (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):31±6
DATES estimate (BCE):
NRMSD=0.201
Ancestry covariance
Genetic Distance(cM)

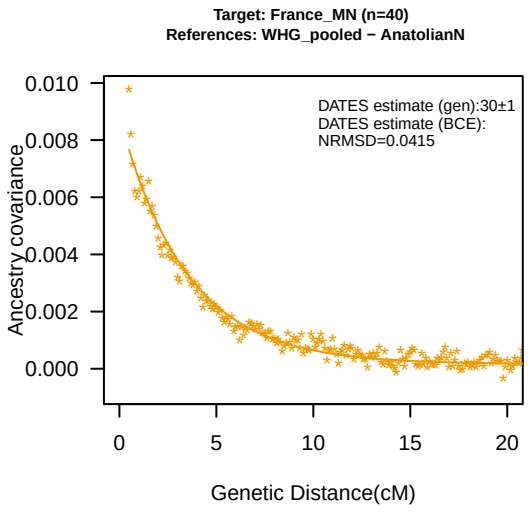

Target: Hungary_LN_Tisza (n=4)
References: WHG_pooled − AnatolianN
DATES estimate (gen):25±7
DATES estimate (BCE):
NRMSD=0.2223
Ancestry covariance
Genetic Distance(cM)

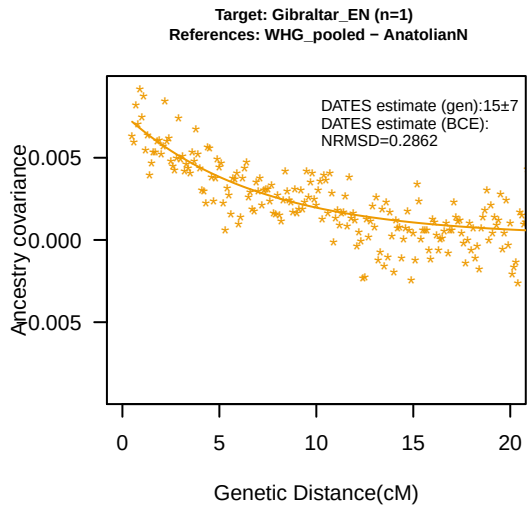

Target: France_HautsDeFrance_MN.SG (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):44±12
DATES estimate (BCE):
NRMSD=0.2085
Ancestry covariance
Genetic Distance(cM)

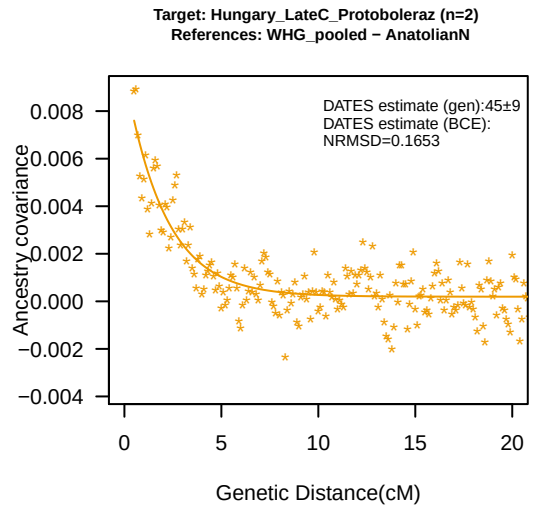

Target: Germany_N_published (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):19±4
DATES estimate (BCE):
NRMSD=0.5338
Ancestry covariance
Genetic Distance(cM)

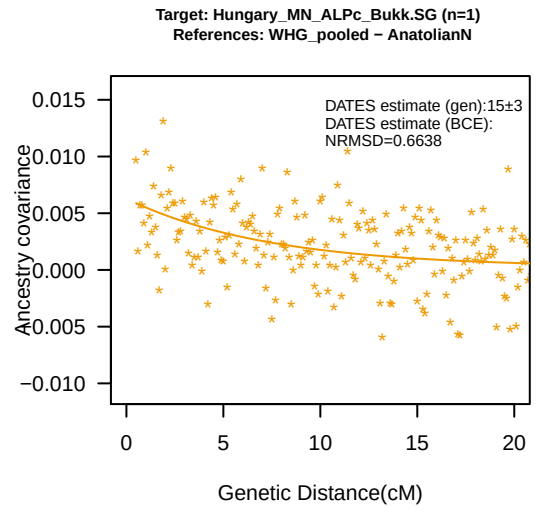

Target: Hungary_LateC_Baden (n=10)
References: WHG_pooled − AnatolianN
DATES estimate (gen):49±9
DATES estimate (BCE):
NRMSD=0.0912
Ancestry covariance
Genetic Distance(cM)

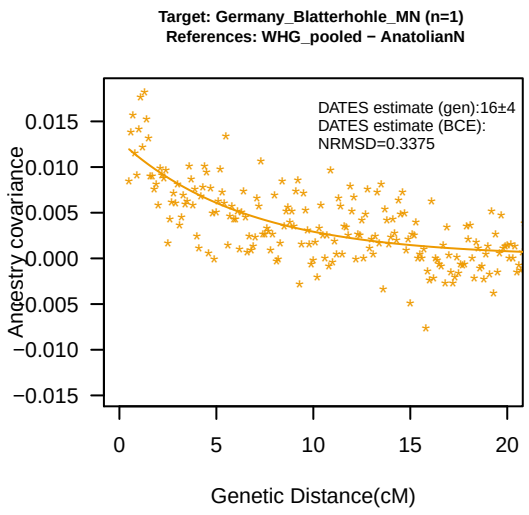

Target: Hungary_MN_ALBK_Szakalhat (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):20±3
DATES estimate (BCE):
NRMSD=0.241
Ancestry covariance
Genetic Distance(cM)

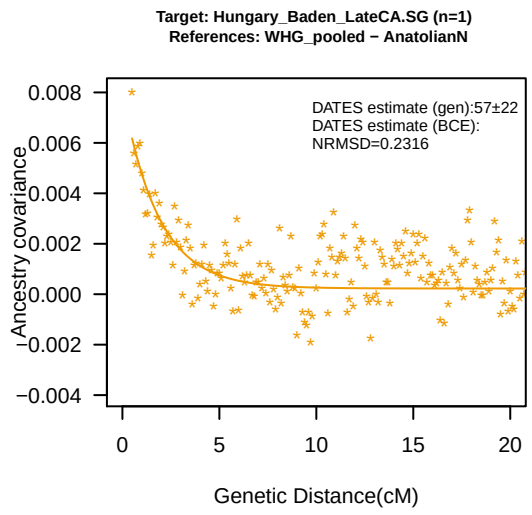

Target: France_MN (n=40)
References: WHG_pooled − AnatolianN
DATES estimate (gen):30±1
DATES estimate (BCE):
NRMSD=0.0415
Ancestry covariance
Genetic Distance(cM)

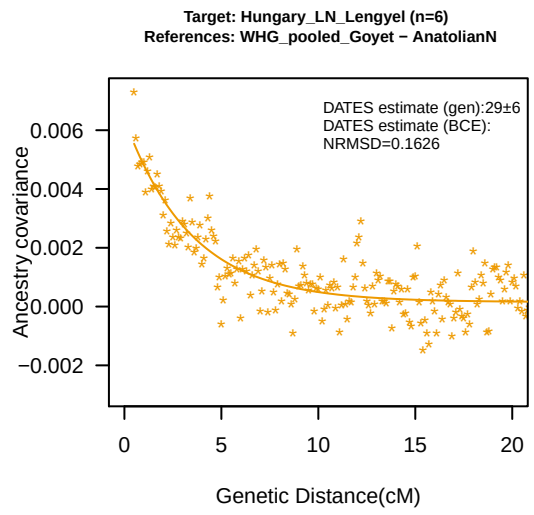

Target: Gibraltar_EN (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):15±7
DATES estimate (BCE):
NRMSD=0.2862
Ancestry covariance
Genetic Distance(cM)

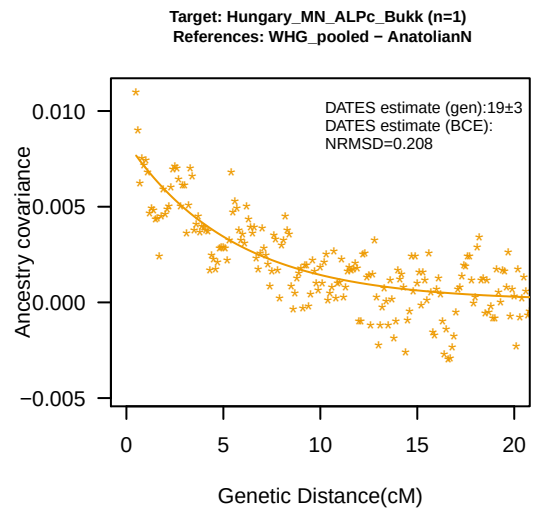

Target: Hungary_LateC_Protoboleraz (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):45±9
DATES estimate (BCE):
NRMSD=0.1653
Ancestry covariance
Genetic Distance(cM)

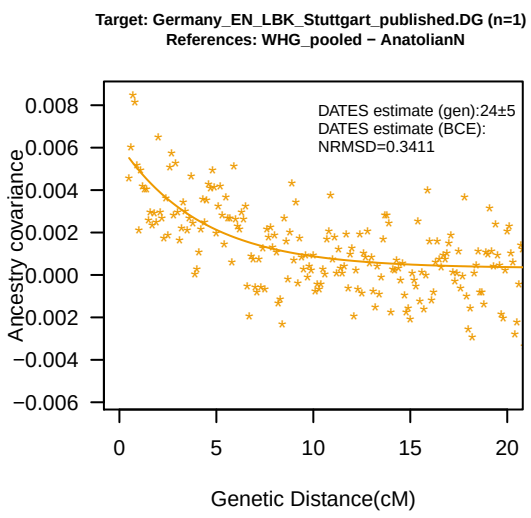

Target: Hungary_MN_ALPc_Bukk.SG (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):15±3
DATES estimate (BCE):
NRMSD=0.6638
Ancestry covariance
Genetic Distance(cM)

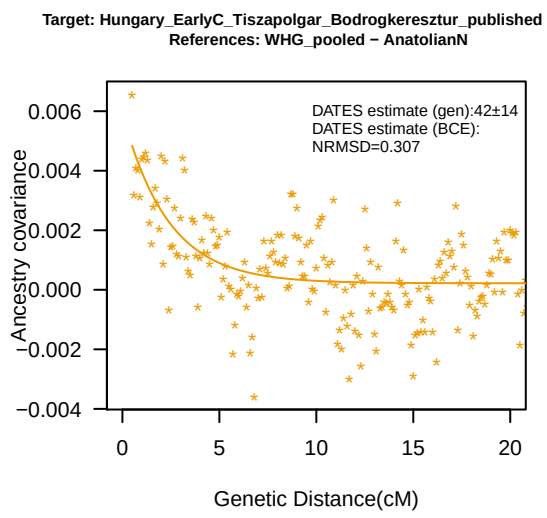

Target: Germany_Blatterhohle_MN (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):16±4
DATES estimate (BCE):
NRMSD=0.3375
Ancestry covariance
Genetic Distance(cM)

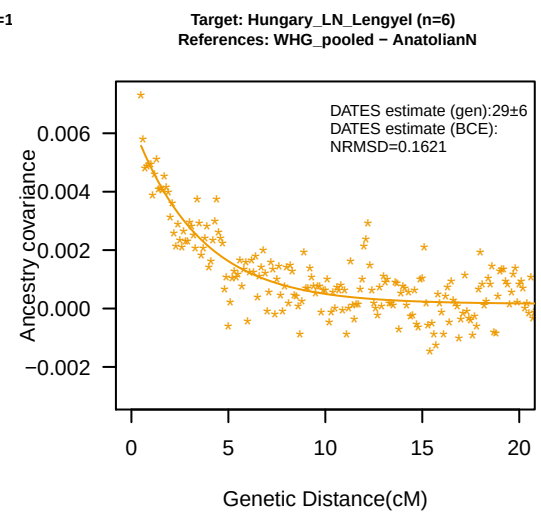

Target: Hungary_Baden_LateCA.SG (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):57±22
DATES estimate (BCE):
NRMSD=0.2316
Ancestry covariance
Genetic Distance(cM)

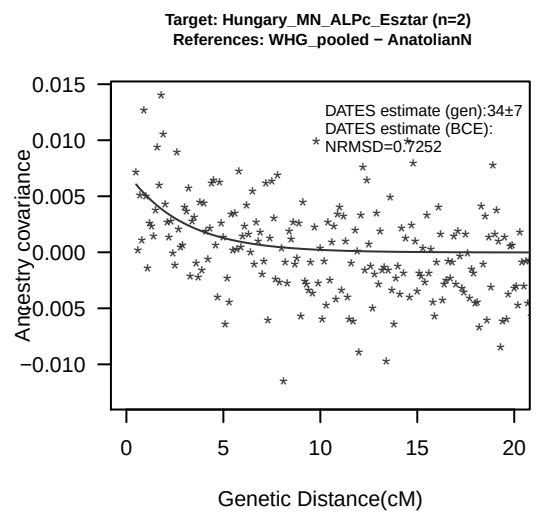

Target: Hungary_LN_Lengyel (n=6)
References: WHG_pooled_Goyet − AnatolianN
DATES estimate (gen):29±6
DATES estimate (BCE):
NRMSD=0.1626
Ancestry covariance
Genetic Distance(cM)

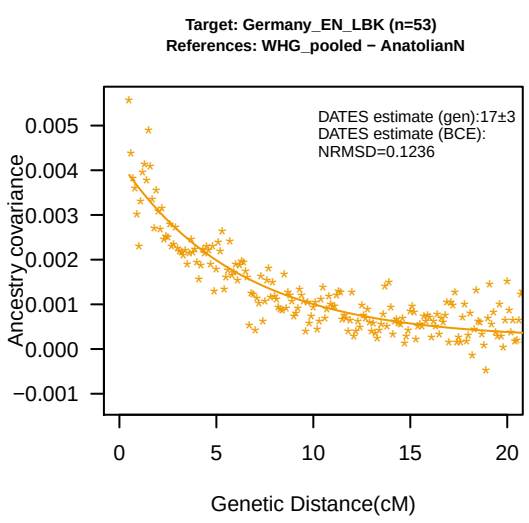

Target: Hungary_MN_ALPc_Bukk (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):19±3
DATES estimate (BCE):
NRMSD=0.208
Ancestry covariance
Genetic Distance(cM)

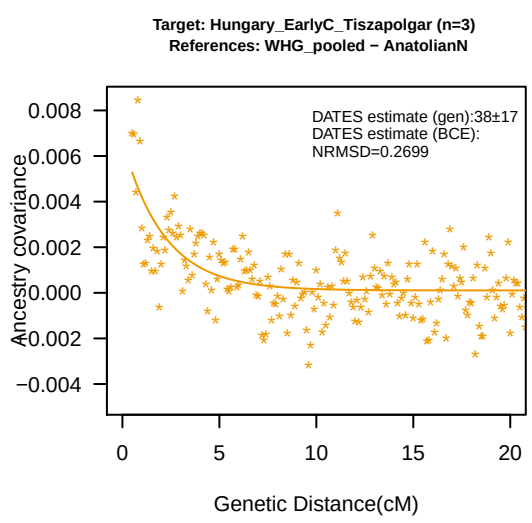

Target: Germany_EN_LBK_Stuttgart_published.DG (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):24±5
DATES estimate (BCE):
NRMSD=0.3411
Ancestry covariance
Genetic Distance(cM)

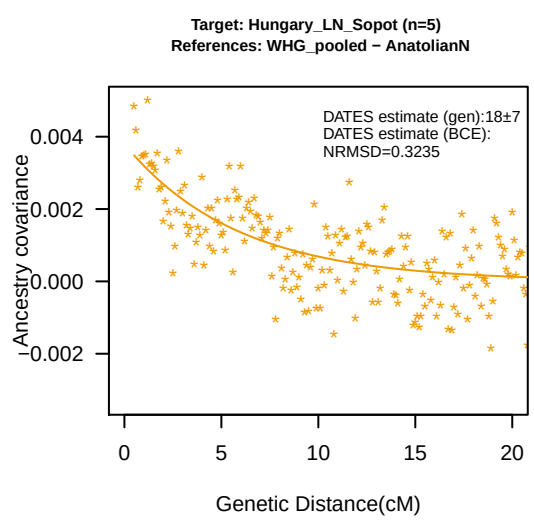

Target: Hungary_EarlyC_Tiszapolgar_Bodrogkeresztur_published (n=1
References: WHG_pooled − AnatolianN
DATES estimate (gen):42±14
DATES estimate (BCE):
NRMSD=0.307
Ancestry covariance
Genetic Distance(cM)

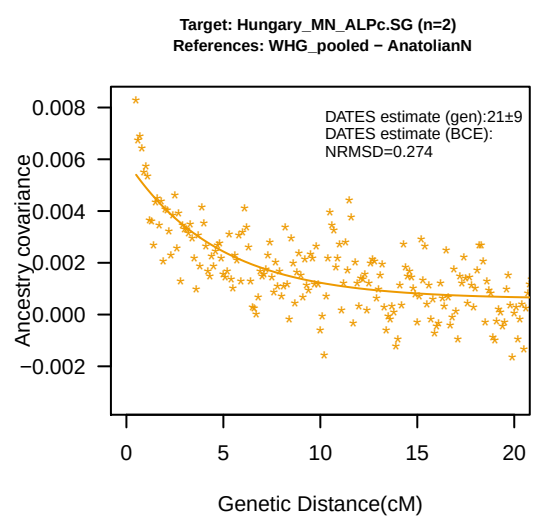

Target: Hungary_LN_Lengyel (n=6)
References: WHG_pooled − AnatolianN
DATES estimate (gen):29±6
DATES estimate (BCE):
NRMSD=0.1621
Ancestry covariance
Genetic Distance(cM)

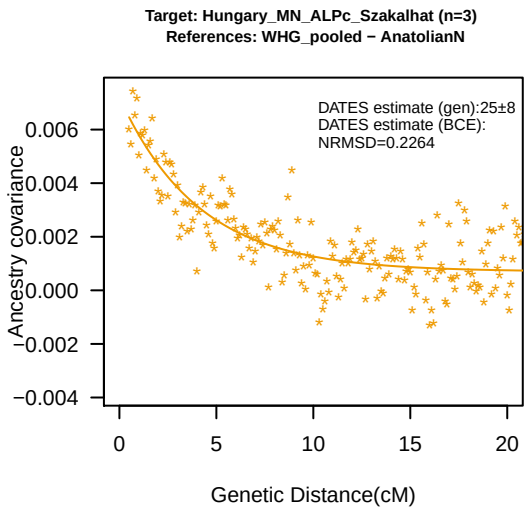

Target: Hungary_MN_ALPc_Esztar (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):34±7
DATES estimate (BCE):
NRMSD=0.7252
Ancestry covariance
Genetic Distance(cM)

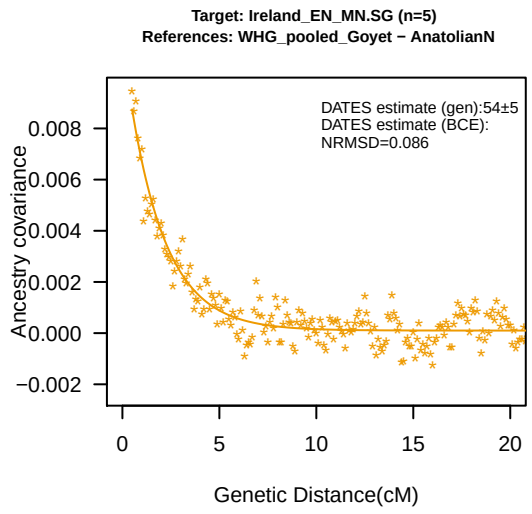

Target: Germany_EN_LBK (n=53)
References: WHG_pooled − AnatolianN
DATES estimate (gen):17±3
DATES estimate (BCE):
NRMSD=0.1236
Ancestry covariance
Genetic Distance(cM)

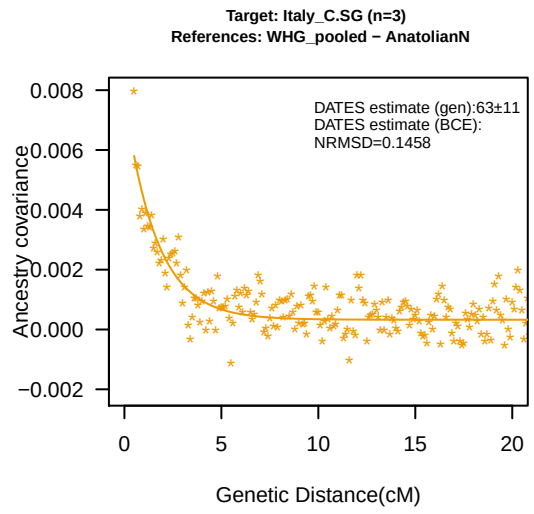

Target: Hungary_EarlyC_Tiszapolgar (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):38±17
DATES estimate (BCE):
NRMSD=0.2699
Ancestry covariance
Genetic Distance(cM)

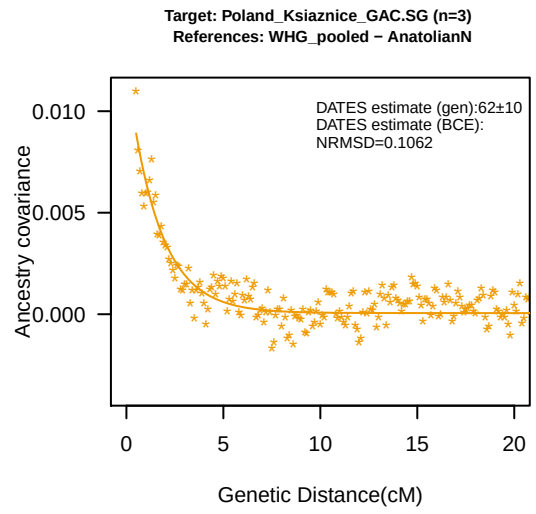

Target: Hungary_LN_Sopot (n=5)
References: WHG_pooled − AnatolianN
DATES estimate (gen):18±7
DATES estimate (BCE):
NRMSD=0.3235
Ancestry covariance
Genetic Distance(cM)

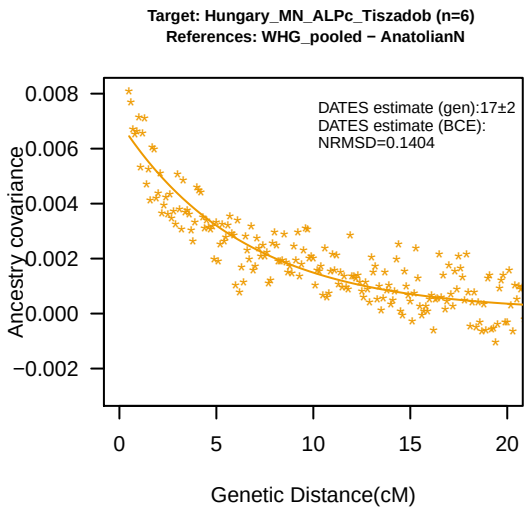

Target: Hungary_MN_ALPc.SG (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):21±9
DATES estimate (BCE):
NRMSD=0.274
Ancestry covariance
Genetic Distance(cM)

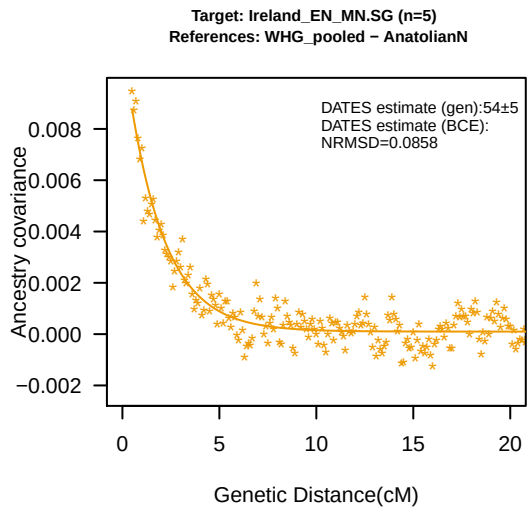

Target: Hungary_MN_ALPc_Szakalhat (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):25±8
DATES estimate (BCE):
NRMSD=0.2264
Ancestry covariance
Genetic Distance(cM)

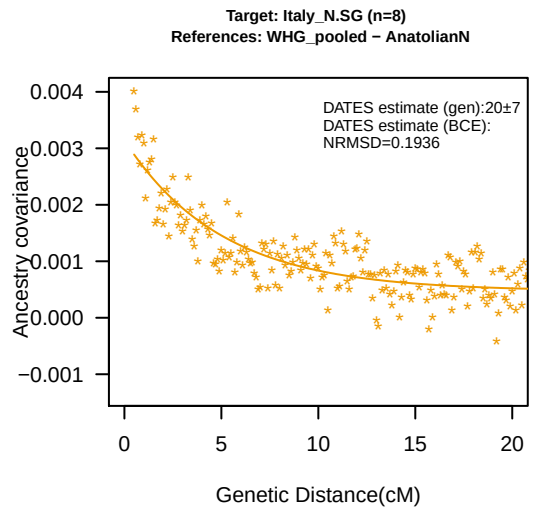

Target: Ireland_EN_MN.SG (n=5)
References: WHG_pooled_Goyet − AnatolianN
DATES estimate (gen):54±5
DATES estimate (BCE):
NRMSD=0.086
Ancestry covariance
Genetic Distance(cM)

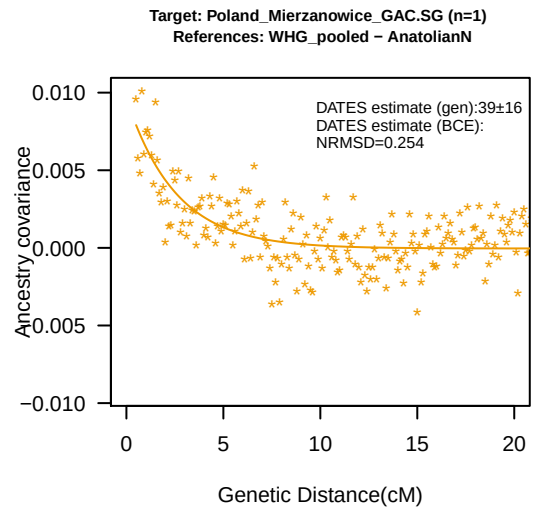

Target: Italy_C.SG (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):63±11
DATES estimate (BCE):
NRMSD=0.1458
Ancestry covariance
Genetic Distance(cM)

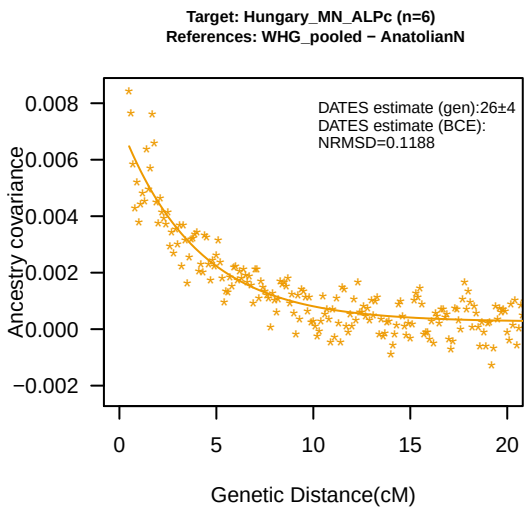

Target: Poland_Ksiaznice_GAC.SG (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):62±10
DATES estimate (BCE):
NRMSD=0.1062
Ancestry covariance
Genetic Distance(cM)

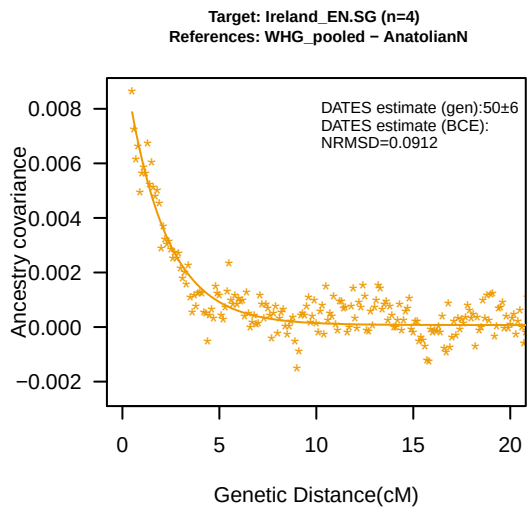

Target: Hungary_MN_ALPc_Tiszadob (n=6)
References: WHG_pooled − AnatolianN
DATES estimate (gen):17±2
DATES estimate (BCE):
NRMSD=0.1404
Ancestry covariance
Genetic Distance(cM)

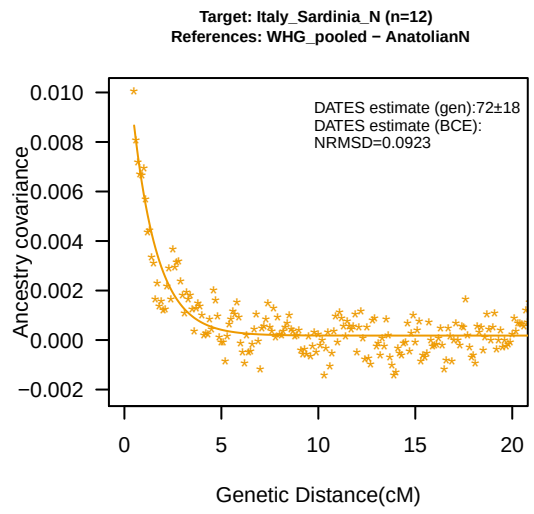

Target: Ireland_EN_MN.SG (n=5)
References: WHG_pooled − AnatolianN
DATES estimate (gen):54±5
DATES estimate (BCE):
NRMSD=0.0858
Ancestry covariance
Genetic Distance(cM)

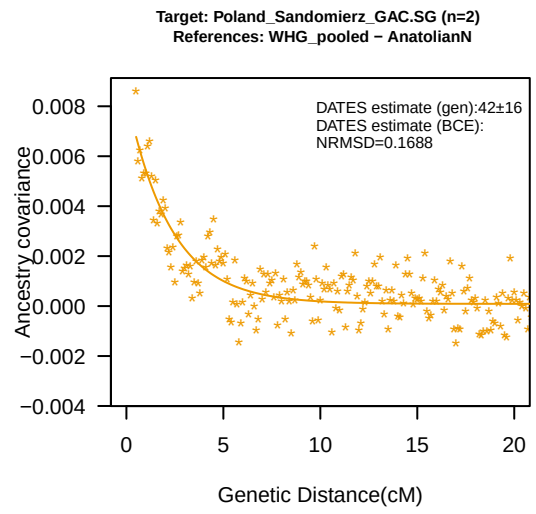

Target: Italy_N.SG (n=8)
References: WHG_pooled − AnatolianN
DATES estimate (gen):20±7
DATES estimate (BCE):
NRMSD=0.1936
Ancestry covariance
Genetic Distance(cM)

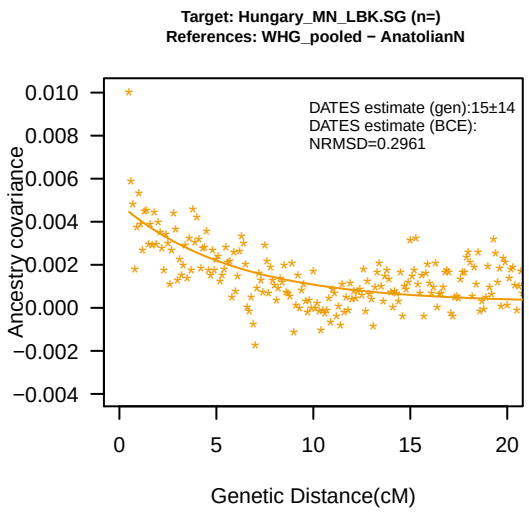

Target: Poland_Mierzanowice_GAC.SG (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):39±16
DATES estimate (BCE):
NRMSD=0.254
Ancestry covariance
Genetic Distance(cM)

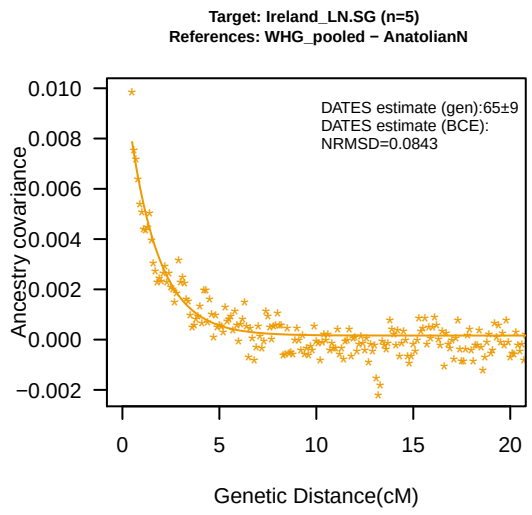

Target: Hungary_MN_ALPc (n=6)
References: WHG_pooled − AnatolianN
DATES estimate (gen):26±4
DATES estimate (BCE):
NRMSD=0.1188
Ancestry covariance
Genetic Distance(cM)

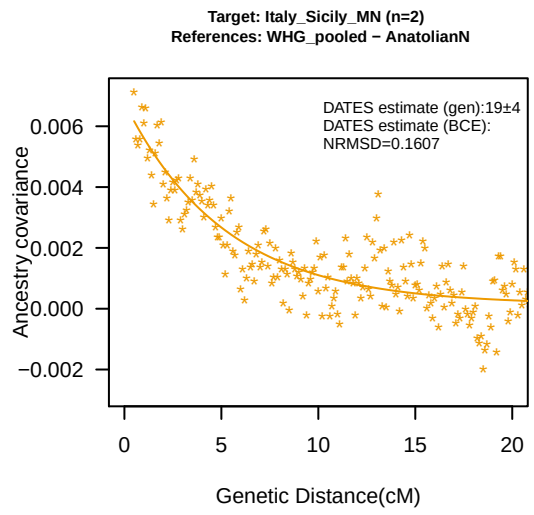

Target: Ireland_EN.SG (n=4)
References: WHG_pooled − AnatolianN
DATES estimate (gen):50±6
DATES estimate (BCE):
NRMSD=0.0912
Ancestry covariance
Genetic Distance(cM)

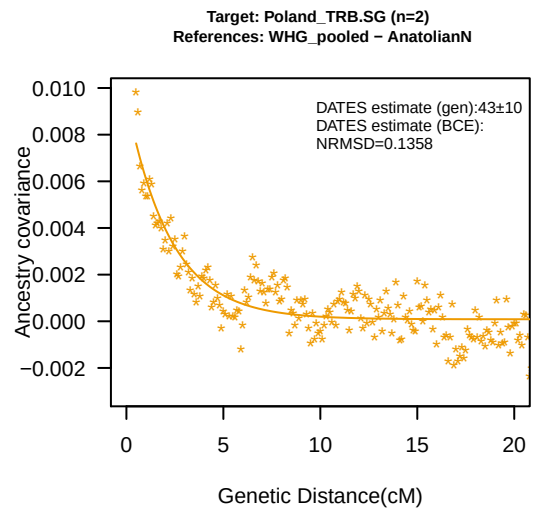

Target: Italy_Sardinia_N (n=12)
References: WHG_pooled − AnatolianN
DATES estimate (gen):72±18
DATES estimate (BCE):
NRMSD=0.0923
Ancestry covariance
Genetic Distance(cM)

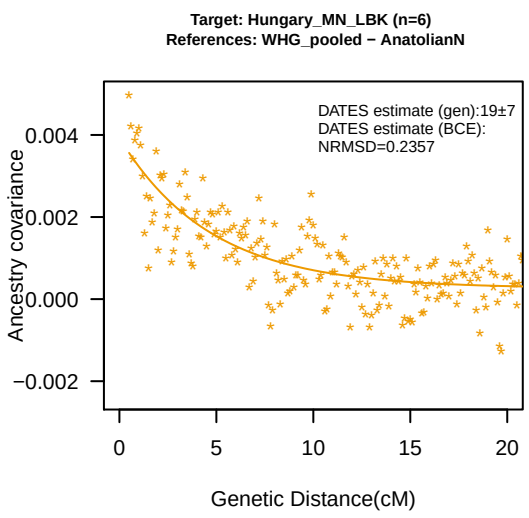

Target: Poland_Sandomierz_GAC.SG (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):42±16
DATES estimate (BCE):
NRMSD=0.1688
Ancestry covariance
Genetic Distance(cM)

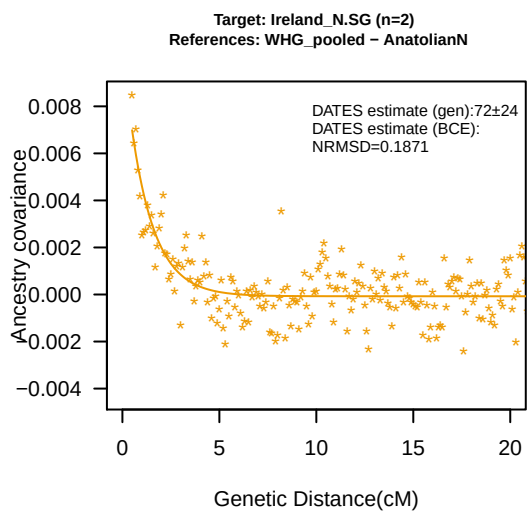

Target: Hungary_MN_LBK.SG (n=)
References: WHG_pooled − AnatolianN
DATES estimate (gen):15±14
DATES estimate (BCE):
NRMSD=0.2961
Ancestry covariance
Genetic Distance(cM)

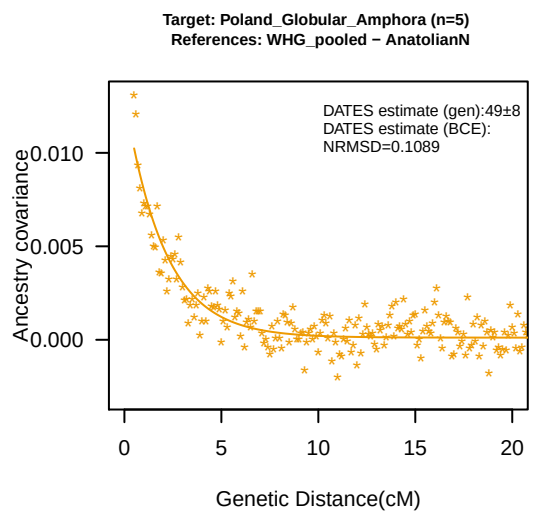

Target: Ireland_LN.SG (n=5)
References: WHG_pooled − AnatolianN
DATES estimate (gen):65±9
DATES estimate (BCE):
NRMSD=0.0843
Ancestry covariance
Genetic Distance(cM)

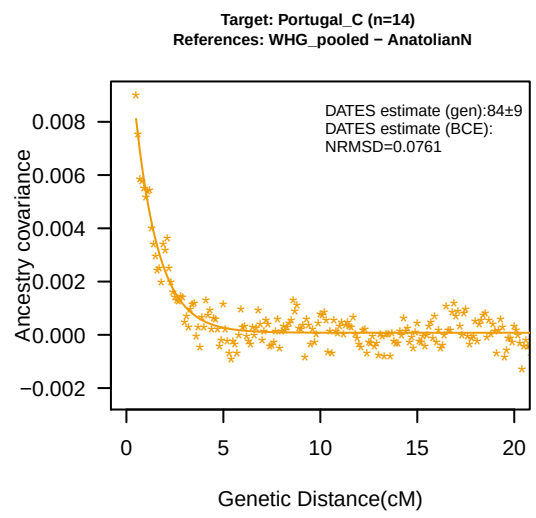

Target: Italy_Sicily_MN (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):19±4
DATES estimate (BCE):
NRMSD=0.1607
Ancestry covariance
Genetic Distance(cM)

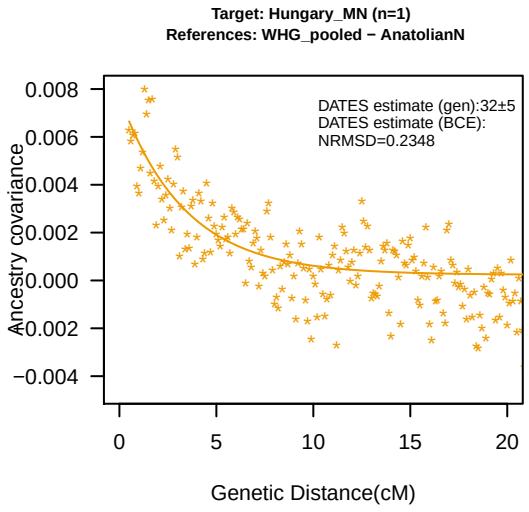

Target: Poland_TRB.SG (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):43±10
DATES estimate (BCE):
NRMSD=0.1358
Ancestry covariance
Genetic Distance(cM)

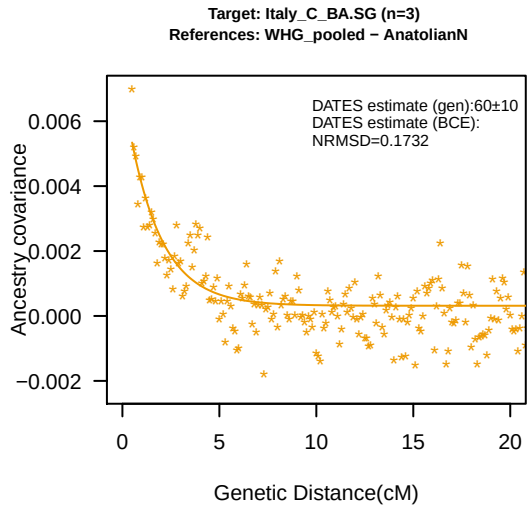

Target: Hungary_MN_LBK (n=6)
References: WHG_pooled − AnatolianN
DATES estimate (gen):19±7
DATES estimate (BCE):
NRMSD=0.2357
Ancestry covariance
Genetic Distance(cM)

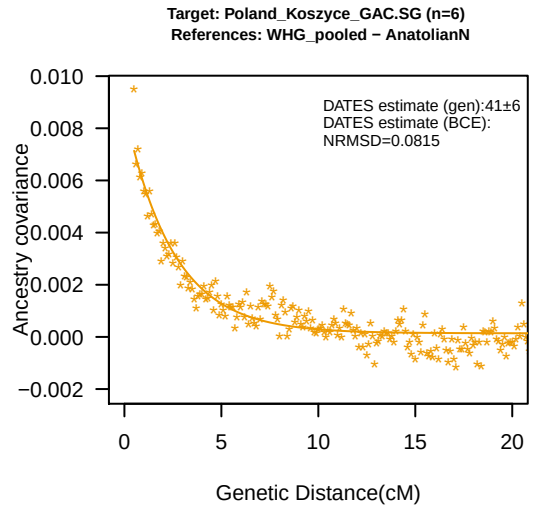

Target: Ireland_N.SG (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):72±24
DATES estimate (BCE):
NRMSD=0.1871
Ancestry covariance
Genetic Distance(cM)

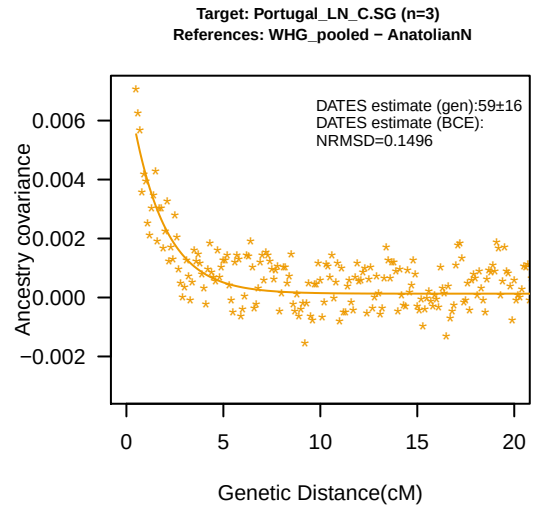

Target: Poland_Globular_Amphora (n=5)
References: WHG_pooled − AnatolianN
DATES estimate (gen):49±8
DATES estimate (BCE):
NRMSD=0.1089
Ancestry covariance
Genetic Distance(cM)

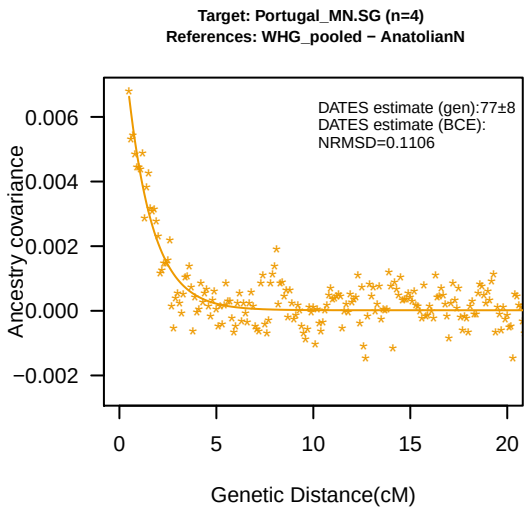

Target: Portugal_C (n=14)
References: WHG_pooled − AnatolianN
DATES estimate (gen):84±9
DATES estimate (BCE):
NRMSD=0.0761
Ancestry covariance
Genetic Distance(cM)

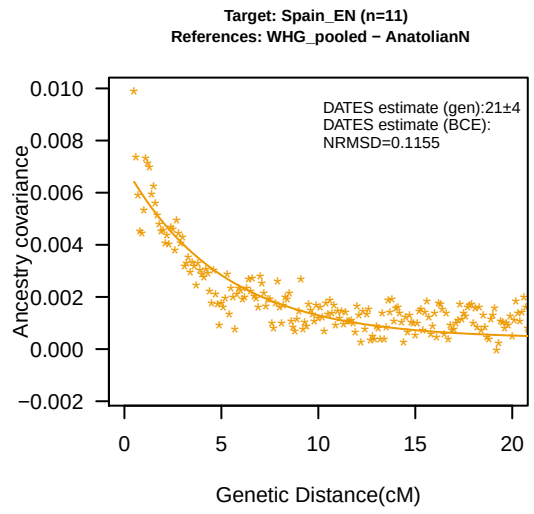

Target: Hungary_MN (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):32±5
DATES estimate (BCE):
NRMSD=0.2348
Ancestry covariance
Genetic Distance(cM)

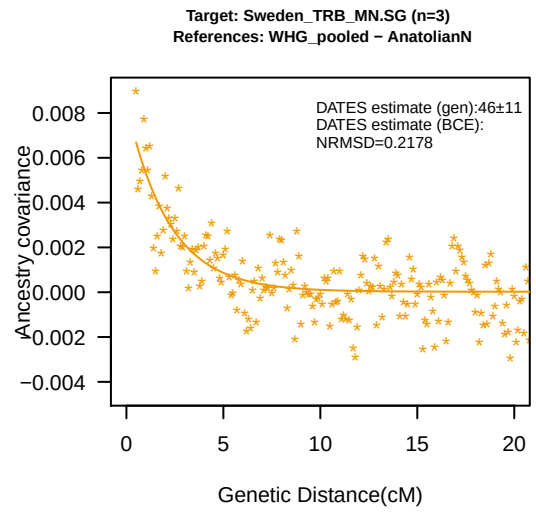

Target: Italy_C_BA.SG (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):60±10
DATES estimate (BCE):
NRMSD=0.1732
Ancestry covariance
Genetic Distance(cM)

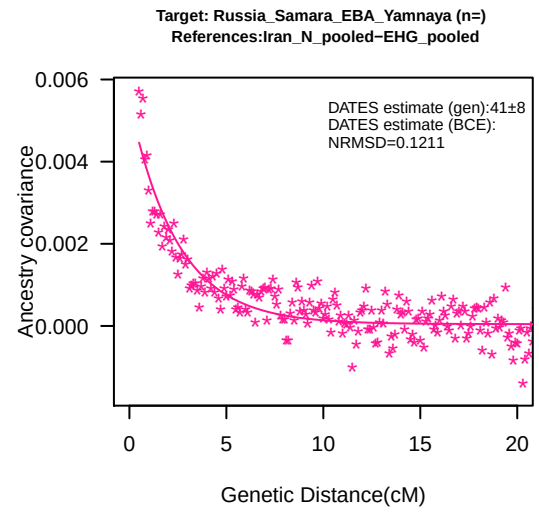

Target: Poland_Koszyce_GAC.SG (n=6)
References: WHG_pooled − AnatolianN
DATES estimate (gen):41±6
DATES estimate (BCE):
NRMSD=0.0815
Ancestry covariance
Genetic Distance(cM)

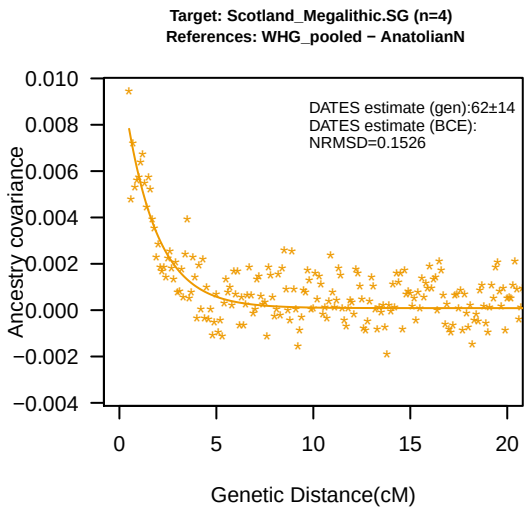

Target: Portugal_LN_C.SG (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):59±16
DATES estimate (BCE):
NRMSD=0.1496
Ancestry covariance
Genetic Distance(cM)

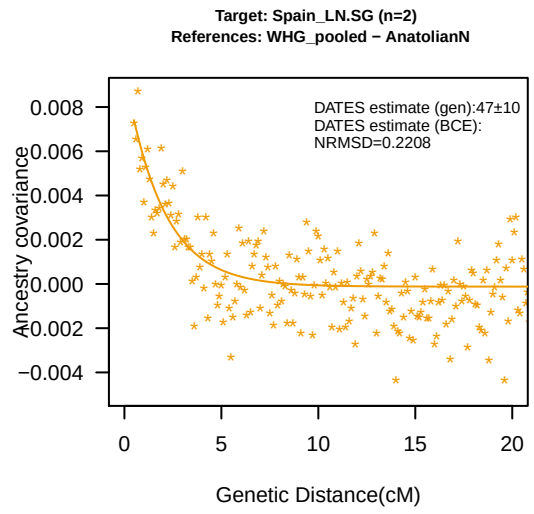

Target: Portugal_MN.SG (n=4)
References: WHG_pooled − AnatolianN
DATES estimate (gen):77±8
DATES estimate (BCE):
NRMSD=0.1106
Ancestry covariance
Genetic Distance(cM)

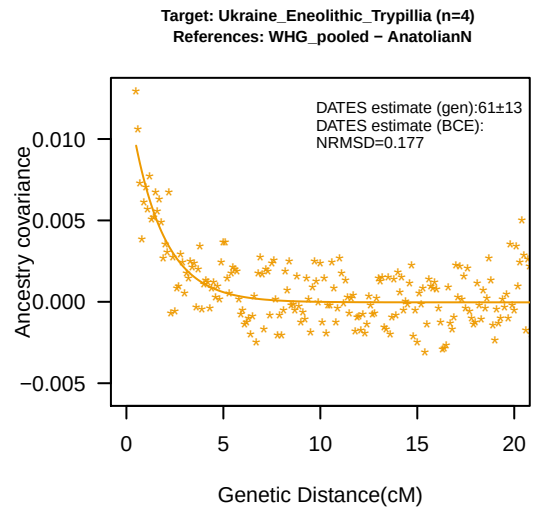

Target: Spain_EN (n=11)
References: WHG_pooled − AnatolianN
DATES estimate (gen):21±4
DATES estimate (BCE):
NRMSD=0.1155
Ancestry covariance
Genetic Distance(cM)

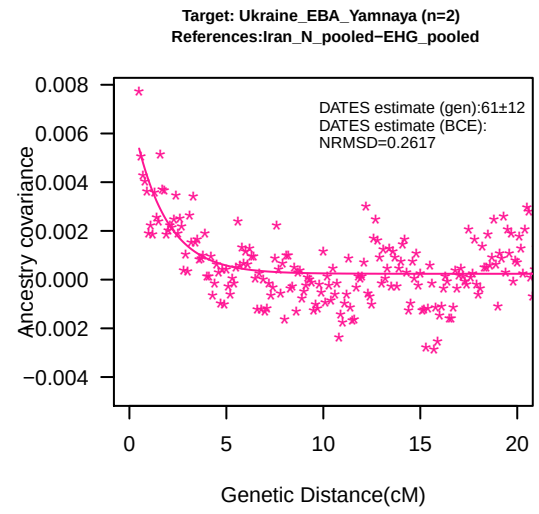

Target: Sweden_TRB_MN.SG (n=3)
References: WHG_pooled − AnatolianN
DATES estimate (gen):46±11
DATES estimate (BCE):
NRMSD=0.2178
Ancestry covariance
Genetic Distance(cM)

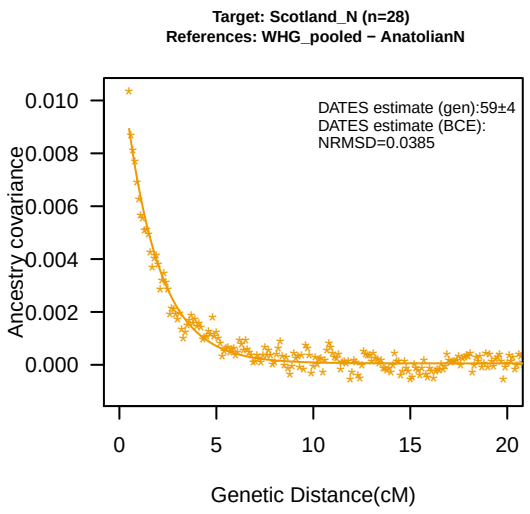

Target: Russia_Samara_EBA_Yamnaya (n=)
References:Iran_N_pooled−EHG_pooled
DATES estimate (gen):41±8
DATES estimate (BCE):
NRMSD=0.1211
Ancestry covariance
Genetic Distance(cM)

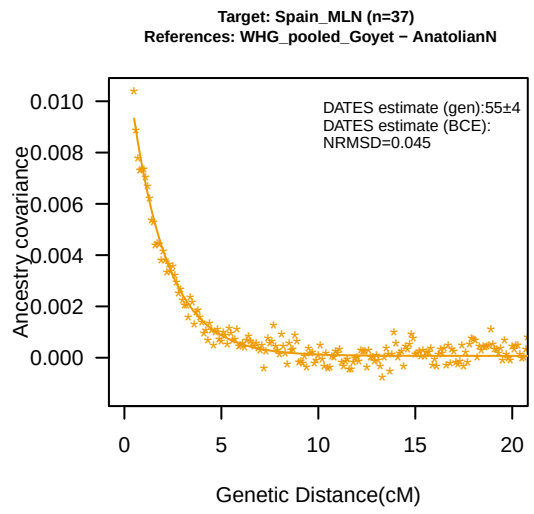

Target: Scotland_Megalithic.SG (n=4)
References: WHG_pooled − AnatolianN
DATES estimate (gen):62±14
DATES estimate (BCE):
NRMSD=0.1526
Ancestry covariance
Genetic Distance(cM)

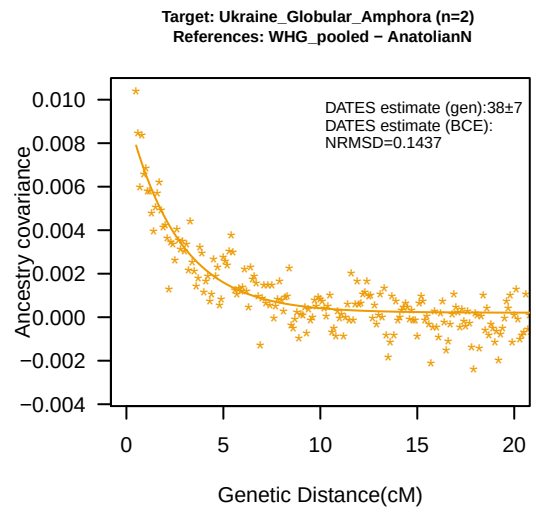

Target: Spain_LN.SG (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):47±10
DATES estimate (BCE):
NRMSD=0.2208
Ancestry covariance
Genetic Distance(cM)

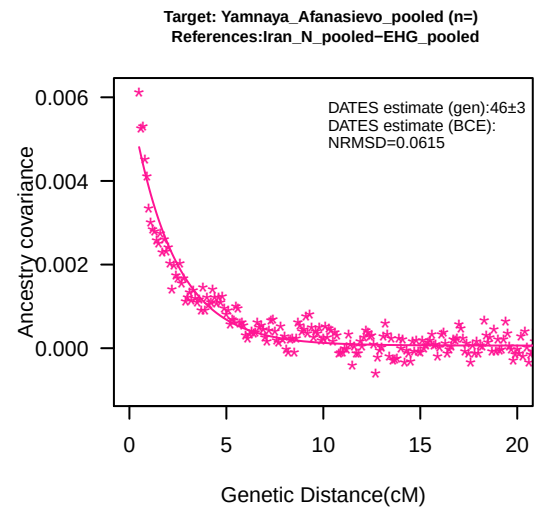

Target: Ukraine_Eneolithic_Trypillia (n=4)
References: WHG_pooled − AnatolianN
DATES estimate (gen):61±13
DATES estimate (BCE):
NRMSD=0.177
Ancestry covariance
Genetic Distance(cM)

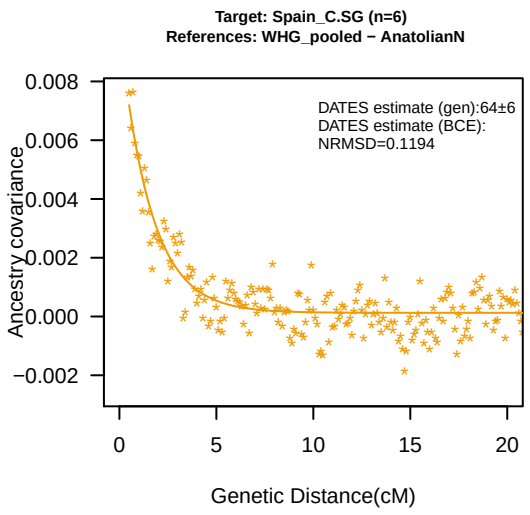

Target: Ukraine_EBA_Yamnaya (n=2)
References:Iran_N_pooled−EHG_pooled
DATES estimate (gen):61±12
DATES estimate (BCE):
NRMSD=0.2617
Ancestry covariance
Genetic Distance(cM)

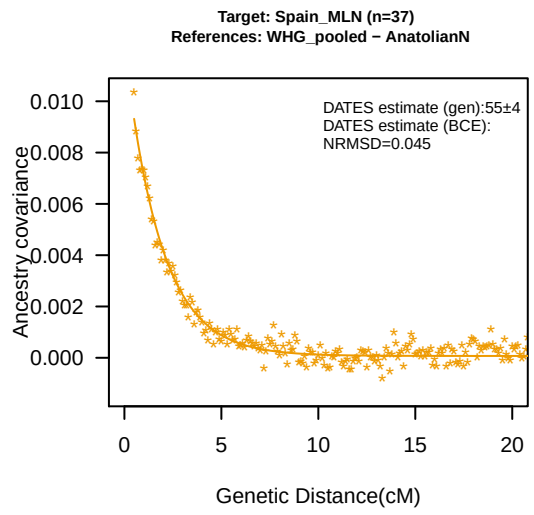

Target: Scotland_N (n=28)
References: WHG_pooled − AnatolianN
DATES estimate (gen):59±4
DATES estimate (BCE):
NRMSD=0.0385
Ancestry covariance
Genetic Distance(cM)

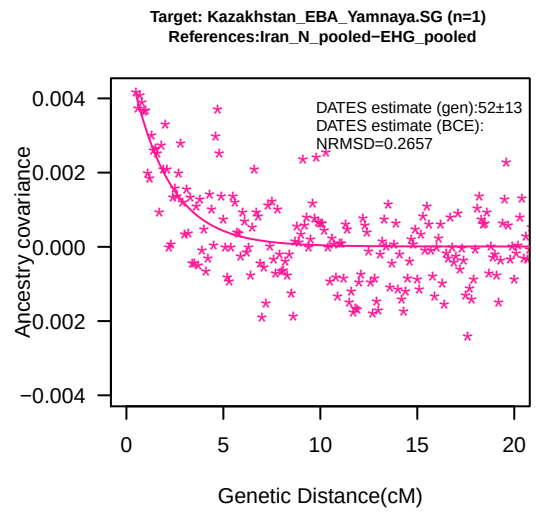

Target: Spain_MLN (n=37)
References: WHG_pooled_Goyet − AnatolianN
DATES estimate (gen):55±4
DATES estimate (BCE):
NRMSD=0.045
Ancestry covariance
Genetic Distance(cM)

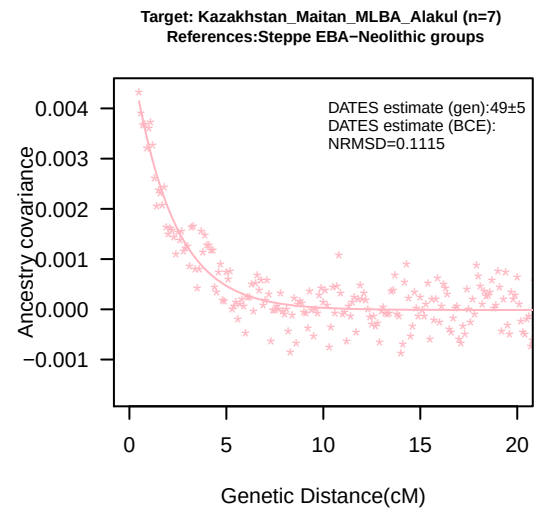

Target: Ukraine_Globular_Amphora (n=2)
References: WHG_pooled − AnatolianN
DATES estimate (gen):38±7
DATES estimate (BCE):
NRMSD=0.1437
Ancestry covariance
Genetic Distance(cM)

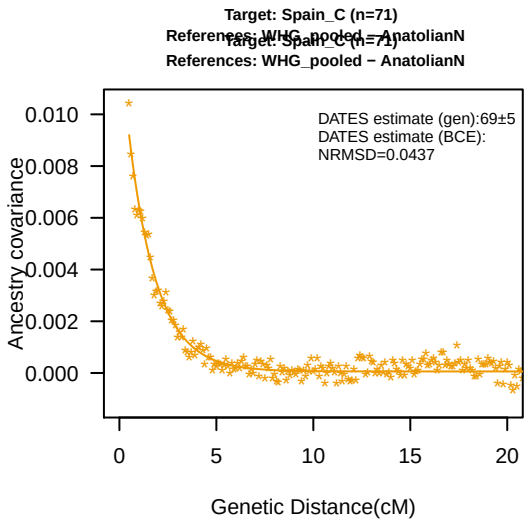

Target: Yamnaya_Afanasievo_pooled (n=)
References:Iran_N_pooled−EHG_pooled
DATES estimate (gen):46±3
DATES estimate (BCE):
NRMSD=0.0615
Ancestry covariance
Genetic Distance(cM)

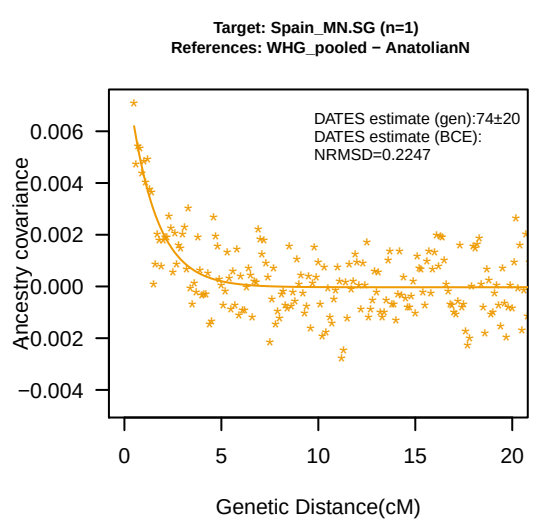

Target: Spain_C.SG (n=6)
References: WHG_pooled − AnatolianN
DATES estimate (gen):64±6
DATES estimate (BCE):
NRMSD=0.1194
Ancestry covariance
Genetic Distance(cM)

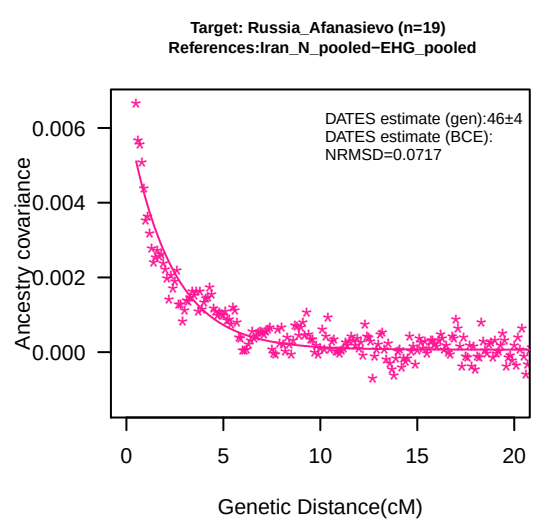

Target: Spain_MLN (n=37)
References: WHG_pooled − AnatolianN
DATES estimate (gen):55±4
DATES estimate (BCE):
NRMSD=0.045
Ancestry covariance
Genetic Distance(cM)

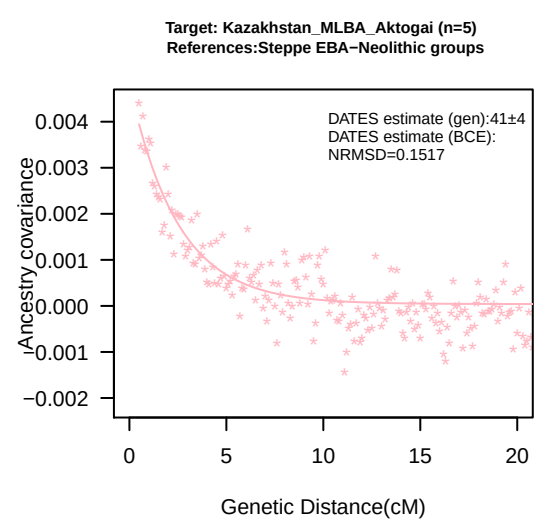

Target: Kazakhstan_EBA_Yamnaya.SG (n=1)
References:Iran_N_pooled−EHG_pooled
DATES estimate (gen):52±13
DATES estimate (BCE):
NRMSD=0.2657
Ancestry covariance
Genetic Distance(cM)

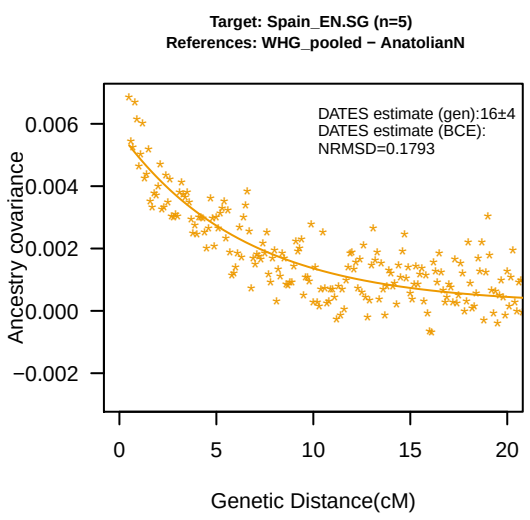

Target: Kazakhstan_Maitan_MLBA_Alakul (n=7)
References:Steppe EBA−Neolithic groups
DATES estimate (gen):49±5
DATES estimate (BCE):
NRMSD=0.1115
Ancestry covariance
Genetic Distance(cM)

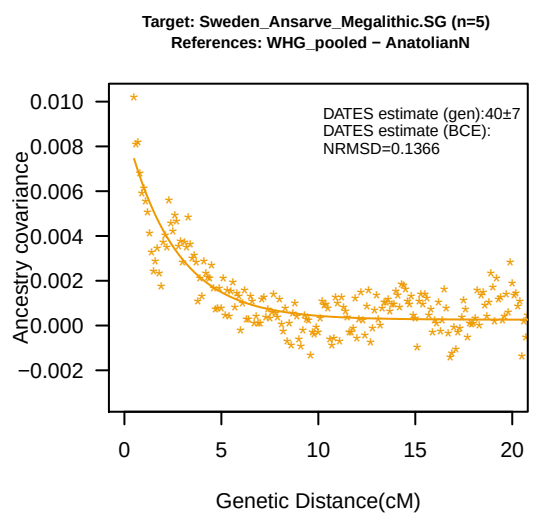

Target: Spain_C (n=71)
References: WHG_pooled − AnatolianN
DATES estimate (gen):69±5
DATES estimate (BCE):
NRMSD=0.0437
Ancestry covariance
Genetic Distance(cM)

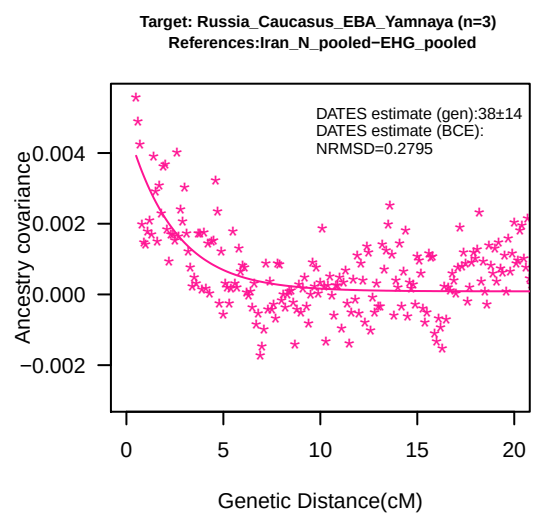

Target: Spain_MN.SG (n=1)
References: WHG_pooled − AnatolianN
DATES estimate (gen):74±20
DATES estimate (BCE):
NRMSD=0.2247
Ancestry covariance
Genetic Distance(cM)

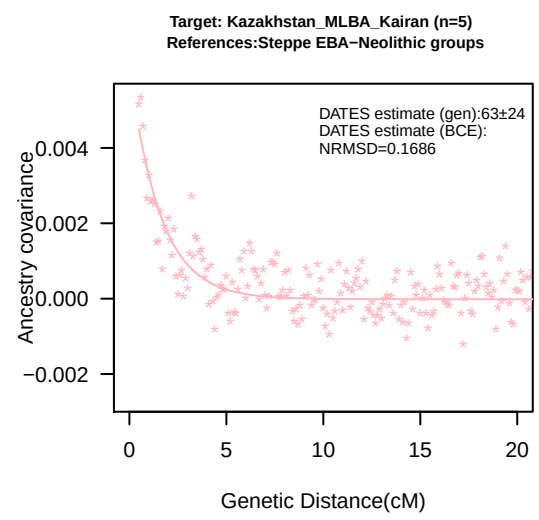

Target: Russia_Afanasievo (n=19)
References:Iran_N_pooled−EHG_pooled
DATES estimate (gen):46±4
DATES estimate (BCE):
NRMSD=0.0717
Ancestry covariance
Genetic Distance(cM)

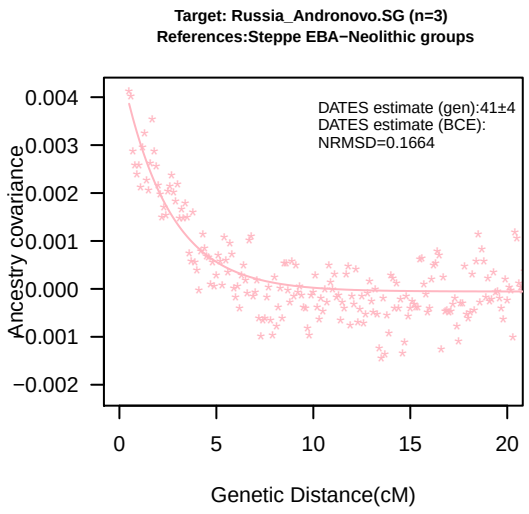

Target: Kazakhstan_MLBA_Aktogai (n=5)
References:Steppe EBA−Neolithic groups
DATES estimate (gen):41±4
DATES estimate (BCE):
NRMSD=0.1517
Ancestry covariance
Genetic Distance(cM)

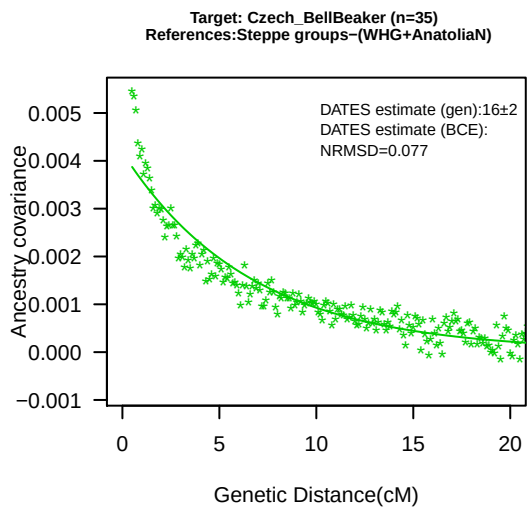

Target: Spain_EN.SG (n=5)
References: WHG_pooled − AnatolianN
DATES estimate (gen):16±4
DATES estimate (BCE):
NRMSD=0.1793
Ancestry covariance
Genetic Distance(cM)

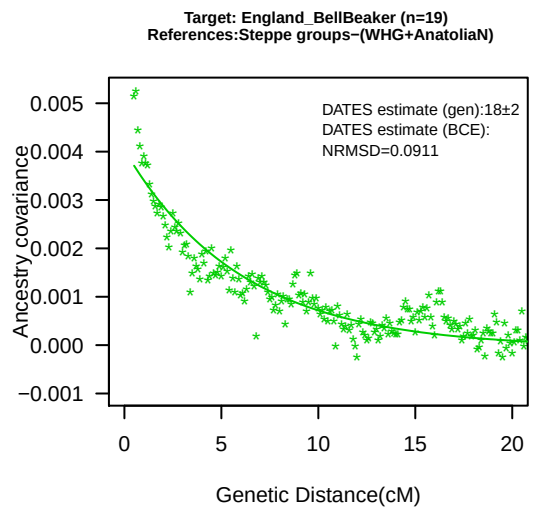

Target: Sweden_Ansarve_Megalithic.SG (n=5)
References: WHG_pooled − AnatolianN
DATES estimate (gen):40±7
DATES estimate (BCE):
NRMSD=0.1366
Ancestry covariance
Genetic Distance(cM)

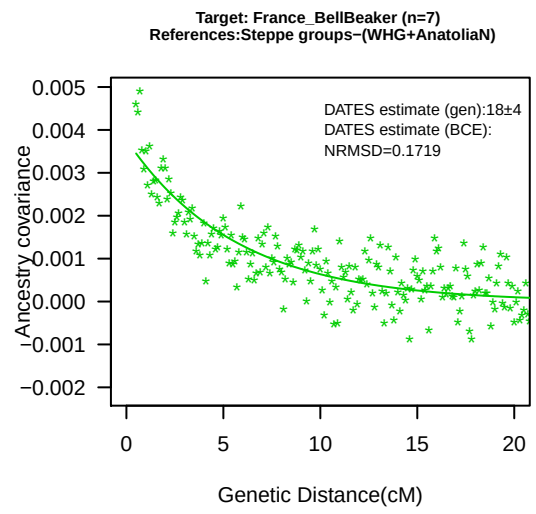

Target: Russia_Caucasus_EBA_Yamnaya (n=3)
References:Iran_N_pooled−EHG_pooled
DATES estimate (gen):38±14
DATES estimate (BCE):
NRMSD=0.2795
Ancestry covariance
Genetic Distance(cM)

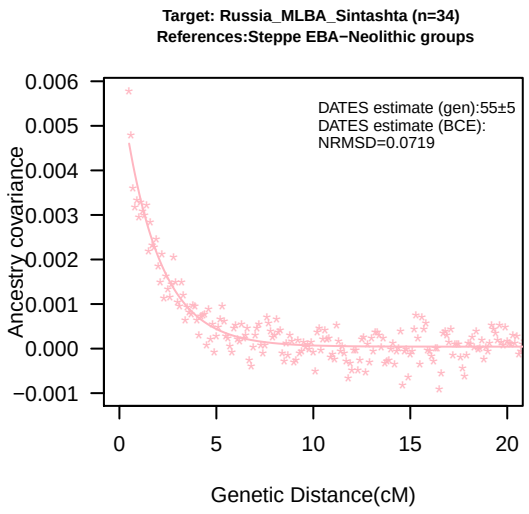

Target: Kazakhstan_MLBA_Kairan (n=5)
References:Steppe EBA−Neolithic groups
DATES estimate (gen):63±24
DATES estimate (BCE):
NRMSD=0.1686
Ancestry covariance
Genetic Distance(cM)

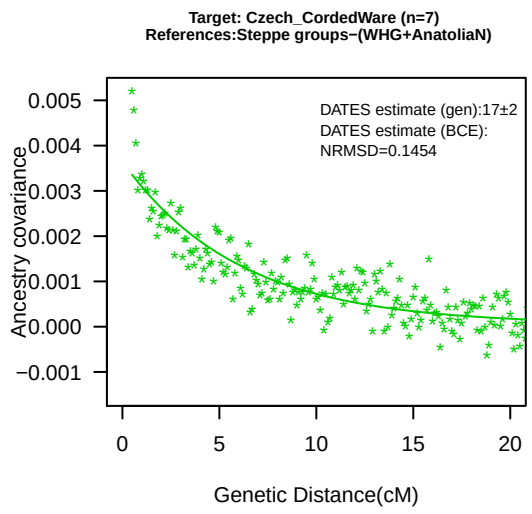

Target: Russia_Andronovo.SG (n=3)
References:Steppe EBA−Neolithic groups
DATES estimate (gen):41±4
DATES estimate (BCE):
NRMSD=0.1664
Ancestry covariance
Genetic Distance(cM)

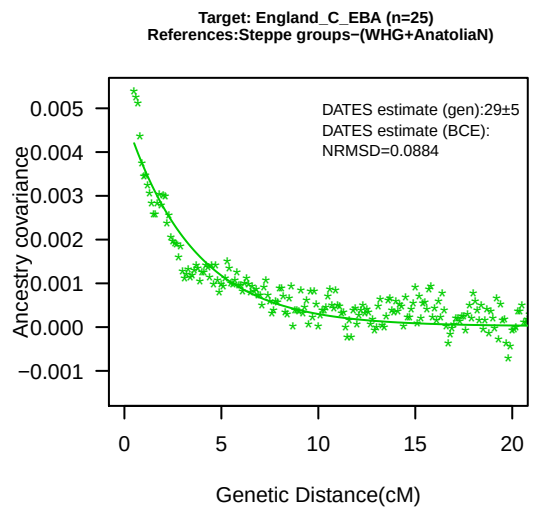

Target: Czech_BellBeaker (n=35)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):16±2
DATES estimate (BCE):
NRMSD=0.077
Ancestry covariance
Genetic Distance(cM)

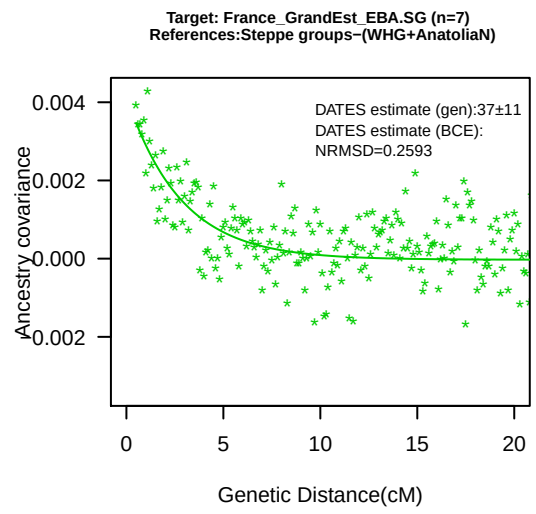

Target: England_BellBeaker (n=19)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):18±2
DATES estimate (BCE):
NRMSD=0.0911
Ancestry covariance
Genetic Distance(cM)

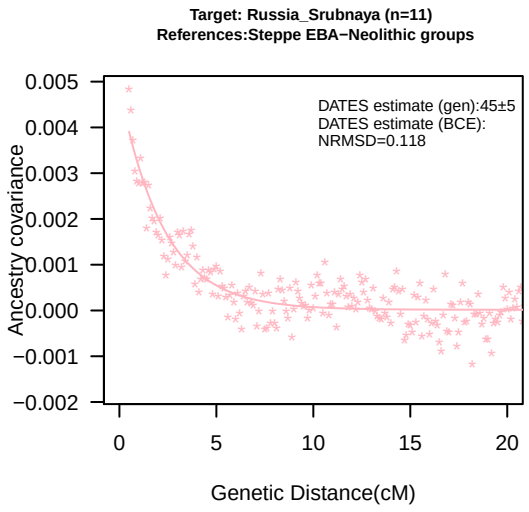

Target: France_BellBeaker (n=7)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):18±4
DATES estimate (BCE):
NRMSD=0.1719
Ancestry covariance
Genetic Distance(cM)

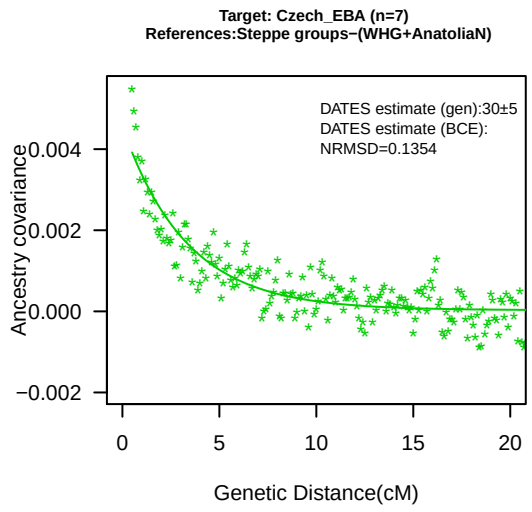

Target: Russia_MLBA_Sintashta (n=34)
References:Steppe EBA−Neolithic groups
DATES estimate (gen):55±5
DATES estimate (BCE):
NRMSD=0.0719
Ancestry covariance
Genetic Distance(cM)

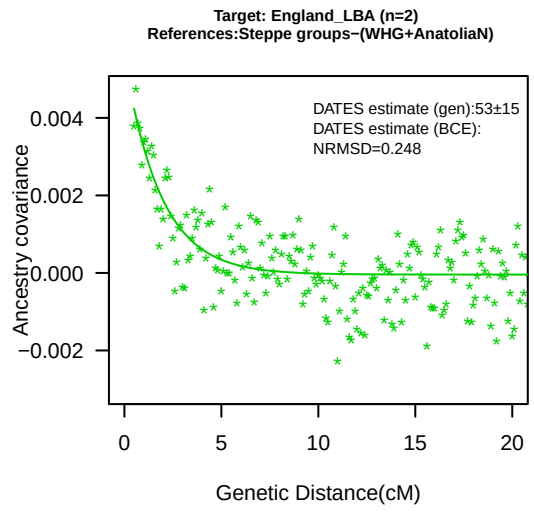

Target: Czech_CordedWare (n=7)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):17±2
DATES estimate (BCE):
NRMSD=0.1454
Ancestry covariance
Genetic Distance(cM)

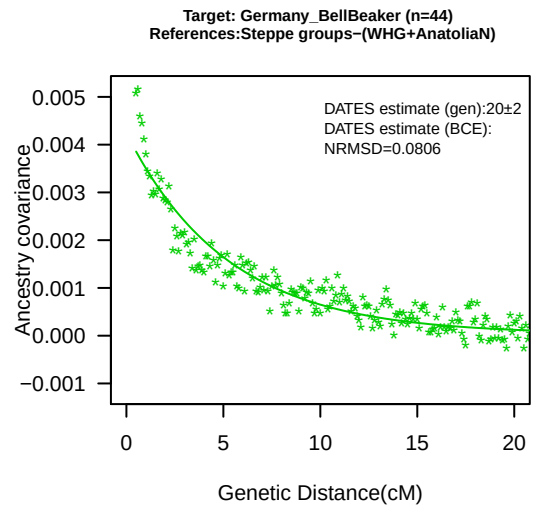

Target: England_C_EBA (n=25)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):29±5
DATES estimate (BCE):
NRMSD=0.0884
Ancestry covariance
Genetic Distance(cM)

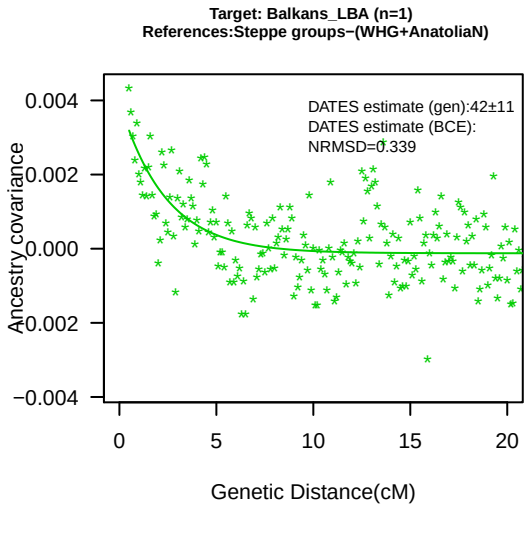

Target: France_GrandEst_EBA.SG (n=7)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):37±11
DATES estimate (BCE):
NRMSD=0.2593
Ancestry covariance
Genetic Distance(cM)

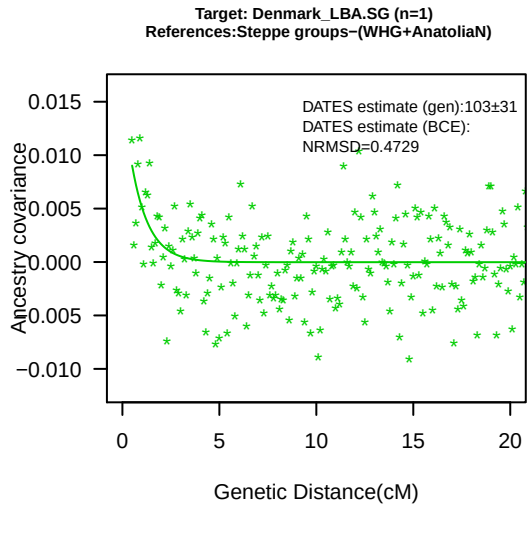

Target: Russia_Srubnaya (n=11)
References:Steppe EBA−Neolithic groups
DATES estimate (gen):45±5
DATES estimate (BCE):
NRMSD=0.118
Ancestry covariance
Genetic Distance(cM)

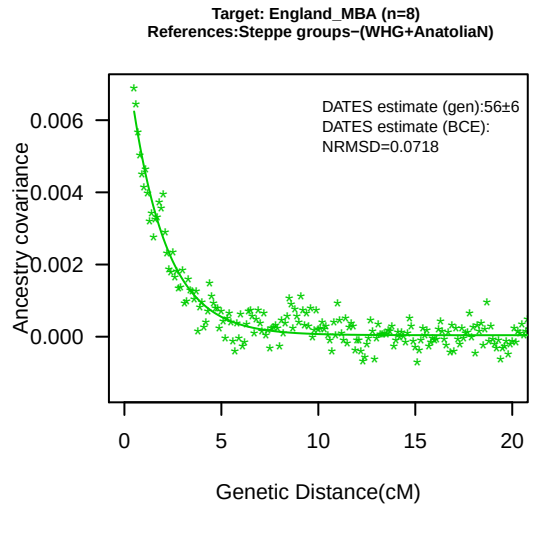

Target: Czech_EBA (n=7)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):30±5
DATES estimate (BCE):
NRMSD=0.1354
Ancestry covariance
Genetic Distance(cM)

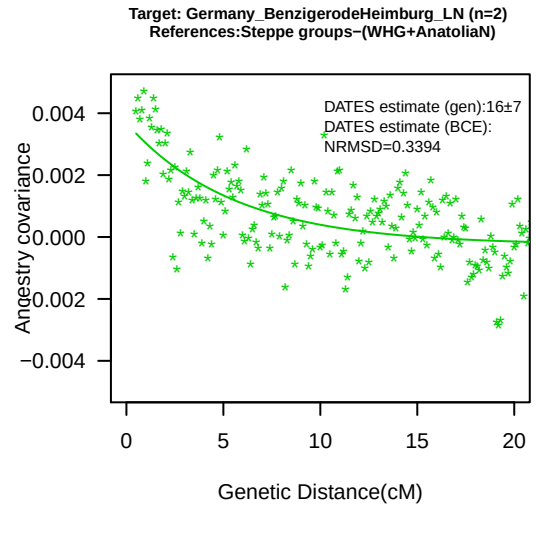

Target: England_LBA (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):53±15
DATES estimate (BCE):
NRMSD=0.248
Ancestry covariance
Genetic Distance(cM)

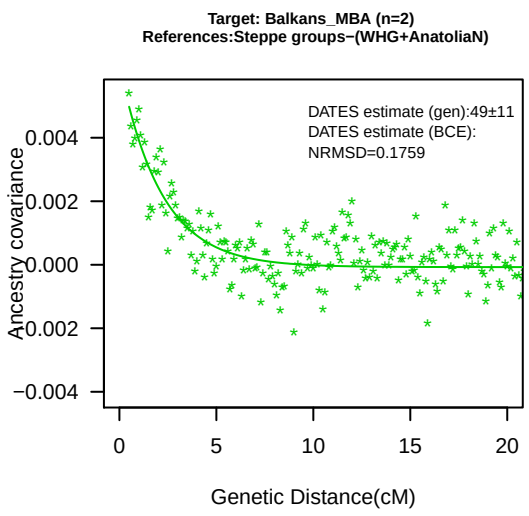

Target: Germany_BellBeaker (n=44)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):20±2
DATES estimate (BCE):
NRMSD=0.0806
Ancestry covariance
Genetic Distance(cM)

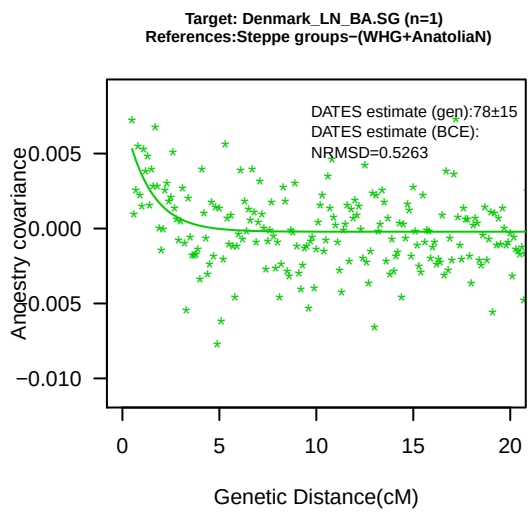

Target: Balkans_LBA (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):42±11
DATES estimate (BCE):
NRMSD=0.339
Ancestry covariance
Genetic Distance(cM)

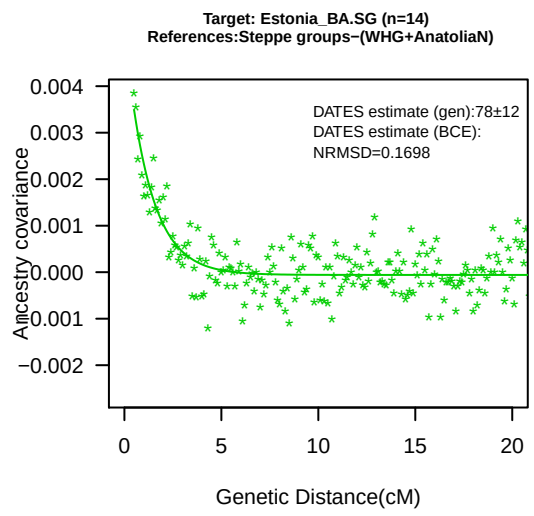

Target: Denmark_LBA.SG (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):103±31
DATES estimate (BCE):
NRMSD=0.4729
Ancestry covariance
Genetic Distance(cM)

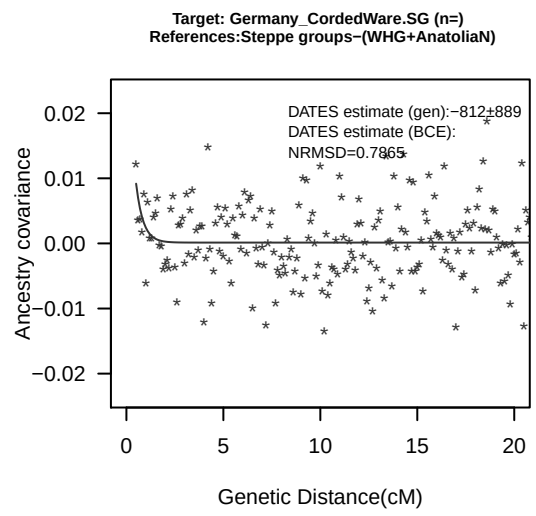

Target: England_MBA (n=8)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):56±6
DATES estimate (BCE):
NRMSD=0.0718
Ancestry covariance
Genetic Distance(cM)

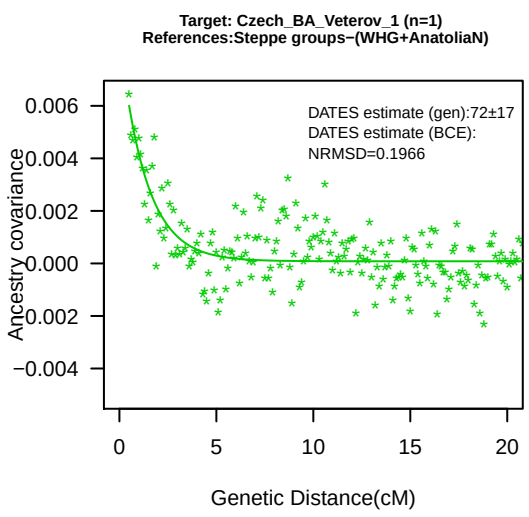

Target: Germany_BenzigerodeHeimburg_LN (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):16±7
DATES estimate (BCE):
NRMSD=0.3394
Ancestry covariance
Genetic Distance(cM)

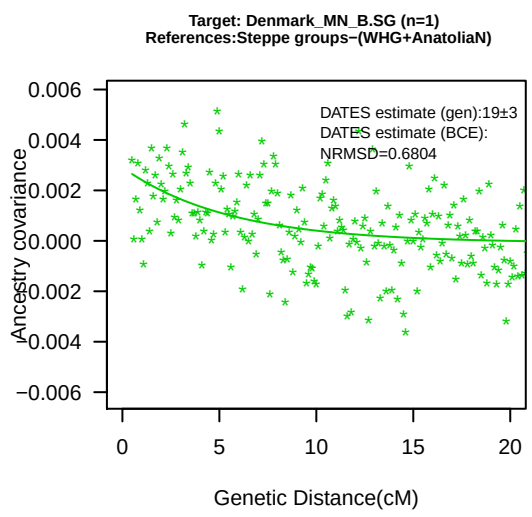

Target: Balkans_MBA (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):49±11
DATES estimate (BCE):
NRMSD=0.1759
Ancestry covariance
Genetic Distance(cM)

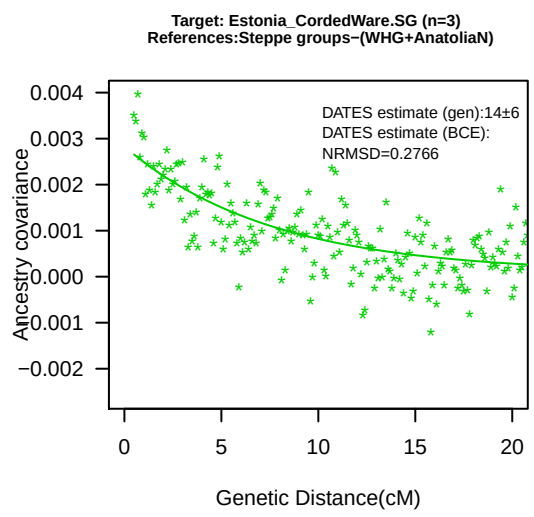

Target: Denmark_LN_BA.SG (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):78±15
DATES estimate (BCE):
NRMSD=0.5263
Ancestry covariance
Genetic Distance(cM)

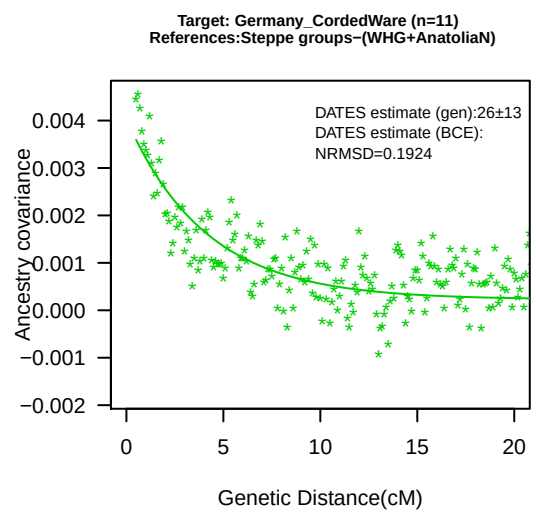

Target: Estonia_BA.SG (n=14)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):78±12
DATES estimate (BCE):
NRMSD=0.1698
Ancestry covariance
Genetic Distance(cM)

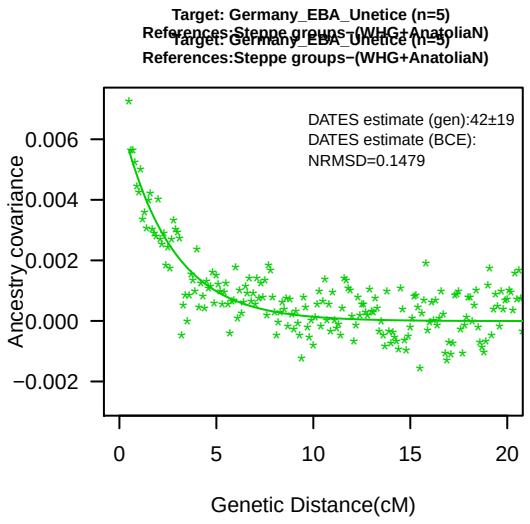

Target: Germany_CordedWare.SG (n=)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):−812±889
DATES estimate (BCE):
NRMSD=0.7865
Ancestry covariance
Genetic Distance(cM)

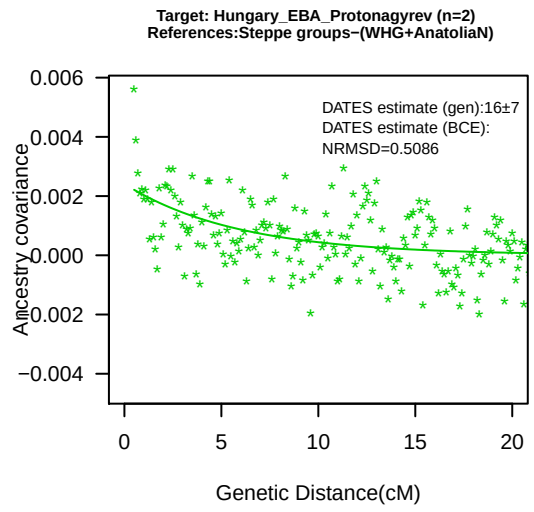

Target: Czech_BA_Veterov_1 (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):72±17
DATES estimate (BCE):
NRMSD=0.1966
Ancestry covariance
Genetic Distance(cM)

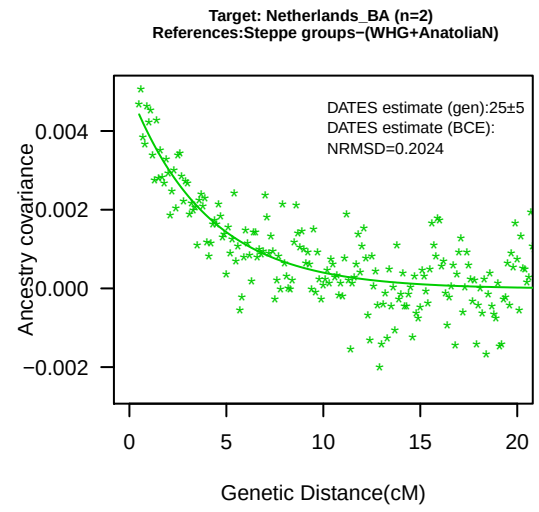

Target: Denmark_MN_B.SG (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):19±3
DATES estimate (BCE):
NRMSD=0.6804
Ancestry covariance
Genetic Distance(cM)

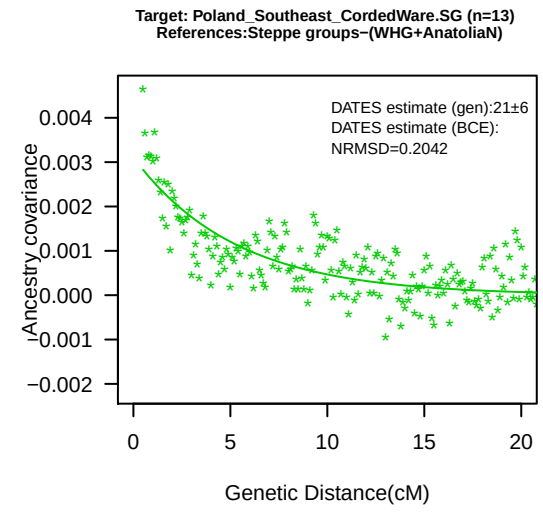

Target: Estonia_CordedWare.SG (n=3)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):14±6
DATES estimate (BCE):
NRMSD=0.2766
Ancestry covariance
Genetic Distance(cM)

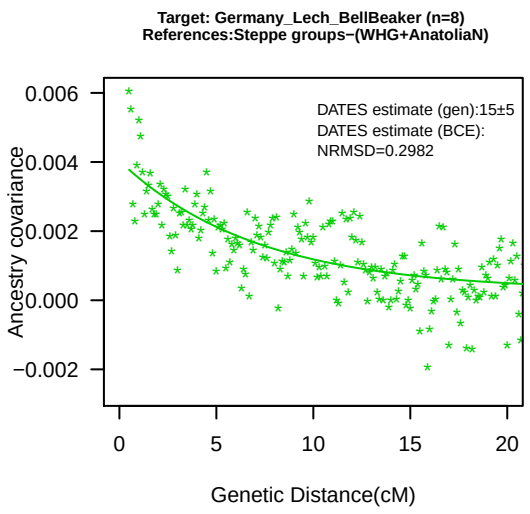

Target: Germany_CordedWare (n=11)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):26±13
DATES estimate (BCE):
NRMSD=0.1924
Ancestry covariance
Genetic Distance(cM)

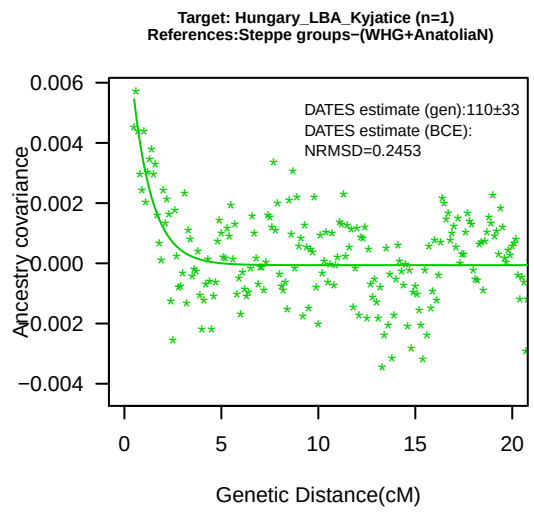

Target: Germany_EBA_Unetice (n=5)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):42±19
DATES estimate (BCE):
NRMSD=0.1479
Ancestry covariance
Genetic Distance(cM)

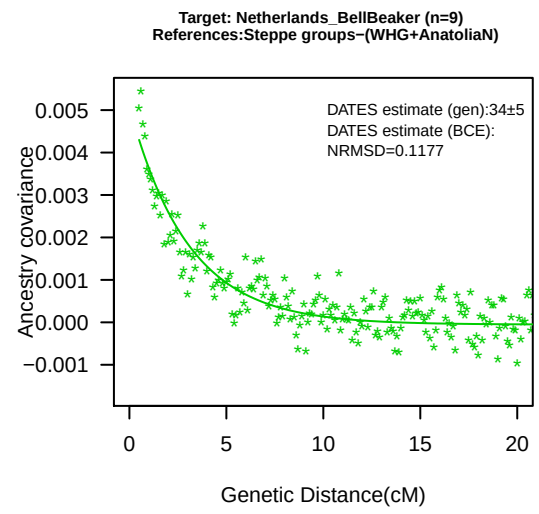

Target: Hungary_EBA_Protonagyrev (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):16±7
DATES estimate (BCE):
NRMSD=0.5086
Ancestry covariance
Genetic Distance(cM)

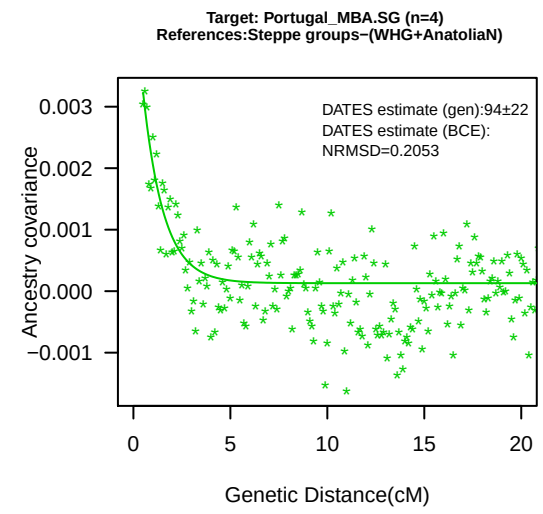

Target: Netherlands_BA (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):25±5
DATES estimate (BCE):
NRMSD=0.2024
Ancestry covariance
Genetic Distance(cM)

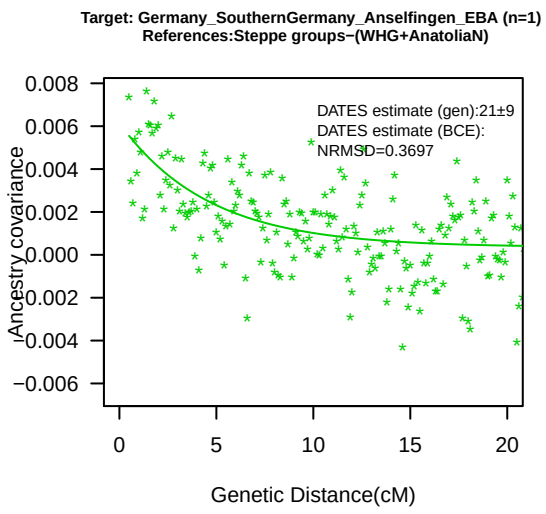

Target: Poland_Southeast_CordedWare.SG (n=13)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):21±6
DATES estimate (BCE):
NRMSD=0.2042
Ancestry covariance
Genetic Distance(cM)

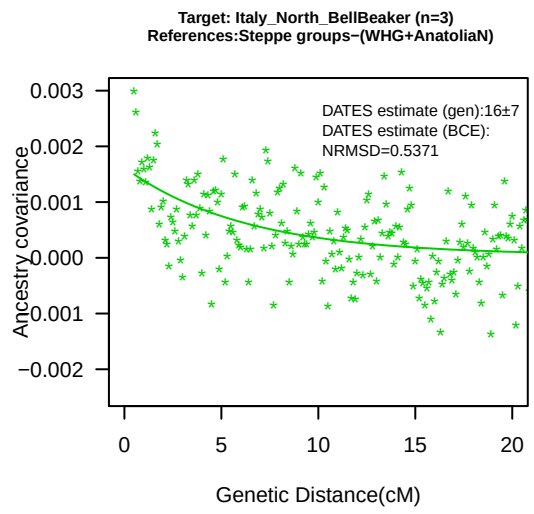

Target: Germany_Lech_BellBeaker (n=8)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):15±5
DATES estimate (BCE):
NRMSD=0.2982
Ancestry covariance
Genetic Distance(cM)

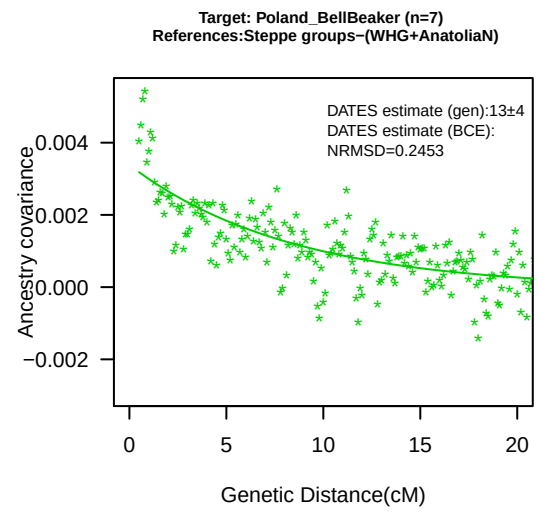

Target: Hungary_LBA_Kyjatice (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):110±33
DATES estimate (BCE):
NRMSD=0.2453
Ancestry covariance
Genetic Distance(cM)

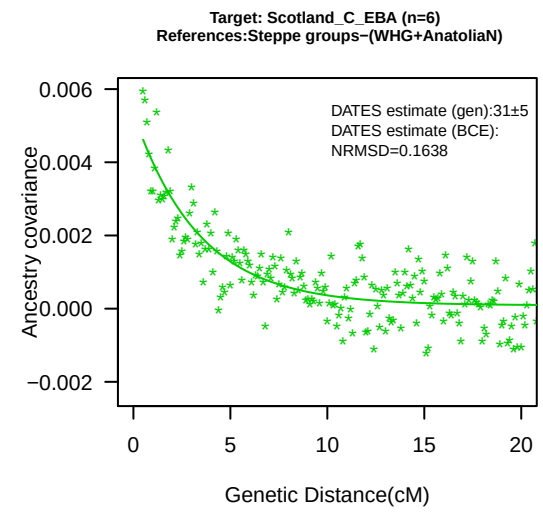

Target: Netherlands_BellBeaker (n=9)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):34±5
DATES estimate (BCE):
NRMSD=0.1177
Ancestry covariance
Genetic Distance(cM)

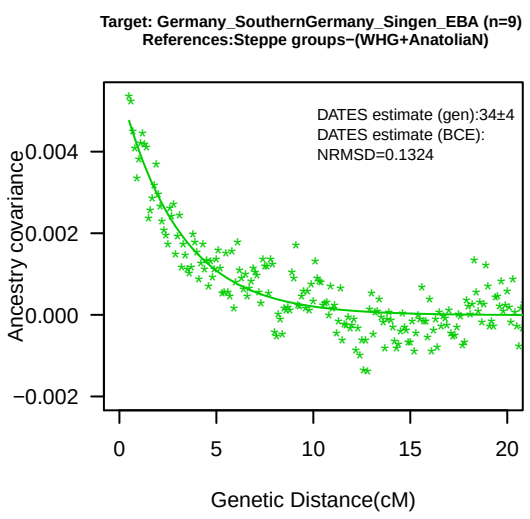

Target: Portugal_MBA.SG (n=4)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):94±22
DATES estimate (BCE):
NRMSD=0.2053
Ancestry covariance
Genetic Distance(cM)

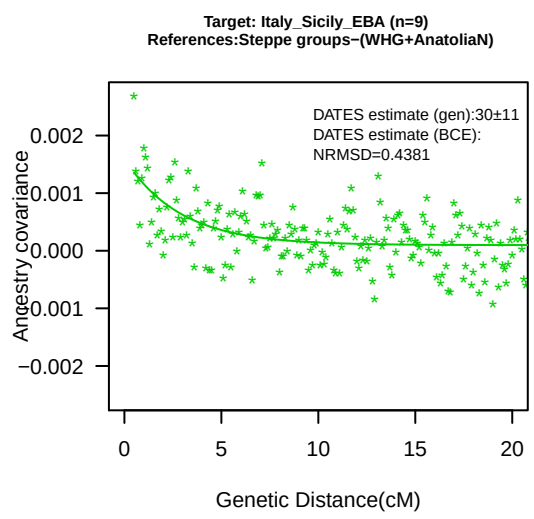

Target: Germany_SouthernGermany_Anselfingen_EBA (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):21±9
DATES estimate (BCE):
NRMSD=0.3697
Ancestry covariance
Genetic Distance(cM)

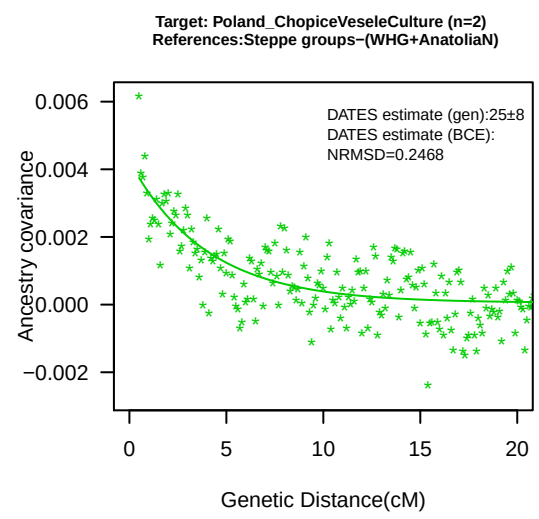

Target: Italy_North_BellBeaker (n=3)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):16±7
DATES estimate (BCE):
NRMSD=0.5371
Ancestry covariance
Genetic Distance(cM)

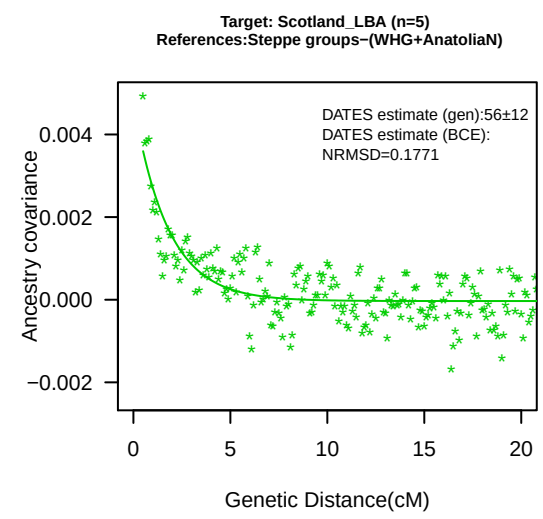

Target: Poland_BellBeaker (n=7)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):13±4
DATES estimate (BCE):
NRMSD=0.2453
Ancestry covariance
Genetic Distance(cM)

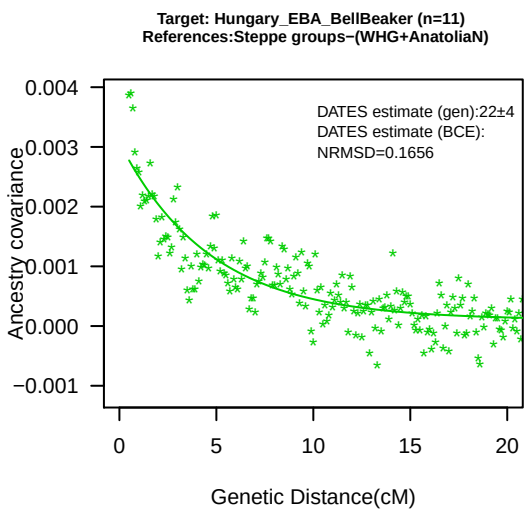

Target: Scotland_C_EBA (n=6)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):31±5
DATES estimate (BCE):
NRMSD=0.1638
Ancestry covariance
Genetic Distance(cM)

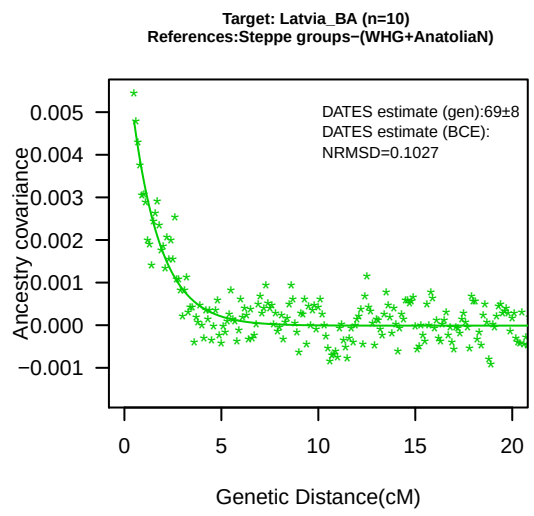

Target: Germany_SouthernGermany_Singen_EBA (n=9)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):34±4
DATES estimate (BCE):
NRMSD=0.1324
Ancestry covariance
Genetic Distance(cM)

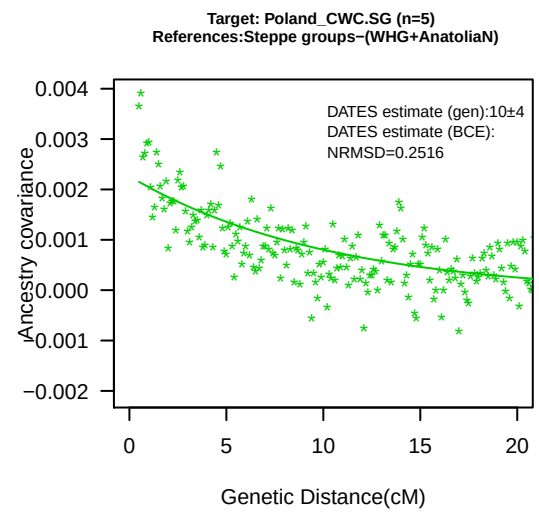

Target: Italy_Sicily_EBA (n=9)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):30±11
DATES estimate (BCE):
NRMSD=0.4381
Ancestry covariance
Genetic Distance(cM)

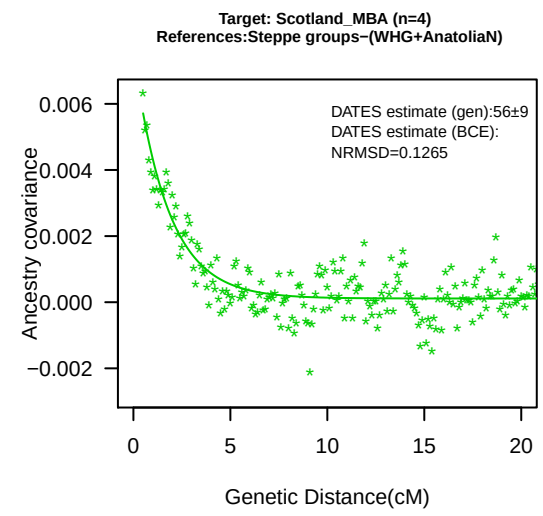

Target: Poland_ChopiceVeseleCulture (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):25±8
DATES estimate (BCE):
NRMSD=0.2468
Ancestry covariance
Genetic Distance(cM)

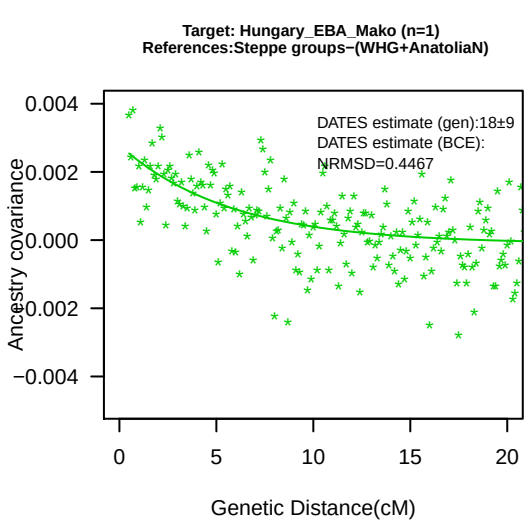

Target: Scotland_LBA (n=5)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):56±12
DATES estimate (BCE):
NRMSD=0.1771
Ancestry covariance
Genetic Distance(cM)

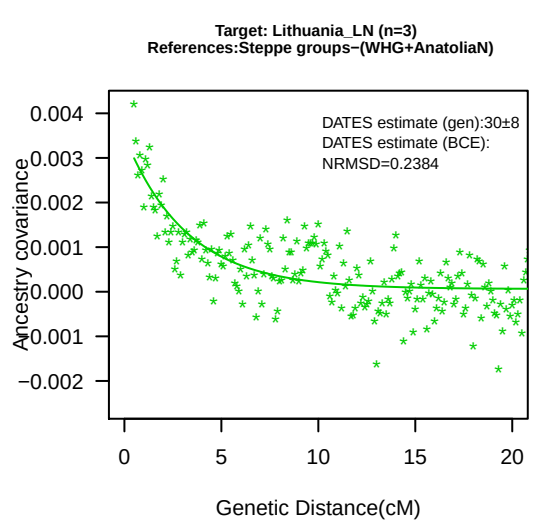

Target: Hungary_EBA_BellBeaker (n=11)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):22±4
DATES estimate (BCE):
NRMSD=0.1656
Ancestry covariance
Genetic Distance(cM)

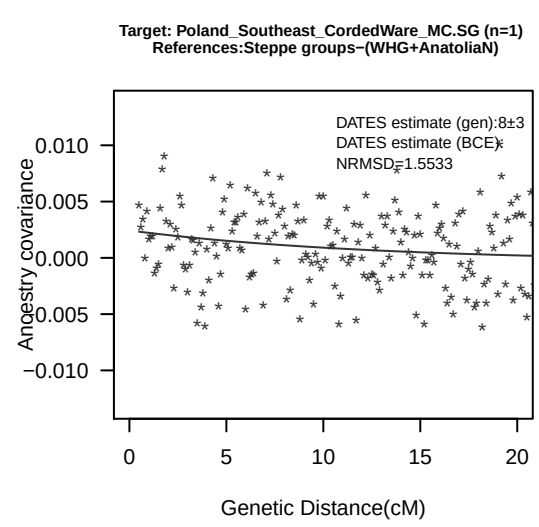

Target: Latvia_BA (n=10)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):69±8
DATES estimate (BCE):
NRMSD=0.1027
Ancestry covariance
Genetic Distance(cM)

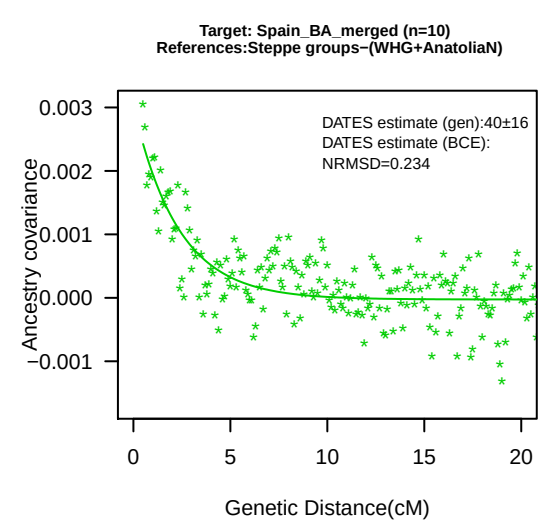

Target: Poland_CWC.SG (n=5)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):10±4
DATES estimate (BCE):
NRMSD=0.2516
Ancestry covariance
Genetic Distance(cM)

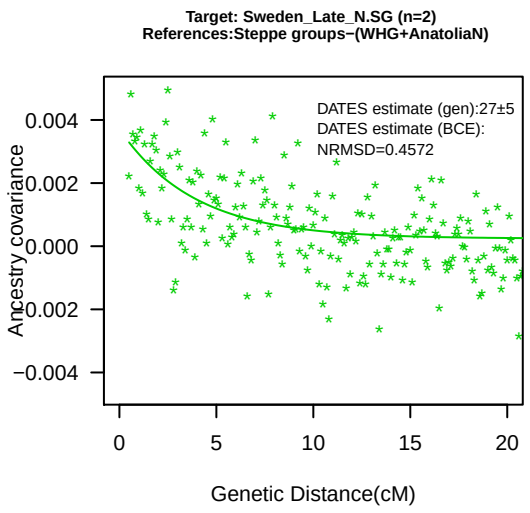

Target: Scotland_MBA (n=4)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):56±9
DATES estimate (BCE):
NRMSD=0.1265
Ancestry covariance
Genetic Distance(cM)

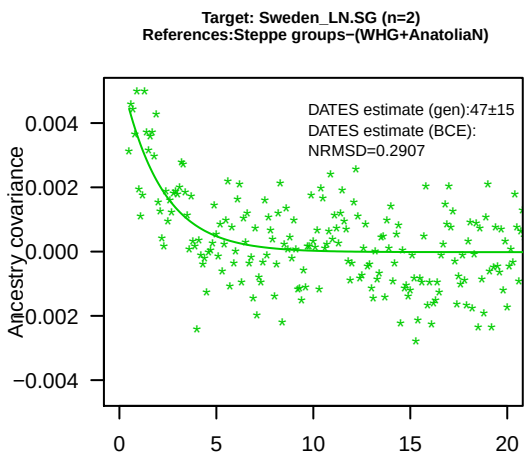

Target: Hungary_EBA_Mako (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):18±9
DATES estimate (BCE):
NRMSD=0.4467
Ancestry covariance
Genetic Distance(cM)

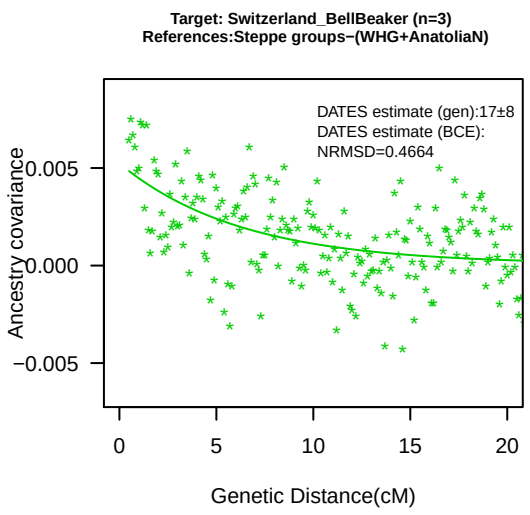

Target: Lithuania_LN (n=3)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):30±8
DATES estimate (BCE):
NRMSD=0.2384
Ancestry covariance
Genetic Distance(cM)

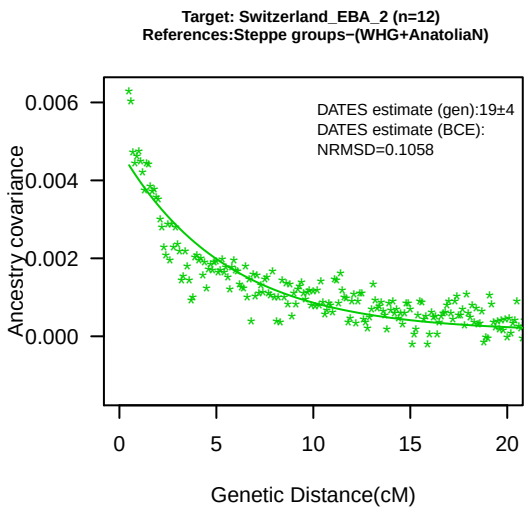

Target: Poland_Southeast_CordedWare_MC.SG (n=1)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):8±3
DATES estimate (BCE):
NRMSD=1.5533
Ancestry covariance
Genetic Distance(cM)


Target: Spain_BA_merged (n=10)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):40±16
DATES estimate (BCE):
NRMSD=0.234
Ancestry covariance
Genetic Distance(cM)

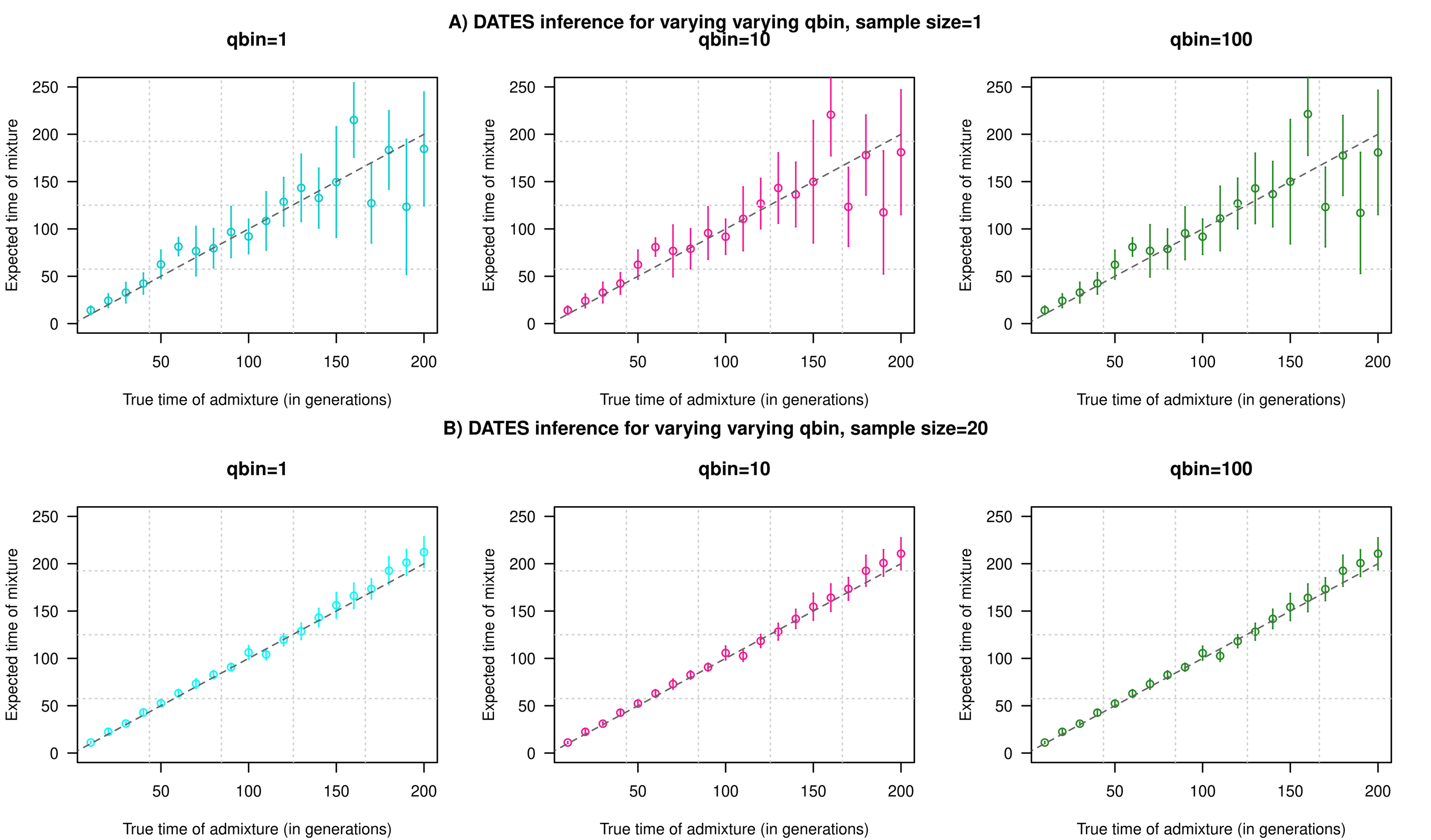

Target: Sweden_Late_N.SG (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):27±5
DATES estimate (BCE):
NRMSD=0.4572
Ancestry covariance
Genetic Distance(cM)

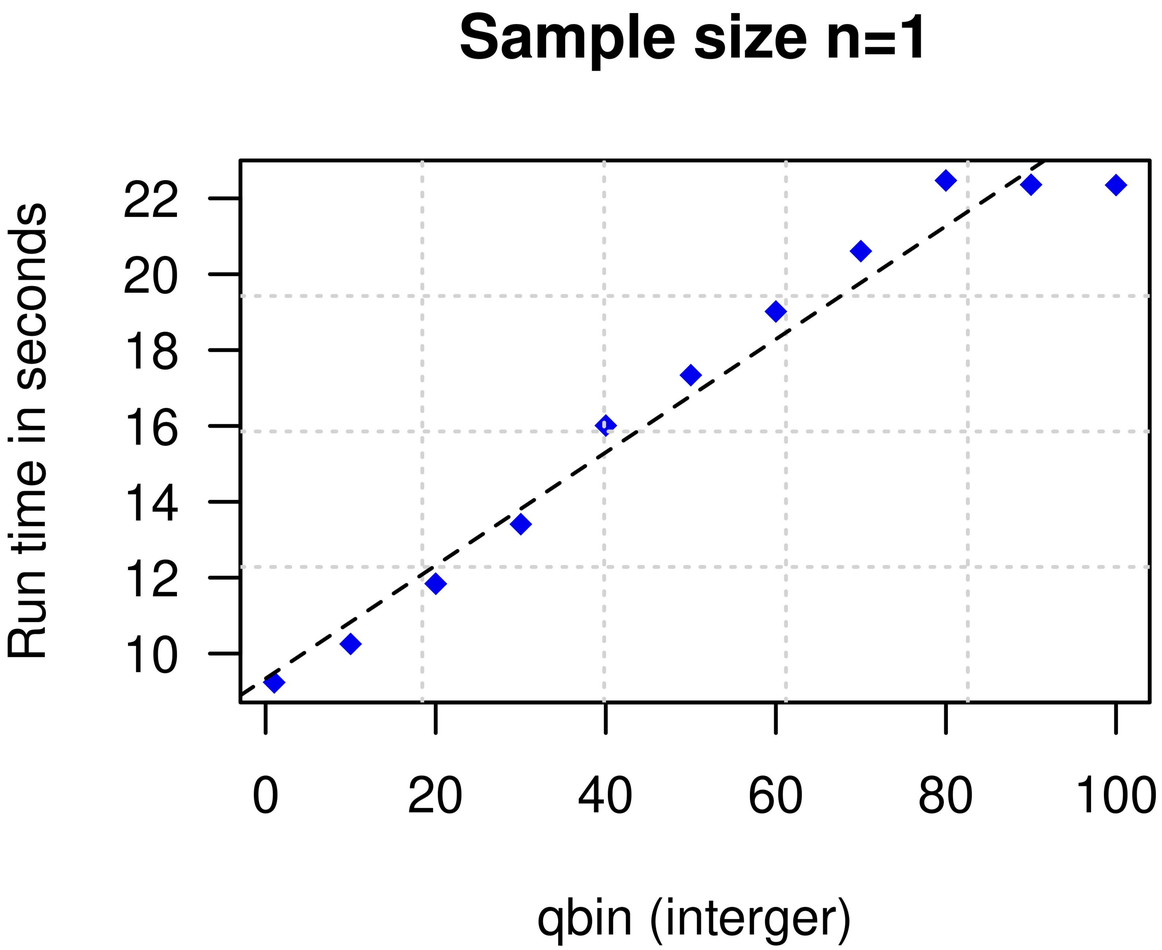

Target: Sweden_LN.SG (n=2)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):47±15
DATES estimate (BCE):
NRMSD=0.2907
Ancestry covariance
Genetic Distance(cM)

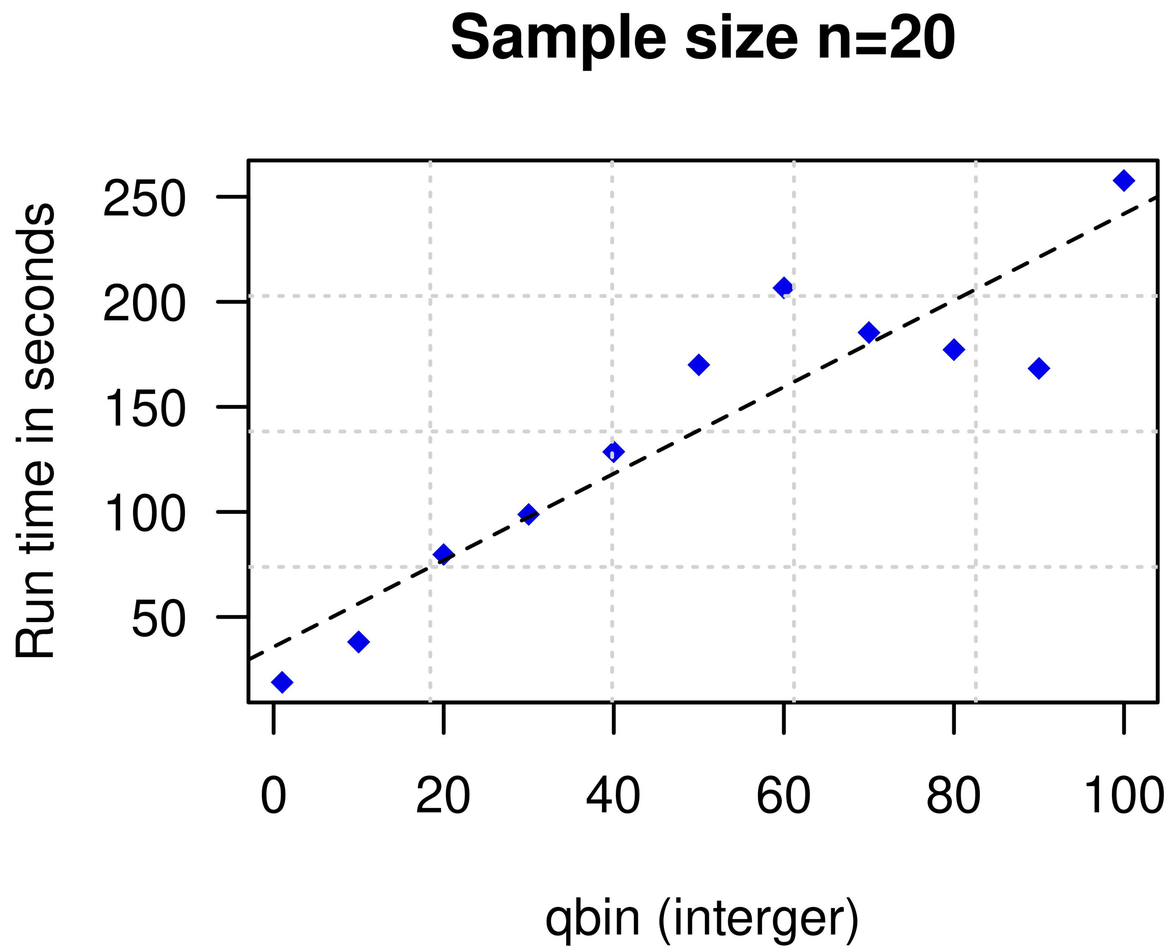

Target: Switzerland_BellBeaker (n=3)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):17±8
DATES estimate (BCE):
NRMSD=0.4664
Ancestry covariance
Genetic Distance(cM)

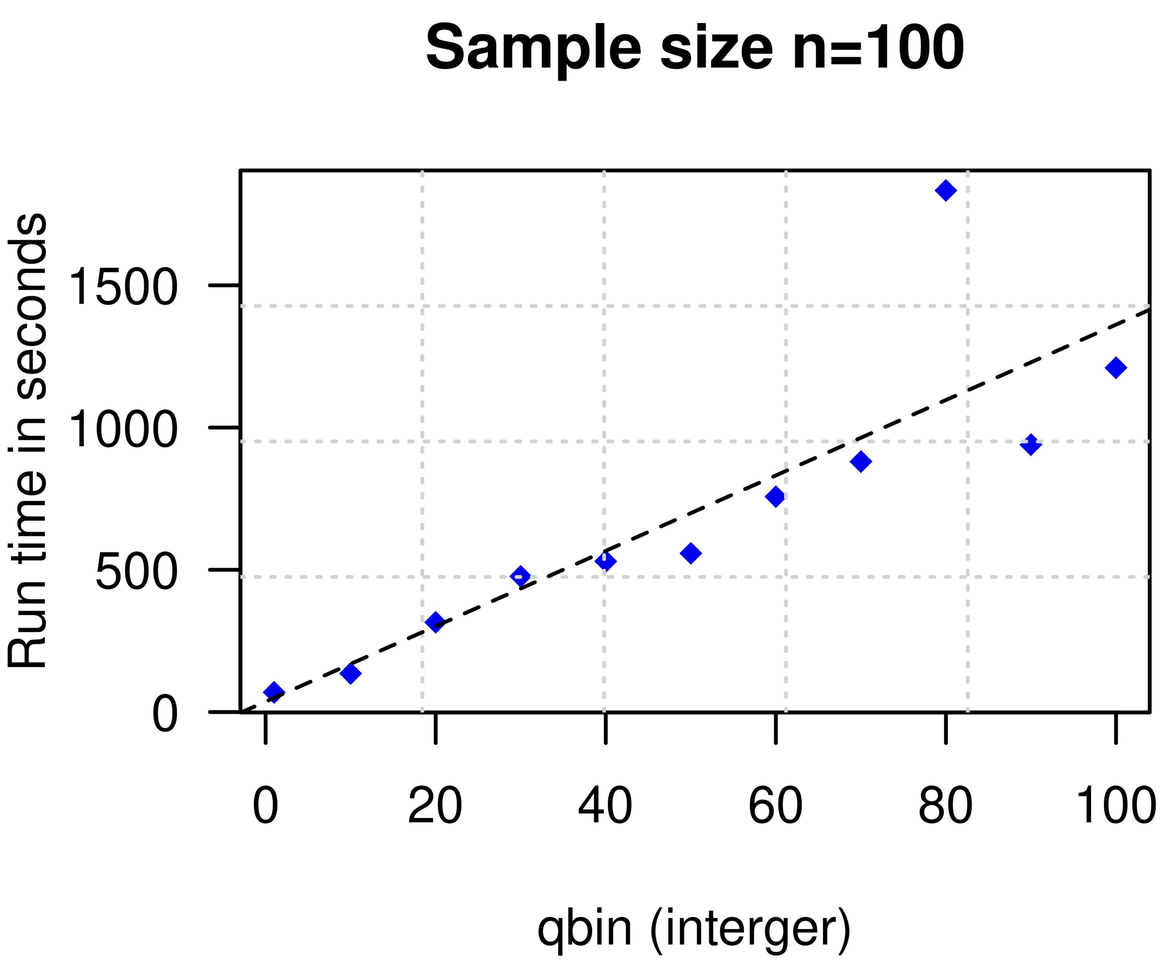

Target: Switzerland_EBA_2 (n=12)
References:Steppe groups−(WHG+AnatoliaN)
DATES estimate (gen):19±4
DATES estimate (BCE):
NRMSD=0.1058
Ancestry covariance
Genetic Distance(cM)

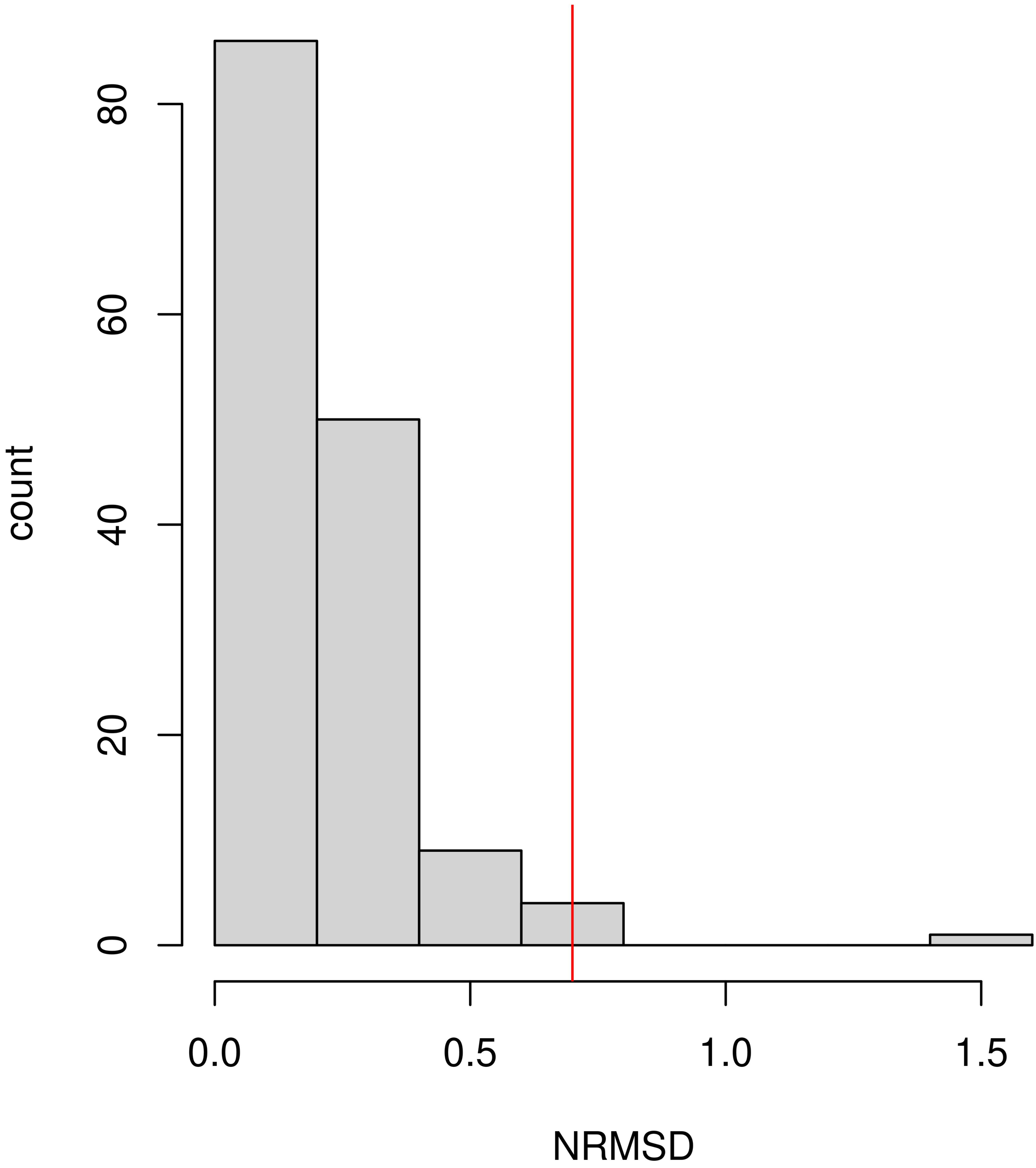
Admixture dates in Iron_Gates samples grouped by c14 age in bins of 500 years
WHG-EHG admixture per c14 bin
WHG-EHG admixture in all samples
average c14 ages
Admixture time in years BCE
15000
10000
5000
5550-6050
6050-6550
6550-7050
7050-7550
7550-8050
8050-8050
ALL-5700-8800

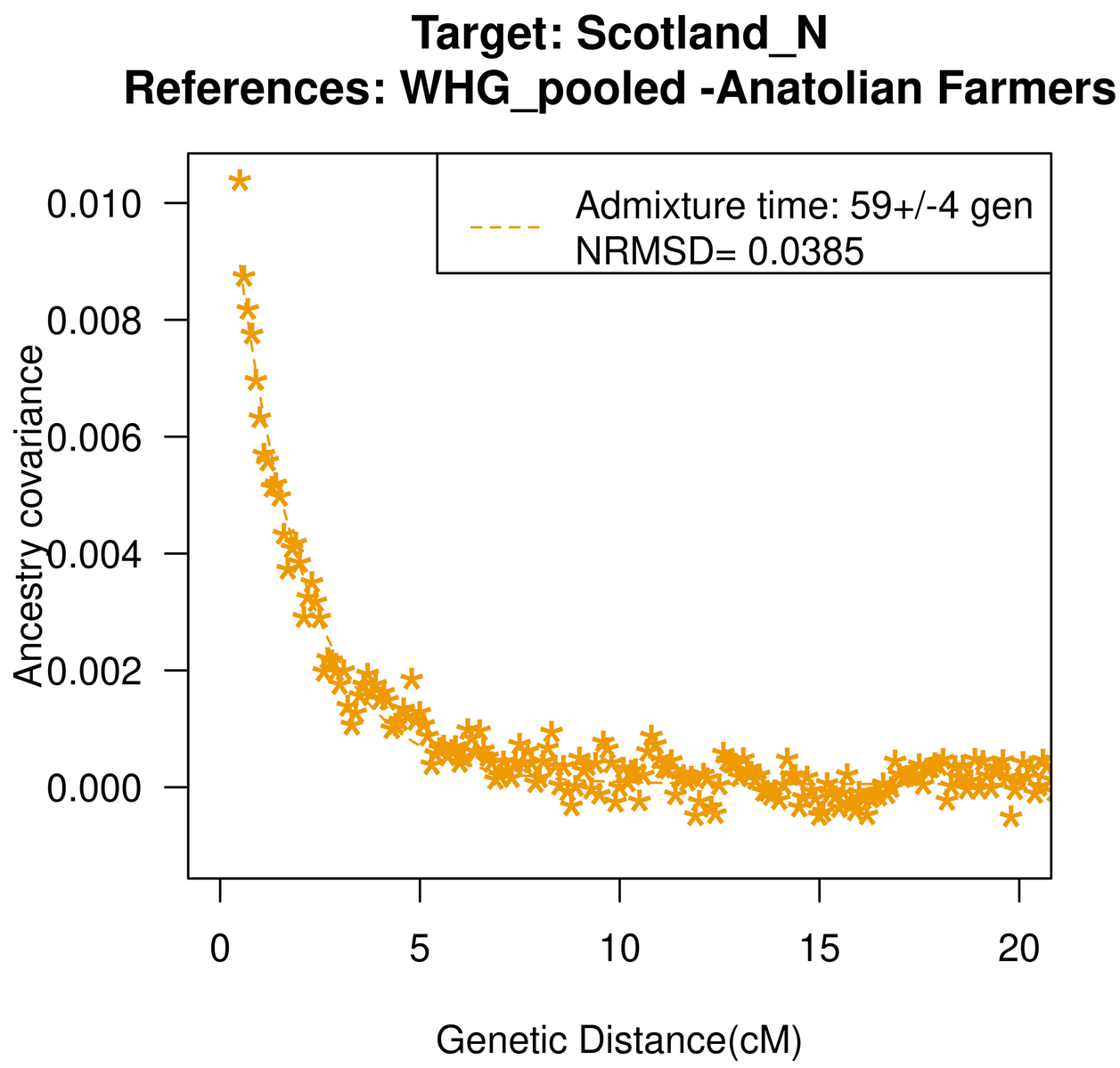

A) DATES inference for varying varying qbin, sample size=1
qbin=1
qbin=10
qbin=100
Expected time of mixture
True time of admixture (in generations)
0
50
100
150
200
250
B) DATES inference for varying varying qbin, sample size=20
qbin=1
qbin=10
qbin=100
Expected time of mixture
True time of admixture (in generations)

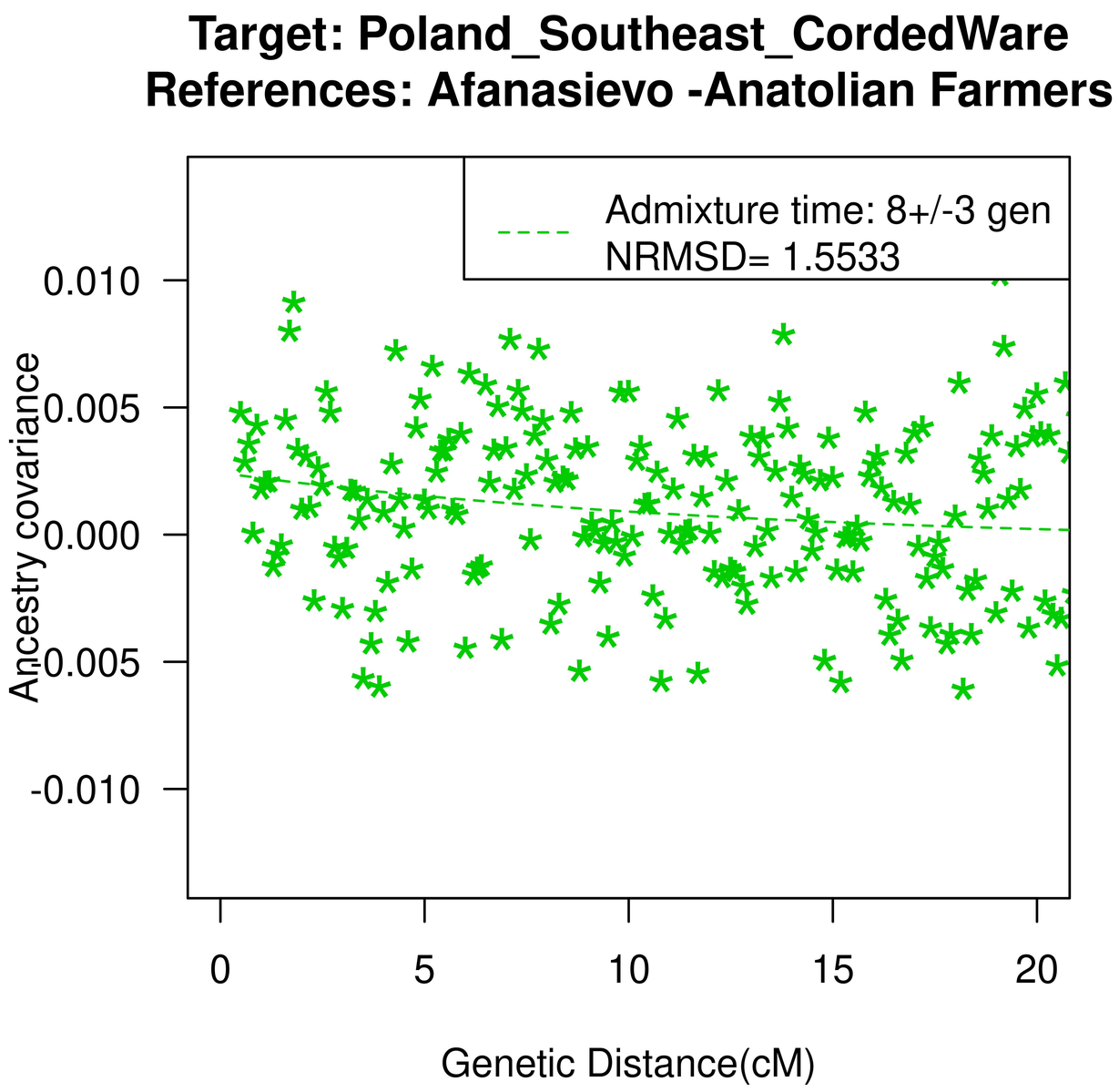

Sample size n=1
Run time in seconds
10
12
14
16
18
20
22
0
20
40
60
80
100
qbin (interger)


Sample size n=20
Run time in seconds
50
100
150
200
250
0
20
40
60
80
100
qbin (interger)


Sample size n=100
Run time in seconds
0
500
1000
1500
0
20
40
60
80
100
qbin (interger)

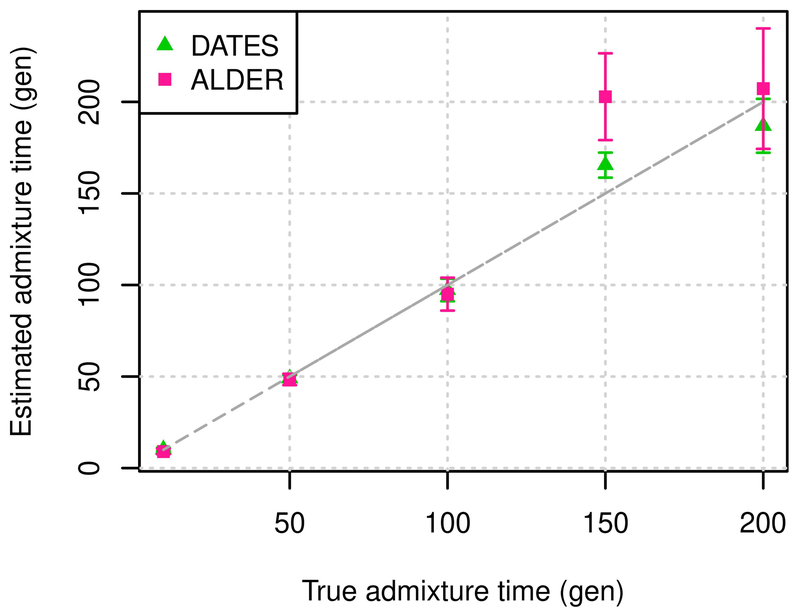

count
0
20
40
60
80
0.0
0.5
1.0
1.5
NRMSD


Target: Scotland_N
References: WHG_pooled -Anatolian Farmers
Admixture time: 59+/-4 gen
NRMSD= 0.0385
Ancestry covariance
0.010
0.008
0.006
0.004
0.002
0.000
0
5
10
15
20
Genetic Distance(cM)

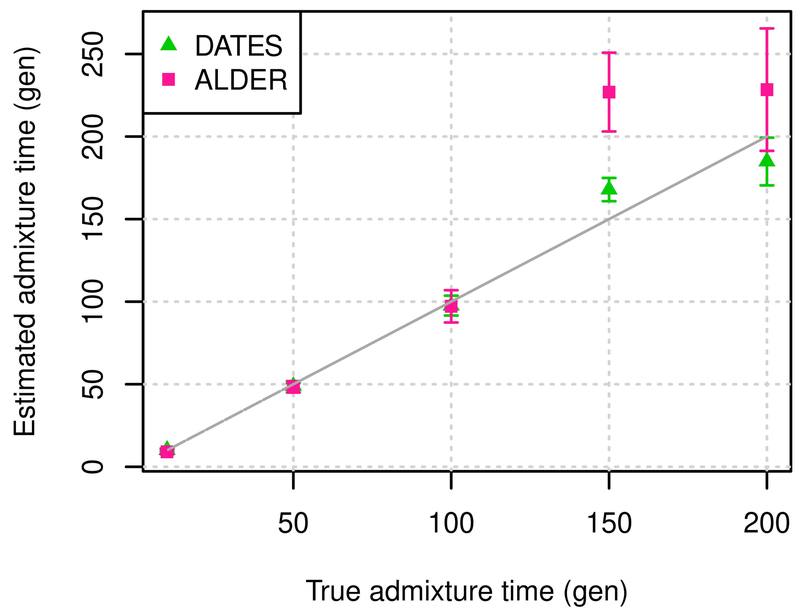

Target: Poland_Southeast_CordedWare
References: Afanasievo -Anatolian Farmers
Admixture time: 8+/-3 gen
NRMSD= 1.5533
Ancestry covariance
0.010
0.005
0.000
-0.005
-0.010
0
5
10
15
20
Genetic Distance(cM)

Effect of missing genotypes on the performance of DATES and ALDER

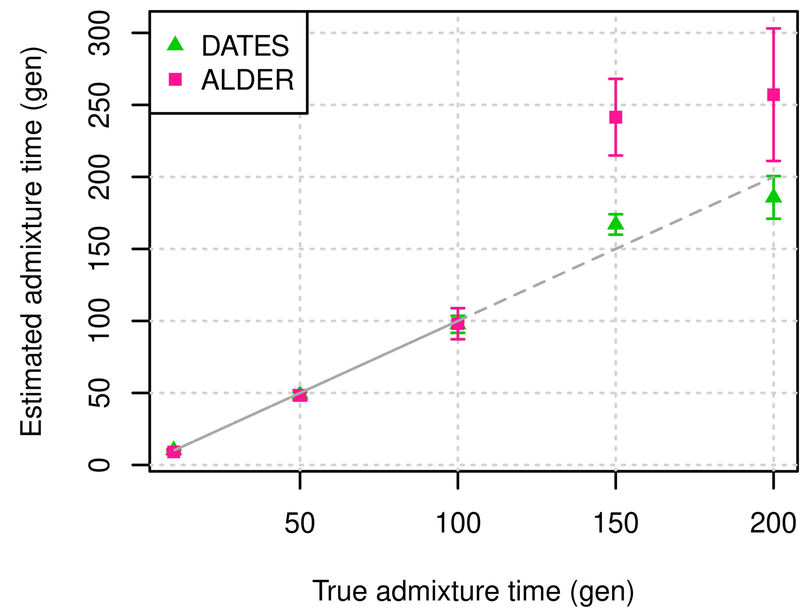
Target genome missing prop=10%



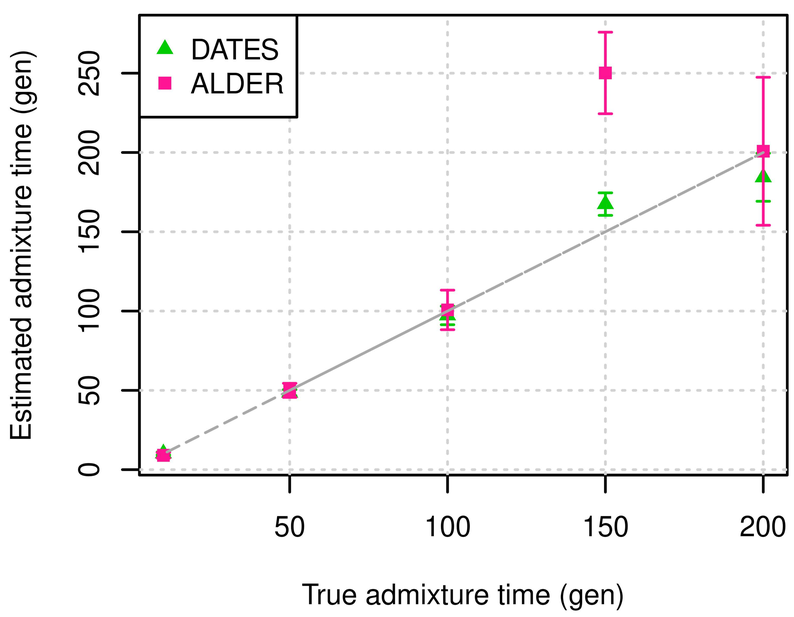
Target genome missing prop=15%



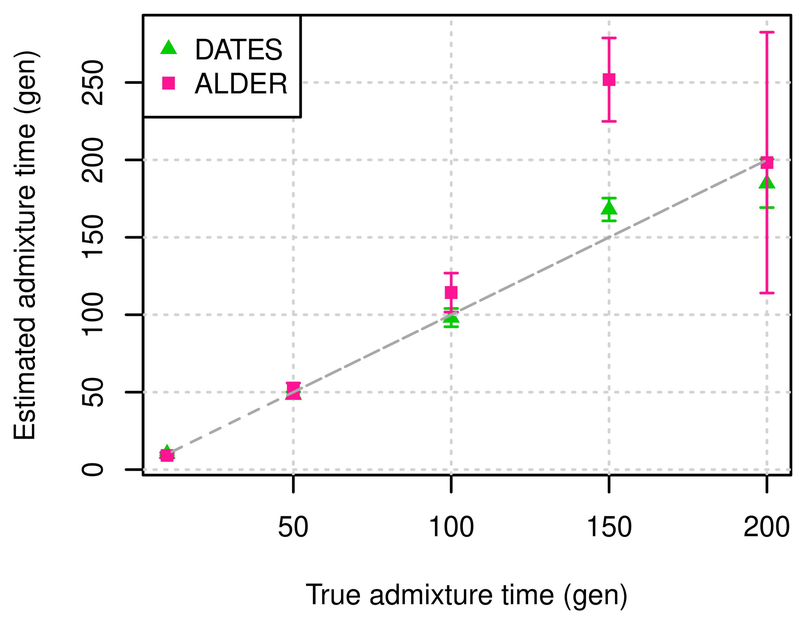
Target genome missing prop=20%



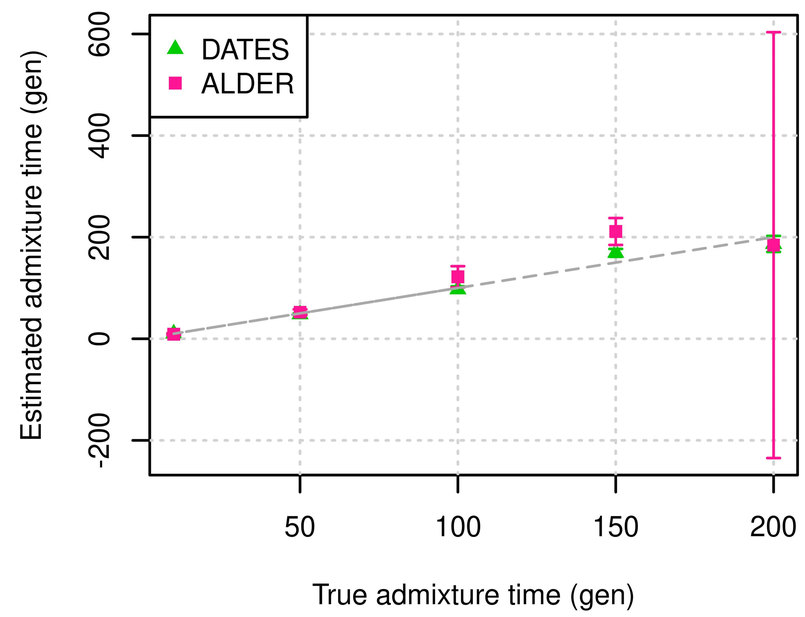
Target genome missing prop=25%



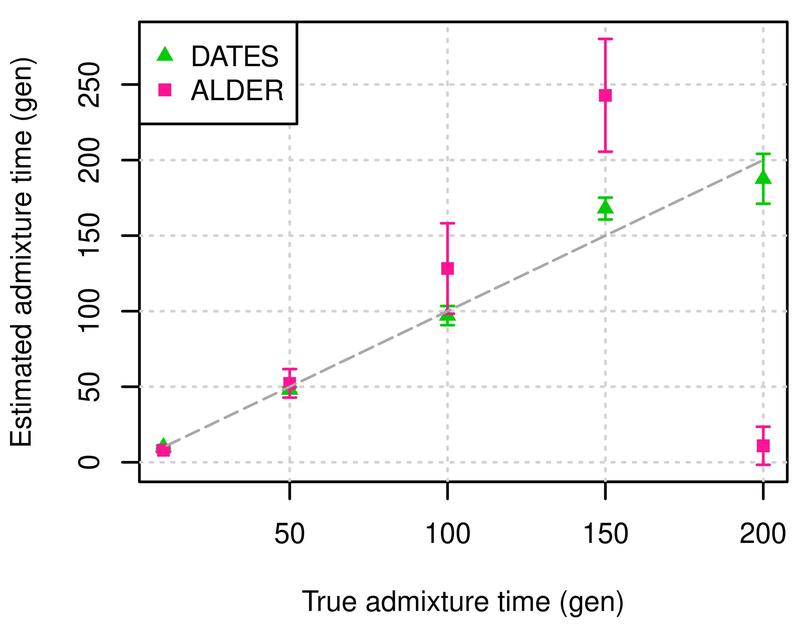
Target genome missing prop=30%



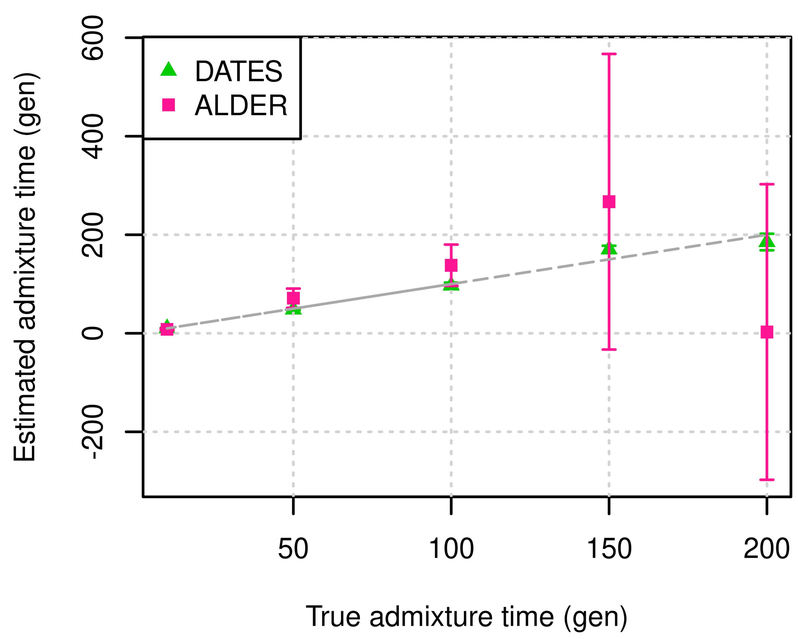
Target genome missing prop=35%



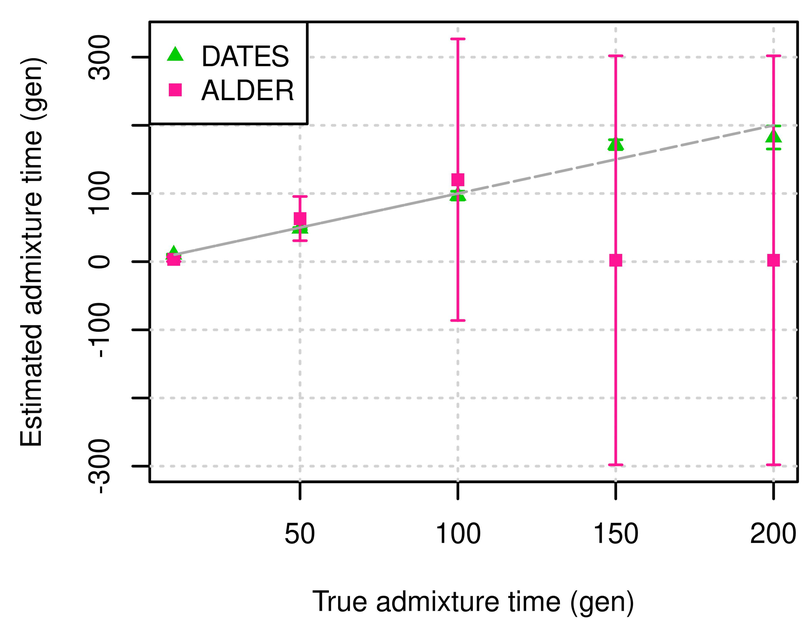
Target genome missing prop=40%



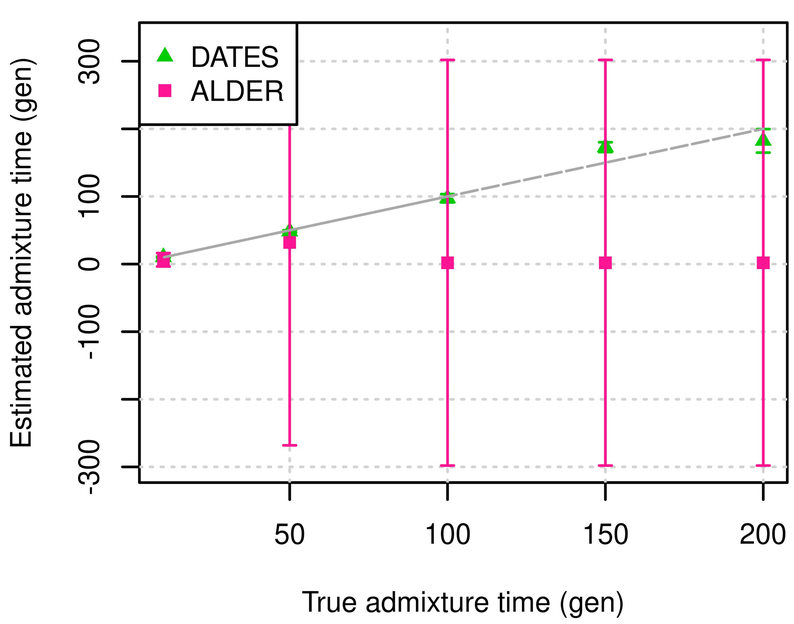
Target genome missing prop=45%



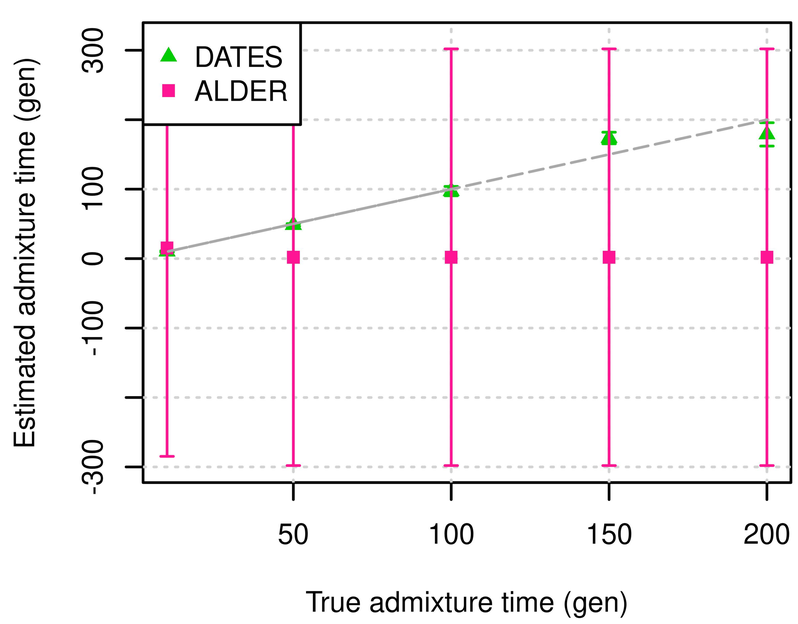
Target genome missing prop=50%

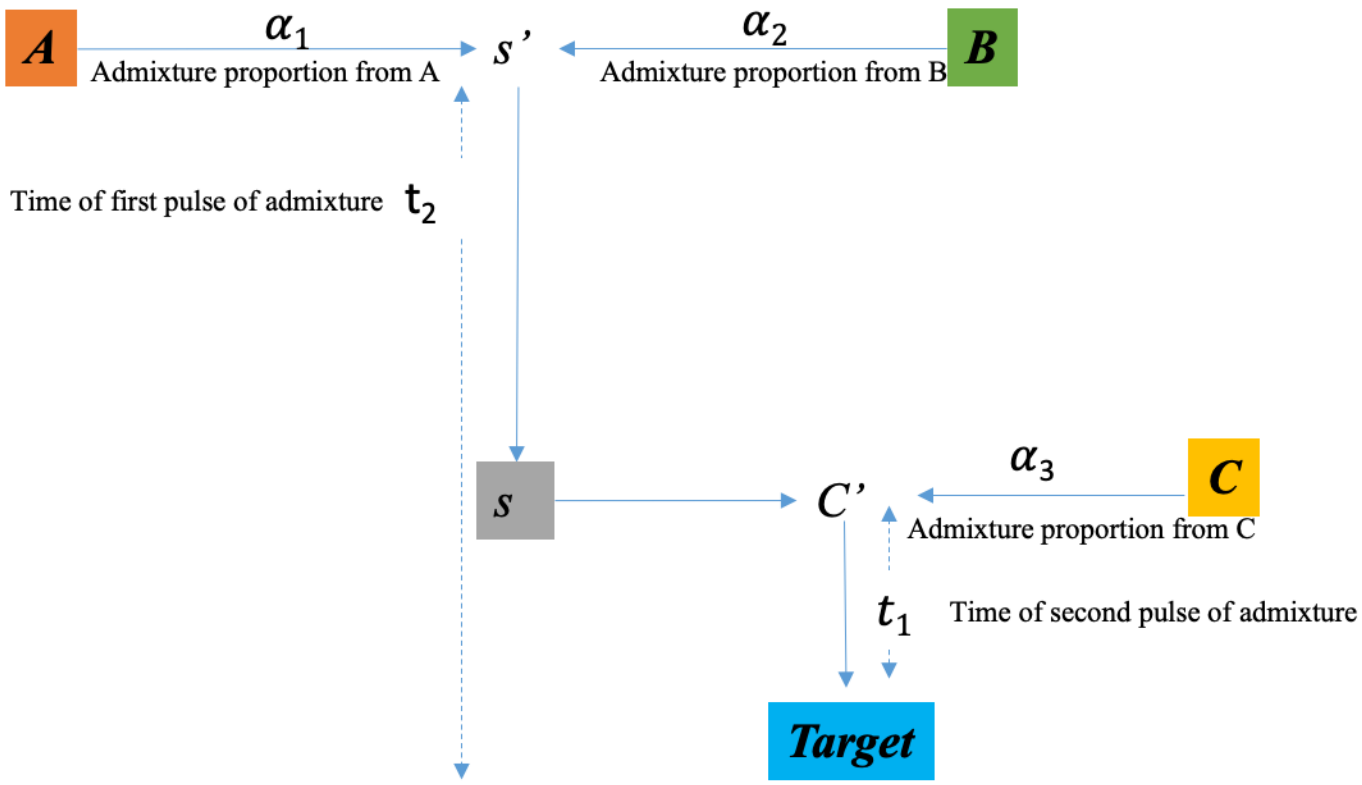


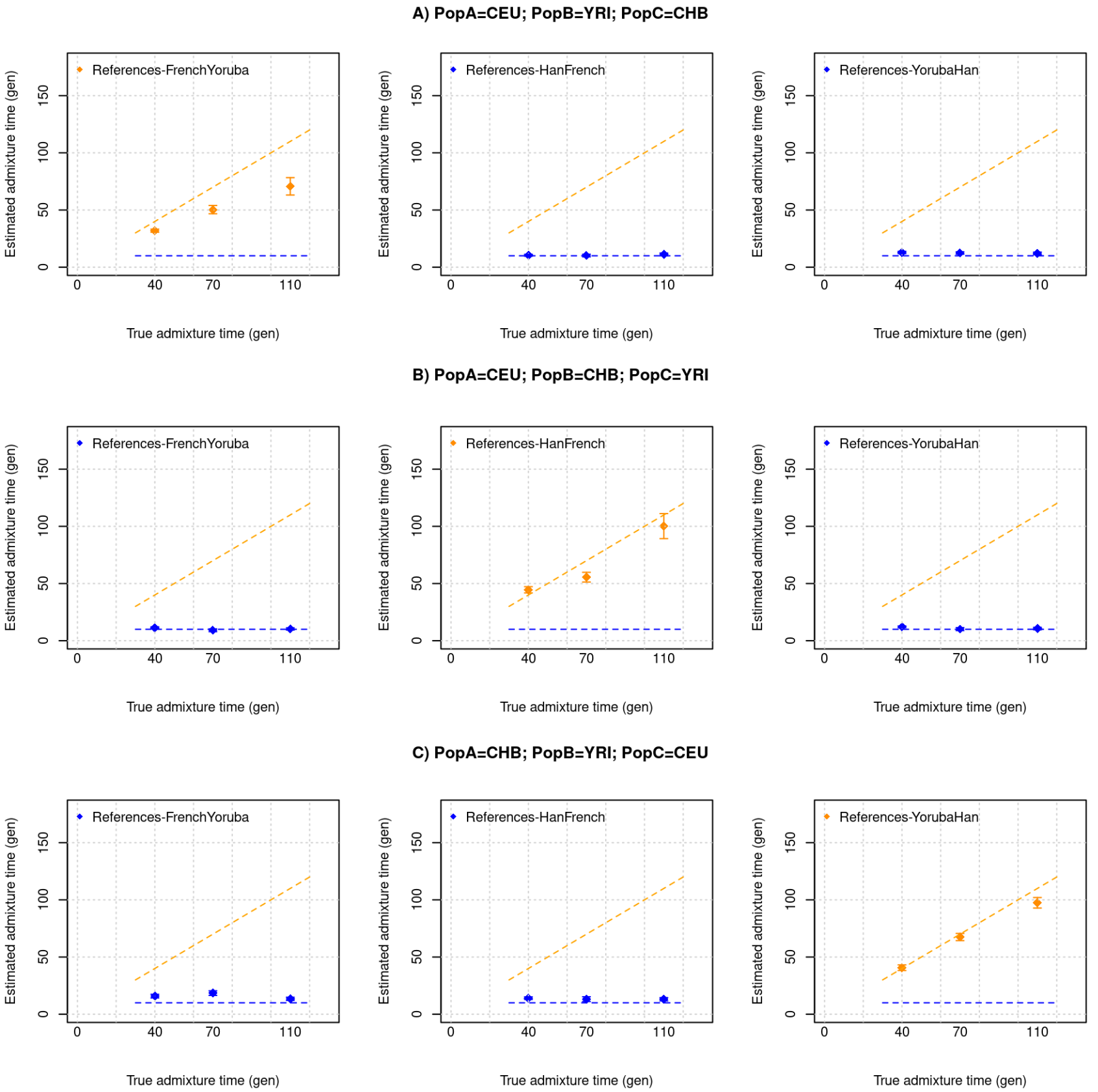
Target genome missing prop=55%

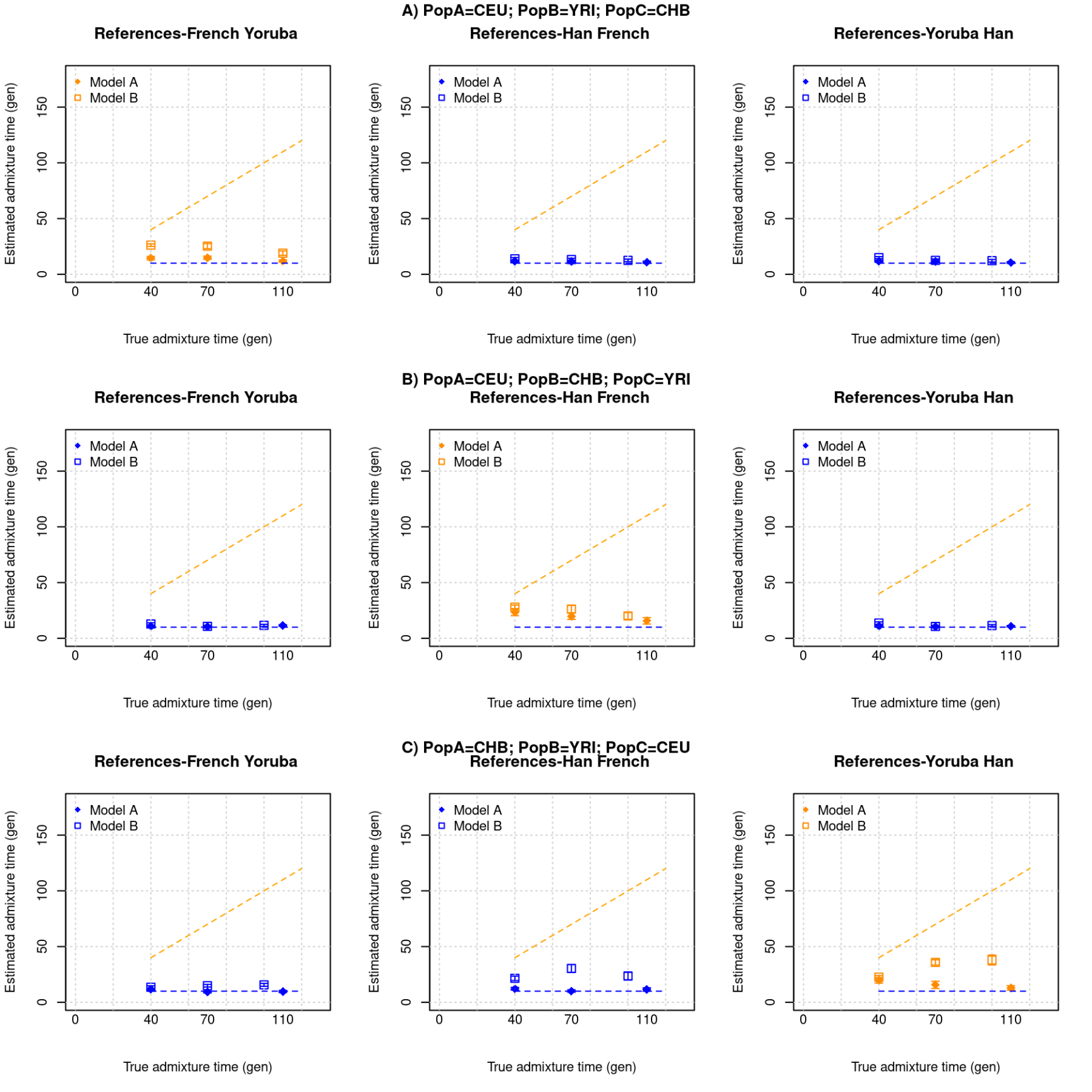


Target genome missing prop=60%

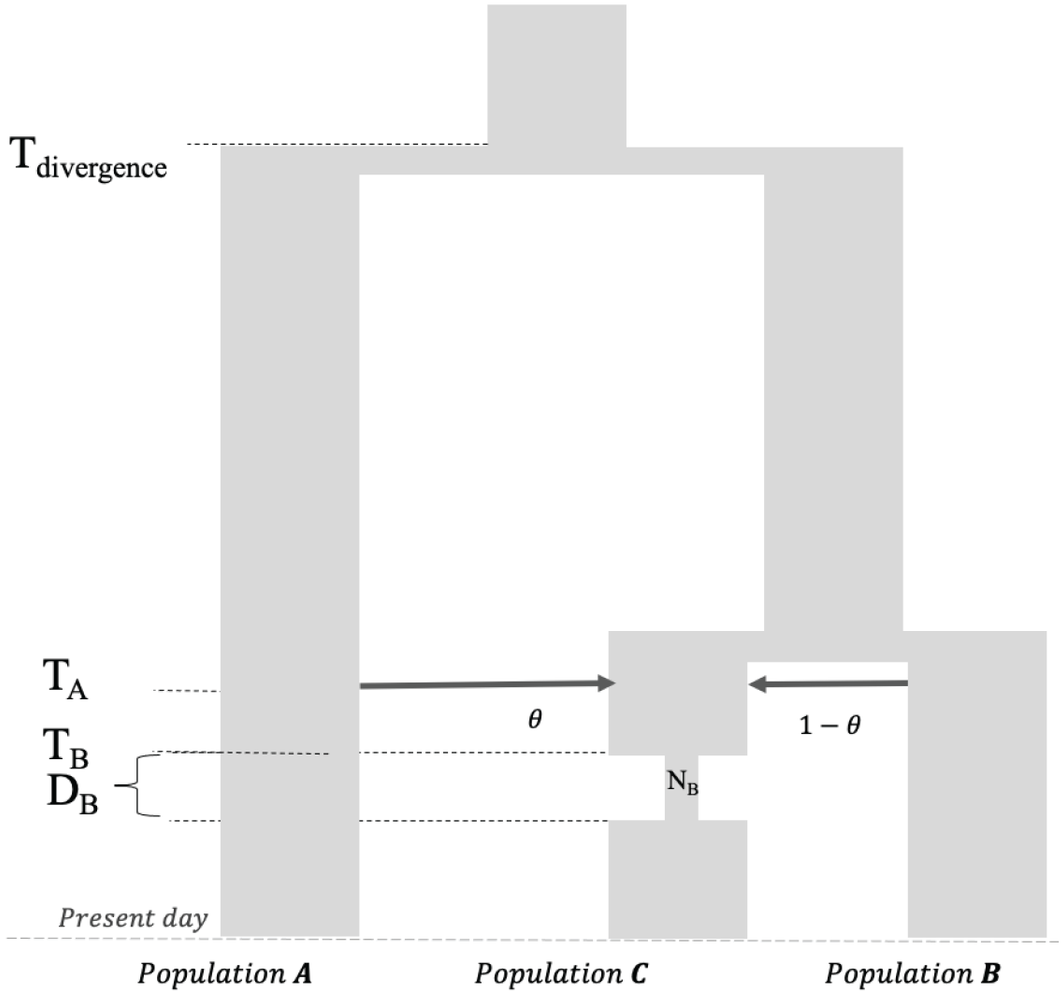



A
α_1
Admixture proportion from A
s'
α_2
Admixture proportion from B
B
Time of first pulse of admixture
t_2
s
C'
α_3
C
Admixture proportion from C
t_1
Time of second pulse of admixture
Target

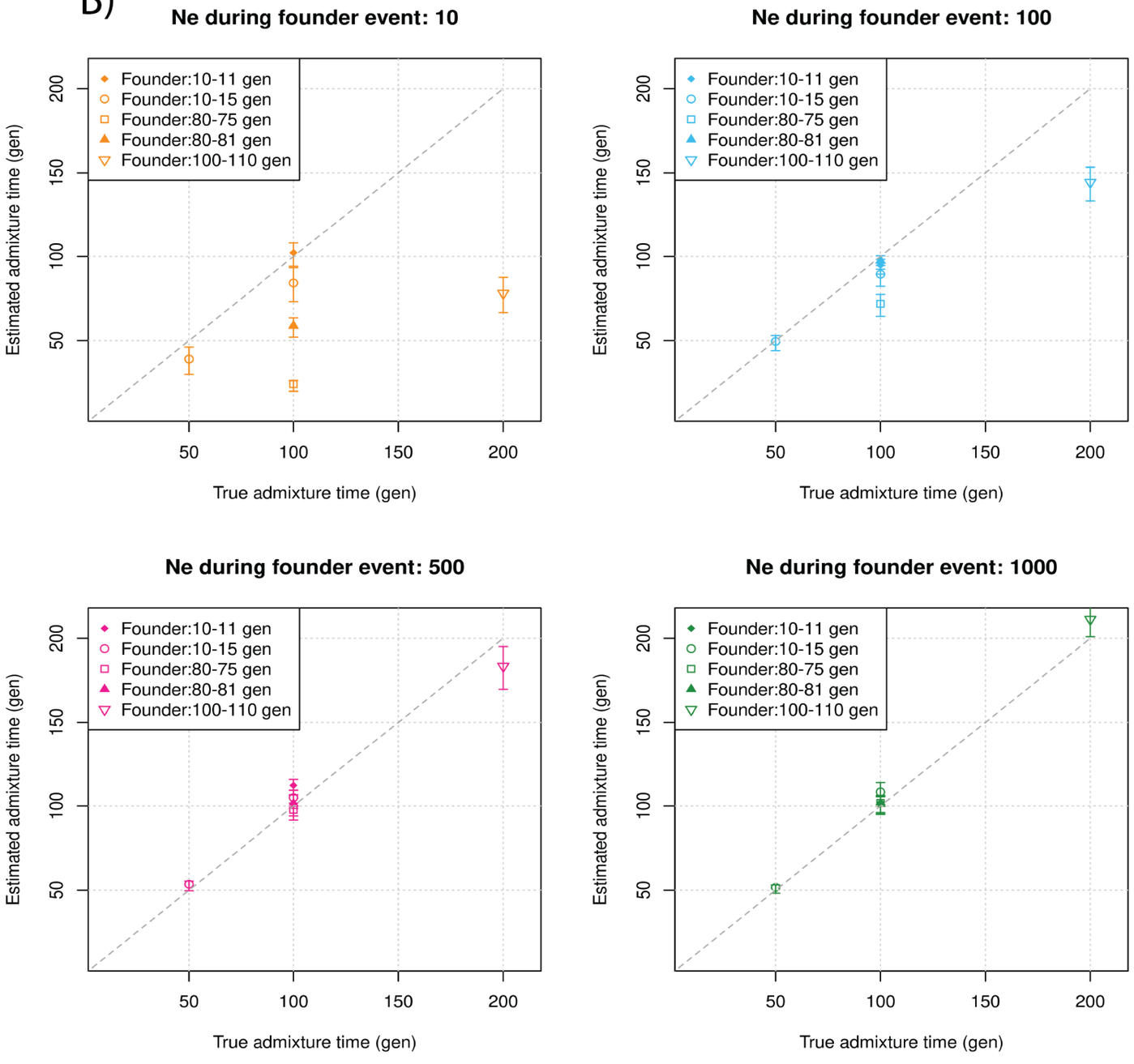

A) PopA=CEU; PopB=YRI; PopC=CHB
References-FrenchYoruba
References-HanFrench
References-YorubaHan
B) PopA=CEU; PopB=CHB; PopC=YRI
References-FrenchYoruba
References-HanFrench
References-YorubaHan
C) PopA=CHB; PopB=YRI; PopC=CEU
References-FrenchYoruba
References-HanFrench
References-YorubaHan
Estimated admixture time (gen)
True admixture time (gen)
0
50
100
150
0
40
70
110


A) PopA=CEU; PopB=YRI; PopC=CHB
References-French Yoruba
References-Han French
References-Yoruba Han
B) PopA=CEU; PopB=CHB; PopC=YRI
References-French Yoruba
References-Han French
References-Yoruba Han
C) PopA=CHB; PopB=YRI; PopC=CEU
References-French Yoruba
References-Han French
References-Yoruba Han
Model A
Model B
Estimated admixture time (gen)
True admixture time (gen)
0
50
100
150
0
40
70
110

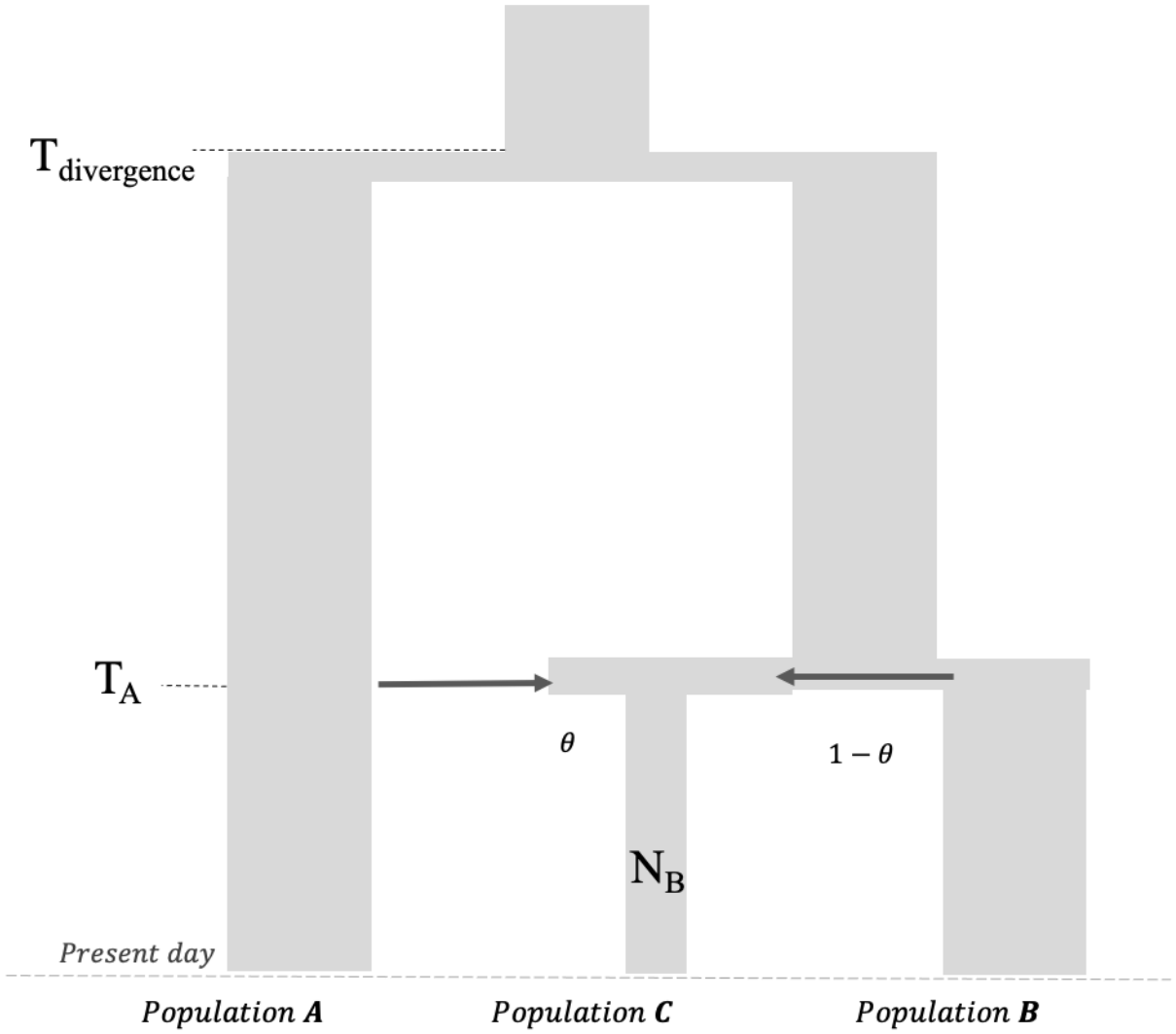
A) Scenario of admixture followed by founder event



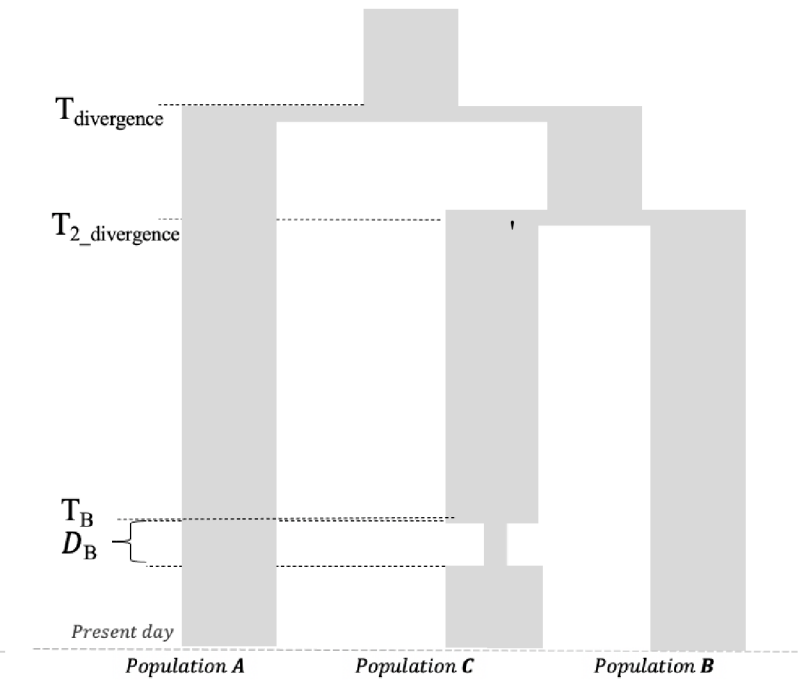
B)



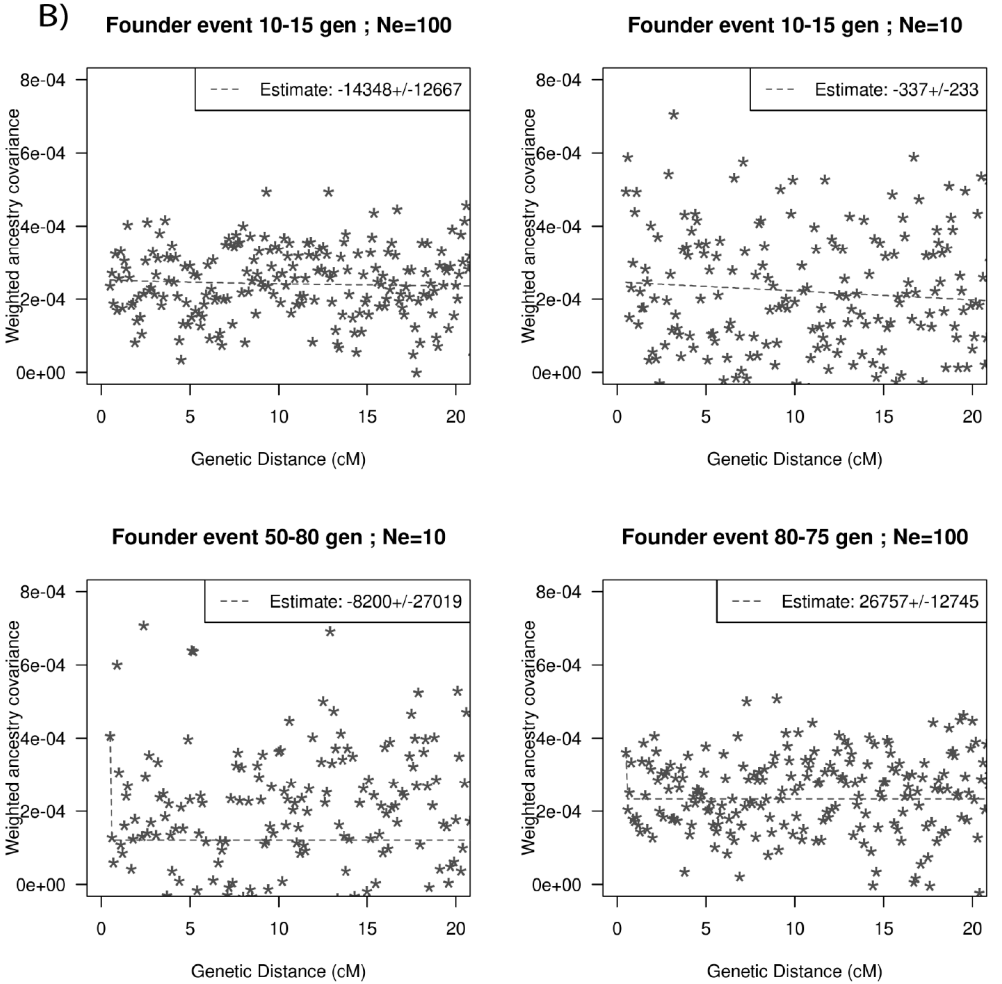
Scenario of drift post admixture



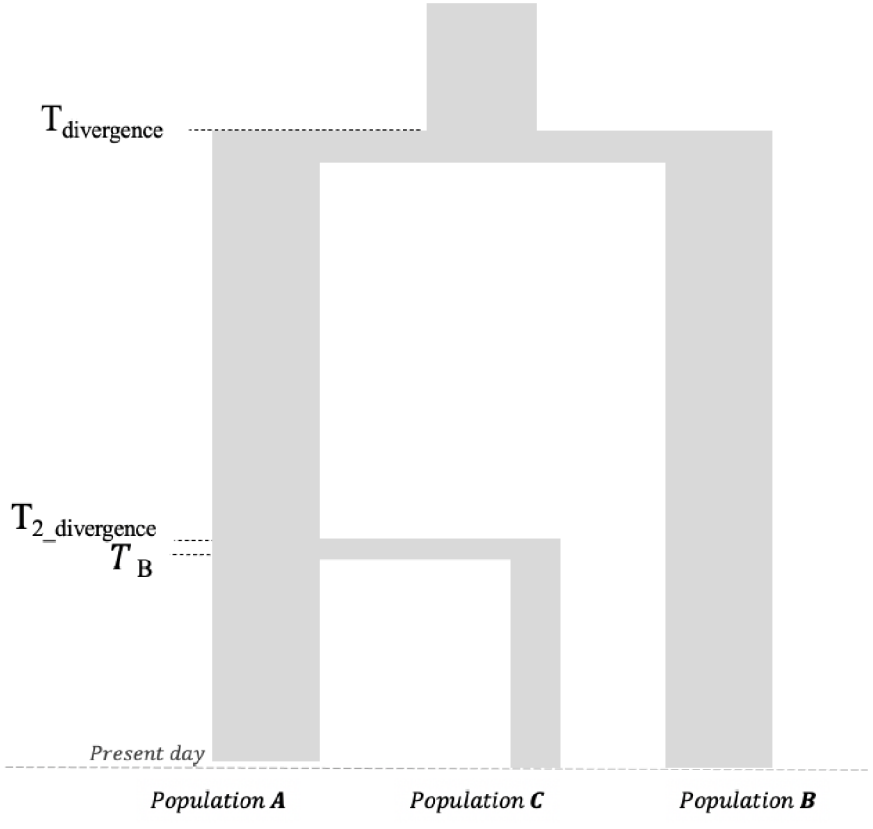
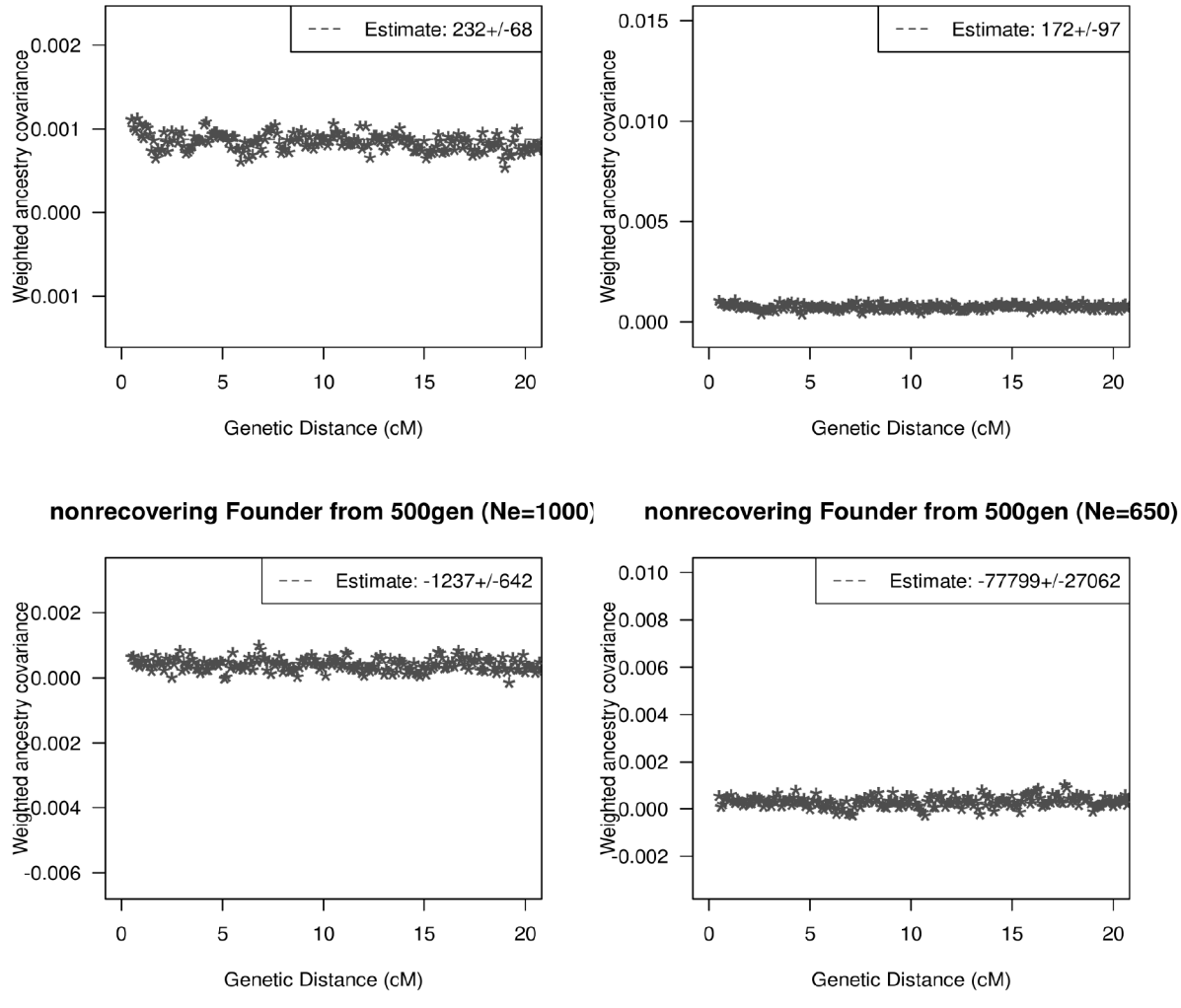
A) Scenario of no admixture and only founder event



B)



A) Scenario of no admixture with strong bottleneck non recovering



B)