Title

FST between archaic and present-day samples

Permalink

https://escholarship.org/uc/item/80g8q4xh

Journal

Heredity, 122(6)

ISSN

0018-067X

Authors

Ortega-Del Vecchyo, Diego
Slatkin, Montgomery

Publication Date

2019-06-01

DOI

10.1038/s41437-018-0169-8

Peer reviewed

Heredity (2019) 122:711–718
https://doi.org/10.1038/s41437-018-0169-8

ARTICLE

F_{ST} between archaic and present-day samples

Diego Ortega-Del Vecchyo[1,2] · **Montgomery Slatkin**[1]

Received: 2 November 2018 / Accepted: 7 November 2018 / Published online: 11 December 2018

Abstract

The increasing abundance of DNA sequences obtained from fossils calls for new population genetics theory that takes account of both the temporal and spatial separation of samples. Here, we exploit the relationship between Wright's F_{ST} and average coalescence times to develop an analytic theory describing how F_{ST} depends on both the distance and time separating pairs of sampled genomes. We apply this theory to several simple models of population history. If there is a time series of samples, partial population replacement creates a discontinuity in pairwise F_{ST} values. The magnitude of the discontinuity depends on the extent of replacement. In stepping-stone models, pairwise F_{ST} values between archaic and present-day samples reflect both the spatial and temporal separation. At long distances, an isolation by distance pattern dominates. At short distances, the time separation dominates. Analytic predictions fit patterns generated by simulations. We illustrate our results with applications to archaic samples from European human populations. We compare present-day samples with a pair of archaic samples taken before and after a replacement event.

Introduction

Genomic sequences obtained from fossils provide new information about the history of present-day species. Already, thousands of partial or complete genomic sequences have been obtained from modern humans and their extinct relatives, and DNA sequences from fossils of numerous other species have been obtained as well (Reich 2018).

Population genetics theory of ancient DNA (aDNA) has focused primarily on the time dimension. Several methods have been developed to test for natural selection and estimate selection coefficients in a time series of samples (Bollback et al. 2008; Malaspinas et al. 2012; Terhorst et al. 2015; Schraiber et al.). Much less effort has gone into incorporating the spatial dimension. The usual approach to analyzing spatially distributed aDNA is to use methods such as principal components analysis (PCA) and f-statistics that were developed for contemporaneous populations and ignore the ages of the fossils from which sequences are obtained. (Slatkin 2016)

There are three papers that have considered the spatial and temporal components of aDNA together. Skoglund et al. (2014) developed the coalescent theory of samples of different age and showed that PCA can reveal the time separation of spatially distributed samples. Duforet-Frebourg and Slatkin (2016) extended the classic Kimura-Weiss (1964) analysis of isolation by distance in a stepping-stone model to predict the decrease in identity by descent with increasing spatial and temporal separation of samples. Silva et al. (2017) carried out an extensive simulation study that showed the importance of considering geographic structure when testing for population continuity. Although all these papers provide some insight into the effects of isolation by distance and time, they did not fully explore the effect on measures of population differentiation.

In this paper, we examine the effects of isolation by distance and time on pairwise F_{ST} values. F_{ST} and related statistics have been widely used to characterize isolation by distance. Using the principles introduced by Skoglund et al. (2014) and Duforet-Frebourg and Slatkin (2016), we will show how pairwise F_{ST} between archaic and present-

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41437-018-0169-8) contains supplementary material, which is available to authorized users.

✉ Montgomery Slatkin
slatkin@berkeley.edu

[1] Department of Integrative Biology, University of California, Berkeley, CA 94720-3140, USA

[2] Present address: International Laboratory for Human Genome Research, National Autonomous University of Mexico, Santiago de Querétaro, Querétaro 76230, Mexico

day samples reflects both the distance and time separating samples in equilibrium populations and in non-equilibrium populations after a partial population replacement.

Pairwise F_{ST}

F_{ST} is useful for characterizing the extent of genetic difference between pairs of populations because it can be predicted analytically for a wide variety of models of population structure. If the per-locus mutation rate is small, F_{ST} computed for pairs of populations is dependent on the average coalescence time of two copies of a gene, one drawn from each population (Slatkin 1991). We consider two populations a and b. We will use the Hudson et al. (1992) estimator of F_{ST}, which Bhatia et al. (2013) have shown has somewhat better properties than either the Weir and Cockerham (1984) or Nei (1986) estimators when applied to genomic data. Hudson et al. (1992) estimated F_{ST} from the expression

$$F_{ST} = 1 - \frac{H_W}{H_B} \tag{1}$$

where H_W is the average number of differences between chromosomes sampled from the same population and H_B is the average number between different populations. That is, H_W is average of the expected per site heterozygosity within each population, which for two populations is $H_w = \left(H_w^{(a)} + H_w^{(b)}\right)/2$. H_B is the expected between-population heterozygosity.

Using the same method as in Slatkin (1991), we can find the expectations of H_W and H_B in terms of average branch lengths and the time between the samples when the per site mutation rate, μ, is small. For two lineages sampled at the same time, the average branch length is twice the average coalescence time. Therefore $E\left(H_W^{(a)}\right) \approx 2\mu\bar{t}_a$ and $E\left(H_W^{(b)}\right) \approx 2\mu\bar{t}_b$ where E denotes the expectation and $\bar{t}_a$ and $\bar{t}_b$ are the average coalescence times of two copies of the locus sampled from populations a and b. Therefore $E(H_W) = \left(E\left(H_W^{(a)}\right) + E\left(H_W^{(b)}\right)\right)/2 \approx \mu(\bar{t}_a + \bar{t}_b)$.

If samples a and b are from different times, then no coalescence is possible until the lineage from the more recent sample reaches the time horizon of the older sample (Skoglund et al., 2014). Assume a and b were sampled T_a and T_b generations in the past, with $T_a<T_b$. Then $E(H_B) \approx \mu(T_b - T_a + 2\bar{t}_{ab})$, where $\bar{t}_{ab}$is the average coalescence time of the a and b lineages starting at T_b. Therefore the expectation of the Hudson et al. estimator of F_{ST} is approximately

$$E[F_{ST}(a,b)] \approx 1 - \frac{\bar{t}_a + \bar{t}_b}{T_b - T_a + 2\bar{t}_{ab}}. \tag{2}$$

In many of the models, $\bar{t}_a$ and $\bar{t}_b$ are the same for all populations while $\bar{t}_{ab}$ depends on the spatial separation of a and b.

It will be convenient to describe patterns of pairwise F_{ST} in terms of the ratio

$$\eta_{ab} = \frac{F_{ST}(a,b)}{1 - F_{ST}(a,b)} = \frac{H_b}{H_w} - 1. \tag{3}$$

This ratio was introduced by Rousset (1997) and denoted by $\beta/(1 - \beta)$. The Users Manual of Arlequin (Excoffier and Lischer 2010) called this ratio "linearized F_{ST}". From Eq. (2), it follows that

$$E(\eta_{ab}) \approx \frac{T_b - T_a + 2\bar{t}_{ab} - (\bar{t}_a + \bar{t}_b)}{(\bar{t}_a + \bar{t}_b)}. \tag{4}$$

Thus, η_{ab} is proportional to the additional average coalescence time between gene copies drawn from different populations attributable to their separation in space and time.

Isolation by time

To illustrate our method, we consider first a single randomly mating diploid population. If the two samples come from the same population, the effect on F_{ST} is easy to calculate. First, assume the population has constant effective size. Standard coalescent theory tells us $\bar{t}_a = \bar{t}_b = \bar{t}_{ab} = 2N$. Therefore

$$E(\eta_{ab}) = \frac{T_b - T_a}{4N}. \tag{5}$$

We compared the analytical estimates of η_{ab} from Eq. (5) with simulations in Supplementary Fig. S2 (Skoglund et al. 2014).

If the population size is a function of time, the result is not quite as simple. Both $\bar{t}_a$ and $\bar{t}_b$ can be computed for an arbitrary demographic model from

$$\bar{t} = \int_T^\infty \exp\left(-\int_0^t \frac{dt'}{2N(t')}\right) dt \tag{6}$$

where $T = T_a$ or T_b, and $\bar{t}_{ab} = \bar{t}_b$. Therefore

$$\eta_{ab} = \frac{(T_b - T_a) + \bar{t}_b - \bar{t}_a}{\bar{t}_b + \bar{t}_a}. \tag{7}$$

We can also obtain analytic results if samples are taken from sister populations. For simplicity, assume all populations are of effective size N and let the time of population divergence be T_C. The two lineages cannot coalesce until

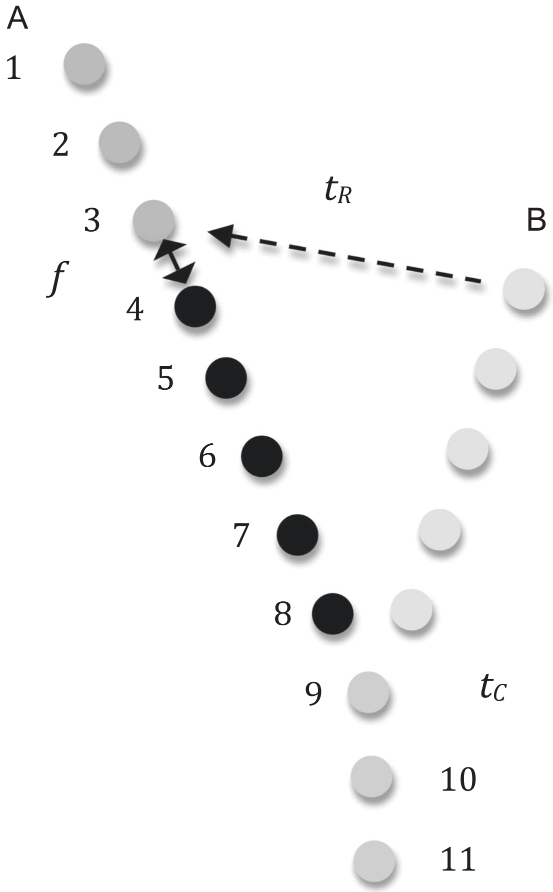

Fig. 1 Illustration of the model of partial population replacement. This model is a generalization of one used by Skoglund et al. (2014). At time t_C in the past, an ancestral population split into two descendent populations, distinguished as light gray and black. Archaic samples are available from the black population at different times in the past, indicated by the numbers. At time t_R in the past, a fraction 1 – f of the black population is replaced by the light gray population. The resulting population survives to the present day

they are in the ancestral population. Therefore, $\bar{t}_a = \bar{t}_b = 2N$ and $\bar{t}_{ab} = (T_C - T_b) + 2N$ (Skoglund et al. 2014) and

$$E(\eta_{ab}) = \frac{2T_C - T_a - T_b}{4N}. \tag{8}$$

Thus, $E(\eta_{ab})$ is proportional to the sum of the branch lengths in the population tree connecting the two samples. If population sizes depend on time in either or both branches, η_{ab} would reflect the coalescence probabilities in the two branches. We compared the analytical estimates of η_{ab} from Eq. (8) with simulations in Supplementary Figs. S1 and S3.

Partial population mixture

We consider a generalization of a model analyzed by Skoglund et al. (2014) and illustrated by Fig. 1. At time t_C in the past the ancestral population splits into two descendent populations, A and B. The numbers in Fig. 1 indicate the times of samples, with sample 1 being from the present day. At time t_R, a fraction 1–f of population A is replaced by individuals from B. The resulting population continues to the present. How this model is described depends on the magnitude of f. If $f = 0$, there was a complete population replacement. If f is small there was partial replacement. If f has an intermediate value, there was a population merger, and if f is nearly 1, there was admixture from B into A.

To illustrate the main result, we assume all populations are of the same size, N. Variable population size can be accounted for in the same way as for a single population. We assume that the present-day population, sample 1, is compared to an archaic sample taken at time τ in the past. The average coalescence time $\bar{t}$of two lineages depends on whether the sample is taken before or after t_R. If τ<t_R, then

$$\begin{aligned}\bar{t} = x\tilde{t} + (1-x)\Big[\Big(f^2 + (1-f)^2\Big)((t_R - \tau) + 2N) \\ +2f(1-f)((t_R - \tau) + (t_C - t_R)) + 2N\Big].\end{aligned} \tag{9}$$

where $x = 1 - e^{-(t_R-\tau)/(2N)}$ is the probability that the two lineages coalesce in the interval (τ, t_R) and $\tilde{t} = 2N - (t_R - \tau)e^{-(t_R-\tau)/(2N)}/\big(1 - e^{-(t_R-\tau)/(2N)}\big)$ is the average time to coalescence given that they coalesce in that time interval. The logic is that if they coalesce before t_R, the average coalescence time is $\tilde{t}$. If they do not coalesce and both lineages go into the same ancestral population, the expected coalescence time is $t_R - \tau + 2N$. If they do not go into the same population, then they have to wait an additional $t_C - t_R$ generations before they can coalesce. If τ>t_R then $\bar{t} = 2N$.

We also need the between-sample coalescence time, $\bar{t}_{ab}$, If τ<t_R, then

$$\bar{t}_{ab} = \bar{t} \tag{10}$$

where $\bar{t}$ is given by Eq. (9). Once the present-day lineage reaches time τ, the average coalescence time is the same as if the two lineages were sampled at the same time. If t_C>τ>t_R,

$$\bar{t}_{ab} = f(2N) + (1-f)(2N + t_C - \tau) \tag{11}$$

because with probability f the present day lineage remains in the same population and with probability 1–f it enters the other population. If τ>t_C, $\bar{t}_{ab} = 2N$.

Substituting these expressions into Eq. (4), we can predict η_{ab} as a function of τ and the other parameters. Some results are shown in Fig. 2. The solid lines show the analytic predictions. The dots show simulation results obtained from using the program scrm (Staab et al. 2015). In these and simulations described later in the paper, 100,000 replicates were run and results accumulated over all segregating sites.

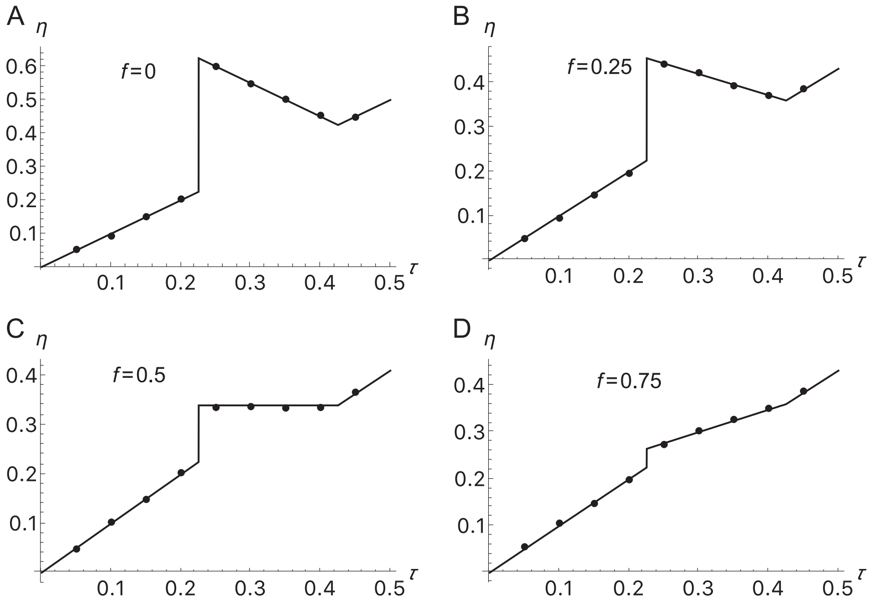

Fig. 2 Comparison of analytic and simulation results quantifying the extent of differentiation (η) between a present-day population (1 in Fig. 1) and an archaic population (2 to 10 in Fig. 1) sampled before or after a partial population replacement. The analytic results indicated by the line were obtained from Eqs. (9)–(11) in the text. The simulation results indicated by the dots were obtained using scrm (Staab et al. 2015), where one chromosome was sampled from the present and the other sampled τ generations in the past, where τ is measured in units of $4N$ generations. The partial replacement occurred at 0.225(x 4N) generations in the past

The mutation rate was chosen so that on average there were ten segregating sites per replicate. With this choice, there were no replicates with no segregating sites.

Isolation by distance and time

Duforet-Frebourg and Slatkin (2016) showed that the combined effects of isolation by distance and time in a stepping-stone model can be understood by considering the movement of lineages ancestral to the more recent sample during the time interval between the two samples. That movement is governed by dispersal patterns during the interval. Coalescence cannot occur until the time of the older sample. For simple models, analytic results can be obtained.

To illustrate, consider a one-dimensional stepping stone model with d demes arranged in a circle, and assume a migration rate m between adjacent demes. The average coalescence time of two genes sampled from i steps apart is

$$\bar{t}_i = 2Nd + \frac{i(d-i)}{4m} \tag{12}$$

where i is counted in a clockwise direction ($0 \le i \le d$–1) (Slatkin 1991). To see the effect of the difference in sampling time, assume one sample is from the present, ($T_a = 0$) and the other from T generations in the past ($T_b = T$). Between 0 and T, the present-day lineage undergoes a random walk on the circle. The probability that the lineage will be in deme j, given that it was initially in deme i is p_{ij}, the jth element of the vector, $p_{ij} = \left(\mathbf{M}^T e_i\right)_j$where e_i is a unit vector with 1 in position i and 0 otherwise and $\mathbf{M}$ is a circular matrix, which has non-zero elements$M_{ii} = 1 - 2m$ and $M_{i,i+1} = M_{i,i-1} = M_{1,d} = M_{d,1} = m$. $\mathbf{M}^T$ denotes the Tth power of $\mathbf{M}$.

Therefore$\bar{t}_a = \bar{t}_b = 2Nd$ and

$$\bar{t}_{ab} = 2Nd + \sum_{j=0}^{d-1} p_{ij}\left(\frac{j(d-j)}{4m}\right) \tag{13}$$

from which we can compute η_i to be

$$\eta_i = \frac{T}{8Nd} + \frac{1}{8Nmd}\sum_{j=0}^{d-1} p_{ij} j(d-j) \tag{14}$$

Figure 3 presents results for η_i using an archaic sample drawn from deme 0 at time $T = 40N$ generations ago and a present-day sample drawn from another deme. Shown for comparison is the equilibrium IBD pattern for contemporaneous samples ($T = 0$). As the age of the archaic sample increases, η_i increases in the neighborhood of the sampled deme. There are two components to this increase. One is the time separation of the samples, represented by the first term in Eq. (14). The other is the averaging of the equilibrium pattern because of the dispersal of the present-day lineage between 0 and T, which is represented by the second term in Eq. (14). Because η_i=0 for two samples from the same population at the same time, the averaging is

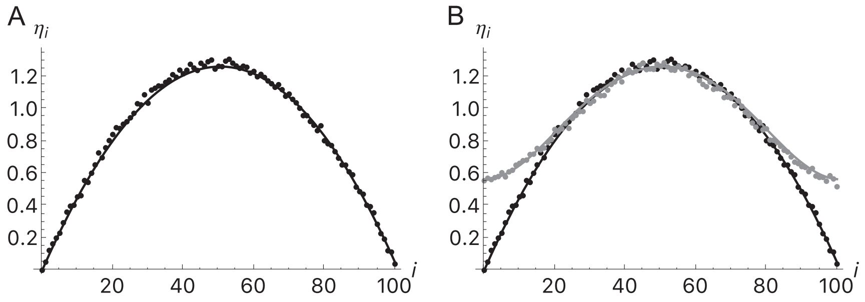

Fig. 3 Comparison of analytic and simulation results quantifying the extent of differentiation (η_i) between populations i steps apart in a circular stepping-stone population sampled at the same time and with a time-separation of $40N$ generations. The analytic results, shown by the solid lines, were obtained using Eq. (14) in the text. The simulation results, shown by the dots, were obtained using scrm (Staab et al. 2015). The model was of a circle of 101 demes with migration rate $4Nm = 10$ between adjacent demes. The black dots and line show the equilibrium pattern of isolation by distance between contemporaneous populations. The gray dots and line show the pattern for a sample taken $40N$ generations in the past from deme 0 against contemporaneous samples from other demes. Simulation results are averages of 100,000 replicates

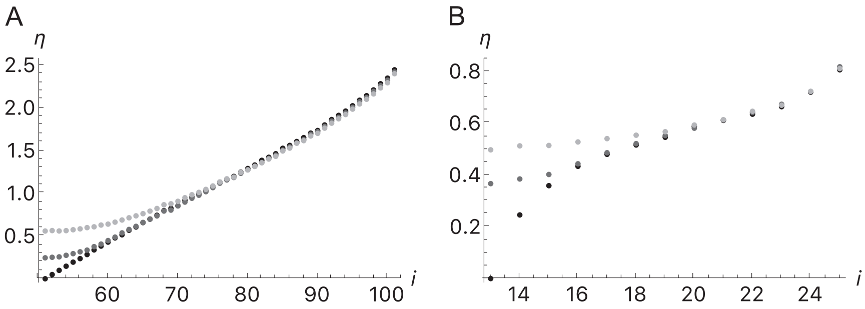

Fig. 4 Isolation by distance patterns in one and two dimensional stepping stone models with symmetric migration. Pairwise values of η were estimated from simulation results obtained with scrm (Staab et al. 2015). Each point is based on 100,000 replicates. **a** 101×1 stepping stone model with $4Nm = 1$ between adjacent demes. The middle population (population 51) was sampled at the present ($t = 0$, black dots), and two times in the past, $t = 8N$ generations (dark gray dots) and $t = 40N$ generations in the past (light gray dots) and compared to each of the present-day populations. The results are symmetric around the middle population. **b** 25×25 stepping stone model with $4Nm = 1$ between adjacent demes. The middle population in the middle row (population 13,13) was sampled at the present ($t = 0$, black dots), and two times in the past, $t = 8N$ generations (dark gray dots) and $t = 40N$ generations in the past (light gray dots) and compared to each of the present-day populations in the middle row (population 13,13 to population 25,13). The results are symmetric around the middle population. Note the difference in scale of the vertical axes in **a** and **b**

over populations for which η_i is positive. That results in a positive contribution. Both terms contribute and their relative magnitudes depend on the parameter values. Note that in this model, as in many models of a subdivided population, the average within-deme coalescence time is twice the total number of individuals in the population, independently of the migration pattern (Strobeck 1987).

Similar results are obtained for one- and two-dimensional symmetric stepping stone models. Figure 4 shows typical examples.

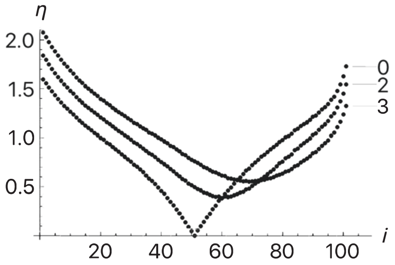

Fig. 5 Simulation results for a 101×1 stepping stone model with asymmetric migration. The model is the same as in **a** of Fig. 4 but with migration rate to the right of $4Nm = 11$ and to the left of $4Nm = 9$. We show values of η for archaic samples taken at $t = 0$, 2 and 3, where time is measured in units of $4N$ generations

Models of symmetric dispersal are a staple of population genetics theory because of their mathematical simplicity. There is no reason to suppose that dispersal in natural populations is actually symmetric either in each generation or when averaged over many generations. Comparison with archaic samples can reveal slight asymmetry in dispersal that may not be apparent when comparing only present-day samples. Figure 5 provides one example. The population is in a 101×1 linear stepping stone model, as in Fig. 4a. The only difference is that $4Nm$ to the right and left are 11 and 9 respectively. As shown in Fig. 5, this difference is not obvious in the isolation by distance pattern of present day populations, but is when a few archaic samples are included.

Range expansion with partial replacement

Range expansions have happened many times in the history of humans and other species. Range expansions create unusual patterns in allele frequencies because of continued founder effects, a phenomenon called "gene surfing" (Excoffier and Ray 2008). Several ways have been proposed for detecting the genetic signatures of range expansions including testing for clines in heterozygosity (Ramachandran et al. 2005) and computing a directionality index (Peter and Slatkin 2013). Range expansion may occur into an area previous unoccupied or an area occupied by another population. In the latter case, admixture between the invading and resident populations will take place (Currat et al. 2008).

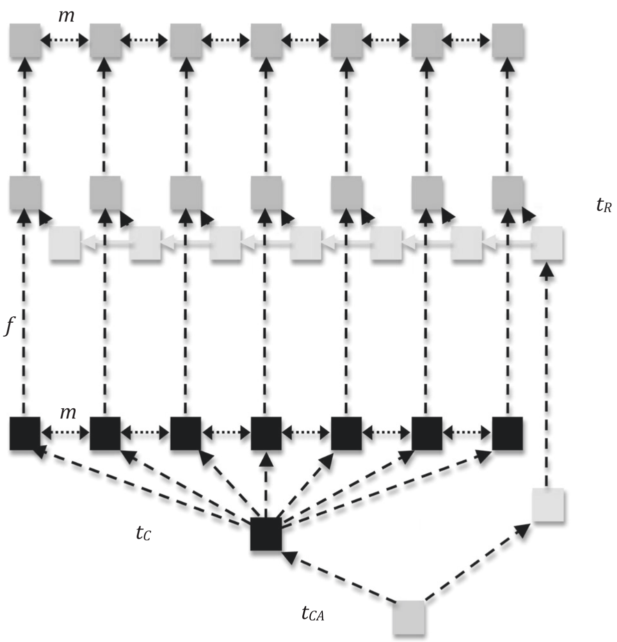

Fig. 6 Illustration of the model of partial population replacement in a stepping stone framework. At time t_{CA} in the past, the ancestral population split into two descendent populations. At time t_C, the one population gave rise to several populations, which then exchange migrants at rate m in a stepping stone model. At time t_R, the descendant of the other population becomes the source for a range expansion. After each colonization event, the expanding population mixes with a resident population to produce a descendent population. Each descendent population is made up of a fraction f of the resident population and a fraction $1-f$ of the expanding population. The resulting populations exchange migrants symmetrically with each neighboring population at rate m until the present day

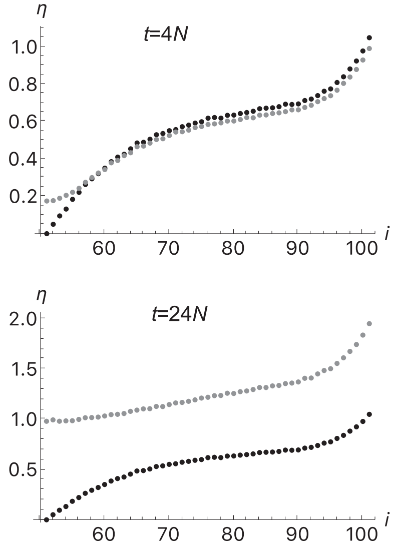

Fig. 7 Representative patterns of isolation by distance seen when archaic samples are taken before and after a partial replacement. The model is as illustrated in Fig. 6. There were 101 populations in a stepping stone configuration with migration at rate $4Nm = 10$ between adjacent demes. At time $t = 16N$ in the past, there was a range expansion beginning with population 1. During each colonization event, the population size was reduced by a factor of 0.01 for $0.002N$ generations. As each colonizing population came into contact with a resident population, the two populations contributed equally to the descendent population ($f = 0.5$). The black dots indicate η_{ab} for the middle present-day population compared to each of the other present-day populations. The gray dots indicate the values for the middle archaic population and each of the present-day populations. The two archaic samples were taken $t = 4N$ and $t = 24N$ generations before the present. The graphs were obtained using scrm. Each point is the average of 100,000 replicate simulations

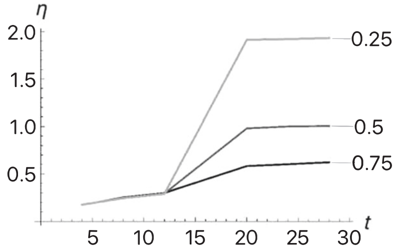

Fig. 8 Comparison of increases in η for different extents of admixture with resident populations. The parameters were the same as in Fig. 7 except that the results for $f = 0.25$, 0.5 and 0.75 are shown. The archaic samples were taken $4N$, $8N$, $12N$, $20N$, $24N$ and $28N$ generations in the past

To determine the effects of range expansion on F_{ST} taken from archaic samples, we simulated a model in which there is a partial replacement of a resident population by an expanding population. Both before and after the range expansion, there is a stepping stone population structure. The model is illustrated in Fig. 6. As in Fig. 1, f is the fraction of each population that is descended from the resident population and $1-f$ is the fraction that is descended

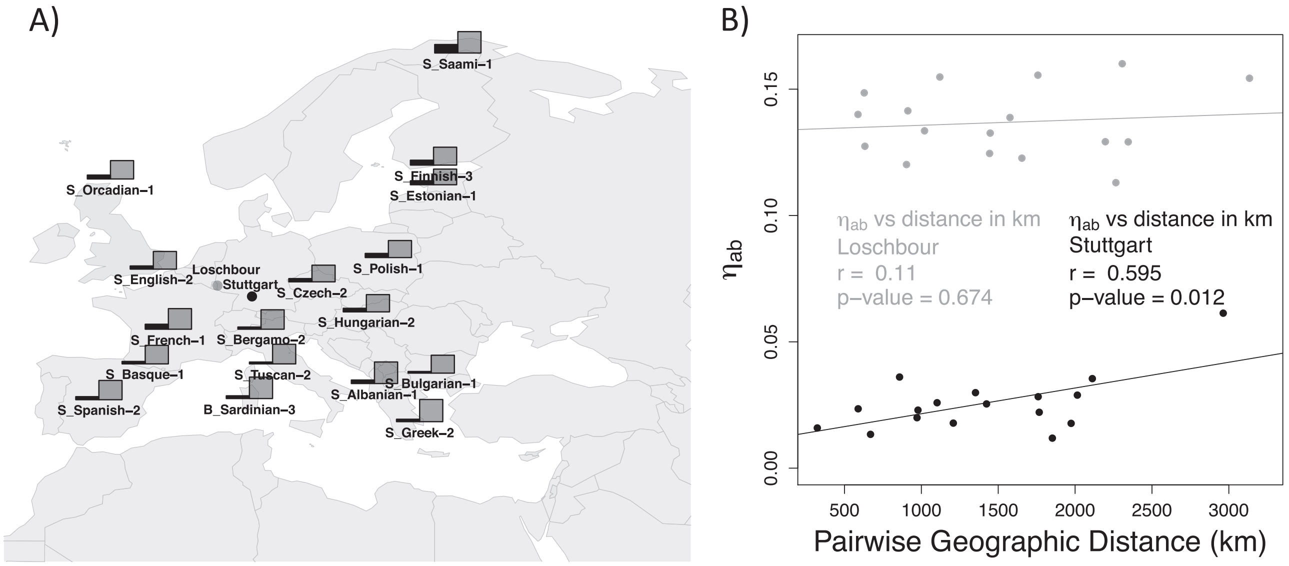

Fig. 9 **a** Comparison of pairwise η_{ab} values computed for two focal samples. All data were taken from the Simons Genome Diversity Project dataset (Mallick et al. 2016), which also contains two ancient human genomes, Stuttgart and Loschbour (Lazaridis et al. 2014). We compare the results for the two focal samples. The black dot indicates the location of the Stuttgart Neolithic farmer skeleton (~7000-years-old) and the gray dot points the location of the Loschbour Neolithic hunter-gatherer skeleton (~8000-years-old). The histogram bars indicate the value of η_{ab} computed between the focal sample of the same color and a present-day sample at each location. **b** Pairwise η_{ab} values of present-day samples and the two focal ancient samples vs. the pairwise geographical distance between the sampling location of the present-day and ancient samples. The correlation coefficient r and the p-value of the null hypothesis that the slope obtained from the linear regression line has a value equal to zero. The p-values were obtained using an F-test comparing the linear model with a non-zero slope to a model with a zero slope

from the invading population in each location. Some simulation results are shown in Figs. 7 and 8. The patterns in pairwise η values are a combination of those seen with partial population replacement and isolation by distance in an equilibrium population. The pattern of isolation by distance and the relationship between archaic and present-day samples are preserved if there is partial replacement (Fig. 7). And the abrupt change created by a partial replacement is evident when comparing archaic samples before and after the replacement event (Fig. 8).

Examples

To illustrate patterns seen in data from human populations, we reanalyzed the data of the Simons Genome Diversity Project (Mallick et al. 2016) and two ancient human genomes (Lazaridis et al. 2014). The ancient genomes come from a Neolithic farmer (the Stuttgart sample, ~7000 years before present) and a Neolithic hunter-gatherer (the Loschbour sample, ~8000ybp). Figure 9 shows the results. The black histograms in Fig. 9a show pairwise values of η_{ab} computed for Stuttgart and several present-day European samples. The gray histograms show η_{ab} computed for Loschbour and the same present-day samples. The results are consistent with two theoretical expectations: The older sample, Loschbour, has larger η_{ab} values. Additionally, the results are consistent with a smaller average ancestry in present-day Europeans coming from hunter-gatherers (Haak et al. 2015). This is in agreement with our partial population replacement model, where comparisons of present-day individuals with ancient samples coming from a population that has been mostly replaced (f close to 0) tend to have larger η_{ab} when the ancient sample was sampled from before the time of replacement, t_R, and after the present-day and ancient populations coalesce to an ancestral population, t_c, (see Fig. 2). The ancient samples we used, Loschbour and Stuttgart, are samples taken from near the time of replacement.

We found a significant positive correlation between the pairwise geographical distance and the pairwise η_{ab} values of present-day samples and the Stuttgart sample (Fig. 9b). This observation is consistent with a pattern of isolation by distance in Neolithic farmers that is retained in present-day populations. In contrast, there is no significant correlation when we do the same analysis with the Loschbour sample. This observation suggests that the replacement of hunter-gatherer populations by early farmers erased any signal of isolation-by-distance in the hunter-gatherer populations, if one was present.

Discussion and conclusion

We present the basic theory of Wright's F_{ST} between samples taken at different times and places. As Skoglund et al. (2014) note, there is an important difference between pairwise F_{ST} and the principal components analysis (PCA). Pairwise F_{ST} values do not depend on what other samples are included in the analysis while principal components do. Although both representations of data reflect pairwise coalescence times (Slatkin 1991; McVean 2009), principal components depend on pairwise coalescence times for a particular pair of samples relative to other pairs of samples. The two ways of looking at data are both useful. Using pairwise F_{ST} values allows a more direct tie to the underlying coalescent process and allows comparison with analytic theory.

The theory we have developed shows that F_{ST} values for pairs of samples of different age depend on numerous parameters in addition to the time separation of the samples. For that reason, pairwise F_{ST} values alone are not suitable for inferring demographic parameters. Both the results presented here and the simulation study of Silva et al. (2017) show that patterns of population differentiation depend in a complex way on the time separation of samples, patterns of dispersal and the extent of population replacement. However, pairwise F_{ST} values could serve a key statistic in an approximate Bayesian computation analysis (Bertorelle et al. 2010) because they directly reflect pairwise coalescence times.

Acknowledgements This research was supported in part by a US NIH grant, R01-GM40282 to MS. We thank L. Excoffier for helpful comments on an earlier version of this paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. Mol Ecol 19:2609–2625

Bhatia G, Patterson N, Sankararaman S, Price AL (2013) Estimating and interpreting FST: The impact of rare variants. Genome Res 23:1514–1521

Bollback JP, York TL, Nielsen R (2008) Estimation of 2Nes from temporal allele frequency data. Genetics 179:497–502

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. Evolution 62:1908–1920

Duforet-Frebourg N, Slatkin M (2016) Isolation-by-distance-and-time in a stepping-stone model. Theor Popul Biol 108:24–35

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 23:347–351

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K et al. (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522:207–211

Hudson RR, Slatkin M, Maddison WP (1992) Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589

Kimura M, Weiss GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics 49:561–576

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M et al. (2014) Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513:409–413

Malaspinas A-S, Malaspinas O, Evans SN, Slatkin M (2012) Estimating allele age and selection coefficient from time-serial data. Genetics 192:599–607

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature 538:201–206

McVean G (2009) A genealogical interpretation of principal components analysis. PLoS Genet 5:e1000686

Nei M (1986) Definition and estimation of fixation indices. Evolution 40:643–645

Peter BM, Slatkin M (2013) Detecting range expansions from genetic data. Evolution 67:3274–3289

Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc Natl Acad Sci 102:15942–15947

Reich D (2018) Who we are and how we got here: Ancient DNA and the new science of the human past. Pantheon Books, New York

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228

Schraiber JG, Evans SN, Slatkin M (2016) Bayesian inference of natural selection from allele frequency time series. Genetics 203:493–511

Silva NM, Rio J, Currat M (2017) Investigating population continuity with ancient DNA under a spatially explicit simulation framework. BMC Genet 18:114

Skoglund P, Sjödin P, Skoglund T, Lascoux M, Jakobsson M (2014) Investigating population history using temporal genetic differentiation. Mol Biol Evol 31:2516–2527

Slatkin M (1991) Inbreeding coefficients and coalescence times. Genet Res 58:167–176

Slatkin M (2016) Statistical methods for analyzing ancient DNA from hominins. Curr Opin Genet Dev 41:72–76

Staab PR, Zhu S, Metzler D, Lunter G (2015) scrm: efficiently simulating long sequences using the approximated coalescent with recombination. Bioinformatics 31:1680–1682

Strobeck C (1987) Average number of nucleotide differences in a sample from a single subpopulation: A test for population subdivision. Genetics 117:149–154

Terhorst J, Schlötterer C, Song YS (2015) Multi-locus analysis of genomic time series data from experimental evolution. PLoS Genet 11:e1005069

Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population structure. Evolution 38:1358–1370