AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 152:11–18 (2013)

Genetic Evidence Does Not Support an Etruscan Origin in Anatolia

Francesca Tassi,[1] Silvia Ghirotto,[1] David Caramelli,[2] and Guido Barbujani [1]*

[1]*Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy*
[2]*Department of Evolutionary Biology, University of Firenze, Ferrara, Italy*

KEY WORDS ancient DNA; mitochondrial DNA; coalescent simulations; approximate Bayesian computation

ABSTRACT The debate on the origins of Etruscans, documented in central Italy between the eighth century BC and the first century AD, dates back to antiquity. Herodotus described them as a group of immigrants from Lydia, in Western Anatolia, whereas for Dionysius of Halicarnassus they were an indigenous population. Dionysius' view is shared by most modern archeologists, but the observation of similarities between the (modern) mitochondrial DNAs (mtDNAs) of Turks and Tuscans was interpreted as supporting an Anatolian origin of the Etruscans. However, ancient DNA evidence shows that only some isolates, and not the bulk of the modern Tuscan population, are genetically related to the Etruscans. In this study, we tested alternative models of Etruscan origins by Approximate Bayesian Computation methods, comparing levels of genetic diversity in the mtDNAs of modern and ancient populations with those obtained by millions of computer simulations. The results show that the observed genetic similarities between modern Tuscans and Anatolians cannot be attributed to an immigration wave from the East leading to the onset of the Etruscan culture in Italy. Genetic links between Tuscany and Anatolia do exist, but date back to a remote stage of prehistory, possibly but not necessarily to the spread of farmers during the Neolithic period. Am J Phys Anthropol 152:11-18, 2013.

The Etruscan civilization is documented in Etruria, roughly corresponding to current Tuscany, starting from the eighth century BC, and is defined by a material culture, by a non-Indo-European language and by an alphabet derived from the Greek alphabet. Availability of copper and iron and ability in seafaring were the main factors leading to an Etruscan expansion over much of Central Italy in the sixty and fifth centuries BC. Later, military defeats and the Roman expansion caused a decline of the Etruscans' political influence. From the first century BC, the Etruscan language disappeared from the archeological record (Barker and Rasmussen, 1998).

Questions about the Etruscans' origins date back to antiquity and are still open. In the fifth century BC, Herodotus described them as a group emigrating from Lydia, in Western Anatolia; by contrast, Dionysius of Halicarnassus regarded them as an indigenous Italic population, and this view is shared by most modern archeologists. The first genetic studies assumed that current inhabitants of Tuscany are the direct mitochondrial descendants of the Etruscans, and their results suggested an evolutionary link with Anatolia that would support Herodotus' view (Achilli et al., 2007; Brisighelli et al., 2009). However, analyses of mitochondrial DNA (mtDNA) sequences from bones excavated in Etruscan necropolis (Vernesi et al., 2004) raised questions about the significance of the similarities observed between modern populations. Indeed, based on ancient DNA data, the Etruscans appeared to represent a single biological population, connected by genetic links across its territory, but showed a limited genetic resemblance with modern people of the same area (Vernesi et al., 2004). Explicit tests comparing mtDNAs of ancient (i.e., Etruscan) and modern inhabitants of Tuscany ruled out the hypothesis that the former might be the latter's direct ancestors (Belle et al., 2006; Guimaraes et al., 2009). Among the possible explanations for these results are the presence of massive errors in the Etruscan sequences (as suggested by Bandelt and Kivisild, 2006), a complete population extinction, and a persistence of the Etruscans' genetic heritage only in isolated localities, whereas most contemporary Tuscans would be descended from different mitochondrial ancestors.

Two recent studies contributed to clarifying this complex picture. First, Bayesian analysis of patterns of mutation showed no evidence of systematic errors in the ancient Etruscan sequences, which might explain the observed differences between modern and ancient inhabitants of Etruria (Mateiu and Rannala, 2008). Errors are always possible and sometimes hard to identify in ancient DNA analyses, but as far as one can test, Vernesi et al.'s (2004) Etruscan sequences comply with the highest quality standards. Second, in a study including a new set of ancient DNA sequences, in which alternative demographic models were explicitly tested by Approximate Bayesian Computation (ABC), the results were compatible with a genealogical continuity between

Grant sponsor: University of Ferrara; Italian Ministry for University and Research (MIUR), PRIN 2012 funds.

*Correspondence to: G. Barbujani, Department of Life Sciences and Biotechnology, University of Ferrara, via Borsari 46, 44121 Ferrara, Italy. E-mail: g.barbujani@unife.it

Received 11 January 2013; accepted 20 May 2013

DOI: 10.1002/ajpa.22319
Published online 30 July 2013 in Wiley Online Library (wileyonlinelibrary.com).

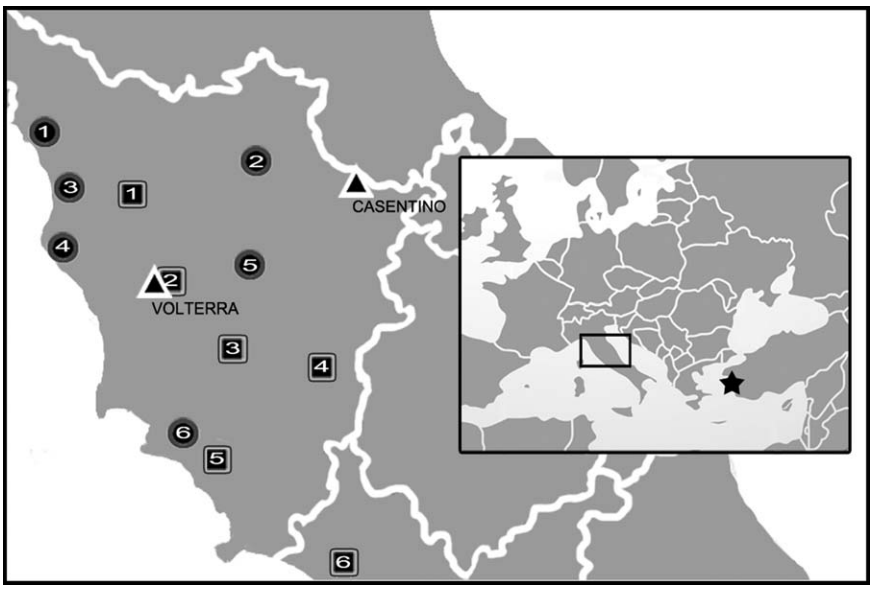


Fig. 1. Geographic location of the samples considered in the ABC analysis. Triangles, contemporary Tuscans (n = 236); Circles, Medieval Tuscans: 1. Massa Carrara (n = 3), 2. Florence (n = 10), 3. Pisa (n = 6), 4. Livorno (n = 3), 5. Siena (n = 4); 6. Grosseto (n = 1); Squares, Etruscans: 1. Castelfranco di Sotto (n = 1); 2. Volterra (n = 3); 3. Casenovole (n = 10); 4. Castelluccio di Pienza (n = 1); 5. Magliano/Marsiliana (n = 6); 6. Tarquinia (n = 9); Star, Turks (n = 35).

the Etruscans and two Tuscan isolates (Volterra and Casentino) that had not been previously compared with them. By contrast, another population of the former Etruscan homeland, Murlo, and a forensic sample from the main city in the area, Florence, showed no special relationships with the Etruscans (Ghirotto et al., 2013). The former finding means that populations separated by short distances may differ in their genetic relationships with ancient populations (as already seen in Sardinia; Ghirotto et al., 2010), and the latter confirms that the Etruscans cannot be regarded as the global ancestors of the people now living in what once was their territory.

In this article, we used all the available ancient mtDNA samples from classical Etruria, the inferential power given by the ABC methods, and the information on the genealogical relationships between the Etruscans and the communities of Volterra and Casentino (Ghirotto et al., 2013) to further investigate the Etruscans' biological origins. Because previous inhabitants of Etruria, associated with the Villanovian culture, cremated their dead, empirical genetic comparisons going further back in time are unfeasible. We then compared the observed genetic data with the results of millions of simulations of modern and ancient mtDNAs, generated under demographic models differing for the homelands of Etruscan people, namely, Western Anatolia or Central Italy. This way, we could test whether or not the genetic links between modern Anatolians and Tuscans may have been established through a process of gene flow occurring approximately between the tenth and eight centuries BC, and thus possibly associated with the onset of the Etruscan civilization in Italy.

MATERIALS AND METHODS

Genetic data

Historical Etruria comprises much of current Tuscany and the northernmost region of current Latium (Fig. 1). Therefore, in this study, we excluded previously published specimens coming from the regions of Etruscan expansion to the North (Adria) and to the South (Capua), because (a) we had no idea of the levels of admixture with non-Etruscan people in these localities and (b) there was no reason to assume that these populations could have contributed to the ancestry of modern Tuscans.

We analyzed 360 bp [positions 16024–16383 of the Cambridge reference sequence (CRS) (Andrews et al., 1999)] within the HVRI (hypervariable I) region of mtDNA. In all statistical analyses, we replaced the nucleotides occupying position 16180–16188 and 16190–16193 with the nucleotides in the CRS, to avoid the stretch of adenines and cytosines known to result in apparent length polymorphism of the mtDNA sequence (Bendall and Sykes, 1995; Bandelt and Kivisild, 2006).

We considered samples of three main historical periods: the Etruscans (with specimens dated around 2,500 years ago, on average), the medieval Tuscans (dated around 900 years ago), and modern subjects. The Etruscan sample is composed of 30 sequences from different necropolis (Vernesi et al., 2004; Ghirotto et al., 2013). The Medieval sample comprises 27 sequences collected in various Tuscan localities (Guimaraes et al., 2009). The modern sample comprises the following: (a) two Tuscan populations [Casentino, 122 sequences and Volterra, 114 sequences (Achilli et al., 2007)] for which we previously demonstrated a high level of genealogical continuity since Etruscan times (Ghirotto et al., 2013) and (b) a population from Western Anatolia [35 sequences (Di Benedetto et al., 2001)], representing the putative Etruscans' homeland according to Herodotus.

Summary statistics

In this study, statistics summarizing genetic diversity were calculated by using Arlequin ver. 3.5.1. (Excoffier and Lischer, 2010). The within-population diversity of each sample is described by (i) sample size, (ii) number

TABLE 1. Statistics summarizing (A) intra- and (B) interpopulation genetic diversity

A					
	Etruscans	Medieval	Casentino	Volterra	Turks
No. of sequences	30	27	122	114	35
No. of distinct Haplotypes	21	14	72	57	29
Mean pairwise difference	2.966	1.972	4.105	3.850	4.689
Haplotype diversity	0.943	0.860	0.976	0.955	0.965
Segregating sites	24	14	62	58	43
B					
Fst					
Etruscans	0.000	0.015	0.020	0.012	0.033
Medieval	0.015	0.000	0.020	0.013	0.045
Allele sharing					
Etruscans	1.000	0.238	0.333	0.238	0.143
Medieval	0.357	1.000	0.500	0.429	0.286

These values were used in the ABC analysis.

of different haplotypes, (iii) mean pairwise difference, (iv) haplotype diversity, and (v) number of segregating sites (Table 1a). In addition, we quantified the relationships between Tuscans and Anatolians calculating, for each comparison, Hudson's F_{ST} (Hudson et al., 1992) an index of genetic diversity particularly suitable for haploid sections of the genome; and a measure of allele sharing, defined as the number of haplotypes of the Anatolian sample also present in each Tuscan sample, scaled by the total number of haplotypes in the latter. In this way, we divided the number of shared haplotypes, respectively, by 21, 14, 72, and 57, namely, the number of haplotypes in the Etruscan, Medieval, Casentino, and Volterra's samples (Table 1b).

Tested demographic scenarios and priors

We designed two models, both assuming a common origin of populations of the Eastern and Northern Mediterranean shores during the Paleolithic dispersal of anatomically modern humans from Africa (see, e.g., Otte, 2000), but differing in the timing of a migration event. We did that by simulating an ancestral population split 1,000 generations ago, roughly equivalent to 25,000 years ago if one assumes, according to Fenner (2005), an average generation time of 25 years. In time, the first resulting lineage gave rise to the Etruscans (100 generations or 2,500 years ago) who, in turn, are the ancestors of the medieval Tuscans (36 generations or 900 years ago) and of current inhabitants of Casentino and Volterra (placed 0 generations ago); the second lineage gave rise to the Western Anatolian population. Both lineages grew exponentially in size after the split. Under Model A, migration from the East, followed by admixture, took place in a relatively remote past [between 6,000 and 10,000 years ago, based on estimates in Ghirotto et al., (2013)], whereas under Model B this event happened just before (and was crucial for) the onset of the Etruscan culture. For each model, we tested various admixture rates.

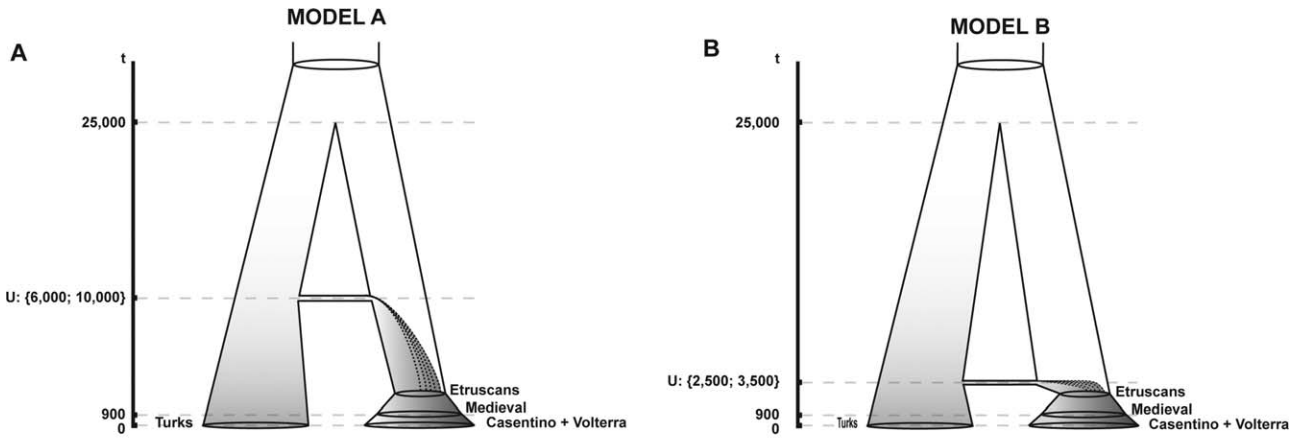
The models are characterized by demographic and evolutionary parameters whose values are independently drawn in each simulation experiment from uniform and wide prior distributions. The ancestral population sizes ranged from 5 to 6,000 individuals and the effective population sizes for modern Tuscans and Anatolians were independently sampled from a prior distribution ranging between 100 and 400,000. The prior for mutation rate was between 0.0003 and 0.0075, in agreement with recent papers based on ABC methods (Sanchez-Quinto et al., 2012; Ghirotto et al., 2013). A schematic outline of the models is in Figure 2 and a complete description of the prior information considered is in Table 2.

Approximate Bayesian computations: Model choice

The two models were compared, and their parameters were estimated, under an ABC framework (Beaumont et al., 2002). The ABC methods combine the analysis of abundant data and realistic models. They allow the probabilistic comparison of different models of evolution accounting for the observed variation, the simultaneous estimation of demographic and evolutionary parameters, and the quantitative evaluation of the results' credibility (Beaumont et al., 2010). ABC method is intuitively very easy: in principle, to test hypotheses on the genealogical relationships between samples, millions of genealogies are generated under different models and assuming different parameter values. The simulations that produce genetic variation patterns close to the observed data are retained and analyzed in detail. Indeed, parameter values and model features in the retained simulations are of course interesting because they are able to generate datasets with some properties found in the observed data. This approach could meet with difficulties because of the large number of parameters needed to fully describe the genealogy underlying the observed data, but the flexibility of ABC makes it possible to evaluate the likelihood also for complex demographic models (Marjoram and Tavarè, 2006). Indeed, the ABC approach allows one to approximate the likelihoods by comparing summary statistics extracted from the data, rather than the DNA sequences themselves, thus reducing the amount of information to account.

The various steps of the ABC procedure and their rationale are described in detail in Bertorelle et al. (2010), and summarized below:

1. For both models, we ran a large number of coalescent-based simulations using the program BayeSSC (Anderson et al., 2005; see http://iod.ucsd.edu/simplex/ssc/BayeSSc.htm); in particular, we



Admixture Rate

	50	60	70	80	90
AR	0.127	0.168	0.176	0.240	**0.289**
LR	0.115	0.158	0.181	0.216	**0.330**

Admixture Rate

	50	60	70	80	90
AR	**0.292**	0.204	0.191	0.155	0.158
LR	**0.334**	0.272	0.188	0.118	0.088

C

	Model A_{90}	Model B_{50}
AR	**0.665**	0.335
LR	**0.698**	0.302

Fig. 2. (**A**) and (**B**) Schematic presentation of the two models tested and ABC results among the admixture rates tested for each model. (**C**) ABC results of the comparison between submodels A_{90} and B_{50}.

generated 2,000,000 simulated datasets for each model (two) and each admixture rate (nine), for a total of 36,000,000 simulations. The values of the other parameters (i.e., population sizes, timing of the demographic events, mutation rates) defining the demographic processes described by the model were drawn from the prior distributions for each simulated dataset.

2. Summary statistics were estimated from both the observed data and from each simulated dataset. Then, after normalization of all statistics, for each simulated dataset, a Euclidean distance between the observed and simulated summary statistics was calculated.
3. Models were then compared for the goodness of their fit to the data. For this procedure, we followed two methods, both based on the *Calmod* function written by M. A. Beaumont (available at PopABC website: https://code.google.com/p/popabc/source/browse/trunk/scripts/calmod.r) for the *R* statistical package. Under the first method, the AR, or acceptance–rejection procedure (Pritchard et al., 1999), the posterior probability of a model is obtained by simply counting the proportion of the simulations in which either model generates statistics arbitrarily close to the observed statistics. This method is considered reliable only when applied to a few simulations showing an excellent fit with the observed data (Beaumont, 2008); therefore, we retained the 750 simulations resulting in the shortest Euclidean distances. The second criterion is a weighted multinomial logistic regression (LR) between the summary statistics and a categorical variable indicating either demographic model (Beaumont, 2008). For this calculation, we retained the 150,000 simulation experiments associated with the shortest Euclidean distances.

Parameter estimation

For the best model, we retained the 5,000 simulations generated under that model showing the shortest associated Euclidean distances, from a total of 5,000,000 simulations. Then, parameters were estimated by a locally weighted multivariate regression (Beaumont et al., 2002), after a *logtan* transformation to prevent the estimate from exceeding the bounds of the prior distribution (Hamilton et al., 2005).

Quality of the estimation: Type I error and posterior predictive test

To check whether the power of the LR and AR procedures is sufficient to actually identify the best model based on the available data, we followed an approach suggested by Fagundes et al. (2007) and Cornuet et al. (2008). First, we simulated 1,000 datasets from the prior distribution under the submodel emerging as the most likely, and we analyzed them as if they were observed datasets in an ABC analysis. We then assigned each of the 1,000 simulated datasets to the model showing the highest posterior probability. Finally, we calculated Type I error as the number of experiments in which the simulated model was not recognized by the model selection procedures.

TABLE 2. Priors, estimates, and R^2 of the parameters estimated under Submodel A_{90}. LowB and UppB are, respectively, the lower and upper bound of the 95% credible interval of the posterior probability distributions

	Priors	Median	Mode	95% HPD LowB	95% HPD UppB	R^2
Time MRCA	[a]	44,424	40,712	9,864	183,807	0.51
Mutation rate	(0.0003–0.0075)	0.0016	0.0016	0.0007	0.0032	0.77
Time admixture	(6,000–10,000)	8,396	10,000	6,007	10,000	0.02
Ne ancestral Tuscans	(5–6,000)	140	36	5	4,729	0.16
Ne ancestral Turks	(5–6,000)	676	229	5	5,814	0.31
Ne Tuscans	(100–400,000)	180,421	128,721	18,257	400,000	0.53
Ne Turks	(100–400,000)	302,391	400,000	6,259	400,000	0.23

HPD, highest posterior density; MRCA, most common recent ancestor.
[a] The time to the most recent common ancestor, Time MRCA, was estimated from the simulated data and not extracted from a prior distribution.

The above-described procedures are suitable to identify the model better reproducing the observed statistics, but do not test whether either model is realistic at all. To that end, we eventually evaluated by a posterior predictive test whether the model we chose has the ability to reproduce the observed data (Gelman et al., 2004). Therefore, we simulated 1,000 datasets according to the model with the highest probability using the estimated posterior parameter distributions. Then, we calculated 20 summary statistics that had not been considered during the previous inferential step, namely, nucleotide diversity and Tajima's *D* within each sample, and five Hudson's F_{ST} and five allele-sharing measures that describe the genetic distance between the Tuscans' samples. We compared these values with the same observed statistics and estimated a posterior predictive *P*-value for each summary statistic. Finally, these probabilities were combined into a global *P*-value, following a procedure described in Ghirotto et al. (2010).

RESULTS

Table 1 shows the statistics summarizing genetic variation in the five samples considered. The posterior probability of the alternative models being compared essentially measures the model's ability to generate data closely resembling the observed data.

Under the ABC framework we actually started by comparing two sets of models. We assumed either (Model A) that the genetic resemblance between Central Italy and Anatolia is because of a relatively ancient gene flow between these geographical regions, or (Model B) that migration from Anatolia brought into Central Italy the immediate ancestors of the Etruscans. The first test we ran was a comparison of the probabilities of various admixture rates (0.50, 0.60, 0.70, 0.80, and 0.90) within each demographic model (as a consequence, each Model was divided into five submodels, namely, A_{50}, A_{60}, A_{70}, A_{80}, A_{90} and B_{50}, B_{60}, B_{70}, B_{80},B_{90}).

We found that the highest posterior probabilities corresponded to an admixture rate of 0.90 for Model A (supported by 29% of the experiments under the LR approach and 33% under the AR approach; Fig. 2A) and 0.50 for Model B (supported by 29% of the experiments under the LR approach and 33% under the AR approach; Fig. 2B). We then proceeded to compare the models keeping constant the admixture rates thus estimated (submodel A_{90} vs. submodel B_{50}), so as to use for both models the most likely admixture level.

Submodel A_{90} proved about twice as likely as the alternative model, regardless of the criterion used for model selection (posterior probabilities were 66% under the LR approach and 70% under the AR approach; Fig. 2C). This result also held when different numbers of simulations were considered to compare models. In addition, the Model A results provided the most probable scenario even when compared with Model B, considering lower admixture rate (data not shown).

Once submodel A_{90} proved able to generate statistics in better agreement with the observed data than the alternative submodel B_{50}, we calculated its parameters' posterior probabilities, here reported in Table 2 and Figure 3, along with the priors. Narrow posterior distributions of the estimates mean that independent simulation experiments suggest similar values, and hence that these estimates are reliable. That seems the case for the ancient populations' sizes, the time to the Most Recent Common Ancestor (MRCA), and the mutation rate; the median for this statistic is 0.17 mutational events per million years per nucleotide, close to the values estimated in previous comparable studies (Hill et al., 2007; Ghirotto et al., 2010). The median time of admixture, 8,396 years ago, is probably an underestimate because the modal value corresponds to the upper limit of the priors we imposed. In other words, had we chosen a broader distribution of priors, we would have likely inferred an older gene flow event. Our purpose, however, was not to estimate a specific date for an event that may well have occurred across many years, or even centuries, but to see if that event has any chance to have closely preceded the appearance of Etruscan artifacts in the archeological record. The answer is that the probability of such a recent event is less than one-third, which means that the alternative is more than twice as likely. The estimates for the archaic population size of Tuscans and Turks are low, not an unexpected finding for the Mediterranean basin in Paleolithic times. On the contrary, the sizes of both modern populations show broad distributions of posterior probabilities. Such a finding is common in studies comparing populations across time (Fagundes et al., 2007; Belle et al., 2009; Laval et al., 2010), and probably reflects the effect of immigration, resulting in incorporation of novel mtDNA variants from external sources that are not easily incorporated into the models. This input of external DNAs increases the internal diversity of populations, and hence the (correlated) estimate of population size.

To be reliable, these results must be supported by evidence that, at the sample sizes we considered, our methods for model selection (AR and LR) were powerful enough to identify the correct model. To answer this question we calculated, for Model A_{90} and Model B_{50},

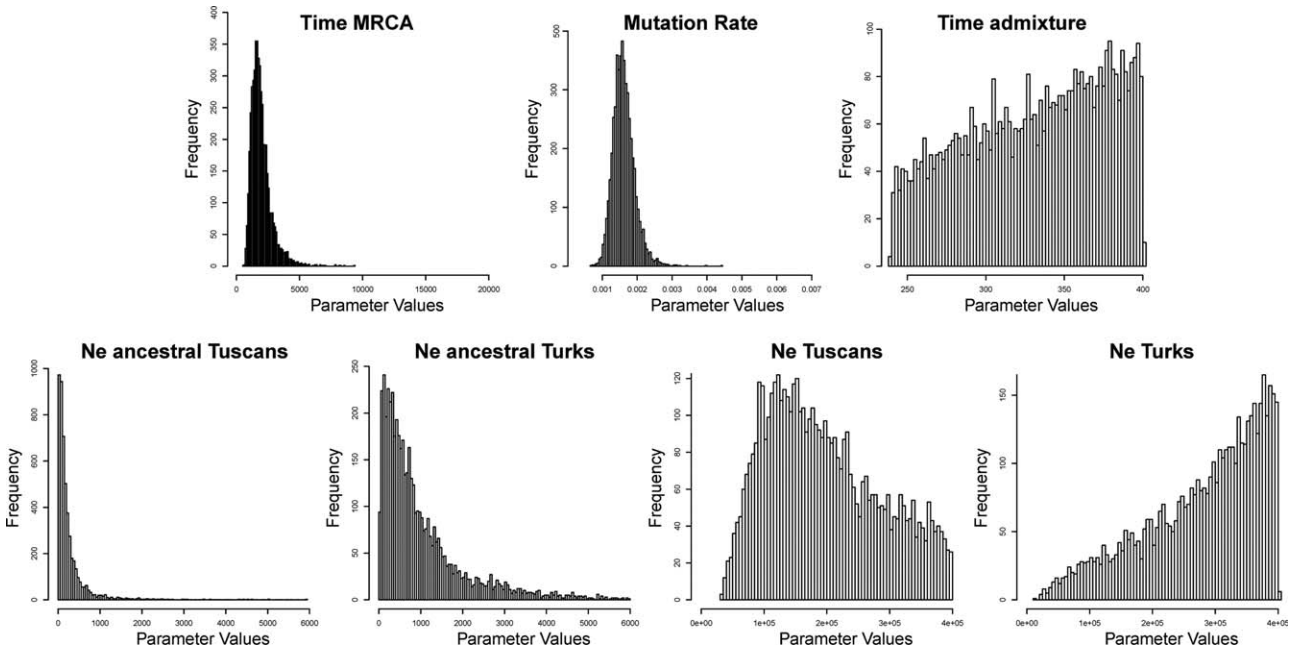


Fig. 3. Posterior distribution of parameters under submodel A_{90}. On the *x*-axis are the parameter values and the width of the axis expresses the range of the (uniform) prior distribution of each parameter, on the *y*-axis their frequencies in the 5,000 best-fitting simulation experiments (out of 5,000,000 performed).

TABLE 3. Type I errors for the two best models emerging from the ABC analysis

	Model A_{90}	Model B_{50}	Type I error
AR			
Model A_{90}	**0.81**	0.19	0.19
Model B_{50}	0.28	**0.72**	0.28
LR			
Model A_{90}	**0.85**	0.15	0.15
Model B_{50}	0.27	**0.73**	0.27

AR, acceptance–rejection criterion; LR, logistic regression criterion. The numbers in bold indicate the percentage of experiments in which the simulated model was successfully recognized.

the Type I error, generating by simulation 1,000 pseudo-observed datasets according to each model, with samples having the same size and age as the observed samples. Analyzing by ABC method these datasets generated under submodels A_{90} and B_{50} as if they were observed datasets, we found that both were in general correctly identified with a probability of recovery ranging from 72% to 84%, so that Type I error was, respectively, 28–16%. As could be expected, the statistical power in this comparison is not very high, because these models are rather similar, differing just for the timing and extent of an admixture event (Table 3).

A related, but different, question would be whether submodel A_{90} can indeed generate patterns of variation compatible with variation in the observed data. The *P*-values calculated by the posterior predictive test led to a global *P*-value for the whole model of 0.48. This probability means that the statistics generated by the model we chose as the best one broadly overlap with, and do not significantly depart from, those in the observed data.

DISCUSSION

The genetic patterns observed at the mtDNA level in the past and present Tuscany have a higher probability of resulting from an ancient migration process from Anatolia than by a migration occurring just before, and associated with, the origins of the Etruscan culture.

This finding is supported by explicit tests of hypotheses against ancient and modern DNA data. It confirms in part previous results based on modern data only, showing that the main separation between the Anatolian and Tuscan mitochondrial pools most likely occurred at least 6,500 years ago, or earlier if the two populations kept exchanging migrants after separation (Ghirotto et al., 2013). Similarly, we did not incorporate into our models the possibility of genetic exchanges between Anatolia and Tuscany after the main admixture event; we see no reason to exclude that these exchanges might have occurred, but, had we considered them, the admixture event would have been placed in a more remote past. Therefore, no genetic evidence, either based on ancient or modern DNA variation, suggests an input of people emigrating from Anatolia into Tuscany as a likely causal factor in the origin of the Etruscan civilization.

The analysis of modern mtDNAs (Ghirotto et al., 2013) and the comparison of ancient and modern DNAs (this study) have another result in common. Despite being based on different methods, and on largely (even if not completely) independent datasets, both dated the contact between the ancestors of current Anatolians and Tuscans at a moment in which gene flow was extensively occurring in Europe, namely, the Neolithic period. Indeed, studies of mitochondrial (Simoni et al., 2000) and nuclear (Chikhi et al., 1998) DNAs in modern Europeans, and comparisons of mitochondrial haplogroups between modern and ancient populations (Bramanti et al., 2009; Fu et al., 2012; Sanchez-Quinto et al., 2012)

show that the Neolithic spread of farming technologies in Europe was accompanied by significant demographic changes. The actual impact of Neolithic processes on European genetic diversity is still debated (e.g., Soares et al., 2010; Arenas et al., 2013) but there is little doubt that a Westward gene flow from the Near East or Anatolia into Europe took place in the Neolithic period (Barbujani and Goldstein, 2004; Barbujani, 2012).

The difference in the probabilities of the two models compared, approximately twofold, is not large. However, one has to consider that (a) genetic differences between populations are minimal in Europe, with the main geographical gradients accounting for some 0.45% of the global diversity (Novembre et al., 2008); (b) because only mtDNA has been typed on a sufficiently large scale to allow for diachronic comparisons, only one locus, albeit a highly variable one, could be considered; (c) a large number of population movements is documented in the ethnohistorical record of Europe (Sokal et al., 1993; Sokal et al., 1996), each of them potentially confounding the genetic pattern left by a more remote event, such as the one we were analyzing; and (d) the two models we were comparing differed only as for the timing of a migration event, and so could not possibly be expected to differ by much in their consequences. In light of these factors, it would have been unrealistic to expect greater levels of statistical significance in this study, and it seems remarkable that we could demonstrate a substantial difference in the models' posterior probabilities. Availability of nuclear DNA sequences from ancient specimens will radically improve our inferential power, but at present only a handful of ancient individuals have been studied at the nuclear level (Sanchez-Quinto et al., 2012).

We are fully aware that the processes that occurred in the Mediterranean area over the long time span considered in this study are far more complex than the one actual model. That is also the reason why we chose to focus mainly on admixture rates equal to or greater than 50%. Although lower values were also taken into consideration, comparing a large number of hypotheses differing for only one parameter, the admixture rate in this case is notoriously complicated (e.g., Konečný et al., 2013). At any rate, smaller Anatolian contributions to the Etruscan gene pool would hardly have been compatible with the notion that the Etruscans are an immigrant Eastern population. In addition, it would have been extremely difficult to tell apart smaller admixture rates from the effects of some of the migration processes that occurred in later prehistoric and historical times. Therefore, we do not give any special importance to the point estimate we obtained for the date of a likely contact between the ancestors of Anatolians and Tuscans, which was admittedly calculated in a rather rough manner. What matters, and has historical relevance, is that this date is clearly earlier than 2,500 or 3,000 years ago, and that a similar date was also inferred from the analysis of only modern DNAs (Ghirotto et al., 2013). More to the point, in this study the posterior probability of Model B increased for older dates of the migration episode, thus suggesting that, if anything, our date might underestimate, certainly not overestimate, the age of the contact.

Therefore, this study shows that inference based on DNA diversity in modern populations is well complemented by ancient DNA studies, and that considering both kinds of data is important if one is to identify the genealogical links of populations. Future studies also considering nuclear DNA diversity in ancient samples will add further details to the general picture, and may possibly lead us to reconsider some of the conclusions of this study. However, the analysis of ancient nuclear genes is still in its infancy and it will take time to accumulate sufficient sample sizes to explicitly test models on the genealogical links between the past and current populations. For the time being, it seems safe to say that, based on the best available data as analyzed by the most advanced biostatistical methods, ancient and modern DNA evidence converges in not suggesting a biological origin of the Etruscans outside Italy. The existing similarities between the Anatolian and Tuscan gene pools (Achilli et al., 2007) can simply be accounted for by the effects of older, or much older, prehistoric contacts, unrelated to the later development of the Etruscan culture.

LITERATURE CITED

Achilli A, Olivieri A, Pala M, Metspalu E, Fornarino S, Battaglia V, Accetturo M, Kutuev I, Khusnutdinova E, Pennarun E, Cerutti N, Di Gaetano C, Crobu F, Palli D, Matullo G, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Semino O, Villems R, Bandelt HJ, Piazza A, Torroni A. 2007. Mitochondrial DNA variation of modern Tuscans supports the Near Eastern origin of Etruscans. Am J Hum Genet 80:759–768.

Anderson CN, Ramakrishnan U, Chan YL, Hadly EA. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. Bioinformatics 21: 1733–1734.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147.

Arenas M, Francois O, Currat M, Ray N, Excoffier L. 2013. Influence of admixture and paleolithic range contractions on current European diversity gradients. Mol Biol Evol 30:57–61.

Bandelt HJ, Kivisild T. 2006. Quality assessment of DNA sequence data: autopsy of a mis-sequenced mtDNA population sample. Ann Hum Genet 70:314–326.

Barbujani G. 2012. Human genetics: message from the Mesolithic. Curr Biol 22:R631–R633.

Barbujani G, Goldstein DE. 2004. Africans and Asians abroad: genetic diversity in Europe. Annu Rev Genomics Hum Genet 5:119–150.

Barker G, Rasmussen T. 1998. The etruscans. Oxford: Blackwell.

Beaumont M. 2008. Joint determination of topology, divergence time and immigration in population trees. Simulations, genetics and human prehistory. Cambridge: McDonald Institute for Archaeological Research. p 135–154.

Beaumont M. 2010. Approximate Bayesian computation in evolution and ecology. Annu Rev Ecol Evol Syst 41:379–406.

Beaumont M, Zhang W, Balding DJ. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Belle EM, Ramakrishnan U, Mountain JL, Barbujani G. 2006. Serial coalescent simulations suggest a weak genealogical relationship between Etruscans and modern Tuscans. Proc Natl Acad Sci USA 103:8012–8017.

Belle EM, Benazzo A, Ghirotto S, Colonna V, Barbujani G. 2009. Comparing models on the genealogical relationships among Neandertal, Cro-Magnoid and modern Europeans by serial coalescent simulations. Heredity 102:218–225.

Bendall KE, Sykes BC. 1995. Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57:248–256.

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.

Bramanti B, Thomas MG, Haak W, Unterlaender M, Jores P, Tambets K, Antanaitis-Jacobs I, Haidle MN, Jankauskas R, Kind CJ, 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.

Brisighelli F, Capelli C, Alvarez-Iglesias V, Onofri V, Paoli G, Tofanelli S, Carracedo A, Pascali VL, Salas A. 2009. The Etruscan timeline: a recent Anatolian connection. Eur J Hum Genet 17:693–696.

Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G. 1998. Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. Proc Natl Acad Sci USA 95:9053–9058.

Cornuet JM, Santos F, Beaumont MA, Robert CP, Marin JM, Balding DJ, Guillemaud T, Estoup A. 2008. Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. Bioinformatics 24:2713–2719.

Di Benedetto G, Erguven A, Stenico M, Castri L, Bertorelle G, Togan I, Barbujani G. 2001. DNA diversity and population admixture in Anatolia. Am J Phys Anthropol 115:144–156.

Excoffier L, Lischer HE. 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.

Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L. 2007. Statistical evaluation of alternative models of human evolution. Proc Natl Acad Sci USA 104:17614–17619.

Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am J Phys Anthropol 128:415–423.

Fu Q, Rudan P, Paabo S, Krause J. 2012. Complete mitochondrial genomes reveal Neolithic expansion into Europe. PLoS One 7:e32473.

Gelman A, Carlin J, Stern H, Rubin D. 2004. Bayesian data analysis. Boca Raton, FL: CRC Press.

Ghirotto S, Mona S, Benazzo A, Paparazzo F, Caramelli D, Barbujani G. 2010. Inferring genealogical processes from patterns of Bronze-Age and modern DNA variation in Sardinia. Mol Biol Evol 27:875–886.

Ghirotto S, Tassi F, Fumagalli E, Colonna V, Sandionigi A, Lari M, Vai S, Petiti E, Corti G, Rizzi E, De Bellis G, Caramelli D, Barbujani G. 2013. Origins and evolution of the Etruscans' mtDNA. PLoS One 8:e55519.

Guimaraes S, Ghirotto S, Benazzo A, Milani L, Lari M, Pilli E, Pecchioli E, Mallegni F, Lippi B, Bertoldi F, Gelichi S, Casoli A, Belle EM, Caramelli D, Barbujani G. 2009. Genealogical discontinuities among Etruscan, Medieval, and contemporary Tuscans. Mol Biol Evol 26:2157–2166.

Hamilton G, Stoneking M, Excoffier L. 2005. Molecular analysis reveals tighter social regulation of immigration in patrilocal populations than in matrilocal populations. Proc Natl Acad Sci USA 102:7476–7480.

Hill C, Soares P, Mormina M, Macaulay V, Clarke D, Blumbach PB, Vizuete-Forster M, Forster P, Bulbeck D, Oppenheimer S, Richards M. 2007. A mitochondrial stratigraphy for island Southeast Asia. Am J Hum Genet 80:29–43.

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

Konečný A, Estoup A, Duplantier JM, Bryja J, Bâ K, Galan M, Tatard C, Cosson JF. 2013. Invasion genetics of the introduced black rat (Rattus rattus) in Senegal, West Africa. Mol Ecol 22:286–300.

Laval G, Patin E, Barreiro LB, Quintana-Murci L. 2010 Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. PLoS One 5: e10284.

Marjoram P, Tavarè S. 2006. Modern computational approaches for analysing molecular genetic variation data. Nature 7:759–770.

Mateiu LM, Rannala BH. 2008. Bayesian inference of errors in ancient DNA caused by postmortem degradation. Mol Biol Evol 25:1503–1511.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, Bustamante CD. 2008. Genes mirror geography within Europe. Nature 456:98–101.

Otte M. 2000. The history of European populations as seen by archaeology. In: Renfrew C, Boyle K, editors. Archaeogenetics: DNA and the population prehistory of Europe. Cambridge: McDonald Institute for Archaeological Research. p 41–44.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16:1791–1798.

Sanchez-Quinto F, Schroeder H, Ramirez O, Avila-Arcos MC, Pybus M, Olalde I, Velazquez AM, Marcos ME, Encinas JM, Bertranpetit J, Orlando L, Gilbert MT, Lalueza-Fox C. 2012. Genomic affinities of two 7,000-year-old Iberian hunter-gatherers. Curr Biol 22:1494–1499.

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. 2000. Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:262–278.

Soares P, Achilli A, Semino O, Davies W, Macaulay V, Bandelt HJ, Torroni A, Richards MB. 2010. The archaeogenetics of Europe. Curr Biol 20:R174–R183.

Sokal RR, Jacquez GM, Oden NL, DiGiovanni D, Falsetti AB, McGee E, Thomson BA. 1993. Genetic relationships of European populations reflect their ethnohistorical affinities. Am J Phys Anthropol 91:55–70.

Sokal RR, Oden NL, Walker J, Di Giovanni D, Thomson BA. 1996. Historical population movements in Europe influence genetic relationships in modern samples. Hum Biol 68:873–898.

Vernesi C, Caramelli D, Dupanloup I, Bertorelle G, Lari M, Cappellini E, Moggi-Cecchi J, Chiarelli B, Castri L, Casoli A, Mallegni F, Lalueza-Fox C, Barbujani G. 2004. The Etruscans: a population-genetic study. Am J Hum Genet 74:694–704.