

## Inference of demographic processes from comparisons of ancient and modern DNAs

Guido Barbujani

*Dept. Biology and Evolution, University of Ferrara, Ferrara, Italy*  
e-mail: g.barbujani@unife.it

Our ability to infer past demographic changes has substantially improved with the development of methods for the reliable typing of DNA from ancient specimens. However, the inferential process remains complicated, because ancient samples are small and the genetic information they yield is generally limited to one marker, mtDNA. Therefore, whenever dealing with ancient DNA evidence, besides asking what is the demographic model best accounting for the observed patterns in the data, one has also to consider whether there is enough statistical power in the data to discriminate among alternative models. To address the main question, one basically compares scenarios of genetic continuity between ancient and modern samples with scenarios in which the samples belong to different branches of the genealogical tree.

Computer simulation of explicit demographic models is an effective means to test hypotheses on the relationships between ancient and modern samples. Serial coalescent approaches, in particular (Anderson *et al.*, 2005), allow one to generate genealogies from the present back to the common ancestor, in which individuals are added at various moments in time, representing modern and ancient samples. By attributing a DNA sequence to the common ancestor of the whole genealogy, and by randomly distributing mutations on the genealogical tree, one thus generates many simulated datasets. The sequences themselves are arbitrary (in fact, strings of 0s and 1s), but their differences are not, as they reflect the consequences of the genealogical and of the mutational

processes. Therefore, one can estimate from them summary statistics, describing how genetic variation would be if the model is true.

Algorithms of Approximate Bayesian Computations (ABC: Beaumont *et al.*, 2002) allow comparisons among models, as well as the estimation of the relevant demographic parameters. In short, genetic diversity in the data is summarized by a number of observed summary statistics. Millions of realizations of the demographic process assumed under each model are generated by Serial coalescent simulation, with parameters sampled from appropriately broad distributions of priors. An arbitrary number (threshold) of simulation experiments showing the shortest Euclidean distance between observed and simulated summary statistics are then retained, and the model parameters are estimated from them. By counting how often each specific model generated data falling within the best-fitting simulation replicates, one estimates a global posterior probability for each model. Algorithms exist for testing whether the parameters estimated under each model depart significantly from the observed statistics, and whether there is enough power in the data to discriminate among models.

Two applications of this method to ancient DNA data from populations of pre-classical Italy, are giving rather different descriptions of the evolution of genetic diversity through a time-bracket of some 2,500 years. In Sardinia, two modern populations separated in space by just 120 km, Ogliastra and Gallura, showed

**Tab. 1 – Posterior probabilities of three models (detailed in the first row of the table) of the genealogical relationships between ancient and modern populations of Sardinia. LR is the logistic regression method, AR is the acceptance-rejection method, and threshold is the number of best-fitting simulations considered for the comparison. For each method and threshold (i.e., for each row of the table), the sum of posterior probabilities is 1.**

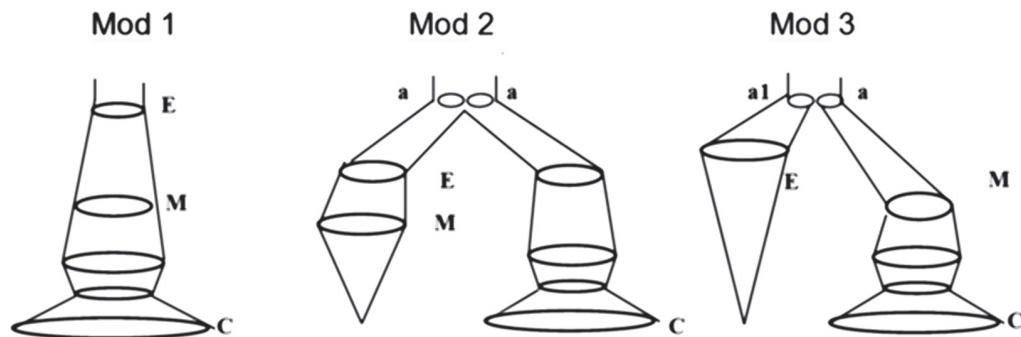
Method	Threshold	Model 1 (Ogliastra in genealogical continuity with ancient Sardinians)	Model 2 (Gallura in genealogical continuity with ancient Sardinians)	Model 3 (Ogliastra and Gallura in genealogical continuity with ancient Sardinians)
LR	50,000	0.956	0.018	0.027
LR	22,500	0.957	0.020	0.023
LR	12,000	0.961	0.019	0.021
LR	6,000	0.970	0.012	0.018
AR	500	0.760	0.120	0.120
AR	100	0.860	0.090	0.050

very different relationships with a sample of 23 individuals from Bronze-age burials. A direct genealogical continuity between Bronze-age Sardinians and the current people of Ogliastra (a genetic isolate), but not Gallura, showed a much higher probability than any alternative scenarios, regardless of the method chosen for comparing models (Table 1). Also, there was evidence that genetic diversity in Gallura evolved largely independently, owing in part to gene flow from mainland Italy (Ghirotto *et al.*, 2009).

In Tuscany, we are currently investigating the demographic scenarios accounting for the observed relationships among modern and ancient (Etruscan) inhabitants of the area. The Etruscans' biological origins are unclear, with ancient historians suggesting either that they immigrated from Anatolia, or alternatively that they represent an autochthonous population (Barker & Rasmussen, 1998); equally obscure are their genealogical relationships with current inhabitants of Tuscany. We had available a set of 20 Etruscan sequences (Vernesi *et al.*, 2004). In general, moderns Tuscans sampled in the areas of highest density of Etruscan sites show some mtDNA resemblance with people of the Eastern Mediterranean shore (Achilli *et al.*, 2007), but not with the Etruscans, and that difference is unlikely to result from

systematic errors in the ancient DNA sequences (Mateiu & Rannala, 2008).

In a preliminary ABC analysis of several modern and ancient samples, the latter comprising Etruscans and Middle-age people from Tuscany (Guimaraes *et al.*, 2009), we compared three basic models of the genealogical relationships among samples (Fig. 1). We found no evidence of genealogical continuity for two Tuscan communities, Murlo and Volterra, for which Model 2 was clearly supported by data. On the contrary, Model 1 received strong statistical support when we compared with the ancient samples a third Tuscan area, Casentino. In addition, we could fit model 1 also to the mtDNA sequences from a population of the Western coast of Anatolia, where Herodotus placed the putative origin of the Etruscans. To make sure that those findings had a biological meaning, we also compared the Etruscans with other modern Italian samples, finding again no evidence of genealogical continuity. The apparent common ancestry does not clearly imply that modern Western Anatolians and Casentino people are both descended from the Etruscans, but rather that they share common ancestors who did not differ much from the Etruscans. Herodotus proposed an origin of the Etruscan culture in a migration episode



**Fig. 1** – Schematic representation of three models of demographic relationships among Etruscan (E), Medieval Tuscan (M) and Contemporary (C) populations. In Models 2 and 3, an ancestral population that underwent a split is designated by an a.

from Anatolia less than 3000 years ago. To test whether genetic data give any support to this interpretation, we are currently estimating by IM methods the likely time of separation of the two modern samples.

In general, the comparisons of ancient and modern DNA suggest that genetic traces of the ancient inhabitants of a region can be found among the modern people, but modern populations are a mosaic of mtDNAs, and cannot be regarded as globally descended from the people who inhabited the same regions in preclassical times.

## References

- 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 A.S., Cavalli-Sforza L.L., Semino O., Villems R., Bandelt H.J., 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, C.N., Ramakrishnan, U., Chan, Y.L. & Hadly, E.A. 2005. Serial SimCoal: a population genetics model for data from multiple populations and points in time. *Bioinformatics*, 21:1733-1734.
- Barker G. & Rasmussen T. 1998. *The Etruscans*. Blackwell, Oxford.
- Beaumont M.A., Zhang W. & Balding D.J. 2002. Approximate Bayesian computation in population genetics. *Genetics*, 162:2025-2035.
- Ghirotto S., Mona S., Benazzo A., Paparazzo F., Caramelli D. & Barbujani G. 2009. Inferring genealogical processes from patterns of Bronze-age and modern DNA variation in Sardinia. *Mol. Biol. Evol.*, 27:875-886.
- 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 E.M., Caramelli D. & Barbujani G. 2009. Genealogical discontinuities among Etruscan, Medieval and contemporary Tuscans. *Mol. Biol. Evol.*, 26:2157-2166.
- Mateiu L.M. & Rannala B.H. 2008. Bayesian inference of errors in ancient DNA caused by postmortem degradation. *Mol. Biol. Evol.*, 25:503-511.
- Vernesi C., Caramelli D., Dupanloup I., Bertorelle G., Lari M., Cappellini E., Moggi-Cecchi J., Chiarelli B., Castrì L., Casoli A., Mallegni F., Lalueza-Fox C. & Barbujani G. 2004. The Etruscans: A population-genetic study. *Am. J. Hum. Genet.*, 74:694-704.

