Characterization of historical demographic expansions from linked microsatellite data

Miguel Navascués, Olivier Hardy

The mismatch distribution expanded to linked microsatellites

Miguel Navascués* & Olivier Hardy

Laboratoire Eco-Ethologie Évolutive, Faculté des Sciences, Université Libre de Bruxelles, Avenue F.D. Roosevelt 50, Campus du Solbosch, 1050 Brussels, Belgium
* Current address: Laboratoire d'Ecologie, École Normale Supérieure, 46 rue d'Ulm 75230 Paris Cedex 05, France. E-mail: m.navascues@gmail.com

The use of genetic data to make inferences on demographic dynamics is one of the common applications of population genetics. Many statistical methods employed on this task are based on coalescent theory which describes the genealogical relationships among a group of genes as a function of population demography (fig 1).

The theoretical background provided by the coalescent describes the distribution of mutations among the sample of genes under different demographic scenarios. However, the applicability of these predictions to empirical genetic data will be hampered by the mutational mechanisms that can erase or distort some of the useful information with back or parallel mutations. The inclusion of mutation models in theoretical predictions can improve the performance of statistical methods.

Li (1977) described the probability distribution for the number of mutations (j) between a pair of haplotyes evolving under a demographic expansion. Schneider & Excoffier (1999) introduced a correction on that equation that allowed describing the distribution of the number of genetic differences (i) for mutation model for DNA sequence polymorphisms (see equation).

Polymorpshim on linked microsatellites (repetitive DNA sequences of short repeat motif, fig 2) is sensitive to demographic expansions (Navascués *et al*. 2006). However large number of homoplasious mutations are expected on these markers which affects the estimation of demographic parameters. In order to improve the accuracy on these estimates we have developed a correction similar to that of Schneider & Excoffier (1999) but using the stepwise mutation model to describe microsatellite evolution.

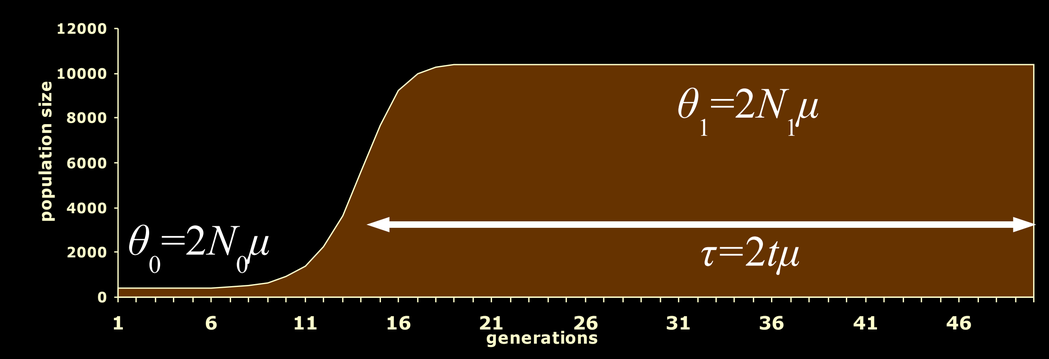
$$P(i;\theta_0,\theta_1,\tau)=\sum_{j=i}^{\infty} P(j;\theta_0,\theta_1,\tau)\times P(i;j)$$

- $P(i;\theta_0,\theta_1,\tau)$: probability distribution of the number genetic differences (i) given a demographic model (θ_0,θ_1,τ)
- $P(j;\theta_0,\theta_1,\tau)$: probability distribution of the number of mutations (j) given a demographic model (θ_0,θ_1,τ)
- $P(i;j)$: probability distribution of the number of genetic differences (i) given a number of mutations (j) under some mutational model

Data was simulated for a set of linked loci evolving under a stepwise mutation model in a population going through a demographic expansion. Four different ages (τ) for the expansion were simulated and maximum likelihood estimates were obtained for the time of expansion using a model without homoplasy [using only Li's $P(j;\theta_0,\theta_1,\tau)$] and a model accounting for homoplasy [using $P(i;\theta_0,\theta_1,\tau)$ with the correction based on the stepwise mutation model]. Table 1 shows how bias and error decrease when using the model corrected for homoplasy, specially for older expansions.

Table 1	Maximum likelihood estimator, model without homoplasy			Maximum likelihood estimator, model with homoplasy		
	mean among estimates	mean squared error	bias	mean among estimates	mean squared error	bias
Sim τ = 1	0.885	0.083	-0.115	1.024	0.154	0.025
Sim τ = 3	2.346	0.564	-0.654	2.856	0.493	-0.144
Sim τ = 5	3.502	2.398	-1.498	4.900	0.872	-0.100
Sim τ = 7	4.361	7.328	-2.639	6.784	2.725	-0.216

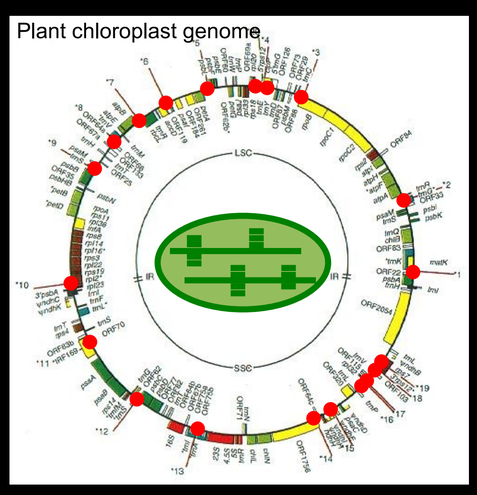
Fig 1 - STAR-LIKE GENEALOGIES

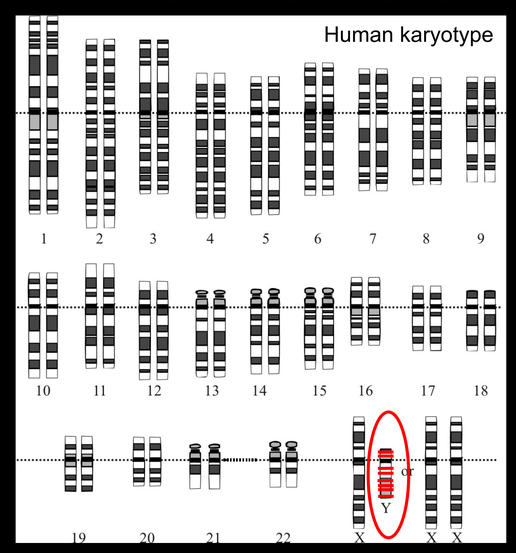


Demographic expansions produce characteristic genealogies where most coalescent events occurred at the time just before the expansion, when population size was small.

Fig 2 - WHERE TO FIND LINKED MICROSATELLITES

Linked microsatellites are typically found on the Y chromosome and on the chloroplast genome.





REFERENCES

- Li (1977) *Genetics* **85**:331–337
- Navascués *et al*. (2006) *Molecular Ecology* **15**:2691–2698
- Schneider & Excoffier (1999) *Genetics* **152**:1079–1089

ACKNOWLEDGMENTS

I am grateful to Brent Emerson & Concetta Burgarella for their support and discussion. This study was partially funded by the European Science Foundation ConGen Exchange Grant 1142