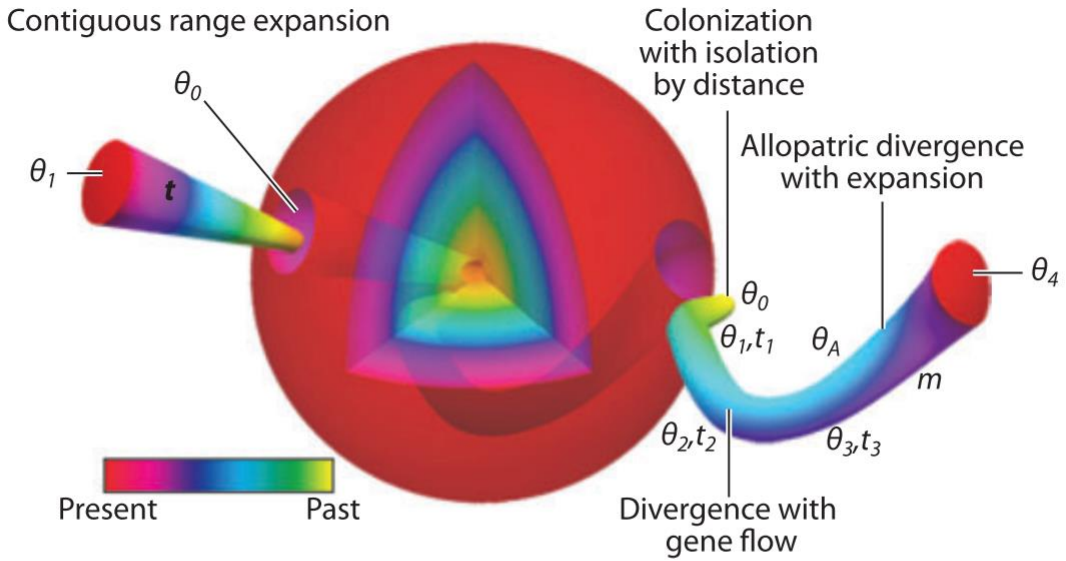
Phylogenetics and Molecular Evolution/Filogenética e Evolução Molecular

Octávio S. Paulo
Computational Biology and Population Genomics Group (CoBiG2)

Análise Demográfica

Sumário:
A análise da história demográfica com base em sequências intraespecíficas.

Demographic history



Demographic history: past processes structuring patterns of genetic variation; includes population size, migration rates, divergence times, and variation in these quantities over time

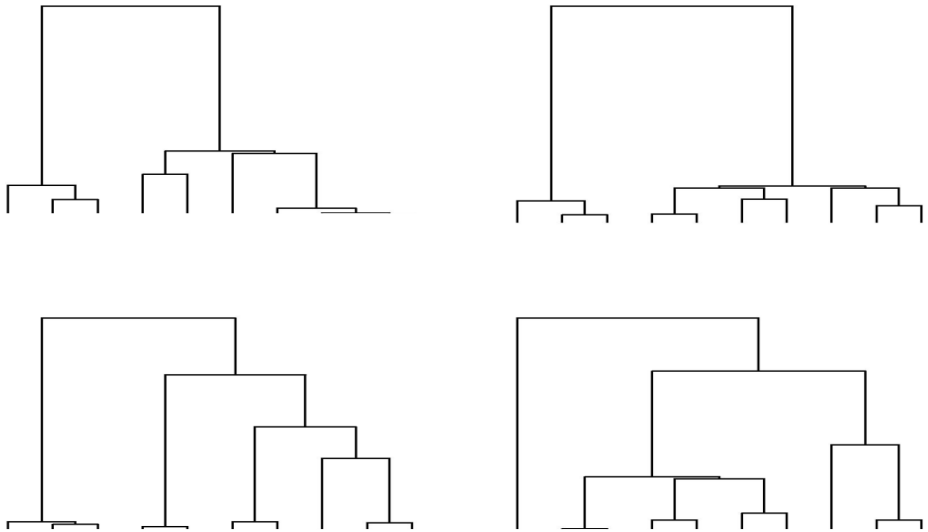
Figure 3

The universe of potential historical scenarios that might be considered in statistical phylogeography is huge. As represented here on the surface of the sphere, there is a continuous array of possible demographic processes and associated parameter values. Depending on the trajectory of a population's history over time (i.e., the path traversed through the sphere), a simple model such as contiguous range expansion (shown on the *left*) might provide an accurate simplified representation, whereas in other cases, a more complex model that incorporates different evolutionary processes at varying time points (shown on the *right*) would be needed for statistical phylogeographic analyses.

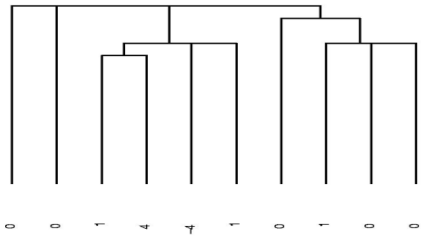
Knowles 2009 ARES

Coalescence and Demography

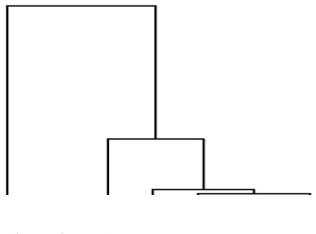
Estável



Expansão



Contração



Demografia e selecção

$$\theta = 4N\mu$$

S the number of segregating sites $\hat{\theta}_S$

π the average number of pairwise differences $\hat{\theta}_\pi$

Neutral /constant $\hat{\theta}_\pi = \hat{\theta}_S$

Shortly after hitchhiking or bottleneck $\hat{\theta}_\pi < \hat{\theta}_S$

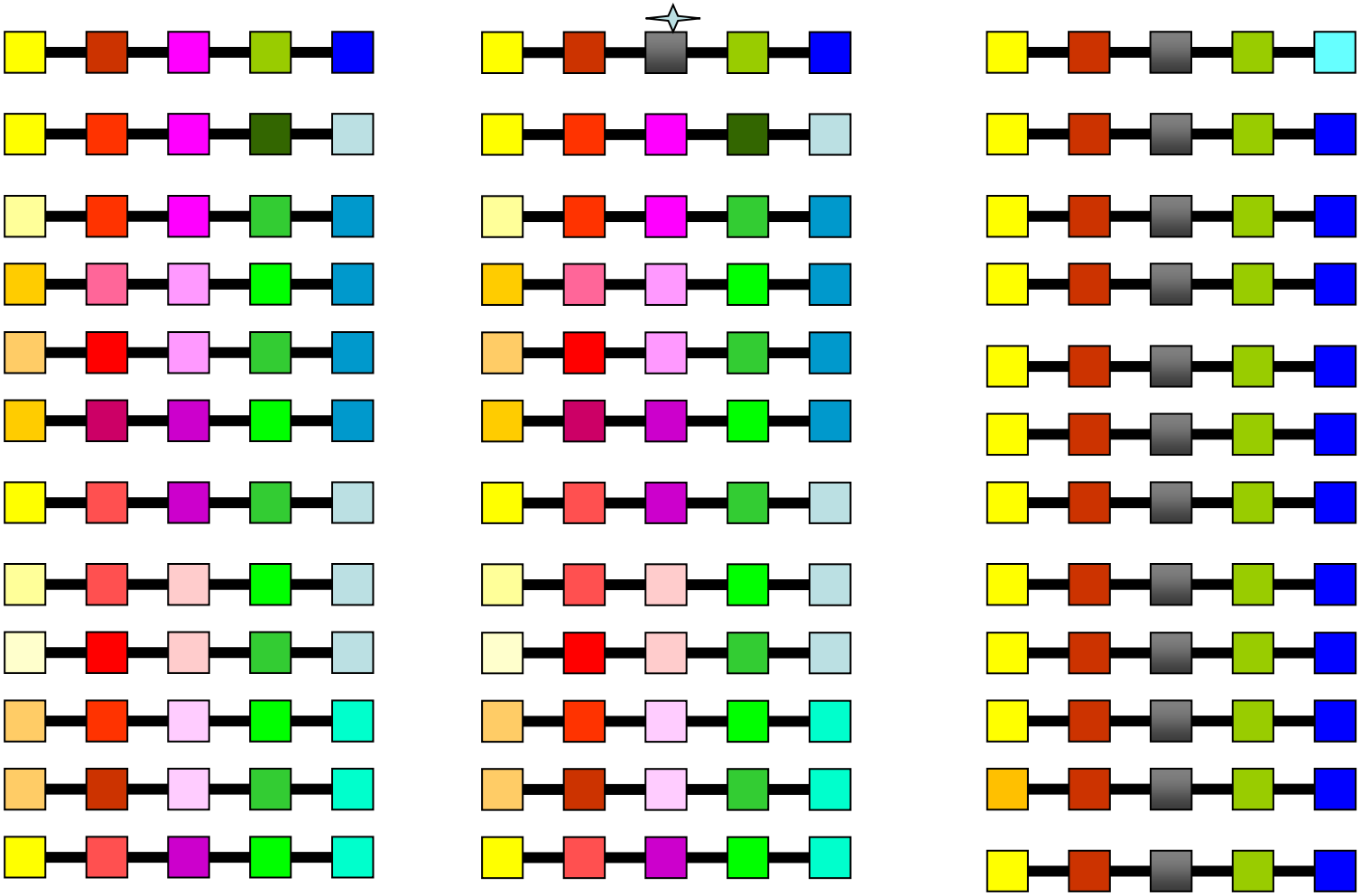
Tajima's *D*

$$D = \frac{\theta_\pi - \theta_S}{\sqrt{\tilde{V}[\theta_\pi - \theta_S]}}$$

The difference between two estimates of θ, divided by an estimate of the standard deviation of that difference.

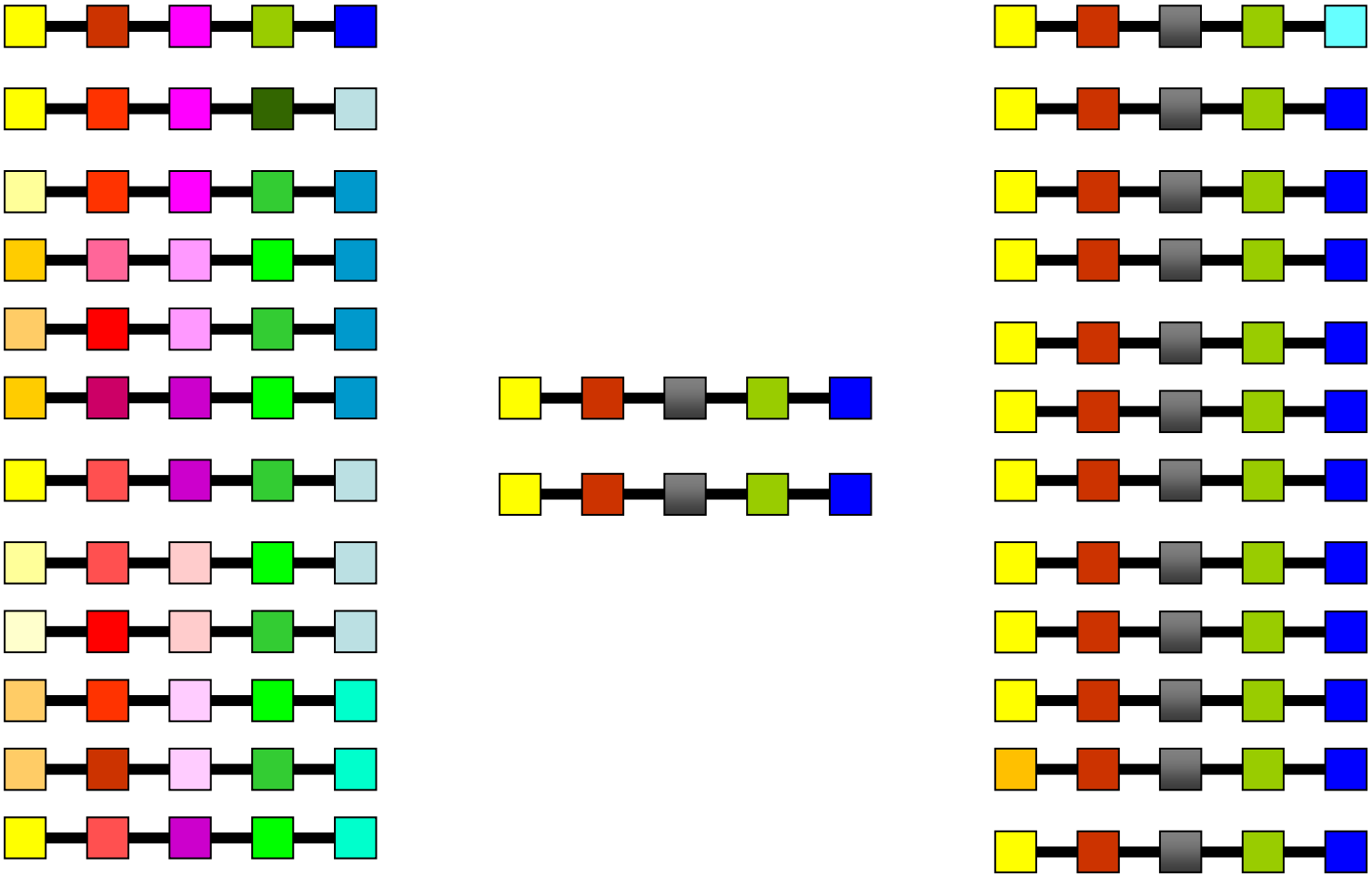
- Negative values indicate positive selection or population growth (excess of rare variants).
- Positive values indicate population decline (sometimes)(excess of intermediate frequency variants).

Selection



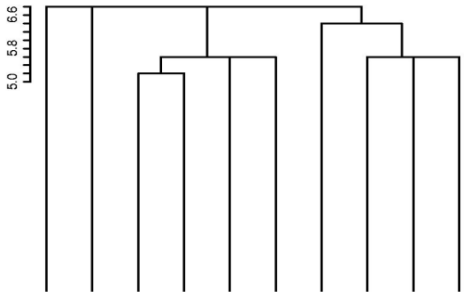
tempo

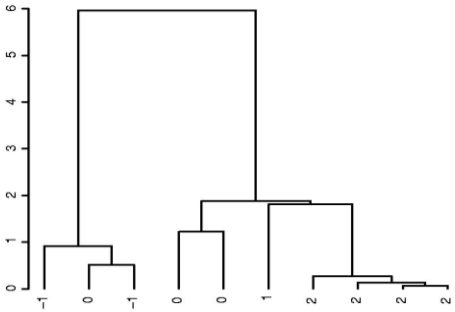
Bottleneck

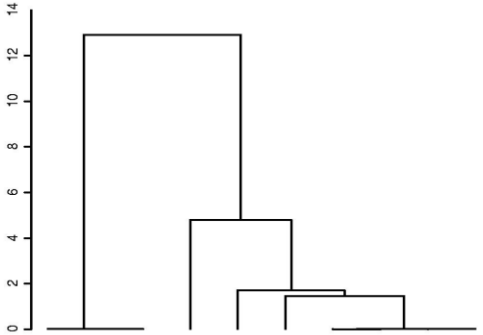


tempo

Demography and Selection



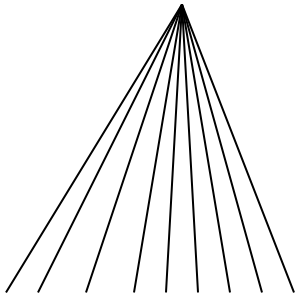


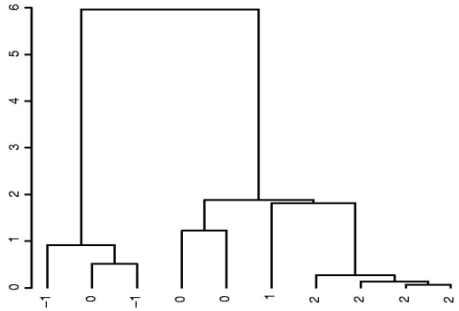


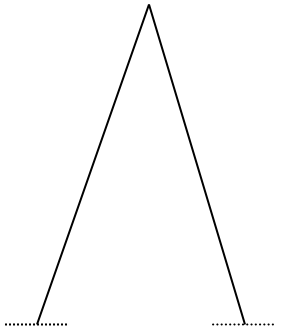
$$\theta_S > \theta_\pi$$

$$\theta_S \approx \theta_\pi$$

$$\theta_S < \theta_\pi$$







Didelot et. al 2021 MBE

Fu's Fs test

above. Here, we evaluate the probability of observing a random neutral sample with a number of alleles similar or smaller than the observed value (see section 7.1.2.3.3 to see how this probability can be computed) given the observed number of pairwise differences, taken as an estimator of θ. In more details, Fu first calls this probability $S' = \Pr(K \geq k_{obs} \mid \theta = \hat{\theta}_{\pi})$ and defines the F_S statistic as the logit of S'

$$F_S = \ln\left(\frac{S'}{1 - S'}\right) \qquad \text{(Fu, 1997)}$$

Fu (1997) has noticed that the F_S statistic was very sensitive to population demographic expansion, which generally lead to large negative F_S values.

Large negative values of Fs will be taken as evidence of expansion

Diversidade haplotipica

Diverisidade Haplotipica (h):

$$h = 1 - \sum f_i^2$$

f_i Frequência do *i* haplótipo

Diversidade haplotipica

$$\hat{H} = \frac{n}{n-1}(1 - \sum_{i=1}^{k} p_i^2)$$

$$\mathrm{V}(\hat{H}) = \frac{2}{n(n-1)}\left\{2(n-2)\left[\sum_{i=1}^{k} p_i^3 - (\sum_{i=1}^{k} p_i^2)^2\right] + \sum_{i=1}^{k} p_i^2 - (\sum_{i=1}^{k} p_i^2)^2\right\},$$

where n is the number of gene copies in the sample, k is the number of haplotypes, and p_i is the sample frequency of the i-th haplotype.

Diversidade nucleotidica

Diversidade nucleotidica (π):

$$\hat{\pi}_n = \frac{\sum_{i=1}^{k} \sum_{j<i} p_i p_j \hat{d}_{ij}}{L}$$

$$\mathrm{V}(\hat{\pi}_n) = \frac{n+1}{3(n-1)L}\hat{\pi}_n + \frac{2(n^2+n+3)}{9n(n-1)}\hat{\pi}_n^2$$

p_i Frequência do *i* haplótipo

p_j Frequência do *j* haplótipo

d_{ij} Sequence divergence entre *i* e *j*

Diversidade haplotipica e Diversidade nucleotidica

h: elevado π: elevado → ou admixture

h: elevado π: baixo →

h: baixo π: elevado → ou admixture

h: baixo π: baixo → Ou SS

Mismatch distribution

The distribution of the number of pairwise differences between haplotypes, from which parameters of a demographic (*NEW*) or spatial population expansion can be estimated

Mismatch distribution

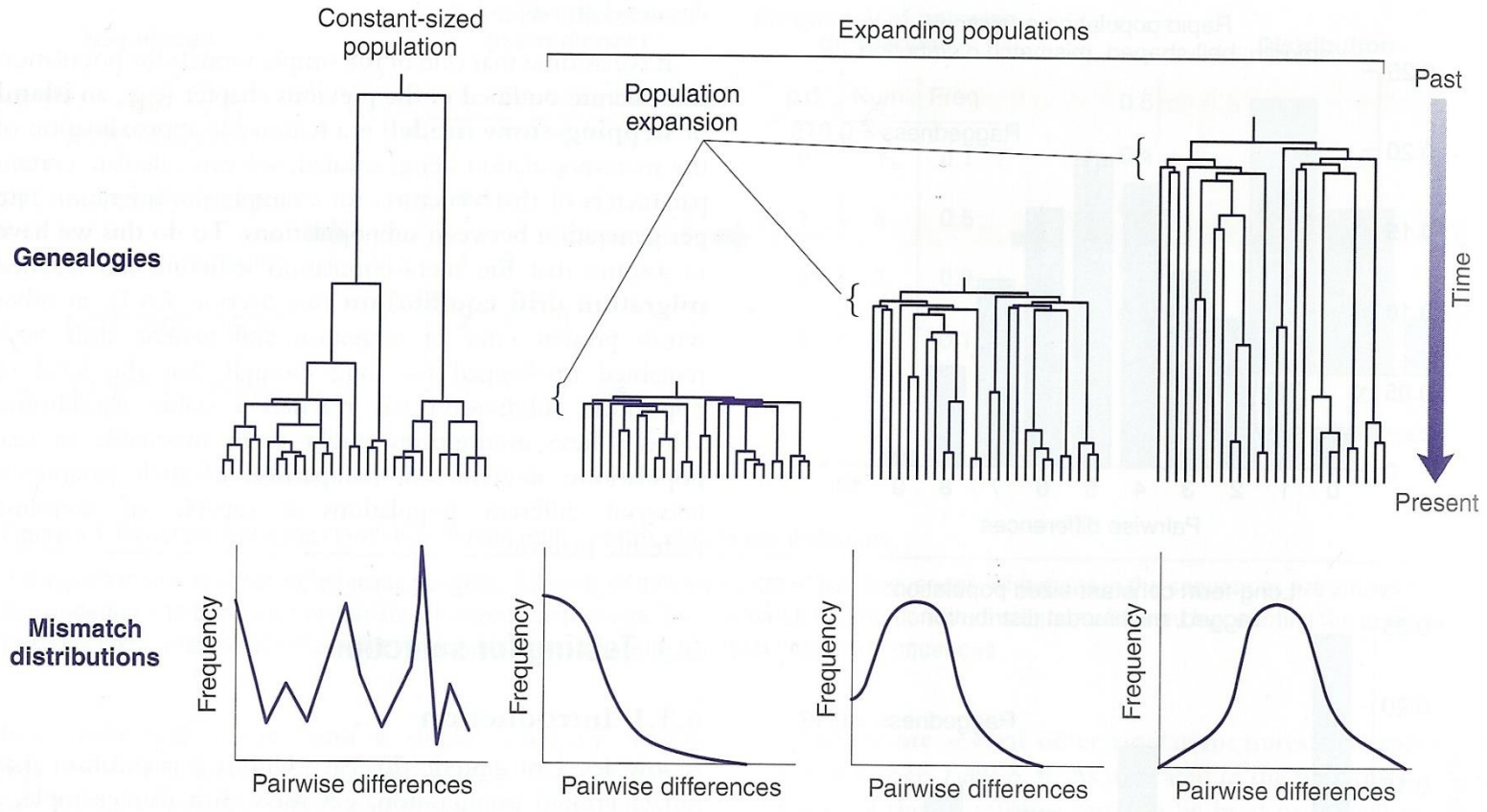


Figure 6.3: Genealogies and mismatch distributions for a constant-sized population and three populations that have undergone population expansions at different times.

The longer branches in the genealogy of the constant population are lineages ancestral to many individuals, whereas the longer branches in the expanding populations are often specific to individuals. Because branch length is indicative of time and therefore the accumulation of mutations, more mutations are shared among individuals in the constant population than in the expanding population, and greater numbers of mutations differentiate individuals in populations with more ancient expansions. Thus it can be seen how the mismatch distributions summarize this information.

Mismatch distribution

Sequences

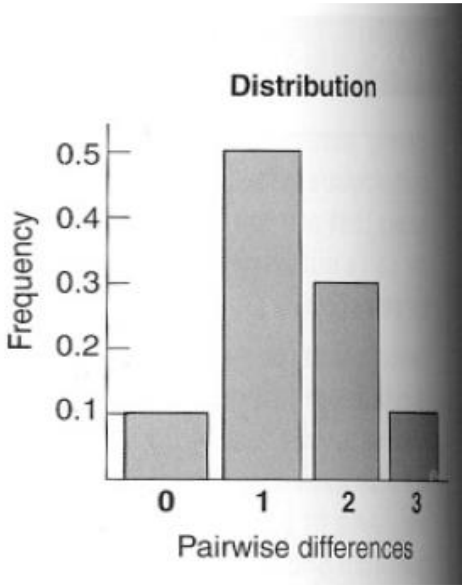
I. AGTCTTACGTATC
II. AGTCTTGCGTATC
III. AGTTTTACGTATC
IV. AGTCTTGCGTGTC
V. AGTCTTACGTATC

Pairwise differences (mismatches)

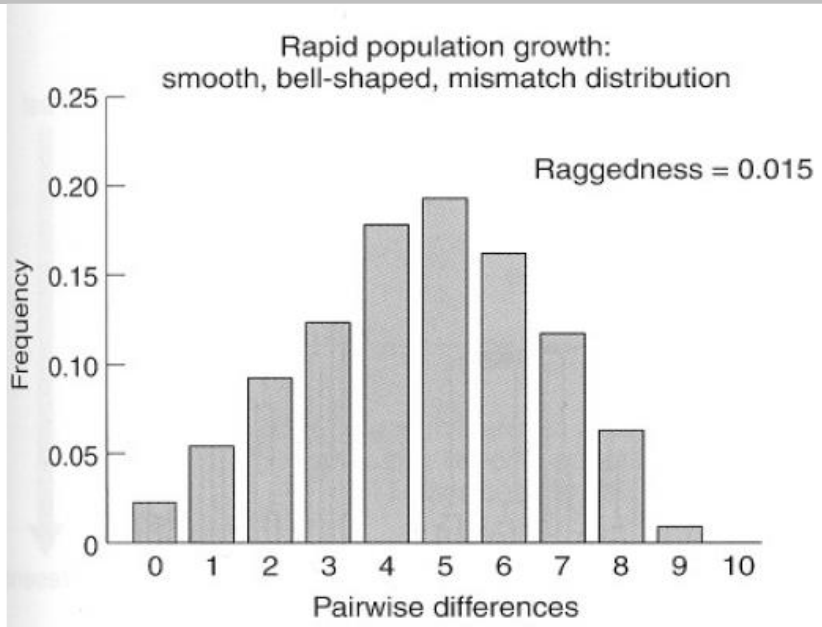
	I	II	III	IV	V
I	–				
II	1	–			
III	1	2	–		
IV	2	1	3	–	
V	0	1	1	2	–

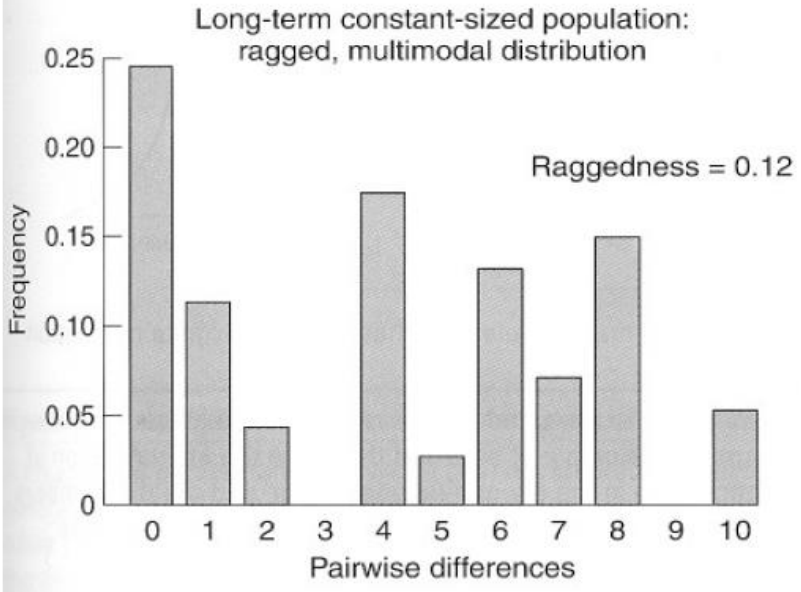
Frequency of pairwise differences

p.d.	Num.	Freq.
0	1	0.1
1	5	0.5
2	3	0.3
3	1	0.1



Mismatch distribution





Mismatch distribution

7.1.2.4 Mismatch distribution

It is the distribution of the observed number of differences between pairs of haplotypes. This distribution is usually multimodal in samples drawn from populations at demographic equilibrium, as it reflects the highly stochastic shape of gene trees, but it is usually unimodal in populations having passed through a recent demographic expansion (Rogers and Harpending, 1992; Hudson and Slatkin, 1991) or though a range expansion with high levels of migration between neighboring demes (Ray et al. 2003, Excoffier 2004).

Mismatch distribution

7.1.2.4.1 Pure demographic expansion

If one assumes that a stationary haploid population at equilibrium has suddenly passed τ generations ago from a population size of N_0 to N_1, then the probability of observing S differences between two randomly chosen non-recombining haplotypes is given by

$$F_S(\tau,\theta_0,\theta_1) = F_S(\theta_1) + exp(-\tau\frac{\theta_1+1}{\theta_1})\sum_{j=0}^{S}\frac{\tau^j}{j!}\left[F_{S-j}(\theta_0) - F_{S-j}(\theta_1)\right], \qquad \text{(Li, 1977)}$$

where $F_S(\theta) = \frac{\theta^S}{(\theta+1)^{S+1}}$ is the probability of observing two random haplotypes with S differences in a stationary population (Watterson, 1975), $\theta_0 = 2uN_0$, $\theta_1 = 2uN_1$, $\tau = 2ut$, and u is the mutation rate for the whole haplotype.

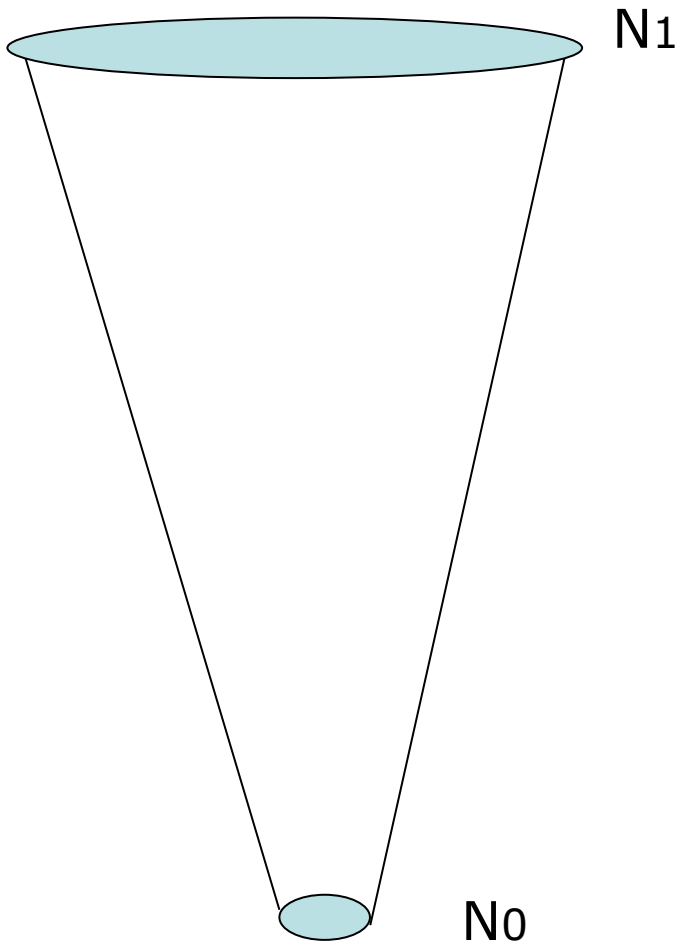
Mismatch distribution

$$\theta_0 = 2uN_0$$

$$\theta_1 = 2uN_1,$$

$$\tau = 2ut$$

Mutational time/
estimator of the time to the expansion



Mismatch distribution

$$\theta_0 = 2uN_0$$

$$\theta_1 = 2uN_1,$$

$$\tau = 2ut$$

$$\hat{\theta}_0 = \sqrt{v - m}$$

$$\hat{\tau} = m - \hat{\theta}_0 \quad ,$$

where m and v are the mean and the variance of the observed mismatch distribution,

Timing of the demographic expansion

$$\theta_0 = 2uN_0$$

$$\theta_1 = 2uN_1,$$

$$\tau = 2ut$$

Timing of the demographic expansion as well as the 95% confidence interval for each mitochondrial haplogroup was estimated by converting the expansion time parameter τ, generated by ARLEQUIN, to time (t) in years using the formula τ = 2ut, where u is the mutation rate per nucleotide per year multiplied by sequence length (i.e. number of nucleotides), and t is the time since population expansion in years [45,49]. We assumed a generation time of one year [24] and the conserved evolutionary rate of 3.54% per million years suggested by [41] for insect mitochondrial gene COI.

$$\hat{\theta}_0 = \sqrt{v - m}$$

$$\hat{\tau} = m - \hat{\theta}_0 \quad ,$$

where m and v are the mean and the variance of the observed mismatch distribution,

Test on the expansion model

The validity of the estimated stepwise expansion model is tested using the same parametric bootstrap approach as described above. We used here the sum of square deviations (*SSD*) between the observed and the expected mismatch as a test statistic.

We obtained its distribution under the hypothesis that the estimated parameters are the true ones, by simulating *B* samples around the estimated parameters. As before, we re-estimated each time new parameters θ_0^*, θ_1^* and τ^*, and computed their associated sums of squares SSD_{sim}. The P-value of the test is therefore approximated by

$$P = \frac{\text{number of } SSD_{sim} \text{ larger or equal to } SSD_{obs}}{B}.$$

Raggedness Index

For convenience, we also compute the raggedness index of the observed distribution defined by Harpending (1994) as

$$r = \sum_{i=1}^{d+1} (x_i - x_{i-1})^2 ,$$

where d is the maximum number of observed differences between haplotypes, and the x's are the observed relative frequencies of the mismatch classes. This index takes larger values for multimodal distributions commonly found in a stationary population than for unimodal and smoother distributions typical of expanding populations. Its significance is tested similarly to that of *SSD*.

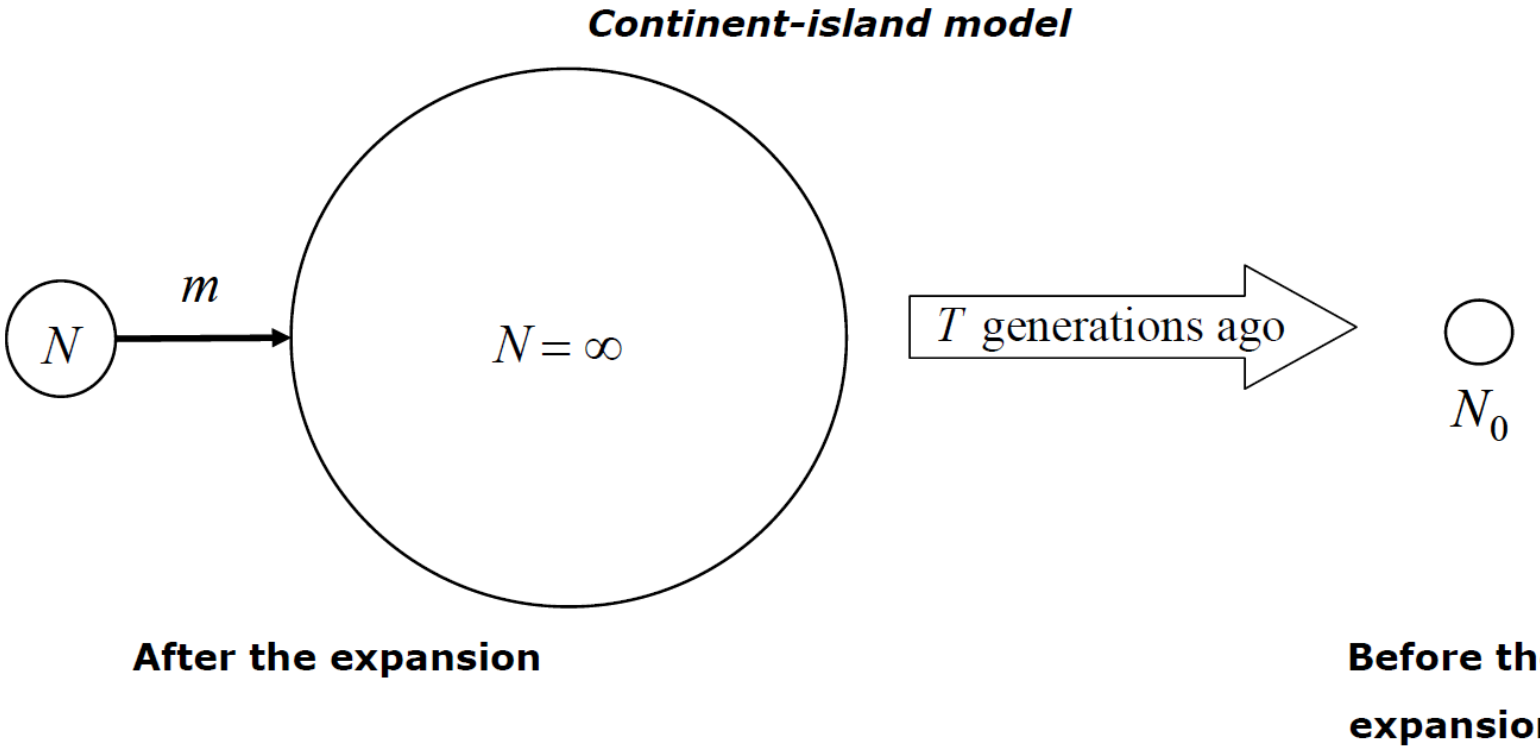
Spatial Expansion

A population spatial expansion generally occurs if the range of a population is initially restricted to a very small area, and then the range of the population increases over time and over space. The resulting population becomes generally subdivided in the sense that individuals will tend to mate with geographically close individuals rather than remote individuals.

Based on simulations, Ray et al. (2003) have shown that a large spatial expansion can lead to the same signal in the mismatch distribution than a pure demographic expansion in a panmictic population, but only if neighboring sub-populations (demes) exchange many migrants (50 or more). The simulations performed in Ray et al. (2003) were

Spatial Expansion

wirh other demes. This infinite-island model is actually equivalent to a continent-island model, where the sampled deme would exchange migrants at rate *m* with a unique population of infinite size. Some *T* generations in the past, the continent-island system would be reduced to a single deme of size N_0, like:



Spatial Expansion

Under this simple model, the probability that two genes currently sampled in the small deme of size *N* differ at *S* sites is given by

$$F_0(S;\ M,\ \theta_0;\ \theta_1,\ \tau) = \frac{\theta_1^S}{A^{S+1}} + \sum_{j=0}^{S}\left(\frac{(Me^{-\tau} + C)\,\theta_0^j\,\tau^{S-j}}{(M+1)(\theta_0+1)^{j+1}(S-j)!} - \frac{\tau^j \theta_1^{S-j} C}{j!\,A^{S-j+1}} \right)$$

, Excoffier (2004)

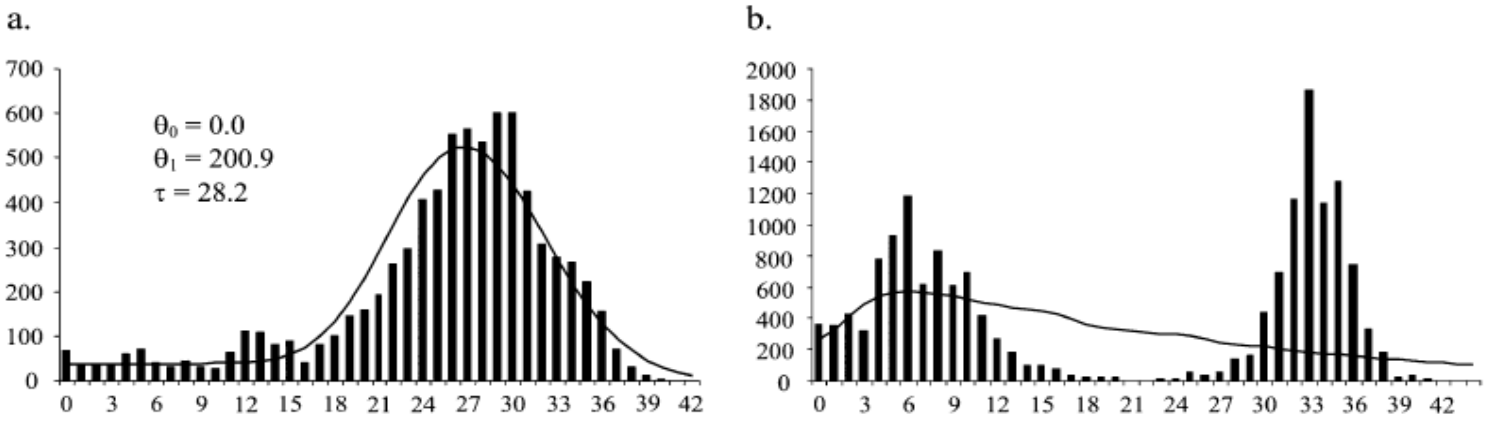
where $\theta_0 = 2N_0\mu$, $\theta_1 = 2N_1\mu$, $\tau = 2T\mu$, and $A = \theta_1 + M + 1$, and $C = e^{-\tau A/\theta_1}$.

In Arlequin, we estimate the three parameters of a spatial expansion, τ, $\theta = \theta_0 = \theta_1$ (here we assume that $N = N_0$), and $M = 2Nm$, using the same least-square method as described in the case of the estimation of the parameters of a demographic expansion (see section 8.1.2.4.1). Like for the demographic expansion, we also provide the expected mismatch distribution and test the fit to the model by coalescent simulations of an instantaneous expansion under the continent-island model defined above.

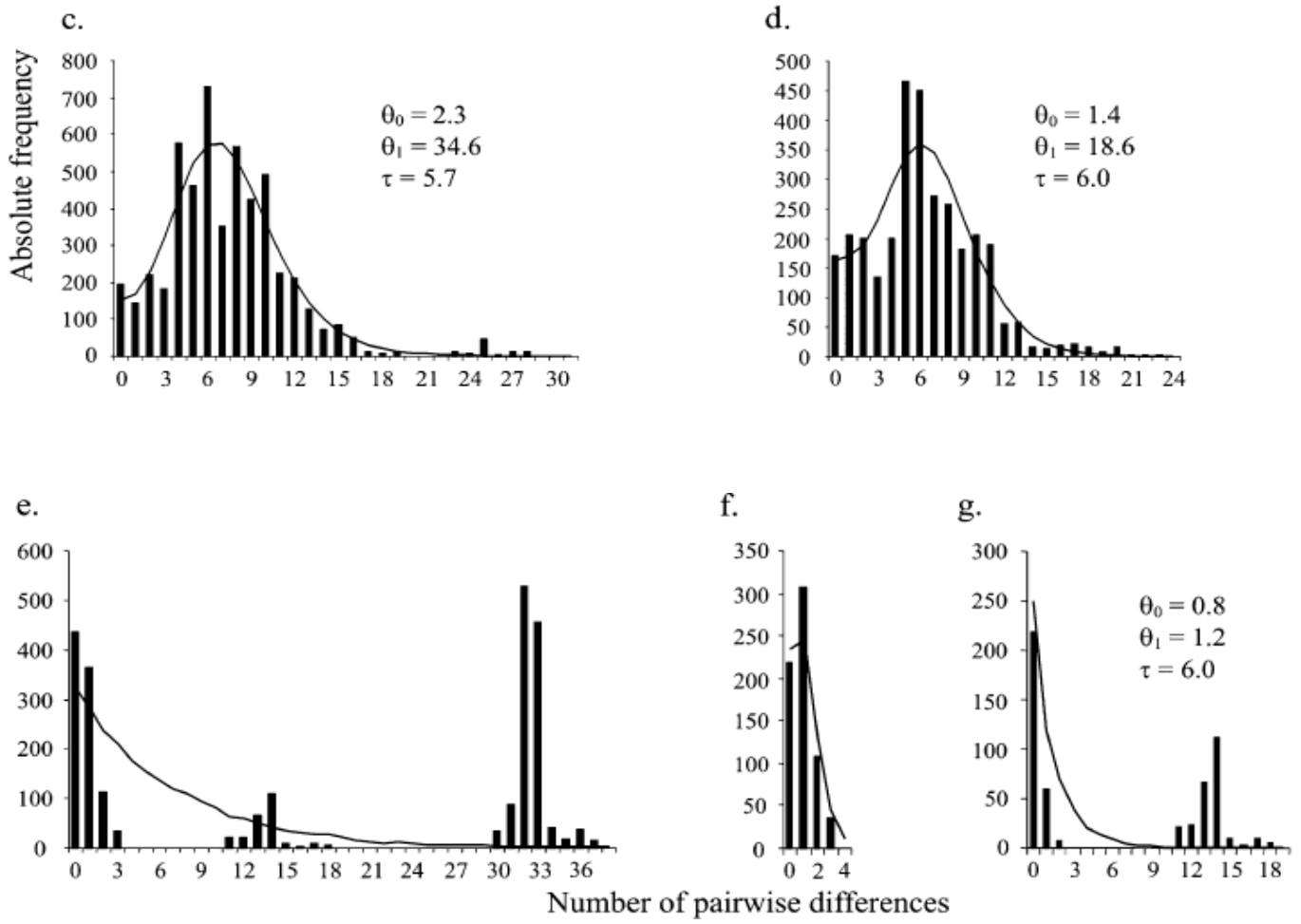
Molecular Ecology (2007) **16**, 605–618 doi: 10.1111/j.1365-294X.2006.03166.x

The rise and fall of the mountain hare (*Lepus timidus*) during Pleistocene glaciations: expansion and retreat with hybridization in the Iberian Peninsula

J. MELO-FERREIRA,*† P. BOURSOT,† E. RANDI,‡ A. KRYUKOV,§ F. SUCHENTRUNK,¶ N. FERRAND* and P. C. ALVES*



Mismatch distribution



Summary statistics and tests

Table 2 Estimates of sequence diversity, neutrality tests and growth rate in native *Lepus timidus* and in *Lepus granatensis*, *Lepus europaeus* and *Lepus castroviejoi* with *L. timidus* mtDNA haplotypes

Group	n_i	n_h	h	π (%)	θ(s) per site (%)	Tajima's D	Fu's F_s	Growth rate
Iberian species								
gra, eur and cas	254	77	0.974 (0.003)	1.9 (0.9)	1.7 (0.4)	—	—	—
gra	183	67	0.978 (0.003)	1.8 (0.9)	1.7 (0.4)	—	—	—
eur	70	11	0.820 (0.026)	1.7 (0.8)	1.0 (0.3)	—	—	—
gra, lineage A	103	34	0.963 (0.006)	0.7 (0.4)	1.2 (0.3)	−1.43	−7.95	152.9 (50.8)†
gra, lineage B	80	33	0.946 (0.013)	0.6 (0.3)	1.0 (0.3)	−1.30	−12.07*	232.2 (52.3)†
eur, lineage A	37	4	0.673 (0.050)	0.1 (0.1)	0.1 (0.1)	−0.05	0.44	611.4(1035.2)
eur, lineage B	33	7	0.587 (0.096)	0.6 (0.3)	0.5 (0.2)	0.31	4.71	−244.6 (108.9)
Native mountain hare								
Total	124	90	0.991 (0.003)	2.3 (1.1)	2.9 (0.7)	−0.70	−23.86*	203.5 (15.0)†
Northern Europe	47	39	0.987 (0.009)	2.0 (1.0)	2.5 (0.7)	−0.73	−11.29*	143.1 (22.0)†
Alps	47	24	0.955 (0.015)	1.9 (1.0)	1.6 (0.5)	0.70	0.82	23.4 (30.8)
Eastern Europe	4	4	1.000 (0.177)	1.6 (1.1)	1.6 (0.9)	−0.17	0.95	288.6 (65.1)†
Eastern Russia	26	23	0.991 (0.013)	2.1 (1.1)	2.3 (0.8)	−0.32	−4.35	236.2 (27.2)†

gra, *Lepus granatensis*; eur, *Lepus europaeus*; cas, *Lepus castroviejoi*; tim, *Lepus timidus*; n_i, number of analysed individuals; n_h, number of observed mtDNA haplotypes; h, haplotype diversity; π, nucleotide diversity; θ(s), computed from the number of segregating sites (Tajima 1983). Standard deviations (SD) are shown in brackets. The significant values are indicated by an asterisk. †indicates $g > 3$(SD).

Bayesian Skyline Plots

Bayesian Coalescent Approaches: Methods like Bayesian skyline plots are part of a broader category of Bayesian coalescent methods. Other methods include Bayesian Skygrid and Bayesian Skyride, which also aim to estimate changes in population size over time.
Extended Bayesian Skyline Plot (EBSP): This is an extension of the skyline plot that allows for more flexible demographic models. It may be useful when populations have complex demographic histories.
Sequentially Markov Coalescent (SMC) methods: These methods are designed for efficient inference of demographic history and genealogy. SMC methods, such as SMC' or MSMC, use sequential importance sampling to estimate population size changes over time.

•**Approximate Bayesian Computation (ABC):** ABC is a method that doesn't rely on explicit likelihood calculations. Instead, it simulates data under different models and compares the simulated data to the observed data. ABC has been used in population genetics for demographic inference

Didelot et. al 2021 MBE

Bayesian Skyline Plots

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 423–434 doi: 10.1111/j.1755-0998.2011.02988.x

INVITED TECHNICAL REVIEW

Skyline-plot methods for estimating demographic history from nucleotide sequences

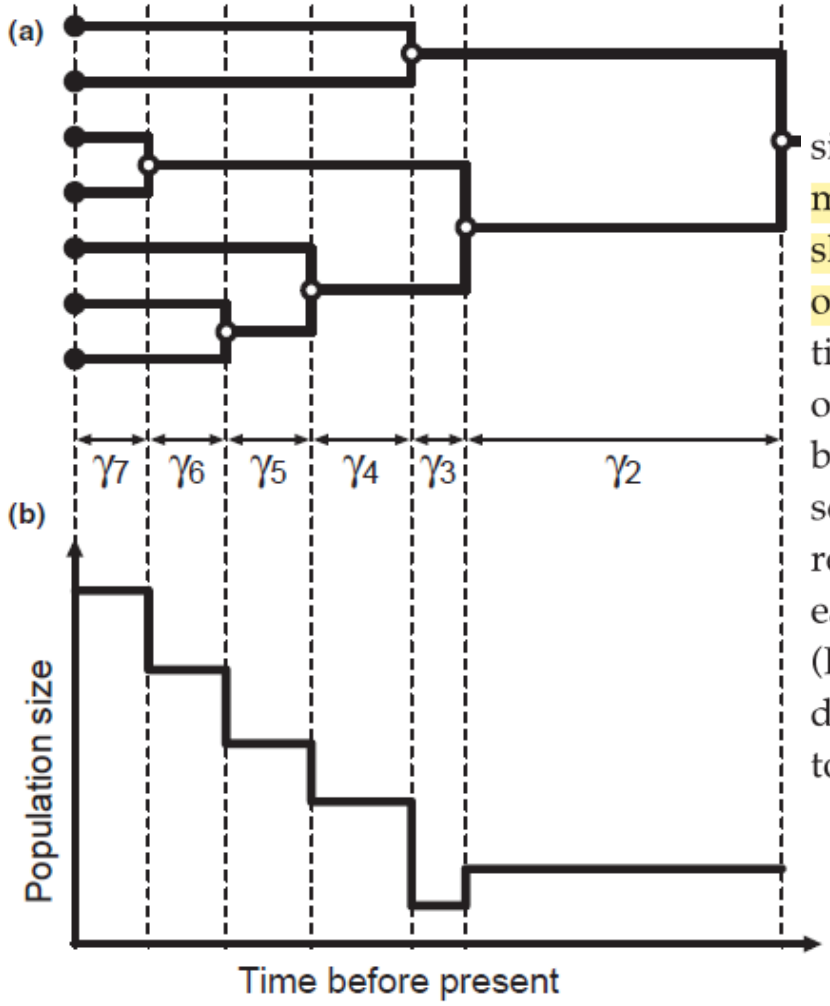
SIMON Y. W. HO*† and BETH SHAPIRO‡
*Centre for Macroevolution and Macroecology, Research School of Biology, Australian National University, ACT 0200, Australia, †School of Biological Sciences, University of Sydney, NSW 2006, Australia, ‡Department of Biology, The Pennsylvania State University, University Park, PA 16802–5301, USA

Reconstructing demographic history from a sequence alignment involves two distinguishable and separable steps:

(i) estimating the genealogy from the sequence data

(ii) estimating the population history based on the genealogy.

Bayesian Skyline Plots



size. To reconstruct demographic history, skyline-plot methods take advantage of a relatively simple relationship between the population size and the expected length of the coalescent interval. Specifically, the mean population size in each interval can be estimated by the product of the interval size (γ_i) and $i(i-1)/2$, where i is the number of genealogical lineages in the interval (Fig. 1a) (Hudson 1982; Kingman 1982b; Tajima 1983). Thus, this relationship gives an estimate of the population size for each coalescent interval in the estimated genealogy (Fig. 1b), producing a piecewise reconstruction of the demographic history that bears a superficial resemblance to the eponymous skyline of a city (Pybus *et al.* 2000). An

Ho & Shapiro 2011 MolEco

Bayesian Skyline Plots

Table 1 Comparison of skyline-plot methods for estimating demographic history from DNA sequence data

Method	Software	Estimation of coalescent error	Estimation of phylogenetic error	Able to analyse heterochronous sequences	Able to analyse multiple loci simultaneously	Reference
Classical skyline	GENIE, APE	No	No	Yes	No	Pybus *et al.* (2000)
Generalized skyline	GENIE, APE	No	No	No	No	Strimmer & Pybus (2001)
Bayesian MCP	APE	Yes	No	No	No	Opgen-Rhein *et al.* (2005)
Bayesian skyline	BEAST	Yes	Yes	Yes	No	Drummond *et al.* (2005)
Bayesian skyride	BEAST	Yes	Yes	Yes	No	Minin *et al.* (2008)
Extended Bayesian skyline	BEAST	Yes	Yes	Yes	Yes	Heled & Drummond (2008)

Ho & Shapiro 2011 MolEco

Bayesian Skyline Plots

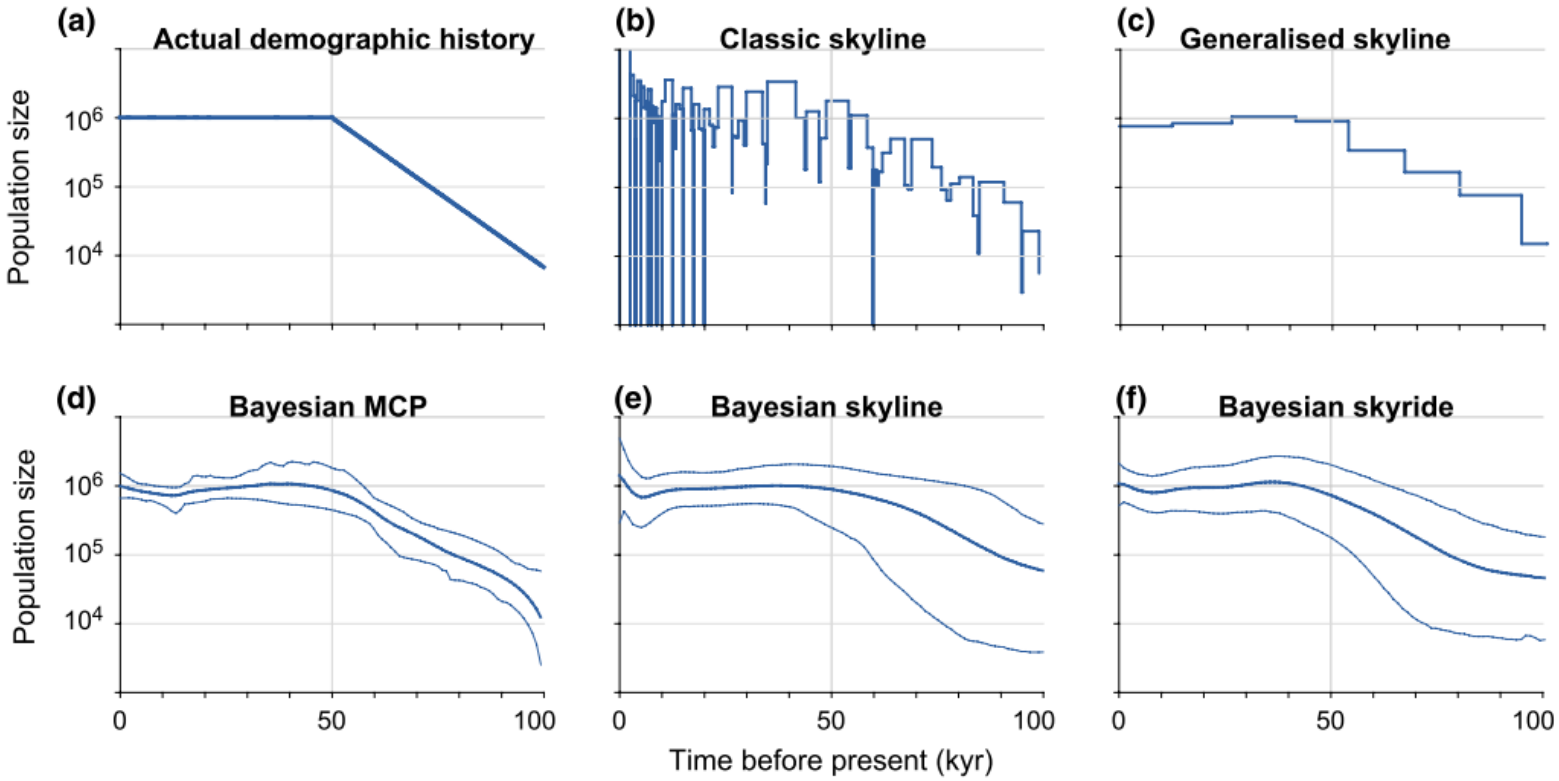


Fig. 2 Performance of different skyline-plot methods for a simulated data set. The actual demographic history used for simulation is shown in panel A, with the faint vertical line at 50 kyr indicating a change-point in the demographic function. The remaining panels show the reconstructions of demographic history by five skyline-plot methods. Note the logarithmic scale on the *y*-axis. Further details of the simulation model are given in the Appendix.

Ho & Shapiro 2011 MolEco

Bayesian Skyline Plots

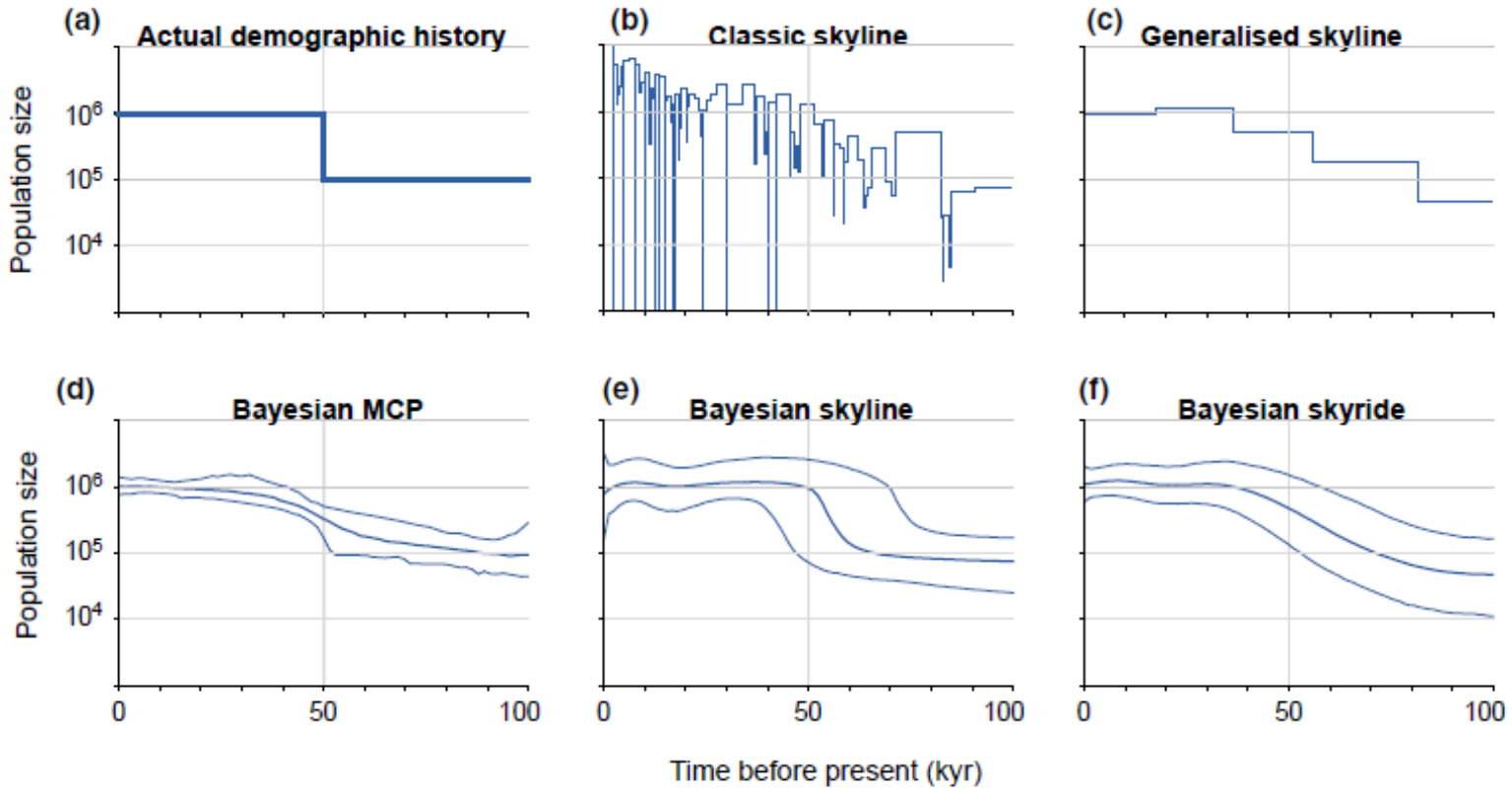


Fig. 3 Performance of different skyline-plot methods for a simulated data set. The actual demographic history used for simulation is shown in panel A, with the faint vertical line at 50 kyr indicating a change-point in the demographic function. The remaining panels show the reconstructions of demographic history by five skyline-plot methods. Note the logarithmic scale on the *y*-axis. Further details of the simulation model are given in the Appendix.

Ho & Shapiro 2011 MolEco

Bayesian Skyline Plots

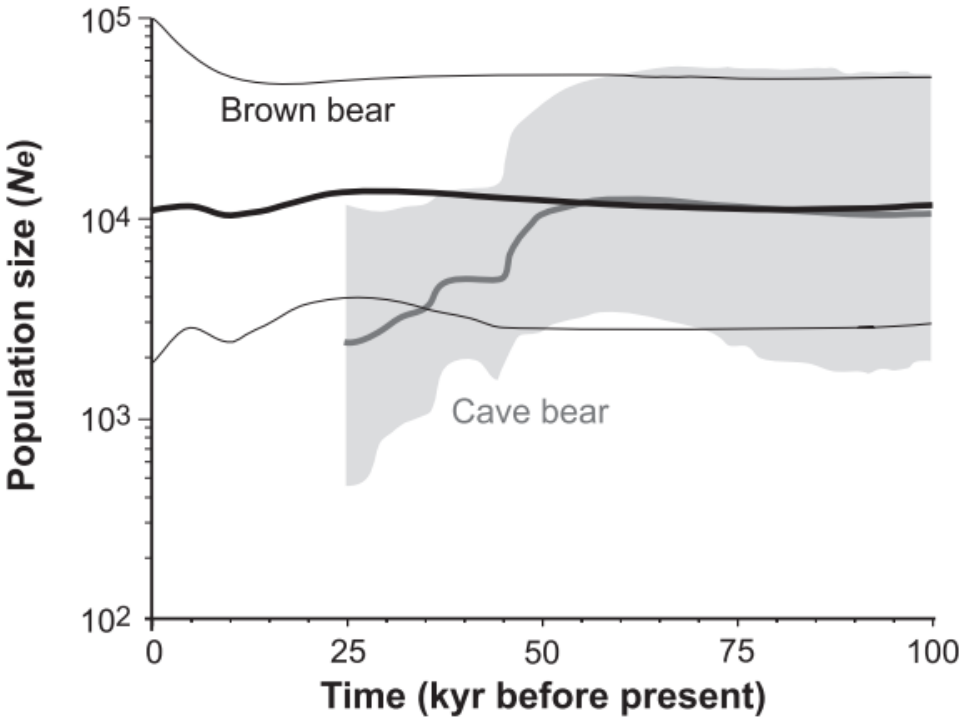


Fig. 4 Demographic histories of brown bear (*Ursus arctos*) and cave bear (*Ursus spelaeus*) estimated using the Bayesian skyline plot from mitochondrial D-loop sequences. The extinction of the cave bear is indicated by the termination of the skyline plot around 24 000 years before present. The figure is modified from Stiller *et al.* (2010).

Ho & Shapiro 2011 MolEco

Bayesian Skyline Plots

Table 2 Units of axes in skyline plots

Calibrating information		Units of x-axis in skyline plot and of branch lengths in the genealogy	Quantity measured on y-axis in skyline plot
Type	Units		
None	n/a	Mutations/site	Population size (N_e) × mutation rate (μ)
Mutation rate	Mutations/site/year	Years	Population size (N_e) × generation time* (τ)
	Mutations/site/generation	Generations	Population size (N_e)
Nodal age(s)	Years	Years	Population size (N_e) × generation time* (τ)
	Generations	Generations	Population size (N_e)

*Generation time measured in years.

Ho & Shapiro 2011 MolEco

Rapid ecological specialization despite constant population sizes

Andrinajoro R. Rakotoarivelo[1,2], Paul O'Donoghue[3], Michael W. Bruford[4] and Yoshan Moodley[1]

[1] Department of Zoology, University of Venda, Thohoyandou, Limpopo, Republic of South Africa
[2] Natiora Ahy, Antananarivo, Madagascar
[3] Specialist Wildlife Services, Specialist Wildlife Services, St Asaph, United Kingdom
[4] Cardiff School of Biosciences, Cardiff University, Cardiff, United Kingdom

ABSTRACT

Background. The bushbuck, *Tragelaphus scriptus*, is a widespread and ecologically diverse ungulate species complex within the spiral-horned antelopes. This species was recently found to consist of two genetically divergent but monophyletic lineages, which are paraphyletic at mitochondrial (mt)DNA owing to an ancient interspecific hybridization event. The Scriptus lineage (*T. s. scriptus*) inhabits the north-western half of the African continent while Sylvaticus (*T. s. sylvaticus*) is found in the south-eastern half. Here we test hypotheses of historical demography and adaptation in bushbuck using a higher-resolution framework, with four nuclear (MGF, PRKCI, SPTBN, and THY) and three new mitochondrial markers (cytochrome b, 12S rRNA, and 16S rRNA).

PeerJ

Rapid ecological specialization despite constant population sizes

Andrinajoro R. Rakotoarivelo[1,2], Paul O'Donoghue[3], Michael W. Bruford[4] and Yoshan Moodley[1]

[1] Department of Zoology, University of Venda, Thohoyandou, Limpopo, Republic of South Africa
[2] Natiora Ahy, Antananarivo, Madagascar
[3] Specialist Wildlife Services, Specialist Wildlife Services, St Asaph, United Kingdom
[4] Cardiff School of Biosciences, Cardiff University, Cardiff, United Kingdom

ABSTRACT

Background. The b ad and ecologically diverse ungulate spe elopes. This species was recently found t onophyletic lineages, which are paraphyle ancient interspecific hybridization event. he north-western half of the African contin l in the south-eastern half. Here we test hy ptation in bushbuck using a higher-resol PRKCI, SPTBN, and THY) and three new RNA, and 16S rRNA).

Peer

spite

. Bruford[4] and

ıth Africa

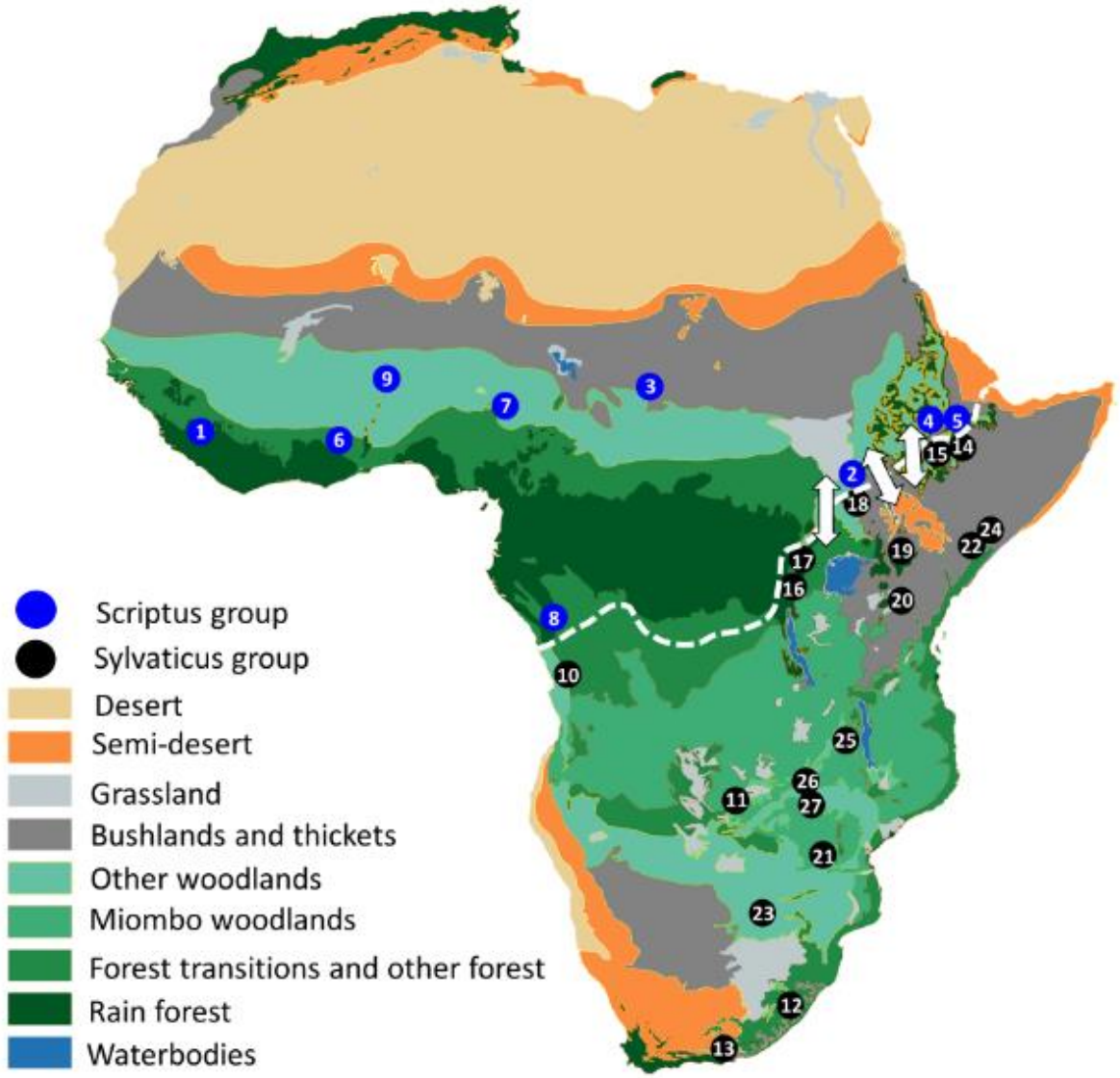


Figure 1 The land cover of Africa reconstructed from remotely sensed data (redrawn from *Mayaux et al., 2004*). The geographical distribution of sampling localities included in the present study are shown on the map. Taxa are plotted as dots and designated either blue for *Scriptus* or black for *Sylvaticus*. Samples are numbered according to Table 1. A dashed white line divides the distributions of both groups and white arrows show zones of potential gene flow.

Full-size DOI: 10.7717/peerj.6476/fig-1

MBE

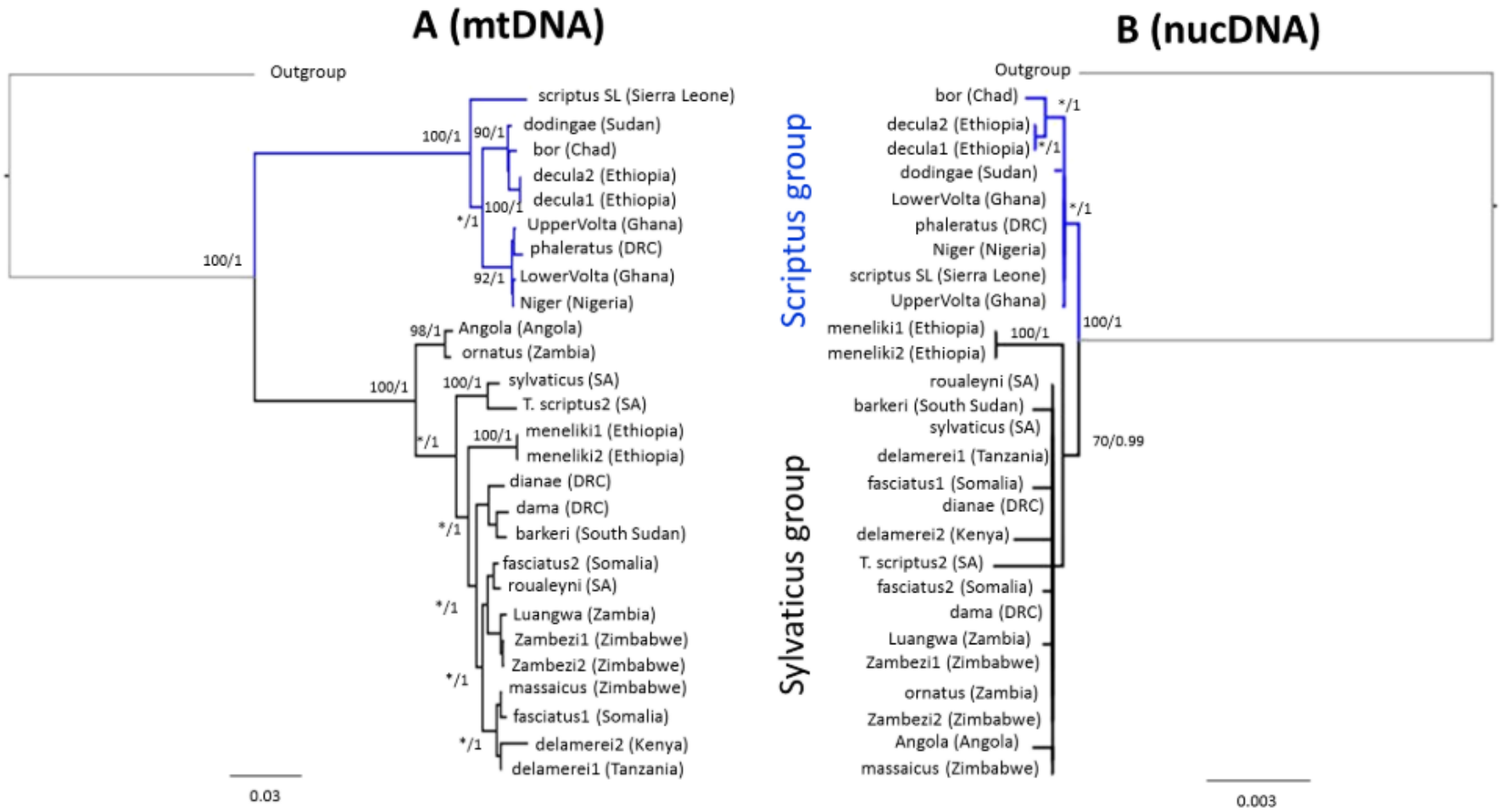


Figure 2 Tree topologies based on maximum likelihood retrieved from (A) the combined mtDNA data and (B) the combined nucDNA data. Values given above the branches represent maximum likelihood bootstrap values and maximum clade probabilities.

Full-size DOI: 10.7717/peerj.6476/fig-2

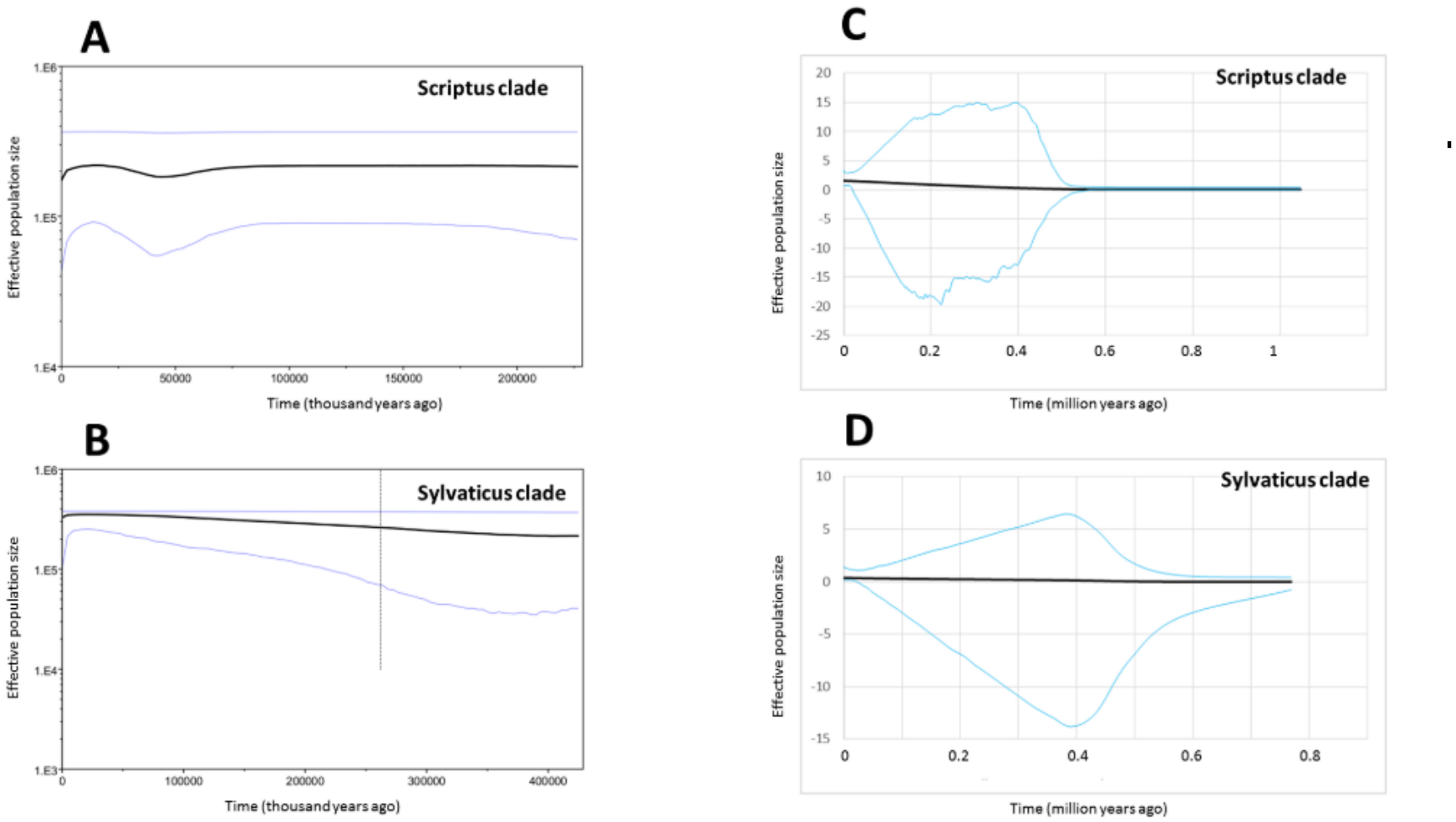


Figure 4 Bayesian Skyline Plots (BSPs) and Extended Bayesian Skyline Plots (EBSPs). (A-B) BSPs represent population size changes over time, inferred with mtDNA and an assumed divergence rate of 0.056 per million years. The *X*-axes are time in thousands of years. *Y*-axes are mean effective population sizes log-scale. Solid black lines represent median height and areas between blue lines encompass the 95% highest posterior density (HPD). (C-D) EBSPs represent population size changes over time in two of the mtDNA clades, inferred by mtDNA and nDNA. *X*-axes are time in millions of years, *Y*-axes are effective population size divided by generation time.

Full-size DOI: 10.7717/peerj.6476/fig-4

Bayesian Skyline Plots

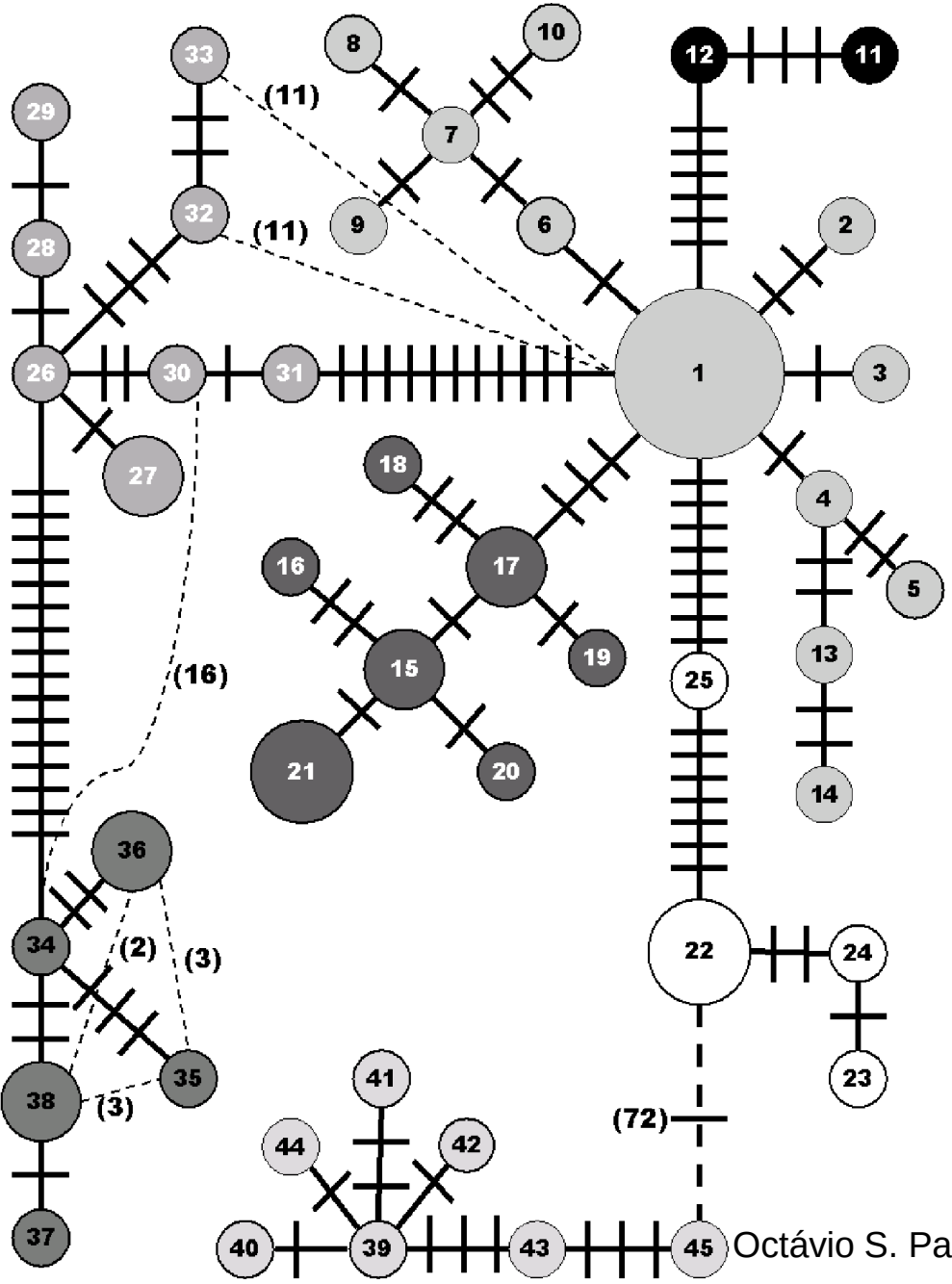
Molecular Ecology (2008) **17**, 1535–1551 doi: 10.1111/j.1365-294X.2008.03706.x

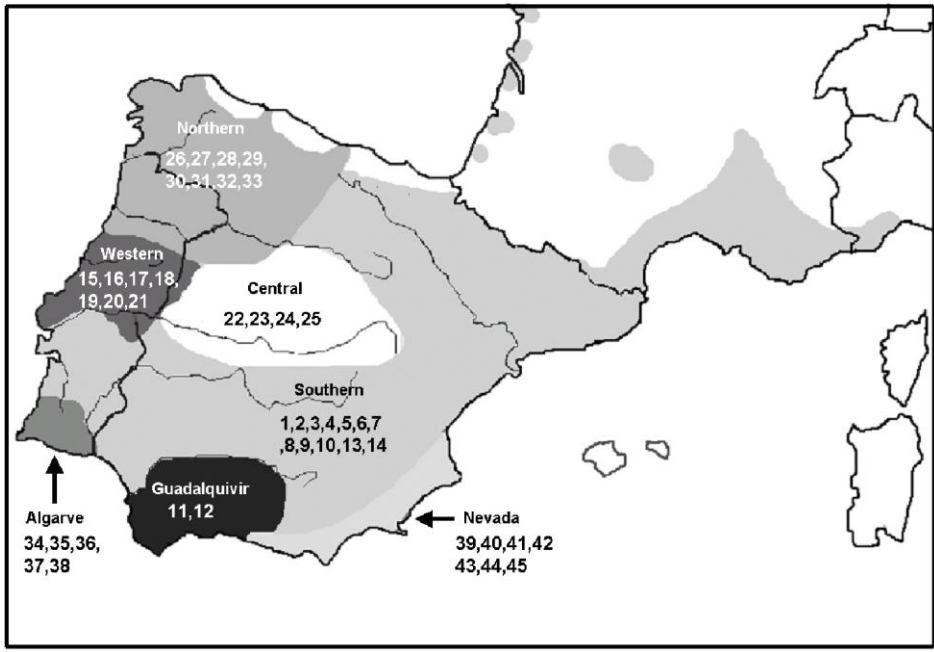
The role of vicariance vs. dispersal in shaping genetic patterns in ocellated lizard species in the western Mediterranean

O. S. PAULO,* J. PINHEIRO,* A. MIRALDO,* M. W. BRUFORD,† W. C. JORDAN‡ and R. A. NICHOLS§
*Centro de Biologia Ambiental/Departamento de Biologia Animal, Faculdade de Ciências da Universidade de Lisboa, P-1749–016 Lisboa, Portugal, †Cardiff School of Biosciences, Cardiff University, Cathays Park, Cardiff CF10 3TL, UK, ‡Institute of Zoology, Zoological Society of London, London NW1 4RY, UK, §School of Biological Sciences, Queen Mary University of London, London E1 4NS, UK

Table 3 Divergence time estimated in million years ago (Ma) for the cytochrome *b* mtDNA gene of ocellated lizard species, for the main five nodes recovered in the phylogenetic analyses (Fig. 3). Divergence times were estimated under three different assumptions of pairwise sequence divergence (2%, 1.7% and 2.5%). For each node, the mean divergence time and the lower and upper values of the highest probability distribution (HPD) are presented under the three different rates and the respective combined value (see Materials and methods section for details)

		2.0%			1.7%			2.5%			Combined		
Node	Clades	Mean	HPD lower	HPD upper	Mean	HPD lower	HPD upper	Mean	HPD lower	HPD upper	Combined	HPD lower	HPD upper
A	Africa/Europe	11.38	7.49	15.48	13.34	9.03	18.54	9.28	6.59	12.56	11.33	6.43	16.19
B	Betic/Iberia	9.64	5.06	14.85	11.02	5.77	16.65	7.61	4.03	11.66	9.43	4.50	15.14
C	Tunisia/Morocco	8.81	5.35	12.00	10.48	5.96	14.80	7.43	4.57	10.53	8.91	4.73	13.13
D	Rif/Atlas	5.43	3.08	7.83	6.44	3.61	9.40	4.54	2.63	6.79	5.47	2.65	8.18
	Clade L splits	1.93	1.28	2.68	2.39	1.50	3.50	1.58	1.00	2.25	1.97	1.05	3.01


33
(11)
8
10
12
11
29
7
32
(11)
9
6
2
28
26
30
31
1
3
27
18
4
16
17
5
(16)
15
19
13
25
21
20
14
36
34
(2)
(3)
22
24
35
23
38
(3)
41
(72)
44
42
37
40
39
43
45


Northern
26,27,28,29,
30,31,32,33
Western
15,16,17,18,
19,20,21
Central
22,23,24,25
Southern
1,2,3,4,5,6,7
,8,9,10,13,14
Guadalquivir
11,12
Algarve
34,35,36,
37,38
Nevada
39,40,41,42
43,44,45

Miraldo *et al. BMC Evolutionary Biology* 2011, **11**:170
http://www.biomedcentral.com/1471-2148/11/170

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones

Andreia Miraldo[1,3*], Godfrey M Hewitt[1], Octavio S Paulo[2] and Brent C Emerson[1,4]

Iberia

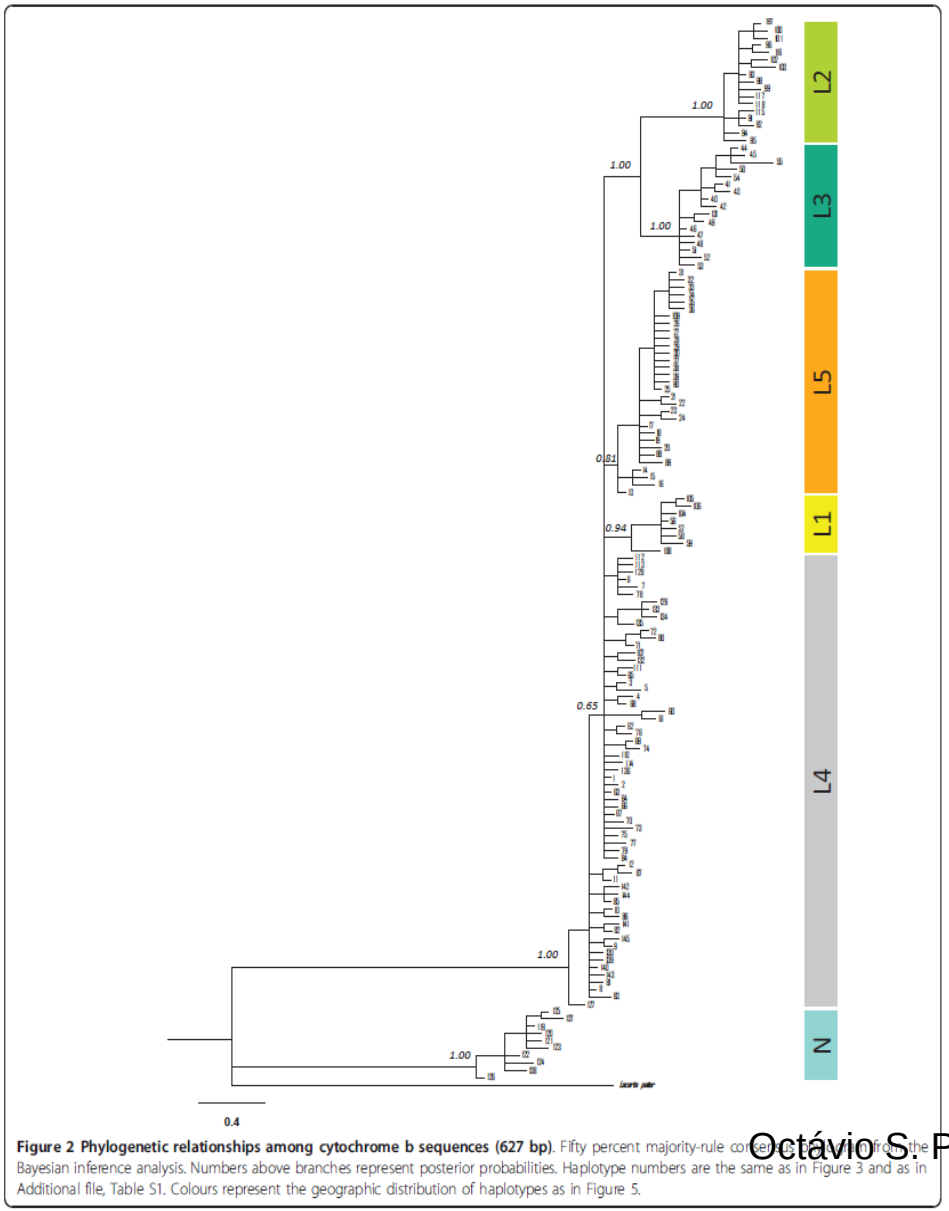
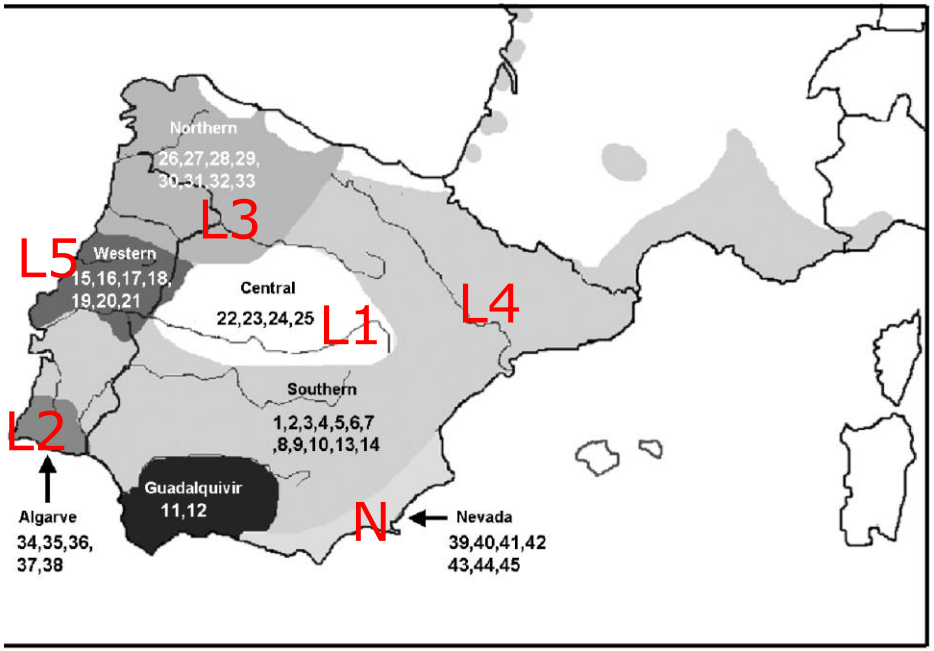


Figure 2 Phylogenetic relationships among cytochrome b sequences (627 bp). Fifty percent majority-rule consensus tree obtained from the Bayesian inference analysis. Numbers above branches represent posterior probabilities. Haplotype numbers are the same as in Figure 3 and as in Additional file, Table S1. Colours represent the geographic distribution of haplotypes as in Figure 5.



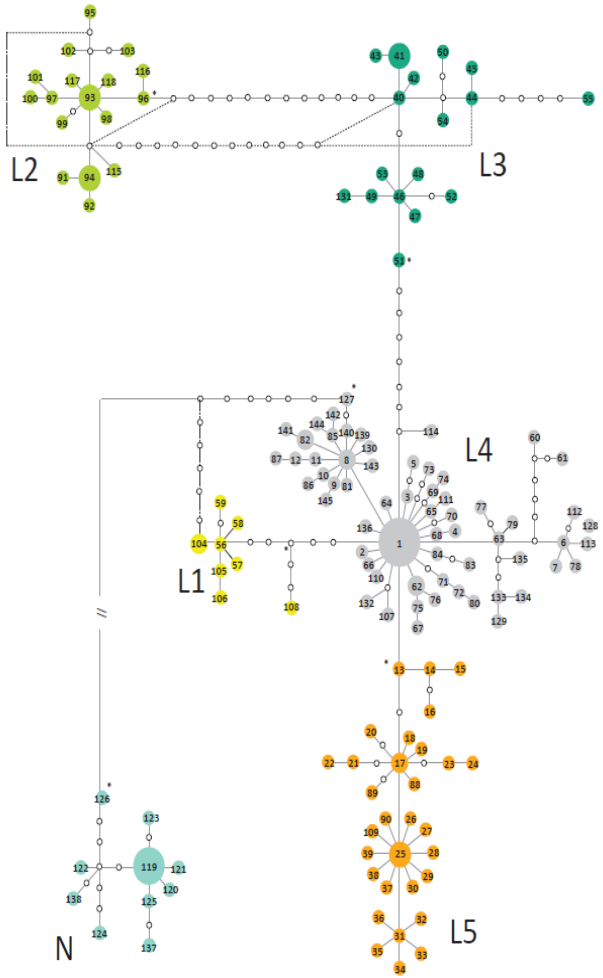


Figure 3 Statistical Parsimony network of *Lacerta lepida* cytochrome b haplotypes from 312 samples. Open circles with no numbers represent unsampled or extinct haplotypes. L1, L2, L3, L4, L5 and N represent different mitochondrial phylogroups. The ancestral haplotype within each phylogroup is marked with an asterisk. Phylogroup N connects to the main network through 65 mutations, represented by an interrupted line.

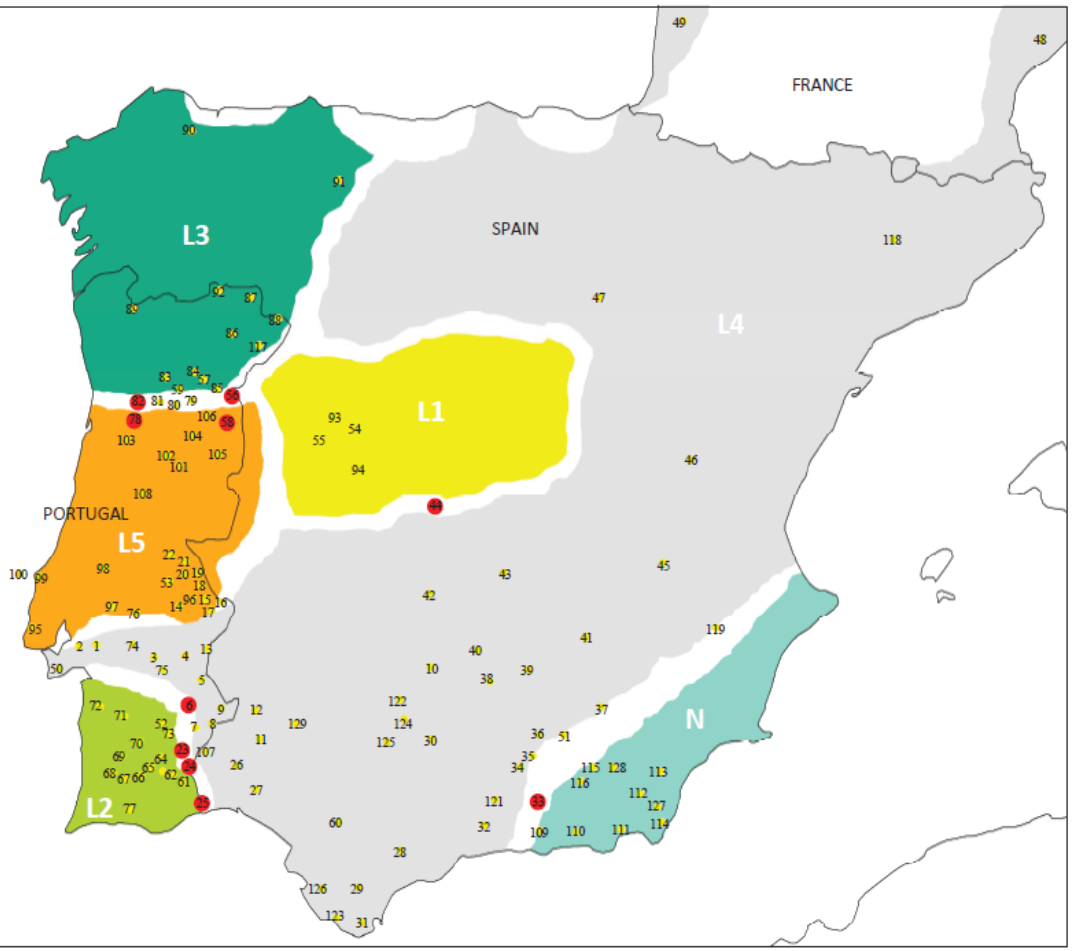


Figure 4 Distribution of *Lacerta lepida* mitochondrial phylogroups based on 627 bp of the cytochrome b gene. Colours are the same as in Figures 2 and 3. Filled red circles represent populations where divergent haplotypes from two or more phylogroups were detected in sympatry. Numbers correspond to sampling localities as in Figure 1 and Additional file 1, Table S1.

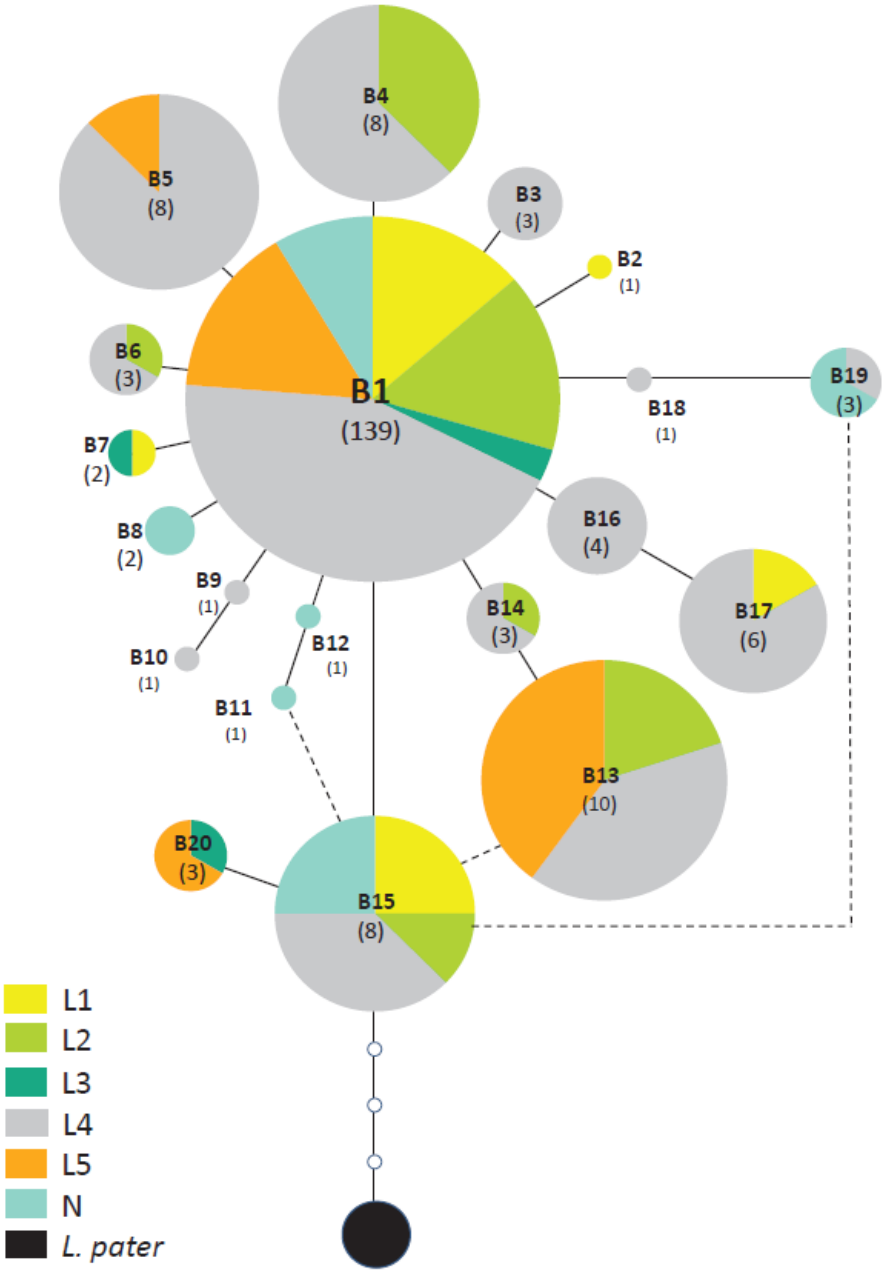


Figure 5 Statistical Parsimony network of *Lacerta lepida* β-Fibrinogen intron 7 alleles from 104 samples. Open circles with no numbers represent unsampled or extinct alleles, and the filled black circle represents the outgroup taxon *L. pater*. Pie charts are scaled to represent the proportion of each allele found within each mitochondrial phylogroup. Colours in pie charts are the same as those used to represent mitochondrial phylogroups in Figure 3. Dashed lines represent ambiguities in the network.

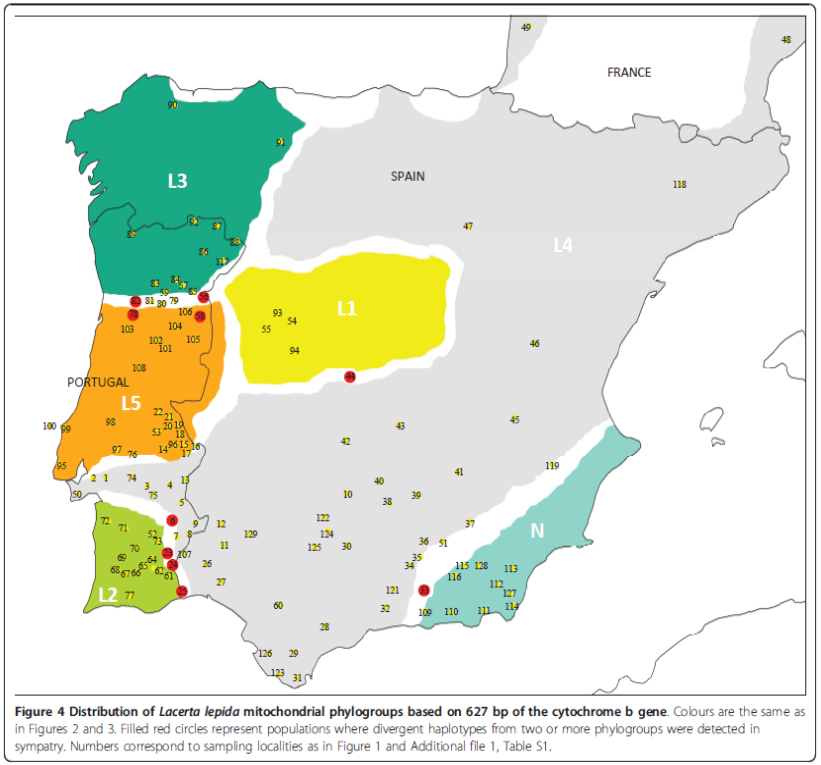


Figure 4 Distribution of *Lacerta lepida* mitochondrial phylogroups based on 627 bp of the cytochrome b gene. Colours are the same as in Figures 2 and 3. Filled red circles represent populations where divergent haplotypes from two or more phylogroups were detected in sympatry. Numbers correspond to sampling localities as in Figure 1 and Additional file 1, Table S1.

Iberia

Table 2 Results from mismatch distribution and neutrality tests for *cytb* mtDNA phylogroups and for the *β-fibint7* nuclear gene

				Mismatch Distribution				Neutrality tests					
		Spatial genetic structure		Sudden-expansion model		Spatial-expansion model							
Locus	Phylogroup	χ^2	p	p (SDD)	p (hg)	p (SDD)	p (hg)	*D*	*p*	*Fs*	p	*R2*	p
	L1	73.78	<0.01	0.49	0.67	0.34	0.63	-1.78	0.04	-2.67	0.04	0.12	**0.18**
	L2	358.07	*0.24*	0.15	0.29	0.31	0.36	-1.72	0.02	-9.92	0.00	0.05	0.00
Cytb	L3	754.10	<0.01	0.13	0.22	0.28	0.46	-1.18	**0.10**	-7.06	0.00	0.07	**0.09**
	L4	3494.93	*0.09*	0.92	0.67	0.95	0.70	-2.35	0.00	-26.33	0.00	0.02	0.00
	L5	946.18	<0.01	0.10	0.09	0.33	0.33	-2.17	0.00	-26.66	0.00	0.03	0.00
	N	105.14	0.03	0.68	0.78	0.86	0.87	-2.00	0.01	-3.90	0.01	0.06	0.00
β-Fib [(a)]	...	148.25	0.02	0.76	0.96	0.58	0.85	-1.69	0.01	-18.52	0.00	0.04	0.02
β-Fib [(b)]	...	-	-	0.74	0.78	0.60	0.81	-1.73	0.01	-18.67	0.00	0.08	0.00

(p(SDD) = sum of square deviations; p(hg) = Harpending's raggedness index; Tajima's D (D) and respective p value; Fu's Fs test (Fs) and respective p value; Ramos-Onsis R2 (R2) and respective p value). Statistics that do not suggest range expansion are shown in bold font. Results for the spatial genetic structure estimated with Geodis are also shown. Statistics that do not show evidence for spatial genetic structure are shown in italic font.

(a) Dataset using haplotypes inferred by Phase with probability threshold > 0.60 (208 haplotypes)

(b) Reduced dataset using haplotypes inferred by Phase with probability threshold > 0.90 (192 haplotypes)

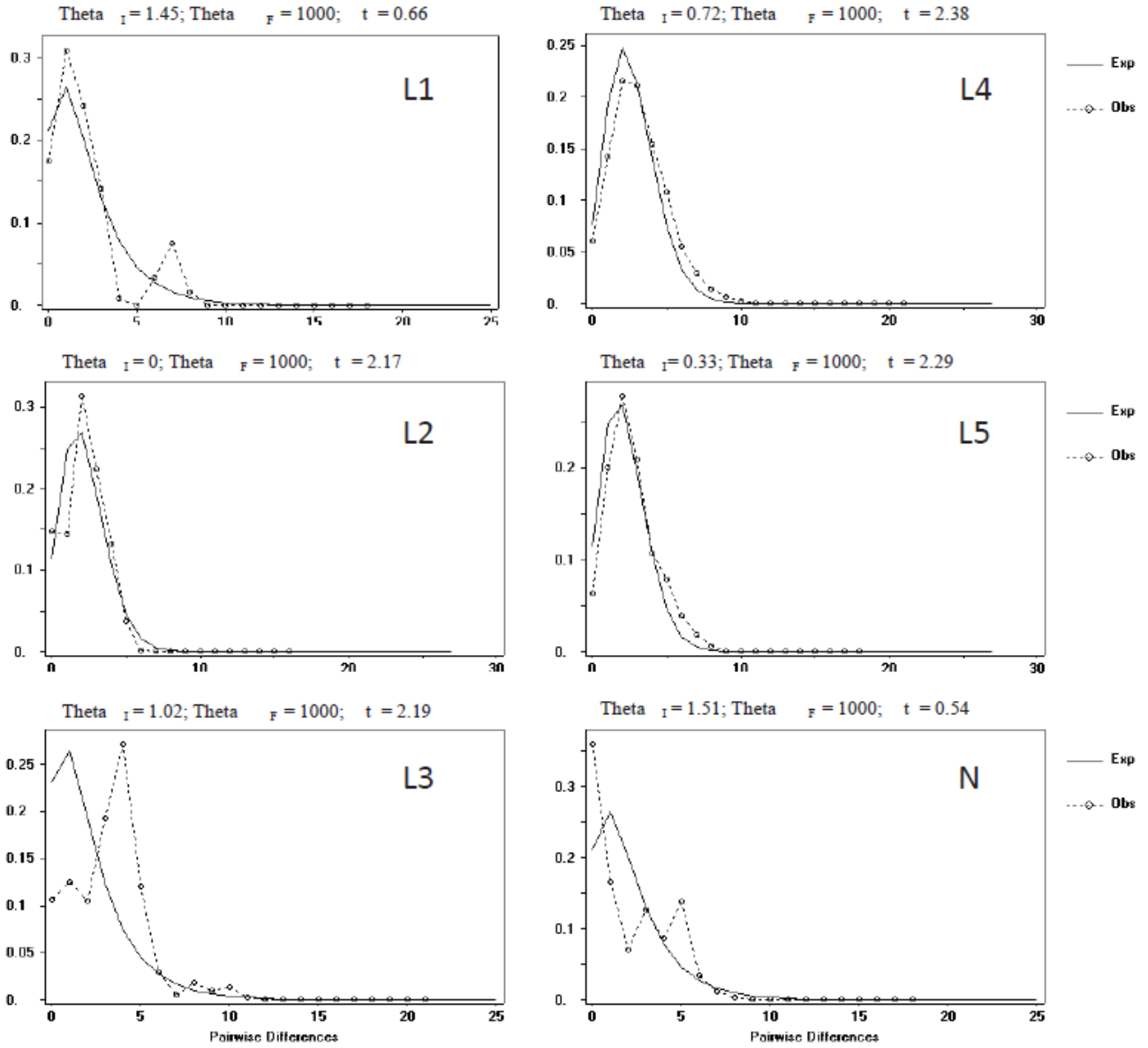


Figure 6 Mismatch distribution of mtDNA haplotypes for each of the 6 *Lacerta lepida* phylogroups. The expected frequency is based on a population growth-decline model, determined using DnaSP v4.50 [69] and is represented by a continuous line. The observed frequency is represented by a dotted line.

Iberia

Table 3 Divergence time estimates in million years (Ma) from the most recent common ancestor (mrca) of all group members from each *Lacerta lepida* phylogroup estimated using the method of Saillard *et al.* (2000) using three different mutation rates; and divergence time estimates for the main nodes recovered in the phylogenetic analysis using a mean mutation rate of 2% in Beast

	Saillard			Beast		
Mutation rate	1%	0.85%	1.15%		1%[(a)]	
mtDNA phylogroups		(mean ± s.d.)		Lower HPD	Mean	Upper HPD
L1	0.64 ± 0.12	0.75 ± 0.14	0.55 ± 0.10	0.28	0.76	1.32
L2	0.45 ± 0.21	0.53 ± 0.24	0.39 ± 0.16	0.21	0.47	0.78
L3	0.59 ± 0.38	0.63 ± 0.40	0.47 ± 0.27	0.32	0.68	1.10
L4	0.92 ± 0.30	1.08 ± 0.35	0.80 ± 0.24	n.a	n.a	n.a
L5	0.59 ± 0.19	0.70 ± 0.22	0.51 ± 0.15	0.29	0.61	0.98
N	0.85 ± 0.34	0.99 ± 0.41	0.74 ± 0.28	0.29	1.04	0.63
L2+L3	n.a	n.a	n.a	0.82	1.50	2.27
L1+L2+L3+L4+L5	n.a	n.a	n.a	1.13	1.96	2.91
All (L+N)	n.a	n.a	n.a	5.58	9.43	13.66

For each node the mean divergence time and the lower and upper values of the highest probability distribution (HPD) are presented. (see text for a detailed explanation of each method).

(a) Divergence times were calculated using mutations rates drawn from a normal distribution with a mean mutation rate of 0.01 mutations/site/million years and a standard deviation of 0.0015.

n.a: not applicable

Bayesian Skyline Plots

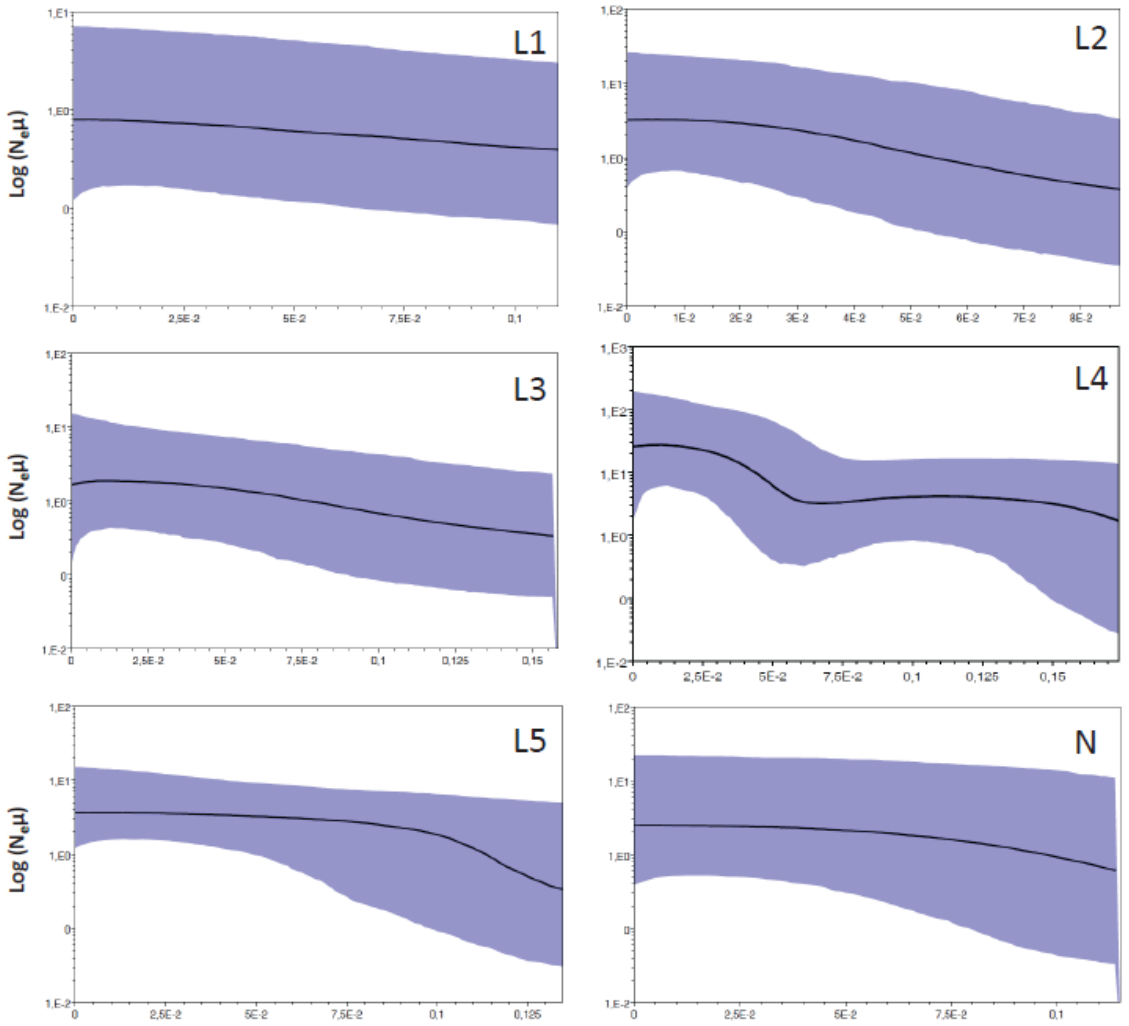


Figure 7 Bayesian skyline plots showing the historical demographic trends for each *Lacerta lepida* mitochondrial phylogroup detected using cytochrome b sequences. Along the y-axis is the expressed population size estimated in units of $N_e\mu$ (N_e: effective population size, μ: mutation rate per haplotype per generation). The y-axis is in a log-scale. Solid lines represent median estimates and shaded areas represent confidence intervals.

OPEN ACCESS Freely available online

PLOS ONE

New Mitochondrial and Nuclear Evidences Support Recent Demographic Expansion and an Atypical Phylogeographic Pattern in the Spittlebug *Philaenus spumarius* (Hemiptera, Aphrophoridae)

Ana S. B. Rodrigues[1]*, Sara E. Silva[1], Eduardo Marabuto[1], Diogo N. Silva[1], Mike R. Wilson[2], Vinton Thompson[3], Selçuk Yurtsever[4], Antti Halkka[5], Paulo A. V. Borges[6], José A. Quartau[1], Octávio S. Paulo[1], Sofia G. Seabra[1]

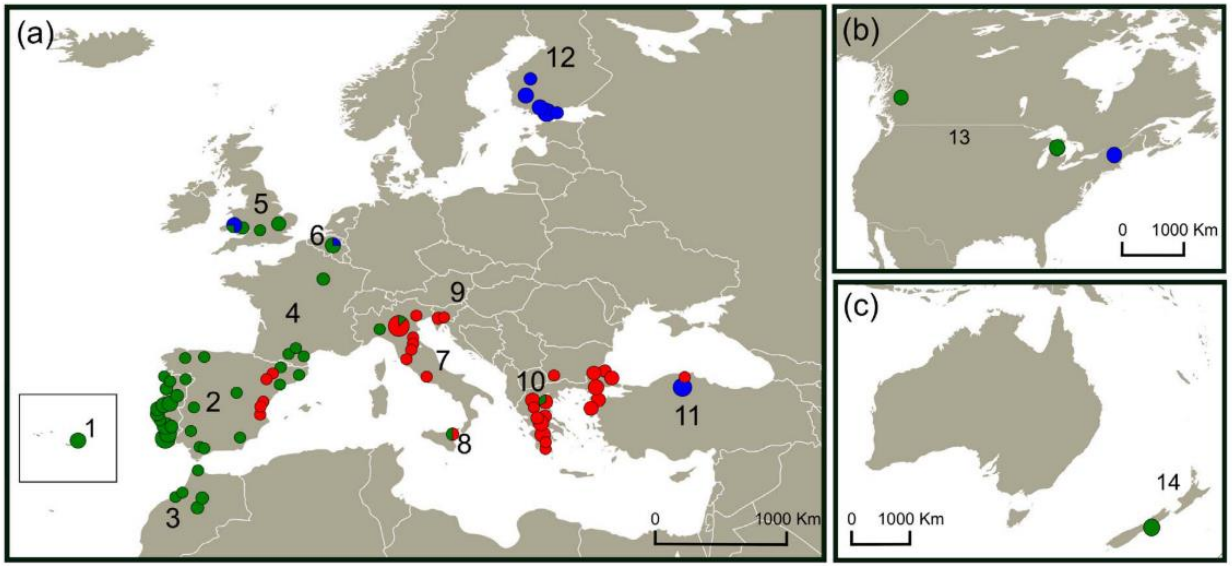
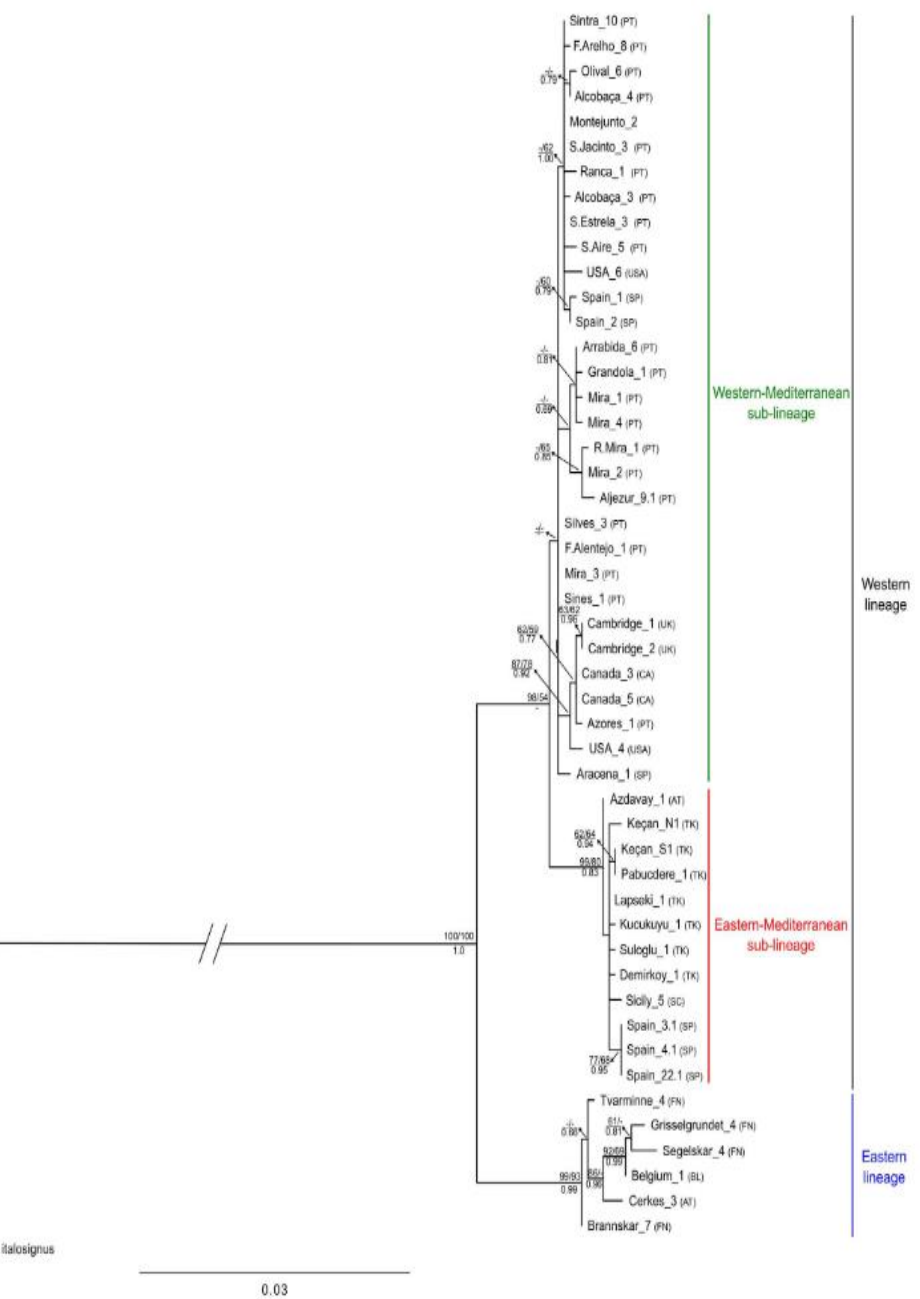


Figure 1. Sampling locations of *Philaenus spumarius* in (a) Europe and Anatolia (b) North America and (c) New Zealand in each geographic region. 1 – Azores; 2 – Iberian Peninsula; 3 – Morocco; 4 – France; 5 – United Kingdom; 6 – Belgium; 7 – Italian Peninsula; 8 – Sicily; 9 – Slovenia; 10 – Balkans (Bulgaria; Greece; European Turkey); 11 – Anatolian Peninsula; 12 – Finland; 13 – North America (Canada and United States of America); 14 – New Zealand. Circle sizes are proportional to the number of individuals. Circles: green – ''western-Mediterranean'' mtDNA group; red – ''eastern-Mediterranean'' mtDNA group; blue – ''eastern'' mtDNA group. Circle sizes are proportional to the number of samples.
doi:10.1371/journal.pone.0098375.g001



e 2. Maximum Likelihood tree based on the 3 concatenated mtDNA genes (COI, COII and cyt *b*) (1527bp). Values above branches pond to MP and ML bootstrap values (only values > 50% are shown) and values below branches correspond to Bayesian posterior probability. ortugal; SP – Spain; UK – United Kingdom; BL – Belgium; FN – Finland; SC – Sicily; TK – European Turkey; AT – Anatolia; USA – United States of ca; CA – Canada.
.1371/journal.pone.0098375.g002

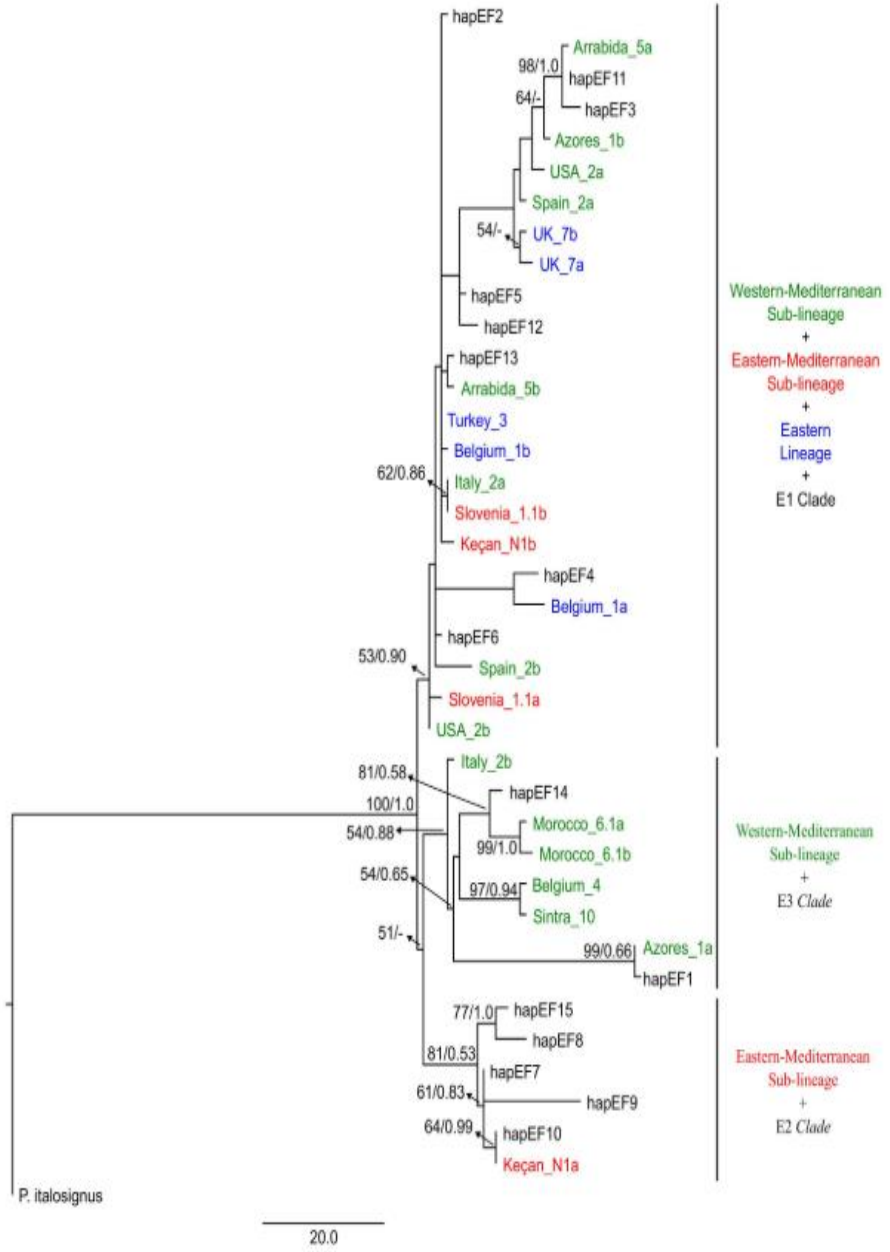


Figure 4. Maximum Parsimony tree based on nuclear gene elongation factor-1α. Values above branches correspond to MP bootstrap (only values > 50% are shown) and Bayesian posterior probability values. Black: GenBank sequences (see [26]); blue individuals correspond to the eastern mtDNA group; red individuals correspond to the eastern-Mediterranean mtDNA group and green individuals to the western-Mediterranean mtDNA group.

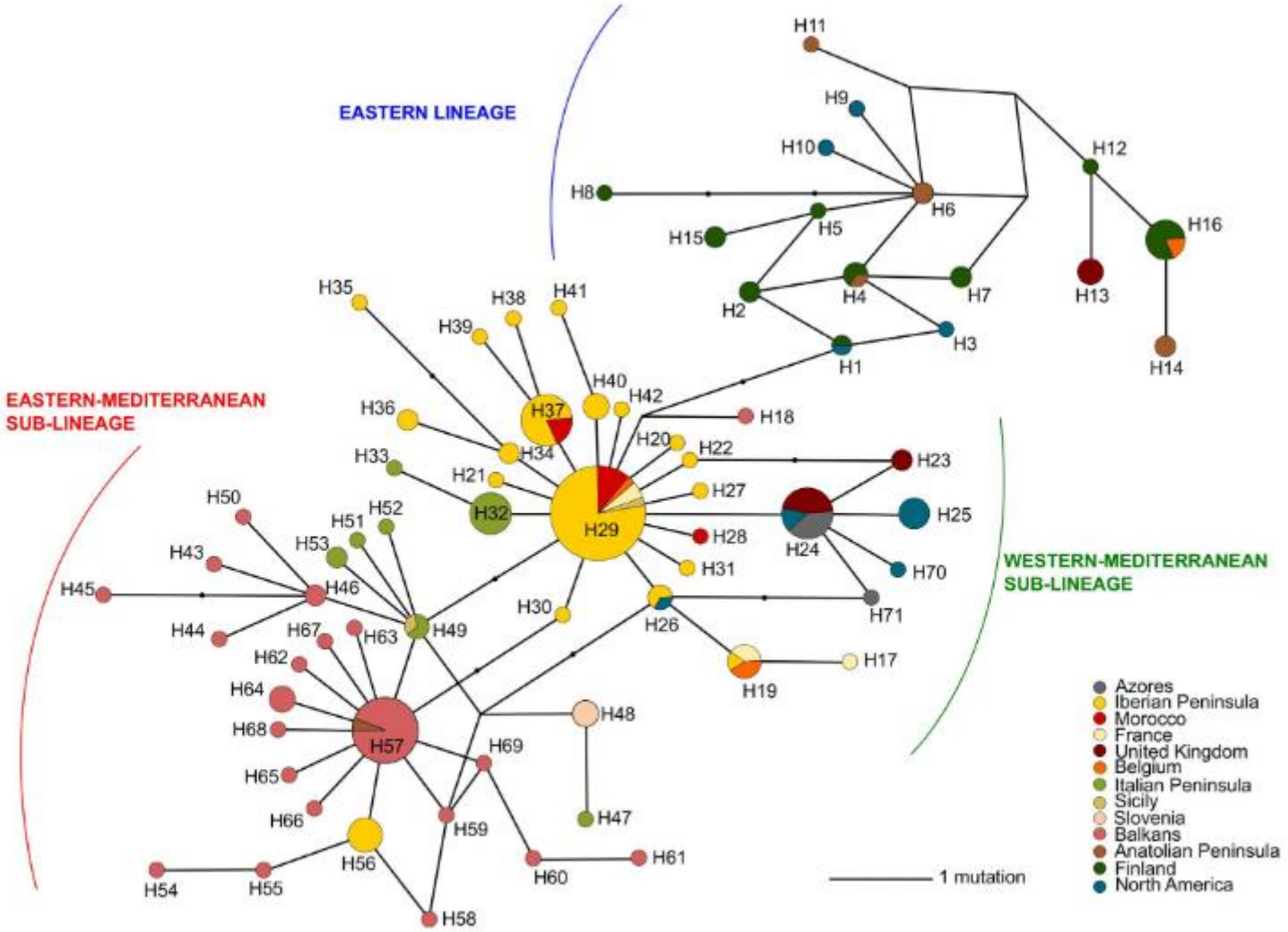


Figure 3. Median-joining haplotype network of *Philaenus spumarius* sampled geographic regions for mitochondrial gene COI (539bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their sizes are in proportion to the number of mutations.
doi:10.1371/journal.pone.0098375.g003

Haplotype Diversity and Nucleotide Diversity

Table 1. Number of individuals, number of haplotypes and genetic diversity indices calculated for geographic regions of *Philaenus spumarius* and for mitochondrial gene Cytochrome *c* oxidase I (COI).

Geographic regions	Number of individuals	Number of haplotypes	Haplotype diversity (*h*)	Nucleotide diversity (π)
Morocco	7	3	0.6667 +/− 0.1598	0.001414 +/− 0.001338
Iberian Peninsula	63	19	0.7798 +/− 0.0493	0.003133 +/− 0.002062
Azores	5	2	0.4000 +/− 0.2373	0.000742 +/− 0.000944
Western Europe	9	4	0.7500 +/− 0.1121	0.006597 +/− 0.004194
Slovenia	3	1	0.0000 +/− 0.0000	0.000000 +/− 0.000000
Italy	17	8	0.8162 +/− 0.0815	0.004666 +/− 0.002957
Balkans	40	21	0.9128 +/− 0.0303	0.003825 +/− 0.002431
Anatolian Peninsula	7	5	0.9048 +/− 0.1033	0.009895 +/− 0.006223
United Kingdom	10	3	0.6889 +/− 0.1038	0.009318 +/− 0.005592
Finland	17	9	0.8971 +/− 0.0534	0.006603 +/− 0.003950
North America	12	8	0.8939 +/− 0.0777	0.008236 +/− 0.004917

Western Europe: Belgium and France; Italy: Italian peninsula and Sicily.
doi:10.1371/journal.pone.0098375.t001

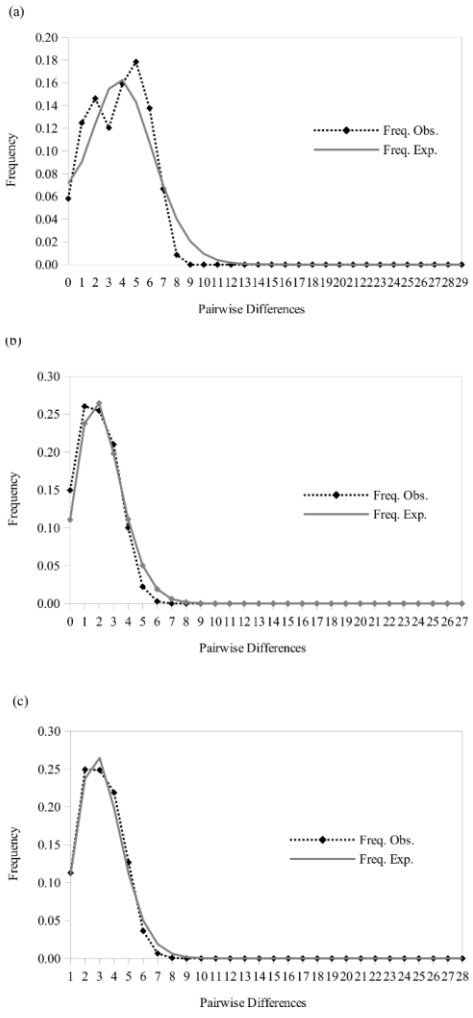
Mismatch distribution

Table 2. Parameters from the mismatch distribution for *Philaenus spumarius* COI groups.

	Mismatch Analysis						
	Demographic Expansion						
	Parameters						
	θ_0 (CI = 95%)	θ_1 (CI = 95%)	τ (CI = 95%)	SSD	P_{SSD}	Raggedness	P_{rag}
Eastern Group	0.00176 (0.000–1.366)	14.41895 (8.628–99999)	4.64844 (1.561–7.461)	0.00620	0.55700	0.01760	0.84400
Western-Mediterranean Group	0.00703 (0.000–0.729)	25.15625 (4.970–99999)	2.11523 (0.801–3.238)	0.00133	0.58200	0.03281	0.65100
Eastern-Mediterranean Group	0.04395 (0.000–0.698)	115.625 (9.687–99999)	2.21094 (1.041–3.016)	0.00144	0.55300	0.03706	0.59400
	Spatial Expansion						
	Parameters						
	θ (CI = 95%)	M (CI = 95%)	τ (CI = 95%)	SSD	P_{SSD}	Raggedness	P_{rag}
Eastern Group	1.26863 (0.001–3.921)	23.44725 (10.230–99999)	3.07576 (1.335–5.833)	0.00933	0.30800	0.01760	0.85800
Western-Mediterranean Group	0.02484 (0.001–1.066)	42.03757 (8.630–99999)	2.05086 (0.863–2.820)	0.00129	0.57000	0.03281	0.66500
Eastern-Mediterranean Group	0.07874 (0.001–0.958)	351.4398 (18.276–99999)	2.15680 (1.062–2.877)	0.00146	0.45200	0.03706	0.60600

Numbers in parenthesis are the upper and lower bound of 95% CI (1000 bootstrap replicates).
θ_0 and θ_1: pre-expansion and post-expansion populations size; τ: time in number of generations elapsed since the sudden/demographic expansion and spatial expansion episodes; SSD: sum of squared deviations; Raggedness: raggedness index following [49]; P_{SSD} and P_{RAG}.: probability that expected mismatch distributions (1000 bootstrap replicates) be larger than observed mismatch distributions.
doi:10.1371/journal.pone.0098375.t002

Mismatch distribution



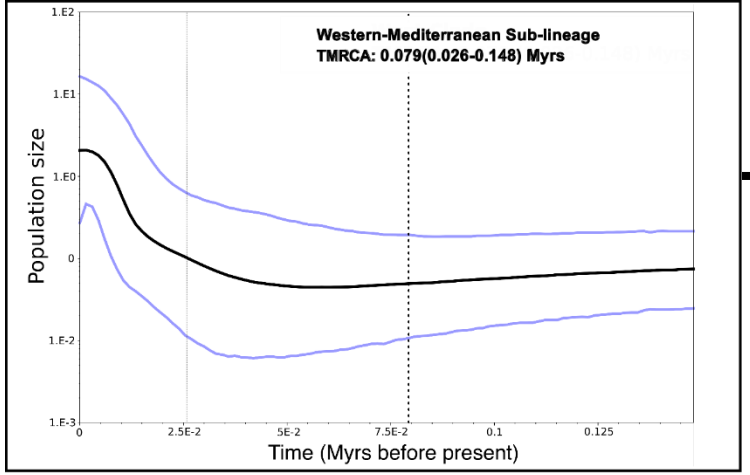
Neutrality Tests

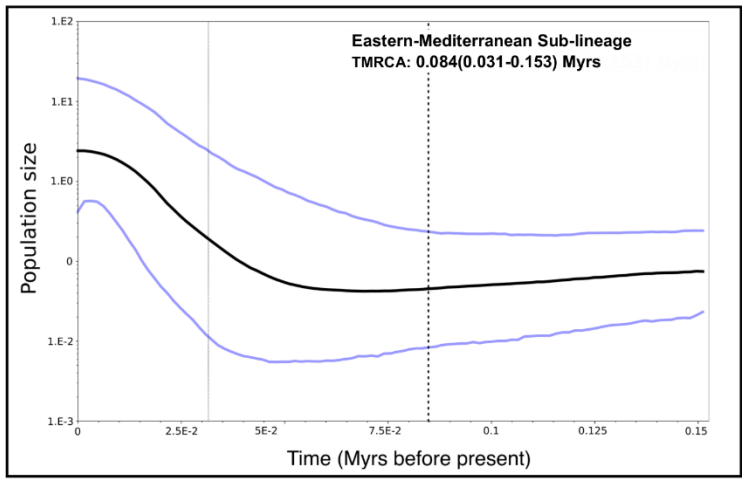
Table 3. Tajima's *D* and [44] Fu's *Fs* test values and their statistical significance for *Philaenus spumarius* Cytochrome *c* oxidase I mtDNA groups.

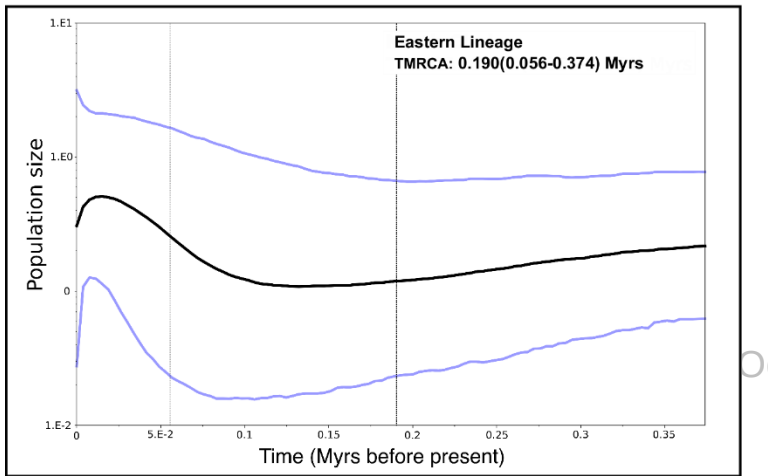
	Neutrality Tests	
	Tajima's *D* test	**Fu's *Fs* test**
Eastern Group	0.15553	−6.28375**
Western-Mediterranean Group	−1.77941*	−23.61561***
Eastern-Mediterranean Group	−1.75709*	−26.22826***

*: indicates significant values at $P<0.05$; **: indicates significant values at $P<0.01$ and ***: indicates significant values at $P<0.001$.
doi:10.1371/journal.pone.0098375.t003

Bayesian Skyline Plots







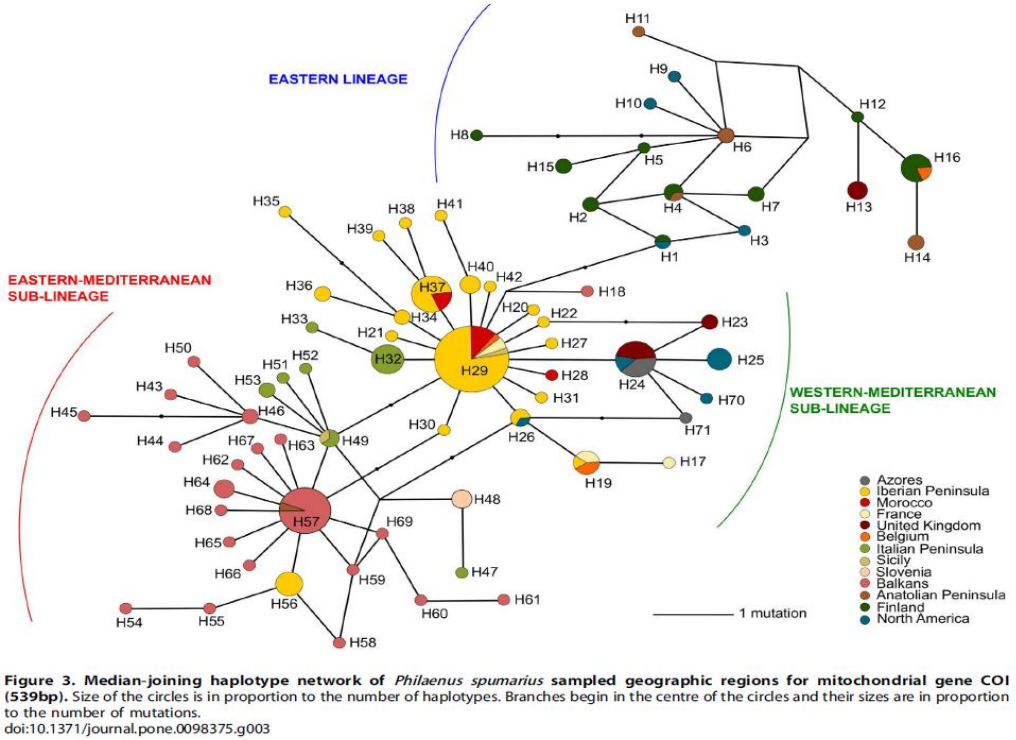


Figure 3. Median-joining haplotype network of *Philaenus spumarius* sampled geographic regions for mitochondrial gene COI (539bp). Size of the circles is in proportion to the number of haplotypes. Branches begin in the centre of the circles and their sizes are in proportion to the number of mutations.
doi:10.1371/journal.pone.0098375.g003

ARTICLE

doi:10.1038/nature11622

Analyses of pig genomes provide insight into porcine demography and evolution

A list of authors and their affiliations appears at the end of the paper

For 10,000 years pigs and humans have shared a close and complex relationship. From domestication to modern breeding practices, humans have shaped the genomes of domestic pigs. Here we present the assembly and analysis of the genome sequence of a female domestic Duroc pig (*Sus scrofa*) and a comparison with the genomes of wild and domestic pigs from Europe and Asia. Wild pigs emerged in South East Asia and subsequently spread across Eurasia. Our results reveal a deep phylogenetic split between European and Asian wild boars ~1 million years ago, and a selective sweep analysis indicates selection on genes involved in RNA processing and regulation. Genes associated with immune response and olfaction exhibit fast evolution. Pigs have the largest repertoire of functional olfactory receptor genes, reflecting the importance of smell in this scavenging animal. The pig genome sequence provides an important resource for further improvements of this important livestock species, and our identification of many putative disease-causing variants extends the potential of the pig as a biomedical model.

ARTICLE

doi:10.1038/nature11622

Analys … nsight into pc … ution

A list of authors and

For 10,000 years … cation to modern breeding practice… ly and analysis of the genome seque… omes of wild and domestic pigs fro… ross Eurasia. Our results reveal a d… o, and a selective sweep analysis in… ted with immune response and olfa… y receptor genes, reflecting the imp… portant resource for further impro… e disease-causing variants extends t…

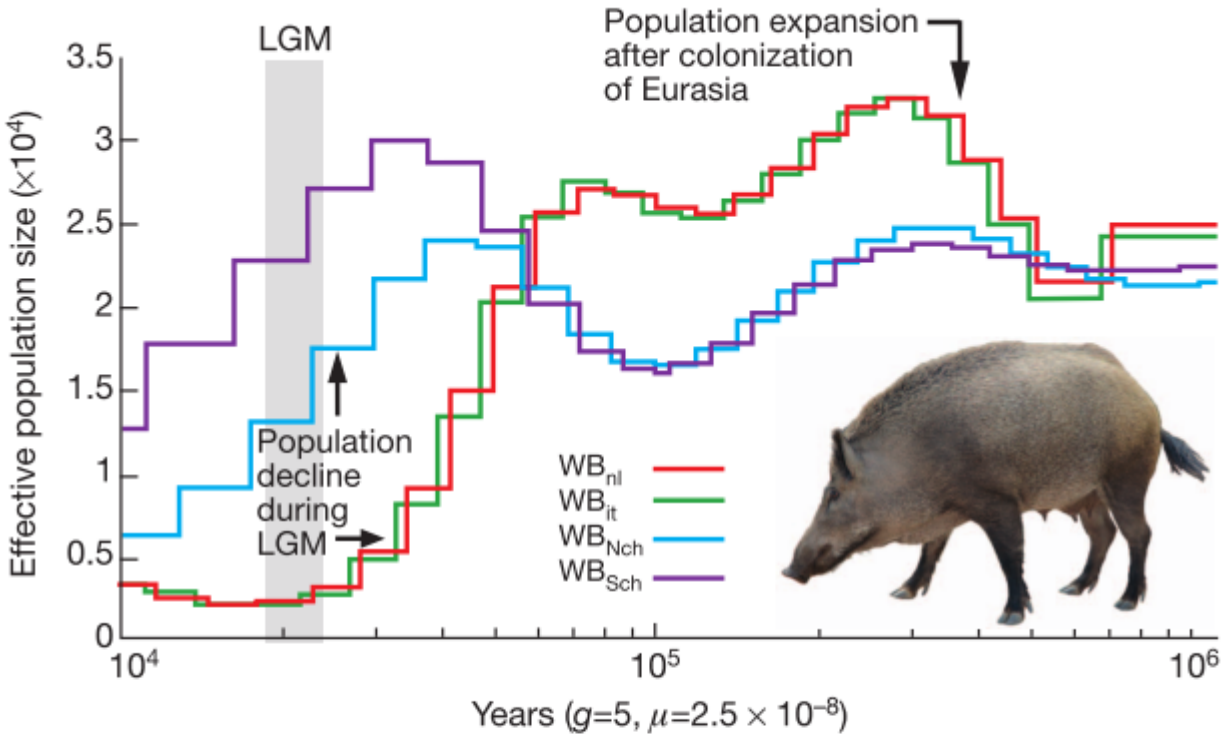


Figure 3 | Demographic history of wild boars. Demographic history was inferred using a hidden Markov model (HMM) approach as implemented in pairwise sequentially Markovian coalescence (PSMC)[45]. In the absence of known mutation rates for pig, we used the default mutation rate for human (μ) of 2.5×10^{-8}. For the generation time (g) we used an estimate of 5 years. The Last Glacial Maximum (LGM) is highlighted in grey. WB_{nl}, wild boar Netherlands; WB_{it}, wild boar Italy; WB_{Nch}, wild boar north China; WB_{Sch}, wild boar south China.

Inferring human population size and separation history from multiple genome sequences

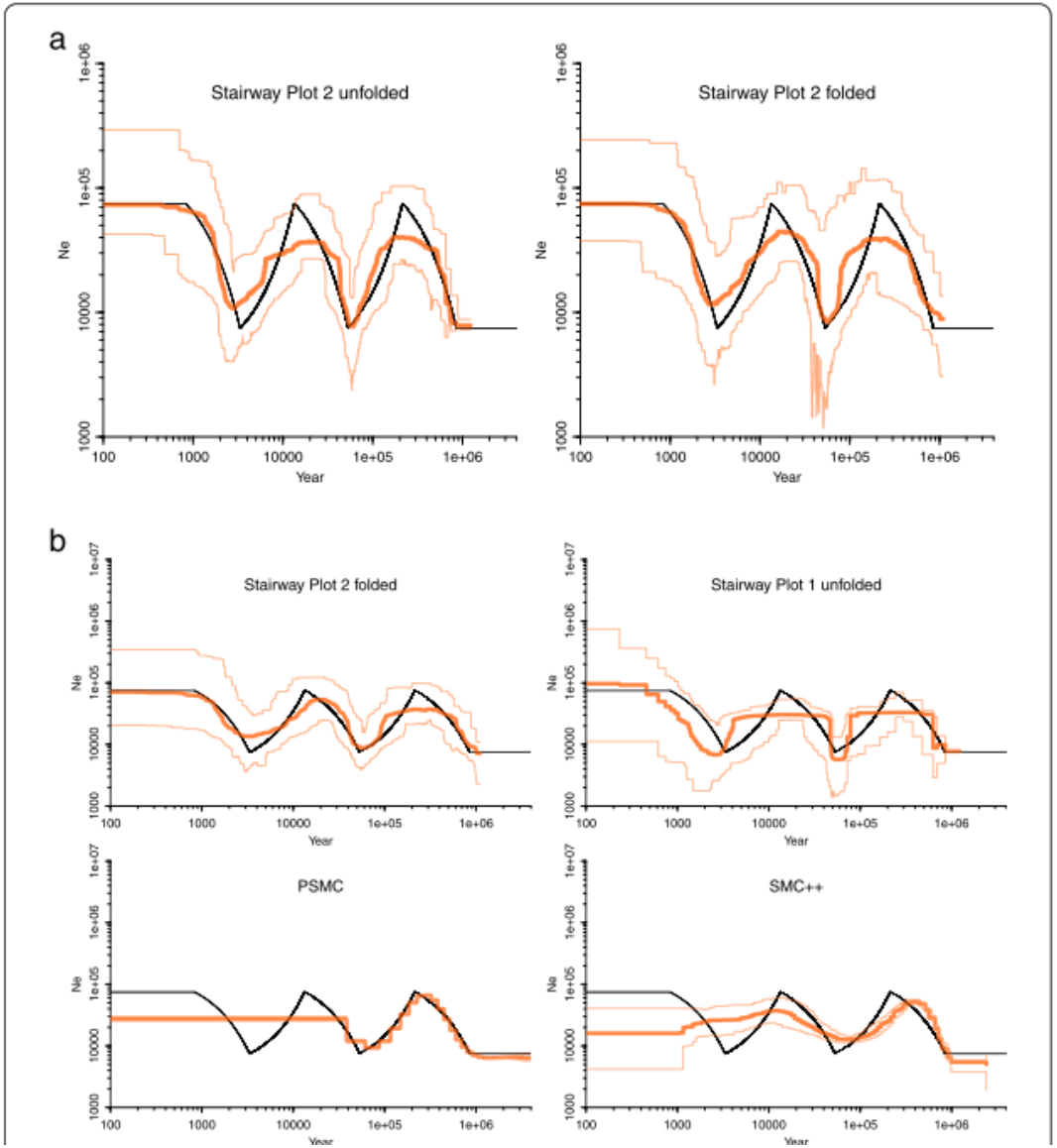
Stephan Schiffels & Richard Durbin

The availability of complete human genome sequences from populations across the world has given rise to new population genetic inference methods that explicitly model ancestral relationships under recombination and mutation. So far, application of these methods to evolutionary history more recent than 20,000–30,000 years ago and to population separations has been limited. Here we present a new method that overcomes these shortcomings. The multiple sequentially Markovian coalescent (MSMC) analyzes the observed pattern of mutations in multiple individuals, focusing on the first coalescence between any two individuals. Results from applying MSMC to genome sequences from nine populations across the world suggest that the genetic separation of non-African ancestors from African Yoruban ancestors started long before 50,000 years ago and give information about human population history as recent as 2,000 years ago, including the bottleneck in the peopling of the Americas and separations within Africa, East Asia and Europe.

Li and Durbin in the pairwise sequential Markovian coalescent (PSMC) model[7], a method that focuses on modeling the two genome sequences in one diploid individual. Because only two sequences are modeled, the coalescent event joining the sequences at the most recent common ancestor almost always occurred more than 20,000 years ago, meaning that PSMC can only infer population size estimates beyond 20,000 years ago. Also, with only two sequences, there is only limited scope for the analysis of population separations.

For more than two sequences, extending PSMC in the natural way by enumerating all possible trees with their branch lengths along the sequences would be very computationally costly, even under the Markovian model. A recent simplification was suggested by Song and colleagues[8–10], which is based on approximating the conditional sampling process for adding an $(n + 1)$th sequence to the distribution of genealogies connecting n sequences. Here we propose an alternative approach that we call multiple sequential Markovian coalescent (MSMC), which overcomes the increase in complexity by introducing a different simplification. We characterize the relationship at a given

Starway Plot 2



Starway Plot 2

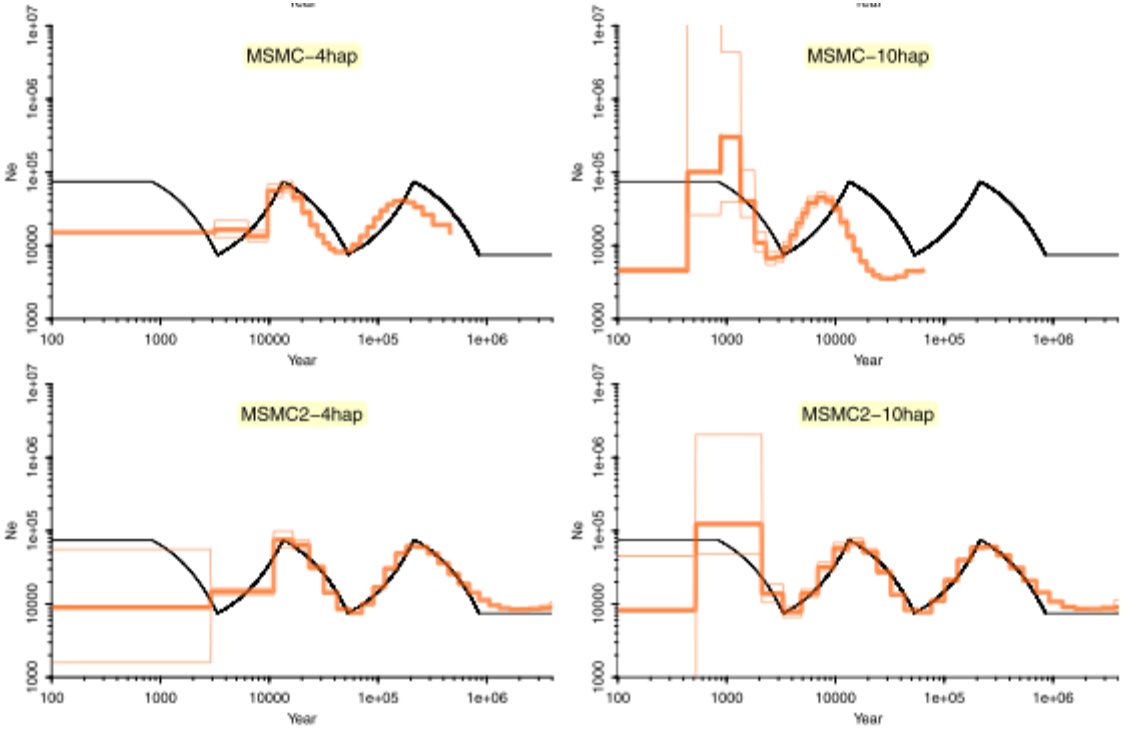


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Comparison of demographic inferences with simulation. **a** Comparison of Stairway Plot 2 with folded or unfolded SFSs using the same average SFS from 200 simulations. **b** Comparison of Stairway Plot 2 with folded SFSs vs. Stairway Plot 1, PSMC, SMC++, MSMC, and MSMC2, using the same simulated sequences from 200 simulations assuming a zig-zag model [2]. Each simulation simulates 100 diploids with 10 chromosomes; each chromosome is 10 MB. Only one estimation for each simulated sample was used for Stairway Plot 1 and Stairway Plot 2. MSMC and MSMC2 group samples with every 4 haplotypes (4hap) or every 10 haplotypes (10hap). Black line: true model. Thick orange line: median of 200 estimations. Thin orange lines: 2.5% and 97.5% confidence limits for 200 estimations