

Supplements

Andreas Wieser

November 8, 2016

Contents

Adjustment of Effective Population Size	1
Models in Detail	2
Migration pathways	2
Constant Demography Model	3
Population Growth Model	4
Approximated Demographic Model	5
Calculation of F_{ST} values	5
References	7

Adjustment of Effective Population Size

To account for different generation times in our populations, we adjusted population sizes for recent epochs:

$$N_E^{adjust} = N_E \frac{G_a}{G_m},$$

where N_E^{adjust} is the adjusted population size, G_a is the number of generations per year and G_m is the mean number of Generations per year over all populations (Table 1, (Oppold et al. 2016)).

We refrained from adjusting population sizes in the distant past, as additional information on local climate and the spacial distribution of the population is not readily available (or even possible to obtain).

Table 1: Populations with generations per year

Population	Abbreviation	Generations per year	θ (Migrate Analysis)
Hessen (G)	MG	7.85	0.0316
Metz (F)	NMF	7.7	0.197
Lyon (F)	MF	9.07	0.396
Piemont (I)	SI	10.57	0.031
Andalucia (S)	SS	14.86	0.025
mean		10.01	

Models in Detail

General settings, consistent in all models:

Number of simulations per model: 200,000 Number of populations: 5

Number of samples: 20 per population

Length of sequence: 1,000 base pairs

Mutation rate per site and generation μ : 4.1×10^{-9}

Recombination rate: 0

Transition bias: 0.595

Simulations were performed using *fastsimcoal* v. 2.5.2 (Excoffier and Foll 2011).

Migration pathways

In all models migration between neighboring populations is allowed (Fig.1 and Table 2).

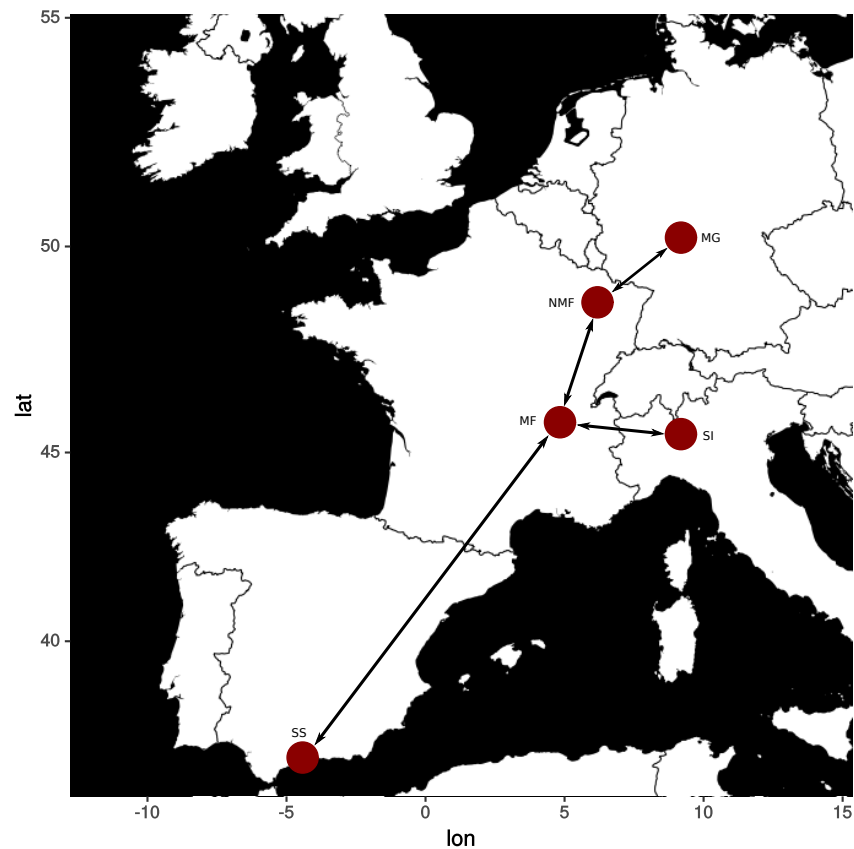


Figure 1: Locations of populations and possible migration routes between them (Kahle and Wickham 2013).

Table 2: Matrix of possible migration between neighboring populations.

	MG	NMF	MF	SI	SS
MG	0	possible	0	0	0
NMF	possible	0	possible	0	0
MF	0	possible	0	possible	possible
SI	0	0	possible	0	0
SS	0	0	possible	0	0

Constant Demography Model

As the simplest option we chose a population split model of constant population sizes and migration rates constant over time (Fig. 2, Table 3). Migration rates and population sizes are based on the results of the Migrate-n analysis (Beerli and Felsenstein 2001; Beerli 2006).

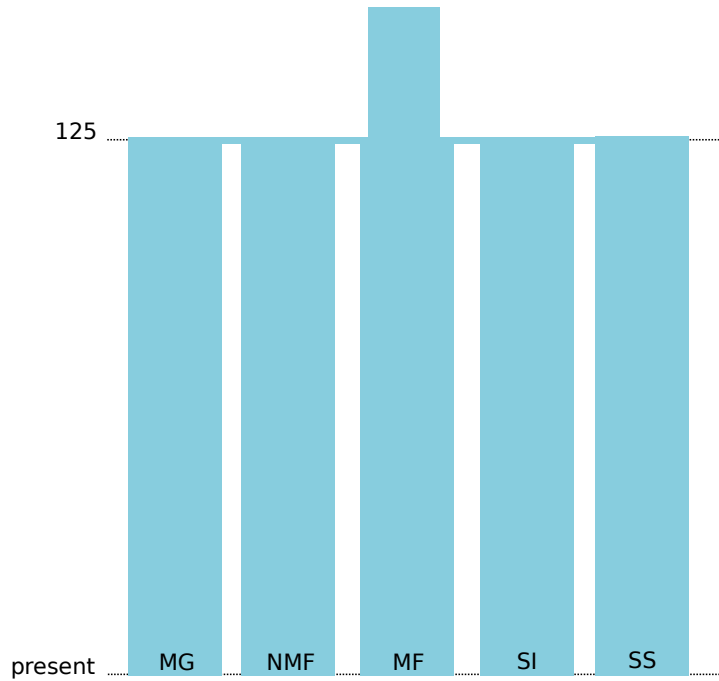


Figure 2: Graphical representation of demographic model: constant demography model (time scales in thousand generations).

Table 3: Population sizes in constant demography model. N_E^{adjust} and $N_E^{initial}$ give number of individuals at present and 125,000 generations ago.

Population	Abbreviation	N_E^{adjust}	$N_E^{initial}$
Hessen (G)	MG	1000857	0
Metz (F)	NMF	6120294	0
Lyon (F)	MF	14491648	28983
Piemont (I)	SI	1322063	0
Andalucia (S)	SS	1498905	0

Table 4: Migration matrix in constant demography model.

	MG	NMF	MF	SI	SS
MG	0	9.1×10^{-4}	0	0	0
NMF	9.1×10^{-4}	0	2.5×10^{-5}	0	0
MF	0	9.1×10^{-4}	0	3.0×10^{-5}	9.1×10^{-4}
SI	0	0	1.9×10^{-5}	0	0
SS	0	0	1.5×10^{-5}	0	0

Population Growth Model

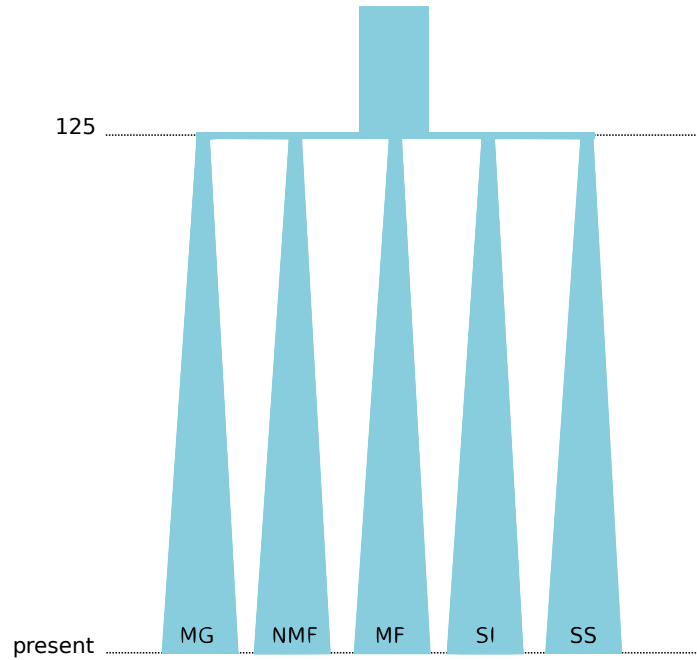


Figure 3: Graphical representation of demographic model: population growth model (time scales in thousand generations).

All parameters of this model are the same as in the Constant Demography Model, except for the addition of a population expansion (Fig.3). The growth rate is $r = 1.0 \times 10^{-5}$ and population growth is given by:

$$N_t = N_0 e^{rt},$$

where N_t equals population size in generation t and N_0 is the initial population size (Excoffier and Foll 2011).

Approximated Demographic Model

Based on the results of our MSMC2 analysis (Schiffels and Durbin 2014) we developed an approximated demographic model (Fig. 2 of main article) of population split, shrinkage and following expansion. Migration rates change over time, first decreasing to near isolation and then rising again, mirroring inferences on cross-coalescence rate from the MSMC2 analysis (Fig. 4, Table 4).

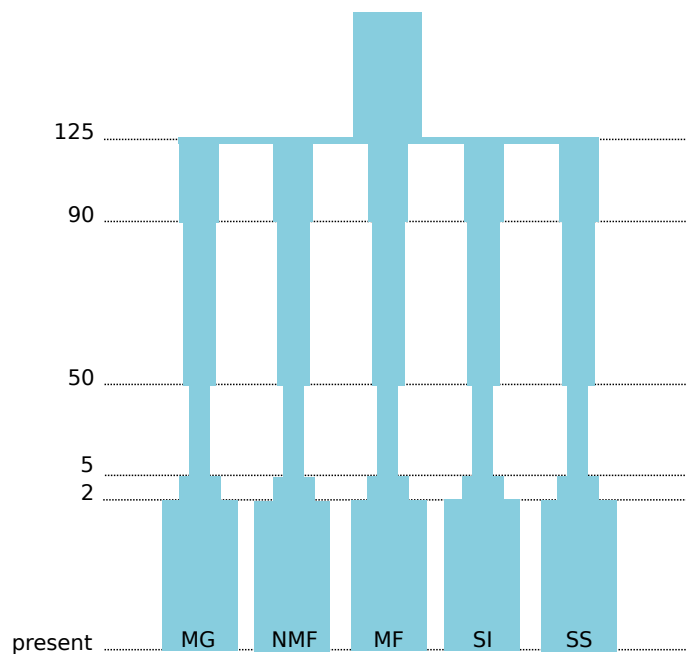


Figure 4: Graphical representation of demographic model: approximated population model (time scales in thousand generations).

Table 5: Population sizes at different time points (in generations) and migration rates (MIG) in these epochs.

	present	2000	5000	20000	50000	90000	125000
MG	282318	39211	18037	16000	20000	27000	0
NMF	615385	38462	16923	12000	19000	25000	0
MF	308072	47117	18122	13000	20000	26000	29000
SI	253427	26399	15839	18000	25000	30000	0
SS	504735	37113	14845	8000	15000	18000	0
MIG	1.02×10^{-5}	2.7×10^{-4}	3.0×10^{-4}	3.7×10^{-3}	5.1×10^{-3}	2.9×10^{-3}	0

Calculation of F_{ST} values

Pairwise F_{ST} values are used to detect short term genetic distances between populations (Excoffier and Lischer 2010; Reynolds, Weir, and Cockerham 1983; Slatkin 1995). We calculated these for all models as well as the empirical data, generated density functions and compared them (Fig. 5). Computation was performed with arlsumstat, the command-line version of Arlequin 3.5 (Excoffier and Lischer 2010).

Kruskal-Wallis tests showed significant differences in all pairings (Hollander, Wolfe, and Chicken 2013).

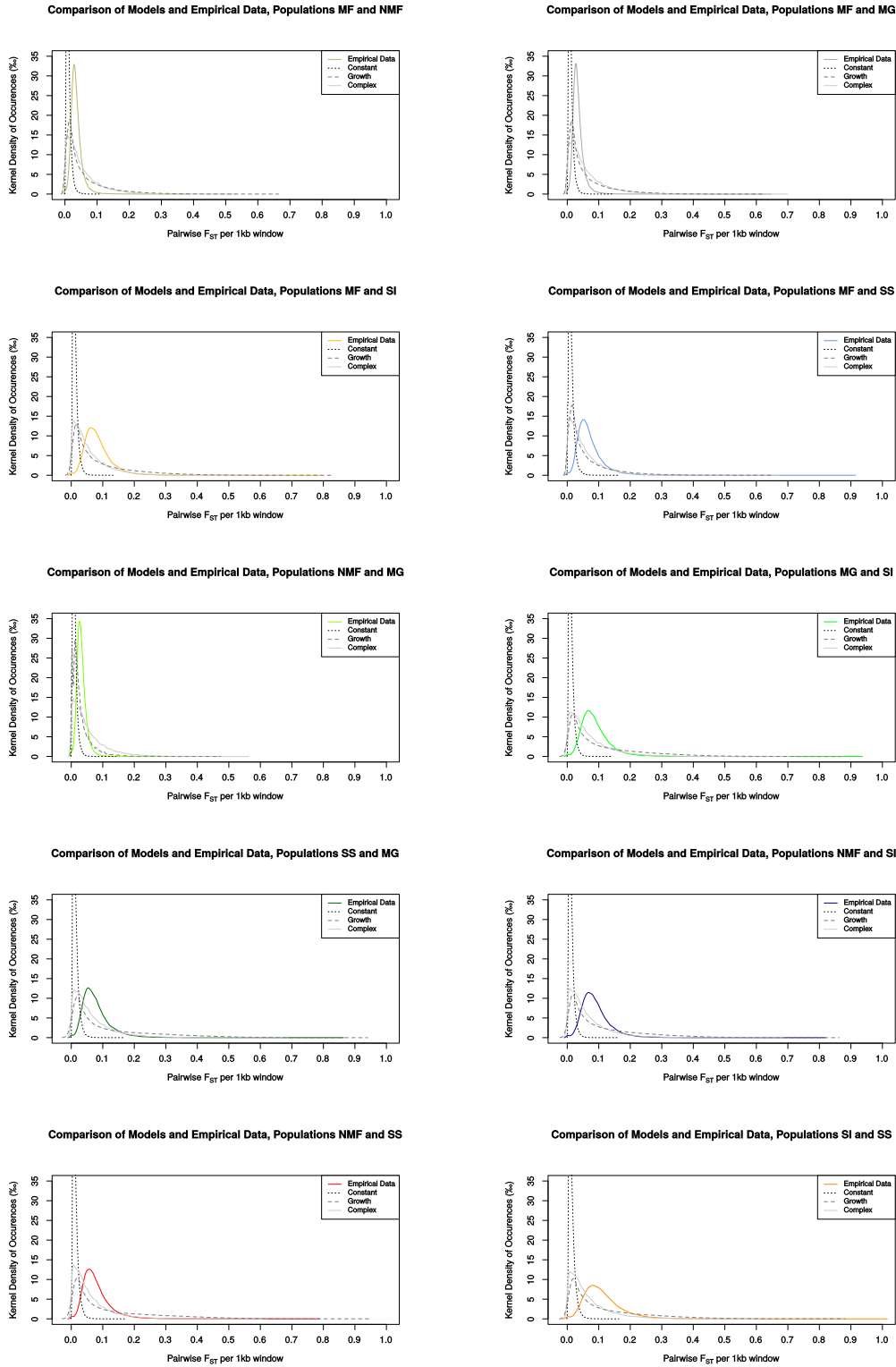


Figure 5: Comparisons of density functions of pairwise F_{ST} values between all pairs of populations

References

- Beerli, Peter. 2006. "Comparison of Bayesian and Maximum-Likelihood Inference of Population Genetic Parameters." *Bioinformatics* 22 (3). Oxford Univ Press: 341–45.
- Beerli, Peter, and Joseph Felsenstein. 2001. "Maximum Likelihood Estimation of a Migration Matrix and Effective Population Sizes in N Subpopulations by Using a Coalescent Approach." *Proceedings of the National Academy of Sciences* 98 (8). National Acad Sciences: 4563–8.
- Excoffier, Laurent, and Matthieu Foll. 2011. "Fastsimcoal: A Continuous-Time Coalescent Simulator of Genomic Diversity Under Arbitrarily Complex Evolutionary Scenarios." *Bioinformatics* 27 (9): 1332–4. doi:10.1093/bioinformatics/btr124.
- Excoffier, Laurent, and Heidi EL Lischer. 2010. "Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses Under Linux and Windows." *Molecular Ecology Resources* 10 (3). Wiley Online Library: 564–67.
- Hollander, Myles, Douglas A Wolfe, and Eric Chicken. 2013. *Nonparametric Statistical Methods*. John Wiley & Sons.
- Kahle, David, and Hadley Wickham. 2013. "Ggmap: Spatial Visualization with Ggplot2." *The R Journal* 5 (1): 144–61. <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- Oppold, Ann-Marie, João AM Pedrosa, Miklós Bálint, João B Diogo, Julia Ilkova, João LT Pestana, and Markus Pfenninger. 2016. "Support for the Evolutionary Speed Hypothesis from Intraspecific Population Genetic Data in the Non-Biting Midge *Chironomus riparius*." In *Proc. R. Soc. B*, 283:20152413. 1825. The Royal Society.
- Reynolds, John, Bruce S Weir, and C Clark Cockerham. 1983. "Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance." *Genetics* 105 (3). Genetics Soc America: 767–79.
- Schiffels, Stephan, and Richard Durbin. 2014. "Inferring Human Population Size and Separation History from Multiple Genome Sequences." *Nature Genetics* 46 (8). Nature Publishing Group: 919–25.
- Slatkin, Montgomery. 1995. "A Measure of Population Subdivision Based on Microsatellite Allele Frequencies." *Genetics* 139 (1). Genetics Soc America: 457–62.