Performance of Approximate Bayesian Computation (ABC) in detecting genetic erosion

The software DIYABC v.2.1.0 was used to perform all the ABC analysis. In order to evaluate the performance of ABC in individuating the correct scenarios a set of simulations were done. In details, a total of 4 cases were tested. Firstly, two different scenarios of population bottleneck were simulated ($N_{e(1)}$, with a 23% reductions per generation starting from N_e=3000; $N_{e(2)}$, with a 7.5% reductions per generation from N_e=3000; $N_{e(3)}$, on a total of 14 generations). Then, two different cases of analytical efforts were evaluated, by assuming one single (recent) sample of 50 specimens (sex ratio 1:1) genotyped at 20 or 50 microsatellites loci. The simulated datasets were analyzed in separated ABCs, with two candidate scenarios: a) a scenario congruent with a stable population evolving with constant Ne through the 14 generations; b) a scenario describing a population showing a potential bottlenecks occurred during the generations. A uniform distribution [10-5000] of N_e was set for population samples at generations 1 and 14. A narrow uniform distribution [13-10] was imposed for N_e variation timing. Finally, under a Generalized Stepwise Mutation model for microsatellites, default values of uniform distribution were set for the mean mutation rate μ [1.0×10^{-4},1.0×10^{-3}], mean P [1.0×10^{-1},3.0×10^{-1}], gamma [1.0×10^{-9},1.0×10^{-4}] and $\mu_{(SNI)}$ [1.0×10^{-8},1.0×10^{-5}]. The chosen summary statistics were the mean number of alleles and the mean genic diversity both one sample and pairwise summary statistics. The model choice was done on the basis of 1.0×10^6 simulations and performed under a polychotomic weighted logistic regression done using the first 10,000 simulations ranked according to their closeness to the observed dataset. The type I (probability to wrongly favor the stationarity in the true case of bottleneck) and type II (probability to wrongly favor the bottleneck in the true case of stationarity) errors were estimated by using the specific analysis "confidence in scenario choice", which used 1000 generated datasets from the built reference table (drawn from priors and from both hypothetical scenarios).

We performed the same analysis also simulating the performance of a hypothetical dataset of 2000 SNPs loci from 50 specimens, leaving unchanged all the other settings. Finally, we performed another analysis evaluating the performance of a temporal repeated sample (7 samples, one each 2 generations) of 50 specimens genotyped at 50 microsatellites loci, leaving unchanged all the other settings.