- Bayesian Project -
Estimating divergence time of Polar bears and Brown bears

CID: 00693808

March 28th, 2017

A brief summary of the statistical methods and results of the Bayesian framework for estimating the divergence time (T) of Polar bears and Brown bears. Genomic data from 18 Polar bears (36 chromosomes) and 7 Brown bears (14 chromosomes) were evaluated using Approximate Bayesian Computation (ABC).
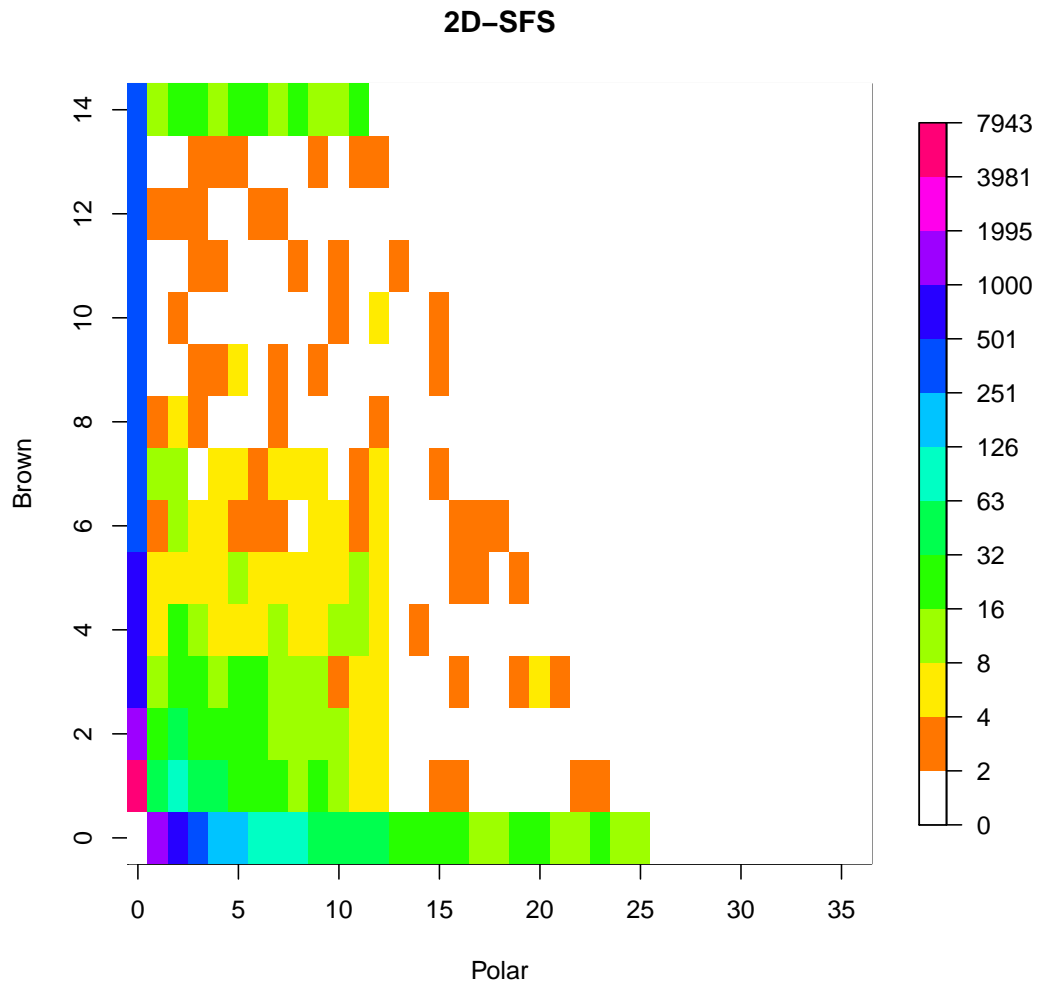

Figure 1: Joint (2-dimensional) site frequency spectrum (SFS) of Polar bears (columns) and Brown bears (rows). A total of 16556 polymorphic sites were analysed.

Figure 1 shows the site frequency spectra of both species. Summary statistics were calculated for observed site frequencies. Simulations were performed to estimate parameter T, drawing random samples from a uniform distribution (MS software). Summary statistics for each of the candidate variables are shown in Figure 2.
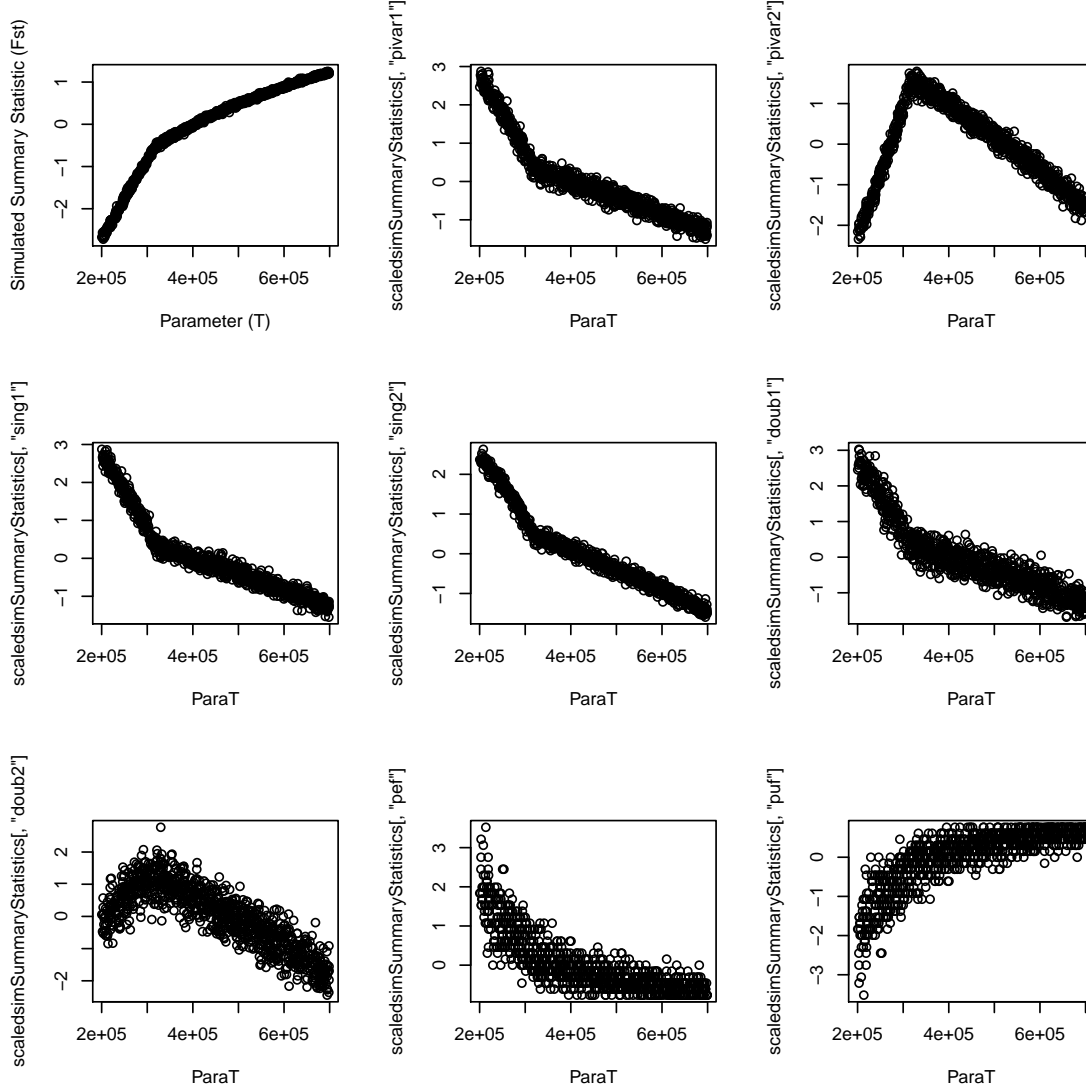

Figure 2: Summary statistics corresponding to nine different measures of genetic diversity.

The majority of statistics appear to be somewhat correlated. Figure 3 shows the pairwise correlations between each of the nine summary statistics. Summary statistics show that "pef" and "puf" appear to be strongly correlated. After performing the first run of simulations with all variables included, tests of correlation between the simulated parameter value and summary statistics were performed. Indeed "pef" and "puf" were perfectly anticorrelated (t = -Inf, df = 998, p-value ¡ 2.2e-16). From tests of correlation between all summary staistics, "Fst" was chosen to be themost representative summary statistic. "Fst" was found to correlate highly with all other summary staistics, with the exception of "pivar2" (cor = -0.03, t = -0.99, df = 998, p-value = 0.32, 95% CI: -0.09 to 0.03).

For this reason, analysis was continued using "Fst" and "pivar2". Simulated summary statistics were scaled separately, with $\mu = 0$ and $\sigma = 1$.
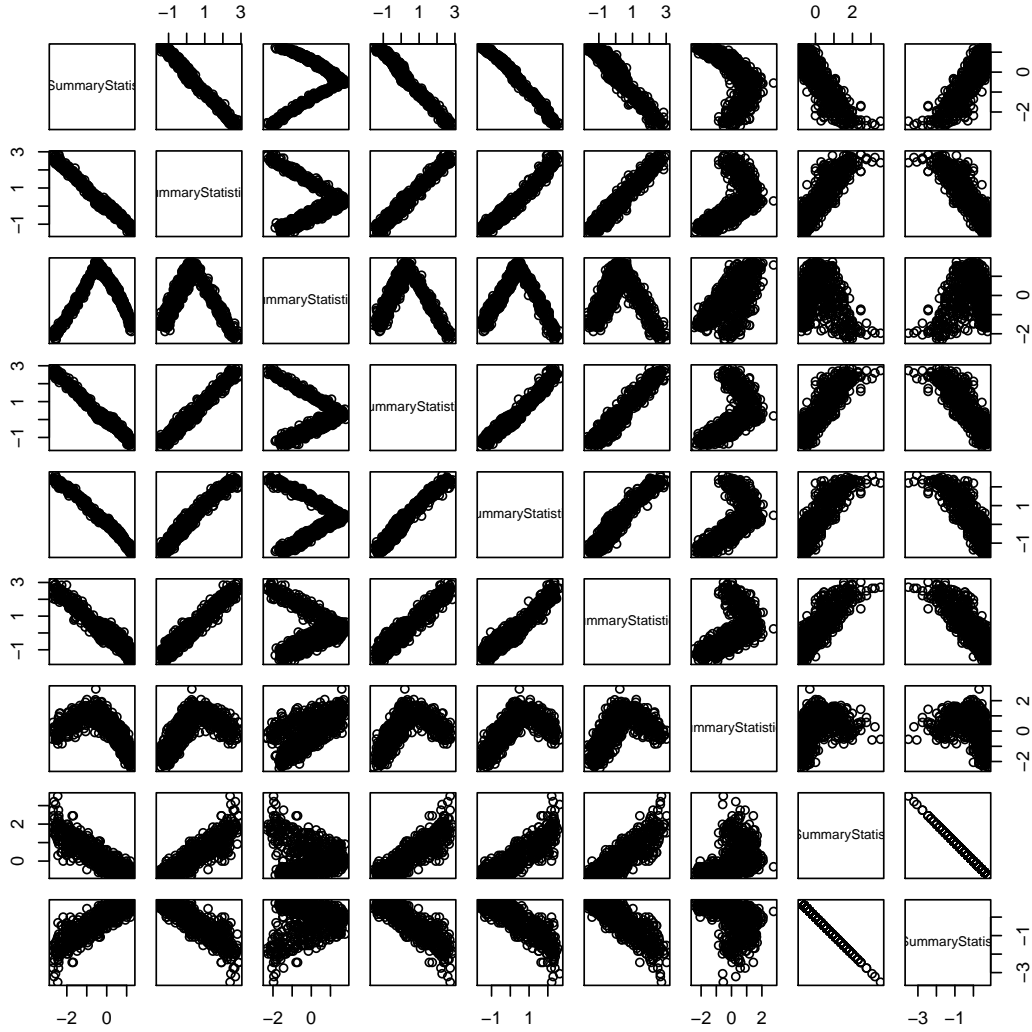

Figure 3: Pairwise correlations between each of the nine summary statistics. High correlation is apparent between the majority of diversity measures.

Approximate Bayesian Computation (ABC) analyses were performed using local linear regression, with correction for heteroscedasticity using the observed summary statistics, and simulated parameter values and associated summary statistics.

Descriptor	Value
Minimum	264871.0
Weighted 2.5%	265744.0
Weighted Median	270215.9
Weighted Mean	270174.0
Weighted Mode	270204.7
Weighted 97.5%	274499.9
Maximum	274555.2

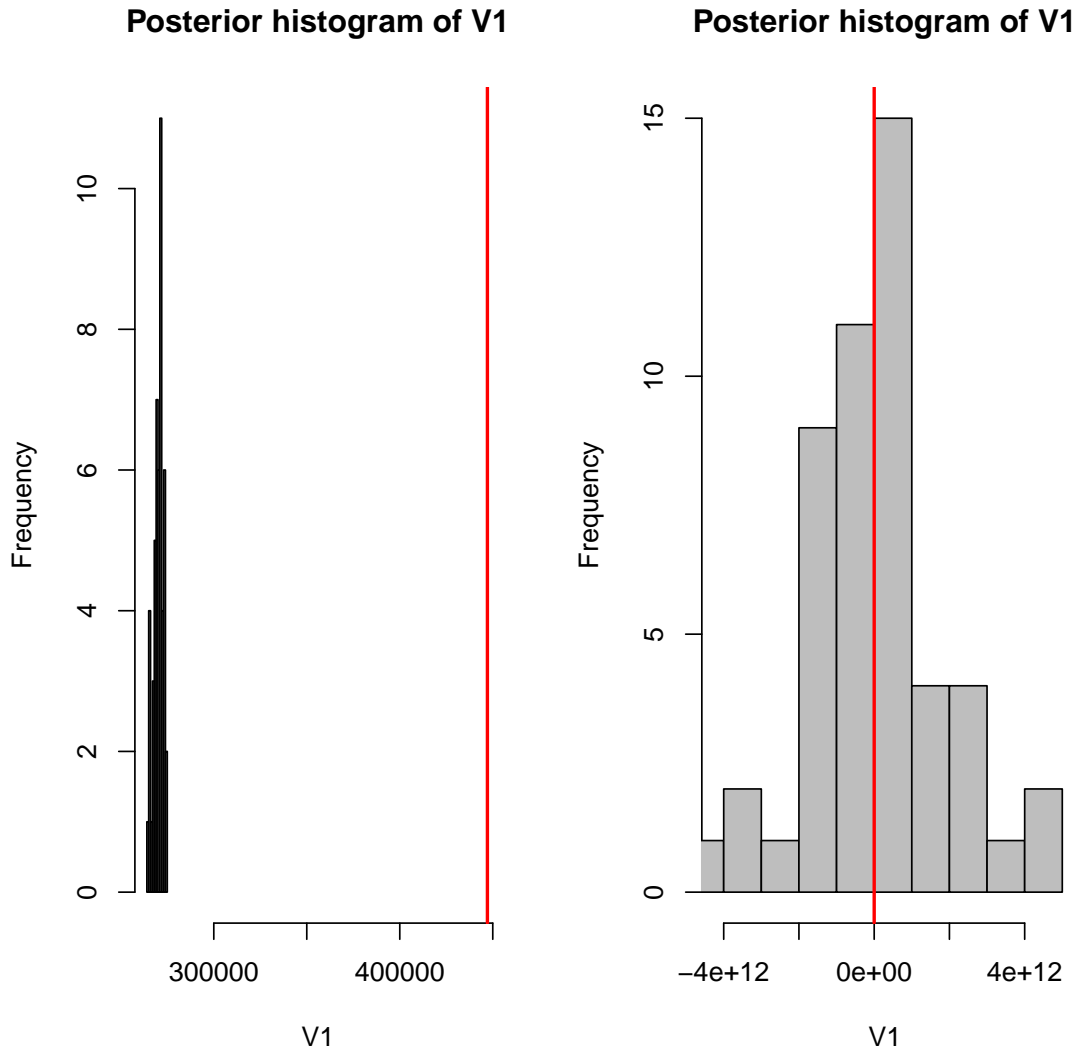Table 1: Fst summary statistics

Figure 4: Histogram of posterior probability density calculated using Approximate Bayesian Computation (ABC) local linear regression analysis. Mean parameter value is shown.

Table 2 shows Highest Posterior Density (HPD) for Fst summary statistic.

Lower	Upper
265229.7	274555.2

Table 2: Posterior distribution. HPD Interval (Fst): p = 0.96
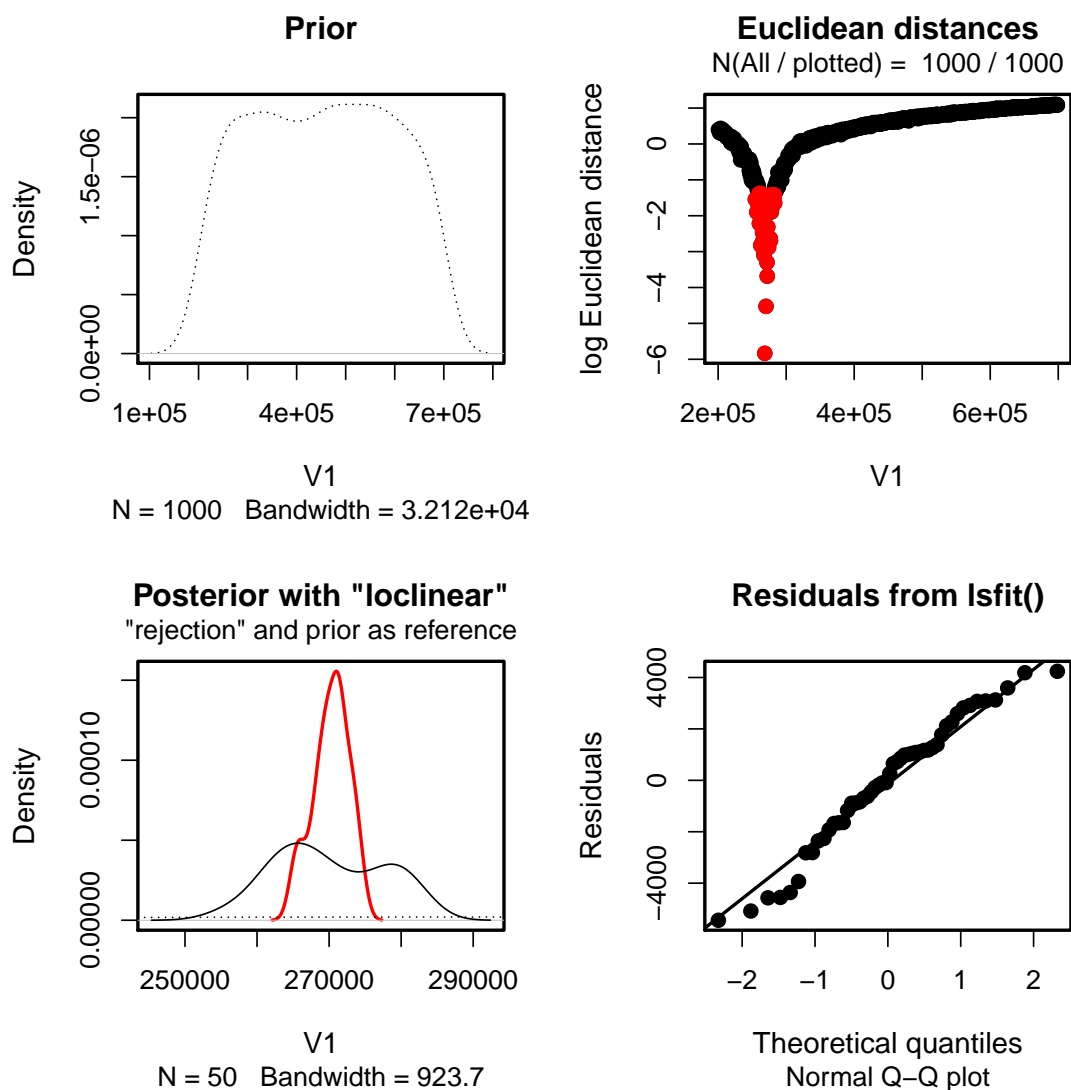

Figure 5: ABC diagnostic plots (local linear regression) show an even distribution of sampled values from across the whole range of the prior.

From this analysis, I conclude that the estimated time since divergence the speciation of Polar bears and Brown bears, based on genetic data, is $270,000 \pm 5,000$ years ago.