Genetic diversity in clustered colonies of an Antarctic marine mesopredator. A role for habitat quality?

Mori et al. *Antarctic Science*

Statistical analyses of microsatellite data

We used the software MicroChecker version 2.2.3 (Van Oosterhout et al. 2003) to assess whether null alleles (i.e. non-amplified), allelic dropouts and errors due to stutter peaks occurred. The software Cervus 3.0.3 (Marshall 1998) was used to determine the Polymorphic Informative Content (PIC) of each locus, a measure of the information related to the expected heterozygosity and, therefore, calculated on the basis of the allelic frequencies. A locus was considered as sufficiently "informative" with PIC $\geq$ 0.4. The software Arlequin vers. 3.11 (Excoffier & Lischer 2010) was used to obtain the number of alleles per locus, the Nei's genetic diversity index and the values of F_{ST} parameter according to Wright's *F*-statistics (i.e. the measure of genetic differentiation among colonies). The genetic diversity index is calculated as an estimate of the expected correct heterozygosity for small samples, a parameter which ranges from 0, in absence of polymorphism, to 1. Allelic frequencies for each locus and the number of unique alleles were obtained with the GenAlEx vers. 6.5 (Smouse & Peakall 2012). The allelic richness corrected for the population size was calculated with the software FStat 2.9.3.2 (Goudet, 2002). Deviations from the Hardy-Weinberg Equilibrium (HWE) and from the Linkage Disequilibrium (LE), in the case of both excess and defect of heterozygotes, were calculated with the software Genepop on the Web (Raymond & Rousset 1995). HWE expectations were tested for each locus and for the total of loci per colony, using randomization procedures with 1000 permutations (Guo & Thompson 1992). The LE was tested through a log-likelihood test, and the empirical distribution was generated by using 10000 permutations (Slatkin & Excoffier 1996). The significance level for HWE and LE was estimated using the Bonferroni correction. The same software was also used to estimate the gene flow, expressed as the number of migrants among colonies per

generation, corrected for population size (Raymond & Rousset 1995), i.e. the average number of migrants successfully entering each population per generation (Whitlock & McCauley 1999). Values of inter-colony genetic distances have been reported on a graph through a 3D Factorial Correspondences Analysis (3D FCA) conducted with the software Genetix vers. 4.03, to show genetic differences among colonies (Belkhir et al. 2001). The degree of differentiation between population pairs was computed through the θ statistics; we assessed its significance against 1000 permutations following the software instructions (Belkhir et al. 2001). Finally, we performed a cluster analysis on the STR genotypic profile of each individual following the Bayesian method, through the software Structure vers. 2.3.4 (Earl 2012). This software carries out a series of simulations to estimate the value of the natural logarithm of the probability that individuals are divided into a K number of clusters, with K ranging between one and eight. Each simulation was based on 106 iterations with a burn-in interval of 104 iterations. The simulation has been repeated three times for each value of K. We computed an *ad hoc* statistic (ΔK) which is based on the change rate of in the log probability between successive K-values: the "best" K value is determined at the highest level of ΔK. The probability that each individual has to belong to each of the clusters defined by the program was calculated (Earl 2012).

References

BELKHIR, K., BORSA, P., CHIKHI, L., RAUFASTE, N., & BONHOMME, F. 2001. *GENETIX, software under WindowsTM for the genetic of populations*. Montpellier, France: Laboratory Genome, Populations, Interactions CNRS UMR, 5000.

EARL, D.A. 2012. Structure Harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conservation Genetics Research*, **4**, 359–361.

EXCOFFIER, L., & LISCHER, H.E. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Res*. 10, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

GOUDET, J., 2002. *FSTAT Software, v. 2.9.3.2*. Available at [http://www2.unil.ch/popgen/softwares/fstat.htm]. Accessed on September 8, 2019.

GUO, S.W., & THOMPSON, E.A. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372. doi: 10.2307/2532296

MARSHALL, T., 1998. *Program Cervus (ver 3.0. 3).* Field Genetics Ltd. www.fieldgenetics.com. Accessed on 12th March 2020.

RAYMOND, M., & ROUSSET, F. 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

SLATKIN, M., & EXCOFFIER, L. 1996. Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. *Heredity*, **76**, 377–383.

SMOUSE, R.P.P., & PEAKALL, R. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, **28**, 2537–2539.

VAN OOSTERHOUT, C., HUTCHINSON, W.F., WILLS, D.P.M., & SHIPLEY, P. 2003. *Micro-Checker, version 2.2.3.* Department of Biological Sciences and Department of Computer Science, University of Hull, Hull, UK.

WHITLOCK, M.C., & MCCAULEY, D.E. 1999. Indirect measures of gene flow and migration: $F_{ST} \neq 1/(4Nm+1)$. *Heredity*, **82**, 117–125.