Evaluating the resolution power of new microsatellites for species identification and stock delimitation in the Cape hakes *Merluccius paradoxus* and *M. capensis* (Teleostei: Merlucciidae)

T. B. HOAREAU*, A. W. KLOPPER, S. M. R. DOS SANTOS, C. J. OOSTHUIZEN AND P. BLOOMER

Molecular Ecology and Evolution Programme, Department of Genetics, University of Pretoria, Private bag X20, Hatfield, 0028, South Africa.

*Author to whom correspondence should be addressed: Tel. +27 12 420 3871; Fax. +27 12 362 5327; email: thoareau@gmail.com

Abstract

The utility of 15 new and 17 previously published microsatellite markers was evaluated for species identification and stock delimitation in the deep-water hake *Merluccius paradoxus* and the shallow-water hake *Merluccius capensis*. A total of 14 microsatellites was polymorphic in *M. paradoxus* and 10 in *M. capensis*. Two markers could individually discriminate the species using Bayesian clustering methods and a statistical power analysis showed that the set of markers for each species is likely to detect subtle genetic differentiation ($F_{ST} < 0.006$), which will be valuable to delimit and characterise genetic stocks.

Key words: Bayesian methods; cross-species amplification; genetic markers; genomic library; power analysis

Both the shallow-water hake *Merluccius capensis* Castelnau, 1861 and the deep-water hake *M. paradoxus* Franca, 1960 are targeted by a valuable demersal fishery along the west coasts of Southern Africa (>100 million USD annually; Butterworth & Rademeyer, 2005), but the intensification of exploitation over recent decades caused a resource decline (Payne & Punt, 1995). Due to their morphological similarity and overlapping distribution, the two species are not distinguished in the commercial landings records (von der Heyden et al., 2007b) and they are combined into geographic managements units, namely Namibia, west coast and south coast of South Africa (Butterworth & Rademeyer, 2005). Previous genetic surveys successfully distinguished the two species (Grant *et al.,* 1987; von der Heyden et al., 2007b; Garcia-Vazquez *et al.*, 2012) and detected population differentiation (Grant *et al.,* 1987; von der Heyden *et al.,* 2007a) using mtDNA markers and allozymes. Since these markers are inadequate to draw final conclusions regarding stock delimitation, highly informative markers such as microsatellites are necessary (Selkoe *et al.,* 2006).

To provide reliable genetic markers for species identification and stock delimitation in the two Cape hake species, the resolution power of newly developed and previously published microsatellites is assessed. The development of *de novo* microsatellite markers is presented. The discriminating power of each microsatellite for the correct identification of species was evaluated. A simulation approach was used to assess the robustness of each set of markers in detecting subtle genetic differentiation.

Total genomic DNA was isolated using the DNeasy tissue extraction kit (Qiagen, www.qiagen.com). A partial genomic library enriched using two sets of four tetranucleotide repeat probes (TATC/AGCA/GCGA/CAGC and GATA/GTCT/GAAA/ACGT) was generated for *M. paradoxus* following Zane *et al.* (2002). A total of 585 clones was selected and sequenced and Msatcommander 0.8.2 (Faircloth, 2008) was used to identify 213 sequences containing repeats and design 141 primers after exclusion of duplicates (Table SI, Supporting Information). A total of 20 loci was chosen for further analyses based on quality/length of the repeats (*e.g.* containing perfect repeats). Additional microsatellites were tested from: Moran *et al.* (1999, five markers), Seibert & Ruzzante (2006, four), D'Amato *et al.* (1999, six) and Rico *et al.* (1997, two).

A total of 15 markers with consistent PCR amplification on 2% agarose gels stained with GelRedTM Acid stain (Biotium, www.biotium.com) were selected. PCRs were prepared with 25-100 ng of genomic DNA, 1X PCR buffer, 0.4U Supertherm *Taq* polymerase (Southern Cross Biotechnologies, http://za.w393.com/27566370205), 1.5 mM $MgCl_2$, 1 pmol forward and reverse primers (Inqaba Biotec, www.inqababiotec.co.za) and 0.2 mM dNTPs (Promega, www.promega.com) in a final volume of 10 µL. The PCR cycling conditions were: 5 min at 94°C, followed by 35 cycles of 30 sec at 94°C, 30 sec at 55°C and 45 sec at 72°C and a final elongation step of 20 min at 72°C. Polymorphism was assessed in 32 adults of each species collected within the same location in 2005 (von der Heyden *et al.,* 2007a). The genotyping was performed by combining the Quantitect Multiplex PCR kit (Qiagen, www.qiagen.com) and four markers fluorescently labelled using the G5 dye set (NED, VIC, PET or FAM) from Applied Biosystems

(www.appliedbiosystems.com). The PCR products were then electrophoresed along with the GeneScan-500 LIZ size standard on an ABI 3100 (Applied Biosystems, www.appliedbiosystems.com) and alleles scored using GeneMarker 1.80 (SoftGenetics, www.softgenetics.com).

Genetic parameters and significance tests were performed using Genetix 4.05 (Belkhir *et al.*, 2004) and included unbiased expected heterozygosity (H_K; Nei, 1978), observed heterozygosity (H_O), number of alleles per locus (A), Hardy-Weinberg (F_{IS}) and linkage (LD) disequilibrium

Table II. Characterization of the 10 polymorphic microsatellite markers for the shallow-water hake *Merluccius capensis*. A: number of alleles; H_E and H_O: expected and observed heterozygosity; F_{IS}: inbreeding coefficient assessing Hardy–Weinberg equilibrium within samples. "*" indicate markers showing significant deficit of heterozygotes (departure from HWE) and potential presence of null alleles according to Micro-Checker.

Locus	Size (bp)	A	H_K	H_O	F_{IS}
MP0051	192-212	5	0.306	0.344	-0.127^{ns}
MP0318	129-149	5	0.310	0.281	0.095^{ns}
MP0374	83-87	2	0.146	0.156	-0.069^{ns}
MP8450	215-275	12	0.870	0.812	0.067^{ns}
MP8478	192-228	10	0.824	0.969	-0.179^{ns}
MP8494	308-320	4	0.305	0.125	0.593*
Mmerhk3b	321-325	3	0.588	0.562	0.044^{ns}
Mmerhk20	215-259	19	0.917	0.844	0.081^{ns}
Mmerhk29	151-183	14	0.877	0.344	0.612*
Mmerhk34b	118-166	18	0.924	0.750	0.190*

(Table I). A total of 14 polymorphic microsatellites (10 *de novo*) were obtained for the deep-water hake (*A* = 4–25; H_K = 0.441–0.967; Table I) and 10 (six *de novo*) for the shallow-water hake (*A* = 2–19; H_K = 0.146–0.924; Table II). No LD was observed among markers. Micro-Checker 2.2.3 (Van Oosterhout et al., 2004) detected null alleles in markers showing departure from Hardy-Weinberg equilibrium in the deep-water hake (*MP9131*, *Mmerhk29*, *Mmer110-8*) and the shallow-water hake (*MP8494*, *Mmerhk29*, *Mmerhk34b*). However, this departure could be linked to genetic subdivision in the samples rather than to the presence of null alleles (e.g. Hoareau *et al.*, 2009).

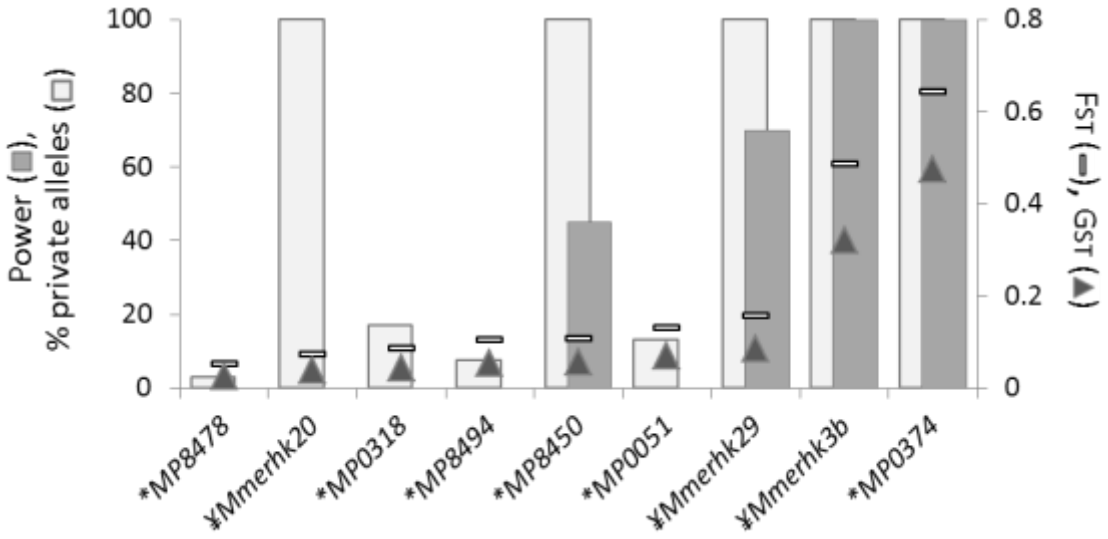

Figure 1. Power of each locus to distinguish between *Merluccius paradoxus* (N = 32) and *M. capensis* (N = 32), their associated F_{ST} and G_{ST} values and cumulated frequency of private alleles observed between the two species. The markers, ranked according to their F_{ST} values, are those giving consistent amplification in both species. The power refers to the percentage of cluster analysis runs implemented in Structure v2.3.1 for which the locus can discriminate the species. ¥ denotes previously published markers and * *de novo* markers.

The resolution power of each locus in distinguishing the two species was evaluated using *F*-statistics and a clustering method. F_{ST} (in Genetix) and G_{ST} (in GenAlEx 6.5; Peakall & Smouse, 2012) were estimated for each locus giving consistent amplification in both species (Fig. 1). The clustering method implemented in Structure 2.3.1 (Pritchard *et al.*, 2000) was applied per locus (no admixture, correlated allele frequency) using 10^6 MCMC replicates following a burn-in period of 10^5. The most likely number of clusters (K) was determined following Evanno *et al.* (2005) over 20 runs for each K = 1–4. The markers *MP0374* and *Mmerhk3b* showed high genetic differentiation ($F_{ST} \geq 0.25$; $G_{ST} \geq 0.15$) with 100% private alleles and could individually assign each genotype to the correct species with high accuracy (Fig. 1; Fig. 2). All the other markers cannot individually discriminate the species even though some show 100% private alleles.

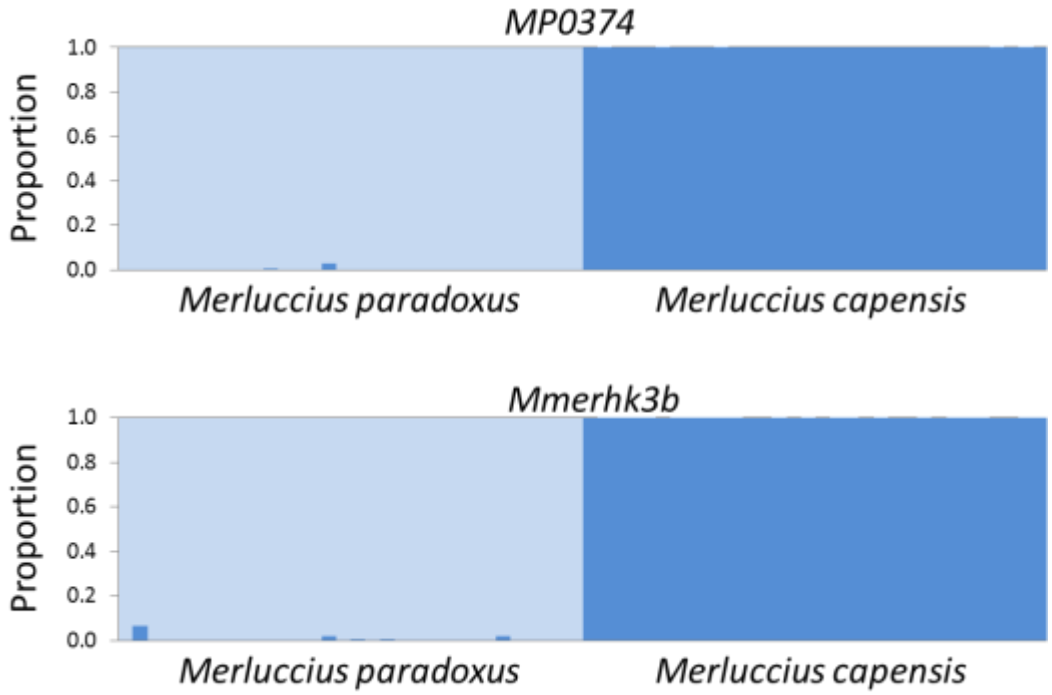

Figure 2. Bayesian cluster analyses illustrating the distinction between samples of *Merluccius paradoxus* (n=32) and *M. capensis* (n=32) at the markers *MP0374* and *Mmerhk3b*. These two markers have the highest F_{ST} values and can identify the species.

The effect of directional/balancing selection on the discriminating power of the microsatellites was investigated using LOSITAN (Antao *et al.*, 2008). The program compares observed F_{ST} values to 10^6 coalescent simulations obtained under mutation–drift equilibrium and a stepwise mutation model. Variation at several loci deviated significantly from neutral expectations but only F_{ST} values of *MP0374* and *Mmerhk3b* were found above the 99% confidence interval (Fig. 3), suggesting directional selection. Considering the long divergence of the species (Grant & Leslie, 2001), the patterns of selection are likely ancient as already observed in other taxa (Narum *et al.*, 2008). Therefore, should divergent populations be included, the two markers will still accurately distinguish the Cape hake species.

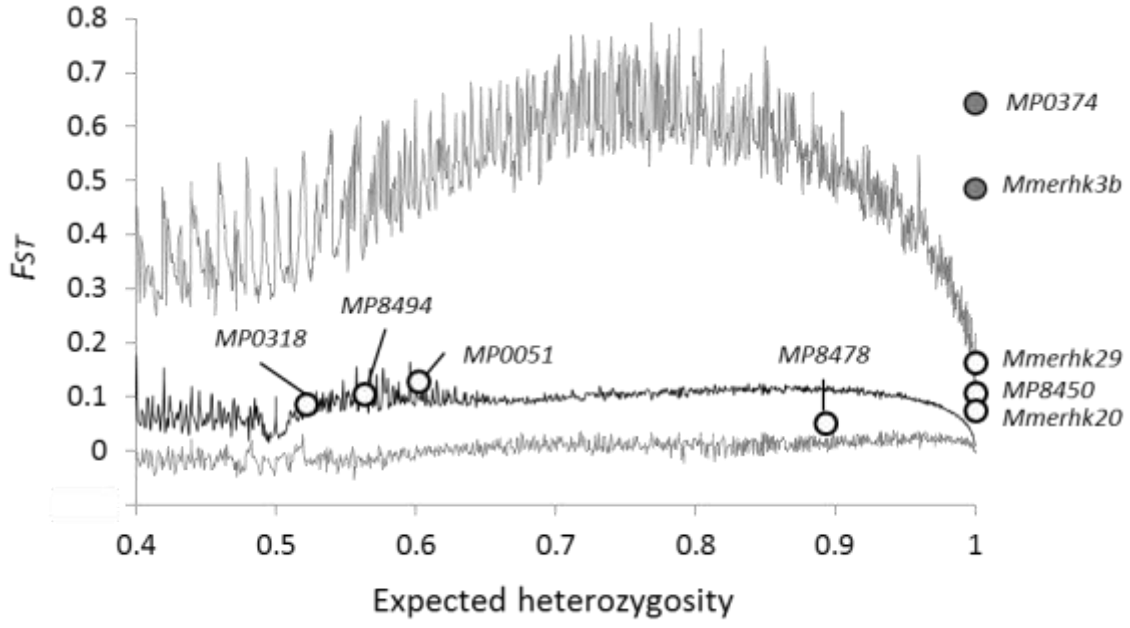

Figure 3. Results of LOSITAN analysis of the two Cape hakes *M. paradoxus* and *M. capensis* for the detection of microsatellites deviating from a model of neutral evolution. The average and 99% confidence limits (three trendlines) of F_{ST} are obtained from simulations under a stepwise mutation model using the weighted mean F_{ST} assumed to be neutral. The circles represent the F_{ST} calculated for each microsatellite marker; the dark grey circles illustrate the markers under directional selection.

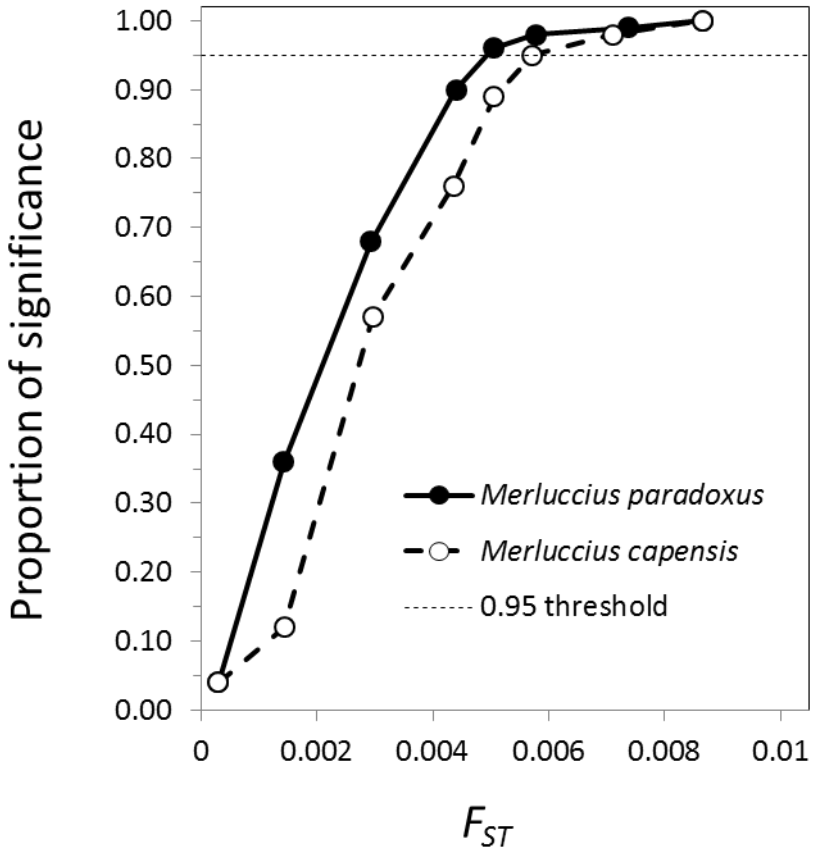

Figure 4. Power analysis of the microsatellite dataset for *Merluccius paradoxus* (14 markers) and *M. capensis* (10 markers) for different F_{ST} values and sampling 50 individuals from each population. The minimum level of genetic differentiation that can be detected with 95% statistical power (resolution of the markers) is F_{ST} = 0.0049 for *M. paradoxus*, F_{ST} = 0.0057 for *M. capensis*.

To evaluate whether the microsatellites are suitable to delimit genetic stocks, the resolution power of each set of markers was assessed using POWSIM 4 (Ryman & Palm, 2006). The test relies on the detection of significant values of F_{ST} estimated at different time points since population divergence under the Wright–Fisher model. For each run, the program randomly assigns alleles among the simulated divergent populations and randomly samples genotypes to estimate F_{ST} values and their significance (Chi-square tests). The parameter sets were $5x10^5$

individuals and 50 samples for each population and 10^3 simulations. The results show that the microsatellite sets are able to detect low genetic differentiation ($F_{ST} < 0.006$) in both deep-water and shallow-water hake (Fig. 4). These F_{ST} values fall within the threshold over which clustering methods can distinguish populations ($F_{ST} = 0.005–0.007$; Waples & Gaggiotti, 2006), and should therefore detect subtle genetic differentiation and demographic independence.

The present study describes molecular tools and results that are relevant for the management of the Cape hakes. First, the two markers with high discriminating power will be important for species identification in several fields including management aspects (distribution area, landings, unintentional harvest, *etc.*), legal actions (*e.g.* poaching, traceability) or ecological surveys (especially ichthyoplankton). Secondly, the sets of microsatellite markers will be used to further investigate the currently recognized management units of Cape hakes. Finally, considering the high level of transferability of microsatellites within (Reid *et al.*, 2012) or across fish families (Carreras-Carbonell *et al.*, 2008), the new markers provide a potential resource for genetic studies of other species within the Gadiformes.

Acknowledgements

The authors thank two reviewers for comments and suggestions. This research was funded by the BCLME/BENEFIT and the Department of Science and Technology African Coelacanth Ecosystem Programme (ACEP) II. DNA samples were provided by S. von der Heyden of

Stellenbosch University. The authors thank Lorraine Harris for laboratory assistance. TB Hoareau was funded by ACEP II and the University of Pretoria's postdoctoral programme.

References

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A. & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a *F(st)*-outlier method. *BMC Bioinformatics* **9**.

Butterworth, D. S. & Rademeyer, R. A. (2005). Sustainable management initiatives for the Southern African hake fisheries over recent years. *Bulletin of Marine Science* **76**, 287-319.

Carreras-Carbonell, J., Macpherson, E. & Pascual, M. (2008). Utility of pairwise mtDNA genetic distances for predicting cross-species microsatellite amplification and polymorphism success in fishes. *Conservation Genetics* **9**, 181-190.

D'Amato, M. E., Lunt, D. H. & Carvalho, G. R. (1999). Microsatellite markers for the hake *Macruronus magellanicus* amplify other gadoid fish. *Molecular Ecology* **8**, 1086-1088.

Evanno, G., Regnaut, S. & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Faircloth, B. C. (2008). MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* **8**, 92-94.

Garcia-Vazquez, E., Machado-Schiaffino, G., Campo, D. & Juanes, F. (2012). Species misidentification in mixed hake fisheries may lead to overexploitation and population bottlenecks. *Fisheries Research* **114**, 52-55.

Grant, W. S. & Leslie, R. W. (2001). Inter-ocean dispersal is an important mechanism in the zoogeography of hakes (Pisces : *Merluccius spp.*). *Journal of Biogeography* **28**, 699-721.

Grant, W. S., Leslie, R. W. & Becker, I. I. (1987). Genetic stock structure of the Southern African hakes *Merluccius capensis* and *Merluccius paradoxus*. *Marine Ecology Progress Series* **41**, 9-20.

Moran, P., Lundy, C., Rico, C. & Hewitt, G. M. (1999). Isolation and characterization of microsatellite loci in European hake, *Merlucius merlucius* (Merlucidae, Teleostei). *Molecular Ecology* **8**, 1357-1358.

Narum, S. R., Banks, M., Beacham, T. D., Bellinger, M. R., Campbell, M. R., Dekoning, J., Elz, A., Guthrie, C. M., Kozfkay, C., Miller, K. M., Moran, P., Phillips, R., Seeb, L. W., Smith, C. T., Warheit, K., Young, S. F. & Garza, J. C. (2008). Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology* **17**, 3464-3477.

Nei, M. (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 583-590.

Payne, A. I. L. & Punt, A. E. (1995). Biology and Fisheries of South African Cape hakes (*M. capensis* and *M. paradoxus*). In Hake: Biology, Fisheries and Markets (Alheit, J. & Pitcher, T. J., eds.), pp. 15-47. London: Chapman & Hall.

Peakall, R. & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537-2539.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Reid, K., Hoareau, T. B. & Bloomer, P. (2012). High-throughput microsatellite marker development in two sparid species and verification of their transferability in the family Sparidae. *Molecular Ecology Resources* **12**, 740-752.

Rico, C., Ibrahim, K. M., Rico, I. & Hewitt, G. M. (1997). Stock composition in North Atlantic populations of whiting using microsatellite markers. *Journal of Fish Biology* **51**, 462-475.

Ryman, N. & Palm, S. (2006). POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes* **6**, 600-602.

Seibert, J. & Ruzzante, D. E. (2006). Isolation and characterization of eight microsatellite loci for white hake (*Urophycis tenuis*). *Molecular Ecology Notes* **6**, 924-926.

Selkoe, K. A. & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9**, 615-629.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.

von der Heyden, S., Lipinski, M. R. & Matthee, C. A. (2007a). Mitochondrial DNA analyses of the Cape hakes reveal an expanding, panmictic population for *Merluccius capensis* and population structuring for mature fish in *Merluccius paradoxus*. *Molecular Phylogenetics and Evolution* **42**, 517-527.

von der Heyden, S., Lipinski, M. R. & Matthee, C. A. (2007b). Species-specific genetic markers for identification of early life-history stages of Cape hakes, *Merluccius capensis* and *Merluccius paradoxus* in the southern Benguela Current. *Journal of Fish Biology* **70**, 262-268.

Waples, R. S. & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology* **15**, 1419-1439.

Zane, L., Bargelloni, L. & Patarnello, T. (2002). Strategies for microsatellite isolation: a review. *Molecular Ecology* **11**, 1-16.

Electronic References

Belkhir, K., Borsa, P., Goudet, J., Chikhi, L. & Bonhomme, F. (2004). GENETIX, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome et Populations, CNRS UPR 9060, Université Montpellier II, Montpellier (France). http://www.univ-montp2.fr/~genetix/genetix.htm.