IN SILICO DETECTION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN EXPRESSED SEQUENCE TAGS OF EUROPEAN SEA BASS

Souche Erika[1], Bart Hellemans[1], Adelino Canario [2], Sven Klages [3], Richard Reinhardt[3] and Filip A.V. Volckaert[1]

[1] Katholieke Universiteit Leuven, Laboratory of Aquatic Ecology, Charles Deberiotstraat 32, B-3000 Leuven, Belgium
E-mail: erika.souche@bio.kuleuven.be

[2] Centro de Ciências do Mar do Algarve (CCMAR), University of Algarbe, 8005-139 Faro, Portugal

[3] Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, D-14195 Berlin, Germany

As a multitude of sequence data are published, discovering polymorphisms bioinformatically becomes a valid option. *In silico* Single Nucleotide Polymorphism (SNP) detection is based on the analysis of multiple alignments. Each column of an alignment is considered a slice containing one base of every sequence aligned. If a mismatch is detected, the slice is further analysed and the mismatch may be reported as a candidate SNP. Around 30,000 European sea bass (*Dicentrarchus labrax*) Expressed Sequence Tags (ESTs) have been sequenced and processed at the Max Planck Institute for Molecular Genetics. Since 55.1% (16,117) ESTs are redundant, they provide a resource for *in silico* SNP discovery. To prevent the detection of sequencing errors, a redundancy of 2 is required for a mismatch to be considered a candidate SNP (Picoult-Newberg *et al.*, 1999). Thus, only contigs containing more than 4 overlapping sequences are analysed. 974 (21.3%) contigs qualify for *in silico* SNPs discovery, representing 5,548 (19%) ESTs and 478,232 base pairs. Various tools are used to detect candidate SNPs; so far, 2 software packages have been tried. 246 candidate SNPs, of which 56 indels, were proposed by SNPServer using default parameters (Savage *et al.*, 2005). Less stringent parameters lead to the discovery of 1027 candidate SNPs, of which 267 indels. PolyBayes (Marth *et al.*, 1999) selected 772 candidates SNPs, of which 231 indels. The Primer3 software is used to design primers flanking each candidate SNP, which are validated in the laboratory by sequencing. Polymorphisms will be mapped, used for selection in aquaculture and the study of adaptation in natural populations.

References

Marth G.T., I. Korf, M.D. Yandell, R.T. Yeh, Z. Gu, H. Zakeri, N.O. Stitziel, P.Y. Hillier Lad. Kwok and W.R. Gish. 1999. A general approach to single-nucleotide polymorphism discovery. Nature Genetics 23:452-456.

Picoult-Newberg L., T.E. Ideker, M.G. Pohl, S.L. Taylor, M.A. Donaldson, D.A. Nickerson and M. Boyce-Jacino. 1999. Mining SNPs from EST Databases. Genome Research 9:167-174.

Savage D., J. Batley, T. Erwin, E. Logan, C.G. Love, G.A.C. Lim, E. Mongin, G. Barker, G.C. Spangenberg and D. Edwards. 2005. SNPServer: a real-time SNP discovery tool. Nucleic Acids Research 33:493-495.