Development of genomic tools in senegalese sole: transcriptome assembly, annotated database, microarray and genome draft

M. Manchado, H. Benzekri, P. Seoane, R. Bautista, J.J. Sánchez, Xavier Cousin, J.V. Planas, L. Rebordinos, M.G. Claros

DEVELOPMENT OF GENOMIC TOOLS IN SENEGALESE SOLE: TRANSCRIPTOME ASSEMBLY, ANNOTATED DATABASE, MICROARRAY AND GENOME DRAFT

Manchado M.[b§], Benzekri H.[b], Seoane P.[b], Bautista R.[c], Sánchez J.J.[d], Cousin X.[e], Planas J.V.[f], Rebordinos L.[g], Claros M.G.[b,c]

[a] *IFAPA Centro El Toruño, Consejeria de Agricultura, Pesca y Medio Rural, 11500 El Puerto de Santa María, Cádiz, Spain*
[b] *Departamento de Biología Molecular y Bioquímica, Universidad de Málaga, Málaga, 29071, Spain*
[c] *Plataforma Andaluza de Bioinformática, Universidad de Málaga, Málaga, 29071, Spain*
[d] *Campus de Ciencias de la Salud, Instituto Nacional de Toxicología y Ciencias Forenses (INT), La Laguna, Spain*
[e] *IFREMER, Laboratoire d'Ecotoxicologie, Place Gaby Coll, BP 7, 17137 L'Houmeau, France.*
[f] *Departament de Fisiologia i Immunologia, Facultat de Biologia, Universitat de Barcelona and Institut de Biomedicina de la Universitat de Barcelona (IBUB), 08028 Barcelona, Spain*
[g] *Laboratorio de Genética. Facultad de Ciencias del Mar y Ambientales, Universidad de Cádiz, Polígono del Río San Pedro, 11510, Puerto Real, Cádiz, Spain*

The Senegalese sole is an economically important flatfish species in fisheries and aquaculture. However, the genomic resources in this species are still scarce. Hence, the transcriptome and genome have been investigated. For transcriptome analysis,more than 1,800 millions reads from different larval and adult tissues were *de novo* assembled resulting in 701,767 tentative transcripts. Orthology analysis using zebrafish as reference identified at least 45,063 putative different transcripts, 18,738 of which were reconstructed with a complete ORF. Moreover, cross-species comparison with the closely-related species *S. solea* and other teleosts identified a set of 14,451 putative transcripts for sole- or lineage-specific genes. As a result, a reference transcriptome including 59,514 transcripts was defined and used to print a sole oligonucleotide microarray containing 43,303 probes. Moreover, a search of molecular markers identified a total of 266,434 SSRs and 337,315 SNPs in the transcriptome. These data and the complete annotation of the transcriptome is available for browsing and downloading at SoleaDB (http://bit.ly/SoleaDB). For genome analysis, ~3.000x10^6 raw reads (including single, paired-end and mate-pair reads from both 454/Roche and Illumina platforms) were processed. Sequences were cleaned using SeqTrimNext, assembled with Ray, reconciled with Gam-NGS, scaffolded with SOAPdenovo2 and SSPACE, and finally gaps closed with Gap closing tool of SOAPdenovo2. Several in-house scripts were developed for contig and scaffold validation and mapping. The 132,712 contigs obtained provide 34,176 scaffolds with a N50 of 85,602 nt. The whole draft genome was ~600 Mb in size and the longest scaffold was 638 kb in length. Mapping of scaffolds onto *Cynoglossus semilaevis* draft genome located 95% of scaffolds (569 Mb in total) onto 21 chromosomes. *In silico* comparison of genetic map markers and scaffold positioning confirmed a linkage correspondence higher than 90%. Moreover, the amplification of 111 predicted SSR markers distributed throught the chromosomes confirmed the accuracy of the assembly obtained. These results confirmed the high similarityo f both flatfish genomes and represent new powerful tools for genomic analysis in *S. senegalensis*. This research was funded by AQUAGENET project program INTERREG IVB SUDOE (ERDF) as well as INIA and EU through the ERDF 2014-2020 "Programa Operativo de Crecimiento Inteligente (RTA2013-00023-C02-01).

Keywords: Sole, transcriptome, genome, draft, database, microarray

[§] Corresponding author. Tel.: +34 671532088; Fax: + 34956011324
E-mail address: + manuel.manchado@juntadeandalucia.es

SANTIAGO DE COMPOSTELA
2015

ISGA XII

INTERNATIONAL SYMPOSIUM ON GENETICS IN AQUACULTURE XII

Santiago de Compostela, Spain
June 21st-27th, 2015