

1 **Mini-review Marine Genomics**

2 **ADVANCES IN EUROPEAN SEA BASS GENOMICS AND FUTURE**

3 **PERSPECTIVES**

4 Bruno Louro<sup>1</sup>, Deborah M. Power<sup>1</sup>, Adelino V.M. Canario<sup>1§</sup>

5

6 <sup>1</sup>CCMAR- Centre of Marine Sciences, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

7 <sup>§</sup>corresponding author; email: [acanario@ualg.pt](mailto:acanario@ualg.pt), phone: +351-289-800925

8

9

10 **Abstract**

11 Only recently available sequenced and annotated teleost fish genomes were restricted to a few model  
12 species, none of which were for aquaculture. Application of Marker Assisted Selection for improved  
13 production traits had been largely restricted to the salmon industry and genetic and Quantitative Trait  
14 Loci (QTL) maps were available for only a few species. With the advent of Next Generation  
15 Sequencing the landscape is rapidly changing and today the genomes of several aquaculture species  
16 have been sequenced. The European sea bass, *Dicentrarchus labrax*, is a good example of a  
17 commercially important aquaculture species in Europe for which in the last decade a wealth of  
18 genomic resources, including a chromosomal scale genome assembly, physical and linkage maps as  
19 well as relevant QTL have been generated. The current challenge is to stimulate the uptake of the  
20 resources by the industry so that the full potential of this scientific endeavour can be exploited and  
21 produce benefits for producers and the public alike.

22

23 **Key words**

24 Genomics; Genetics; Aquaculture; Selective breeding program; European sea bass.

25

## 26 **1. Introduction**

27 As little as a decade ago the only available fish genomes were from model fish species: *Fugu*  
28 *rubripes* [1], *Tetraodon nigroviridis* [2], *Danio rerio* [3], *Oryzias latipes* [4] and *Gasterosteus*  
29 *aculeatus* [5]. With the “next generation sequencing” revolution, the flood of genomic and genetic  
30 data has grown exponentially and recently several genetics and genomics resources, including  
31 transcriptomes and genomes of economically relevant fish species have been published, e.g. [6-9]  
32 and [10] for a review. Despite these advances, so far the impact on aquaculture of new technologies  
33 in genome analysis coupled to a parsimonious breeding program is still limited [11]. This is  
34 particularly true in the Mediterranean area where intensive models of production have only recently  
35 been adopted and few documented examples of structured selective breeding programs exist. The  
36 objective of the present review is to evaluate the status of genomic and genetic tools for the  
37 European sea bass, *Dicentrarchus labrax*, and discuss a conceptual approach for the efficient  
38 application by industry of genomic information into selective breeding programs for this species. The  
39 strategy proposed for implementation of genomic data in a production setting may also be applicable  
40 to newly adopted aquaculture species of interest for which available resources may be limited.

## 41 **2. European sea bass aquaculture history and genetics resources**

42 The European sea bass is a gonochoristic marine teleost fish, distributed in temperate European  
43 coastal areas of the Northeast Atlantic Ocean and Mediterranean Sea. Its intensive exploitation as an  
44 aquaculture species is relatively recent and production is concentrated predominantly in the  
45 Mediterranean basin. It was initially cultivated in semi-intensive lagoon systems but since the 1980's  
46 production has become progressively more intensive due to its high commercial value. Total  
47 production of European sea bass was 126 thousand tonnes in 2010, with a market value of 500  
48 million Euro [12]. The expansion of European sea bass aquaculture production throughout Europe  
49 and the associated increase in its commercial importance has been the catalyst that has led in a

50 relatively short space of time to a significant body of scientific and technical knowledge about this  
51 species. The bulk of the research carried out on the European sea bass has largely occurred over the  
52 past twenty years and encompasses basic biology through to modern day genetics and genomics.  
53 The European sea bass has in the last 10 years moved to the forefront of aquaculture species in terms  
54 of availability of genetic and genomic resources. The production of genomics and genetics tools for  
55 this species has been a community wide effort that has involved numerous scientists in Europe and in  
56 a large part has been driven by European Commission funded consortia. Outputs from such  
57 European projects include high density linkage and synteny maps, a radiation hybrid map,  
58 transcriptome data [13-20], a high quality draft genome sequence (NCBI bioproject  
59 accession: PRJEA39865) [19, 21, 22] and mapped QTLs for economic traits [15, 16, 23-25]. Table 1  
60 lists publicly available genetic, genomic and/or transcriptomics resources for European sea bass and  
61 the source reference. Clearly the next important step is to apply these tools to a long-term and  
62 sustainable breeding program for European sea bass analogous to what has been developed for  
63 terrestrial farm animal production [26, 27].

### 64 **3. Genetics & Genomics trends in research & industry**

65 Selective breeding in aquaculture is mostly done by mass selection of the previous generation, or  
66 through family based selection. While mass selection is based only on selected parentage phenotypic  
67 values to identify the best individuals (selection candidates) in terms of their genetic potential for the  
68 desired traits, within family selection is based on breeding values (calculated through phenotype  
69 measurements and pedigree information) of the fish that is the target of selection and incorporating  
70 information on its relatives [28, 29]. Selection based on genomic information is still a novelty in  
71 aquaculture, and there are relatively few examples of marker assisted selection (MAS) [30, 31]. One  
72 example of successful application of MAS is in salmonids, in which a major quantitative trait locus

73 (QTL) affecting resistance to infectious pancreatic necrosis was selected by incorporating marker  
74 information in the selective breeding program [32].

75 Alternative or complimentary approaches and strategies are required to MAS, which despite its  
76 utility has inherent weaknesses linked to the limited number of QTL flanking markers used which  
77 means only a fraction of the total genetic variance is captured [33]. An alternative approach to  
78 tracing a limited number of QTLs with markers is to trace all the QTL genome wide. This can be  
79 done by dividing the entire genome into chromosome segments, by adjacent markers with such  
80 density that the population-wide linkage disequilibrium between markers and QTL is utilised to  
81 generate the predicted genetic merit of the individual. This method has been termed genomic  
82 selection (GS) [34], but needs a dense set of markers across the genome. Thus, genomic selection  
83 integrated with next-generation-sequencing (NGS) promises to be of great potential to create  
84 genomic information of added value for the accuracy of genomic prediction and genome wide  
85 associations studies (e.g. finding causal mutations). The GS approach can potentially be done either  
86 by genotyping with Restriction site Associated DNA (RAD) [35], Genotyping-by-Sequencing  
87 (GBS) [36], or by whole genome re-sequencing [37] methodologies as illustrated by the 1000 bull  
88 genomes project (<http://www.1000bullgenomes.com/>). Simulations based upon standard aquaculture  
89 breeding practices of the gains (improved growth, disease resistance, etc.) suggest that genome-wide  
90 selection will result in high genetic gain for a typical family [38].

### 91 **3.1 Genomic selection approach**

92 GS can be seen as a new form of scale-up MAS with genetic markers densely covering the whole  
93 genome identifying the full suite of QTLs of a given trait genome-wide. With the ease of production  
94 of large single nucleotide polymorphism (SNP) markers data and lower the genotyping costs, the  
95 limitation today may be in the initial steps, namely of obtaining a reference population with robust  
96 phenotype data (and posterior prediction tuning) for the genomic prediction of phenotypes and

97 breeding values with higher accuracies and better control of inbreeding [39]. To calculate the  
98 genomic estimated breeding value (gEBV), a reference population is genotyped and phenotyped in  
99 order to obtain a prediction equation which basically is the sum of the substitution effects over all  
100 SNPs. Selection candidates can then be screened through genotyping to choose the breeders by and  
101 obtain predictions of the phenotypes [34]. This approach is particularly useful in aquaculture species  
102 that reproduce by mass spawning, as it eliminates the steps of pedigree recording because all  
103 pedigree information and inbreeding control can be inferred from the SNP data [40]. The viability of  
104 applying GS in aquaculture will come from a balance between the cost of dense genotyping and the  
105 added gains the approach delivers compared to traditional mass or family-based selection. The cost  
106 of genotyping will depend on the cost per individual and how many individuals have to be  
107 genotyped, both initially in the test population to obtain marker estimates, and then on the number of  
108 candidates selected based on estimated breeding values.

### 109 **3.2 Transcriptomic approaches**

110 NGS is not only revolutionizing genetics by changing the scale and density of genotyping genome  
111 wide, it is also facilitating the identification of QTL through genome expression studies identify  
112 underlying genes and their respective levels of expression in order to understand the genetic  
113 pathways that underlie the traits effects. The European sea bass whole genome assembly [21] is an  
114 important tool for re-sequencing or gene expression studies. For example, SuperSAGE (Serial  
115 Analysis Gene Expression) combined with next generation sequencing was found to be an effective  
116 means of single nucleotide variant (SNV) calling [41] (Figure 1). SuperSAGE tags, obtained in the  
117 context of a study to analyse growth rates, were mapped to the draft genome sequence of the  
118 European sea bass and resulted in the identification of 506 SNVs and 257 one base indels that were  
119 directly linked to genes, mostly in the 3' UTR of the gene region. From the 763 candidate markers it  
120 was possible to obtain the identity of the genes for about half (387) of the identified polymorphisms.  
121 A total of approx. 11 million edited tags (26bp) that corresponded on average 2 million tags per

122 SAGE library (brain, liver, white muscle), that represented 47.071 unique transcripts were identified.  
123 SuperSAGE or digital gene expression (DGE) potentially has more quantification depth than  
124 RNAseq (mRNA) for the same amount of edited reads output from a sequence run, since in  
125 SuperSAGE one tag corresponds to one transcript molecule count while with RNAseq one transcript  
126 molecule can generate several reads. This is set from library construction, where SuperSAGE reads  
127 are always originated from a single position loci (normally endonuclease EcoP15I restriction site) in  
128 the a cDNA [42], while in RNAseq the reads are randomly originated at any position of the mRNA.  
129 This superSAGE data was obtained by SOLID4 sequencing technology. For SNV calling, the most  
130 relevant characteristic of SOLID sequencing is the double encoded nature of colour space sequences.  
131 This together with fact that read quality could be confirmed by the presence of the adaptor A  
132 sequence at the end of the unedited tag, increasing considerably the confidence regarding SNV false  
133 positive calling. If there was any sequence error within the tag there would be a sequence misframe  
134 and the end adaptorA would not be present.If it had been possible to use RNA samples from  
135 individuals instead of pools of individuals it would have been possible to genotype individuals with  
136 thousands of SNV at the same time obtain the gene expression quantification to be used as trait *per*  
137 *si*. This would discriminate genes allele-specific expression (ASE) and another very interesting  
138 possibility would be to apply genetical-genomics [43] in a single step using the gene expression  
139 values as the trait and the genotypes, instead of methodologies in which gene expression is measured  
140 and genotyping performed separately. RNA-seq can be used even to greater advantage for such  
141 purpose of all in one gene expression Quantitative Trait Locus (eQTL) studies [44]. RNA-seq  
142 delivers ASE data, unlike microarrays, and also RNA-isoform expression discrimination [45].

#### 143 **4. Conclusion**

144 European sea bass is now a member of the restrict club of aquaculture species rich in genomic and  
145 genetic resources, which until recently only included salmonids [46, 47]. The molecular tools and  
146 approach can be applied in selective breeding programs, especially in the case where traits of interest

147 cannot be or are difficult to measure directly in the selection candidate fish, as for example disease  
148 resistance or fillet quality. Not only can they be applied to improve selection accuracy but they can  
149 also be used to manage, and control on-going breeding programs, such as to monitor the  
150 maintenance of suitable levels of additive genetic variation within the broodstocks so that selection  
151 for rapid genetic gain does not lead to decreased heterozygosity and increased inbreeding [48].

152 While until recently the lack of molecular genomic tools has hindered progress for the  
153 implementation of a successful high yield selective breeding program, the limitations have now  
154 shifted onto structural and economic issues related to the lack of appropriate facilities and  
155 infrastructures to maintain established genetic groups, phenotyping methodologies and trained  
156 personnel [46]. In salmon aquaculture the transfer of knowledge from research to industry is a long  
157 term achievement that took years of investment of money and research and the establishment of a  
158 strong collaboration between industry and R&D [11, 29] . The salmon story is decades long and is  
159 one of success with the genetic gain achieved being estimated at approximately 14% per generation  
160 with a global benefit/cost ratio estimated at 15/1 [11]. The current challenge is to stimulate the  
161 uptake of the resources by the European sea bass industry so that the full potential of this scientific  
162 endeavour can be exploited and produce benefits for producers and the public alike.

163

## 164 **Acknowledgements**

165 The research reported has been supported by the European Regional Development Fund (ERDF)  
166 through the COMPETE - Operational Competitiveness Programme and national funds through the  
167 Portuguese Foundation for Science and Technology (FCT) under the project “PEst-  
168 C/MAR/LA0015/2011. BL benefited from a FCT postdoctoral fellowship (SFRH/BPD/89899/2012).

169





171 **References**

- 172 1. Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.-m., Dehal, P., Christoffels, A., Rash,  
173 S., Hoon, S., Smit, A., et al. (2002). Whole-genome shotgun assembly and analysis of the  
174 genome of *Fugu rubripes*. *Science* 297, 1301-1310.
- 175 2. Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L.,  
176 Fischer, C., Ozouf-Costaz, C., Bernot, A., et al. (2004). Genome duplication in the teleost fish  
177 *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431, 946-957.
- 178 3. Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E.,  
179 Humphray, S., McLaren, K., Matthews, L., et al. (2013). The zebrafish reference genome  
180 sequence and its relationship to the human genome. *Nature* 496, 498-503.
- 181 4. Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu,  
182 Y., Doi, K., Kasai, Y., et al. (2007). The medaka draft genome and insights into vertebrate  
183 genome evolution. *Nature* 447, 714-719.
- 184 5. Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R.,  
185 Pirun, M., Zody, M.C., White, S., et al. (2012). The genomic basis of adaptive evolution in  
186 threespine sticklebacks. *Nature* 484, 55-61.
- 187 6. Chen, S., Zhang, G., Shao, C., Huang, Q., Liu, G., Zhang, P., Song, W., An, N., Chalopin, D.,  
188 Volff, J.-N., et al. (2014). Whole-genome sequence of a flatfish provides insights into ZW sex  
189 chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 46, 253-260.
- 190 7. Star, B., Nederbragt, A.J., Jentoft, S., Grimholt, U., Malmstrom, M., Gregers, T.F., Rounge,  
191 T.B., Paulsen, J., Solbakken, M.H., Sharma, A., et al. (2011). The genome sequence of Atlantic  
192 cod reveals a unique immune system. *Nature* 477, 207-210.

- 193 8. Guyon, R., Rakotomanga, M., Azzouzi, N., Coutanceau, J.P., Bonillo, C., D'Cotta, H., Pepey, E.,  
194 Soler, L., Rodier-Goud, M., D'Hont, A., et al. (2012). A high-resolution map of the Nile tilapia  
195 genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* *13*, 222.
- 196 9. Wang, S., Furmanek, T., Kryvi, H., Krossoy, C., Totland, G., Grotmol, S., and Wargelius, A.  
197 (2014). Transcriptome sequencing of Atlantic salmon (*Salmo salar* L.) notochord prior to  
198 development of the vertebrae provides clues to regulation of positional fate, chordoblast  
199 lineage and mineralisation. *BMC Genomics* *15*, 141.
- 200 10. Canario, A.V.M., Bargelloni, L., Volckaert, F., Houston, R.D., Massault, C., and Guiguen, Y.  
201 (2008). Genomics toolbox for farmed fish. *Reviews in Fisheries Science* *16*, 3-15.
- 202 11. Gjedrem, T. (2010). The first family-based breeding program in aquaculture. *Reviews in*  
203 *Aquaculture* *2*, 2-15.
- 204 12. FAO (2012). *The state of the world fisheries and aquaculture*, (Rome: Food and Agriculture  
205 Organization of the United Nations).
- 206 13. Chistiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G.,  
207 Bertotto, D., Libertini, A., and Volckaert, F.A. (2005). A microsatellite linkage map of the  
208 European sea bass *Dicentrarchus labrax* L. *Genetics* *170*, 1821-1826.
- 209 14. Chistiakov, D., Tsigenopoulos, C., Lagnel, J., Guo, Y., Hellemans, B., Haley, C., Volckaert, F.,  
210 and Kotoulas, G. (2008). A combined AFLP and microsatellite linkage map and pilot  
211 comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. *Anim. Genet.* *39*,  
212 623-634.
- 213 15. Massault, C., Hellemans, B., Louro, B., Batargias, C., Van Houdt, J.K.J., Canario, A., Volckaert,  
214 F.A.M., Bovenhuis, H., Haley, C., and De Koning, D.J. (2010). QTL for body weight,  
215 morphometric traits and stress response in European sea bass *Dicentrarchus labrax*. *Anim.*  
216 *Genet.* *41*, 337-345.

- 217 16. Volckaert, F., Hellemans, B., Batargias, C., Louro, B., Massault, C., Van Houdt, J., Haley, C., de  
218 Koning, D.-J., and Canario, A. (2012). Heritability of cortisol response to confinement stress  
219 in European sea bass *dicentrarchus labrax*. *Genetics Selection Evolution* 44, 15.
- 220 17. Guyon, R., Senger, F., Rakotomanga, M., Sadequi, N., Volckaert, F.A.M., Hitte, C., and  
221 Galibert, F. (2010). A radiation hybrid map of the European sea bass (*Dicentrarchus labrax*)  
222 based on 1581 markers: Synteny analysis with model fish genomes. *Genomics* 96, 228-238.
- 223 18. Louro, B., Passos, A.L.S., Souche, E.L., Tsigenopoulos, C., Beck, A., Lagnel, J., Bonhomme, F.,  
224 Cancela, L., Cerdà, J., Clark, M.S., et al. (2010). Gilthead sea bream (*Sparus auratus*) and  
225 European sea bass (*Dicentrarchus labrax*) expressed sequence tags: Characterization, tissue-  
226 specific expression and gene markers. *Marine Genomics* 3, 179-191.
- 227 19. Kuhl, H., Beck, A., Wozniak, G., Canario, A., Volckaert, F., and Reinhardt, R. (2010). The  
228 European sea bass *Dicentrarchus labrax* genome puzzle: comparative BAC-mapping and low  
229 coverage shotgun sequencing. *BMC Genomics* 11, 68.
- 230 20. Kuhl, H., Tine, M., Hecht, J., Knaust, F., and Reinhardt, R. (2011). Analysis of single  
231 nucleotide polymorphisms in three chromosomes of European sea bass *Dicentrarchus*  
232 *labrax*. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 6, 70-  
233 75.
- 234 21. Tine, M., Kuhl, H., Gagnaire, P.-A., Louro, B., Desmarais, E., Martins, R.S.T., Hecht, J., Knaust,  
235 F., Belkhir, K., Klages, S., et al. (2014). The genome of European sea bass (*Dicentrarchus*  
236 *labrax*) provides insights into euryhalinity adaptation and Atlantic-Mediterranean  
237 divergence mechanisms. *Nat Commun*, submitted.
- 238 22. Kuhl, H., Tine, M., Beck, A., Timmermann, B., Kodira, C., and Reinhardt, R. (2011). Directed  
239 sequencing and annotation of three *Dicentrarchus labrax* L. chromosomes by applying

- 240 Sanger- and pyrosequencing technologies on pooled DNA of comparatively mapped BAC  
241 clones. *Genomics* 98, 202-212.
- 242 23. Chatziplis, D., Batargias, C., Tsigenopoulos, C.S., Magoulas, A., Kollias, S., Kotoulas, G.,  
243 Volckaert, F.A.M., and Haley, C.S. (2007). Mapping quantitative trait loci in European sea  
244 bass (*Dicentrarchus labrax*): The BASSMAP pilot study. *Aquaculture* 272, S172-S182.
- 245 24. Dupont-Nivet, M., Vandeputte, M., Vergnet, A., Merdy, O., Haffray, P., Chavanne, H., and  
246 Chatain, B. (2008). Heritabilities and GxE interactions for growth in the European sea bass  
247 (*Dicentrarchus labrax* L.) using a marker-based pedigree. *Aquaculture* 275, 81-87.
- 248 25. Saillant, E., Dupont-Nivet, M., Haffray, P., and Chatain, B. (2006). Estimates of heritability  
249 and genotype–environment interactions for body weight in sea bass (*Dicentrarchus labrax*  
250 L.) raised under communal rearing conditions. *Aquaculture* 254, 139-147.
- 251 26. de Koning, D.-J., Archibald, A., and Haley, C.S. (2007). Livestock genomics: bridging the gap  
252 between mice and men. *Trends Biotechnol.* 25, 483-489.
- 253 27. Sellner, E.M., Kim, J.W., McClure, M.C., Taylor, K.H., Schnabel, R.D., and Taylor, J.F. (2007).  
254 Applications of genomic information in livestock. *J Anim Sci* 85, 3148-3158.
- 255 28. Falconer, D., and Mackay, T. (1996). *Introduction to quantitative genetics*, 4th Edition,  
256 (Harlow, UK: Longman).
- 257 29. Gjedrem, T. ed. (2005). *Selection and breeding programs in aquaculture* (Dordrecht ; New  
258 York: Springer).
- 259 30. Fernando, R., and Grossman, M. (1989). Marker assisted selection using best linear unbiased  
260 prediction. *Genetics Selection Evolution* 21, 467 - 477.
- 261 31. Sonesson, A.K. (2007). Marker-assisted selection in aquaculture breeding programs.  
262 *Aquaculture* 272, *Supplement 1*, S311.

- 263 32. Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J.,  
264 and Bishop, S.C. (2008). Major quantitative trait loci affect resistance to infectious  
265 pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics* 178, 1109-1115.
- 266 33. Dekkers, J.C.M. (2004). Commercial application of marker- and gene-assisted selection in  
267 livestock: Strategies and lessons. *J Anim Sci* 82, E313-E328.
- 268 34. Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). Prediction of total genetic value  
269 using genome-wide dense marker maps. *Genetics* 157, 1819-1829.
- 270 35. Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., and Johnson, E.A. (2007). Rapid and  
271 cost-effective polymorphism identification and genotyping using restriction site associated  
272 DNA (RAD) markers. *Genome Res.* 17, 240-248.
- 273 36. Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell,  
274 S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity  
275 species. *PLoS ONE* 6, e19379.
- 276 37. Stratton, M. (2008). Genome resequencing and genetic variation. *Nat. Biotechnol.* 26, 65-66.
- 277 38. Sonesson, A., and Meuwissen, T. (2009). Testing strategies for genomic selection in  
278 aquaculture breeding programs. *Genetics Selection Evolution* 41, 1-9.
- 279 39. Daetwyler, H.D., Calus, M.P.L., Pong-Wong, R., de los Campos, G., and Hickey, J.M. (2013).  
280 Genomic prediction in animals and plants: Simulation of data, validation, reporting, and  
281 benchmarking. *Genetics* 193, 347-365.
- 282 40. Power, D.M., Louro, B., Houston, R., Anjos, L., and Cardoso, J.C.R. (2011). Genomic-  
283 proteomic research in Sparidae and its application to genetic improvement. In *Sparidae*.  
284 (Wiley-Blackwell), pp. 359-381.
- 285 41. Li, H., Ruan, J., and Durbin, R. (2008). Mapping short DNA sequencing reads and calling  
286 variants using mapping quality scores. *Genome Res.* 18, 1851-1858.

- 287 42. Matsumura, H., Urasaki, N., Yoshida, K., Krüger, D., Kahl, G., and Terauchi, R. (2012).  
288 SuperSAGE: Powerful Serial Analysis of Gene Expression. In RNA Abundance Analysis,  
289 Volume 883, H. Jin and W. Gassmann, eds. (Humana Press), pp. 1-17.
- 290 43. Jansen, R.C., and Nap, J.-P. (2001). Genetical genomics: the added value from segregation.  
291 Trends Genet. 17, 388-391.
- 292 44. Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010).  
293 High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science  
294 329, 643-648.
- 295 45. Sun, W., and Hu, Y. (2013). eQTL mapping using RNA-seq data. Statistics in Biosciences 5,  
296 198-219.
- 297 46. Fjalestad, K.T., Moen, T., and Gomez-Raya, L. (2003). Prospects for genetic technology in  
298 salmon breeding programmes. Aquaculture Research 34, 397-406.
- 299 47. Moen, T., Baranski, M., Sonesson, A., and Kjøglum, S. (2009). Confirmation and fine-mapping  
300 of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo*  
301 *salar*): population-level associations between markers and trait. BMC Genomics 10, 368.
- 302 48. Muir, J. (2005). Managing to harvest? Perspectives on the potential of aquaculture. Philos T  
303 R Soc B 360, 191-218.
- 304 49. Magnanou, E., Klopp, C., Noirot, C., Besseau, L., and Falcón, J. (2014). Generation and  
305 characterization of the sea bass *Dicentrarchus labrax* brain and liver transcriptomes. Gene  
306 544, 56-66.
- 307 50. Sarropoulou, E., Galindo-Villegas, J., García-Alcázar, A., Kasapidis, P., and Mulero, V. (2012).  
308 Characterization of European sea bass transcripts by RNA Seq after oral vaccine against *V.*  
309 *anguillarum*. Marine Biotechnology 14, 634-642.

- 310 51. Ferraresso, S., Milan, M., Pellizzari, C., Vitulo, N., Reinhardt, R., Canario, A., Patarnello, T.,  
311 and Bargelloni, L. (2010). Development of an oligo DNA microarray for the European sea  
312 bass and its application to expression profiling of jaw deformity. *BMC Genomics* *11*, 354.
- 313 52. Geay, F., Ferraresso, S., Zambonino-Infante, J., Bargelloni, L., Quentel, C., Vandeputte, M.,  
314 Kaushik, S., Cahu, C., and Mazurais, D. (2011). Effects of the total replacement of fish-based  
315 diet with plant-based diet on the hepatic transcriptome of two European sea bass  
316 (*Dicentrarchus labrax*) half-sibfamilies showing different growth rates with the plant-based  
317 diet. *BMC Genomics* *12*, 522.
- 318 53. Darias, M.J., Zambonino-Infante, J.L., Hugot, K., Cahu, C.L., and Mazurais, D. (2008). Gene  
319 expression patterns during the larval development of European sea bass (*Dicentrarchus*  
320 *Labrax*) by microarray analysis. *Marine Biotechnology* *10*, 416-428.
- 321 54. Souche, E.L., Hellemans, B., Van Houdt, J.K., and Volckaert, F.A. (2012). Characterisation  
322 and validation of single-nucleotide polymorphism markers in expressed sequence  
323 tags of European sea bass. *Molecular Ecology Resources* *in press*.
- 324 55. Homer, N., Merriman, B., and Nelson, S.F. (2009). BFAST: An alignment tool for large scale  
325 genome resequencing. *PLoS ONE* *4*, e7767.
- 326 56. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G.,  
327 Durbin, R., and Subgroup, G.P.D.P. (2009). The sequence alignment/map format and  
328 SAMtools. *Bioinformatics* *25*, 2078-2079.
- 329 57. Goecks, J., Nekrutenko, A., Taylor, J., and Team, T.G. (2010). Galaxy: a comprehensive  
330 approach for supporting accessible, reproducible, and transparent computational research  
331 in the life sciences. *Genome Biology* *11*, R86.

332

333



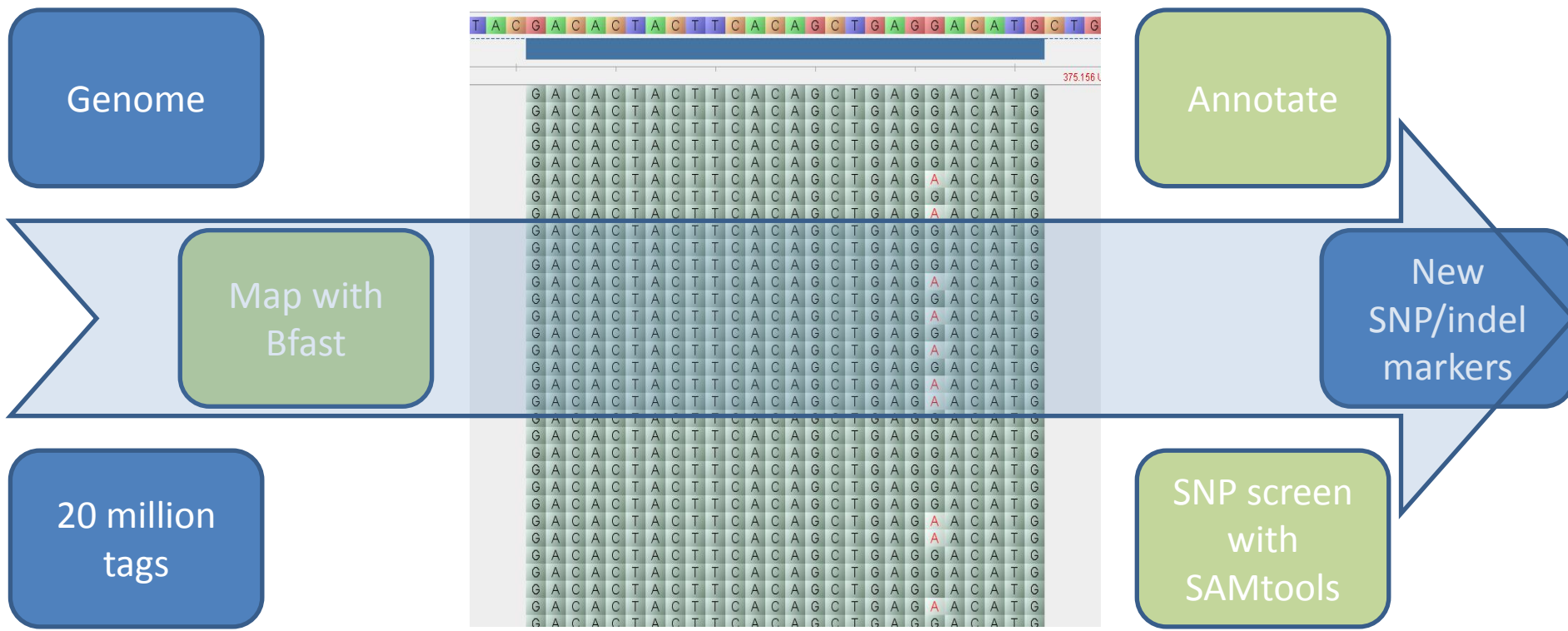


335 **Figure legends**

336 **Figure 1.** SNV discovery and mapping with SuperSAGE. Solid4 colour-space 26bp SuperSAGE  
337 reads were aligned with the European sea bass genome using bfast-0.6.4e [55]. A SAM mapping  
338 file for aligned and matched tags was sorted, indexed and converted to BAM format using “faidx”,  
339 “sort”, “index”, and “view” scripts from SAMtools program [<http://samtools.sourceforge.net>; 56].  
340 SNV and indels (only one nucleotide) were filtered from the genome tag alignments containing more  
341 than 10 overlapping tag sequences. SAMtools scripts (“pileup” and “varFilter”) were used to select  
342 SNV and indel candidates over a threshold of 20 and 50 Phred-scaled likelihood (aka SNP quality)  
343 respectively resulting in 506 SNVs and 257 indel. Identified and retrieved polymorphisms were  
344 annotated using their positional coordinate against European sea bass GFF3 genomic gene annotation  
345 file. This was done with “Operate on Genomic Intervals” tool in the Galaxy server  
346 [<https://main.g2.bx.psu.edu>; 57].

347

Figure1



**Table 1.** Genetic, genomic and transcriptomics publicly available resources.

Resource type	Resource description	Year	Reference	Accession #
Genomic	Genome project	2011, 2014	[21, 22]	PRJEA39865
Genomic	Comparative BAC ends mapping	2010	[19]	FN436279 to FN538968
Genomic	Radiation hybrid map	2010	[17]	-  FM178562 to FM178778,
Transcriptomics	ESTs and <i>de-novo</i> RNA-seq assemblies	2010, 2012, 2014	[18, 49, 50]	SRA050000, E-MTAB- 1867
Transcriptomics	Oligo DNA microarray	2008, 2010, 2011	[51-53]	PRJNA120433 PRJNA120529 PRJNA138507
Genetic	Growth and stress related QTLs	2007, 2010	[15, 23]	-
Genetic	Growth and stress related heritability's estimations	2006, 2008, 2012	[16, 24, 25]	-
Genetic	1 <sup>st</sup> and 2 <sup>nd</sup> generation linkage maps	2005, 2008	[13, 14]	Notes at PMC1449790
Genetic	SNV calling	2011, 2012	[20, 54]	FQ310506 to FQ310508
Transcriptomics	Oligo DNA microarray. Immune response to stressor	2011	-	PRJNA138797
Transcriptomics	RNA-seq/ <i>de novo</i> assembly	2014	-	PRJEB4602
Metagenomics	Gut metagenome	2012	-	PRJNA171730