View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Stirling Online Research Repository (RIOXX)

1 **A description of the origins, design and performance of the TRAITS / SGP Atlantic salmon**

2 **(*Salmo salar* L.) cDNA microarray**

3

4 J. B. TAGGART *†, J. E. BRON *, S. A. M. MARTIN ‡, P. J. SEEAR §#, B. HØYHEIM ¶, R.

5 TALBOT ¦, L. VILLENEUVE *, G. E. SWEENEY §, D. F. HOULIHAN ‡, C. J. SECOMBES ‡, D.

6 R. TOCHER * AND A. J. TEALE *

7

8 ** Institute of Aquaculture, University of Stirling, Stirling, FK9 4LA, Scotland, U.K., ‡ Scottish Fish*

9 *Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Tillydrone*

10 *Avenue, Aberdeen, AB24 2TZ, Scotland, U.K., § School of Biosciences, Cardiff University, Museum*

11 *Avenue, Cardiff, CF10 3US, U.K., ¶ Norwegian School of Veterinary Science, BasAM-Genetics, PO*

12 *Box 8146 DEP, NO-0033, Oslo, Norway and ¦Roslin Institute, Roslin, Midlothian, EH 25 9PS,*

13 *Scotland, U. K.*

14

15

16 Running Head: TRAITS / SGP cDNA microarray

17

18

19

20

21

22 # Present address: British Antarctic Survey, Biological Sciences Division, High Cross, Madingley

23 Road, Cambridge, CB3 0ET, U.K.

24

25 † Author to whom correspondence should be addressed. Tel No.: +44 1786 467927;

26 fax +44 1784 472133; email: j.b.taggart@stir.ac.uk

27

27 **Abstract**

28 The origins, design, fabrication and performance of an Atlantic salmon microarray is described. The
29 microarray comprises 16,950 Atlantic salmon derived cDNA features, printed in duplicate and
30 mostly sourced from pre-existing EST collections (SALGENE and SGP) but also supplemented with
31 cDNAs from suppression-subtractive hybridisation (SSH) libraries and candidate genes involved in
32 immune response, protein catabolism, lipid metabolism and the parr-smolt transformation. A
33 preliminary analysis of a dietary lipid experiment identified a number of genes known to be involved
34 in lipid metabolism. Significant fold-change differences (as low as 1.5×) were apparent from the
35 microarray analysis and were confirmed by qRT-PCR. The study also highlighted the potential for
36 obtaining artifactual expression patterns as a result of cross-hybridisation of similar transcripts.
37 Examination of the robustness and sensitivity of the experimental design employed, demonstrated the
38 importance of biological replicates over technical (dye flip) replicates for the studied system. The
39 TRAITS / SGP microarray has been proven, in a number of studies, to be a powerful tool for the
40 study of key traits of Atlantic salmon biology. It is now available for use by researchers in the wider
41 scientific community.

42

43

44

45 Key Words: Atlantic salmon; DNA microarray; gene expression; lipid metabolism; immune
46 response; smoltification.

47

47 **INTRODUCTION**

48

49 The Atlantic salmon (*Salmo salar* L.) is an important farmed fish species throughout its native range
50 (western Europe and east coast North America). Over the past 10 years a substantial Atlantic salmon
51 mariculture industry has also been established in Chile. In indigenous areas the wild species also
52 underpins valuable sectors of the rural economy founded on sport and ecotourism. Its biology is
53 unusual in that, as an anadromous species, it adapts to very different environments in terms of
54 temperature and salinity at different stages in its life-cycle. Atlantic salmon is a high value food
55 source providing quality protein and oils, and together with other oily fish it is the most important
56 source of essential "omega 3" fatty acids in the human diet. Data compiled from FAO UN database
57 (Fishstat Plus) indicates that farmed Atlantic salmon production in the world exceeds 1.2 million
58 tonnes per annum.
59 Currently, the four most important constraints on commercial production of this species are a) supply
60 of dioxin-free highly unsaturated oils for the salmon diet, b) protein growth efficiency, c) infectious
61 disease, and d) a long and complex lifecycle. In 2002, research groups from three UK universities
62 (Aberdeen, Cardiff and Stirling), specialising in different aspects of Atlantic salmon biology, formed a
63 partnership to develop and exploit transcriptomics resources to explore the molecular basis of the
64 biology underlying these constraints. The goal of TRAITS (TRanscriptome Analysis of Important
65 Traits of Salmon, www.abdn.sfirc/salmon) was to bolster the sustainability of Atlantic salmon farming
66 through identification of genes and metabolic pathways influencing traits that are important in terms of
67 a) efficiency and sustainability of farm production, b) welfare of farmed stocks and c) quality and
68 nutritional value of salmon products for the consumer. This goal was to be achieved through selection
69 of a set of key "indicator" genes associated with the traits of interest, in order to form the basis of a
70 prototype DNA chip for monitoring salmon health and performance. The underlying strategy (Fig 1.)
71 was to design a primary cDNA microarray based on extant EST collections together with novel ESTs
72 derived from genes shown to respond, in expression terms, to relevant laboratory / field "challenges".

73 RNA samples derived from these and other challenges would be interrogated by the cDNA array to
74 identify candidate responder genes. A second more focused oligonucleotide array, comprising mainly
75 responder genes, would then be fabricated and initially validated by interrogation of the same samples
76 that were hybridised to the cDNA array.
77
78 The development of the TRAITS cDNA microarray was initially reliant upon a *c.* 11K EST collection
79 from the EC funded SALGENE project ("Construction of a genetic body map for Atlantic salmon";
80 FAIR CT98 4314), in which Stirling had been a partner. However, prior to project start-up a formal
81 collaboration with the Norwegian Salmon Genome Project (SGP, www.salmongenome.no) was
82 developed that allowed access to a more extensive physical EST resource (Hagen-Larsen *et al*.,
83 2005; Adzhubei *et al*., 2007).
84
85 Several cDNA microarray platforms have been established for salmonid fish with varying numbers
86 of features; Koskinen *et al*., 2004, (1380 features), Rise *et al*., 2004 (3700 features), Ewart *et al*.,
87 2005 (4104 features), Jordal *et al*., 2005 (79 features) von Schalburg *et al*., 2005 (16008 features). A
88 number of different biological processes have been examined by microarray in salmonid fish
89 including immune responses to bacterial infections (Ewart *et al*., 2005; MacKenzie *et al*., 2006;
90 Martin *et al*., 2006; Rise et al., 2004; von Schalburg *et al*., 2005), viral infections (Purcell *et al*.,
91 2006), fungal infections (Roberge *et al*., 2007), and to physiological states such as nutrition (Jordal *et*
92 *al.,* 2005), mobilisation of energy reserves (Salem *et al*., 2006) and stress (Cairns *et al*., 2007). In the
93 above studies varying complexities of experimental design were employed, using different numbers
94 of replicates and often only genes showing two-fold or higher differences in expression were selected
95 for further scrutiny / characterisation.
96
97 It is widely accepted that the supply of fish meal and oils for the diets of farmed fish is not
98 sustainable (Sargent & Tacon, 1999). One consequence of this has been an increased emphasis on the

99 development of diets based on vegetable oil, which are suitable for aquaculture. A key aspect of
100 salmonid lipid metabolism that is being extensively investigated is this regard is the conversion of C_{18}
101 polyunsaturated fatty acids (PUFAs), found in vegetable oils to the C_{20} and C_{22} highly unsaturated fatty
102 acids (HUFAs), eicosapentaenoate (EPA) and docosahexaenoate (DHA), that are the specific omega-3
103 fatty acids responsible for the health-promoting properties of fish and fish oil. These components are
104 critical to the maintainance of nutritional quality of farmed fish. Two key enzymes involved in this
105 pathway, Δ5 and Δ6 fatty acyl desaturase (FAD), have been characterised in depth (Hastings *et al.*,
106 2004; Zheng *et al.*, 2005*a*). Both these genes have been shown, by quantitative real-time PCR analysis
107 (qRT-PCR), to exhibit diet-dependent differential gene expression (Zheng *et al.*, 2004, 2005*a, b*),
108 though detected fold-change differences are minimal (1.3 – 2.0).
109
110 In this paper, design and fabrication details relating to the TRAITS / SGP cDNA microarray are
111 reported. Results of a preliminary transcriptomic analysis from a dietary lipid feeding trial are used
112 to explore both the robustness and the sensitivity of analysis that may be achieved by the use of this
113 microarray.
114
115 **MATERIALS & METHODS**
116
117 cDNA RESOURCES
118
119 *Archived ESTs*
120
121 Two main EST collections (SALGENE *c.* 11K clones and SGP *c.* 30K clones) were available to the
122 TRAITS consortium in 2004. In all cases, fish were sourced from farm stocks of European origin and
123 library construction began with mRNAs from tissues being used as template for oligo(dT)-primed

124 reverse transcription. The SALGENE resource comprised ESTs from seven tissue-specific non-
125 normalised libraries and two tissue-specific normalised libraries, with all cDNAs being directionally
126 cloned into vectors. Details of tissues used, life-history stage (juvenile *i.e.* freshwater phase or adult *i.e.*
127 marine phase) and cloning systems employed are given in Table I. Non-normalised library construction
128 has been detailed elsewhere (Martin *et al*., 2002; Hagen-Larsen *et al*., 2005; Adzhubei *et al*., 2007).
129 Insert size varied among libraries but ranged from 300 bp – 4 kbp. Single pass sequence data (5' end)
130 were available for all clones. Normalised libraries were made in M.B. Soares' laboratory, University of
131 Columbia, USA, following their standard methodology (Bonaldo *et al*., 1996). Single pass sequence
132 data available for these clones was a mixture of both 5' and 3' end reads. SGP clones were derived
133 from 14 tissue-specific non-normalised libraries; brain, eye, gill, head-kidney, heart, intestine, kidney,
134 liver, white muscle, ovary, skin, spleen, swim-bladder and testis. All tissues were sampled from parr
135 (freshwater phase). The cDNAs were directionally cloned into pBlueScript II SK(+) XR phagemid
136 vector and transformed into XL10–Gold host cells (Hagen-Larsen et al., 2005; Adzhubei *et al*., 2007).
137 EST data comprised single pass 5'-end sequences.

138

139 *Trait-specific enriched libraries*

140

141 All enrichments used a standard approach – suppression subtractive hybridisation (SSH; PCR-select
142 cDNA subtraction kit, Clontech). Non-directional cloning was subsequently used to insert subtracted
143 cDNA fragments into the pGEM T-Easy vector (Promega). Resultant cDNA fragments generally
144 ranged in size between 150-700 bp.

145

146 Immune response genes

147

A bacterial challenge was undertaken using *Aeromonas salmonicida,* the bacterial pathogen responsible for furunculosis of salmon. Three tissue-specific enriched libraries (head kidney, gill, and liver) were constructed (Martin *et al*., 2006). Fish were anaesthetised with benzocaine (Sigma 20 mg L^{-1}) and injected intraperitoneally with 100 μL (10^9 CFU mL^{-1}) of a genetically attenuated strain ($aroA^-$) of *A. salmonicida* (Brivax II, Marsden *et al*., 1996). in phosphate buffered saline (PBS) or 100 μL of PBS as control. Brivax II is a non virulent strain, but acts in a similar manner to the intact virulent pathogen, however, after several rounds of replication the fish clears the bacteria. Intraperitoneal injection of Brivax II induces a protective immune response, with fish resistant to a later challenge with virulent strains. The two groups of fish were kept separately and RNA was pooled from 10 'challenged' fish and 10 control fish at 24 h and 48 h post injection. In each case the challenged RNA was "tester" and the control "driver". Approximately 500 clones from each library were sequenced, a mean redundancy of *c.* 33% being observed. Following BLASTx sequence homology matching, 20%, 23% and 50 % of genes had sequences homologous to immune-associated genes for head-kidney, gill and liver respectively (Martin *et al*., 2006).

Protein catabolism genes

Two tissue-specific enriched libraries (white muscle and liver) were constructed following a starvation trial. RNA was pooled from 10 fish starved for 14 days and from 10 fish fed *ad libitum*. RNA from the starved pool was used as "tester" and RNA from fed fish as "driver". For genes enriched following short term starvation, a highly heterogeneous group of genes was found, as many different biological processes were altered by this treatment including those related to protein turnover. For the library generated from liver, 92% of the sequences were found to have homologies following BLASTx searches. Key groups of genes represented in the library encoded metabolic enzymes, serum proteins and immune response genes, with other minor groupings being iron-binding proteins, globins and factors involved in transcription and translation. For the genes enriched in

174 muscle following starvation, 77% had BLASTx homologies, with key groups of sequences encoding
175 metabolic enzymes, structural proteins and transcription and translation factors (10%) and minor
176 groups including heat shock proteins.
177
178 Diet responsive genes
179
180 Atlantic salmon were fed from first-feeding on diets containing either fish oil (FO; capelin oil) or a
181 25% FO : 75% blended vegetable oil (VO) diet (detailed in 'feed trial' section below). Four
182 subtracted liver cDNA libraries (two timepoints – 52 weeks (pre-smolt / freshwater) and 55 weeks
183 (post-smolt / seawater) and two directions, FO driver and VO driver) were made. Pooled RNA from
184 12 (pre-smolt) and 4 (post-smolt) female fish on each diet were used. A total of 768 clones were
185 sequenced. All four libraries were found to be highly redundant; Ten fragments comprised *c.* 40% of
186 all sequences. BLASTx analyses gave significant hits (e-value < e^{-20}) for 79 (54%) of the 145
187 different sequences, though only one of these (catfish fatty-acid binding protein) appeared to be
188 directly related to lipid metabolism.
189
190 Smoltification responsive genes
191
192 Four tissue-specific SSH libraries enriched in genes up-regulated in brain, pituitary, kidney and gill
193 of smolts (seawater phase) in comparison to parr (freshwater phase) were made. Tissues were
194 dissected from parr and smolt in November / December 2002 and April / May 2003. A total of 380
195 clones from each of the four libraries was sequenced. Between 32% and 50% of the sequenced
196 clones were identified by BLASTx sequence homology searches. All libraries had considerable
197 redundancy (28% brain, 56% gill, 56% kidney, 86% pituitary). Significantly, the single most
198 abundant sequence in the gill subtractive library corresponded to Na^+/K^+ ATPase whose levels of
199 activity are a key indicator of smoltification status.

200

201 *Candidate and other genes*

202

203 A third, minor source of cDNAs for the array, was a small collection of full length genes or gene
204 fragments in plasmid constructs that were already possessed by the partners. These included both
205 candidate genes and also other genes with no known relevance to the specific traits of interest (Table
206 II). Being better characterised than the EST clones these constructs were potentially useful as reference
207 genes on the microarray. Of note, with regard to the research reported in this paper, is the inclusion of
208 three cDNA fragments (the ORF and two 3' UTR fragments) from both (Δ5 FAD and Δ6 FAD)
209 Atlantic salmon fatty acyl desaturase genes.

210

211 CLUSTER ANALYSIS AND PROBE SELECTION

212

213 All sequence data derived from the above resources, together with 57K Atlantic salmon sequences
214 available *in silico* from Genbank in July 2004, were clustered using the TGI clustering tools (Pertea
215 *et al*., 2003). The process of clone selection for the cDNA microarray is summarised in Fig. 2.
216 Approximately equal numbers (*c.* 9 K) of contigs and singletons were identified as having accessible
217 clones. Since the overall number (*c.* 18 K clones) was comfortably within the printing capacity of the
218 microarray spotter no further clone selection / refinement was undertaken. A single representative
219 clone from each contig was selected for inclusion on the microarray. Where possible a SALGENE
220 clone was selected in preference to an SGP clone, since the entire SALGENE resource was archived
221 at the printing site (ARK Genomics, Roslin Institute). Also, clones were selected from non-
222 normalised or normalised libraries in preference to SSH derived cDNAs in order to take advantage of
223 longer transcripts.

224

225 MICROARRAY FABRICATION

226

227 Bacterial cultures were grown from archived samples and cDNA inserts amplified directly from
228 these using vector-specific primers. Initially various primer sets were used, dependent on the specific
229 vector employed. However, a generic set (BSKS-F: CGATTAAGTTGGGTAACGC; BSKS-R:
230 CAATTTCACACAGGAAACAG) was found to work with all but one vector. For pBK-CMV
231 constructs, T3 (AATTAACCCTCACTAAAGGGA) and T7 (TAATACGACTCACTATAGGG)
232 primers were employed. PCR reactions (50 µL) were purified by passing them through a 384-well
233 multiscreen filter plate (Millipore). Amplicons were electrophoresed through a 1% agarose gel, and
234 ethidium bromide stained. Only PCR products producing a clear singlet band were considered for
235 spotting. Amplicons were quantified using PicoGreen assay (Invitrogen). Features were printed onto
236 amino-silane coated glass slides (Corning GAPS II) using a MicroGrid II printer (Genomic
237 Solutions). DNA was resuspended in printing buffer (150 mM sodium phosphate buffer, 0.1% SDS,
238 pH8.0) to a concentration of *c.* 150 ng $mL^{-1}$and spotted with a 48 pin tool (Biorobotics 10K pins).
239 Mean spot diameter was 110 microns. The slide format comprised 48 sub-arrays, each consisting of
240 27 columns and 28 rows. Each cDNA was printed in duplicate, with duplicate features being printed
241 non-adjacently within the same sub-array. A number of control features were printed across the
242 microarray; sonicated Atlantic salmon genomic DNA (96 features); sheared salmon sperm DNA
243 (*Oncorhynchus* derived – Sigma; 96 features); four different SpotReport® (Stratagene) controls –
244 namely PCR product 1-3 (Cab, RCA, rbcL genes from *Arabidopsis thaliana*), and human β-actin
245 PCR fragment (20 features each) and spotting buffer (192 features). In addition each sub-array had
246 two Cy3 spots (landing lights) located at the upper left corner to aid orientation of the slide during
247 grid placement and spot finding procedures. The remaining 1826 locations on the 36,288 spot grid
248 were left blank. Following printing DNA spots were fixed by baking at 80°C for 2 h. Prior to
249 hybridisation microarray slides were treated using succinic anhydride and 1-methyl-2-pyrrolidinone
250 (Sigma) to block unbound amino groups (slide manufacturer's recommended protocol) and

251 denatured by incubation in 95°C MilliQ water for 2 min. Slides were then rinsed twice in
252 isopropanol, centrifuged to dry and stored in a desiccated environment until required. Details of the
253 TRAITS cDNA microarray have been submitted to the ArrayExpress platform
254 (www.ebi.ac.uk/arrayexpress) under accession number A-MEXP-664. The GAL file is also available
255 for inspection from the TRAITS website (www.abdn.ac.uk/sfirc/salmon/). All clones used on the
256 microarray are archived in 384 well plates as glycerol stocks in two locations (ARK Genomics
257 facility, Roslin Institute UK; SGP Genetics laboratory, University of Oslo).
258
259 EST ANNOTATION
260
261 Gene identification was carried out by BLAST searches (Altschul *et al*., 1990) of appropriate
262 databases (NCBI nr nucleotide and protein databases) and interrogation of the TIGR Atlantic salmon
263 gene index, release 3.0, (biocomp.dfci.harvard.edu/tgi/cgi-bin/tgi/gimain.pl?gudb=salmon). Gene
264 ontology identifiers were obtained through Blast2GO (Conesa *et al*., 2005).
265
266 SENSITIVITY OF MICROARRAY PROTOCOL
267
268 *Feed trial*
269
270 The effect of replacing fish oil (FO) with vegetable oil (VO) at a replacement level of 75% was
271 investigated in Atlantic salmon in a trial conducted over an entire two-year production cycle
272 (Torstensen *et al.*, 2005). Briefly, the two diets were fed to triplicate tanks / cages at Marine Harvest
273 Ltd. facilities at Invergarry (freshwater) and Loch Duich, Lochalsh (seawater), Scotland. Atlantic
274 salmon fry were distributed randomly into 6 tanks (3m x 3m, depth 0.5m) at a stocking level of 3000
275 / tank, and weaned onto extruded feeds containing 20% added oil which was either FO (capelin oil)
276 or a VO blend, containing rapeseed, palm and linseed oils in a 3.7 : 2 : 1 ratio, replacing 75% of the

277 FO. This VO blend was formulated to mimic fish oil in saturated and monounsaturated fatty acid
278 content *but* with C_{18} PUFA replacing n-3 HUFA. Fish were fed the diets described above for one year
279 until seawater transfer, at which point fish (mean weight ~ 50g) were transferred into 5m × 5m net
280 pens at 700 fish / pen. The fish were fed the same diet in seawater as in freshwater although the
281 dietary oil levels were increased to 25% (3mm pellet) rising to 32% (9mm pellets) through the year-
282 long seawater phase. The diets aimed to be practical, and were formulated and manufactured by
283 Skretting ARC, Stavanger, Norway according to current practices in the salmon feed industry. All
284 diets were formulated to satisfy the nutritional requirements of salmonid fish (NRC, 1993)
285
286 *Sample preparation*
287
288 Fish fed on each diet were sampled at two timepoints during freshwater rearing (at 36 & 52 weeks
289 post-hatch, the latter just one week before transfer to sea) and at a further two time-points in seawater
290 (at 55 and 86 weeks post-hatch). Twenty four liver samples per dietary treatment and timepoint were
291 collected. Total RNA was isolated by organic solvent extraction (TriReagent, Sigma) following the
292 manufacturer's protocol. Spectrophotometry (Nanodrop) and electrophoresis (Bioanalyser 2100,
293 Agilent Technologies) were used to quantify and assess quality of the RNAs respectively. For the
294 transcriptomics analysis, equal amounts of RNA from four individuals (2 males + 2 females) were
295 pooled to produce six biological replicates per diet per time-point. Each pooled RNA sample was
296 further cleaned by mini spin-column purification (RNeasy, Qiagen) and was re-quantified and quality
297 assessed as above.
298
299 *Experimental design*
300
301 Each biological replicate was co-hybridised in a two dye experiment with a single pooled reference
302 sample. This design permits valid statistical comparisons across both diets and time-points to be

303 made. The pooled reference sample comprised equal amounts of RNA from each of the 24 biological
304 replicate samples. A dye-swap procedure was incorporated to mitigate selective binding and
305 scanning artifacts. Thus the entire experiment comprised 96 separate hybridisations (2 diets × 4 time-
306 points × 6 biological replicates × 2 dye-swaps).
307
308 *Labelling and hybridisation protocols*
309
310 Due to the large number of hybridisations, not all hybridisations could be completed at the same
311 time. Samples were therefore randomised and processed in two batches one week apart. RNA was
312 reverse transcribed and labelled with either Cy3 or Cy5 fluors using the FAIRPLAY II cDNA
313 indirect labelling kit (Stratagene) according to the manufacturer's instructions. Briefly 20μg total
314 RNA was reverse transcribed after being primed with oligo dT, which incorporated aminoallyl-dUTP
315 into the synthesised cDNA strand. The RNA template was then hydrolysed using 1M NaOH for 15
316 min and neutralised with 1M HCl. The cDNA was NaAce / ethanol precipitated overnight. cDNA
317 pellets were washed in 80% ethanol and air dried before being resuspended in 5 μL 2X coupling
318 buffer. Once the cDNA had fully dissolved (after at least 30 min) 5 μL of either Cy3 or Cy5 dye was
319 added and the samples incubated in the dark for 30 min. The Cy3 and Cy5 dyes (Amersham-
320 Pharmacia) were dissolved in 45μL DMSO prior to being added to the coupling buffer. To remove
321 unincorporated dye, the labelled cDNA (total volume 10μL) was passed through a SpinEX column
322 (Qiagen). Dye incorporation was assessed by separating 1μL of the sample on a mini agarose gel and
323 visualising fluorescent products on a microarray scanner (Perkin Elmer ScanArray 5000XL). No pre-
324 hybridisation step was required. For hybridisation the remainder of each labelled cDNA (7-9 μL; 16-
325 30 pmol each dye) was added to 85 μL hybridisation buffer (UltraHyb, Ambion), 10 μL poly$(A)_{80}$
326 (10 mg mL^{-1}; Sigma) and 5 μL ultrapure BSA (10 mg mL^{-1}; Ambion). The hybridisation mixture was
327 heated to 95°C for 3 min, then cooled to 60°C before being applied to the microarray. Hybridisations
328 (48 slides per day) were performed on a Gene TAC Hyb Station (Genomic Solutions) for 16 h at

45°C. Slides were then automatically washed with 2×SSC, 0.5% SDS for 10 min at 60°C; 0.2×SSC, 0.5% SDS for 10 min at 42°C; and finally 0.2 X SSC 10 min at 42°C. Slides were then manually rinsed in isopropanol and dried by centrifugation before being scanned.

Data acquisition and analysis

Hybridised slides were scanned at 10 μm resolution using a Perkin Elmer ScanArray Express HT scanner. BlueFuse software (BlueGnome) was then used to visualise and identify features. Following a manual spot removal procedure and fusion of duplicate spot data (BlueFuse proprietary algorithm) the resulting intensity values and quality annotations were exported into the GeneSpring GX version 7.3.1 (Agilent Technologies) analysis platform. Data transformation, normalisation and quality filtering were as follows: 1) all intensity values less than 0.01 were set to 0.01; 2) a 'per spot per chip' intensity dependent (Lowess) normalisation was undertaken using software defaults (20% smoothing / cutoff 10); 3) data were filtered using a BlueFuse spot confidence value > 0.1 in ≥ 24 slides and BlueFuse spot quality of ≥ 0.5 in ≥ 24 slides. This gave a final list of 11,800 genes which were eligible for statistical analysis. Experimental annotations complied fully with MIAME guidelines (Brazma et al., 2001).

qRT-PCR validation

qRT-PCR was performed as described by Villeneuve *et al.* (2005). Primer details are given in Table III. Relative expression ratios were statistically compared between diet samples following normalisation against three housekeeping genes, using REST software (Pfaffl, 2001; Pfaffl *et al.* 2002). Five thousand random allocations were performed and differences were considered to be significant at P<0.05.

355 **RESULTS**

356

357 MICROARRAY FEATURE OVERVIEW

358

359 cDNAs derived from fifteen different tissue sources are represented among the 16,950 Atlantic
360 salmon gene features printed on the TRAITS / SGP cDNA microarray (Table I). Of these,
361 approximately 9% are from SSH libraries. Brain cDNAs predominate (15% of total). This reflects
362 the large number of ESTs generated for this tissue, due to 1) brain libraries being included in all three
363 of the EST resources available, 2) increased sequencing effort being focused on these libraries due to
364 perceived diversity of gene expression within brain tissue, 3) availability of a normalised brain EST
365 library. BLASTx homology searches of contig and singleton sequences (Table IV) revealed close to
366 40% of features having a weak hit (e-value $\geq e^{-10}$) or no hit at all to the NCBI non-redundant protein
367 database. Among the 10,399 features with a significant BLASTx hit (e-value $< e^{-10}$), 6762 (65%)
368 nominally different genes were identified. GO annotations were obtained for 7749 features and can
369 be accessed from the TRAITS website.

370

371 DIETARY LIPID EXPERIMENT

372

373 The main findings of this microarray-based investigation will be reported in detail elsewhere. Here,
374 data from preliminary analyses are presented to illustrate the degree of robustness and sensitivity that
375 the microarray / experimental design achieved.

376

377 Hybridisation data were analysed by two-way ANOVA, which examined the explanatory power of
378 the variables 'time-point' and 'diet' and the interaction between the two and incorporated a
379 Benjamini-Hochberg (1995) multiple test correction ($P \leq 0.05$). This analysis identified 4,142
380 features showing significant differential expression over the year long experimental time course,

381 demonstrating considerable temporal changes which may be related to a large number of biological
382 factors. In contrast only 15 significant diet-responsive features were detected (and a further 10
383 features with significant time-point × diet interactions) Of the 15 significant diet-associated genes, 10
384 with BLASTx or other known sequence homologies had functions associated with either HUFA or
385 cholesterol biosynthetic pathways (Table V). The differential expression of these genes in these
386 samples was confirmed by qRT-PCR. The identity of the remaining five of the 15 probes remains to
387 be established.

388

389 To explore the consequences of analysing smaller numbers of microarrays, the significance of
390 expression of the nine Δ5 / Δ6 fatty acyl desaturase (FAD) probes present on the microarray was
391 used as a proxy indicator for the performance of a given experimental design. Two strategies were
392 employed to reduce microarray number, 1) omission of dye swap, 2) reduction of biological
393 replicates. Two-way ANOVAs were undertaken (without multiple test correction) and gene lists
394 ordered by ascending (diet) P-value. The position of the nine probes in each of the lists is shown in
395 Table VI. In a full analysis incorporating all six biological replicates and a dye swap (12 replicate
396 microarrays per condition) the nine desaturase probes were all present in the top 25 of this list.
397 Reducing microarray numbers, whilst retaining a dye swap, decreased apparent sensitivity, but only
398 markedly when the number of biological replicates was reduced to three. At this replication level the
399 results of the analyses appeared to be sensitive to the particular biological replicates selected, as
400 demonstrated by the duplicate analyses (biological replicates 1-3 vs 4-6). Omission of a dye swap
401 gave mixed results depending upon the dye selected for the pooled reference sample. With a Cy5
402 pooled reference the results (all nine desaturase probes in top 27 of list) closely matched those of the
403 full 12 microarray design. However, with a Cy3 pooled reference only three probes were in the top
404 25 of the list. One probe was not in the list at all and others had dropped as low as position 284.

405

406 Expression profiles derived from the nine FAD probes over the four time-points and for both diets
407 are depicted in Fig. 3. Two distinct patterns are apparent, which, from 3' UTR probe results, clearly
408 correspond to expression of the two different FAD genes (Δ5 and Δ6). However, the expression
409 profiles derived from two of the Δ5 FAD probes (the ORF PCR fragment and full length EST) mimic
410 those of the Δ6 FAD probes.

411
412

413

414 **DISCUSSION**

415

416 TRAITS / SGP MICROARRAY FABRICATION

417

418 The TRAITS / SGP cDNA microarray was conceived as a preliminary tool, contributing towards the
419 goal of developing a more focused DNA chip for routine health and performance monitoring in
420 Atlantic salmon. A number of existing EST collections were used as the basis for its design and
421 construction, and whilst this approach had obvious advantages in reducing the time-frame and costs
422 associated with resource development, it also made microarray fabrication all the more challenging.
423 Not only did probe preparation involve a range of different host / vector combinations, but also, the
424 need to track and annotate clones from different library resources compounded the difficulties. Most
425 of the ESTs available comprised 5' end reads. Although this improved the likelihood of successful
426 probe annotation, use of mainly 5' sequence data will inevitably have compromised the cluster
427 analysis as not all the clones will have been full length. The extent of gene redundancy on the
428 microarray can only be confidently established from analysis of 3' end sequence data. Presently,
429 there are no plans to re-sequence the microarray resource.

430

431 Enriching the microarray for potential trait-specific genes has proven to be a worthwhile strategy. In
432 a number of studies (Martin *et al.*, 2007; and in preparation), SSH derived probes have been

433 identified as significant responders in immune, protein catabolism, and smoltification studies.
434 However, in lipid metabolism studies to date, no probes from SSH clones (or contigs containing SSH
435 clones) have been identified as responding differentially between diets. One possible explanation for
436 the apparent failure of the SSH procedure in this case follows from the subsequent microarray
437 analysis of the dietary lipid experiment reported here. Overall fold changes in diet-responsive genes
438 were found to be quite low, with the greatest changes being only 3-fold, and there were no data to
439 support differential expression of these genes at the two time-points (50 & 53 weeks) used for SSH
440 library construction.

441

442 There are no plans to radically improve upon the probe set in the TRAITS / SGP microarray. In its
443 existing format there is still spare capacity for feature printing, and additional candidate gene probes
444 have, and will, be added on an *ad hoc* basis. For example, probes for 12 genes associated with lipid
445 metabolism and 10 immune-related genes, not known to be on the microarray, have recently been
446 included to facilitate two specific experimental studies. Over the past few years printing technology
447 has advanced significantly. Currently the TRAITS / SGP microarray is being printed using non-
448 contact inkjet based technology (ArrayJet Ltd), which delivers more consistent spot and slide
449 uniformity than contact pin printing, and which should improve the reliability of generated
450 microarray data and increase the sensitivity of detection.

451

452 DIETARY LIPID EXPERIMENT

453

454 By identifying candidate genes, the results of this experiment provide reassurance of the clone-
455 tracking accuracy of the microarray. They also confirm that the selected microarray design and
456 fabrication technology, together with the experimental methodology employed, provide the capacity
457 for sensitive detection of differential expression. The fold change differences in expression of Δ5

458 FAD and Δ6 FAD between fish fed VO *vs* fish fed FO based diets closely match those obtained by
459 qRT PCR analysis in a previous study (Zheng *et al.*, 2005*b*).
460
461 Microarray analyses are expensive and time consuming to perform and there is often financial
462 pressure to minimise the number of microarrays used in an experiment. Here, reducing the slide
463 number from 12 to 8 (while retaining the dye swap) had little apparent effect on the ability to detect
464 significant differential expression in FAD genes known to show a variable response according to
465 diet. Using just six slides per condition (one per biological replicate and no dye swap) gave different
466 outcomes according to the dye / target combination used. When the experimental sample was
467 labelled with Cy3 and the pooled reference with Cy5, the results were comparable with the full 12
468 slide (including dye swap) analysis. However, there was much less agreement when the experimental
469 sample was Cy5 labelled and the pooled reference was Cy3 labelled. The reason for this marked dye-
470 dependent disparity remains to be established. However, it has been noted in this and other related
471 studies conducted by TRAITS partners that 1) pooled reference samples produce higher background
472 intensities and 2) the Cy3 channel consistently displays higher background values compared to the
473 Cy5 channel. The combination of these factors may, at least in part, account for this phenomenon.
474 The relevance of these observations to other studies is difficult to assess. The results are largely
475 dependent on the homogeneity of the system under study and the absolute and relative expression
476 levels of any differentially responding genes. For similar reasons, it is also difficult to critically
477 assess published studies more generally. It is likely that the use of sex-balanced pools for
478 experimental biological replication in this study, rather than individual samples, contributed to the
479 discriminatory capabilities of this microarray experiment. Where cost considerations are an
480 important factor, interrogation of reduced numbers of microarrays may be the only viable option
481 available. If background fluorescence can confidently be controlled, omitting a dye swap, as opposed
482 to reducing biological replication, may be a preferable route to follow in such cases.
483

484 The FAD expression profiles demonstrated the potential for obtaining artifactual results due to cross-
485 hybridisation of similar transcripts. Expression profiles derived from two Δ5 FAD probes (the ORF
486 PCR fragment and full length EST) closely matched those obtained from all Δ6 FAD probes. While
487 the 3' UTRs are very distinct (Δ5 FAD 3' UTR = 1072 bp; Δ6 FAD 3' UTR = 457 bp; sequence
488 similarity *c.* 30%), the two ORFs are very similar to their Δ6 counterparts (both 1365 bp; *c.* 95%
489 sequence similarity). It has also been reported from qRT-PCR analysis, that Δ6 FAD gene expression
490 is approximately four-fold higher than Δ5 FAD expression, in liver tissue of farmed salmon fed on
491 fish oil (Zheng *et al*,. 2005*b*). Thus it would appear that cross-hybridisation of the more abundant Δ6
492 FAD transcript targets is masking the true expression profiles derived from these two Δ5 FAD
493 probes. The failure of the microarray analysis to correctly distinguish between Δ5 FAD and Δ6 FAD
494 expression profiles in all cases clearly illustrates one of the inherent weaknesses of cDNA
495 microarray-based studies *i.e.* the inability to distinguish between highly similar message transcripts.
496 This is of particular relevance in interpreting transcriptomic data from salmonid species, since fish of
497 the family Salmonidae have undergone a relatively recent whole-genome duplication *c.* 25-100
498 million years ago (Allendorf & Thorgaard, 1984), such that simultaneous expression of duplicate
499 genes is a commonly observed phenomenon.
500
501 OTHER STUDIES.
502
503 As part of the original funded project, the TRAITS partners have already used the cDNA array to
504 explore transcriptomic responses in experiments targeting the four key traits identified as being
505 important for sustained salmon aquaculture *i.e.* a) supply of dioxin-free highly unsaturated oils for
506 the salmon diet, b) protein growth efficiency, c) infectious disease, and d) a long and complex
507 lifecycle. In addition to the diet work outlined in this paper, other research at Stirling University,
508 funded by the European Commission, is comparing gene expression both within and between
509 families of fish fed on fish oil or vegetable oil-based diets. Other studies, investigating the

510 transcriptomic response of Atlantic salmon to infection by infectious pancreatic necrosis virus
511 (IPNV) in both seawater and freshwater environments are similarly in progress. Researchers at
512 Aberdeen University are using the cDNA microarray to study the immune response in Atlantic
513 salmon *in vivo* following infection by *A.salmonicida* and *in vitro* to examine the response to
514 recombinant fish cytokines. Additionally, short-term starvation trials have been used to explore
515 protein catabolism pathways. Finally the Cardiff partner has identified genes and gene pathways
516 from three tissues (brain, gill and kidney) that are involved in the parr-smolt transformation. The
517 TRAITS / SGP microarray has also been supplied to an Australian research group to examine gene
518 expression responses to amoebic gill disease. To date its performance has not been critically assessed
519 for use with other closely related species. Cross-species hybridisation between salmonids has been
520 shown to be extremely high (Rise *et al*., 2004; von Schalburg *et al*., 2005), suggesting that
521 microarrays could be used with similar confidence for both Atlantic salmon and rainbow trout due to
522 the high level of sequence homology between the two species.
523
524 FUTURE DIRECTIONS
525
526 The final stage of the TRAITS project, *i.e.* development and validation of a focused oligonucleotide
527 array is currently in progress. Approximately 1000 differentially responding genes have been
528 identified from cDNA microarray interrogations by the TRAITS partners and unique 70mer
529 oligonucleotides designed and synthesised for these and appropriate control genes. These form the
530 basis of the TRAITS / SGP secondary oligochip, the performance of which is currently being
531 evaluated. Irrespective of the outcome, the TRAITS / SGP cDNA microarray will continue to be a
532 valuable tool and be available for use to the wider scientific community. Enquiries regarding
533 purchase / use of this microarray should be directed to ARK Genomics (www.ark-genomics.org) in
534 the first instance.
535

536 Acknowledgments

537

538 The TRAITS / SGP microarray was constructed by ARK Genomics and funded by Biotechnology
539 and Biological Sciences Research Council (BBSRC) grants 98/EGA17674 (University of Stirling,
540 UK), 98/EGA17675 (University of Aberdeen, UK), 72/EGA17676 (Cardiff University, UK), and
541 with grant 139617/140 "Salmon Genome Project" by the Research Council of Norway. The dietary
542 trial was funded by the EU FP5 Project Q5RS-2000-30058 (RAFOA). We thank Prof. A. Brass and
543 Dr. Y. Fang (Manchester University) for key discussions relating to experimental design and
544 analysis.

545

546 **References**

547

548 Adzhubei, A. A., Vlasova, A. V., Hagen-Larsen, H., Ruden, T. A., Laerdahl, J. K., Høyheim, B.
549 (2007). Annotated expressed sequence tags (ESTs) from pre-smolt Atlantic salmon (*Salmo salar*) in
550 a searchable data resource. *BMC Genomics* **8,** 209. doi:10.1186/1471-2164-8-209

551

552 Allendorf, F. W. & Thorgaard, G.W. (1984). Tetraploidy and the evolution of salmonid fishes. In
553 *Evolutionary Genetics of Fishes* (Turner, B.J., ed.), pp. 1–46. New York: Plenum Press.

554

555 Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment
556 search tool. *Journal of Molecular Biology* **215**, 403-410.

557

558 Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and
559 powerfulapproach to multiple testing. *Journal of the Royal Statistical Society* **B57**, 289-300.

560

561 Bonaldo, M. D. F., Lennon, G. & Soares, M. B. (1996). Normalisation and subtraction: two
562 approaches to facilitate gene discovery. *Genome Research* **6,** 791-806.
563

564 Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J.,
565 Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P,. Holstege, F. C. P., Kim,
566 I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S.,
567 Stewart, J., Taylor, R., Vilo, J. & Vingron, M. (2001). Minimum information about a microarray
568 experiment (MIAME) - toward standards for microarray data. *Nature Genetics* **29**, 365-371.
569

570 Cairns, M. T., Johnson, M. C., Talbot, A. T., Pemmasani, J. K., McNeill, R. E., Houeix, B., Sangrador-
571 Vegas, A. & Pottinger, T. G. (2007). A cDNA microarray assessment of gene expression in the liver
572 of rainbow trout (*Oncorhynchus mykiss*) in response to a handling and confinement stressor
573 *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*
574 doi:10.1016/j.cbd.2007.04.009
575

576 Conesa, A., Gotz, S., Garcia-Gomez, J. M., Terol, J., Talon, M. & Robles, M. (2005). Blast2GO: a
577 universal tool for annotation, visualization and analysis in functional genomics research.
578 *Bioinformatics* **21**, 3674-3676.
579

580 Ewart, K. V., Belanger, J. C., Williams, J., Karakach, T., Penny, S., Tsoi, S. C. M., Richards, R. C. &
581 Douglas, S. E. (2005). Identification of genes differentially expressed in Atlantic salmon (*Salmo
582 salar*) in response to infection by *Aeromonas salmonicida* using cDNA microarray technology.
583 *Developmental and Comparative Immunology* **29**, 333-347.
584

585 Hagen-Larsen, H., Laerdahl, J. K., Panitz, F., Adzhubei, A. & Høyheim, B. (2005). An EST-based
586 approach for identifying genes expressed in the intestine and gills of pre-smolt Atlantic salmon
587 (*Salmo salar)*. *BMC Genomics* **6**, 171. doi:10.1186/1471-2164-6-171
588
589 Hastings, N., Agaba, M. K., Tocher, D. R., Zheng, X., Dickson, C. A., Dick, J. R. & Teale, A.J.
590 (2004). Molecular cloning and functional characterization of fatty acyl desaturase and elongase
591 cDNAs involved in the production of eicosapentaenoic and docosahexanoic acids from α-
592 linolenic acid in Atlantic salmon (*Salmo salar*). Marine Biotechnology **6,** 463-474.
593
594 Jordal, A.-E. O., Torstensen, B. E., Tsoi, S., Tocher, D. R., Lall, S. P. & Douglas, S.E. (2005).
595 Dietary rapeseed oil affects the expression of genes involved in hepatic lipid metabolism in
596 Atlantic salmon (*Salmo salar* L.). *Journal of Nutrition* **135,** 2355-2361.
597
598 Koskinen, H., Pehkonen, P., Vehniainen, E., Krasnov, A., Rexroad, C., Afanasyev, S., Molsa, H. &
599 Oikari, A. (2004). Response of rainbow trout transcriptome to model chemical contaminants.
600 *Biochemical and Biophysical Research Communications* **320**, 745-753.
601
602 MacKenzie, S., Iliev, D., Liarte, C., Koskinen, H., Planas, J. V., Goetz, F. W., Molsa, H., Krasnov, A.
603 & Tort, L. (2006). Transcriptional analysis of LPS-stimulated activation of trout (*Oncorhynchus*
604 *mykiss*) monocyte / macrophage cells in primary culture treated with cortisol. *Molecular*
605 *Immunology* **43**, 1340-1348.
606
607 Marsden, M. J., DeVoy, A., Vaughan, L., Foster, T. J. & Secombes C. J. (1996). Use of a genetically
608 attenuated strain of *Aeromonas salmonicida* to vaccinate salmonid fish. *Aquaculture International*
609 **4**, 55–66.
610

611 Martin, S.A., Caplice, N.C., Davey, G.C. & Powell, R. (2002). EST-based identification of genes
612 expressed in the liver of adult Atlantic salmon (*Salmo salar*). *Biochemical and Biophysical*
613 *Research Communications* **293**, 578-585.

614

615 Martin, S. A. M., Blaney, S. C., Houlihan, D. F. & Secombes, C. J. (2006). Transcriptome response
616 following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*). *Molecular*
617 *Immunology* **43**, 1900-1911.

618

619 Martin, S. A. M., Taggart, J. B., Seear, P. J., Bron, J. E., Talbot, R., Teale, A. J., Sweeney, G. E.,
620 Hoyheim, B., Houlihan, D. F., Tocher, D. R., Zou, J, & Secombes CJ. (2007). Interferon type I and
621 type II responses in an Atlantic salmon (*Salmo salar*) SHK-1 cell line using the salmon
622 TRAITS/SGP microarray. *Physioogical Genomics* doi:10.1152/physiolgenomics.00064.2007.

623

624 National Research Council (NRC). (1993). *Nutrient Requirements of Fish.* Washington DC: Nation
625 Academy Press.

626

627 Pertea, G., Huang, X. Q., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J.,
628 Cheung, F., Parvizi, B., Tsai, J. & Quackenbush, J. (2003). TIGR Gene Indices clustering tools
629 (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics* **19**: 651-652.

630

631 Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR.
632 *Nucleic Acids Research* **29**, 2002-2007.

633

634 Pfaffl, M. W., Horgan, G. W. & Dempfle, L. (2002). Relative expression software tool (REST) for
635 group-wise comparison and statistical analysis of relative expression results in real-time PCR.
636 *Nucleic Acids Research* **30**, 9-36.

637

638 Purcell, M. K., Nichols, K. M., Winton, J. R., Kurath, G., Thorgaard, G. H., Wheeler, P., Hansen, J. D.,
639 Herwig, R. P. & Park, L. K. (2006). Comprehensive gene expression profiling following DNA
640 vaccination of rainbow trout against infectious hematopoietic necrosis virus. *Molecular Immunology*
641 **43**, 2089-2106.

642

643 Rise, M. L., von Schalburg, K. R., Brown, G. D., Mawer, M. A., Devlin, R. H., Kuipers, N., Busby,
644 M., Beetz-Sargent, M., Alberto, R., Gibbs, A. R. Hunt, P., Shukin, R., Zeznik, J. A., Nelson, C.,
645 Jones, S. R., Smailus, D. E., Jones, S. J., Schein, J. E., Marra, M. A., Butterfield, Y. S., Stott, J. M.,
646 Ng, S. H., Davidson, W. S. & Koop, B. F. (2004). Development and application of a salmonid EST
647 database and cDNA microarray: Data mining and interspecific hybridization characteristics.
648 *Genome Research* **14**, 478-490.

649

650 Roberge, C., Paez, D. J., Rossignol, O., Guderley, H., Dodson, J. & Bernatchez, L. (2007). Genome-
651 wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular*
652 *Immunology* **44**, 1374-1383.

653

654 Salem, M., Kenney, P. B., Rexroad, C. E. & Yao, J. B. (2006). Microarray gene expression analysis in
655 atrophying rainbow trout muscle: a unique nonmammalian muscle degradation model.
656 *Physiological Genomics* **28**, 33-45.

657

658 Sargent, J. R. & Tacon, A. G. R. (1999). Development of farmed fish: a nutritionally necessary
659 alternative to meat. *Proceedings of the .Nutrition Society* **58,** 377-383

660

661 Torstensen, B. E., Bell, J. G., Rosenlund, G., Henderson, R. J., Graff, I. E., Tocher, D. R., Lie, Ø. &
662 Sargent, J. R. (2005). Tailoring of Atlantic salmon (*Salmo salar* L.) flesh lipid composition and

663 sensory quality by replacing fish oil with a vegetable oil blend. *Journal of Agricultural and Food*
664 *Chemistry*. **53,** 10166 - 10178.
665

666 von Schalburg, K. R., Rise, M. L., Cooper, G. A., Brown, G. D., Gibbs, A. R., Nelson, C. C.,
667 Davidson, W. S. and Koop, B. F. (2005). Fish and chips: Various methodologies demonstrate utility
668 of a 16,006-gene salmonid microarray. *BMC Genomics* 6. 126. doi:10.1186/1471-2164-6-126
669

670 Villeneuve, L., Gisbert, E., Zambonino-Infante, J. L., Quazuguel, P. & Cahu, C. L. (2005). Effect of
671 nature of dietary lipids on European sea bass morphogenesis: implication of retinoid receptors.
672 *British Journal of Nutrition* **94,** 877-884.
673

674 Zheng, X., Tocher, D. R., Dickson, C. A., Bell, J. G. and Teale, A. J. (2004). Effects of diets containing
675 vegetable oil on expression of genes involved in polyunsaturated fatty acid biosynthesis in liver
676 of Atlantic salmon (*Salmo salar*). *Aquaculture* **236,** 467-483.
677

678 Zheng, X., Tocher, D.R., Dickson, C.A., Dick, J.R., Bell, J.G. and Teale, A.J. (2005*a*) Highly
679 unsaturated fatty acid synthesis in vertebrates: new insights with the cloning and characterisation
680 of a Δ6 desaturase of Atlantic salmon. *Lipids* **40,** 13-24.
681

682 Zheng, X., Torstensen, B.E., Tocher, D.R., Dick, J.R., Henderson, R.J. and Bell, J.G. (2005*b*).
683 Environmental and dietary influences on highly unsaturated fatty acid biosynthesis and
684 expression of fatty acyl desaturase and elongase genes in liver of Atlantic salmon (*Salmo salar*).
685 *Biochimica et Biophysica Acta* **1734,** 13-24.
686

686 **FIGURE LEGENDS**

687

688 Fig. 1. Overview of TRAITS strategy to generate both a general cDNA and a focused

689 oligonucleotide- microarray.

690

691

692 Fig. 2. Summary of clustering procedure and probe selection for the cDNA microarray.

693

694

695 Fig 3. Fatty acyl desaturase expression profiles (two diets over four time-points) derived from

696 microarray analysis of nine different Δ5 and Δ6 FAD probes. Blue lines denote Δ6 FAD

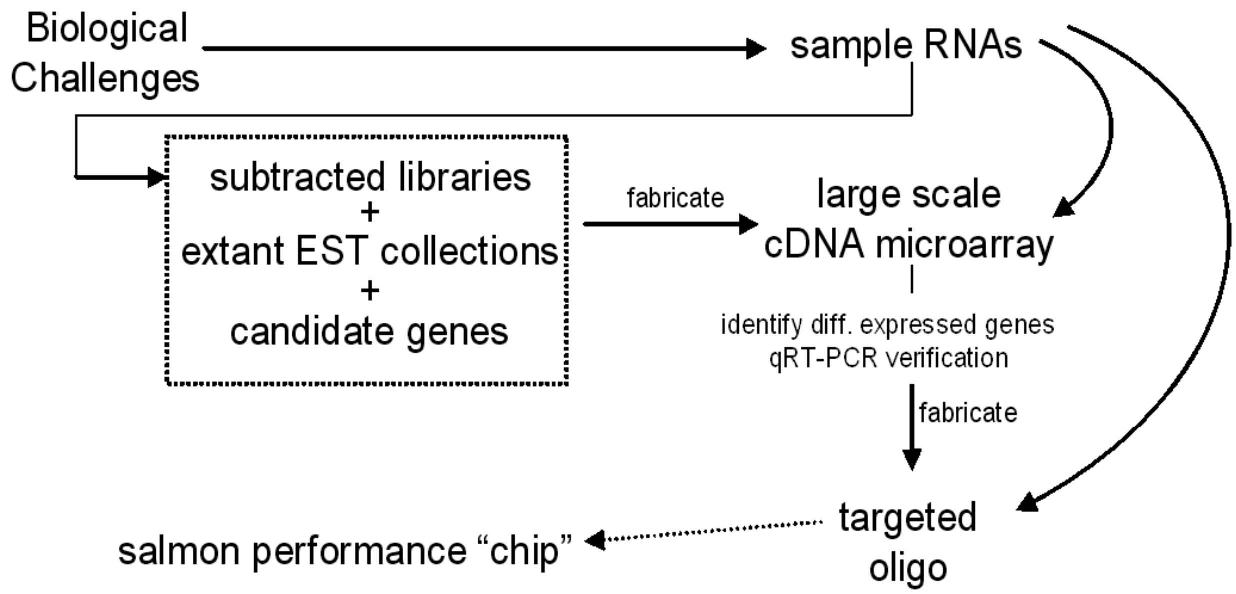
697 probes; red lines denote Δ5 FAD probes. Dashed red lines represent two Δ5 FAD probes

698 (ORF PCR fragment and clone bra_snb_04D02 – full length EST sequence) and show

699 expression profiles that mimic those of the Δ6 FAD probes.

700

700 Fig.1



701 Fig.2.

Clustering (July 04)

Input:	**Source**	**No. Clones**	**Remarks**
	SALGENE	7,492	non-normalised; 7 tissue types; all 5' reads
	SALGENE	4,236	normalised; 2 tissue types, mix of 5' & 3' reads
	SGP	30,564	non-normalised 15 tissue types; all 5' reads
	TRAITS	3,515	SSH derived gene fragments 6 tissue types
	GENBANK	57,536	mix of 5' & 3' reads (*in silico* only)
	TOTAL	**103,343**	

Output:	**Type**	**All**		**Available Clones**
	CONTIGS	13,555	→	9,042
	SINGLETONS	20,826	→	9,145
	TOTAL	**34,381**	→	**18,187**

Removed:	Duplicate clone data (5' & 3' reads) Clones that failed to grow Sub-optimal insert amplifications	- 1,412
Added:	TRAITS selected candidate clones	+ 31
	Later available ESTs	+ 144

GRAND TOTAL: 16,950

736

737

737 Fig.3.

739

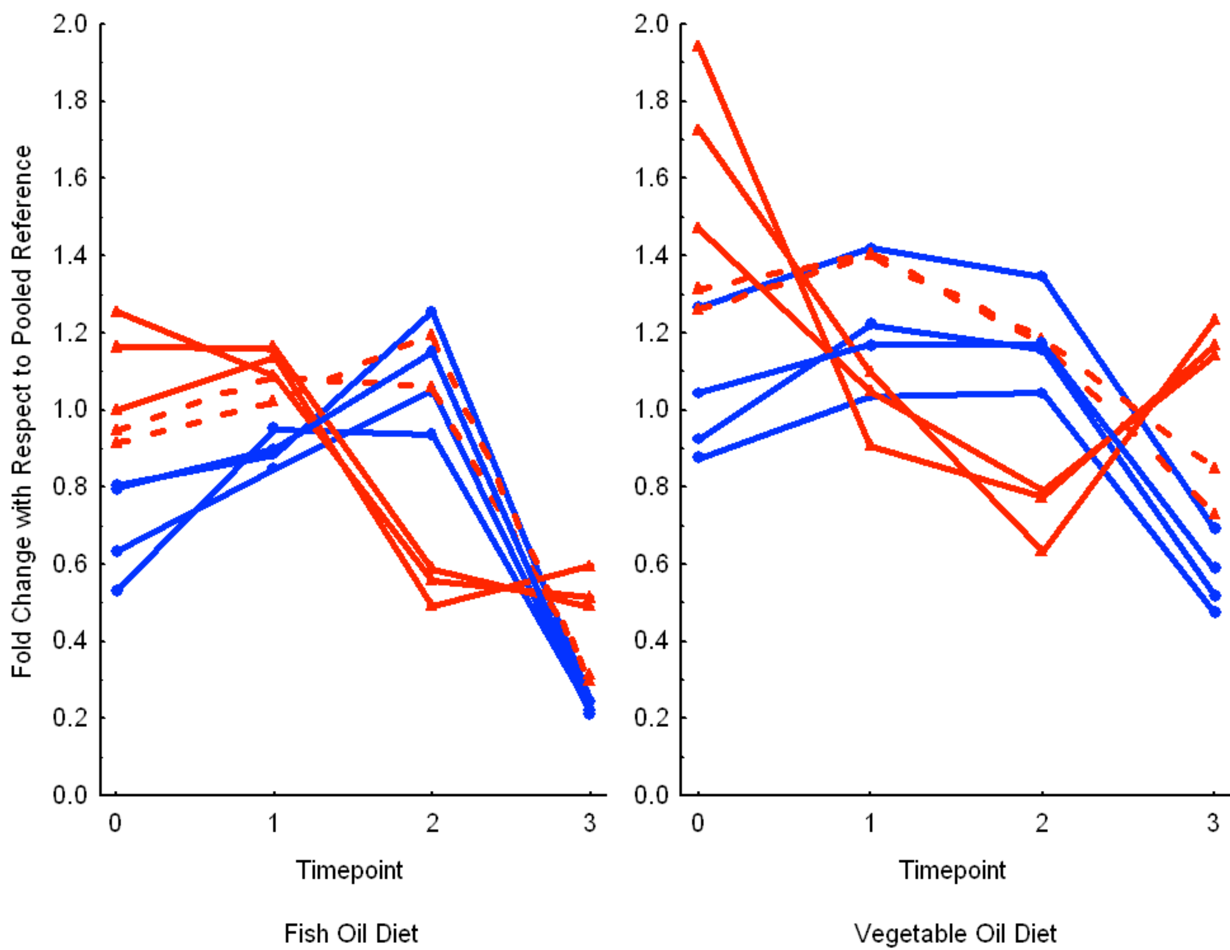


Table I. Details of the EST libraries used to construct the TRAITS / SGP cDNA microarray.

Source	**Tissue**	**Environment**	**Host Cells**	**Vector**
SALGENE	Liver, Testis, Ovary	Seawater	XLOLR	pBK-CMV
SALGENE	Spleen, Kidney	Seawater	SOLR	pBlueScript II SK(-)
SALGENE	Gill, Intestine	Freshwater	XL10 - Gold	pBlueScript II SK(+) XR
SALGENE	White muscle, Brain	Seawater	SURE	pT7T3-Pac
TRAITS	Liver, Kidney, Gill, White muscle	Freshwater	ElectroTen Blue	pGEM T-easy
TRAITS	Liver, Kidney, Brain, Pituitary, Gill	Seawater	ElectroTen Blue	pGEM T-easy
SGP	Brain, Eye, Gill, Head kidney, Heart, Intestine, Kidney, Liver, White muscle, Ovary, Skin, Spleen, Swim-Bladder, Testis	Freshwater	XL10 - Gold	pBlueScript II SK(+) XR

743 Table II. Pre-identified candidates and other reference genes

Gene Name	cDNA length	Notes
Apolipoprotein B	1402 bp	partial with 3’UTR
Carnitine Palmityltransferase1	823 bp	partial
Carotene dioxygenase	872 bp	partial
Estrogen receptor alpha	2900 bp	partial with 3’UTR
Growth hormone receptor	340 bp	Partial, PCR fragment
Glyceraldehyde phosphate dehydrogenase	1086 bp	full length
Heat shock protein P70	830 bp	partial with 3’UTR
Homogenistate dioxygenase	952 bp	partial with 3’UTR
Insulin-like growth factor-1	230 bp	partial
Interferon gamma	1132 bp	full length
Iinterleukin-1 beta	790 bp	ORF
NGF1-B	224 bp	partial (RACE fragment)
Pituitary specific transcription factor 1	250 bp	partial
Peroxisome proliferator-activated receptor alpha	1644 bp	full length
Peroxisome proliferator-activated receptor beta 1	1462 bp	full length
Peroxisome proliferator-activated receptor beta 2	779 bp	partial
Peroxisome proliferator-activated receptor gamma	1665 bp	full length
PUFA elongase	950 bp	ORF
Retinoic acid receptor alpha	840 bp	ORF (RACE fragment)
Retinoic acid receptor gamma	440 bp	Partial inc 5’ UTR
Retinaldehyde dehydrogenase type2	922 bp	partial
TNFa induced adipose related protein	483bp	partial
Thyroid hormone receptor alpha	*c.* 2000 bp	partial inc 5’ UTR
Thyroid hormone receptor beta	*c.* 900 bp	partial inc 5’ UTR
Vitamin D3 receptor	360 bp	partial inc 5’ UTR
Δ5 fatty acyl desaturase	408 bp	3' UTR fragment
Δ5 fatty acyl desaturase	881 bp	3' UTR fragment
Δ5 fatty acyl desaturase	1365 bp	ORF
Δ6 fatty acyl desaturase	384 bp	3' UTR fragment
Δ6 fatty acyl desaturase	401 bp	3' UTR fragment
Δ6 fatty acyl desaturase	1365 bp	ORF

744