ACADEMIC PRESS

Biochemical and Biophysical Research Communications 293 (2002) 578–585

BBRC

www.academicpress.com

EST-based identification of genes expressed in the liver of adult Atlantic salmon (*Salmo salar*)

Sarah A. Martin, Nicole C. Caplice, Grace C. Davey, and Richard Powell*

Department of Microbiology, National University of Ireland, Galway, Ireland

Received 25 March 2002

Abstract

A list of genes expressed in the liver of Atlantic salmon was compiled using the expressed sequence tag (EST) strategy. 733 ESTs, derived from 170 abundant and 563 rare mRNA encoding liver cDNA clones, were determined. Bioinformatic analysis revealed that 390 (53%) of the salmon liver ESTs could be ascribed to the transcriptional products of 93 identified genes including 7 previously described in the Atlantic salmon. The identified Atlantic salmon genes were classified with respect to cellular role which showed that 33 (36%) of the identified genes encoded proteins associated with primary liver functions such as transport, acute phase response, and blood clotting. Furthermore, comparative analysis revealed that 12 of the 16 salmon genes that were shown to encode abundant mRNA transcripts in liver had homologues that have also been shown to be highly expressed in mammalian liver systems. Finally, two cDNA variants corresponding to the two cDNA forms of the apolipoprotein A-I gene previously identified in rainbow trout were also found in Atlantic salmon.

Keywords: Atlantic salmon (*Salmo salar*); Expressed sequence tag (EST); Liver; Gene expression; Apolipoprotein A-I gene

The liver is a dynamic organ playing an important role in carbohydrate, lipid, steroid, amino acid, and prostaglandin metabolism; in detoxification and in seroprotein and biliary acid production. In mammalian species, the cell types unique to the liver are hepatocytes, biliary cells, and sinusoidal cells with the hepatocytes constituting approximately 78% of liver volume and approximately 76% of liver cells [1]. With an estimated 130 million cells per gram of liver, the hepatocyte fulfills the majority of the organ functions including the bulk production of plasma proteins (e.g. proteins involved in binding and transport, the blood clotting cascade, and the acute phase response), detoxification, and ATP production [2]. Due to its vital function and relative lack of complexity, the liver is a model for mammalian gene expression studies, with gene regulation being primarily exerted at the transcriptional level [3].

By comparison, far less is known of gene expression in fish liver tissue. With respect to Atlantic salmon, previous studies have focused on the description of individual genes including strongly expressed genes such as serum albumin [4], apolipoprotein A-I [5], transferrin [6], and α-1-microglobulin/bikunin [7]. Currently, and in addition to the mitochondrion genome [8], the international databases contain sequence information describing 75 full length Atlantic salmon genes, 14 of which have been shown to be expressed in salmon liver.

Of specific interest to many salmonids, hepatic ultrastructure change has been associated with the parr-smolt transformation in anadromous Atlantic salmon [9]. Other features of smoltification including increased hepatic enzyme activities [10] and alterations in fatty acid content [11] also tacitly imply a role for the liver in this adaptation process. At the transcriptional level, one previous report studied liver-specific expression of five known salmon genes during smoltification revealing no change in mRNA levels for β-fibrinogen and apolipoprotein A-I, slight increases in mRNA levels for complement C3 and hemopexin, and a major 200-fold increase in serum albumin mRNA levels [12]. However, the further study of specific features such as salmonid smoltification, or more generally, the comparison of fish

* Corresponding author. Fax: +353-91-525700.
E-mail address: richard.powell@nuigalway.ie (R. Powell).

0006-291X/02/$ - see front matter
PII: S0006-291X(02)00263-2

and mammalian liver gene expression is currently limited by the lack of identified salmon genes expressed in liver cells.

This study reports an expressed sequence tag (EST)-based gene identification analysis of 733 Atlantic salmon cDNA clones derived from an adult mixed-sex Atlantic salmon liver cDNA library. One-hundred and seventy of these cDNA clones represented abundant liver mRNA transcripts while the remaining 563 cDNA clones encoded rare liver mRNA transcripts. Gene identification was based on homology searches of the NCBI nr protein database. The identified salmon genes were then characterised in terms of cellular role and by comparison to data derived from homologues described in mammalian liver systems. Finally, two cDNA variants were identified for the Atlantic salmon apolipoprotein A-I gene which seem to correspond directly to the two cDNA forms of this gene, i.e., apoA-I-1 and apoA-I-2, previously identified in rainbow trout [13].

Materials and methods

cDNA library construction and cDNA clone selection. The cDNA library was constructed from mRNA prepared from equivalent amounts of liver tissue dissected from one male and one female, approximately 3 kg, adult Atlantic salmon (MOWI strain) obtained from an Irish fish farm. The details of the cDNA library construction using the λ Zap Express cDNA synthesis/Gigapack cloning kit (Stratagene Cloning Systems, CA, USA) have been previously described [14]. The identification of abundant and rare mRNA encoding cDNA clones was performed following the methodology previously described [15]. Briefly, 3000–4000 phages from the liver cDNA library were screened with a total cDNA probe reverse transcribed from 5 μg of total liver RNA primed with a mixture of anchored oligo-$(dT)_{12}$ primers containing all 12 possible dinucleotide combinations at the 3′-terminus (Sigma Genosys, Cambridgeshire, UK). The reverse transcription was performed at 37 °C for 30 min before the addition of 2 μl of 2.5 mM dNTP and incubation at 37 °C for a further 30 min. The cDNA probe was purified using a High Pure PCR Product Purification kit (Roche Diagnostics GmbH, Mannheim, Germany) and in situ hybridisation was performed using conventional procedures [16]. Following autoradiography at −70 °C for 48 h, hybridisation-positive and -negative plaques, representing abundant and rare mRNA encoding cDNA clones, respectively, were catalogued and stored separately.

DNA sequencing and bioinformatic analysis. Single-pass sequencing of the 5′-termini of 960 selected salmon liver cDNA clones in phagemid form was performed using the ABI 3700 automatic DNA sequencer (PE Applied Biosystems, CA, USA) and the ABI prism Big Dye Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems). In order to identify identical sequences, all ESTs >270 bp in length after elimination of vector sequence were aligned together using the Clustal X programme [17] and the longest EST was taken as the unique representative member of each EST cluster. Subsequently, all the unique ESTs were submitted to the NCBI nr protein database [18] using the Blast X programme located on the NCBI Blast homepage (http://www.ncbi.nim.nih.gov/BLAST/). The first 50 most homologous sequences were listed and the best 10 alignments were examined. The identification of an Atlantic salmon EST was based on a minimum amino acid sequence identity of >50% over a contiguous series of >50 amino acids. All unique ESTs have been deposited in the GenBank dbEST under accession numbers BI468016-BI468193, BI544051-BI544053, and BI544216-BI544217. Appropriate phagemid infected cells have been placed in long-term storage at −70 °C.

Results and discussion

Identification of abundant and rare mRNA encoding cDNA clones

The liver cDNA library used in this study had previously been assessed as containing 3.6×10^5 primary clones with a parental background of 2%, an average cDNA fragment length of 1.7 kb and with 95% of clones containing cDNA fragments >0.5 kb in length [14]. Several 1000 salmon liver cDNA clones were then screened by in situ hybridisation with a total liver cDNA probe in order to identify cDNA clones corresponding to either abundant or rare mRNA encoding genes. As this type of cDNA probe has been reported to only identify cDNA clones derived from mRNA transcripts that exceed 0.06% of the total mRNA population [15], the 49% of the phage plaques in the liver cDNA library that hybridised with the cDNA probe was catalogued as encoding abundant liver mRNA transcripts. The remaining 51%, non-hybridising phage plaques, was catalogued as encoding rare liver mRNA transcripts.

Atlantic salmon liver EST determination and gene identification

Nine-hundred and sixty randomly chosen cDNA clones, composed of 192 and 768 cDNA clones of the abundant and rare mRNA classes, respectively, were submitted to single-pass DNA sequencing targeting the 5′-terminus of each cDNA fragment. After elimination of the vector sequence, only ESTs > 270 bp in length were chosen for subsequent bioinformatic analysis and this constituted 170 abundant and 563 rare mRNA encoding salmon liver ESTs. Cluster analysis was then performed on the 733 salmon liver ESTs using the criterium that ESTs showing a <1% nucleotide mismatch over a length of 300 nucleotides were encoded by the same gene. The result showed that the 733 liver ESTs was composed of 246 unique ESTs (derived from 128 clusters and 118 singletons), 68 and 178 of which encoded abundant and rare mRNA transcripts, respectively.

Due to the comparative lack of identified fish genes and as protein sequences have been shown to be more suitable to detect homology over long periods of evolutionary time [19], a protein-based homology strategy was used to screen for homologous genes in the international databases. The 246 unique salmon liver ESTs

Table 1
Classified list of identified Atlantic salmon liver ESTs

Atlantic salmon ESTs			Matching sequences				
Accession	Name	mRNA Class[a]	Species	Accession number	*E*-value	Identity (%)	Amino acid overlap
I. Plasma proteins							
BI468086	Beta-globin	R	*Salmo salar*	CAA65953	1.0e – 66	99	125/126
BF228472	Haptoglobin[b]	A	*Oncorhynchus mykiss*	AAF87767	4.3e – 46	88	97/109
BI468016	Haptoglobin fragment 1[b]	A	*Oncorhynchus mykiss*	AAG30004	2.2e – 33	86	76/88
BE518589	Hemopexin-like protein	A	*Oncorhynchus mykiss*	CAA92147	4.2e – 103	91	187/205
BI468020	Hemopexin-like protein	A	*Oncorhynchus mykiss*	CAA92147	4.0e – 39	68	80/116
BI468065	Hemopexin-like protein	A	*Oncorhynchus mykiss*	CAA92147	2.0e – 48	59	83/139
BI468078	Pentraxin	R	*Salmo salar*	CAA67765	1.0e – 116	98	207/210
BE518593	Serotransferrin I precursor[b]	A	*Salmo salar*	P80426	1.0e – 110	97	189/193
BE518594	Serotransferrin II precursor[b]	A	*Salmo salar*	P80429	1.0e – 104	95	186/195
BI468094	Serotransferrin II precursor[b]	A	*Salmo salar*	P80429	1.0e – 69	99	100/101
BI468084	Serotransferrin II precursor[b]	A	*Salmo salar*	P80429	3.0e – 83	96	151/156
BI544051	Serotransferrin II precursor[b]	A	*Salmo salar*	P80429	1.0e – 121	97	209/215
BE518596	Serum albumin 1 precursor[b]	A	*Salmo salar*	P21848	6.0e – 97	89	175/195
BI468021	Serum albumin 2 precursor[b]	A	*Salmo salar*	Q03156	1.0e – 48	97	95/97
II. Protease inhibitors							
BE518580	Alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor[b]	A	*Salmo salar*	JC2556	6.4e – 91	81	172/211
BI544216	Alpha-1-microglobulin/inter-alpha-trypsin inhibitor precursor[b]	A	*Salmo salar*	JC2556	2.8e – 97	100	178/178
BI468029	Antithrombin	R	*Salmo salar*	CAB64714	5.0e – 55	95	108/113
BI468058	Heparin cofactor II	R	*Gallus gallus*	AAC16324	1.0e – 63	59	117/197
BI468082	Inter-alpha-trypsin inhibitor heavy chain 3 precursor[b]	R	*Homo sapiens*	S30350	1.0e – 61	56	114/203
BI468072	Serine proteinase inhibitor CP9[b]	R	*Cyprinus carpio*	I50494	1.0e – 41	68	83/122
III. Coagulation factors							
BF228485	Angiopoietin-related protein 3	R	*Mus musculus*	AAD45920	4.5e – 42	51	96/185
BE518584	Beta-fibrinogen precursor[b]	A	*Homo sapiens*	AAA52429	4.0e – 56	64	96/148
BI468018	Fibrin beta	A	*Homo sapiens*	0401173A	3.0e – 54	71	88/123
BE518587	Fibrinogen B-beta subunit precursor[b]	A	*Xenopus laevis*	AAA85283	1.1e – 36	50	75/149
BI544052	Fibrinogen B-beta subunit precursor[b]	A	*Xenopus laevis*	AAA85283	3.0e – 61	54	106/193
BI468019	Fibrinogen gamma polypeptide[b]	A	*Rattus norvegicus*	NP_036691	1.0e – 21	70	50/71
BE518591	Prothrombin[b]	A	*Struthio camelus*	BAA89046	1.2e – 46	53	98/182
BI468077	Thrombin-B chain	R	*Oncorhynchus mykiss*	G42696	2.0e – 93	99	163/164
IV. Complements							
BI468022	Chemotaxin	A	*Oncorhynchus mykiss*	AAG28030	2.0e – 41	82	80/97
BE518585	Complement C3-1[b]	A	*Oncorhynchus mykiss*	P98093	1.0e – 73	95	125/131
BI468073	Complement C3-1[b]	A	*Oncorhynchus mykiss*	P98093	5.0e – 96	95	176/184
BI468074	Complement C3-1[b]	A	*Oncorhynchus mykiss*	P98093	1.0e – 119	95	212/221
BE518598	Complement C3-1[b]	A	*Oncorhynchus mykiss*	P98093	6.0e – 84	91	147/161
BI468035	Complement C3-1[b]	A	*Oncorhynchus mykiss*	P98093	5.0e – 99	92	179/194
BI518586	Complement component C3-3[b]	A	*Oncorhynchus mykiss*	AAC60015	1.0e – 78	77	151/196
BI468048	Complement component C3-3[b]	A	*Oncorhynchus mykiss*	AAC60015	1.0e – 79	83	150/180
BI468051	Complement component C3-3[b]	A	*Oncorhynchus mykiss*	AAC60015	2.0e – 60	92	116/125
BI468034	Complement component C3-3[b]	A	*Oncorhynchus mykiss*	AAC60015	1.0e – 102	81	185/227
BI468017	Complement C4B[b]	A	*Cyprinus carpio*	BAB03285	4.0e – 30	61	61/99
BI468031	Complement C4B[b]	A	*Cyprinus carpio*	BAB03285	7.0e – 53	54	115/212
BI468049	Complement C4B[b]	A	*Cyprinus carpio*	BAB03285	2.0e – 51	62	113/182
BI480050	Complement C4B[b]	A	*Cyprinus carpio*	BAB03285	9.0e – 53	56	106/186
BI468093	Complement component C8 beta	R	*Paralichthys olivaceus*	BAA86877	2.0e – 86	72	146/201
BI468023	Complement component C9	A	*Oncorhynchus mykiss*	P06682	5.0e – 88	95	145/152
BE518599	Complement factor B/C2-B	R	*Oncorhynchus mykiss*	BAB19788	1.0e – 114	90	206/228
BF228496	Complement factor B/C2-B	R	*Oncorhynchus mykiss*	BAB19788	1.0e – 75	86	138/159
BI468056	Complement factor Bf-1	R	*Oncorhynchus mykiss*	AAC83699	1.0e – 104	85	182/212
BI468028	Complement factor Bf-2	R	*Oncorhynchus mykiss*	AAC83698	1.0e – 105	91	186/204
BI468052	Orla C3-1	R	*Oryzias latipes*	BAA92285	5.0e – 48	52	103/197
BI468037	Orla C4	R	*Oryzias latipes*	BAA92287	6.0e – 67	61	122/200

Table 1 (*continued*)

Atlantic salmon ESTs			Matching sequences				
Accession	Name	mRNA Class[a]	Species	Accession number	*E*-value	Identity (%)	Amino acid overlap
V. Lipoproteins							
BE518581	Apolipoprotein A-1 precursor[b]	A	*Salmo trutta*	AAA88542	2.0e – 81	89	156/175
BE518583	Apolipoprotein A-I-1 precursor[b]	A	*Oncorhynchus mykiss*	O57523	3.0e – 89	72	173/239
BE518582	Apolipoprotein A-I-1 precursor[b]	A	*Salmo trutta*	Q91488	9.0e – 38	72	86/119
BF228481	Apolipoprotein CII[b]	A	*Oncorhynchus mykiss*	AAG11410	1.0e – 42	79	89/112
BI468076	Lipoprotein lipase	R	*Pagrus major*	BAB20996	4.0e – 79	77	141/182
VI. Detoxificants							
BI468047	Cytochrome P450 2P2	R	*Fundulus heteroclitus*	AAF21999	8.0e – 84	70	150/213
VII. Glycolysis and gluconeogenesis							
BI468053	Triosephosphate isomerase	R	*Macaca mulatta*	P15426	6.0e – 44	79	86/108
VIII. Ribosomal proteins							
BI468046	Ribosomal protein L34	R	*Homo sapiens*	XP_034711	1.0e – 59	95	112/117
BI468066	40S Ribosomal protein S2	R	*Mus musculus*	P25444	3.0e – 38	95	78/82
BI468061	60S Ribosomal protein L13	R	*Rattus norvegicus*	P41123	1.0e – 69	83	134/161
BI468041	60S Ribosomal protein L13A	R	*Salmo trutta*	Q91487	9.0e – 89	90	166/184
BI468063	60S Ribosomal protein L3	R	*Rattus rattus*	CAA44095	1.6e – 62	95	120/126
BE518608	60S Ribosomal protein L6	R	*Rattus norvegicus*	P21533	3.0e – 78	66	154/233
BE518592	Ribosomal protein S13	A	*Gillichthys mirabilis*	AAG13286	1.1e – 39	91	68/74
IX. Metabolism							
BI468087	Carboxylesterase precursor	R	*Mesocricetus auratus*	BAA23604	4.0e – 26	55	59/106
F228470	Cysteine proteinase	A	*Oncorhynchus mykiss*	AAG30006	4.1e – 27	67	62/92
BI468071	Diamine acetyltransferase	R	*Sus scrofa*	Q28999	2.0e – 52	72	93/129
BI468067	Flavin containing mono-oxygenase 5	R	*Homo sapiens*	XP_001664	1.0e – 61	63	117/183
BE518590	3-hydroxy-3-methylglutaryl-coenzyme A reductase	A	*Homo sapiens*	NP_000850	2.0e – 73	76	142/185
BE518600	Glucosamine-fructose-6-phosphate aminotransferase	R	*Homo sapiens*	NP_002047	1.1e – 100	83	194/233
BE518601	Glucosidase II alpha subunit	R	*Homo sapiens*	AAF66685	1.0e – 62	65	95/144
BE518588	Glutathione peroxidase 3	A	*Mus musculus*	NP_032187	1.0e – 49	66	89/134
BE518602	Guanidinoacetate N-methyltransferase	R	*Homo sapiens*	NP_000147	3.0e – 88	71	151/212
BI468095	Phosphatidylinositol 3-kinase	R	*Rattus norvegicus*	NP_075247	3.0e – 85	73	170/232
BE518606	Protein phosphatase 1	A	*Homo sapiens*	NP_002700	1.0e – 133	98	226/229
BI468088	Retinol dehydrogenase type 6	R	*Mus musculus*	NP_033066	1.0e – 64	59	119/201
BI468043	Sepiapterin reductase	R	*Takifugu rubripes*	AAC60296	9.0e – 56	55	116/210
BI468064	Serine protease-like protein precursor	R	*Salvelinus fontinalis*	AAC17927	3.0e – 89	87	161/185
BI468025	Tyrosine aminotransferase	R	*Rattus norvegicus*	NP_036800	7.0e – 97	73	162/220
X. Housekeeping Genes							
BI468054	CCCH zinc finger protein C3H-2	R	*Xenopus laevis*	AAD24208	1.0e – 31	51	89/174
BI468068	CCT	R	*Carassius auratus*	BAA89277	9.0e – 78	82	147/178
BI468040	DEAD-box protein abstrakt	R	*Homo sapiens*	NP_057306	1.0e – 87	80	156/195
BI468079	Elongation factor 2	R	*Gallus gallus*	Q90705	1.0e – 99	91	170/185
BI468081	Heat shock protein 108	R	*Gallus gallus*	CAA28629	5.0e – 89	79	157/198
BI468080	Heat shock protein hsp90 beta	R	*Salmo salar*	AAD30275	1.0e – 80	69	156/223
BI468062	Mini chromosome maintenance deficient 6	R	*Mus musculus*	NP_032593	6.4e – 43	59	101/169
BI468069	146D nuclear protein	R	*Xenopus laevis*	T30887	1.0e – 105	96	191/197
BI468036	Nuclear receptor coactivator 4; RFG	R	*Mus musculus*	NP_062718	2.0e – 32	50	73/146
BI468085	Proteasome activator subunit 2	R	*Danio rerio*	AAF05817	1.0e – 73	71	142/200
BI468075	Translation elongation factor	R	*Xenopus laevis*	I51237	6.0e – 70	74	126/170
XI. Mitochondrion							
BI468060	Cytochrome b	R	*Acantholingua orhidana*	AAF25872	6.0e – 63	86	123/142
BI468083	Cytochrome c oxidase subunit I	R	*Diplophos taenia*	NP_073653	2.0e – 63	62	135/217
BI468044	NADH dehydrogenase subunit 1	R	*Salvelinus alpinus*	NP_008673	1.0e – 65	67	139/207
BI468045	Peptidyl-prolyl cis-trans isomerase mitochondrial precursor	R	*Rattus norvegicus*	P29117	1.0e – 64	67	122/182
XII. Cell signalling/communication							
BI468057	Activated protein kinase C receptor	R	*Mus musculus*	AAG29506	1.0e – 119	97	201/207

Table 1 (*continued*)

Atlantic salmon ESTs			Matching sequences				
Accession	Name	mRNA Class[a]	Species	Accession number	*E*-value	Identity (%)	Amino acid overlap
BI468042	Calreticulin	R	*Danio rerio*	AAF13700	1.0e – 114	85	183/213
BE518603	Integrin beta-1 precursor	R	*Xenopus laevis*	P12606	4.0e – 96	83	165/198
BI468059	IQ motif containing GTPase activating protein 2	R	*Homo sapiens*	NP_006624	2.0e – 90	77	156/202
BI468070	Probable calcium-binding protein	R	*Homo sapiens*	JS0027	3.0e – 20	63	51/80
BI468038	ras homolog gene family, member A	R	*Homo sapiens*	NP_001655	1.0e – 99	97	177/181
BE518607	RAS-related protein RAB-8	R	*Discopyge ommata*	P22128	2.0e – 47	88	64/72
BE518610	Transthyretin precursor	R	*Sparus aurata*	AAC26108	1.0e – 51	72	94/129
XIII. Unclassified							
BI468033	ABCA1	R	*Homo sapiens*	AAF86276	4.0e – 86	82	151/183
BE518597	Alpha tubulin	A	*Chionodraco*	AAG15366	3.0e – 65	85	119/139
BI468027	Cathepsin L	R	*Danio rerio*	CAA69623	4.0e – 57	63	95/149
BI468055	C-type lectin 2-1	R	*Oncorhynchus mykiss*	AAG30024	2.0e – 28	92	53/57
BI468092	Hepatocyte growth factor-like 1	R	*Danio rerio*	AAK54207	6.0e – 75	90	118/131
BI468039	High-mobility group protein 4	R	*Homo sapiens*	XP_013062	5.0e – 66	75	118/156
BI468091	hnRNP protein (pre-mRNA binding K protein)	R	*Xenopus laevis*	S41224	2.0e – 81	81	152/186
BI468032	Hypoxia-inducible gene 1	R	*Gillichthys mirabilis*	AAG13326	7.0e – 30	72	61/84
BE518604	Kal1.1	R	*Danio rerio*	AAF25779	3.0e – 76	61	144/235
BE518605	Progesterone receptor-related protein p23	R	*Gallus gallus*	B56211	4.0e – 51	60	99/163
BI468024	Prosaposin precursor	A	*Danio rerio*	AAG32919	2.0e – 37	59	75/126
BI468090	Sec 61 alpha form A	R	*Oncorhynchus mykiss*	AAK29081	1.0e – 117	99	211/212
BI468089	15 KDa selenoprotein	R	*Homo sapiens*	CAC04186	2.0e – 38	66	76/115
BI468026	Signal sequence receptor beta subunit	R	*Xenopus laevis*	AAK15544	5.0e – 82	93	148/158
BE518609	Striatin	R	*Mus musculus*	NP_035630	5.0e – 55	80	104/129

[a] mRNA Class: A, abundant mRNA encoding gene; R, rare mRNA encoding gene.
[b] Genes with highly expressed homologues in human and mouse liver [20].

were translated in all six reading frames and used to search for amino acid homology in the NCBI nr protein database. Table 1 lists the 117 (48%) Atlantic salmon ESTs that showed sufficient homology along with the details of the best match sequences, homology values, and membership of abundant or rare mRNA classes. Based on our subjective criteria for unambiguous salmon gene identification, and accounting for EST redundancy whereby different clusters of EST corresponded to different regions of the same mRNA transcript, 93 different salmon genes were identified. In terms of mRNA class, these constituted 23 abundant and 70 rare mRNA encoding genes (Table 1). The 129 unidentified Atlantic salmon ESTs were composed of 104 salmon ESTs that did not reach the criteria for unequivocal gene identification, 7 salmon ESTs which did show sufficient homology but to, as yet, unidentified mouse and human genes, and 18 salmon ESTs which showed no database match.

Analysis of identified Atlantic salmon genes expressed in liver

Of the 93 identified Atlantic salmon genes, only 4 were mitochondrion-encoded and a further 7 represented ribosomal protein genes. Twenty-five of the identified genes had previously been reported in salmonids including 7 genes cloned from the Atlantic salmon (Table 1). The 93 identified salmon genes were classified according to cellular role by comparison with the more comprehensive mammalian liver gene expression data available [20]. The results showed that genes associated with typical mammalian liver function were also well represented in the salmon liver cDNA library, e.g., 33 of the 93 identified salmon genes encoded either plasma proteins, protease inhibitors, coagulation factors, complements, lipoproteins, detoxificants, or enzymes involved in glycolysis and gluconeogenesis (Table 1). Furthermore, 16 of these 33 salmon genes were shown to encode abundant mRNA transcripts in salmon liver, including 12 genes for which homologues have been identified that also showed strong mRNA expression profiles in mouse and human liver (Table 1). These genes included those encoding the serum albumin, haptoglobin, and serotransferrin transport proteins; the apolipoprotein A-I and C-II lipoproteins; the complements C3-1, C3-3, and C4B acute phase response proteins; the β- and γ-fibrinogen subunits and prothrombin blood clotting cascade proteins; and the α-1-microglobulin/inter-α-trypsin protease inhibitor. This conservation of

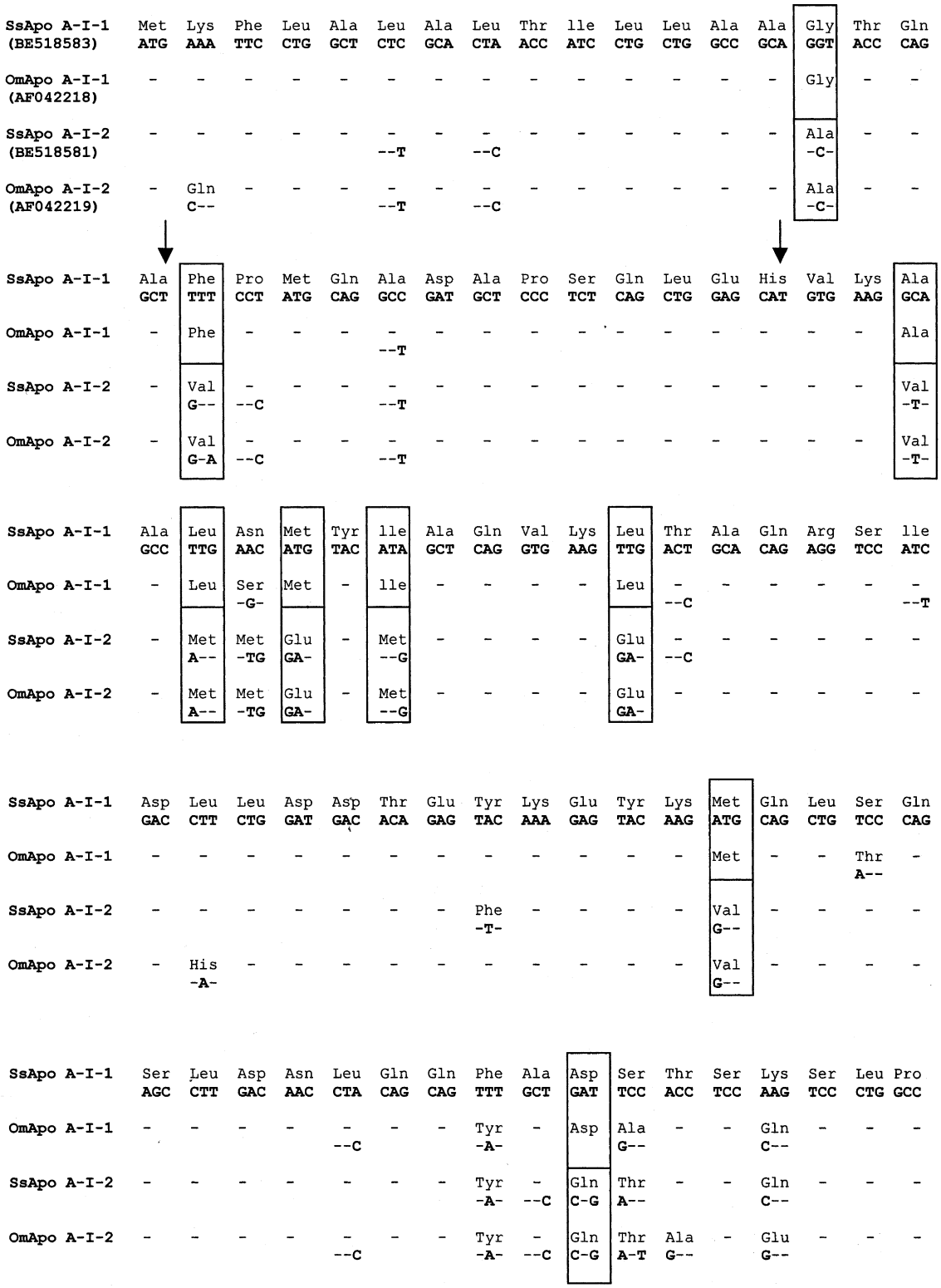

Fig. 1. Comparative alignment of the 85 inferred N-terminal amino acids of the Atlantic salmon (SsApo) and rainbow trout (OmApo) Apolipoprotein A-I forms. Only the amino acid substitutions and nucleotide mutations are indicated. The boxed residues indicate the substitutions/mutations conserved between both ApoA-I forms. The two downward arrows indicate the predicted cleavage sites for the signal peptide and prepeptide.

highly expressed genes suggests that many components of the primary liver functions evolved before the teleost fish-tetrapod divergence 400–450 million years ago. It also indicates that the range of salmon liver ESTs identified in this study spans the known liver functions and therefore should be useful to monitor liver gene expression under different physiological conditions (e.g., smoltification, development, pathogen infection).

Atlantic salmon genes with two cDNA forms

Genes represented by two forms of cDNA have previously been reported in some salmonid species and this is thought to reflect the high incidence of gene duplication in salmonids after a common tetraploid event about 100 million year ago [21]. In Atlantic salmon, two cDNA variants have been described for serum albumin [22], serotransferrin [6], IgM heavy chain [23], α- and β-globin [24], and parvalbumin [25]. This study found between 2 and 5 EST clusters for 11 of the 93 identified salmon genes, i.e., haptoglobin, hemopexin-like protein, serotransferrin, serum albumin, α-1-microglobulin/inter-α-trypsin inhibitor, β-fibrinogen, complements C3-1, C3-3, C4B, and B/C2-B, and apolipoprotein A-I (and a unique EST representing each cluster has been deposited in the GenBank dbEST and listed in Table 1). Analysis of the homology alignments showed that while both reported cDNA forms of the serotransferrin and serum albumin genes were identified by individual EST clusters, all but one of the remaining 9 salmon genes identified by >1 cluster reflected ESTs derived from different regions of the mRNA transcript. The exception was the apolipoprotein A-I gene where two of the three EST clusters (accession numbers BE518581 and BE518583) showed homology to the same region located at the 5′ terminus of the gene. While nucleotide variation between the ESTs within each cluster was <1% over the mRNA region which codes for the initial 85 amino acids of Apolipoprotein A-I, the nucleotide variation between both clusters was 9%. This intra-cluster variation is composed of 24 nucleotide differences conserved within each EST cluster which result in 14 amino acid changes between both inferred protein sequences over this region (Fig. 1). This is similar to that reported for the rainbow trout (*Oncorhynchus mykiss*) apolipoprotein A-I gene where two cDNA forms, A-I-1 and A-I-2, have been identified which showed about 10% nucleotide variation and 14 amino acid substitutions in the 85 amino acid N-terminal region [13]. A comparison of the N-terminal region of the inferred trout and salmon Apolipoprotein A-1 (ApoA-I) amino acid sequences showed that each of the salmon ApoA-I forms is most homologous to one of the trout ApoA-I forms, i.e. salmon EST BE518583 equates to trout A-I-1, while salmon EST BE518581 equates to trout A-I-2. Nine of the 14 amino acid changes (and 12 of the 13 relevant nucleotide differences) that distinguish between the two Apo-I forms in both species are conserved between both species (Fig. 1). This suggests that apolipoprotein A-I can be added to the list of Atlantic salmon genes for which the presence of two cDNA variants has been described.

Acknowledgments

This research was supported by the EU FAIR Programme SALGENE project (Contract FAIR-CT94-4314), the Enterprise Ireland Basic Research Grant Scheme (Contract SC/97/320), and the Irish National Centre for Biomedical Engineering Science (PRTLI-1, DP-05). NCC was supported by NUI Galway and Enterprise Ireland postgraduate fellowships.

References

[1] A. Blouin, R.P. Bolender, E.R. Weibel, Distribution of organelles and membranes between hepatocytes and nonhepatocytes in the rat liver parenchyma. A stereological study, J. Cell Biol. 72 (1977) 441–455.

[2] G. Feldmann, Tissue and cellular organisation of the liver, in: F. Tronche, M. Yaniv (Eds.), Liver Gene Expression, R.G. Landes Company, Austin, TX, 1994, pp. 17–34.

[3] E. Derman, K. Krauter, L. Walling, C. Weinberger, M. Ray, D.E. Darnell, Transcriptional control in the production of liver specific mRNAs, Cell 23 (1981) 731–739.

[4] L. Byrnes, F. Gannon, Atlantic salmon (*Salmo salar*) serum albumin: cDNA sequence, evolution and tissue expression, DNA Cell Biol. 9 (1990) 647–655.

[5] R. Powell, D.G. Higgins, J. Wolff, L. Byrnes, M. Stack, P.M. Sharp, F. Gannon, The salmon gene encoding apolipoprotein A-I: cDNA sequence, tissue expression and evolution, Gene 104 (1991) 155–161.

[6] A.M. Kvingedal, K.-A. Rorvik, P. Alestrom, Cloning and characterisation of Atlantic salmon (*Salmo salar*) serum transferrin cDNA, Mol. Mar. Biol. Biotechnol. 2 (1993) 233–238.

[7] S. Hanley, R. Powell, Sequence of a cDNA clone encoding the Atlantic salmon α-1-microglobulin/bikunin protein, Gene 147 (1994) 297–298.

[8] C.D. Hurst, S.E. Bartlett, W.S. Davidson, I.J. Bruce, The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*, Gene 239 (1999) 237–242.

[9] J.C. Robertson, T.M. Bradley, Hepatic ultrastructure changes associated with the parr-smolt transformation of Atlantic salmon (*Salmo salar*), J. Exp. Zool. 260 (1991) 135–148.

[10] R.L. Blake, F.L. Roberts, R.L. Saunders, Parr-smolt transformation of Atlantic salmon (*Salmo salar*): activities of two respiratory enzymes and concentrations of mitochondria in the liver, Can. J. Fish. Aquat. Sci. 41 (1984) 199–203.

[11] M.S. Sheridan, Alterations in lipid metabolism accompanying smoltification and sea water adaptation in salmonid fish, Aquaculture 82 (1989) 191–203.

[12] G. Hardiman, L. Byrnes, F. Gannon, An analysis of highly expressed salmon liver genes during smoltification in the Atlantic salmon (*Salmo salar* L.), Mol. Mar. Biol. Biotechnol. 3 (1994) 51–56.

[13] G.P. Delcuve, J.M. Sun, J.R. Davie, Expression of rainbow trout apolipoprotein A-I genes in liver and hepatocellular carcinoma, J. Lipid Res. 33 (1992) 251–262.

[14] G.C. Davey, N.C. Caplice, S.A. Martin, R. Powell, A survey of genes in the Atlantic salmon (*Salmo salar*) as identified by expressed sequence tags, Gene 263 (2001) 121–130.

[15] M.B. Dworkin, I.B. Dawid, Use of a cloned library for the study of abundant poly(A) + RNA during *Xenopus laevis* development, Dev. Biol. 76 (1980) 449–464.
[16] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.
[17] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.
[18] D.A. Benson, M.S. Boguski, D.J. Lipman, J. Ostell, B.F. Francis Ouellette, GenBank, Nucleic Acids Res. 26 (1998) 1–7.
[19] W.R. Pearson, Identifying distantly related protein sequences, Comput. Appl. Biosci. 13 (1997) 325–332.
[20] S. Kawamoto, Y. Matsumoto, K. Mizuno, K. Okubo, K. Matsubara, Expression profiles of active genes in human and mouse livers, Gene 174 (1996) 151–158.
[21] F.W. Allendorf, G.H. Thorgaard, Tetraploidy and the evolution of salmonid fishes, in: B.J. Turner (Ed.), Evolutionary Genetics of Fishes, Plenum Press, New York, 1984, pp. 1–53.
[22] L. Byrnes, F. Gannon, Sequence analysis of a second cDNA encoding Atlantic salmon (*Salmo salar*) serum albumin, Gene 120 (1992) 319–320.
[23] I. Hordvik, A.M. Voie, J. Glette, R. Male, C. Endresen, Cloning and sequence analysis of 2 isotypic IgM heavy-chain genes from Atlantic salmon, *Salmo salar* L, Eur. J. Immunol. 22 (1992) 2957–2962.
[24] T. McMorrow, A. Wagner, F. Deryckere, F. Gannon, Structural organisation and sequence analysis of the globin locus in Atlantic salmon, DNA Cell Biol. 15 (1996) 407–414.
[25] C.D.-V. Lindstrom, T. Do, I. Hordvik, C. Endresen, S. Elsayed, Cloning of two distinct cDNAs encoding parvalbumin, the major allergen of Atlantic salmon (*Salmo salar*), Scand. J. Immunol. 44 (1996) 335–344.