K. Tsukamoto, T. Kawamura, T. Takeuchi, T. D. Beard, Jr. and M. J. Kaiser, eds.
Fisheries for Global Welfare and Environment, 5th World Fisheries Congress 2008, pp. 77–86.

Genomics and the Genome Duplication in Salmonids

Ben F. Koop[1] and William S. Davidson[2]

[1]*Department of Biology*
University of Victoria
Victoria, BC V8W 2Y2 Canada

[2]*Department of Molecular Biology & Biochemistry*
Simon Fraser University
Burnaby, BC V5A 1S6 Canada

*E-mail: bkoop@uvic.ca

Salmonids have been classified into nine genera and roughly sixty-eight closely related species. Considerable basic biological knowledge of trout and salmon has been developed as a result of their widespread use in scientific research, as an environmental sentinel species and their use as a food and sport fish. More is known about the biology of salmonids than nearly any other fish group. More recently genomic studies have contributed gene lists and genetic and physical maps as well as new technologies to integrate gene expression patterns with traditional ecological, evolutionary and physiological studies. Of particular interest is that the common ancestor of salmonids purportedly experienced a whole genome duplication event between 25 and 100 MYA. Given that gene duplication may one of the most important mechanisms of generating molecular and physiological diversity, the study of salmonid genomes provides an excellent opportunity to examine genetic and ultimately species diversity. Here we review some of the recent genomic data and suggest areas of further research.

KEYWORDS Salmonids; genomics; gene duplication; diversity; evolution; repeats; ESTs

1. Introduction

Our long-term objectives are to study the impact of genome duplication on the modes and rates of change in gene number, gene function and evolution particularly in the purported genome duplication in salmonids. This entails the study of loss, dosage effects, neofunctionalization and subfunctionalization of gene functions. It also entails the study of mechanisms of genome structure restabilization after genome duplication and the potential role of interspersed repetitive elements

in restructuring the genome through rearrangements, insertions and deletions. This should be relatively straight-forward but recent EST data suggest a much more complicated history than can be explained by a single genome duplication in salmonids. Given the large volume of work done in salmonids and that "gene duplication is probably the most important mechanism for generating new genes and new biochemical processes that have facilitated the evolution of complex organisms from primitive ones," (Li 1983) the first and most important objective of our studies is to test and verify an ancestral salmonid genome duplication and then, if our initial data are correct, document the extent to which very recent segmental gene duplications have occurred. Genome duplications are thought to be defining genetic events in the evolution of complex vertebrates. In our studies we have focused efforts on documenting and characterizing; i) global versus regional duplication patterns, ii) gene family duplications (immune systems), and iii) the role of interspersed repeats in genome stabilization. The purpose of this chapter is review some of the broad and very general interpretations of recent data pertaining to the purported genome duplication in salmonids and suggest areas of further research.

The Salmonidae family includes; whitefish and ciscos (subfamily Coregoninae); graylings (Thymallinae); and trout, salmon and charr (Salmoninae), Fig. 1.

These fish have been further classified into nine genera and roughly sixty-eight species (Nelson 2006) that are more than 92% similar at the DNA level. They are native to the cooler climates of the Northern Hemisphere but have been widely introduced around the world.

Considerable basic biological knowledge of trout and salmon has been developed as a result of their widespread use in scientific research, as an environmental sentinel species and their use as a food and sport fish. More is known about the physiology and biology of salmonids than any other fish group. In the past 20 years there have been over 20,000 reports on the ecology, behavior, physiology and genetics of these species (Thorgaard *et al*. 2002). These studies provide data from an economically important and phylogenetically distinct group of fish. Recent genome studies have targeted Ostariophysi (zebrafish, catfish, flathead minnow, etc), or Acanthopterygii (cod, ciclids, fugu, medaka, sticklebacks, rockfish) euteleostei lineages that have been separated from Protacanthopterygii (salmonids) 200–300 million years ago [MYA] (Ichiguro *et al*. 2003; Nelson 2006; Yamanoue *et al*. 2006; Inoue *et al*. 2006). Salmonids, with their genome duplication and wealth of biological data, are excellent model organisms for studying comparative genomics, evolutionary processes, fates of duplicate genes and genetic architecture of complex phenotypes, as well as carcinogenesis, toxicology,

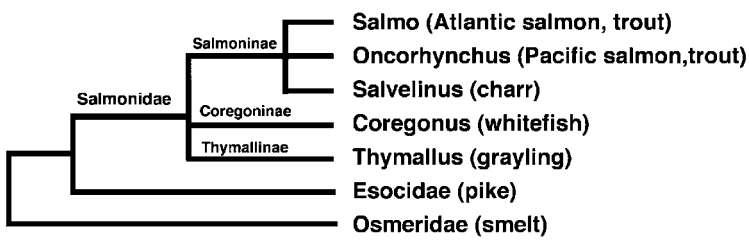

Fig. 1. General relationships of the major salmonid and closest sister groups (Ichiguro *et al.* 2003; Nelson 2006; Yamanoue *et al.* 2006; Inoue *et al.* 2006).

comparative immunology, disease, ecology, physiology, nutrition and many other genetic and physiological processes (Thorgaard *et al*. 2002).

2. Genome Duplications

The common ancestor of salmonids purportedly experienced a whole genome duplication event between 25 and 100 MYA (Ohno 1970; Allendorf and Thorgaard 1984). Extant species would therefore be the result of reversion back to a stable diploid state. Evidence for the ancestral salmonid autotetraploid genome duplication includes; multivalent chromosome formation during meiosis and evidence for tetrasome segregation at some loci; one of the largest euteleost genome sizes (3–4.5 pg) with double that of sister orders Esociformes (0.8–1.8 pg, pike) and Osmeriformes (0.7 pg, smelt) (Gregory 2005); homeologous chromosomal segments based on recent genetic maps and comparative studies using microsatellite markers (Leder *et al*. 2004), and duplicated gene family studies such as Hox, MHC, growth hormone, many allozymes and others (Thorgaard *et al*. 2002; McKay *et al*. 2004; Moghadam *et al*. 2005; Hoegg and Meyer 2005; Shiina *et al*. 2005).

The genome duplication in salmonids is just the most recent of a series of genome duplications in this lineage. There is good evidence primarily from sequenced zebrafish, and pufferfish genome sequences for tetraploidization/rediploidization early in the ray-finned fish lineage (350–400 MYA) (Hoegg *et al*. 2004; Volff 2004; Steinke *et al*. 2006; Crow *et al*. 2006). In addition, several studies have supported Ohno's original 2R hypothesis that one and possibly two rounds of genome duplication occurred in early ancestral vertebrates (~600 MYA) (Taylor and Raes 2004). The 2R hypothesis (1 to 2 to 4 gene rule) is the prevalent model that is used to explain the evolution of gene families and vertebrate genomes. The 3R hypothesis further explains gene families in ray-fined fish (1–2–4–8). While evidence from Hox, sodium channels, glycolytic enzymes and other genes (Novak *et al*. 2006) support the 3R hypothesis, debate continues (Hughes 1999) and it may be impossible to deduce events that happened so long ago using current methods. Interestingly, in salmonids, 14 Hox clusters have been recently identified thus supporting a 4R hypothesis (1–2–4–8–16 with 2 losses) (Allendorf and Thorgaard 1984).

The importance of understanding the role of genome duplications lies in that vertebrate species diversity and body plan diversity have commonly been linked to genome duplications although there is some debate on how well we can draw these conclusions based on the very old genome duplications commonly studied (Donoghue 2006). The number of salmonid species and the relatively recent genome duplication make salmonids ideal for examining recent events that could have played such a pivotal role in generating gene diversity and possible species diversity found in modern vertebrates.

The most commonly discussed mechanism by which organisms acquire new functions has been Ohno's classical model (Ohno 1970) which is based on the principle that duplication events provide gene redundancy upon which natural selection may be relaxed and new functions evolve. This model predicts that one of the genes remains under conservative selection while the most likely fate of the other gene duplicate is nonfunctionalization via mutations. However, occasionally, rare beneficial mutations at the redundant loci followed by positive natural selection may give rise to a gene with a novel function (neo-functionalization), thus preserving both duplicates (Ohno 1970). This model predicts that retention of duplicates is a rare event. However, what is observed in the case of many genes is that duplicated genes are commonly retained (Hughes 1999).

Previous estimates of 30–70% duplicate retention occur in salmonids (est. time of duplication is 25–100 MYA), 20% retention in teleost duplicates (~350 MYA), 50% retention in xenopus duplicates (30 MYA) and 72% retention in rice (11 MYA). To account for this high level of duplicate retention, Hughes (1999) suggested that gene duplications that lead to functionally distinct proteins are ordinarily preceded by a period of gene sharing: that is, a period in which a single generalist gene performs two or more distinct functions, perhaps in different tissues or in different developmental stages. Upon gene duplication, these functions are able to take on separate specialist functions through positive selective forces. This model is used to explain some of the very large immunoglobulin, olfactory receptor, and defensin gene family structures in which selection for sub-functionalization, or genes expressed differently in various tissues or developmental stages appears to precede gene duplication. More recently, the duplication/degeneration/complementation (DDC) model (Force *et al*. 1999; Lynch 2000; Hughes 2005) proposes partitioning of the expression patterns of the original gene between the duplicates (sub-functionalization) via complementary degenerative mutations, particularly in regulatory regions. This does not require positive selection pressure to preserve both copies in the genome, only complementary mutations in the duplicates. Larger numbers of small populations are particularly important variables in this model. Salmonids offer particularly valuable life histories in this regard.

3. Expressed Sequence Data

One very puzzling observation arising from preliminary EST (expressed sequence tag) analysis of the Atlantic salmon (436,000 Atlantic salmon ESTs; GenBank, September 2007) is that the number of expressed duplicate transcripts (presumed paralogues) decreases more regularly with respect to percent divergence than expected (see Fig. 2).

Over 81,000 contigs derived from 436,000 ESTs (cGRASP) were compared to each other by BLAST analysis and the number of alignment pairs with e-values < 1e-25 and lengths > 200 bp were plotted against % identity. A similar analysis of Atlantic salmon EST contigs compared with rainbow trout EST contigs showed a peak at approximately 94% identity and 92% when compared to lake whitefish ESTs (unpublished data, Brown 2008). Within this overall context, the data from Fig. 2 suggest that many of the Atlantic salmon contig consensus sequences are more similar to each other than Atlantic salmon contigs consensus sequences were to rainbow trout or whitefish consensus sequences. These preliminary data indicate that there may be more gene duplicates of recent origin than expected from a single ancestral genome duplication in the common ancestor of salmonids. These observations raise very intriguing questions, i) why are there so few paralogous alignment pairs among expressed genes and ii) why is there no paralogous peak at ~80–90% similarity?

The distinguishing feature between multiple segmental genome duplications and a genome duplication is that a whole genome duplication affects all genes at the same time and predicts similar levels of divergence among duplicates; within the bounds of differential selection pressures. Multiple segmental duplications on the other hand affect only portions of the genome and several may occur over a broader time period. In the later instance we expect a broader range of divergence values between duplicates, perhaps similar to the pattern shown in Fig. 2. Extended genomic sequence data are only beginning to become available and the problems associated with resolving extensive duplications requires diligence. Some regions clearly show an 80–90% divergence in noncoding sequences [MHC class I regions;

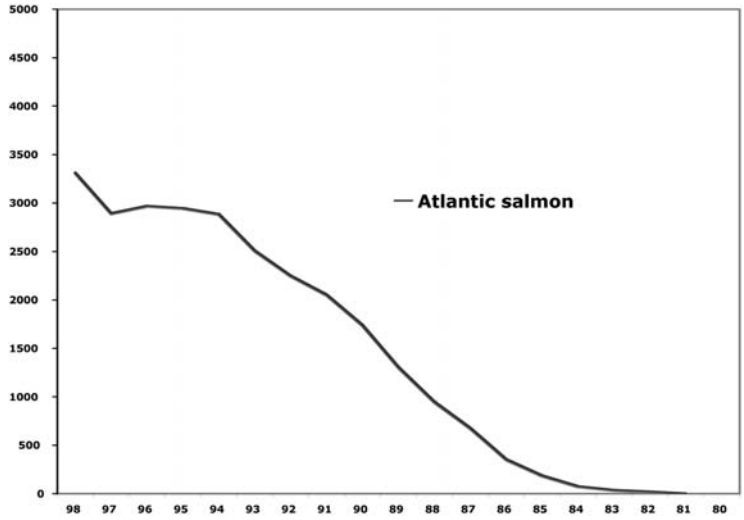

Fig. 2. The number of Atlantic salmon EST contigs that have a BLAST hit (E-value < 1e–25; length > 200 bp) against another Atlantic salmon EST contig is plotted with the percent identity (x-axis). Modified from GD Brown thesis (Brown 2008).

>500 kb comparisons (Shiina *et al*. 2005; Lukacs *et al*. 2007) and others (TCRG regions; >250 kb comparisons (Yazawa *et al*. 2008)] show no similarity at all outside of coding regions. However since most gene targets were originally *a-priori* chosen to verify the ancestral salmonid genome duplication hypothesis, we need to reexamine our initial assumptions and look for examples of other gene/region duplications. Specifically we need to;

i) Identify more duplicate gene sets that are 92–96% similar in inter-genic regions, isolate and characterize corresponding BAC clones and use FISH analysis to test for segmental duplications (on the same chromosome) or a hallmark of a genome duplication (on different chromosomes).
ii) Examine all genes in duplicated genomic regions (existing plus new regions) for changes in rates of evolution among duplicates (*Ka*/*Ks* ratios), determine frequency of gene loss, and identify levels of altered transcription in tissues and developmental life stages. This will enable us to more accurately identify patterns of duplicate death, dosage effects, rates of gene evolution and possible sub/neo-functionalization.
iii) Use the total EST data from rainbow trout, chinook, sockeye, brook trout, lake whitefish, grayling, northern pike and rainbow smelt to; a) build contigs and consensus transcript sequences from all species b) identify bins of similar transcripts, c) for each bin, generate common alignments and identify largest common alignment regions, d) for each bin generate phylogenies and analyze them for species relationships and gene duplications, e) examine common patterns that support or refute the ancestral salmonid duplication hypothesis. The use of species relationships will enable us to clearly identify gene duplications in ancestral salmonids and those occurring more recently.
iv) Continue building a full-length salmonid gene database. This will be done using

a combination of EST assemblies and sequenced full-length cDNA clones. Efforts to identify 8,000–10,000 full-length cDNAs will need to expand and the results made available for all salmonid researchers.

4. Repeated Regions

A second puzzling observation arises from an analysis of salmonid transposons. We initially hypothesized that subsequent to an ancestral salmonid genome duplication and during the rediploidization process, that we would expect an increase in transposable element (TE) activity to facilitate the restabilization of the salmonid genome back to a diploid state required for successful cell replication. Transposable elements (TEs) are sequences capable of integrating into new sites within the genome and are classified into retrotransposons (use RNA intermediates and reverse transcription) or transposons (no reverse transcription). TEs can alter or disrupt gene expression depending on specific insertion locations. Several observations that have been made on TEs specifically in fish are: i) the diversity of TEs is higher in fish than mammals; ii) there tends to be a higher turnover of TEs in teleosts; and iii) TEs tend to be localized especially to heterochromatic areas of the chromosomes. It was hypothesized that subsequent to an ancestral salmonid genome duplication, during the rediploidization process, that there would be an increase in TE activity to facilitate the restabilization of the salmonid genome. For non-coding TE DNA, this would correspond to sequence divergence values greater than that between salmon/trout and whitefish (~8%). While the specific role of transposable elements in speciation, genome duplication and subsequent genome restabilization remains uncertain (Kazazian 2004; Sverdlov 2000) there is no doubt that TEs can be very important drivers of genome evolution. Certainly the enormous impact of TE's is evident in the more than 30% composition of the total human genome.

When we did a phylogenetic comparison of several hundred TEs (avg. 1.5 kb in length and consisting of over 30% of the 7 MB of genomic DNA thus far analyzed; Fig. 3), we found that one of the waves of TE activity (comprising ~one third of the total) roughly corresponded to the time of the *Salmo*/*Oncorhynchus* speciation period and another wave (again one third of the total) seems to represent very recent and ongoing TE activity (de Boer *et al*. 2007). the majority of new TE families correspond to recent waves of activity (~6–8% and 2–4% divergence) that correspond to times and levels of divergence seen among salmonid species. We had expected the majority of transposon activity to have occurred immediately after an ancestral genome duplication and during a restabilization period (or and estimated 10–20% divergence). These observations also raise an intriguing question. As transposons are both an indicator and facilitator of extensive genome-wide changes, why do so few transposon duplications appear to occur pre-salmonid speciation during which a purported genome duplication and subsequent rediploidization/restabilization process occurred?

What we see from our initial data is that one of the waves of TE activity roughly corresponds to the time of the *Salmo*/*Oncorhynchus* speciation period (corresponding to 94% identity or 6% difference) and another wave seems to represent extensive ongoing TE activity. It is not clear whether the observed TE activity is associated with speciation processes, or is involved in ongoing restabilization efforts to eliminate occasional tetravalent structures evident during cell division. What we need to do is to obtain better estimates of when and where bursts of TE activity occurred by expanding the number of Tc1-like elements in not only Atlantic salmon but also other salmonid species. We also need to test whether the same

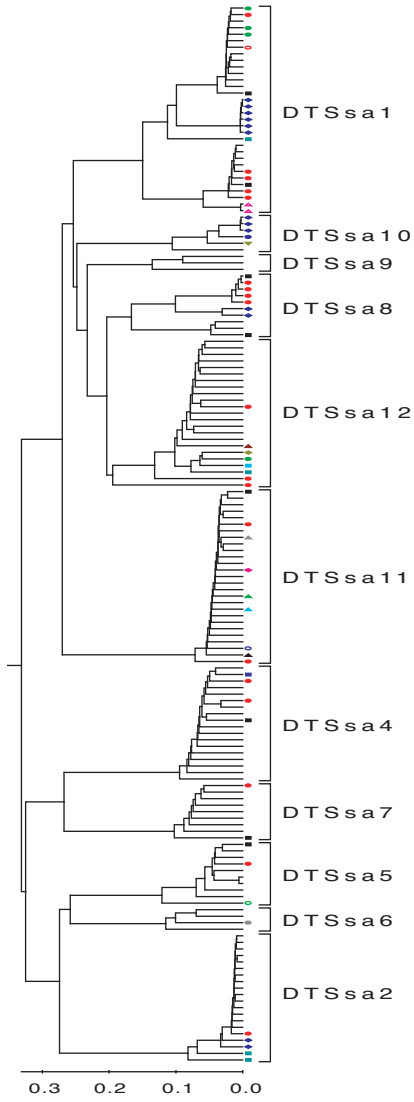

Fig. 3. Recent ancestry of salmonid transposons. Salmon DNA transposons, SALT1 and Tss sequences, transposons obtained from approximately 5 Mbp of in-house *Salmo salar* BAC clone sequences and sequences recovered from GenBank searches were aligned and a UPGMA phylogenetic tree was generated. The absences of a marker on terminal branch links indicates *Salmo salar* repeat elements and colored markers indicated repeats from other fish species (see de Boer *et al.* 2007 for details). Distances (%) are indicated by the bottom bar. Figure modified from deBoer *et al.* (2007).

pattern is found in other repeat families. Initial observations from SINE and LINE elements suggest a similar pattern. One of the benefits of these studies will be a repeat library that will be essential in genome sequence analysis and resolution of repeated regions.

Still another intriguing result comes about from the identification of *Xenopus*, catfish, and lamprey Tc1 transposons all within genomic BAC sequences that show 95–99% similarity to *Salmo salar* Tc1 transposons (de Boer *et al*. 2007). Additional GenBank searches further identified >97% identity to several partial (480 bp) Tc1-like transposons in EST clones of *Schistosoma japonicum* (Melamed *et al*. 2004). This exciting observation strongly suggests that there may be horizontal transfer of these elements. Testing this idea will require confirming the initial observations by obtaining much broader TE sequence representation in salmonids as well as different fish species. Extensive phylogenetic studies will be needed to determine the spread and timing of elements within species. In addition, two open reading frames with high similarity to transposase have been found, so we will try to confirm RNA and protein expression. This will be very important in the further development of vertebrate gene transfer biotechnologies, like the Sleeping Beauty vector (Wadman *et al*. 2005). The potential for horizontal DNA transfer between frogs, lampreys, catfish and salmonids is particularly intriguing in developing our understanding the role of repeat elements in times and regions of genomic stress.

5. Discussion

Salmon, trout and charr comprise a group of fish that are of great economic and societal importance to coastal, rural and aboriginal communities of many northern countries. Although Atlantic salmon is the main aquaculture species, there are also vibrant

commercial fisheries for wild salmonids. As salmonid aquaculture continues to develop and expand, it must find ways of minimizing its impact on wild fisheries and the environment. There is a real need for domesticated broodstocks that maximize disease resistance, optimize adaptation to local environments and minimize escape viability and impact on wild populations. A better understanding of how natural populations of salmonids adapt to local conditions will benefit agencies that have to make management decisions concerning stock assessment and harvesting plans. The aquaculture industry and enhancement schemes can also take advantage of this knowledge so that all commercial activities relating to salmonids can develop in a complementary manner.

One of our objectives is to demonstrate the power of genomics to conduct scientifically exciting research that will yield practical benefits for salmonid production and provide sound advice for managing wild stocks and the environment. Expansion of existing genomic research efforts of Canada, Norway, the United States, Chile and the United Kingdom will: (i) expand genomic resources for Atlantic salmon and rainbow trout; (ii) extend genomic resources to other salmonids including, chinook, sockeye, whitefish, grayling, and brook charr; and (iii) use the existing and expanded genomic resources as tools to answer questions that are of biological, economic and social importance to aquaculture, conservation, and the environment.

Physical map resources, gene identification (ESTs) and BAC end sequencing provide the starting resources for many genomic activities. Genomic efforts have built a physical genomic framework for Atlantic salmon consisting of approximately 4,200 contigs (Ng *et al*. 2005), and have provided key BAC libraries for both Atlantic salmon, rainbow trout and rainbow smelt (cGRASP; Ng *et al*. 2005; von Schalburg *et al*. 2008a). Contig BAC end sequences and end sequences from more than 100,000 Atlantic salmon BACs enable genome characterization, and provide resources for consolidating the physical map and integrating it with the linkage map. Completion of over 700,000 ESTs from Atlantic salmon, rainbow trout, chinook, sockeye, brook trout, Arctic whitefish, Arctic grayling, northern pike and rainbow smelt provide an excellent foundation for the identification of genes and polymorphic variation in genic regions (cGRASP). Atlantic salmon and rainbow trout have the 17th and 27th largest EST representation of any species to date (October, 2007). This foundation facilitates the identification of full-length coding sequences of genes. These resources provide an excellent gene representation for application tools such as microarrays, and provide tens of thousands of potential polymorphic markers for genetic maps. The distribution of more than 3,000 16K microarray (cGRASP; von Schalburg *et al*. 2008b) slides in the last year to over 40 laboratories around the world points to the impact of these resources on the fish community. These arrays facilitate unsurpassed assessment of gene expression of thousands of genes and provide whole new avenues of studying duplicated gene families. The completion of several megabases of finished genomic has enabled a more thorough understanding of key functions such as repeat families, immunity (MHC, TCR, IL-2) and growth (GH1, GH2) as well as suggesting candidate genes for sex determination and upper temperature tolerance.

At sequence divergences corresponding to a purported ancestral salmonid genome duplication we find fewer gene duplicates and less transposon activity than expected. Conversely, we find greater gene duplication and greater transposon activity at sequence divergences corresponding to salmonid speciation events and more recent times. Given that an ancestral salmonid genome duplication is assumed in thousands of independent studies and is considered a fundamental tenant in our current understanding of salmonid genetics, physiology and

biology, these two puzzling observations need to be carefully examined.

Genomic and EST sequences have also provided key insights into the nature of the genome duplication in salmonids. As a result, we now have a much better understanding of the enormous complexity and reorganization that occurred in salmonid genomes during the rediploization process (e.g., extensive segmental duplication, transposons/repeat element expansion). We also have a much better appreciation of its impact on the generation of new species and their variation in response to environmental conditions, pathogens and disease.

References

Allendorf FW, Thorgaard GH. Tetraploidy and the evolution of salmonid fishes. *In*: Turner, BJ (ed). *Evolutionary Genetics of Fishes*. Plenum Press, New York. 1984; 55–93.

Brown, GD. An analysis of salmonid RNA sequences and implications for salmonid evolution. University of Victoria, PhD Thesis. [Online] http://hdl.handle.net/1828/331 [Accessed 2008].

cGRASP website; www.uvic.ca/grasp & grasp.mbb.sfu.ca [Accessed 2008].

Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. *Mol. Biol. Evol.* 2006; **23**: 121–136.

de Boer JG, Yazawa R, Davidson WS, Koop BF. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics* 2007; **8**: 422.

Donoghue PCJ, Purnell MA. Genome duplication, extinction and vertebrate evolution. *TREE* 2006; **20**: 312–319.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. *Gene*tics 1999; **151**: 1531–1545.

Gregory, TR. Animal Genome Size Database. http://www.genomesize.com [Accessed 2005].

Hoegg S, Brinkmann H, Taylor JS, Meyer A. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* 2004; **59**: 190–203.

Hoegg S, Meyer A. Hox clusters as models for vertebrate genome evolution. *Trends in Genetics* 2005; **21**: 421–424.

Hughes AL. *Adaptive Evolution of Genes and Genomes*. Oxford Univ. Press. 1999.

Hughes, AL. Gene duplication and the origin of novel proteins. *PNAS* 2005; **102**: 8791–8792.

Inoue JG, Miya M, Venkatesh B, Nishida M. The mitochondrial genomie of Indonesian coelacanth *Latimeria menadoensis* (Sarcopterygii: Coelacanthiformes) and divergence time estimation between the two coelacanths. *Gene* 2006; **349**: 227–235.

Ishiguro, NB, Miya M, Nishida M. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". *Molecular Phylogenetics and Evolution* 2003; **27**: 476–488.

Kazazian HH. Mobile elements: Drivers of genome evolution. *Science* 2004; **303**: 1626–1632.

Leder EH, Danzmann RG, Ferguson MM. Comparison of GNRH3 genes across salmonid genera. *Anim. Genet.* 2004; **35**: 126–129.

Li, W-H. Evolution of duplicate genes and pseudogenes. *In*: Nei M, Koehn RK (eds). *Evolution of Genes and Proteins*. Sinauer. 1983; 14–37.

Lukacs MF, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, Bakke HG, Phillips RB, Miller KM, Davidson WS, Koop BF. Genomic organization of duplicated Major Histocompatibility Complex class I regions in Atlantic salmon (*Salmo salar*). *BMC Genomics*. 2007; **8**: 25.

Lynch M, Force A. The origin of interspecific genomic incompatibility via gene duplication. *Am. Nat.* 2000; **156**: 590–605.

Melamed P, Chong KL, Johansen MV. Evidence for lateral gene transfer from salmonids to two *Schistosome* species. *Nat. Genet.* 2004; **36**: 786–787.

McKay SJ, Trautner J, Smith MJ, Koop BF, Devlin RH. Evolution of duplicated growth hormone genes in autotetraploid salmonid fishes. *Genome* 2004; **47**: 714–723.

Moghadam HK, Ferguson MM, Danzmann RG. Evidence for Hox gene duplication in rainbow trout (*Oncorhynchus mykiss*): a tetraploid model species. *J. Mol. Evol.* 2005; **61**: 804–818.

Nelson, JS. *Fishes of the World*. 4th ed. John Wiley & Sons, New York. 2006.

Ng SHS , Artieri, CG, Bosdet IE, Chiu R, Danzmann RG, Davidson WS, Ferguson MM, Fjell CD, Hoyheim B, Jones SJM, Jong PJ, Koop BF, Krzywinski MI, Lubieniecki K, Marra MA, Mitchell L, Mathewson C, Osoegawa K, Parisotto SE, Phillips RB, Rise ML, vonSchalburg KR, Schein JE, Shin H, Siddiqui A, Thorsen J, Wye N. Towards a physical map of the genome of Atlantic salmon, *Salmo salar*. *Genomics*, 2005; **86**: 396–404.

Novak AE, Jost MC, Lu Y, Taylor AD, Zakon HH, Ribera AB. Gene duplications and evolution of vertebrate voltage-gated sodium channels. *J. Mol. Evol.* 2006; **63**: 208–221.

Ohno S. *Evolution by Gene Duplication*. Springer-Verlag, New York, 1970.

Shiina T, Dijkstra JM, Shimizu S, Watanabe A, Yanagiya K, Kiryu I, Fujiwara A, Nishida-Umehara C, Kaba Y, Hirono I, Yoshiura Y, Aoki T, Inoko H, Kulski JK, Ototake M. Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (*Oncorhynchus mykiss*), a species with a presumably recent tetraploid ancestry. *Immunogenetics* 2005; **56**: 878–893.

Steinke D, Salzburger W, Braasch I, Meyer A. Many genes in fish have species-specific asymmetric rates of molecular evolution. *BMC Genomics* 2006; **7**: 20.

Sverdlov, ED. Retroviruses and primate evolution. *Bioessays* 2000; **22**: 161–171.

Taylor JS, Raes J. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* 2004; **38**: 615–643.

Thorgaard GH, Bailey GS, Williams D, Buhler DR, Kaattari SL, Ristow SS, Hansen JD, Winton JR, Bartholomew JL, Nagler JJ, Walsh PJ, Vijayan MM, Devlin RH, Hardy RW, Overturf KE, Young WP, Robison BD, Rexroad C, Palti Y. Status and opportunities for genomics research with rainbow trout. *Comp Biochem Physiol B Biochem Mol Biol.* 2002; **133**: 609–646.

Volff J-N. Genome evolution and biodiversity in teleost fish *Heredity* 2005; **94**: 280–294.

von Schalburg KR, Leong J, Cooper GA, Robb A, Beetz-Sargent MR, Lieph R, Holt RA, Moore R, Ewart KV, Driedzic WR, Ten Hallers BF, Zhu B, de Jong PJ, Davidson WS, Koop BF. Rainbow smelt (*Osmerus mordax*) genomic library and EST resources. *Mar. Biotechnol*. 2008a; **10**(5): 487–491.

von Schalburg KR, Cooper GA, Leong J, Robb A, Lieph R, Rise ML, Davidson WS, Koop BF. Expansion of the genomics research on Atlantic salmon (*Salmo salar* L.) project (GRASP) microarray tools. *J. Fish Biology* 2008b; **72**(9): 2051–2070.

Wadman SA, Clark KJ, Hackett PB. Fishing for answers with transposons. *Mar. Biotech.* 2005; **7**: 135–141.

Yamanoue Y, Miya M, Inoue JG, Matsuura K, Nishida M. The mitochondrial genome of spotted green pufferfish *Tetraodon nigroviridis* (Teleostei: Tetraodontiformes) and divergence time estimation among model organisms in fishes. *Genes Genet. Syst.* 2006; **81**: 29–39.

Yazawa R, Cooper GA, Beetz-Sargent M, Robb A, McKinnel L, Phillips RB, Davidson WS, Koop BF. Functional adaptive diversity of the Atlantic Salmon T-cell receptor gamma locus. *Molecular Immunology* 2008; **45**: 2150–2157.