Differential Evolution of the 13 Atlantic Salmon Hox Clusters

Sutada Mungpakdee,* Hee-Chan Seo,*† Anna Rita Angotzi,* Xianjun Dong,‡ Altuna Akalin,‡ and Daniel Chourrout*

*Sars International Centre for Marine Molecular Biology, University of Bergen, Thormøhlensgt. 55, Bergen, Norway; †Department of Molecular Biology, University of Bergen, Thormøhlensgt. 55, Bergen, Norway; and ‡Computational Biology Unit, University of Bergen, Thormøhlensgt. 55, Bergen, Norway

Hox cluster organization represents a valuable marker to study the effects of recent genome duplication in salmonid fish (25–100 Mya). Using polymerase chain reaction amplification of cDNAs, BAC library screening, and genome walking, we reconstructed 13 Hox clusters in the Atlantic salmon containing 118 Hox genes including 8 pseudogenes. Hox paralogs resulting from the genome duplication preceding the radiation of ray-finned fish have been much better preserved in salmon than in other model teleosts. The last genome duplication in the salmon lineage has been followed by the loss of 1 of the 4 HoxA clusters. Four rounds of genome duplication after the vertebrate ancestor salmon Hox clusters display the main organizational features of vertebrate Hox clusters, with Hox genes exclusively that are densely packed in the same orientation. Recently, duplicated Hox clusters have engaged a process of divergence, with several cases of pseudogenization or asymmetrical evolution of Hox gene duplicates, and a marked erosion of identity in noncoding sequences. Strikingly, the level of divergence attained strongly depends on the Hox cluster pairs rather than on the Hox genes within each cluster. It is particularly high between both HoxBb clusters and both HoxDa clusters, whereas both HoxBa clusters remained virtually identical. Positive selection on the Hox protein–coding sequences could not be detected.

Introduction

Genome duplications have occurred several times during the evolution of chordates (Holland et al. 1994; Hoegg et al. 2004). After 2 rounds of genome duplication predating the origin of vertebrates, a fish-specific third round (3R) genome duplication has been uncovered using a combination of synteny data and genome-wide phylogenetic analysis of duplicated genes in zebrafish, tetraodon, medaka, and other teleosts (Jaillon et al. 2004; Naruse et al. 2004; Woods et al. 2005). This “fish genome duplication” occurred approximately 320–350 Mya (Christoffels et al. 2004). Salmonid fish (e.g., rainbow trout and Atlantic salmon) have undergone a fourth round of genome duplication much more recently (~25–100 Mya) (Allendorf et al. 1984). Gene segregation studies and the presence of meiotic multivalents in males showed that the diploidization of salmonid genomes is not yet completed (Allendorf and Thorgaard 1984; Phillips et al. 2006). Thus salmonids together with other model teleosts provide an interesting opportunity to evaluate the effects of genome duplications at different timescales.

Several models predicting the fate of duplicated genes have been proposed since Ohno’s “classical model” (Ohno 1970), according to which 1 of 2 gene duplicates is more prone to accumulate mutations leading to loss or gain of functions as long as the other duplicate copy keeps the original function. As deleterious mutations are more frequent than beneficial ones, most duplicates get lost or silenced through a process named nonfunctionalization. Far less frequently, 1 copy can acquire a new function (neofunctionalization) emerging under positive Darwinian selection. Hughes (1994) and Force et al. (1999) proposed the “subfunctionalization model” in which degenerative mutations accumulating in both duplicates lead them to retain partial and complementary subfunctions. A third model called “duplication–degeneration–complementation” predicts that degenerative mutations in the regulatory elements of duplicates increase the probability of their long-term preservation (Force et al. 1999). Recent studies (He and Zhang 2005; Rastogi and Liberles 2005) yielded the “subneofunctionalization” model, in which rapid subfunctionalization is often accompanied by prolonged and substantial rates of neofunctionalization, for a large proportion of gene duplicates.

Hox genes encode homeodomain-containing transcription factors which control pattern formation along the anterior–posterior of body axis of various animals, including arthropods and vertebrates. Hox clusters contain up to 14 Hox genes (Powers and Amemiya 2004). The cluster organization may aim at ensuring the temporal coordination of Hox gene expression, at least in vertebrates (Kmita and Duboule 2003). In vertebrates, Hox clusters have evolved toward a higher level of compaction and organization, through a so-called consolidation process (Duboule 2007), which was possibly favored by genome duplications. Together with the rest of the genome, the Hox cluster was duplicated twice in early vertebrates, leading to 4 Hox clusters (HoxA–D) which after losing a variable number of genes still contained a total of 39 genes in amniotes (Schughart et al. 1989; Ruddle et al. 1994). The third round of duplication in ray-finned fish eventually led to 7 Hox clusters, after the loss of either 1 HoxC cluster (tetraodon, fugu, and medaka) or 1 HoxD cluster (zebrafish) and 45–49 genes after the loss of numerous gene duplicates. The fourth salmonid-specific genome duplication has received strong support after the recent cloning and mapping of 25 Hox genes from Atlantic salmon and rainbow trout, which suggested the preservation of 4 HoxA, 4 HoxB, 4 HoxC, and at least 2 HoxD clusters (Moghadam et al. 2005). A larger number of Hox clusters in teleosts and the differential evolution of their contents has been seen as a possible explanation for their considerable anatomical diversity, but this speculation meets clear contradictions (Crow et al. 2006).

Positive Darwinian selection has affected the Hox-coding sequences, at least for the HoxB5, HoxB6, Hox7, and
HoxA11 vertebrate paralogs (Van de Peer et al. 2001; Fares et al. 2003; Crow et al. 2006), and more particularly regions of the homeodomain involved in protein–protein interactions (Lynch et al. 2006).

In the present study, systematic cloning and physical mapping of Atlantic salmon Hox genes and their genomic environments allowed us to reconstruct 13 Hox clusters postdating the fourth round of genome duplication. The evolutionary rates of recently duplicated Hox genes have been estimated and compared with those of Hox genes in various vertebrate outgroups. As the salmonid-specific genome duplication is fairly recent, the estimates of these rates are expected to be relatively small and more accurate than after long postduplication times, during which nucleotide and amino acid substitution rates can attain saturation. Overall, we found that fewer Hox paralogs have been lost in the salmon lineage than in those of other teleosts. However, a significant divergence has appeared between the recent duplicates of Hox cluster and individual Hox genes, in coding and noncoding sequences. The process of divergence, which is sometimes asymmetrical but does not necessarily involve positive selection, has affected the Hox clusters with a variable intensity.

Materials and Methods
Gene Cloning, Sequencing, and Physical Mapping

Partial homeobox sequences were cloned by using degenerate primers based on conserved homeodomain sequences on genomic DNA extracted from blood of a single salmon male. Putative Hox gene homeobox sequences were extended by polymerase chain reaction (PCR) using genomic libraries based on the Universal GenomeWalker kit (Clontech, Hampshire, UK). Hox gene–specific probes were used to screen a BAC library (18× coverage; average insert size, 188 kb), which was made available by the Salmon Genome Project (a Norwegian consortium). Probes (300- to 600-bp long) were prepared by incorporating digoxigenin-11-deoxyuridine triphosphate (Roche, Mannheim, Germany) using the PCR method. Information from the fingerprinting analysis of BAC inserts (from the Canadian consortium Genome Research on All Salmon Project [GRASP]) helped the identification of clones containing entire or almost entire Hox clusters. Clones were sequenced commercially (MWG-Biotech, Ebersberg, Germany) by the shotgun method. Genes missing in some clusters and all genes of HoxA clusters were linked to previously identified genes using genome walking and long-range PCR (Bio-X-ACT, Bioline, Luckenwalde, Germany). Primers were designed in the regions that have diverged enough to distinguish recent salmon-specific duplicates. PCR amplification of cDNAs from 170- to 250-day-old embryos allowed us to check for expression of each Hox gene and to deduce their intron–exon organization.

Gene Annotation

Protein-coding genes in and immediately outside Hox clusters were predicted by combining 3 types of information: (1) cDNA and EST sequences, (2) BlastX alignment of Atlantic salmon sequences against the nr protein database at the National Center for Biological Information, and (3) ab initio gene prediction tools GeneMark (http://exon.gate- ch.edu/GeneMark/) and GENSCAN (http://genes.mit.edu/ GENSCAN.html). In a pair of recently duplicated salmon Hox clusters, the gene richer and the gene poorer clusters were named α and β, respectively. We also calculated the %GC content and %GC at the third position of codons of genes by using the program CodonW (http://mobyle.pasteur.fr/gibin/mobyportal/portal.py?form=codonw). Most genes in α clusters have slightly higher %GC and %GC3. Micro-RNAs in the Hox clusters were identified by BlastN search against miRBase (http://microrna.sanger.ac.uk/sequences).

Analysis of Coding Sequence Evolution

Codon alignment was generated from a multiple sequence alignment of predicted amino acid sequences and the corresponding DNA sequences by PAL2NAL program (http://coot.embl.de/pal2nal). Synonymous and nonsynonymous substitutions of each gene pair were also determined by this program, which is based on the codon model program in PAML. HYPHY (Pond et al. 2005) was used to evaluate the variability of Ks/Ka (ω) ratio or rates of nonsynonymous substitution among tree branches. The “ParalogSelectionComparison” script used for detecting selective strengths on the 2 paralogs is available in supplementary text (Supplementary Material online). Global pairwise alignment of large genomic sequences was performed using AVID in VISTA web interface software (http://genome.lbl.gov/vista/mvista/submit.shtml).

Results
Identification and Reconstruction of Salmon Hox Clusters

PCR amplification from genomic DNA of 1 Atlantic salmon male using degenerate Hox gene–specific primers yielded 95 homeobox sequences, which were classified into 81 Hox and 14 non-Hox genes. More than 1,500 PCR clones were sequenced. Sequences flanking homeodomains and Hox genes were obtained through genome walking approach and used to prepare probes for BAC library screening. We thus established 10 linkage groups representing a minimum of 10 distinct Hox clusters, including 4 HoxA clusters (using HoxB6a, B3a, B1a, and B1b probes), 4 HoxC clusters (using HoxC12a, C12b, C8b, and C5a probes), and 2 HoxDa clusters (using HoxD9a and D4a probes). No HoxDb clusters were revealed through PCR survey or cross-hybridization in BAC library screening. In addition to these 10 linkage groups, we found 5 singleton BAC clones which hybridized with 7 Hox probes (Hox-A1a, A2a, A2b, A9a, A11a, A13a, and A13b), but these contained small and unstable inserts (data not shown). Several BAC inserts were chosen for full sequencing by the shotgun method, and most of them were shown to contain entire Hox clusters. The HoxBaβ cluster was obtained from 2 BAC clones, one containing HoxB1a to HoxB9a and the other containing HoxB10a and HoxB13a. HoxC4βx was linked to the HoxCbα cluster through genome walking from the BAC end. The HoxCaα cluster was reconstructed from
BAC sequences (from HoxC13a to HoxC6a) and genes isolated by genome walking (HoxC5a, C4a, C3a, and C1a). Exon1 of HoxC3a and HoxC1a are still incomplete. HoxC1a appears to be a pseudogene due to a premature stop codon in the exon2. The entire HoxA clusters could not be cloned but were reconstructed from each individual HoxA gene by genome walking. Two paralogs were found for each HoxAa gene, indicative of 2 HoxAa clusters. In contrast, only 1 paralog was detected for each of 5 HoxAb genes, suggesting a single HoxAb cluster. Linkage of all genes within 3 HoxA clusters, HoxAa, HoxAb, and Hox-Ab, could indeed be established using long-range PCR amplification and end sequencing of resulting fragments (fig. 1). Approximate sizes of the HoxAa, Aaβ, and Abβ clusters are 91, 89, and 40 kb, respectively.

Hox Cluster Contents

In total, we identified 118 Hox genes and allocated them to 13 clusters named HoxAaα, Aaβ, Abα, Abβ, Baα, Baβ, Caα, Caβ, Cbα, Cbβ, Daα, and Daβ (fig. 1). Eight of these Hox genes (HoxB8bβ, B3bβ, B1bβ, C1α, C3aβ, C10bβ, D13aβ, and D1aβ) appeared to be pseudogenes. In these, remnants of the Hox-coding sequence, especially stretches of nucleotides encoding the third helix of the homeodomain (including WFQNRR), were easily identified, whereas they are generally undetectable in intervals between active Hox genes of other teleosts. Most salmon pseudogenes have nucleotide insertions or deletions in the exons, but keep the same intron–exon boundaries as their paralogs. The presence of only 1 nucleotide transversion in HoxC10bβ, which causes a premature stop codon in the homeodomain, suggests a particularly recent inactivation. Results from systematic cDNA cloning showed that at least 108 Hox genes (including 3 pseudogenes) were expressed during early development (170 and 250 dd). In contrast, only 5 Hox genes were recorded in the salmon EST database (http://www.salmongenome.no/cgi-bin/sgp.cgi#Blast), which was mainly generated from adult tissues. Alignment of cDNA sequences with genomic sequences allowed us to identify the number of original alternative splice sites (data not shown), including for transcribed pseudogenes.

Without counting the most recent Hox duplicates, Atlantic salmon displays far more Hox genes than the other teleosts, with at least 59 genes retained from the fish radiation versus only 45–49 genes in zebrafish (Amores et al. 2004), medaka (Kurosawa et al. 2006), or fugu (Aparicio et al. 2002): the HoxA clusters of salmon contain as many Hox paralogs as the HoxA clusters of medaka, its HoxB and HoxC clusters contain as many paralogs as those of zebrafish, and its HoxDa clusters contain 1 more paralog than the zebrafish hoxda cluster and 2 more paralogs than the medaka and fugu HoxDa cluster. Characterization of zebrafish/medaka/fugu Hox clusters showed that gene loss after the ancient fish genome duplication predominantly affected one of each cluster duplicate (defined as “b”) either entirely lost or with around half of the genes lost). Gene loss and pseudogenization after the last genome duplication predating...
salmonoids also lost 1 of the 2 cluster duplicates (defined as α) essentially intact. The gene content of the other cluster (β) was altered at very variable levels, little for \( \text{HoxAα}/\text{HoxAβ} \) or \( \text{HoxCα}/\text{HoxCβ} \), severely for \( \text{HoxBβ}/\text{HoxDβ} \), and entirely for \( \text{HoxAβ} \). It is noteworthy, as already noted in a comparison of mouse and zebrafish \text{Hox} \) cluster gene contents (Duboule 2007), that the posterior genes are particularly well retained in the salmon \text{Hox} \) clusters. The only exceptions are both \text{HoxBb} \) clusters which, like the zebrafish \( \text{hoxbb} \) cluster, have lost all of them. In the 11 other \text{Hox} \) clusters of salmon, \( \text{Hox13} \) is present (though \( \text{HoxDβ}_1/\text{HoxEβ} \) seems to be a pseudogene) and associated with 2 (in 2 clusters), 3 (in 4 clusters), or 4 (in 5 clusters) other posterior \text{Hox} \) genes.

In the available \text{Hox} \) cluster sequences, we identified 7 candidate \text{mir-10} \) microRNA genes located between \( \text{Hox5} \) and \( \text{Hox9} \) in salmon. The \( \text{mir-10} \) cluster of zebrafish and salmon contains 2 (in 2 clusters), 3 (in 4 clusters), and 4 (in 5 clusters) genes. In all gnathostomes, \text{mir-10} \) was identified in \( \text{HoxA}, \text{HoxC}, \) and \( \text{HoxD} \) clusters, whereas \text{mir-196} \) appeared in salmon and zebrafish, but not in the mouse or fugu. The \text{mir-196} \) is found neither in the 2 salmon \( \text{HoxBb} \) clusters nor in the fugu \( \text{HoxBb} \) cluster; it has to be noted, though, that these 3 clusters have lost all their posterior \text{Hox} \) genes including \( \text{Hox9} \) and \( \text{Hox10} \). In all gnathostomes, \text{mir-10} \) was identified in \( \text{HoxB}, \text{HoxC}, \) and \( \text{HoxD} \) clusters, whereas \text{mir-196} \) appeared in \( \text{HoxA}, \text{HoxB}, \) and \( \text{HoxC} \) clusters. Gnathostomes lost 1 \text{mir-10} \) in the \( \text{HoxA} \) cluster and 1 \text{mir-196} \) in the \( \text{HoxD} \) cluster after 2 rounds of initial duplication. After the third round of duplication in ray-finned fish, \( \text{6 mir-10s} \) and \( \text{5 mir-196s} \) were identified. The \text{mir-196} \) in the \( \text{HoxBb} \) cluster may have been lost together with the posterior \text{Hox} \) genes. Due to a lack of upstream sequence of \( \text{HoxAβa}, \text{HoxBβa}, \) and \( \text{HoxCβa} \), we ignore whether \text{mir-10} \) in \( \text{HoxCα} \) cluster and \text{mir-196} \) in \( \text{HoxAβ} \) and \( \text{HoxBα} \) clusters are still present. However, we tend to believe so as their neighbor and putative target genes were kept in these clusters. The putative target site of \text{mir-196} \) is found in the \( \text{3' untranslated region} \) of \( \text{HoxB8α} \), which is identical to that of \( \text{HoxB8aβ} \).

Detailed analysis of gene sequences surrounding the salmon \text{Hox} \) clusters was performed to evaluate the degree of synteny conservation with other vertebrate and fish species (supplementary text and figure S1, Supplementary Material online). Numerous sequences related to known DNA transposons and retrotransposons were also detected. Most of them were located outside the \( \text{Hox} \) cluster regions. Transposons within the \( \text{Hox} \) clusters were located where genes were lost.

**Gene Density in Hox Clusters**

We performed a detailed comparison of \( \text{Hox} \) cluster total lengths and physical distances between adjacent \( \text{Hox} \) genes of salmon, zebrafish, and mouse (fig. 2; detailed data not shown). The \( \text{HoxA} \) clusters of salmon are larger than those of the zebrafish, which have lost several \( \text{Hox} \) genes, and slightly smaller than the mouse \( \text{HoxA} \) cluster. The absence of \( \text{HoxA6} \) in salmon would be sufficient to explain the latter size difference. Major gene losses have affected the \( \text{HoxBb} \) clusters of both fish species, and their total size is also considerably reduced compared with the \( \text{HoxB} \) cluster of the mouse. The salmon \( \text{HoxBβ} \) cluster is slightly larger than the zebrafish \( \text{hoxbb} \) cluster, but it contains one more gene. The salmon \( \text{HoxBβ} \) cluster is almost twice larger, due to expanded space between \( \text{HoxB3}ββ \) and \( \text{HoxB5}ββ \) (note that \( \text{HoxB1β}ββ \) and \( \text{HoxB3β} \) are pseudogenes). The \( \text{HoxBa} \) clusters of zebrafish and salmon contain 11 \( \text{Hox} \) genes, one more than the mouse \( \text{HoxB} \) cluster which has lost \( \text{HoxB10} \) in addition to \( \text{HoxB11} \) and \( \text{HoxB12} \). The larger total size of the mouse \( \text{HoxB} \) cluster is essentially due to the exceptionally large distance between the 2 remaining
posterior genes (HoxB13 and HoxB9), whereas the level of compaction of all clusters between HoxB1 and HoxB9 is virtually equivalent in all 3 species. HoxCa clusters of zebrafish and salmon have more genes than the mouse HoxC cluster due to the persistence of anterior genes. In the region which can be compared between the 3 species (Hox13 to Hox6), the zebrafish cluster is the smallest, and the salmon clusters are also slightly more compact than the mouse HoxC cluster. In the region of the most anterior genes, the HoxCaβ cluster of salmon is slightly more expanded than the zebrafish hoxca cluster (physical distances are unknown for the HoxCaα cluster), but again, it may be important to note that HoxC3αβ seems to be a pseudogene. The zebrafish hoxcb cluster has lost many genes, and the remaining 4 genes are located very close to each other. Interestingly, both salmon HoxCb clusters have retained all genes identified in the mouse HoxC cluster (from HoxC4 to HoxC13), and the total cluster size is very similar between the mouse and the salmon. Finally, the remaining HoxD clusters of zebrafish and salmon (all HoxDa) are significantly smaller than the mouse HoxD cluster. This may be correlated with the absence of HoxD8 in both fish species and HoxD1 in zebrafish. However, with one gene more than the zebrafish hoxda cluster and one gene less than the mouse HoxD cluster, both HoxDa clusters of salmon appear more compact.

Evolutionary Rates of Salmon Hox Genes and Cluster Divergence

We first evaluated the level of sequence identity between each recent Hox gene duplicate of salmon, and we compared it with the identity level of the same paralogs among distinct teleosts (zebrafish, medaka, fugu, and salmon). Overall the identity level greatly varied from gene to gene and from cluster to cluster. Rank correlations (calculated for Hox clusters which contain more genes) (fig. 3 and supplementary table S1, Supplementary Material online) indicated that the more dissimilar the salmon duplicates were from each other, the more divergent they also were from their paralogs in other teleosts, and the more these paralogs in other teleosts diverged from each other. The Hox paralogs diverging faster after salmonid genome duplication and after teleost speciation were among those preferentially lost or pseudogenized in some teleost lineages. To evaluate how Hox genes evolved in the salmon lineage, we estimated the synonymous (Ks) and nonsynonymous (Ka) substitution rates of each recently duplicated gene pair (α and β; pseudogenes excluded). Average $K_a$ and $K_s$ were 0.04 ± 0.02 and 0.15 ± 0.07 for recent duplicates, respectively, and 0.22 ± 0.15 and 2.11 ± 2.41 for ancient duplicates, respectively. Average $K_a$ and $K_s$ of other genes flanking the Hox clusters were also within the same
ranges. Important variations of $K_a$, $K_s$, and their ratio $K_a/K_s$ were observed between distinct Hox genes, including within a given cluster (fig. 4). However, comparisons with standard t-test showed that genes of HoxAa clusters have significantly higher average $K_a$ values than those of HoxDa and HoxCa clusters and that HoxBa clusters have significantly lower $K_a$ and $K_s$ values than the others. Global pairwise alignments between the entire Hox cluster duplicates showed an extremely strong conservation of both coding and noncoding sequences of both HoxBa clusters (fig. 5A), whereas other pairs of Hox clusters resulting from the same recent duplication were markedly divergent in the Hox genes and in the noncoding sequences (fig. 5B–D).

To get insight into the mechanisms leading to the retention of duplicated genes, we evaluated selective
pressures through branch-specific test by comparisons 1) between the 1-ratio model with no change in ω ratio after duplication and 2-ratio model where increased ω ratio of paralogs was allowed and 2) between the 2-ratio model with no difference in Ka of paralogs and the free-ratio model with different rates of Ka in each paralog. In the latter test, the paralog with significantly higher Ka value also had higher ω ratio than the other paralog and than the nonduplicated ortholog. As genome data for fish species most closely related to salmonids are lacking, zebrafish was chosen as outgroup for most Hox genes and medaka was used for HoxAa genes which were massively lost in zebrafish. Due to the possibility of substitution rate saturation since ancient gene duplication, amino acid substitution rates of salmon genes and of their nonduplicated counterparts of the outgroups were compared using maximum likelihood relative rate test as implemented in HYPHY. We also evaluated the asymmetric divergence of duplicated genes using *Xenopus tropicalis* or human as outgroups (human sequences were used when *X. tropicalis* sequences were not available). It would have been more appropriate to use bowfin fish (*Amia calva*) and bichir (*Polypterus* sp.), but only a few Hox genes have been cloned in these species. All recently duplicated salmon Hox genes except HoxC8a (supplementary table S2, Supplementary Material online) showed an average ω ratio significantly higher than those of their nonduplicated orthologs, suggesting a relaxation of purifying selection following the latest genome duplication. The increased rate of evolution after duplication was not the consequence of increased mutation rates. Significant asymmetric divergence between recent duplicates of salmon Hox genes was detected for 6 out of 41 of gene pairs (15%), which all belonged to HoxDa and HoxC clusters (fig. 6 and supplementary table S3, Supplementary Material online). Comparison of nonsynonymous substitution rates for genes of HoxA and HoxC clusters also revealed an almost systematic increase of evolutionary rate after the ancient fish genome duplication (Ka/salmon/Ka/outgroup > 1; supplementary table S4, Supplementary Material online). In comparison, asymmetric divergence after the ancient fish genome duplication for genes of clusters HoxA/HoxB/HoxC were detected in 8 out of 19 gene pair comparisons of Ka (fig. 7 and supplementary table S5, Supplementary Material online). Although these calculations were based on not more than 60 duplicated pairs, they are consistent with larger scale studies (Kondrashov et al. 2002). Altogether, our results support that asymmetrical divergence is a gradual process that occurs at an early stage of postduplication divergence. After recent duplication, genes of a given cluster evolved at different rates (figs. 6 and 7). However, all β paralogs of HoxDa cluster evolved faster than the α paralogs and 2 of them showed a significantly higher rate of nonsynonymous substitution. After ancient duplication, all “b” paralogs of HoxA and HoxB clusters evolved faster than their “a” paralogs.

Asymmetric evolution of gene duplicates can be explained by at least 2 models (Kellis et al. 2004; He and Zhang 2005). To evaluate whether positive selection or asymmetric subfunctionalization acted on the faster evolving paralogs in the salmon lineage, we examined asymmetrically evolving gene pairs further (6 and 8 Hox gene pairs from recent and ancient fish genome duplication, respectively) using the ParalogSelectionComparison script, implemented in the HYPHY package. The analysis is based on a likelihood ratio test (LRT) on 3 sequences (2 paralogs...
and 1 outgroup from different species). The selected outgroup was assumed to have invariant $K_{a}/K_{s}$ ratio from site to site. Results revealed that none of the tested gene pairs are significantly different from each other in the overall proportion of positively selected sites. However, the weakness of LRT method for the detection of positive selection, when only 3 sequences are used, is expected. The limitation of $K_{a}/K_{s}$ ratios for the detection of adaptive evolution is due to the fact that only a small fraction of the sites are under positive selection, whereas the majority of them is under strong purifying selection. It is noteworthy that in our data set, higher values of $K_{a}$ are generally associated with a parallel increase of $K_{s}$, thus $\omega$ ratios were consistently $<1$.

**Discussion and Conclusions**

In this work, we presented evidence for 13 Hox clusters in Atlantic salmon. Comparison of the gene contents of these clusters with the vertebrate ABCD orthologs strengthens the 4R hypothesis, with a recent salmonid-specific genome duplication. Strong support for the 4R hypothesis was provided in a previous report that assigned 25 Hox genes from Atlantic salmon and rainbow trout to 14 genomic loci (Moghadam et al. 2005). Although 2 HoxA2b loci were identified and allocated to distinct linkage groups in rainbow trout, we could identify only 1 HoxAb cluster in Atlantic salmon. Two HoxAb loci had also been proposed for the Atlantic salmon based on the segregation of PCR fragments, but these displayed identical nucleotide sequences. A difference in Hox clusters between Atlantic salmon and rainbow trout would not be very surprising as they belong to distinct genera, which underwent considerable karyotype rearrangements during their divergence. The karyotype of Atlantic salmon indeed displays an exceptionally small number of chromosome arms (NF = 74) compared with most other salmonids including the rainbow trout (NF =100). After important efforts of cDNA and genomic cloning, we also propose the absence of HoxDb clusters in Atlantic salmon. As the hoxdb locus of zebrafish has been reduced to the mir-10b gene (Woltering and Durston 2006), we assume that the HoxDb cluster was already lost in the common ancestor of salmon and zebrafish. This observation, together with the recognition of 2 HoxC clusters in salmon, is not unexpected as Salmoniformes (Protacanthopterygii) are considered phylogenetically closer to Cypriniformes (Ostariophysi, which include zebrafish) than to Beloniformes and Tetraodontiformes (Acanthopterygii), which include medaka and pufferfish) (Nelson 1994; Inoue et al. 2003). However, important Hox gene loss occurred in the zebrafish lineage after its split from the salmon lineage. Gene loss in the salmon lineage essentially consisted of the elimination and pseudogenization of several recent duplicates and loss of 1 HoxAb cluster. A less expected observation is that all Hox paralogs thus far recognized in one or the other of the well-studied teleosts (zebrafish, medaka, and pufferfish) are present in the salmon clusters. In other words, from the common ancestor of the 4 teleost species, fewer Hox genes have been lost in the salmon lineage. Our prediction for the Hox gene content of the common teleost ancestor is essentially unchanged after cloning the salmon Hox genes. This ancestor had 8 Hox clusters containing at least 62 post-3R paralogs (fig. 8) and 45 post-2R paralogs. Until new fish Hox complements will be uncovered, we can still assume that most losses of post-2R Hox paralogs occurred before the main radiations of jawed vertebrates.

A very striking pattern of Hox cluster evolution after the “ancient” fish genome duplication is the preferential loss of genes in 1 of the 2 Hox cluster duplicates, which invariably resulted in a gene-rich and a gene-poor duplicate (named “a” and “b,” respectively). A similar pattern seems
Asymmetric divergence between anciently duplicated gene pairs has also been reported in yeast, *Drosophila*, and *Cae- nohabditis elegans* in a genome-wide study (Conant and Wagner 2003; Kellis et al. 2004). Recent work on mammalian genomes concluded that genome colocalization might determine similar $K_a$ and $K_s$ evolutionary rates via common meiotic recombination features (Lercher et al. 2004). When the recombination features of regions containing each salmon *Hox* cluster will be known, it will be interesting to see whether the extreme similarity of the 2 *HoxBa* clusters can reflect their persistent interaction during male meiosis.

In most salmon *Hox* clusters, duplications probably relaxed the constraints on gene evolution before purifying selection started to inactivate a subset of the *Hox* gene duplicates. The 2 *HoxBa* clusters in salmon have evolved much slower than the others, both in the number of their genes and in their sequences, for unclear reasons. Positive selection, which allows new functions to appear in duplicated genes, is not detected in salmon *Hox* clusters from measurements of $K_a/K_s$ ratios. It may have been retarded by a globally slower evolution, the postduplication time has been too short, or our tools for revealing positive selection are simply not potent enough. In this respect, we must stress that no evidence for completely novel functions in post-3R *Hox* paralogs of fish have been brought thus far, even though there are indications subfunctionalization in zebrafish *Hox* genes.

Finally, a question that is not directly addressed in this report is how the function of *Hox* genes has evolved in the salmon lineage. We have obtained expression patterns of several anterior *Hox* genes using whole mount in situ hybridization on early embryos, which clearly suggest as expected that *Hox* genes pattern the anteroposterior axis of salmon (Munghakdee S, Seo H-C, Chourrout D, unpublished results) as they do in all other vertebrates examined thus far. We have not been able to observe obvious differences of spatial expression between these salmon genes and the same zebrafish paralogs or between the recently duplicated *Hox* genes of salmon. However, the study needs to proceed further with an increased resolution in space and time before we will be able to answer whether recent duplications have been followed by a functional specialization of the duplicates. Interestingly, the global organization of the *Hox* clusters themselves appears to be strongly maintained in salmon, as already observed and thoroughly commented for the zebrafish in a comparison with mouse *Hox* clusters. The salmon *Hox* clusters have indeed conserved the typical and presumably highly constrained vertebrate-like organization, with a high density of *Hox* genes all in the same orientation, exclusion of non-*Hox* genes, and a marked persistence of posterior genes including *Hox13* in almost all clusters. The salmon does not make exception to the vertebrate rule, despite its highest number of *Hox* clusters reported for an animal and 2 extra rounds of genome duplication.

**Supplementary Material**

Supplementary text, tables S1–S5, and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).
Acknowledgments

The authors wish to thank Boris Lenhard for helpful comments on the manuscript and useful advices for the work progress. Bjorn Hoyheim and Jim Thorsen kindly provided the Atlantic salmon BAC library and BAC clones. Physical mapping results were produced by GRASP and kindly provided by William S. Davidson. Marit Flo Jensen provided genomic DNA, and Ragnhild Eskeland prepared the GenomeWalker libraries. The ParalogSelectionComparison script implemented in HYPHY was kindly provided by Sergej L. Kosakovsky Pond. This research was supported by the Salmon Genome Project from the Norwegian Research Council. The sequences of Atlantic salmon reported in this paper have been deposited in the GenBank database (accession number EF695248–EF695352, EU025680–EU025719, and EU221640–221655). H.-C.S. and D.C. conceived the project and cosupervised its progress; S.M. carried out most experiments and analyses; A.R.A. collaborated through periodic advice and provision of biological material; X.D. and A.A. contributed to genomic alignments; and S.M. and D.C. wrote the manuscript with contribution from H.C.S.

Literature Cited


Woltering JM, Durston AJ. 2006. The zebrafish hoxDb cluster has been reduced to a single microRNA. Nat Genet. 6:601–602.


William Jeffery, Associate Editor

Accepted March 25, 2008