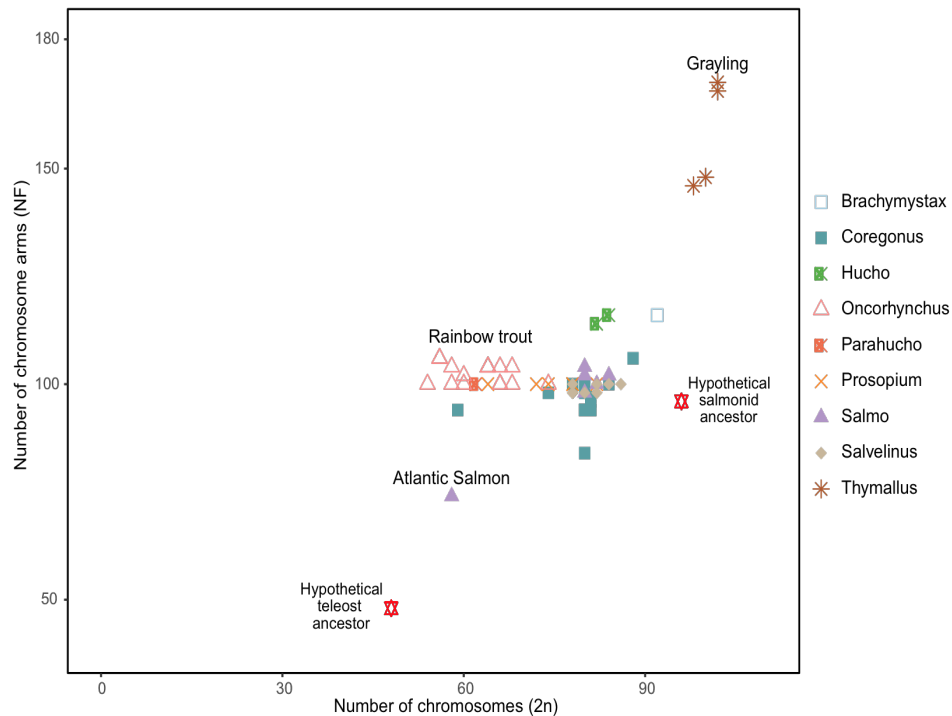# Supplementary figures and tables

Supplementary Figures

Figure S1: Karyotype evolution in salmonids. Chromosome number (2N) plotted against Number of chromosome arms (NF). Data from (Hartley et al. 1987). Based on the karyotype of most extant teleosts (48-50 acrocentric chromosomes), it is hypothesized that the salmonid ancestor had a karyotype of around 96-100 uni-armed chromosomes (NF 100). While most of the salmonids have a karyotype consisting of chromosome number of 52 to 102 and NF of 72-170, Atlantic salmon and grayling seem to be the exceptions on the opposite extremes. It has also been seen that the bi-armed metacentric chromosomes in grayling are much smaller than those in other salmonids. Thus, it has been hypothesized that, while most salmonids have reduced the chromosome number and retained NF close to the ancestral karyotype through translocations and fusions, the grayling karyotype has evolved through inversions (Phillips and Ráb 2007; Ocalewicz et al. 2013).
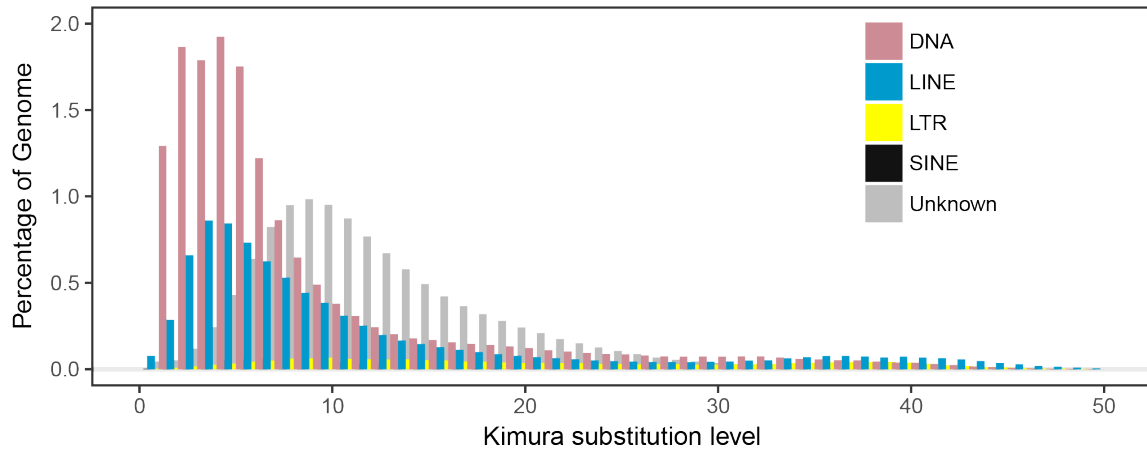
Figure S2: Repeat landscape of grayling genome based on Kimura distance. X-axis represents divergence from repeat consensus sequence and y-axis represents the proportion of the transposable element family in the genome (where LTR stands for long terminal repeats, LINE represents long interspersed nuclear elements and SINE stands for short interspersed nuclear elements).
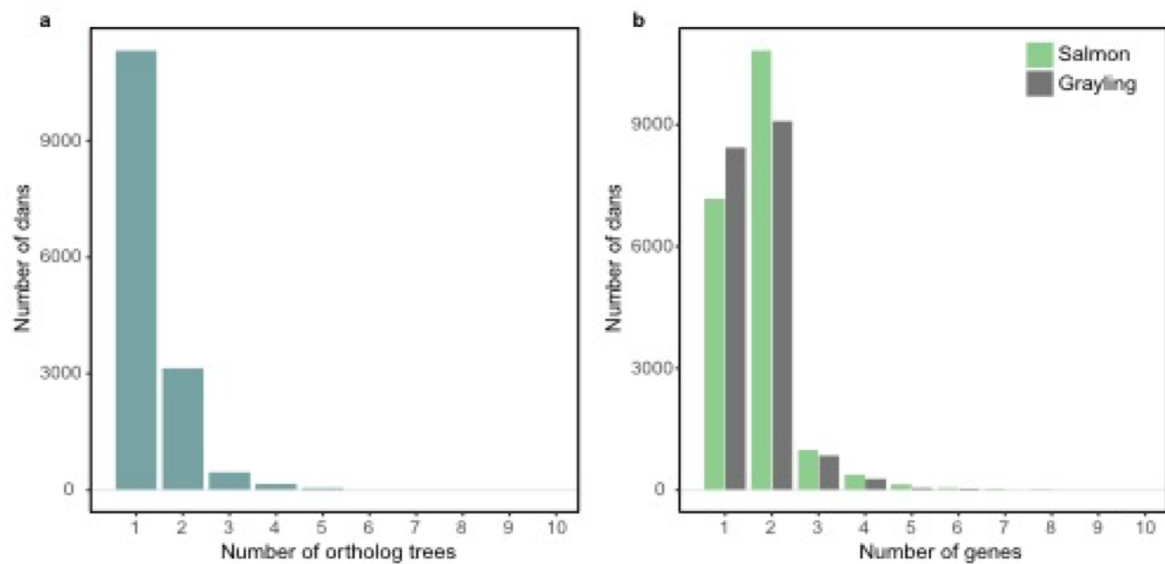

Figure S3: Orthologs in clans. a) Distribution of clans per ortholog tree. b) Number of Atlantic salmon and grayling genes per clan.
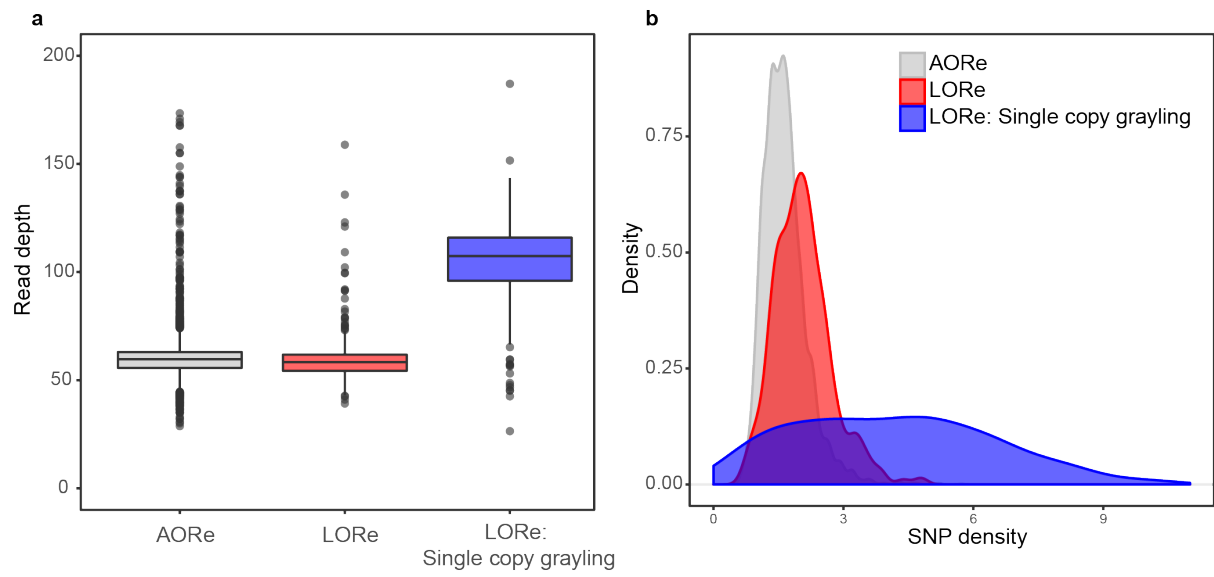
Figure S4: Distribution of (a) mapped read depth and (b) SNP density per Kb across all Ss4R duplicates in grayling grouped by the ohnolog resolution models.
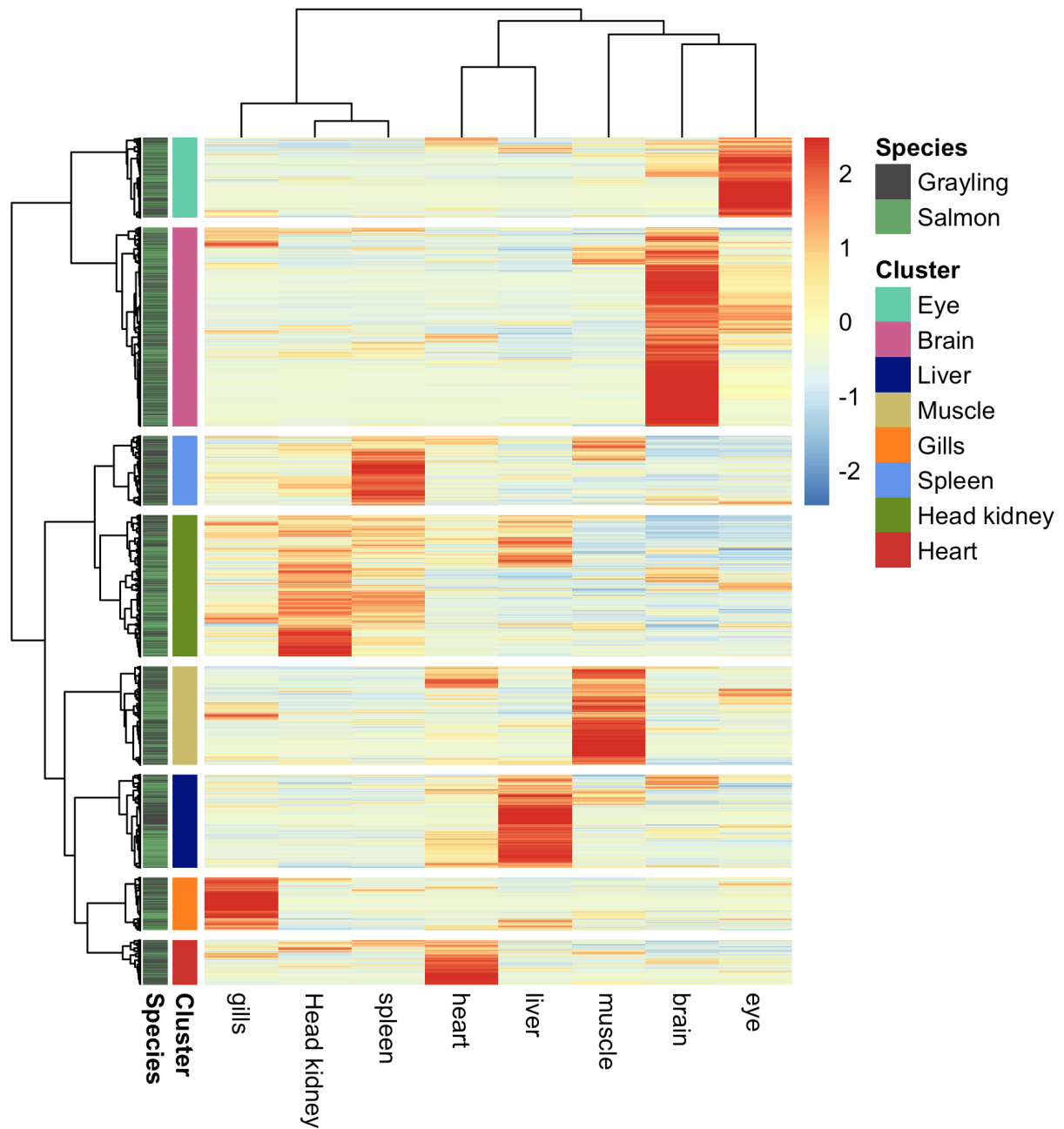
Figure S5: Heatmap of tissue expression clusters from Atlantic salmon and grayling. Tissue expression profile of ohnologs from Atlantic salmon and grayling using hierarchical clustering. The color scale of the heatmap corresponds to the relative abundance of the transcript across all the tissues within the two species. The first vertical bar ('Cluster') represents the 8 distinct 'tissue-specific' clusters. The 'Species' bar represents the respective species corresponding to the gene.
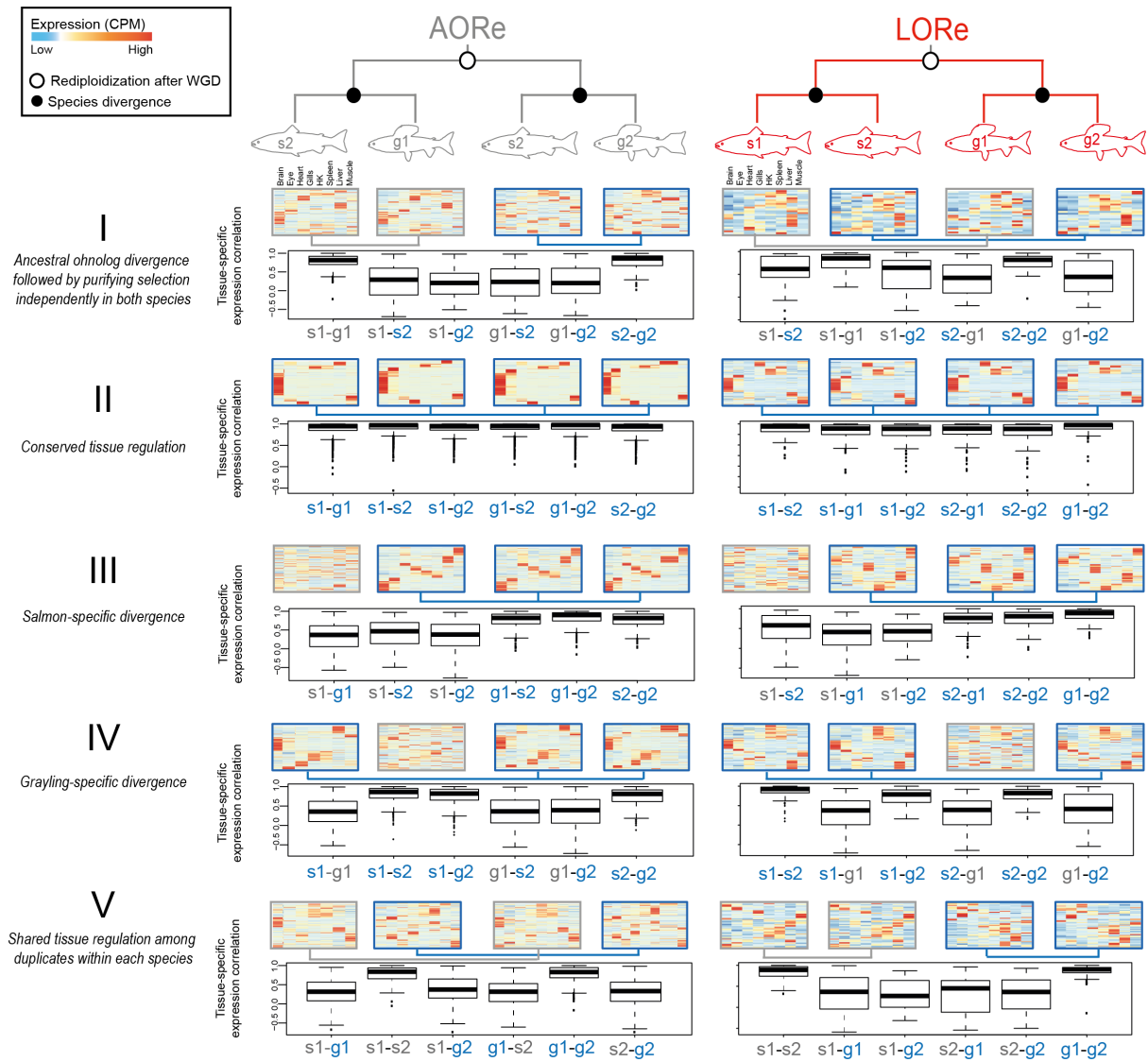
Figure S6: Expression pattern evolution of ohnologs in LORe and AORe regions. An extension of Figure 3 with heatmaps showing 5 expression evolutionary scenarios (see Table 2) reflecting differential selection on tissue expression regulation after Ss4R WGD over genes with LORe and AORe histories. The color scale of the heatmaps correspond to the relative abundance of the transcript across all the tissues within the two species, in terms of counts per million (CPM). Each row across the four heatmaps represents one ortholog group of an ohnolog-tetrad. Connecting lines below heatmaps indicate duplicates belonging to same tissue clusters (conserved). Below the heatmaps are the boxplots representing expression correlation between and within duplicates in Atlantic salmon and grayling. The ohnologs in Atlantic salmon and grayling are represented as s1, s2 and g1 and g2 respectively.
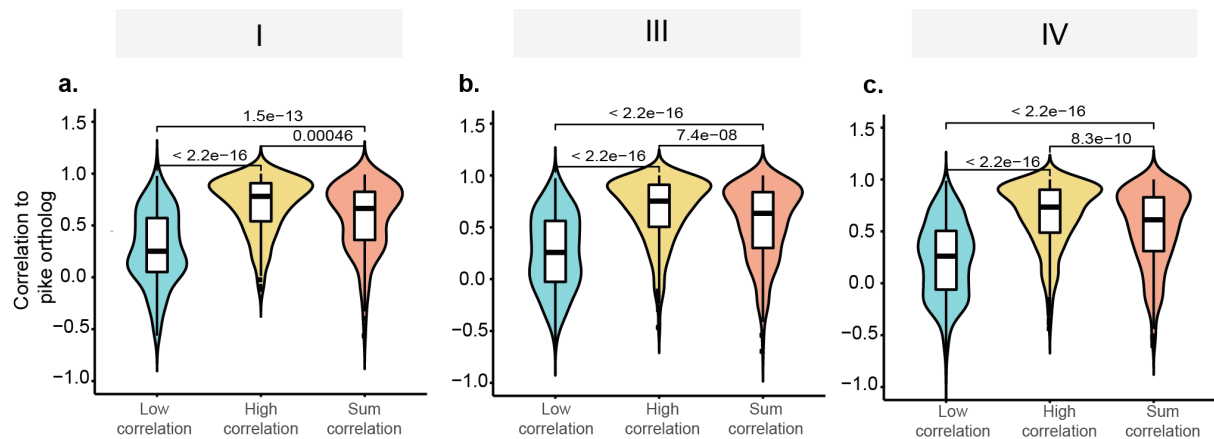
Supplementary figure S7: Violin plots (a-c) show the distributions of expression correlations between each of the duplicates, and the sum of the duplicates, to the corresponding pike ortholog for scenarios I, III and IV respectively. Duplicates belonging to each of the classes are divided based on correlation to the pike ortholog into High and Low groups. For class I and III, the plot represents the distributions of correlations of salmon duplicates, while for class (IV) grayling duplicates are plotted. P-values from pairwise comparisons using Wilcoxon test are indicated.
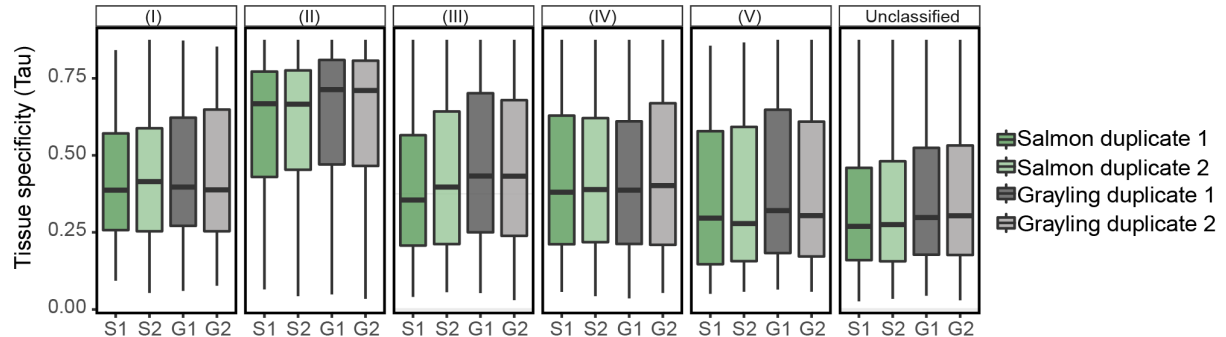
Figure S8: Tissue-specificity of ohnologs. Overall tissue specificity (Tau) distribution for each of the five evolutionary scenarios. The ohnologs in Atlantic salmon and grayling are represented as S1, S2 and G1 and G2 respectively.
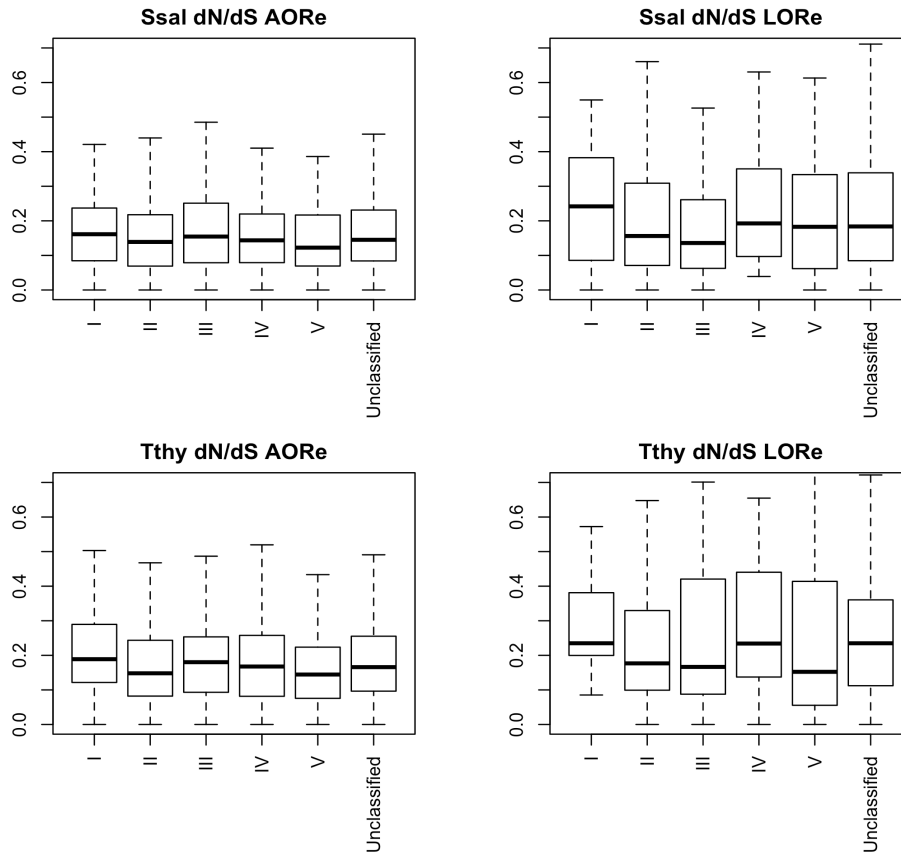
Figure S9: Coding sequence evolution. Distribution of dN/dS, representing coding sequence evolution, across different expression evolution scenarios across the LORe and AORe regions in Atlantic salmon and grayling.
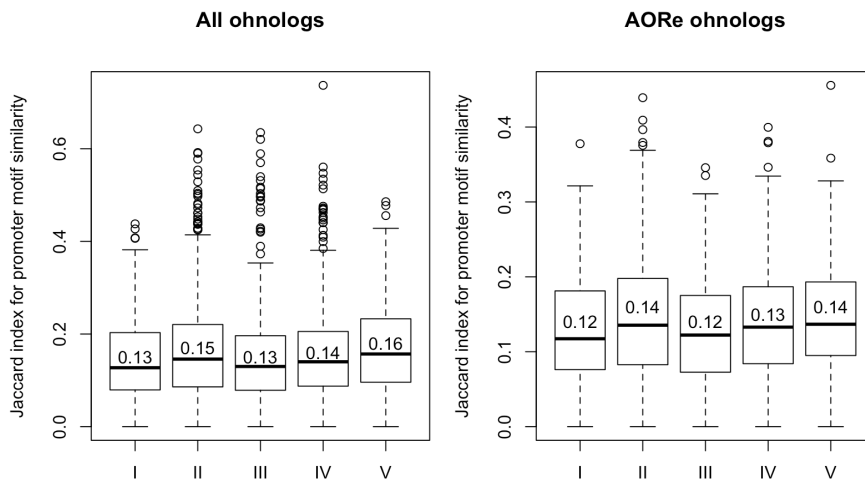
Figure S10: **Salmon ohnolog promoter motif conservation.** Boxplots of the similarity of promoter motif presence between ohnolog pairs in salmon in AORe regions is shown as Jaccard Index for the evolutionary scenarios I-V. Median Jaccard Indexes are indicated within boxes. The categories (I) and (III), where tissue regulation is diverged among salmon ohnologs, have significantly lower rank sum means in a wilcoxon test compared to the ohnologs in categories (II, IV and V) (p=0.01-0.002) which have conserved expression in salmon.
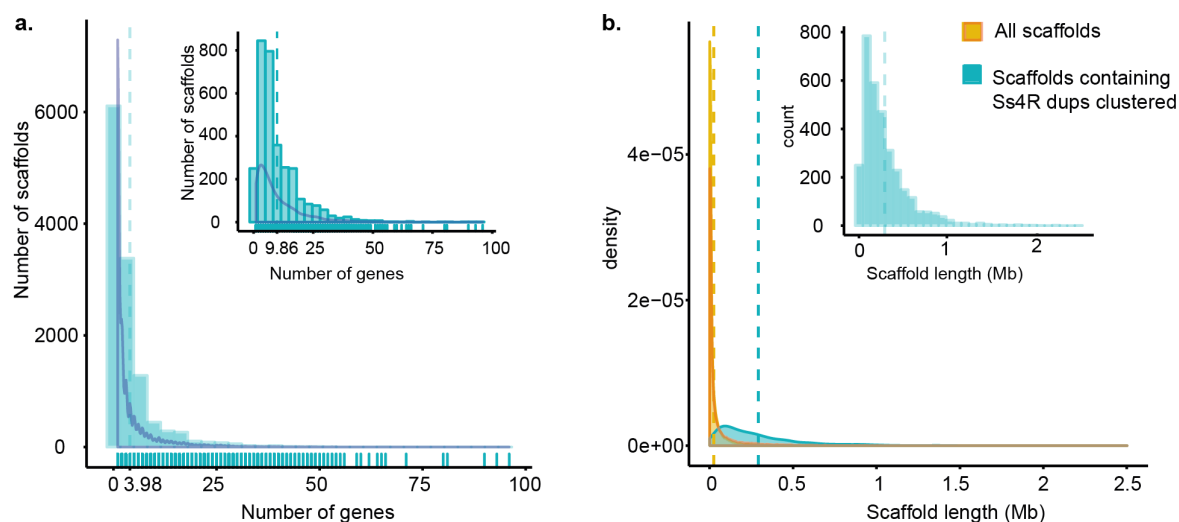
Figure S11: a. Number of genes per scaffold in the *T. thymallus* draft assembly. The inset plot shows the distribution of number of genes per scaffold for the genes used in the expression evolution analyses. Despite the overall fragmented nature of the assembly, most genes in our expression analysis are located on larger scaffolds. **b. Scaffold length distribution.** The density plot represents scaffold length distribution and the inset histogram plot represents the scaffold length distribution for the scaffolds containing the Ss4R genes used in the expression analyses.

Supplementary tables

Table S1: Sequencing libraries and data produced.

Insert size (bp)	Read length (bp)	Number of bases	Coverage*
180	150	103,520,976,600	57.51
3kb	100	85,304,163,600	47.39
3kb	100	42,464,706,800	23.59
6kb	100	90,372,684,400	50.21

* Based on a genome size estimate of 1.8 Gbp

Table S2: Summary of RNAseq data generated.

Tissue	Number of bases
Liver	17,086,217,700
Muscle	26,352,566,700
Spleen	27,942,424,800
Heart	28,336,371,600
Headkidney	23,154,448,800
Gonad	19,270,169,100
Eye	19,541,207,700
Brain	21,784,344,900
Gills	25,275,737,700

Table S3: Repeats and transposable elements in the assembly.

	Number of elements	Length occupied (bp)	Percentage of sequences
SINES	69,830	8566799	0.58 %
ALU	0	0	0.0 %
MIRs	216	8,430	0.0 %
LINES	316,144	117,430,749	8.0 %
LINE1	11,346	4,292,742	0.29 %
LINE2	120,455	40,284,357	2.74 %
L3/CR1	2,850	463,243	0.03 %
LTR elements	86,307	22,365,017	1.52 %
ERVL	91	13,152	0.0 %
ERVL-MaLRs	8	488	0.0 %
ERV_classI	9,443	2,223,741	0.15 %
ERV_classII	3,252	195,686	0.01 %
DNA elements	830,457	235,731,278	16.05 %
hAT-Charlie	11,902	3,316,142	0.23 %
TcMar-Tigger	141	41,487	0.0 %
Unclassified	777,695	167,400,996	11.40 %
Total interspersed repeats		551,494,839	37.55 %
Small RNA	1,867	150,973	0.01 %
Satellites	18,039	2,929,358	0.20 %
Simple repeats	599,983	43,271,286	2.95 %
Low complexity	64,487	4,726,681	0.32 %

Table S4. Distribution of tissue-expression clusters in tetrads of different regulatory evolution. Red cells represent genes in tissue expression clusters that were disproportionately represented compared to 'all' tetrads.

Expression evolution scenarios	Brain	Eye	Gills	Heart	Head kidney	Liver	Muscle	Spleen
I	124 (17%)	50 (7%)	118 (16%)	36 (5%)	124 (17%)	144 (20%)	68 (9%)	56 (8%)
II	1948 (49%)	328 (8%)	404 (10%)	52 (1%)	336 (8%)	368 (9%)	508 (13%)	56 (1%)
III	363 (20%)	150 (8%)	234 (13%)	109 (6%)	205 (12%)	321 (18%)	294 (17%)	104 (6%)
IV	578 (24%)	173 (7%)	276 (12%)	149 (6%)	453 (19%)	363 (15%)	228 (10%)	164 (7%)
V	162 (16%)	112 (11%)	96 (10%)	62 (6%)	148 (15%)	120 (12%)	144 (15%)	140 (14%)
All	4113 (26%)	1357 (9%)	1848 (12%)	996 (6%)	2134 (13%)	2231 (14%)	2057 (13%)	1220 (8%)
Red = Fisher test Bonferreoni corrected p-value <= 0.05								

Table S5: Differential expression analysis using the liver expression data. Number of ohnolog-tetrads with at least one ohnolog assigned to liver expression cluster, separated by evolutionary scenario class (see Table 2). Given are the number and percentage of those ohnolog-tetrads with significant (FDR adjusted p-value < 0.001) fold change (FC >2 or <0.5) in liver expression for grayling only, Atlantic salmon only, and for both species ohnologs.

Evolutionary scenario class	Ohnolog-tetrads with liver clustering	Significant FC in salmon only	Significant FC in grayling only	Significant FC in both species	No significant FC in any species
I	72	7 (10%)	14 (19%)	41 (57%)	10 (14%)
II	92	13 (14%)	23 (25%)	29 (32%)	27 (29%)
III	163	52 (32%)	20 (12%)	60 (37%)	30 (19%)
IV	165	22 (13%)	63 (38%)	58 (35%)	22 (13%)
V	60	8 (13%)	16 (27%)	16 (27%)	20 (33%)
Total	552	102 (18%)	136 (25%)	204 (37%)	110 (20%)