Genome Biology

CrossMark

Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification

Fiona M. Robertson[1], Manu Kumar Gundappa[1], Fabian Grammes[2], Torgeir R. Hvidsten[3,4], Anthony K. Redmond[1,5], Sigbjørn Lien[2], Samuel A. M. Martin[1], Peter W. H. Holland[6], Simen R. Sandve[2] and Daniel J. Macqueen[1*]

Abstract

Background: The functional divergence of duplicate genes (ohnologues) retained from whole genome duplication (WGD) is thought to promote evolutionary diversification. However, species radiation and phenotypic diversification are often temporally separated from WGD. Salmonid fish, whose ancestor underwent WGD by autotetraploidization ~95 million years ago, fit such a 'time-lag' model of post-WGD radiation, which occurred alongside a major delay in the rediploidization process. Here we propose a model, 'lineage-specific ohnologue resolution' (LORe), to address the consequences of delayed rediploidization. Under LORe, speciation precedes rediploidization, allowing independent ohnologue divergence in sister lineages sharing an ancestral WGD event.

Results: Using cross-species sequence capture, phylogenomics and genome-wide analyses of ohnologue expression divergence, we demonstrate the major impact of LORe on salmonid evolution. One-quarter of each salmonid genome, harbouring at least 4550 ohnologues, has evolved under LORe, with rediploidization and functional divergence occurring on multiple independent occasions >50 million years post-WGD. We demonstrate the existence and regulatory divergence of many LORe ohnologues with functions in lineage-specific physiological adaptations that potentially facilitated salmonid species radiation. We show that LORe ohnologues are enriched for different functions than 'older' ohnologues that began diverging in the salmonid ancestor.

Conclusions: LORe has unappreciated significance as a nested component of post-WGD divergence that impacts the functional properties of genes, whilst providing ohnologues available solely for lineage-specific adaptation. Under LORe, which is predicted following many WGD events, the functional outcomes of WGD need not appear 'explosively', but can arise gradually over tens of millions of years, promoting lineage-specific diversification regimes under prevailing ecological pressures.

Keywords: Whole genome duplication, Rediploidization, Species radiation, Lineage-specific Ohnologue Resolution (LORe), Duplicate genes, Functional divergence, Autotetraploidization, Salmonid fish

* Correspondence: daniel.macqueen@abdn.ac.uk
[1]Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK
Full list of author information is available at the end of the article

Robertson *et al. Genome Biology* (2017) 18:111

Page 2 of 14

Background

Whole genome duplication (WGD) has occurred repeatedly during the evolution of vertebrates, plants, fungi and other eukaryotes (reviewed in [1–4]). The prevailing view is that despite arising at high frequency, WGD is rarely maintained over macroevolutionary (i.e. millions of years (Myr)) timescales, but that, nonetheless, ancient WGD events are over-represented in several species-rich lineages, pointing to a role in long-term evolutionary success [1, 5]. WGD events provide an important source of duplicate genes (ohnologues) with the potential to diverge in protein functions and regulation during evolution [6, 7]. In contrast to the duplication of a single or small number of genes, WGD events are unique in allowing the balanced divergence of whole networks of ohnologues. This is thought to promote molecular and phenotypic complexity through the biased retention and diversification of interactive signalling pathways, particularly those regulating development [8–10].

As WGD events dramatically reshape opportunities for genomic and functional evolution, it is not surprising that an extensive body of literature has sought to identify causal associations between WGD and key episodes of evolutionary history, for example species radiations. Such arguments are clearly appealing and have been constructed for WGD events ancestral to vertebrates [11–15], teleost fishes [16–19] and angiosperms (flowering plants) [10, 20–22]. Nonetheless, it is now apparent that the evolutionary role of WGD is complex, often lineage-dependent and without a fixed set of rules. For example, some ancient lineages that experienced WGD events never underwent radiations, including horseshoe crabs [23] and paddlefish (e.g. [24]), while other clades radiated explosively immediately post-WGD, for example the ciliate *Paramecium* species complex [25]. In addition, apparent robust associations between WGD and the rapid evolution of species or phenotypic-level complexity may disappear when extinct lineages are considered, as proposed for WGDs in the stem of vertebrate and teleost evolution [26, 27].

Such findings either imply that the causative link between WGD and species radiations is weak, or demand alternative explanations. In the latter respect, it is has become evident that post-WGD species radiations may commonly arise following extensive time-lags. For example, major purported species radiations occurred >200 Myr after a WGD in the teleost ancestor ('Ts3R') ~320–350 million years ago (Ma) [3, 28, 29]. In angiosperms, similar findings have been reported in multiple clades [30, 31]. Such findings led to the proposal of a 'WGD Radiation Lag-Time' model, where some, but not all, lineages within a group sharing ancestral WGD diversified millions of years post-WGD, due to an interaction between a functional product of WGD (e.g. a novel trait) and lineage-specific ecological factors [30]. Within vertebrates, salmonids provide a textbook case of delayed species radiation following an ancestral WGD event ~95 Ma ('Ss4R'), where a role for ecological factors has been implied [32]. In this respect, salmonid diversification was strongly associated with climatic cooling and the evolution of a life-history strategy called anadromy [32] that required physiological adaptations (e.g. in osmoregulation [33]) enabling migration between fresh and seawater. Importantly, a convincing role for WGD in such cases of delayed post-WGD radiation is yet to be demonstrated, weakening hypothesized links between WGD and evolutionary success. Critically missing in the hypothesized link between WGD and species radiations is a plausible mechanism that constrains the functional outcomes of WGD from arising for millions or tens of millions of years after the original duplication event. Here we provide such a mechanism and uncover its potential impacts on adaptation.

Following all WGD events, the evolution of new molecular functions with the potential to influence long-term diversification processes depends on the physical divergence of ohnologue sequences. This is fundamentally governed by the meiotic pairing outcomes of duplicated chromosomes during the cytogenetic phase of post-WGD rediploidization [11, 34, 35]. Depending on the initial mechanism of WGD, rediploidization can be resolved rapidly or protracted in time. For example, after WGD by allotetraploidization, as recently described in the frog *Xenopus leavis* [36], WGD follows a hybridization of two species and recovers sexual incompatibility [11]. The outcome is two 'sub-genomes' within one nucleus that segregate into bivalents during meiosis [35]. In other words, rediploidization is resolved instantly, leaving ohnologues within the sub-genomes free to diverge as independent units at the onset of WGD. The other major mechanism of WGD, autotetraploidization, involves a spontaneous doubling of exactly the same genome. In this case, four identical chromosome sets will initially pair randomly during meiosis, leading to genetic exchanges (i.e. recombination) that prohibit the evolution of divergent ohnologues and enable an ongoing 'tetrasomic' inheritance of four alleles [35]. Crucially, rediploidization may occur gradually over tens of millions of years after autotetraploidization [35, 37].

Salmonid fish provide a vertebrate paradigm for delayed rediploidization post-autotetraploidization (reviewed in [37]). The recent sequencing of the Atlantic salmon (*Salmo salar* L.) genome revealed that rediploidization was delayed for one-quarter of the duplicated genome and associated with major genomic reorganizations such as chromosome fusions, fissions, deletions or inversions [38]. In addition, large regions of salmonid genomes still behave in a tetraploid manner in extant species (e.g. [38–40]), despite the passage of ~95 Myr since the Ss4R WGD [32].

In light of our understanding of salmonid phylogeny [32, 41], we can also be certain that rediploidization has been ongoing throughout salmonid evolution [38] and was likely occurring in parallel to lineage-specific radiations [32, 42]. However, the outcomes of delayed rediploidization on genomic and functional evolution remain uncharacterized in salmonids and other taxa. In the context of the commonly reported time-lag between WGD events and species radiation, this represents a major knowledge gap. Specifically, as explained above, a delay in the rediploidization process will cause a delay in ohnologue functional divergence, theoretically allowing functional consequences of WGD to be realized long after the original duplication.

Here we propose 'Lineage-specific Ohnologue Resolution' or 'LORe' as a mechanism to address the role of delayed rediploidization on the evolution of sister lineages sharing an ancestral WGD event (Fig. 1). It builds on and unifies ideas/data presented by Macqueen and Johnston [32], Martin and Holland [43] and Lien et al. [38] and is a logical outcome when rediploidization and speciation events occur in parallel. Under LORe, the rediploidization process is not completed until after a speciation event, which will result in the independent divergence of ohnologues in sister lineages (Fig. 1). This leads to unique predictions compared to the alternative scenario, where ohnologues began to diverge in the ancestor to sister lineages due to ancestral rediploidization (hereafter the 'Ancestral Ohnologue Resolution', or 'AORe' model). Under LORe, the evolutionary mechanisms allowing functional divergence of gene duplicates [6, 7, 11] become activated independently under lineage-specific selective pressures (Fig. 1). Conversely, under AORe, ohnologues share ancestral selection pressures, which hypothetically increases the chance that similar gene functions will be conserved in different lineages by selection (Fig. 1). A phylogenetic implication of LORe is a lack of 1:1 orthology between ohnologue pairs from different lineages (Fig. 1), leading to the definition of the
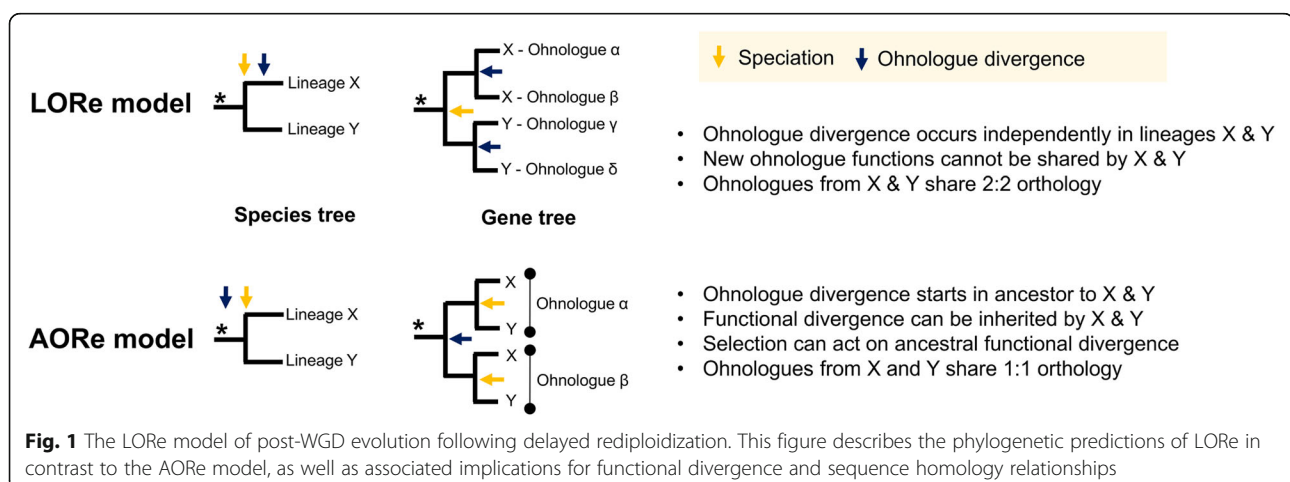
term 'tetralog' to describe a 2:2 homology relationship between ohnologues in sister lineages [43]. Thus, LORe may be mistaken for small-scale duplication if the underlying mechanisms are not appreciated. Despite this, LORe ohnologues have unique phylogenetic properties (Additional file 1: Figure S1) and are distinguished from small-scale gene duplication by their location within duplicated (or 'homeologous') blocks on distinct chromosomes sharing collinearity [38, 44, 45].

In this study, we demonstrate that one-quarter of retained salmonid ohnologues (conservatively, 4550 unique genes) have evolved under LORe, which has had a major impact on salmonid fish evolution at multiple levels of genomic and functional organization. Our findings allow us to propose that LORe offers a more broadly applicable mechanism to explain time-lags between many WGD events and subsequent lineage-specific diversification regimes.

Results

Extensive LORe followed the Ss4R WGD

To understand the extent and dynamics of lineage-specific rediploidization in salmonids, we used in-solution sequence capture [46] to generate a genome-wide ohnologue dataset spanning the salmonid phylogeny [32, 41]. Note, here we use the term ohnologue, but elsewhere 'homeologue' has been used to describe gene duplicates retained from the Ss4R WGD event [38]. In total, 383 gene trees were analysed (sum of aligned sequence data, 155,166 bp; mean/standard deviation (SD) alignment length, 405/208 bp), sampling every Atlantic salmon chromosome continuously at regular intervals and including ohnologues from at least seven species spanning all the major salmonid lineages plus a sister species (northern pike, *Esox lucius*) that did not undergo the Ss4R WGD [47] (Additional file 2). All the gene trees included verified Atlantic salmon ohnologues based on their location within duplicated (homeologous) blocks sharing common rediploidization histories [38]. Salmonids are split into three

Fig. 1 The LORe model of post-WGD evolution following delayed rediploidization. This figure describes the phylogenetic predictions of LORe in contrast to the AORe model, as well as associated implications for functional divergence and sequence homology relationships

Robertson *et al. Genome Biology* (2017) 18:111
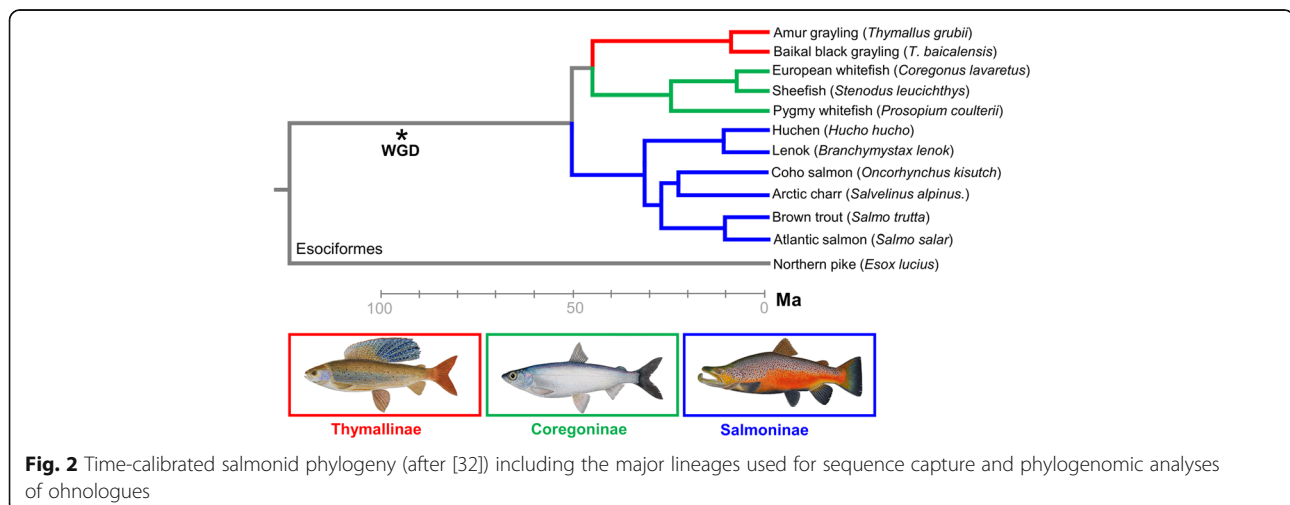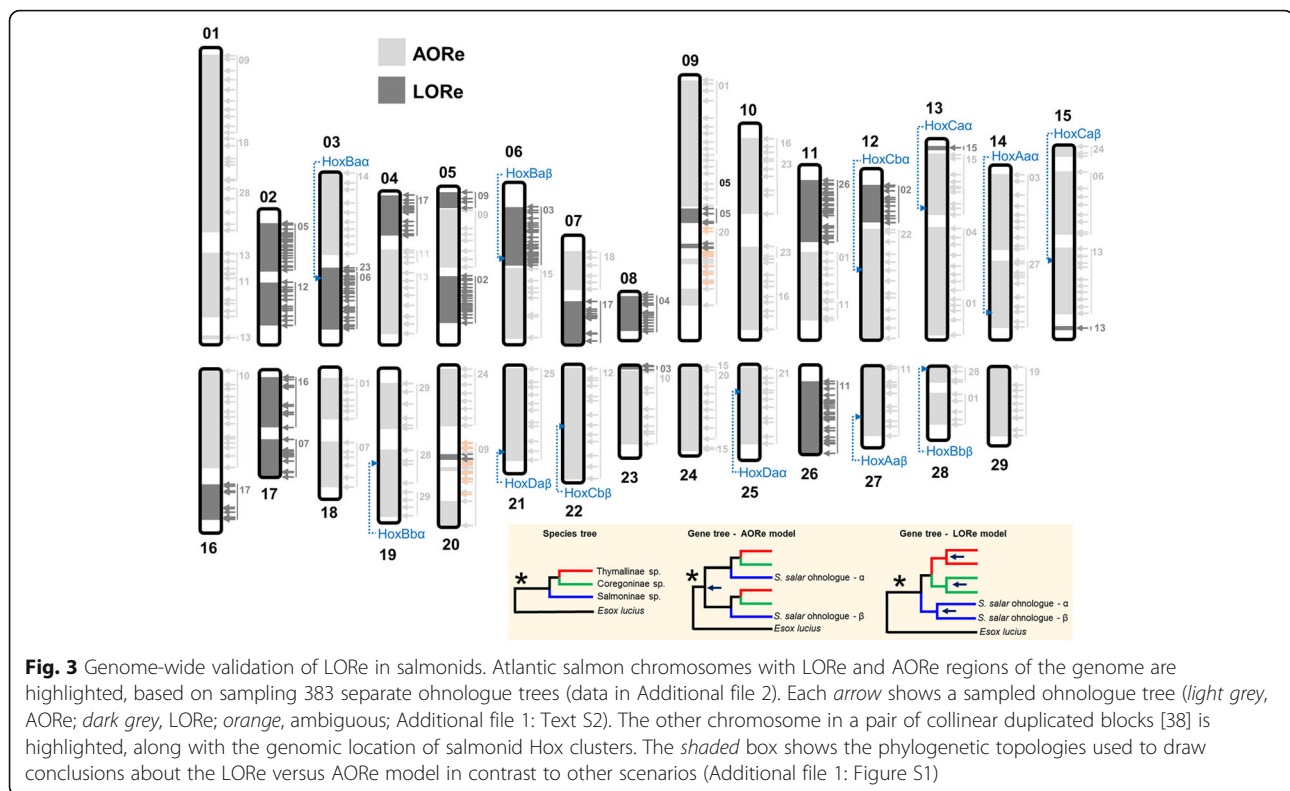
Page 4 of 14

subfamilies, Salmoninae (salmon, trout, charr, taimen/huchen and lenok spp.), Thymallinae (grayling spp.) and Coregoninae (whitefish spp.), which diverged rapidly between ~45 and 55 Ma (Fig. 2). Hence, phylogenetic signals of LORe are evidenced by subfamily-specific ohnologue clades (Fig. 1; Additional file 1: Figure S1). In accordance with this, our analysis revealed a consistent phylogenetic signal shared by large continuous duplicated blocks of the genome, with 97% of trees fitting predictions of either the LORe ($n = 151$ trees) or the AORe ($n = 219$ trees) model (Fig. 3; Additional file 2; Additional file 1: Text S2). This finding demonstrates a strong phylogenetic signal of either LORe or AORe, irrespective of the relatively short alignment length that was possible using our sequence capture approach.

The LORe regions defined by phylogenomic analysis represent around one-quarter of the genome and overlap fully with seven pairs of chromosome arms (homeologous arms '2p–5q', '2q–12qa', '3q–6p', '4p–8q', '7q–17qb', '11qa–26' and '16qb–17qa', according to Atlantic salmon nomenclature [38]) known to have undergone delayed rediploidization [38]. Lien et al. [38] reported that each of these chromosome arms shares a higher similarity among ohnologous sequences compared to the rest of the duplicated genome. Our gene tree sampling also revealed two additional, relatively small LORe regions (Fig. 3). The rest of the genome fits to the AORe model in our analysis (Fig. 3) and overlaps fully with collinear blocks located on chromosomes previously concluded to have experienced rediploidization in the salmonid ancestor [38]. Considering the near perfect congruence between our definitions of LORe and AORe and data presented in Lien et al. [38], we can robustly extrapolate that, among 16,786 high-confidence ohnologues identified within genomic regions covered by our analysis (see "Methods"), 27.1% (4550 genes) and 72.9% (12,236 genes) evolved under LORe and AORe, respectively.

To complement our genome-wide overview, we performed a finer-resolution phylogenetic analysis of Hox genes included in our sequence capture study. Hox genes are organized into genomic clusters located across multiple chromosomes and have been used to confirm separate WGD events in the stem of the vertebrate, teleost and salmonid lineages [43, 48, 49]. Phylogenetic analyses of Hox clusters (HoxBa) residing within predicted LORe regions in Atlantic salmon (Fig. 3) strongly supported the LORe model, considering either individual gene trees within a duplicated Hox cluster or trees built from combining separate ohnologue alignments sampled within clusters (e.g. Fig. 4a; Additional file 1: Text S1 and Figures S2–S10). Our data indicate that two salmonid-specific Hox cluster pairs underwent rediploidization as single units, either once independently in the common ancestor of each salmonid subfamily for HoxBa (Fig. 4a) or twice in Coregoninae for HoxAb (Additional file 1: Figure S9 and Text S1). These results cannot be explained by small-scale gene duplication events under any plausible scenario (Additional file 1: Text S1). Thus, HoxAb and HoxBa clusters were in regions of the genome that remained tetraploid until after the major salmonid lineages diverged ~50 Ma (Fig. 2). Phylogenetic analyses of the HoxAa, HoxBb, HoxCa, HoxCb and HoxDa cluster pairs strongly supported the AORe model (e.g. Fig. 4b; Additional file 1: Figure S10 and Text S1), as predicted by genomic location (Fig. 3).

We also studied proteins encoded within Hox clusters to contrast patterns of sequence divergence under the AORe and LORe models (Additional file 1: Figures S11 and S12). As our phylogenetic reconstructions were performed with nucleotide data, we wanted to rule out the possibility that the underlying sequence changes were predominantly synonymous, with little impact at the functional level. The data support our predictions (Fig. 1), as LORe has allowed many amino acid replacements to
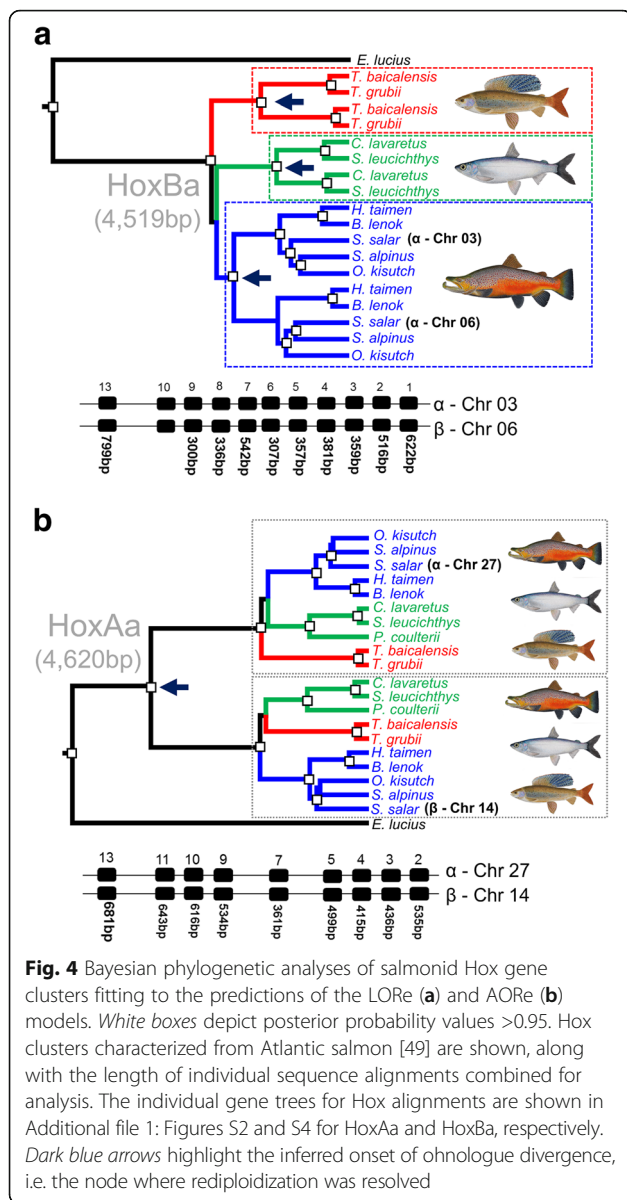

Fig. 2 Time-calibrated salmonid phylogeny (after [32]) including the major lineages used for sequence capture and phylogenomic analyses of ohnologues

Robertson *et al. Genome Biology* (2017) 18:111

Page 5 of 14

Fig. 3 Genome-wide validation of LORe in salmonids. Atlantic salmon chromosomes with LORe and AORe regions of the genome are highlighted, based on sampling 383 separate ohnologue trees (data in Additional file 2). Each *arrow* shows a sampled ohnologue tree (*light grey*, AORe; *dark grey*, LORe; *orange*, ambiguous; Additional file 1: Text S2). The other chromosome in a pair of collinear duplicated blocks [38] is highlighted, along with the genomic location of salmonid Hox clusters. The *shaded* box shows the phylogenetic topologies used to draw conclusions about the LORe versus AORe model in contrast to other scenarios (Additional file 1: Figure S1)

become independently fixed among Hox ohnologues within each salmonid subfamily (Additional file 1: Figure S12). These changes are typically highly conserved across species, suggesting lineage-specific purifying selection within a subfamily (Additional file 1: Figure S12). Conversely, under the AORe model, numerous amino acid replacements that distinguish Hox ohnologues arose in the common salmonid ancestor and have been conserved across all the major salmonid lineages (Additional file 1: Figure S11).

Distinct rediploidization dynamics across salmonid lineages

Our data also reveal distinct temporal dynamics of rediploidization across different salmonid lineages. First, using a Bayesian approach, the onset of divergence for the HoxBa-α and -β clusters of Salmoninae, Coregoninae and Thymallinae (i.e. Fig. 4a tree) was estimated at ~46, 25 and 34 Ma (posterior mean values), respectively (95% posterior density intervals of 36–57, 15–37 and 21–47 Ma, respectively). While the confidence intervals on these estimates overlap, the major difference in the mean posterior estimates is consistent with a scenario where the genomic regions containing these Hox clusters experienced rediploidization at substantially different times for the major salmonid lineages.

Further evidence of divergent rediploidization dynamics among salmonid lineages was observed through gene tree sampling (Fig. 3; Additional file 2), which allowed the number of inferred rediploidization events to be mapped along a time-calibrated salmonid phylogeny [32] (Fig. 5a). This was done by recording the divergence between ohnologue pairs (i.e. inferred onset of rediploidization) within each salmonid subfamily in all LORe trees sampled across the genome ($n = 151$; Additional file 2). In Salmoninae, 60/151 (40%) of the sampled genes trees indicated that rediploidization was completed in the stem of this subfamily, before the radiation of extant lineages (Fig. 5a). Assuming 4550 LORe genes (i.e. 2275 ohnologue pairs) underwent rediploidization during Salmoninae evolution, as estimated for the Atlantic salmon genome (i.e. 27.1% of 16,786 genes; see above), and that the Salmoninae stem branch is 19.5 Myr long [32] (Fig. 5a), we extrapolate that ~47 ohnologue pairs underwent rediploidization per Myr (i.e. 40% of 2275 ohnologue pairs/19.5 Myr) during the initial stages of Salmoninae evolution, leading up to the point when anadromy evolved [42]. In contrast, for the whitefish lineage, only 14% of the same LORe gene trees indicated that rediploidization was complete in the stem of extant lineages (Fig. 5a). Assuming the same number of ohnologue pairs were present in the whitefish ancestor and that the relevant stem branch is 25.5 Myr long [32] (Fig. 5a), we extrapolate that ~12 ohnologue pairs underwent rediploidization per Myr (i.e. 14% of 2275 ohnologue pairs/25.5 Myr) in the early stages of whitefish evolution, a rate four times lower than

Robertson *et al. Genome Biology* (2017) 18:111

Page 6 of 14

Fig. 4 Bayesian phylogenetic analyses of salmonid Hox gene clusters fitting to the predictions of the LORe (**a**) and AORe (**b**) models. *White boxes* depict posterior probability values >0.95. Hox clusters characterized from Atlantic salmon [49] are shown, along with the length of individual sequence alignments combined for analysis. The individual gene trees for Hox alignments are shown in Additional file 1: Figures S2 and S4 for HoxAa and HoxBa, respectively. *Dark blue arrows* highlight the inferred onset of ohnologue divergence, i.e. the node where rediploidization was resolved

a salmonid subfamily, we mapped our findings back to Atlantic salmon chromosomes (Fig. 5b). This showed that the number of inferred rediploidization events within a LORe region is consistent across large genomic regions (Fig. 5b; Additional file 1: Figure S13). Overall, these data support past observations that the rediploidization process is dependent on chromosomal location [38], while emphasizing distinct dynamics of rediploidization in different salmonid subfamilies.

Regulatory divergence under LORe

To better understand the functional implications of LORe, we contrasted the level of expression divergence between Atlantic salmon ohnologue pairs from AORe and LORe regions (Fig. 6). This was done in multiple tissues under controlled conditions (Fig. 5a, b) and also following 'smoltification' [33], a physiological remodelling that accompanies the life-history transition from freshwater to saltwater in anadromous salmonid lineages (Fig. 6c). In regions of the genome covered by our analysis, ohnologue expression was more correlated within LORe than AORe regions, both across tissues (Fig. 6b; Wilcoxon test, $P = 2.2e-16$) and considering differences in regulation between fresh and saltwater (Fig. 6c; Wilcoxon test, $P = 5.1e-10$). A recent analysis [38] suggested that 28% of salmonid ohnologues fit a model of expression divergence where one duplicate maintained the ancestral tissue expression (as observed in northern pike) and the other acquired a new expression pattern (i.e. 'regulatory neofunctionalization' [38]). We extended this analyses by partitioning ohnologue pairs from LORe and AORe regions of the genome. Among 2021 ohnologue pairs displaying regulatory neofunctionalization, ~19 versus ~81% were located in LORe and AORe regions, respectively, constituting a significant enrichment in AORe regions compared to the background expectation (i.e. 27.1 versus 72.9%; hypergeometric test, $P = 2e-13$). The average higher correlation in expression and lesser extent of regulatory neofunctionalization for ohnologues in LORe regions is expected, as they have had less evolutionary time to diverge in terms of sequences controlling mRNA-level regulation. Nonetheless, many ohnologues in LORe regions have diverged in expression (Fig. 6), which may have contributed to phenotypic variation available solely for lineage-specific adaptation.

Role of LORe in lineage-specific evolutionary adaptation

To better understand the role of LORe in adaptation, we performed an in-depth analysis of Atlantic salmon genes with established or predicted functions in smoltification [33], which we hypothesize represent important factors for the lineage-specific evolution of anadromy. Interestingly, LORe regions contain ohnologues for many genes from master hormonal systems regulating smoltification,
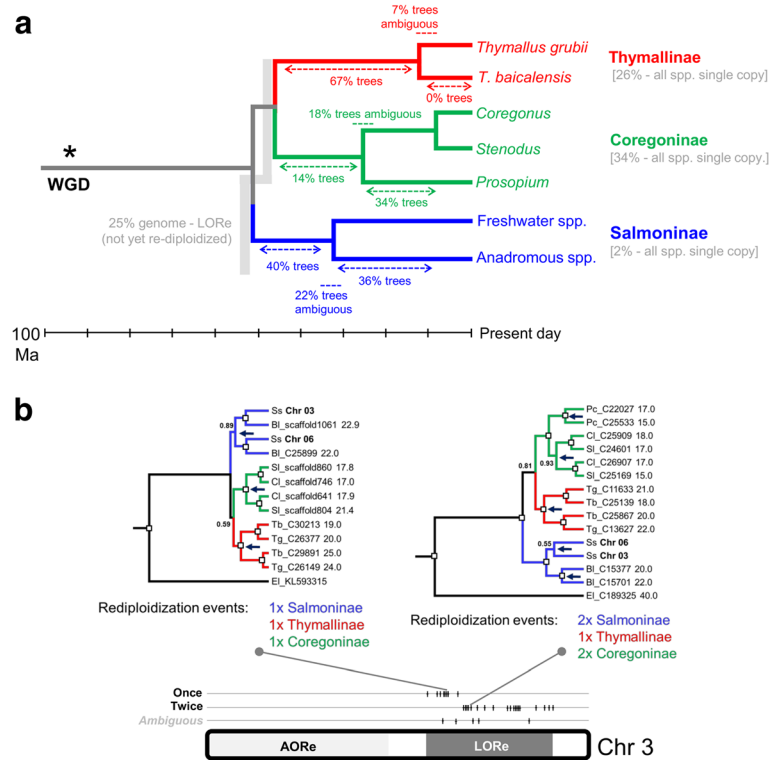
Salmoninae. It is impossible to estimate the rediploidization rate during the equivalent early stages of grayling evolution, as extant lineages diverged within the last 15 Myr [50]. Nonetheless, our data indicate that two-thirds of LORe ohnologues experienced rediploidization in the common ancestor to extant grayling spp. (Fig. 5a).

Interestingly, one-third of all sampled gene trees included a single ohnologue copy for whitefish and grayling, which were clustered along chromosomes in the genome (Additional file 2). As these regions have experienced delayed rediploidization, this likely reflects the 'collapse' of highly similar sequences in the assembly process into single contigs [38], rather than the evolutionary loss of an ohnologue. For two LORe regions with evidence of multiple rediploidization events within

Robertson *et al. Genome Biology* (2017) 18:111

Page 7 of 14

Fig. 5 Divergent rediploidization dynamics in different salmonid lineages. **a** Time-tree of species relationships [32] showing the fraction of 383 gene trees supporting independent rediploidization events at different nodes. **b** LORe region on chromosome 03 (paired with an ohnologous region on chromosome 06), where the number of independent rediploidization events inferred within Salmoninae (shown) is consistent along contiguous regions of the genome. Example trees are shown for genomic regions with distinct rediploidization histories. Abbreviations: *Ss S. salar*, *Bl Brachymystax lenok*, *Sl Stenodus leucichthys*, *Cl Coregonus lavaretus*, *Pc Prosopium coulterii*, *Tb Thymallus baicalensis*, *Tg T. grubii*
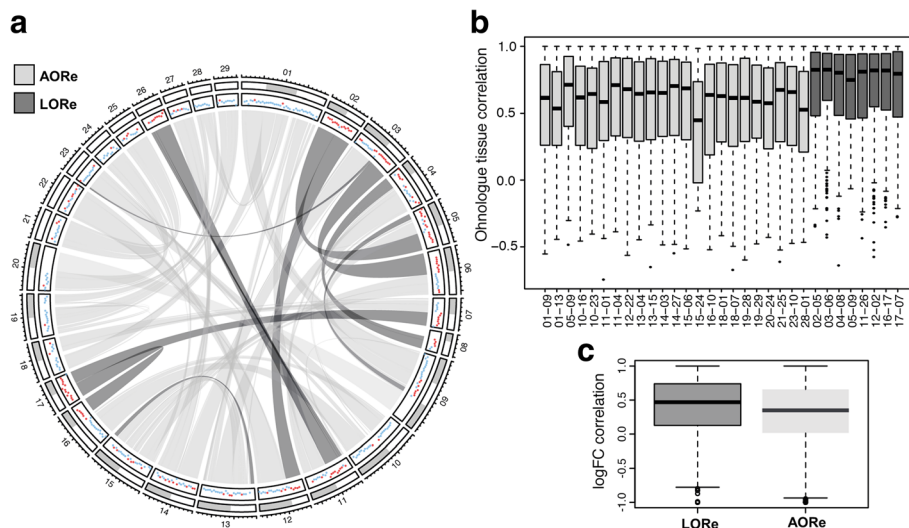
Fig. 6 Global consequences of LORe for ohnologue expression evolution. **a** Circos plot of Atlantic salmon chromosomes highlighting LORe and AORe regions defined by phylogenomics. The panel with *coloured dots* indicates expression similarity among ohnologue pairs: each *dot* represents the correlation of ohnologue expression across a 4-Mb window. *Red* and *blue dots* show correlations ≥0.6 and <0.6, respectively. **b** Correlation in expression levels across 15 tissues for ohnologue pairs in AORe and LORe regions. Different collinear blocks are shown [38] containing at least ten ohnologue pairs. **c** The overall correlation in the expression responses of ohnologues from LORe and AORe regions (2505 and 6853 pairs, respectively) during the physiological transition from fresh to saltwater. The correlation was calculated for log fold-change responses across nine tissues

Robertson *et al. Genome Biology* (2017) 18:111

Page 8 of 14

including the insulin-like growth factor (IGF), growth hormone (GH), thyroid hormone (TH) and cortisol pathways (Additional file 1: Table S1) [33, 51–53]. Notably, the actual master hormones from the IGF and GH pathways, i.e. encoding IGF1 and GH, which are together crucial for the development of seawater tolerance [33, 51], represent LORe ohnologues. We also identified many LORe ohnologues within a large set of genes involved in osmoregulation and cellular ionic homeostasis, key for saltwater tolerance, including Na+, K+-ATPases (targets for the above mentioned hormones [33, 51]), along with members of the ATP-binding cassette transporter, solute carrier and carbonic anhydrase families (Additional file 1: Table S1). Several additional genes from the same systems were represented by ohnologues in AORe regions (Additional file 1: Table S1).

To characterize the regulatory evolution of ohnologues with roles in smoltification, we compared equivalent tissue expression 'atlases' from Atlantic salmon in fresh and saltwater (Fig. 7; Additional file 3). The extent of regulatory divergence was variable for ohnologues in both LORe and AORe regions, ranging from conserved to unrelated tissue responses (Fig. 7a; Additional file 3). Several pairs of ohnologues from both LORe and AORe regions showed marked expression divergence in tissues of established importance for smoltification (examples in Fig. 7b; Additional file 3). For example, a pair of LORe

ohnologues encoding IGF1 located on chromosomes 07 and 17 (i.e. homeologous arms 7q–17qb under Atlantic salmon nomenclature [38]), despite differing by only a single conservative amino acid replacement, were differentially regulated in several tissues (Fig. 5b). The differential regulation of IGF1 ohnologues in gill and kidney is especially notable, as both tissues are vital for salt transport and, in gill, this hormone stimulates the development of chloride cells and the upregulation of Na+, K+-ATPases, together required for hypo-osmoregulatory tolerance [54, 55]. Thus, key expression sites for IGF1 are evidently fulfilled by different LORe ohnologues and these divergent roles have evolved specifically within the Salmoninae lineage, 40–50 Myr post WGD [32]. In contrast to IGF1, LORe ohnologues encoding GH showed highly conserved regulation during smoltification (Additional file 3). Overall, these findings demonstrate that many Atlantic salmon ohnologues in both LORe and AORe regions are differentially regulated under a physiological context that recaptures lineage-specific adaptations linked to anadromy.

To further characterize the role of LORe in lineage-specific adaptation, we performed gene ontology (GO) enrichment analysis contrasting all ohnologues present in LORe versus AORe regions (Additional file 4). Remarkably, ohnologues in LORe versus AORe regions were enriched for 99.9% non-overlapping GO terms,
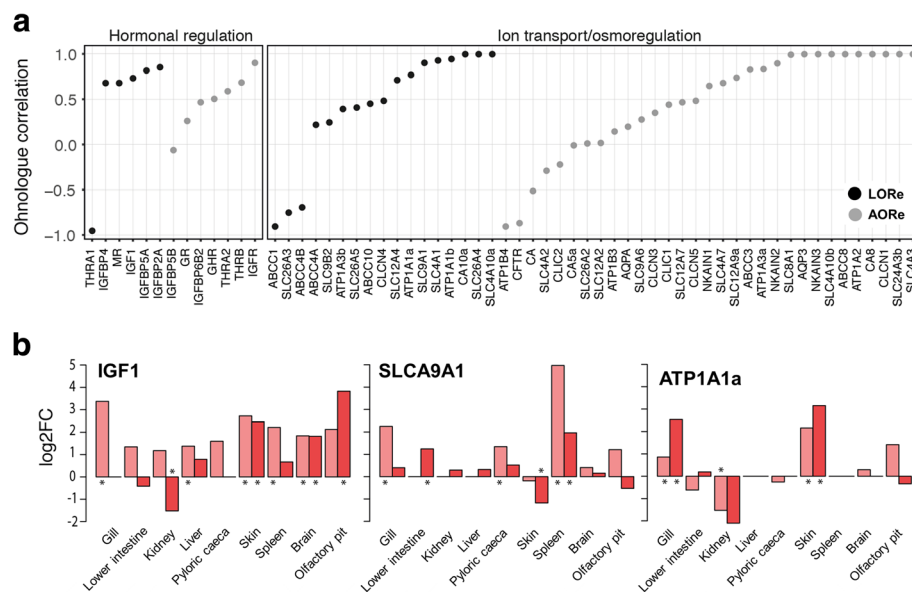

Fig. 7 Regulatory evolution of salmonid ohnologues implied in anadromy defined within a lineage-specific context of physiological adaptation. **a** Correlation in expression responses for ohnologues from LORe versus AORe regions during the fresh to saltwater transition in Atlantic salmon. Each name on the *x-axis* is a pair of ohnologues (details in Additional file 1: Table S1). The data are ordered from the most to least correlated ohnologue expression responses. Correlation was performed using Pearson's method. Data for additional ohnologues where correlation was impossible due to a restriction of expression to a limited set of tissues are provided in Additional file 3. **b** Example ohnologues showing a multi-tissue differential expression response to the fresh to saltwater transition. The *asterisks* highlight significant expression responses. Equivalent plots for all genes shown in **a** are provided in Additional file 3

Robertson *et al. Genome Biology* (2017) 18:111

Page 9 of 14

suggesting global biases in encoded functions (Additional file 4). The most significantly enriched GO terms for LORe ohnologues were 'indolalkylamine biosynthesis' and 'indolalkylamine metabolism' (Additional file 4). This is notable as 5-hydroxytryptamine is an indolalkylamine and the precursor to serotonin, which plays an important role controlling the master pituitary hormones that govern smoltification [51, 56]. An interesting feature of rediploidization is the possibility that functionally related genes residing in close genomic proximity (e.g. due to past tandem duplication) started diverging into distinct ohnologues as single units, for example Hox clusters (Fig. 4). We found that LORe ohnologues contributing to enriched GO terms ranged from being highly clustered in the genome to not at all clustered (Additional file 1: Table S2). In the latter case, we can exclude biases linked to regional rediploidization history. In the former case, we noted that two clusters of globin ohnologues on chromosomes 03 and 06 (i.e. homeologous arms '3q-6p' under Atlantic salmon nomenclature [38]) explain the enriched term 'oxygen transport' (Additional file 1: Table S2). This is interesting in the context of lineage-specific adaptation, as haemoglobin subtypes are regulated during smoltification to increase oxygen-carrying capacity and meet the higher aerobic demands of the oceanic migratory phase of the lifecycle [33]. Other GO terms enriched for LORe ohnologues included pathways regulating growth and protein synthesis, immunity, muscle development, proteasome assembly and the regulation of oxidative stress and cellular organization (Additional file 4).

Discussion

Here we define the LORe model and characterize its impacts on multiple levels of organization, adding a novel layer of complexity to our understanding of evolution after WGD. While past analyses have highlighted the quantitative extent of delayed rediploidization for a single salmonid genome [38], our study is the first to establish the genome-wide functional impacts of LORe and is unique in revealing divergent long-term rediploidization dynamics across the major salmonid lineages. Our results show that salmonid ohnologues can have strikingly distinct evolutionary 'ages', both for different genes located within the same genome (Figs. 3 and 4) and when comparing the same genes in phylogenetic sister lineages sharing the same ancestral WGD (Fig. 5). Our data also indicate that thousands of LORe ohnologues have diverged in regulation or gained novel expression patterns tens of Myr after WGD, likely contributing to lineage-specific phenotypes (Fig. 7). Hence, in the presence of highly delayed rediploidization, all ohnologues are not 'born equal' and many will have opportunities to functionally diverge under unique environmental and ecological contexts, for example during different phases of Earth's climatic and biological

evolution in the context of salmonid evolution [32]. It is also notable that ohnologues retained in LORe and AORe regions of the genome are enriched for different functions (Additional file 4), suggesting unique roles in adaptation, similar to past conclusions gained from comparison of ohnologues versus small-scale gene duplicates (e.g. [57, 58]). However, LORe is quite distinct from small-scale duplication, considering that large blocks of genes with common rediploidization histories will get the chance to diverge in functions in concert, meaning selection on duplicate divergence can operate on a multi-genic level.

LORe is possible whenever speciation precedes (or occurs in concert) to rediploidization (Fig. 1). This scenario is probable whenever rediploidization is delayed, most relevant for autotetraploidization events, which have occurred in plants [59], fungi [2] and unicellular eukaryotes (e.g. [25]) and was the likely mechanism of WGD in the stem vertebrate and teleost lineages [35, 43, 60]. However, LORe is not predicted under a strict definition of allotetraploidization, as cytogenetic rediploidization is resolved immediately. Nonetheless, after some allotetraploidization events, the parental genomes have high regional similarity (i.e. segmental allotetraploidy [61]), allowing prolonged tetrasomic inheritance in some genomic regions, leading to potential for LORe. Interestingly, past studies have provided indirect support for LORe outside salmonids, including following WGD in the teleost ancestor [43]. A recent analysis of duplicated Hox genes from the lamprey *Lethenteron japonicum* failed to provide evidence of 1:1 orthology comparing jawed and jawless vertebrates, leading to the radical suggestion of independent, rather than ancestral vertebrate WGD events [62]. However, if rediploidization was delayed until after the divergence of these major vertebrate clades, which occurred no more than 60–100 Myr after the common vertebrate ancestor split from 'unduplicated' chordates [63, 64], such findings are parsimoniously explained by LORe. In other words, WGD events may be shared by all vertebrates [60, 65], but some ohnologues became diploid independently in jawed and jawless lineages. Gaining unequivocal support for LORe beyond salmonids will require careful phylogenomic approaches akin to those employed here.

Our findings also reveal a possible mechanism to explain why some lineages experienced delayed post-WGD species radiations, i.e. the WGD radiation lag-time model [30–32]. This is a topical subject, given the recent suggestion that teleosts radiated at a similar rate to their sister lineage (holosteans) in the immediate wake of the teleost-specific WGD (Ts3R) [27], but nonetheless experienced later radiations [28, 29]. Our results suggest that, in the presence of delayed rediploidization, the functional outcomes of WGD need not arise 'explosively', but can be mechanistically delayed for tens of Myr. For

Robertson *et al. Genome Biology* (2017) 18:111

Page 10 of 14

example, tissue expression responses for master genes required for saltwater tolerance are evidently fulfilled by one member of a salmonid ohnologue pair that first began to diverge in functions 40–50 Myr post-WGD (Fig. 7). Hence, in light of evidence for delayed rediploidization after Ts3R [43], an alternative interpretation is that teleosts gained an increasing competitive advantage through time compared to their unduplicated sister group, via the drawn-out creation of functionally divergent ohnologue networks that provided greater scope for adaptation to on-going environmental change. Similar arguments apply for delayed radiations in angiosperm lineages sharing WGD with a sister clade that diversified at a lower rate [30, 31], offering a worthy area of future investigation.

For salmonids, climatic cooling likely provided a selective pressure promoting the lineage-specific evolution of anadromy, which, according to formal diversification rate tests, facilitated speciation in the long-term [32]. Interestingly, we observed an elevated rediploidization rate in Salmoninae compared to other lineages leading up to the time that anadromy evolved. Taken with the lineage-specific regulatory divergence of LORe ohnologues regulating smoltification (Figs. 5 and 7), we hypothesize that LORe contributed to the evolution of lineage-specific adaptations that promoted species radiation. However, the role of LORe in adaptation is likely complex, occurring in a genomic context where an existing substrate of AORe ohnologues (that have had greater opportunity to diverge in function) can also contribute to lineage-specific adaptation. This is evident in our data, as many relevant ohnologues from AORe regions of the genome show extensive regulatory divergence in the context of smoltification (Fig. 7; Additional file 3). A realistic scenario for lineage-specific adaptation involves functional interactions between networks of newly diverging LORe ohnologues and 'older' AORe ohnologues that have already diverged in function from the ancestral state. Nonetheless, even though all ohnologues may undergo lineage-specific functional divergence, only during the initial stages of LORe will neofunctionalization and subfunctionalization [6, 7, 11] arise without the influence of purifying selection on past functional divergence (Fig. 1). In the future, follow-up questions on the roles of both classes of salmonid ohnologues (and indeed other types of gene duplicate) in lineage-specific adaptation will become possible through comparative analysis of multiple salmonid genomes, done in a phylogenetic framework spanning the evolutionary transition to anadromy [66].

Conclusions

Our results empirically validate the LORe model and demonstrate its unappreciated significance as a nested component of genomic and functional divergence following WGD. LORe should now be considered within

future investigations into the role of WGD as a driver of evolutionary adaptation and diversification, including delayed post-WGD radiations.

Methods

Target-enrichment and Illumina sequencing

To generate a genome-wide ohnologue set for phylogenomic analyses in salmonids, we used in-solution sequence capture with the Agilent SureSelect platform prior to sequencing on an Illumina HiSeq2000. Full methods were recently detailed elsewhere, including the source and selection of 16 study species [46]. While this past study was a small-scale investigation of a few genes [46], here we up-scaled the approach to 1293 unique capture probes (Additional file 5; Additional file 1: Text S3 provides details on probe design). 120mer oligomer baits were synthesised at fourfold tiling across the full probe set and a total of 1.5 Mbp of unique sequence data were produced in each capture library. The captures were performed on randomly fragmented gDNA libraries, meaning that the recovered data represent exons plus flanking genomic regions [46]. We recovered 21.7 million reads per species on average after filtering low-quality data (SD, 0.8 million reads; >99.1% paired-end data; Additional file 1: Table S3), which were assembled using SOAPdenovo2 [67] with a K-mer value of 91 and merging level of 3 (otherwise default parameters). Species-specific BLAST databases [68] were created for downstream analyses. Assembly statistics were assessed via the QUAST webserver [69] (Additional file 1: Table S3). We used BLAST and mapping approaches to confirm that the sequence capture worked efficiently with high specificity and that pairs of ohnologues had been routinely recovered, even when a single ohnologue was used as a capture probe (full details given in Additional file 1: Text S3).

Phylogenomic analyses

This work was split into a genome-wide investigation and a detailed study of Hox clusters. For both approaches, sequence data were sampled from our capture databases for different salmonid spp. using BLASTn [68] and aligned with MAFFT v.7 using the default automatic strategy [70]. Northern pike was used as the outgroup to the Ss4R WGD in all analyses; this species was included in our target-enrichment study, but pike sequences were captured slightly less efficiently compared to salmonids [46]. Thus, we supplemented pike sequences using the latest genome assembly [47] (ASM72191v2; NCBI accession CF_000721915). All phylogenetic tests were done at the nucleotide level within the Bayesian Markov chain Monte Carlo (MCMC) framework BEAST v.1.8 [71],

Robertson *et al. Genome Biology* (2017) 18:111

Page 11 of 14

specifying an uncorrelated lognormal relaxed molecular clock model [72] and the best-fitting substitution model (inferred by maximum likelihood in Mega v.6.0 [73] for individual alignments and PartitionFinder [74] for combined alignments). The MCMC chain was run for 10-million generations and sampled every 1000th generation. TRACER v.1.6 [75] was used to confirm adequate mixing and convergence of the MCMC chain (effective sample sizes >200 for all estimated parameters). Maximum clade credibility trees were generated in TreeAnnotator v.1.8 [71]. All sequence alignments and Bayesian gene trees are provided in Additional file 2, including details on ohnologues sampled from the Atlantic salmon genome, alignment lengths and the best-fitting substitution model.

For the genome-wide study, the 1293 unique capture probes were used in BLASTn searches against the Atlantic salmon genome (ICSASG_v2; NCBI accession GCF_000233375) via http://salmobase.org/. This provided a genome-wide overview of the location of ohnologue alignments that could be generated via our capture assemblies and confidence that the targeted genes were true ohnologues retained from the Ss4R WGD, based on their location within collinear duplicated (homeologous) blocks [38]. In total, 383 ohnologue alignments were generated, using the appropriate probes as BLAST queries against our capture databases to acquire the sequence data. The selection of gene trees sampled among those available from the sequence capture data was done to maximize the overall representation of each salmon chromosome, with a higher sampling effort performed in putative LORe regions, i.e. chromosome arms with a known history of delayed rediploidization [38]. Each tree contained a pair of verified ohnologues from Atlantic salmon and putative ohnologues captured from at least one species per each of the most distantly related lineages within the three salmonid subfamilies.

For the Hox study, we used 89 Hox genes from Atlantic salmon [53] as BLASTn queries against our capture assemblies. The longest captured regions were aligned, leading to 54 alignments (accounted for within the 383 ohnologue alignments mentioned above) spanning all characterized Hox clusters [53]. We performed individual-level phylogenetic analyses on each dataset, revealing a highly consistent phylogenetic signal across different Hox genes from each Hox cluster (Additional file 1: Figure S2–S8), allowing alignments to be combined to the level of whole Hox clusters. To estimate the timing of rediploidization in the duplicated HoxBa cluster of salmonids [49], we employed the dataset combining all sequence alignments (i.e. tree in Fig. 4a). However, the analysis was done after setting calibration priors at four nodes according to MCMC posterior estimates of divergence times from a previous fossil-calibrated analysis [32]. The calibrations were made for the ancestor to two salmonid-specific HoxBa ohnologue clades for Salmoninae and Coregoninae. For Salmoninae, we set the prior for the common ancestor to *Hucho*, *Brachymystax*, *Salvelinus*, *Salmo* and *Oncorhynchus* (normally distributed, median = 32.5 Ma; SD, 3.5 Ma; 97.5% interval, 25–39 Ma). For Coregoninae, we set the prior for the common ancestor to *Stenodus leucichthys* and *Coregonus lavaretus* (normally distributed, median = 4.2 Ma; SD, 0.9 Ma; 97.5% interval, 2.4–5.7 Ma). We ran the calibrated BEAST analysis without data to confirm the intended priors were recaptured in the MCMC sampling.

RNAseq analyses

To analyse ohnologue regulatory divergence in an appropriate physiological context to explore the evolution of anadromy, we performed RNAseq on nine Atlantic salmon tissues (gill, lower intestine, kidney, liver, pyloric caeca, skin, spleen, brain, olfactory pit) sampled before and after smoltification (see Additional file 6 for detailed information on samples and mapping statistics). Six fish (three males and three females) were sampled from both freshwater (i.e. pre-smoltification, $n = 6$; mean/SD length, 18.6/0.5 cm) and saltwater (i.e. post-smoltification, $n = 6$; mean/SD length, 25.8/0.8 cm) at AquaGen facilities (Trondheim, Norway). RNA extraction was performed on each tissue and its purity and integrity were assessed using a Nanodrop 1000 spectrophotometer (Thermo-Scientific) and 2100 BioAnalyzer (Agilent), respectively. Subsequently, libraries were produced from 2 μg of total RNA using a TruSeq stranded total RNA sample Kit (Illumina, USA) according to the manufacturer's instructions (Illumina #15031048 Rev.E). Sequencing was performed on a MiSeq instrument using a v.3 MiSeq Reagent Kit (Illumina) generating 2×300 bp, strand-specific, paired-end reads. For each tissue, the sequenced individuals were pooled into two sets of three individuals of each sex in both freshwater and saltwater (hence, any reported responses are common to males and females; sex-specific differences were not considered in this study). For the global analysis of ohnologue expression divergence in different tissues under controlled conditions (i.e. Fig. 6a, b), we employed high-coverage Illumina transcriptome reads previously generated for 15 Atlantic salmon tissues (described in [38]).

In both RNAseq analyses, raw Illumina reads were subjected to adapter and quality trimming using cutadapt [76], followed by quality control with FastQC, before mapping to the RefSeq genome assembly (ICSASG_v2) using STAR v.2.3 [77]. Uniquely mapped reads were counted using the HTSeq python script [78] in combination with a modified RefSeq.gff file. The .gff file was modified to contain the attribute "gene_id" (file accessible at http://salmobase.org/download.html).

Robertson *et al. Genome Biology* (2017) 18:111

Page 12 of 14

Expression levels were calculated as counts per million total library counts in EdgeR [79]. Total library sizes were normalised to account for bias in sample composition using the trimmed mean of m-values approach [77]. For the smoltification study, log-fold expression changes were calculated, contrasting samples from freshwater and saltwater, done separately for each tissue using EdgeR [79]. Genes showing a false discovery rate-corrected *P* value ≤0.05 were considered differentially expressed.

To identify salmonid-specific ohnologue pairs in AORe and LORe regions of the Atlantic salmon genome, a self-BLASTp analysis was done using all annotated RefSeq proteins, keeping only proteins coded by genes within verified collinear (homeologous) regions retained from the Ss4R WGD [38] with >50% coverage and >80% identity to both query and hit. Statistical analyses on expression data were performed using various functions within R [80]. Expression divergence was estimated using Pearson correlation in all cases. The Circos plot (Fig. 6a) was generated using the circlize library in R [81].

GO enrichment analyses

GO annotations for Atlantic salmon protein-coding sequences were obtained using Blast2GO [82]. The longest predicted protein for each gene was blasted against SwissProt (http://www.ebi.ac.uk/uniprot) and processed with default Blast2GO settings [83]. The results have been bundled into an R package (https://gitlab.com/cigene/R/Ssa.RefSeq.db). Protein-coding genes were tested for enrichment of GO terms belonging to the sub-ontology 'biological process' using a Fisher test implemented in the Bioconductor package topGO [83]. The analysis was restricted to terms of a level higher than four, with more than 10 but less than 1000 assigned genes. Enrichment analyses were done separately for all ohnologue pairs with annotations retained in LORe (2002 pairs) and AORe (5773 pairs) regions of the RefSeq genome assembly. We recorded the chromosomal locations of LORe ohnologues for the most significantly enriched GO terms, including the number of unique LORe regions they occupy in the genome (Additional file 1: Table S2). The rationale was to establish the extent to which ohnologues underlying an enriched GO term are physically clustered. We devised a 'clustering index', quantifying the total number of cases where $n \geq 2$ ohnologues present within the relevant genomic regions are located within 500 kb of each other, expressed as a proportion of $n - 1$ the total number of ohnologues located within those regions. A respective clustering index of 1.0, 0.5 and 0.0 means that all, half or zero of the ohnologues accounting for an enriched GO term are located within 500 kb of their next nearest gene within the same genomic region; 500 kb was considered a conservative distance to capture genes expanded by tandem duplication.

Additional files

Additional file 1: Supporting text, figures and tables. Contains **Text S1–S3**, **Figures S1–S13** and **Tables S1–S3**. (PDF 3119 kb)

Additional file 2: Phylogenomic analysis. Full data associated with the 383 ohnologue gene trees used to define LORe and AORe regions, including: (i) NCBI accession number and gene/protein details for sequence capture probes, (ii) genomic location of ohnologues in the Atlantic salmon genome, (iii) the length of each sequence alignment, (iv) the substitution model used for Bayesian phylogenetic analysis, (iv) the number of inferred rediploidization events (v), each gene tree (nexus format), and (vi) alignments used to generate each gene tree (fasta format). (XLSX 1984 kb)

Additional file 3: Full ohnologue expression response data (summarized in Fig. 7). Multi-tissue expression responses accompanying the freshwater to saltwater transition for candidate Atlantic salmon ohnologues with implied functions in smoltification and anadromous life-history. (PDF 1844 kb)

Additional file 4: Ohnologue GO enrichment analyses. Full enrichment data for GO terms belonging to the sub-ontology 'biological process', done comparing protein-coding ohnologues located in LORe and AORe regions. (XLSX 7524 kb)

Additional file 5: Sequence capture probes. Full data associated with the 1293 sequence probes used to capture a genome-wide dataset of ohnologues across 15 salmonid species, including: (i) NCBI accession and encoded protein product for each sequence capture probe, (ii) the salmonid species from which each probe sequence was taken, (iii) 'Probe status', where 'Singleton' means only a single gene duplicate was included in the probe set, even if other duplicates were identified and 'Duplicate' means both gene duplicates were present in the probe set, and (iv) 'Selection status', where 'Pre-selected' means the sequence capture probe was selected a priori and 'Selected randomly' means the genes were chosen at random. (XLSX 588 kb)

Additional file 6: Detailed RNAseq information. Full data associated with transcriptomic analysis of ohnologue expression responses accompanying the freshwater to saltwater transition in Atlantic salmon, including: (i) sample metadata, (ii) read and mapping statistics, (iii) gene/feature count statistics, and (iv) NCBI SRA accession numbers. (XLSX 18 kb)

Robertson *et al. Genome Biology* (2017) 18:111

Page 13 of 14

Authors' contributions
DJM and PWHH defined the LORe model. DJM designed the sequence capture study. FMR performed lab work for the sequence capture study and contributed to probe design. FMR, MKG and DJM performed phylogenetic analyses. FG, TRH and SRS performed the expression analyses. FG performed GO enrichment analyses. DJM and MKG interpreted GO enrichment analyses. FMR, MKG, FG, SRS and DJM designed figures/tables. DJM drafted the manuscript. All authors interpreted data and contributed to the writing of the final manuscript. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Consent for publication
Not applicable.

Ethics approval and consent to participate
Atlantic salmon tissues were sampled by AquaGen (Trondheim, Norway) following Norwegian national guidelines with regard to the ethical aspects of animal welfare.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 2TZ, UK. [2]Centre for Integrative Genetics (CIGENE), Faculty of Biosciences, Norwegian University of Life Sciences, Ås NO-1432, Norway. [3]Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1432 Ås, Norway. [4]Umeå Plant Science Centre, Department of Plant Physiology, Umeå Plant Science Centre, Umeå University, SE-90187 Umeå, Sweden. [5]Centre for Genome-Enabled Biology & Medicine, University of Aberdeen, Aberdeen AB24 2TZ, UK. [6]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

References
1. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplication. Nat Rev Genet. 2009;10:725–32.
2. Albertin W, Marullo P. Polyploidy in fungi: evolution after whole-genome duplication. Proc Biol Sci. 2012;279:2497–50.
3. Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol Genet Genomics. 2014;289:1045–60.
4. Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. Curr Opin Genet Dev. 2015;35:119–25.
5. Comai L. The advantages and disadvantages of being polyploid. Nat Rev Genet. 2005;6:836–46.
6. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 2008;9:938–50.
7. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010;11:97–108.
8. Freeling M, Thomas BC. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. Genome Res. 2006;16:805–14.
9. Huminiecki L, Heldin CH. 2R and remodeling of vertebrate signal transduction engine. BMC Biol. 2010;8:146.
10. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473:97–100.
11. Ohno S. Evolution by gene duplication. New York: Springer-Verlag; 1970. See: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2006-7-5-r43.
12. Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. Dev Suppl. 1994;125-33. See: https://genomebiology.biomedcentral.com/articles/10.1186/gb-2006-7-5-r43.
13. Sidow A. Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev. 1996;6:715–22.
14. Aburomia R, Khaner O, Sidow A. Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. J Struct Funct Genomics. 2003;3:45–52.
15. Blomme T, Vandepoele K, De Bodt S, Simillion C, Maere S, Van de Peer Y. The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol. 2006;7:R43.
16. Wittbrodt J, Meyer A, Schartl M. More genes in fish? BioEssays. 1998;20:511–5.
17. 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.
18. Meyer A, Van de Peer Y. From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays. 2005;27:937–45.
19. 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;23:121–36.
20. De Bodt S, Maere S, Van de Peer Y. Genome duplication and the origin of angiosperms. Trends Ecol Evol. 2005;20:591–7.
21. Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, et al. Polyploidy and angiosperm diversification. Am J Bot. 2009;96:336–48.
22. Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. Curr Opin Plant Biol. 2016;30:159–65.
23. Nossa CW, Havlak P, Yue JX, Lv J, Vincent KY, Brockmann HJ, Putnam NH. Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. Gigascience. 2014;3:9.
24. Crow KD, Smith CD, Cheng JF, Wagner GP, Amemiya CT. An independent genome duplication inferred from Hox paralogs in the American paddlefish–a representative basal ray-finned fish and important comparative reference. Genome Biol Evol. 2012;4:937–53.
25. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, et al. Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. Nature. 2006;444:171–8.
26. Donoghue PC, Purnell MA. Genome duplication, extinction and vertebrate evolution. Trends Ecol Evol. 2005;20:312–9.
27. Clarke JT, Lloyd GT, Friedman M. Little evidence for enhanced phenotypic evolution in early teleosts relative to their living fossil sister group. Proc Natl Acad Sci U S A. 2016;113:11531–6.
28. Santini F, Harmon LJ, Carnevale G, Alfaro ME. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes BMC Evol Biol. 2009;9:194.
29. Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc Natl Acad Sci U S A. 2009;106:13410–4.
30. Schranz ME, Mohammadin S, Edger PP. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr Opin Plant Biol. 2012;15:147–53.
31. Tank DC, Eastman JM, Pennell MW, Soltis PS, Soltis DE, Hinchliff CE, Brown JW, Sessa EB, Harmon LJ. Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications. New Phytol. 2015;207:454–67.
32. Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc Biol Sci. 2014;281:20132881.
33. Björnsson BT, Stefansson SO, McCormick SD. Environmental endocrinology of salmon smoltification. Gen Comp Endocrinol. 2011;170:290–8.
34. Wolfe KH. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet. 2001;2:333–41.
35. Furlong RF, Holland PW. Were vertebrates octoploid? Philos Trans R Soc Lond B Biol Sci. 2002;357:531–4.
36. Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, Fukui A, Hikosaka A, et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. Nature. 2016;538:336–43.
37. Allendorf FW, Thorgaard GH. Tetraploidy and the evolution of salmonid fishes. In: Turner BJ, editor. Evolutionary genetics of fishes. New York: Plenum Press; 1984. p. 1–53.
38. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, et al. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016;533:200–5.

Robertson *et al. Genome Biology* (2017) 18:111

Page 14 of 14

39. Allendorf FW, Bassham S, Cresko WA, Limborg MT, Seeb LW, Seeb JE. Effects of crossovers between homeologs on inheritance and population genomics in polyploid-derived salmonid fishes. J Hered. 2015;106:217–27.

40. Waples RK, Seeb LW, Seeb JE. Linkage mapping with paralogs exposes regions of residual tetrasomic inheritance in chum salmon (*Oncorhynchus keta*). Mol Ecol Resour. 2016;16:17–28.

41. Campbell MA, López JA, Sado T, Miya M. Pike and salmon as sister taxa: detailed intraclade resolution and divergence time estimation of Esociformes + Salmoniformes based on whole mitochondrial genome sequences. Gene. 2013;530:57–65.

42. Alexandrou MA, Swartz BA, Matzke NJ, Oakley TH. Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. Mol Phylogenet Evol. 2013;69:514–23.

43. Martin KJ, Holland PW. Enigmatic orthology relationships between Hox clusters of the African butterfly fish and other teleosts following ancient whole-genome duplication. Mol Biol Evol. 2014;31:2592–611.

44. Van de Peer Y. Computational approaches to unveiling ancient genome duplications. Nat Rev Genet. 2004;5:752–63.

45. Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nat Commun. 2014;22(5):3657.

46. Lappin FL, Shaw RL, Macqueen DJ. Targeted sequencing for high-resolution evolutionary analyses following recent genome duplication: proof of concept for key components of the salmonid insulin-like growth factor axis. Mar Genomics. 2016. doi:10.1016/j.margen.2016.06.003.

47. Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, von Schalburg KR, Lemon C, Bird NH, et al. The genome and linkage map of the northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS One. 2014;9:e102089.

48. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, et al. Zebrafish hox clusters and vertebrate genome evolution. Science. 1998;282:1711–4.

49. Mungpakdee S, Seo HC, Angotzi AR, Dong X, Akalin A, Chourrout D. Differential evolution of the 13 Atlantic salmon Hox clusters. Mol Biol Evol. 2008;25:1333–43.

50. Ma B, Jiang H, Sun P, Chen J, Li L, Zhang X, Yuan L. Phylogeny and dating of divergences within the genus Thymallus (Salmonidae: Thymallinae) using complete mitochondrial genomes. Mitochondrial DNA A DNA MappSeq Anal. 2015;27:3602–11.

51. Stefansson SO, Björnsson BT. Ebbesson, LOE, McCormick SD. Smoltification. In: Finn N, Kappor BG, editors. Fish larval physiology. Florida: CRC Press; 2008. p. 639–681.

52. Harada M, Yoshinaga T, Ojima D, Iwata M. cDNA cloning and expression analysis of thyroid hormone receptor in the coho salmon *Oncorhynchus kisutch* during smoltification. Gen Comp Endocrinol. 2008;155:658–67.

53. Kiilerich P, Kristiansen K, Madsen SS. Hormone receptors in gills of smolting Atlantic salmon, *Salmo salar*: expression of growth hormone, prolactin, mineralocorticoid and glucocorticoid receptors and 11beta-hydroxysteroid dehydrogenase type 2. Gen Comp Endocrinol. 2007;152:295–303.

54. Seidelin M, Madsen SS, Byrialsen A, Kristiansen K. Effects of insulin-like growth factor-I and cortisol on Na+, K+-ATPase expression in osmoregulatory tissues of brown trout (*Salmo trutta*). Gen Comp Endocrinol. 1999;113:331–42.

55. Seidelin M, Madsen SS. Endocrine control of Na+, K+-ATPase and chloride cell development in brown trout (Salmo trutta): interaction of insulin-like growth factor-I with prolactin and growth hormone. J Endocrinol. 1999; 162:127–35.

56. Balsa JA, Sánchez-Franco F, Pazos F, Lara JI, Lorenzo MJ, Maldonado G, Cacicedo L. Direct action of serotonin on prolactin, growth hormone, corticotropin and luteinizing hormone release in cocultures of anterior and posterior pituitary lobes: autocrine and/or paracrine action of vasoactive intestinal peptide. Neuroendocrinology. 1998;68:326–33.

57. Hakes L, Pinney JW, Lovell SC, Oliver SG, Robertson DL. All duplicates are not equal: the difference between small-scale and genome duplication. Genome Biol. 2007;8:R209.

58. Carretero-Paulet L, Fares MA. Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. Mol Biol Evol. 2012;29:3541–51.

59. Ramsey J, Schemske DW. Neopolyploidy in flowering plants. Annu Rev Ecol Evol Syst. 2002;33:589–639.

60. Smith JJ, Keinath MC. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. Genome Res. 2015;25:1081–90.

61. De Storme N, Mason A. Plant speciation through chromosome instability and ploidy change: Cellular mechanisms, molecular factors and evolutionary relevance. Curr Plant Biol. 2014;1:10–33.

62. Mehta TK, Ravi V, Yamasaki S, Lee AP, Lian MM, Tay BH, Tohari S, Yanai S, Tay A, Brenner S, Venkatesh B. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). Proc Natl Acad Sci U S A. 2013;110:16044–9.

63. Benton MJ, Donoghue PC. Paleontological evidence to date the tree of life. Mol Biol Evol. 2009;24:26–53.

64. Erwin DH, Laflamme M, Tweedt SM, Sperling EA, Pisani D, Peterson KJ. The Cambrian conundrum: early divergence and later ecological success in the early history of animals. Science. 2011;334:1091–7.

65. Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, et al. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nat Genet. 2013;45:415–21.

66. Macqueen DJ, Primmer CR, Houston RD, Nowak BF, Bernatchez L, Bergseth S, Davidson WS, Gallardo-Escarate C, et al. Functional Analysis of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. bioRxiv. doi: https://doi.org/10.1101/095737.

67. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 2012;1:18.

68. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

69. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29:1072–5.

70. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

71. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012;29:1969–73.

72. Drummond AJ, Ho SY, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. PLoS Biol. 2006;4:e88.

73. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30:2725–9.

74. Lanfear R, Calcott B, Ho SY, Guindon S. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol. 2012;29:1695–701.

75. Rambaut A, Suchard MA, Xie D, Drummond AJ. Tracer v1.6. 2014. http://beast.bio.ed.ac.uk/Tracer.

76. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–2.

77. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012;29:15–21.

78. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31:166–9.

79. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

80. R Core Team. R: A language and environment for statistical computing. 2013. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

81. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize implements and enhances circular visualization in R. Bioinformatics. 2014;30:2811–2.

82. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.

83. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22:1600–7.