# The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications

Jeramiah J. Smith and Melissa C. Keinath

*Department of Biology, University of Kentucky, Lexington, Kentucky 40506, USA*

It is generally accepted that many genes present in vertebrate genomes owe their origin to two whole-genome duplications that occurred deep in the ancestry of the vertebrate lineage. However, details regarding the timing and outcome of these duplications are not well resolved. We present high-density meiotic and comparative genomic maps for the sea lamprey (*Petromyzon marinus*), a representative of an ancient lineage that diverged from all other vertebrates ~550 million years ago. Linkage analyses yielded a total of 95 linkage groups, similar to the estimated number of germline chromosomes (1n ~ 99), spanning a total of 5570.25 cM. Comparative mapping data yield strong support for the hypothesis that a single whole-genome duplication occurred in the basal vertebrate lineage, but do not strongly support a hypothetical second event. Rather, these comparative maps reveal several evolutionarily independent segmental duplications occurring over the last 600+ million years of chordate evolution. This refined history of vertebrate genome duplication should permit more precise investigations of vertebrate evolution.

[Supplemental material is available for this article.]

It is generally accepted that the ancestral lineage of all or most extant vertebrates experienced two separate whole-genome duplications (WGD) during evolution. Each of these duplications doubled the number of chromosomes and are often invoked as major factors underlying the structure and function of extant vertebrate genomes (Abi-Rached et al. 2002; Kasahara 2007). These duplications are colloquially referred to as the 2R hypothesis (Holland et al. 1994; Hughes 1999). Signatures of ancient vertebrate WGDs were apparent in early investigations of vertebrate genomes. Based on genome size, chromosome morphology, and isozyme counts, Susumu Onho proposed as early as 1970 that at least one WGD must have occurred prior to the diversification of the ancestral amniote or tetrapod lineage (Ohno 1970). Support for the 2R hypothesis rests partially on phylogenetic analyses of gene families and discrete synteny groups that have retained a significant number of genes that presumably trace to two WGDs (Larsson et al. 2008; Canestro et al. 2009; Kuraku et al. 2009) (but see Hughes 1999; Asrar et al. 2013). Additional support for 2R comes from studies that examined the number and distribution of homologous segments within the human genome (Dehal and Boore 2005) among jawed vertebrates (gnathostomes) (Nakatani et al. 2007; Murat et al. 2012) and between gnathostomes and amphioxus (*Branchiostoma floridae*: a cephalochordate) (Abi-Rached et al. 2002; Putnam et al. 2008). Studies comparing tetrapod and fish genomes can reconstruct features of an ancestral genome that existed ~400 million years ago (Mya) (Nakatani et al. 2007; Murat et al. 2012) and studies using amphioxus can reconstruct features of an ancestral genome that existed >600 Mya (Putnam et al. 2008), but it is clear that these respective ancestors substantially pre- and post-date the vertebrate WGDs. As such, analyses of duplication patterns are confounded to some degree by fissions, fusions, and gene/segmental duplications that have occurred over deep evolutionary time, both prior to and following WGD events.

Lampreys provide a novel perspective on the deep evolutionary history of vertebrate genomes, having diverged from the majority of other vertebrates (i.e., the lineage that gave rise to the gnathostomes) shortly after the most recent WGD, ~550 Mya (Fig. 1). Comparative studies using lampreys can therefore provide critical perspective on the nature and timing of evolutionary events that occurred at or near the base of the vertebrate phylogeny, including WGDs. Analyses of the sea lamprey genome assembly revealed that the duplication content of the lamprey genome is similar to that of other vertebrates and detected patterns of conserved synteny consistent with duplication and extensive paralog loss in both lineages (Smith et al. 2013). Such data are consistent with the most recent duplication event having occurred prior to the divergence of lamprey and gnathostomes, predating the evolution of several anciently derived features such as jaws, paired appendages, and neuronal myelination (Smith et al. 2013). Analyses of a second lamprey genome (*Lethenteron japonicum*) provided further confirmation that the duplication content of lamprey genomes is similar to that of jawed vertebrates and suggested the possibility that the lamprey lineage had undergone a third whole-genome duplication, subsequent to 2R (discussed in more detail below) (Mehta et al. 2013). However, it is important to recognize that both of these studies relied on the conventional wisdom that gnathostomes have undergone two rounds of WGD and did not explicitly test alternative evolutionary models. Unfortunately, existing lamprey assemblies do not permit the resolution of chromosome-scale patterns of homology or the robust reconstruction of ancestral vertebrate chromosomes, which are necessary to test assumptions of the 2R hypothesis.

To better resolve the chromosome-scale structure of the lamprey genome, we constructed the first meiotic map for a lamprey species by restriction-site associated DNA sequencing (RAD-seq) (Miller et al. 2007; Baird et al. 2008; Amores et al. 2011) of siblings
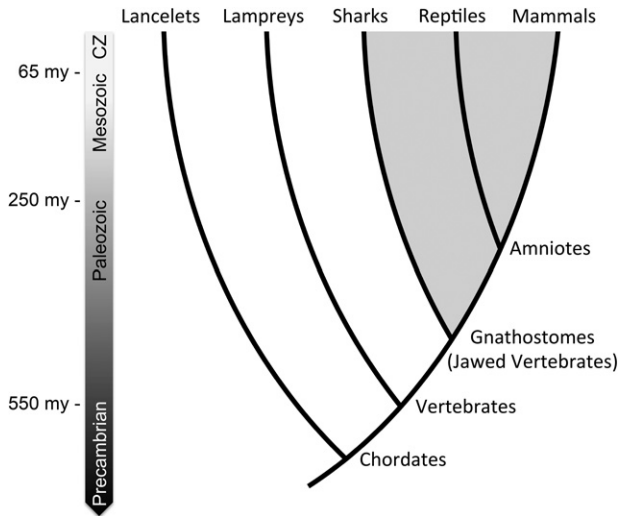
**Figure 1.** An abridged phylogeny of the vertebrate lineage. Labels denote taxa for which comparative analyses were performed. The approximate timing of key radiation events is shown to the *left*. Extinct lineages and some extant lineages (e.g., Ascidians, hagfish, coelacanth, lungfish) have been omitted for simplicity. Lancelets, lampreys, sharks, reptiles, and mammals are, respectively, represented by *Branchiostoma floridae* (Florida lancelet), *Petromyzon marinus* (sea lamprey), *Callorhinchus milii* (elephant shark), *Gallus gallus* (chicken), and *Homo sapiens* (human). (CZ) Cenozoic.

from a controlled cross between two wild-captured *Petromyzon marinus*. Integration of mapping data greatly improved the resolution of comparative maps, which provide strong support for a single ancient WGD but only weak support to a proposed second duplication. Comparative mapping data reveal the signatures of specific events (segmental duplications and translocations) that provide a simpler explanation for the patterns of synteny that have previously been referenced in support of the 2R hypothesis. The improved resolution of our study clarifies the duplication history of vertebrate genomes, and specifically the broadscale distribution of paralogous segments in gnathostome genomes.

## Results

### Linkage mapping and analysis

Our RAD-seq approach yielded 7215 segregating markers that could be directly assigned to a scaffold from the lamprey genome assembly (AEFG01; http://www.ncbi.nlm.nih.gov/genome/) or used to cross-validate and merge maternal and paternal maps. The ability to characterize large numbers of markers of a diverse segregation phase was facilitated in part by a relatively high frequency of segregating polymorphism, characteristic of *P. marinus* (Smith et al. 2013). This permitted

the construction of parent-specific and parent-averaged linkage maps. In total, 5275 markers could be confidently placed within linkage groups containing at least 10 markers linked at a minimum LOD score (log of odds) of 3.0, yielding a total of 95 linkage groups (LGs), which is similar to the estimated number of germline chromosomes (1n ~ 99) (Smith et al. 2010). These linkage groups span a total of 5570 cM of the parent-averaged map at an average marker density of one marker per 1.3 cM (Supplemental Figs. S1, S2; Supplemental Table S1). In total, 50% of orthology informative genes present in the somatic genome assembly (5483 of 10,891 in the lamprey/chicken comparative map) and 39% of all non-gap sequence (including 94 of the 100 longest scaffolds) could be placed on the 95 primary linkage groups. The high marker density and cross-validation of maternal and paternal genetic maps provided by this approach yielded robust chromosome-scale scaffolding of the lamprey genome assembly and comparative maps (Supplemental Tables S2, S3).

### Genome-wide patterns of conserved synteny

Lamprey/gnathostome (human and chicken) comparative maps reveal relatively strong conservation of chromosome-scale synteny across deep evolutionary time (Fig 2; Supplemental Fig. S3; Supplemental Table S4). These comparative maps also bear the signatures of large-scale duplication events, as a majority of the lamprey LGs that showed a significant enrichment of homologs from one gnathostome (chicken) chromosome also showed a significant enrichment of homologs from a second chicken chromosome (61%). This pattern was even more prominent among the 30 largest linkage groups, with 90% of these linkage groups showing statistically significant enrichment of orthologs from two or more chicken chromosomes. Likewise, 93% of chicken chromosomes
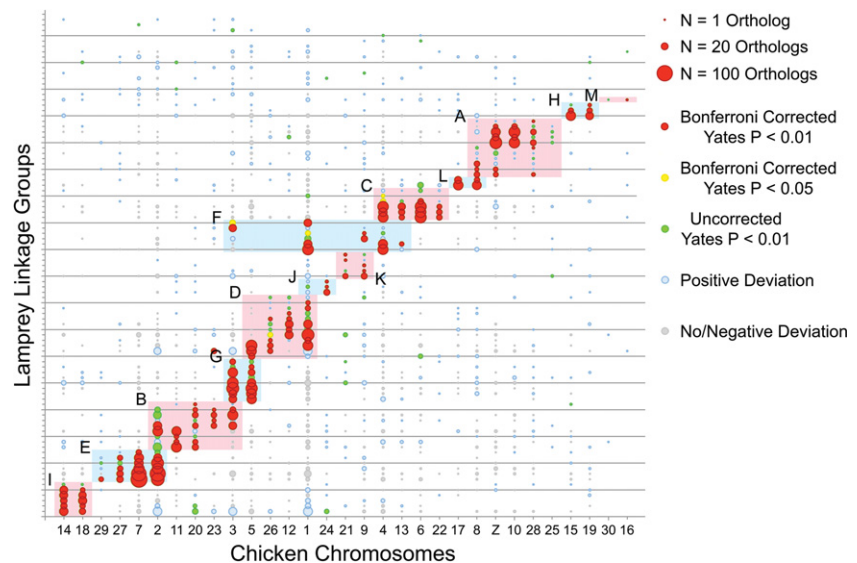


**Figure 2.** The chromosomal distribution of lamprey/chicken orthologs reveals conserved syntenic segments and paralogous chromosomes that derive from individual ancestral chromosomes. Lamprey linkage groups are oriented along the *y*-axis, and chicken chromosomes are oriented along the *x*-axis. Circles reflect counts of syntenic orthologs on the corresponding lamprey LG and chicken chromosome, with the size of each circle being proportional to the number of orthologous genes. The color of each circle represents the degree to which the number of observed orthologs deviates from null expectations under a uniform distribution across an identical number of LGs, chromosomes, and genes per LG and chromosome. Shaded regions of the plot designate homology groups that correspond to presumptive ancestral chromosomes, marked A–M (Supplemental Table S1). The ordering of lamprey LGs along the *y*-axis is provided in Supplemental Table S4.

that showed a significant enrichment of homologs from a lamprey LG showed a significant enrichment of homologs from a second lamprey LG. Figure 3 shows the distribution of presumptive orthologs and paralogs among one interrelated group of chicken and lamprey chromosomes (labeled "I" in Fig. 2). Similar patterns are also visible in the lamprey/human comparative map, albeit obscured by the effects of increased rates of intrachromosomal rearrangement in the mammalian lineage (Supplemental Fig. S3; Bourque et al. 2005; Smith and Voss 2006; Alfoldi et al. 2011; Voss et al. 2011; Amemiya et al. 2013). Given the assumption that selective convergence in chromosomal gene content is not a major driving force in vertebrate karyotype evolution, enrichment of homologs between chicken and lamprey is interpreted as strong evidence that the respective gnathostome and lamprey chromosomes/chromosomal segments are derived from the same individual ancestral chromosome.

The ancestral groupings that are resolved by the lamprey/gnathostome comparative maps largely recapitulate a set of 10 ancestral chromosomes previously proposed for the pre-1R ancestor (ancA-J) (Supplemental Table S5; Nakatani et al. 2007; Murat et al. 2012). In addition to these 10 previously identified chromosomes, the lamprey/chicken comparative map reveals three additional groupings that likely represent independent chromosomes from the presumptive pre-1R ancestor (Fig. 2). Comparative mapping data from amphioxus and elephant shark show signatures

of conserved synteny that are consistent with the existence of these ancestral chromosomes, although these genomes currently lack the long-range scaffolding information required to fully evaluate patterns on a chromosomal scale (Fig. 4). Regions of nonindependence between presumptive ancestral chromosomes (i.e., overlaps of highlighted regions on the x- or y-axes of Fig. 2) reflect fusion or translocation events that occurred in the 550 million years since the divergence of lamprey and gnathostome lineages or false mergers of linkage groups/scaffolds. The effect of these factors is relatively minor with respect to the lamprey/chicken comparative map, suggesting relatively low rates of interchromosomal rearrangement in both lineages and accuracy of ortholog assignments. Barring other patterns within the data, the observed relationship between ancestral and derived chromosomes could potentially result from a combination of large-scale duplication events and rearrangements (fissions, fusions, and translocations).

By examining the distribution of orthologous loci among ancestrally associated segments, it is possible to more precisely reconstruct the evolutionary history of derived chromosomes. In both chicken and lamprey, conserved syntenies from each ancestral chromosome are typically distributed across two or more paralogous chromosomes. With respect to the chicken genome, a majority of groups show a 1:2 ratio of ancestral:derived (a:d) chromosomes (see Fig. 2; e.g., $N = 7$ for ancG–M). Three other orthology groups (ancA, ancC, and ancE) show ~1:4 ratios, consisting of two macrochromosomes and two microchromosomes. The remaining three groups (ancB, ancD, and ancF) show more complex patterns consistent with rearrangements and small duplications having occurred subsequent to 1R. Notably, the corresponding chicken chromosomes show clear breaks in synteny, suggesting that these groups formerly constituted 1a:2d or 1a:3d ratios (Supplemental Fig. S4).

Comparative mapping data from amphioxus appear to support the interpretation that ancB, ancD, and ancF were shaped by post-WGD rearrangements (Fig. 4). Discreet substructures are apparent within these three ancestral chromosomes, providing additional evidence that these chromosomes were shaped by post-WGD fusions and translocations. Overall, we interpret the presence of an overarching 1:2 (or more) ratio of ancestral to derived chromosomes as consistent with a scenario wherein all ancestral chromosomes were affected by at least one WGD.

## Genomic distribution of orthologs and paralogs

To further evaluate the patterns of duplication that are revealed by lamprey comparative maps, we examined the distribution of orthologies within lamprey and chicken derivatives of the 13 ancestral chromosomes. The extensive loss and (rarer) retention of paralogs yield distinctive genomic signatures in the conserved syntenic structure of duplicated chromosomes. The independent nature of paralog loss results in a situation wherein two derived chromosomes retain a unique subset of ancestral genes, which form an interdigitated pattern when mapped to their homologous (unduplicated) ancestral chromosome (Kellis et al. 2004; Amores et al. 2011; Smith et al. 2013). Lamprey/gnathostome comparative maps show a slightly more complex pattern of conserved synteny. Specifically, gnathostome (chicken) chromosomes retain largely independent subsets of genes located on several lamprey chromosomes and similarly lamprey chromosomes retain largely independent subsets of genes located on several (typically two) gnathostome chromosomes (Figs. 2, 3; Supplemental Table S4). We interpret such a pattern as consistent with divergence of basal
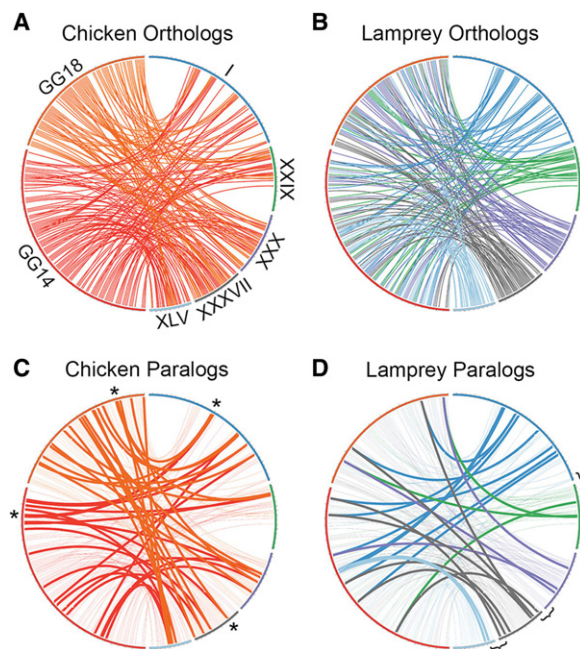


**Figure 3.** Distribution of individual orthologous genes across paralogous chicken and lamprey chromosomes. (*A*) Orthologs from chicken chromosomes GG14 and GG18 are distributed across five lamprey LGs (the complete set of lamprey linkage groups with significant enrichment for orthologs on these two chicken chromosomes). (*B*) Orthologs from these LGs are distributed across GG14 and GG18 (the complete set of chicken chromosomes with significant enrichment for orthologs on these lamprey linkage groups). (*C*) Orthologies plotted in *A*, with retained chicken paralogs denoted by bold lines. (*D*) Orthologies plotted in *B*, with retained lamprey paralogs denoted by bold lines. Roman numerals designate lamprey LGs. Asterisks mark the location of four conserved paralogs that are derived from a single gene located on ancI. Brackets denote presumptive ancestral linkages that are supported by the distribution of paralogous genes.
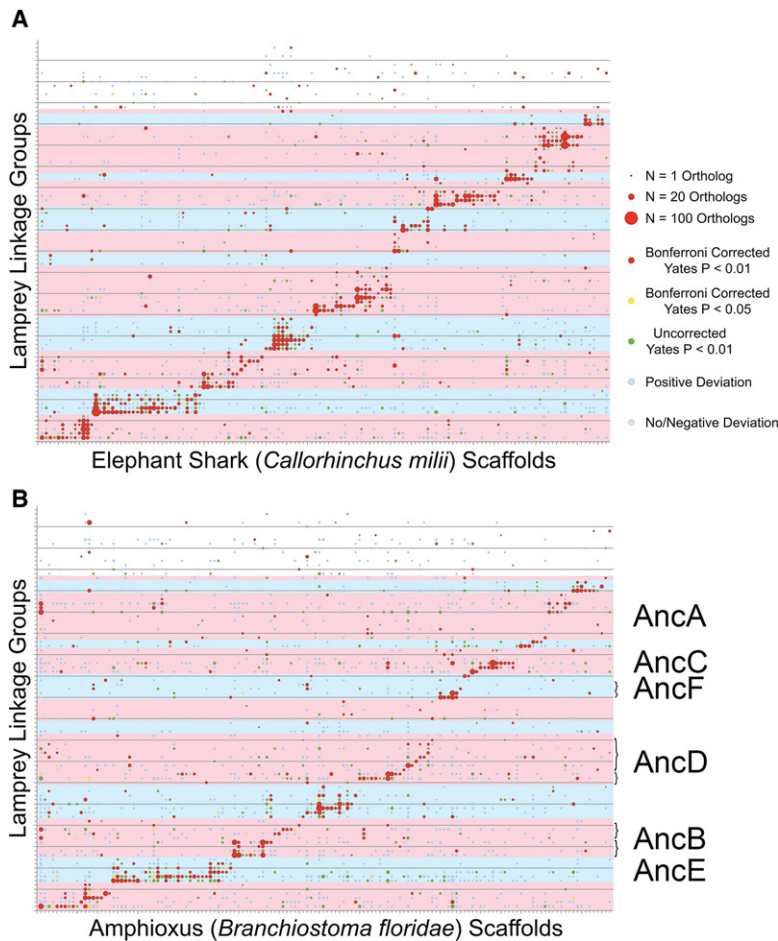
**Figure 4.** Comparative mapping with amphioxus and elephant shark reveals conserved syntenic segments that provide additional support for the proposed set of ancestral (pre-duplication) chromosomes. (*A*) The distribution of orthologs across lamprey linkage groups and elephant shark scaffolds. (*B*) The distribution of orthologs across lamprey linkage groups and amphioxus scaffolds. Lamprey linkage groups are oriented along the *y*-axis and reference scaffolds are oriented along the *x*-axis. Circles reflect counts of syntenic orthologs on the corresponding linkage group and scaffold, with the size of each circle being proportional to the number of orthologous genes. The color of each circle represents the degree to which the number of observed orthologs deviates from null expectations under a uniform distribution across an identical number of LGs, chromosomes, and genes per LG and chromosome. Shaded regions of the plot designate homology groups that correspond to presumptive ancestral chromosomes, marked A–M (Supplemental Table S1). The ordering of lamprey LGs along the *y*-axis is identical to Figure 2 and is provided in Supplemental Table S4. Brackets in *B* denote discreet sets of orthologous segments that lend support to post-WGD rearrangements.

chicken ($N = 503$) genomes indicate that lamprey and gnathostome genomes retain similar numbers of ancestrally syntenic paralogs, consistent with previous studies (Smith et al. 2013). The distribution of retained paralogs in lamprey and chicken genomes can be further leveraged to resolve fission events that have fractured chromosomes subsequent to an ancestral WGD (Fig. 3D); such fission events appear to have contributed significantly to the karyotypic evolution in the ancestral lineage that gave rise to *P. marinus*, which has a karyotype that is similar to all other members of the family Petromyzontidae (Hardisty 1971).

## Hypothesis testing

The identification of ancestral chromosomes and their derivative genomic segments provides a framework for testing hypotheses as to the origin of duplicated segments in gnathostome genomes (Fig. 5). Below we address three previously proposed explanations for the distribution of paralogous regions in gnathostome genomes: (A) two rounds of WGD (either before or after the divergence of the lamprey lineage) with essentially random loss of duplicated genes between/among derived chromosomes (Ohno 1970; Kasahara 2007), but not necessarily relative to biological function (Davis and Petrov 2004; Brunet et al. 2006; Putnam et al. 2008); (B) two rounds of WGD followed by extensive loss of large duplicated segments/chromosomes (Nakatani et al. 2007; Murat et al. 2012), (C) only segmental duplication (Asrar et al. 2013); and a fourth hypothesis that arose from examination of lamprey comparative maps (D) segmental duplications pre- or post-dating a single WGD. We reasoned that the plausibility of models B–D might be tested by measuring the degree to which the distribution of losses or seg-

lamprey and gnathostome lineages shortly after WGD. Under this model, paralog losses would have occurred through a largely independent series of mutational events in the two lineages (i.e., shared duplication and independent divergence).

The chromosome-wide distribution of retained paralogs (within species) also supports a scenario of shared duplication and independent divergence. The lamprey linkage map (in conjunction with chromosome-scale data from chicken) allows us to directly trace the ancestry of paralogs to their pre-duplication chromosome, thereby confirming their evolutionary origin via large-scale duplication. We detect synteny-supported paralogs on lamprey and chicken derivatives of all proposed ancestral chromosomes, with the exception of the apparently gene-poor ancM. Given that ~1/2 of the lamprey genome is anchored to the linkage map, counts of retained paralogs with lamprey ($N = 266$) and

mental duplications across ancestral orthology groups conforms to expected frequencies under a simple random mutational model (i.e., distributed Poisson for rare events). All four models and corresponding tests are summarized in Figure 5.

Observed ratios of ancestral to derived chromosomes (a:d) deviate strongly from expectations under the strict 2R model, even given a liberal interpretation that all 1a:>2d ratios ($N = 5$) are supportive of 2R (Supplemental Table S6). As mentioned above, large-scale chromosome loss has been invoked in some reconstructions as a significant contributing factor to post-2R assortment of chromosomal segments in order to explain observed deviations from 1a:4d (Nakatani et al. 2007; Murat et al. 2012). Depending on the timing of loss relative to 1R and 2R, the number of losses that are necessary to generate the observed patterns ranges from nine to 18 (see Methods for additional details). Corresponding
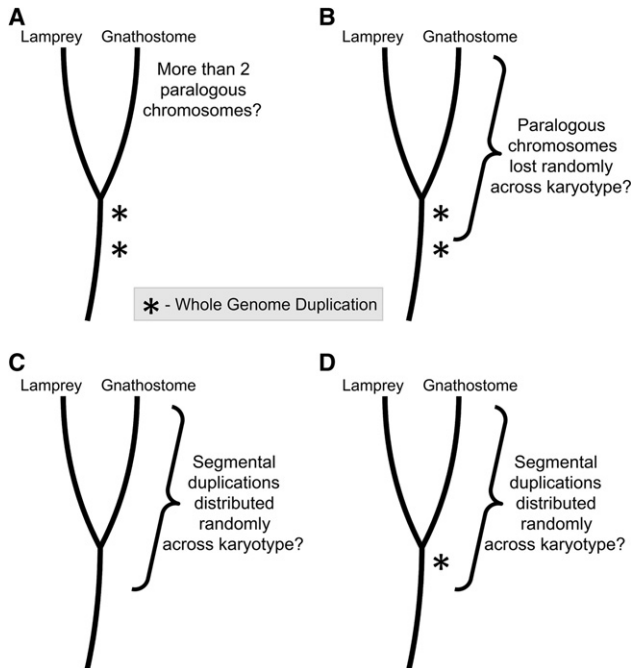
**Figure 5.** A summary of statistical tests aimed at assessing the feasibility of several evolutionary models that have been proposed to explain the distribution of paralogous regions in gnathostome genomes: (*A*) A liberal test of a simple model invoking two rounds of whole-genome duplication (WGD). (*B–D*) Tests of alternate scenarios wherein chromosome losses or segmental duplications occur in the context of a specified number of WGD events. (*B*) Two rounds of WGD followed by extensive loss of large duplicated segments/chromosomes. (*C*) A model invoking only segmental duplication. (*D*) A model invoking segmental duplications pre- or post-dating a single WGD. Asterisks denote whole-genome duplication events.

statistical tests for goodness of fit reveal that the distribution of deletion events across ancestral chromosomes is inconsistent with expected patterns, under both the liberal and conservative mutational models (Supplemental Table S7). However, by considering all possible permutations of liberal and conservative counting schemes across ancestral chromosomes, it is possible to identify a range of scenarios wherein observed counts do not reject a random mutational model. The best fit to the "2R plus deletion" model requires 12 chromosome losses, with three convergent losses of paralogous chromosomes post-2R (Supplemental Table S7). The simplest model failing to reject an underlying Poisson distribution at $P < 0.05$ involved 10 chromosome losses, with one parallel loss.

A similar scheme can be used to assess the distribution of duplications under a scenario wherein observed duplications are strictly the product of sporadic duplication events that individually affect only single chromosomes or large chromosomal segments. As above, the numbers of duplication events necessary to recover patterns in the lamprey/gnathostome comparative map were estimated using conservative (down-weighting recurrent events) and liberal (up-weighting recurrent events) schemes. Under the "segmental duplication only" model, all counting schemes and permutations thereof yield distributions that reject a random mutational model (Supplemental Tables S8, S9). Thus, segmental/chromosomal duplication alone appears to be insufficient to explain observed patterns of duplication.

Finally, this same logical framework was also used to test the fit of distributions deriving from a single WGD event, with segmental duplications preceding or following the event. After accounting for 1R, the distribution of remaining presumptive segmental duplicates is consistent with (i.e., fails to reject for all counts and permutations) models invoking sporadic segmental/chromosomal duplication (Supplemental Tables S8, S10).

In summary, two classes of models failed to reject expectations under a model of random mutation: (1) those invoking two rounds of whole-genome duplication and between 10 and 16 independent chromosome losses, and (2) those invoking a single round of whole-genome duplication and five to six segmental duplications or derived translocation events.

## Discussion

Taken together, analyses of lamprey/gnathostome comparative maps resolve and refine the complement of chromosomes that existed in the pre-1R vertebrate ancestor and support an evolutionary scenario wherein the divergence of two basal vertebrate lineages (that respectively gave rise to lampreys and gnathostomes) occurred shortly after a single whole-genome duplication event. Although comparative maps do reveal a clear global signature of one whole-genome duplication event, we do not observe global signatures consistent with a second WGD in the ancestry of the gnathostome lineage. Rather, these comparative maps suggest alternate evolutionary origins of paralogous regions that have been previously cited in support of the 2R duplication. Specifically, these patterns of conserved synteny are consistent with previously cited examples of fourfold conserved synteny (including *Hox* paralogy regions) being the product of ancient segmental duplication, followed by a single round of WGD.

### Chromosome loss versus segmental duplication

As described above, we interpret patterns that are visually apparent in the lamprey/chicken comparative map as consistent with being generated by one WGD with additional large paralogy regions being the product of rare segmental duplications occurring both before and after WGD. Moreover, the distribution of duplicated segments under this scenario is consistent with expectations given a simple random mutational model. Other scenarios invoking two rounds of WGD followed by chromosomal loss also yield counts that are consistent with a random mutational model, but require a substantially larger number of mutational steps to conform to the model. Specifically, models invoking one WGD require as few as six mutational steps (one WGD plus five segmental duplications/fissions), whereas models invoking two WGD events require between 12 and 18 steps (14 for the optimal permutation: two WGD events and 12 deletions). A scenario of one WGD pre- and post-dated by sporadic segmental duplications appears to provide the most parsimonious explanation for the distribution of large paralogous segments across gnathostome genomes. Below we discuss some of the salient evolutionary details of this scenario and relate these to previous observations that were conceptualized under the 2R hypothesis.

Our analyses indicate that both WGD and large segmental duplications likely played key roles in the deep history of vertebrate genome evolution. Among the duplication and fission events apparent in the comparative map, one recurrent pattern is particularly striking. All of the statistically supported groupings of ancestrally associated chicken chromosomes that deviate from a

pattern of 1a:2d chromosomes share a common feature in that they involve groups of macro- and microchromosomes. Previous studies have shown that microchromosomes are present in most major gnathostome lineages and that they represent distinct evolutionarily conserved entities at least to the base of the tetrapod lineage (~350 Mya) (Voss et al. 2011; Uno et al. 2012). Several of these microchromosomes: GG28 (ancA), GG20, 23 (ancB), GG13, 22 (ancC), GG26 (ancD), and GG27, 29 (ancE), bear the signature of duplication and fission events that are temporally distinct from 1R. The "single WGD" model identifies three sets of macro- and microchromosomes that experienced duplication prior to 1R (specifically, derivatives of ancA, ancC, and ancE). Notably, these three proposed duplications account for essentially all of the fourfold paralogous conserved syntenies that have been classically studied in the context of the 2R hypothesis, including large synteny groups that are exemplified by paralogs of *Hox* and *RAR* (ancE), *MHC* and *ALDH1* (ancA), and *NPYR* (ancC) loci (Larsson et al. 2008; Canestro et al. 2009; Kuraku et al. 2009). Previous reconstructions of the *ALDH1*-syntenic region also strongly implicate a pre-1R intrachromosomal duplication of ancA followed by chromosomal fission (Canestro et al. 2009), consistent with the single WGD model.

Patterns of conserved synteny in ancC and ancE suggest plausible mechanisms underlying the pre-1R duplication of these 1a:4d orthology groups that are distinct from those previously described for ancA (Canestro et al. 2009). The observed pattern for ancE is consistent with one of the expected segregation products following a Robertsonian fusion between an ancestral *Hox*-bearing chromosome and another (presumably larger) chromosome. Following such a fusion, normal disjunctions can give rise to two chromosomes that are respectively similar to the ancE-derived micro- (GG27, 29) and macro- (GG2, 7) chromosomes, with the larger possessing material duplicated from the smaller. A similar mechanism seems equally viable for the ancE (*NPYR*-bearing) chromosomes, although other mutations could yield similar patterns of duplication (Moore and Best 2001). The proposed scenario of duplication of the ancestral *Hox*-bearing chromosome early in the course of vertebrate evolution (before 1R) is consistent with the presence of four *Hox* clusters in gnathostomes and recent studies either supporting (Smith et al. 2013) or confirming (Mehta et al. 2013) the existence of more than two paralogous *Hox* clusters in lamprey genomes. The segmental duplication model is therefore consistent with both well-defined mutational mechanisms and studies confirming the existence of paralogous conserved syntenies shared by gnathostome and lamprey lineages. Altogether, we interpret patterns of conserved synteny between lamprey and gnathostomes as strongly supporting the occurrence of a single WGD predating the lamprey/gnathostome split, with other paralogous regions (in excess of 1a:2d) being the product of a small number of independent segmental duplications.

The above interpretation provides a simple explanation that integrates observations from lamprey/gnathostome comparative maps with previous analyses of vertebrate syntenic regions. However, we acknowledge that the simplest explanation may not always be correct. If microchromosome-associated duplications are, in reality, the product of a gnathostome-specific WGD (i.e., 2R), then the evolutionary assortment of post-2R chromosomes must have progressed in a manner that is substantially different from other vertebrate whole-genome duplication events (i.e., in the *Xenopus* lineage) [Evans et al. 2004; Uno et al. 2013], in the salmonid lineage [Berthelot et al. 2014], or near the base of the teleost fish lineage [Amores et al. 2011]). Under a hypothetical

2R, the evolutionary assortment of post-2R chromosomes would have involved large-scale loss of chromosomes and chromosomal segments, rather than the gradual diversification and nearly random mutational loss of paralogs. Analyses of other vertebrate lineages that have undergone more recent large-scale duplication and chromosomal loss might ultimately reveal a candidate mechanism for post-2R loss. Studies of recently duplicated plants have revealed that some chromosomes may retain fewer single-copy paralogs than their paralogous chromosomes (Woodhouse et al. 2010; Schnable et al. 2011; Garsmeur et al. 2014; Renny-Byfield et al. 2015). Taken to an extreme, such biases could conceivably mimic the apparent chromosome losses necessary to support the 2R model.

It is also notable that our analyses treat chromosome losses and segmental duplications as functionally equivalent with respect to their mutational effect. It is important to recognize, however, that the fitness effects of a specific mutation (whole-genome duplication, deletion, segmental duplication, or otherwise) are contingent on the genetic background and environment in which the mutation occurred, with the probability of fixation being further contingent on the population structure. Studies examining the effects of segmental duplication and chromosome loss across diverse vertebrate taxa will ultimately provide some insight into the relative probabilities of fixation, but it seems certain that these parameters will never be known for the ancestral vertebrate lineage. Given the relative simplicity of incorporating previous observations into the model, we assert that the distribution of paralogous segments in vertebrate genomes is currently best explained by one WGD and the evolution of microchromosome-associated paralogy regions via segmental duplication (Fig. 6).

## IR in the context of previous studies

Although the evolutionary model inspired by analysis of the lamprey linkage map differs from more classical versions of the 2R model that have been supported by other studies, the "1R plus segmental duplication" model does not seem to be at odds with the primary findings of those studies. For example, studies examining the distribution and depth of syntenic paralogous regions in the
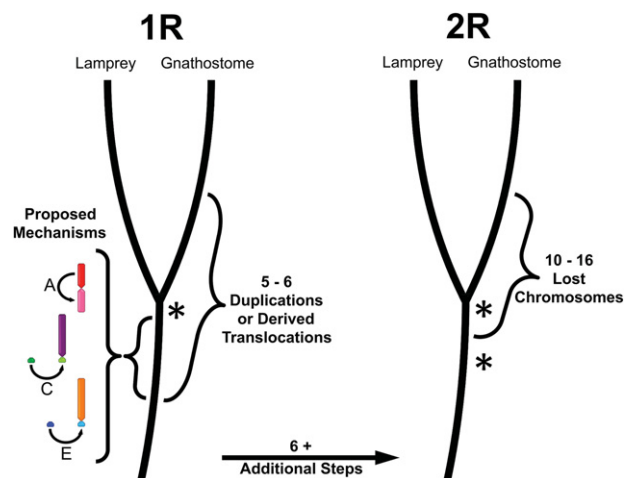


**Figure 6.** A summary of alternate evolutionary models explaining the distribution of paralogous segments in gnathostome genomes. Asterisks denote whole-genome duplication events proposed under two alternate sets of evolutionary models. Mechanisms underlying three pre-R1 duplications are depicted under the "1R" model.

human genome estimated that each region of the human genome contained paralogous copies that existed in at least three other distinct locations (Dehal and Boore 2005; Putnam et al. 2008). Our analyses reveal that extensive intrachromosomal rearrangement has occurred post-1R (Supplemental Fig. S4) and corroborate numerous studies demonstrating that rates of interchromosomal rearrangement are much higher in mammals than in other vertebrate lineages (Supplemental Fig. S3; International Chicken Genome Sequencing Consortium 2004; Smith and Voss 2006; Alfoldi et al. 2011; Voss et al. 2011; Amemiya et al. 2013). As discussed above, duplication and fission/translocation events can both distribute orthologous segments in similar patterns, and the same holds true for paralogous segments within a genome. It seems clear that interchromosomal rearrangement has played a major role in structuring the distribution of conserved syntenic regions between pre-1R ancestral chromosomes and their human derivatives, as can be seen by comparing Figure 2 and Supplemental Figure S3. Thus, patterns of conserved synteny in the human genome do not seem to be at odds with the findings of the current study.

Other studies have leveraged comparisons among gnathostomes (Nakatani et al. 2007; Murat et al. 2012) or between gnathostomes and chordates (Putnam et al. 2008) to reconstruct pre-1R orthology groups. Notably, pre-1R chromosomes reconstructed via the lamprey meiotic map closely resemble those proposed by these previous studies. However, these studies did not explicitly address alternate evolutionary models or departures from 1a:4d (expected under 2R). Still other studies have focused on a smaller number of genomic regions that are thought to represent the largest and best-conserved regions of fourfold paralogy (Larsson et al. 2008; Canestro et al. 2009; Kuraku et al. 2009), although the existence of these well-defined regions is consistent with several proposed models. Finally, two recent studies used draft lamprey genomes to examine the relative timing of WGD and divergence events, but relied on the assumption that gnathostome genomes were indeed the product of two rounds of whole-genome duplication (Mehta et al. 2013; Smith et al. 2013). In general, the "1R plus segmental duplication" model appears to be consistent with the primary findings of all studies of 2R performed to date (discussed in detail below), though it seems that this alternate model was not readily apparent in the absence of chromosome-scale data from lamprey.

### Ancestral versus independent duplication

Two recent studies (Mehta et al. 2013; Smith et al. 2013) provide strong evidence that ancient WGD(s) impacted the lamprey genome, and that subsequent paralog losses occurred through a largely independent series of mutational events in lamprey and gnathostome lineages. However, these studies differ in their interpretation of duplication patterns relative to the divergence of the basal lineages that gave rise to lampreys and gnathostomes. It should be noted that the timing of these ancient duplication and divergence events does not strongly affect the identification of ancestral chromosomes in the current study. The observation of nonindependent distributions of paralog losses was considered evidence that duplication preceded divergence (Smith et al. 2013), whereas differences in the 4DTv (transversions at fourfold degenerate sites) distributions for lamprey versus gnathostome paralogs were considered evidence for independent duplication in the lamprey lineage (Mehta et al. 2013). The current study verifies that estimates of paralog retention/loss rates in anc-derived paralogous

regions are consistent with those estimated from the entire genome assembly, lending some support to the idea that divergence followed duplication. Moreover, it should be noted that estimates of 4DTv are highly contingent on patterns of substitution (Tang et al. 2008) and that long-term substitution bias in the lamprey lineage is known to have driven the lamprey genomes to exceedingly high GC content, especially within coding regions (Qiu et al. 2011; Mehta et al. 2013; Smith et al. 2013). This bias (although not the GC content itself) is seemingly sufficient to explain variation in 4DTv between lampreys and gnathostomes. Indeed, greater degrees of intraspecific variation in 4DTv have been observed among dicot plants in the wake of the ancestral γ hexaploidization event (Tang et al. 2008). We interpret the bulk of available evidence as indicating that 1R (and a finite number of large segmental duplications) predated the divergence of the basal lineages that gave rise to lampreys and gnathostomes, with a relatively small number of large segmental duplications occurring subsequent to 1R. It seems possible, however, that material deleted from somatic cells during programmed genome rearrangement (Smith et al. 2009, 2012) (and therefore not represented in the current study), could be the product of large-scale duplication (or more limited duplication) in the ancestral lamprey/cyclostome lineage. We anticipate that the development of genome assemblies for other taxa (especially germline genomes from hagfish and divergent lamprey species) will improve the temporal resolution of ancestral duplication and divergence events.

## Conclusions

By resolving the deep evolutionary history of vertebrate genomes, analyses leveraging the anchored lamprey genome provide new context for understanding the ancestry and evolutionary diversification of vertebrates. Although the most common fate of duplicated genes is mutational degradation of one copy (paralog loss), gene duplication also provides raw material for the evolution of new gene functions and the evolutionary tuning of old functions (Ohno 1970; Taylor and Raes 2004; Conant and Wolfe 2008; Hahn 2009); thus it seems likely that these duplication events have had a major impact on the early evolutionary trajectory of the vertebrate lineage and establishment of the basal condition from which all vertebrates evolved. We further anticipate that incorporation of a more accurate duplication history into future studies and revision of previous evolutionary studies that have conceptualized vertebrate evolution in light of other hypotheses (including 2R) will provide a more robust understanding of the functional evolution of vertebrate genomes and aid in translating information between vertebrates and nonvertebrate species.

This study also underscores the importance of generating genome assemblies and incorporating chromosome-scale scaffolding data for divergent vertebrate lineages (including amphioxus, hagfish, other lamprey species, and sharks). We anticipate that such data will shed additional light on the deep evolutionary history of vertebrate karyotypes and permit more robust tests of the alternate evolutionary models presented here.

## Methods

### RAD-seq genotyping

Lamprey embryos were produced by in vitro fertilization (Nikitina et al. 2009) of eggs from a single female with sperm from a single male. DNA was extracted from 141 lamprey embryos (21 d

post-fertilization) via standard phenol/chloroform extraction (Sambrook and Russell 2001) and evaluated by gel electrophoresis. A total of 30 samples, containing highly intact DNA were selected for RAD-seq, along with parental DNAs that were extracted from muscle tissue. RAD-seq was performed by Floragenex, Inc., yielding a total of 147 million 95-bp sequence fragments (3,590,251–5,489,313 per individual) anchored to SbfI restriction sites. The *Stacks* software package (Catchen et al. 2011) was used to identify segregating polymorphisms and generate genotype calls for parents and offspring. For this study, the average depth of coverage per locus was 187.3 reads (range 156.0–217.3 per individual). Because sequence information is generated from DNA extracted from whole embryos, we expect to primarily sample DNA from somatic cells. Therefore, the current data set is missing ∼20% of the germline genome and 1–2000 genes that are specifically retained in the genome of lamprey germ cells (Smith et al. 2009, 2012). Similarly, the method may also fail to produce genotypes from genomic regions with high insertion/deletion frequencies. Given the relatively small number of genes that are eliminated by programmed genome rearrangement, we do not expect exclusion of these regions to systematically bias our comparative analyses, especially as they relate to the distribution of duplicated segments in gnathostome genomes.

## Linkage analysis

Linkage analysis was performed via maximum likelihood mapping using JoinMap software package and default parameters, except that the number of optimization rounds was increased from three to five to ensure accurate internal ordering of large numbers of tightly linked markers (Stam 1993; Van Ooijen 2011). In order to circumvent limitations of the software package related to computational overhead, efficiently anchor the map to the existing assembly, and permit robust integration of female and male maps, we limited our analyses to markers that (1) aligned to the published genome assembly (>95% identity over ≥90 bp via BLAST) (Altschul et al. 1990), (2) yielded informative segregation phasing, and (3) yielded genotypic information for at least 20 of 30 offspring. These were further supplemented with markers that were specifically informative for one or both parents, regardless of alignment to the genome assembly, permitted that they were genotyped for at least 27 of 30 offspring for biallelic markers (llxlm, nnxnp, or hkxhk) (Maliepaard et al. 1997; Van Ooijen 2011) or at least 25 of 30 offspring for tri- and tetra-allelic markers (efxeg or abxcd) (Maliepaard et al. 1997; Van Ooijen 2011). The maximum likelihood mapping algorithm for mapping of a full sib family from outbred parents (CP) (Van Ooijen 2011) was used to generate a linkage map from a total of 7215 potentially informative markers. Linkage groups were manually curated to break linkages at >30 cM, except in four cases where markers targeting alternate alleles of the same locus were within 40 cM in both parental maps. Rarely, markers mapping to a single scaffold were assigned to two different linkage groups. These discrepancies could be due to misassemblies in the draft somatic genome, genotyping errors, or programmed genome rearrangements. A majority of these (totaling 2.8% of mapped scaffolds) were readily resolved by majority rule, with multiple mapped SNPs supporting an assignment to a specific linkage group. A smaller fraction of scaffolds could not be resolved by majority rule (1.5%) and were placed arbitrarily. If misplaced, these scaffolds would be expected to contribute to (low level) background noise in the comparative maps, but when placed properly they should contribute to the signal of conserved synteny (again to a minor degree). Markers not incorporated into the map may represent distal portions of chromosomes, uninformative segregation phases, or genotyping errors.

## Statistical analysis of conserved syntenic regions

Methods for identifying putatively orthologous loci are described elsewhere (Smith et al. 2013). Lamprey gene annotations used for this study were previously published and can be downloaded and browsed through the UCSC Genome Browser, http://genome.ucsc.edu/ (Kent et al. 2002; Smith et al. 2013). Operationally, a locus from the reference genome is considered to be orthologous to a lamprey gene if all three of the following conditions are met: (1) the TBLASTN (Altschul et al. 1990) alignment bitscore between the lamprey gene and a given locus from the reference genome is within 90% of the best alignment bitscore for that lamprey gene, (2) there are six or fewer paralogs detected in the lamprey genome, and (3) there are six or fewer paralogs detected in the reference genome. This method circumvents many of the confounding effects of extreme substitution bias in the lamprey lineage, independent paralog loss between gnathostome and lamprey lineages, and the proximity of duplication and divergence events, all of which prevent accurate tree-based orthology assignment (Qiu et al. 2011; Mehta et al. 2013; Smith et al. 2013). In order to reflect the biological reality of whole-genome duplication and subsequent evolutionary assortment of paralogs, "orthologs" refer to the set of loci that share a most recent common ancestor in the preduplicated/predivergence state.

Counts of orthologs on all pairwise combinations of lamprey LGs and gnathostome (chicken: GCA_000002315.1 or human: GCA_000001405.1, http://www.ncbi.nlm.nih.gov) chromosomes or genomic scaffolds (amphioxus: ftp://ftp.jgi-psf.org/pub/JGI_data/Branchiostoma_floridae/v1.0/ Branchiostoma_floridae_v2.0.assembly.fasta.gz or elephant shark: GCA_000165045.2) were tabulated and compared with expected values based on random sampling from LGs and chromosomes (or scaffolds) with the same number of loci. Because these comparisons involve a large number of pairwise combinations of lamprey LGs and gnathostome chromosomes that often possess a small number of putative orthologs (i.e., most cells in the contingency table correspond to nonorthologous segments, especially in comparisons between species with extensive conservation of synteny), the distribution of orthologs was evaluated using a $\chi^2$ incorporating Yates' correction for continuity and Bonferroni corrections for multiple testing. Notably, the locations of conserved syntenic regions with marginal statistical support (Yates' $P < 0.01$ and Bonferroni corrected Yates' $P > 0.05$) are heavily biased to proposed duplicates of ancestral chromosomes. In total, 40 marginally supported regions fell within 143 remaining pairs of chromosomes that were associated with an ancestral chromosome at Bonferroni corrected Yates' $P > 0.05$. In contrast, 12 marginally supported regions fell within 2136 pairs of chromosomes that were not associated with an ancestral chromosome. Marginally supported syntenies are therefore heavily enriched among chromosomes that are derived from our 13 proposed ancestors (49.8× more common), implying that few paralogous segments have escaped detection in the lamprey/chicken comparative map.

Hypotheses as to mode of gene duplication were evaluated using goodness of fit tests; however, it should be noted that the small number of ancestral chromosomes presents a fundamental limitation to the power of these tests. Given that two rounds WGD are expected to yield ratios in excess of 1a:2d for chromosomes or chromosomal segments (expected values are zero for 1a:1d and 1a:2d), categorical test statistics become infinitely large for a test of goodness of fit to a model of 1a:4d. We therefore used a more conservative threshold of 1a:>2d for testing the strict "2R" hypothesis and applied a continuity correction to expected values in order to permit calculation of conservative $G$-tests ($G_{adj(c)}$) (Sokal and Rohlf 1995).

For models involving deletion or segmental duplication, expected distributions were based on Poisson distributions with mean values equivalent to the observed number of proposed events divided by the number of ancestral chromosomes. Two models were used to count the number of events necessary to generate observed ancestral:derived ratios for each chromosome: (1) a "conservative" model based on the minimal set of events necessary to generate the observed patterns (i.e., when all 1a:2d ratios are considered to be the product of deletion of a 1R paralog; Supplemental Table S8), and (2) a "liberal" count based on the maximal set of irreversible events that could potentially generate the observed patterns (i.e., when 1a:2d ratios are considered to be the product of recurrent deletions of different paralogous segments post 2R; Supplemental Table S8). Statistical tests of goodness of fit to models involving deletion and segmental duplication utilized a G-test with Williams' correction for a small sample size ($G_{adj(W)}$) (Sokal and Rohlf 1995).

## Data access

Sequence data from this study have been submitted to the NCBI BioProject (http://www.ncbi.nlm.nih.gov/bioproject) under accession number PRJNA232586.

## Acknowledgments

## References

Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat Genet* **31:** 100–105.

Alfoldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, et al. 2011. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477:** 587–591.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215:** 403–410.

Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N, et al. 2013. The African coelacanth genome provides insights into tetrapod evolution. *Nature* **496:** 311–316.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* **188:** 799–808.

Asrar Z, Haq F, Abbasi AA. 2013. Fourfold paralogy regions on human HOX-bearing chromosomes: role of ancient segmental duplications in the evolution of vertebrate genome. *Mol Phylogenet Evol* **66:** 737–747.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* **3:** e3376.

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noel B, Bento P, Da Silva C, Labadie K, Alberti A, et al. 2014. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* **5:** 3657.

Bourque G, Zdobnov EM, Bork P, Pevzner PA, Tesler G. 2005. Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res* **15:** 98–110.

Brunet FG, Roest Crollius H, Paris M, Aury JM, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M. 2006. Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol* **23:** 1808–1816.

Canestro C, Catchen JM, Rodriguez-Mari A, Yokoi H, Postlethwait JH. 2009. Consequences of lineage-specific gene loss on functional evolution of surviving paralogs: ALDH1A and retinoic acid signaling in vertebrate genomes. *PLoS Genet* **5:** e1000496.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)* **1:** 171–182.

Conant GC, Wolfe KH. 2008. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet* **9:** 938–950.

Davis JC, Petrov DA. 2004. Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol* **2:** E55.

Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3:** e314.

Evans BJ, Kelley DB, Tinsley RC, Melnick DJ, Cannatella DC. 2004. A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. *Mol Phylogenet Evol* **33:** 197–213.

Garsmeur O, Schnable JC, Almeida A, Jourda C, D'Hont A, Freeling M. 2014. Two evolutionarily distinct classes of paleopolyploidy. *Mol Biol Evol* **31:** 448–454.

Hahn MW. 2009. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered* **100:** 605–617.

Hardisty MW. 1971. *The biology of lampreys*. Academic Press, New York.

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. *Dev Suppl* 125–133.

Hughes AL. 1999. Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. *J Mol Evol* **48:** 565–576.

International Chicken Genome Sequencing Consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432:** 695–716.

Kasahara M. 2007. The 2R hypothesis: an update. *Curr Opin Immunol* **19:** 547–552.

Kellis M, Birren BW, Lander ES. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428:** 617–624.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Res* **12:** 996–1006.

Kuraku S, Meyer A, Kuratani S. 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol Biol Evol* **26:** 47–59.

Larsson TA, Olsson F, Sundstrom G, Lundin LG, Brenner S, Venkatesh B, Larhammar D. 2008. Early vertebrate chromosome duplications and the evolution of the neuropeptide Y receptor gene regions. *BMC Evol Biol* **8:** 184.

Maliepaard C, Jansen J, Van Ooijen JW. 1997. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. *Genet Res* **70:** 237–250.

Mehta TK, Ravi V, Yamasaki S, Lee AP, Lian MM, Tay BH, Tohari S, Yanai S, Tay A, Brenner S, et al. 2013. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). *Proc Natl Acad Sci* **110:** 16044–16049.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res* **17:** 240–248.

Moore CM, Best RG. 2001. Chromosomal genetic disease: structural aberrations. In *Encyclopedia of Life Sciences*, Wiley Online Library.

Murat F, Van de Peer Y, Salse J. 2012. Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes. *Genome Biol Evol* **4:** 917–928.

Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res* **17:** 1254–1265.

Nikitina N, Bronner-Fraser M, Sauka-Spengler T. 2009. The sea lamprey *Petromyzon marinus*: a model for evolutionary and developmental biology. In *Emerging model organisms: a laboratory manual*, Vol. 1, pp. 405–429. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ohno S. 1970. *Evolution by gene duplication*. Springer-Verlag, New York.

Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453:** 1064–1071.

Qiu H, Hildebrand F, Kuraku S, Meyer A. 2011. Unresolved orthology and peculiar coding sequence properties of lamprey genes: the KCNA gene family as test case. *BMC Genomics* **12:** 325.

Renny-Byfield S, Gong L, Gallagher JP, Wendel JF. 2015. Persistence of subgenomes in paleopolyploid cotton after 60 my of evolution. *Mol Biol Evol* **32:** 1063–1071.

Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc Natl Acad Sci* **108:** 4069–4074.

Smith JJ, Voss SR. 2006. Gene order data from a model amphibian (Ambystoma): new perspectives on vertebrate genome structure and evolution. *BMC Genomics* **7:** 219.

Smith JJ, Antonacci F, Eichler EE, Amemiya CT. 2009. Programmed loss of millions of base pairs from a vertebrate genome. *Proc Natl Acad Sci* **106:** 11212–11217.

Smith JJ, Stuart AB, Sauka-Spengler T, Clifton SW, Amemiya CT. 2010. Development and analysis of a germline BAC resource for the sea lamprey, a vertebrate that undergoes substantial chromatin diminution. *Chromosoma* **119:** 381–389.

Smith JJ, Baker C, Eichler EE, Amemiya CT. 2012. Genetic consequences of programmed genome rearrangement. *Curr Biol* **22:** 1524–1529.

Smith JJ, Kuraku S, Holt C, Sauka-Spengler T, Jiang N, Campbell MS, Yandell MD, Manousaki T, Meyer A, Bloom OE, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet* **45:** 415–421.

Sokal FJ, Rohlf RR. 1995. *Biometry: the principles and practice of statistics in biological research*. W.H. Freeman, NewYork.

Stam P. 1993. Construction of integrated genetic-linkage maps by means of a new computer package—joinmap. *Plant J* **3:** 739–744.

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res* **18:** 1944–1954.

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

Uno Y, Nishida C, Tarui H, Ishishita S, Takagi C, Nishimura O, Ishijima J, Ota H, Kosaka A, Matsubara K, et al. 2012. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS One* **7:** e53027.

Uno Y, Nishida C, Takagi C, Ueno N, Matsuda Y. 2013. Homoeologous chromosomes of *Xenopus laevis* are highly conserved after whole-genome duplication. *Heredity (Edinb)* **111:** 430–436.

Van Ooijen JW. 2011. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Res (Camb)* **93:** 343–349.

Voss SR, Kump DK, Putta S, Pauly N, Reynolds A, Henry RJ, Basa S, Walker JA, Smith JJ. 2011. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res* **21:** 1306–1312.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol* **8:** e1000409.