


# Analysis of Paralogons, Origin of the Vertebrate Karyotype, and Ancient Chromosomes Retained in Extant Species

Trevor D. Lamb \*

Eccles Institute of Neuroscience, John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia

\*Corresponding author: E-mail: Trevor.Lamb@anu.edu.au.

Accepted: 1 March 2021

## Abstract

A manually curated set of ohnolog families has been assembled, for seven species of bony vertebrates, that includes 255 four-member families and 631 three-member families, encompassing over 2,900 ohnologs. Across species, the patterns of chromosomes upon which the ohnologs reside fall into 17 distinct categories. These 17 paralogons reflect the 17 ancestral chromosomes that existed in our chordate ancestor immediately prior to the two rounds of whole-genome duplication (2R-WGD) that occurred around 600 Ma. Within each paralogon, it has now been possible to assign those pairs of ohnologs that diverged from each other at the first round of duplication, through analysis of the molecular phylogeny of four-member families. Comparison with another recent analysis has identified four apparently incorrect assignments of pairings following 2R, along with several omissions, in that study. By comparison of the patterns between paralogons, it has also been possible to identify nine chromosomal fusions that occurred between 1R and 2R, and three chromosomal fusions that occurred after 2R, that generated an ancestral bony-vertebrate karyotype comprising 47 chromosomes. At least 27 of those ancestral bony-vertebrate chromosomes can, in some extant species, be shown not to have undergone any fusion or fission events. Such chromosomes are here termed “archeochromosomes,” and have each survived essentially unchanged in their content of genes for some 400 Myr. Their utility lies in their potential for tracking the various fusion and fission events that have occurred in different lineages throughout the expansion of bony vertebrates.

**Key words:** vertebrate karyotype, chromosome evolution, 2R whole-genome duplication, archeochromosomes, micro-chromosomes, zebra finch.

## Significance

1) The vertebrate karyotype is postulated to have arisen from an ancestral chordate karyotype comprising 17 chromosomes, through two rounds of whole-genome duplication, but it remains unclear what chromosomal fusions and fissions may have occurred. 2) Through an analysis of quadruplicated genes in extant vertebrate species, the present study provides clear evidence for the existence of 17 such ancestral chromosomes, and in doing so it exposes the chromosomal fusions that occurred following the first and second rounds of genome duplication. A by-product of this analysis is the discovery that >20 chromosomes in two species (zebra finch and spotted gar) have undergone no fusions or fissions since bony vertebrates first appeared. 3) The study thereby provides a clear description of the transition (some 600–400 Ma) from an ancestral chordate karyotype of 17 chromosomes into the 47 chromosome karyotype of the first bony vertebrates.

## Introduction

It is arguable that the event of greatest importance in setting vertebrates apart from other animals was the successive occurrence, over a relatively short interval some 500–600 Ma, of two rounds of whole-genome duplication (2R-WGD), as

proposed 50 years ago by Susumu Ohno (1970). The resulting paralogs, in sets of up to four members, have been termed “ohnologs.” A substantial body of work supports the 2R-WGD hypothesis (Pébusque et al. 1998; Abi-Rached et al. 2002; Larhammar et al. 2002; Hokamp et al. 2003; Lundin

© The Author(s) 2021. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

et al. 2003; Dehal and Boore 2005; Nakatani et al. 2007; Putnam et al. 2008; Smith et al. 2013; Singh et al. 2015; Sacerdot et al. 2018; Singh and Isambert 2020; Simakov et al. 2020), though an alternative description of the second duplication, as arising from block duplications, has been proposed (Smith and Keinath 2015; Smith et al. 2018). An alternative scenario without any whole-genome duplications has been suggested by Naz et al. (2017). Although Ohno envisaged the first duplication to have occurred prior to the divergence of tunicates from our lineage (see Holland 1999), it is now clear that 1R occurred after that divergence.

A number of studies have attempted to reconstruct the ancestral “pre-1R” karyotype that existed in our chordate ancestor, immediately prior to the first round of genome duplication. Some of those investigations supported the existence of 10–13 pre-1R chromosomes (Kohn et al. 2006; Nakatani et al. 2007; Smith et al. 2013, 2018), whereas others found evidence for 17 pre-1R chromosomes (Putnam et al. 2008; Simakov et al. 2015, 2020; Sacerdot et al. 2018). The analysis of Sacerdot et al. (2018) began with the construction of an ancestral amniote karyotype, through analysis of the genomes of 61 species selected from Ensembl 69 (October 2012). That study then analyzed the ohnolog composition of the ancestral amniote karyotype, and extracted 17 “contiguous ancestral region tetrads” (CARTs), that showed similarity to the 17 ancestral “chordate linkage groups” (CLGs) previously identified by Putnam et al. (2008) in their investigation of the amphioxus genome. Sacerdot et al. (2018) concluded that they had identified the 17 pre-1R chromosomes; the relatively minor differences from the 17 CLGs of Putnam et al. (2008) were likely to have arisen from one fusion, one fission, and several translocations, during the interval between the divergence of lancelets and 1R.

Here, I adopted a very different approach from previous studies, by assembling a manually curated set of ohnolog quartets and trios, across the genomes of seven species of bony vertebrates. This curated ohnolog resource shows unmistakable grouping into 17 paralogons of ohnolog quadruplicates, where each paralogon corresponds one-to-one with the 17 pre-1R chromosomes. Crucially, for each paralogon, I have been able to determine which pair of ohnolog families diverged from which other pair at 1R, through analysis of sequence phylogeny for 34 ohnolog families that retain all four members.

One outcome of the present study is the identification in extant species of chromosomes that appear to have undergone not a single case of fusion or fission since the appearance of bony vertebrates, roughly 400 Ma. Thus, apart from the more-or-less random loss of individual genes or local regions, the overall complement of coding genes on these chromosomes appears to have remained essentially unchanged; such chromosomes are here termed “archeochromosomes.” They are prevalent in zebra finch and spotted gar, and in many cases are congruent with

microchromosomes (i.e., no longer than ~20 Mb). Between zebra finch and spotted gar, at least 26 of the 47 ancestral bony-vertebrate chromosomes can still be identified, essentially unchanged in their content of ohnologs.

These conclusions rely on analyses of “synteny” in the sense originally coined by Renwick (1971) as “presence together on the same chromosome”; they do not involve consideration of the order of genes. Throughout this article the term synteny will be used with that original meaning.

## Results

### A Curated Set of Ohnolog Quartets and Trios

I constructed a curated set of ohnolog families for a restricted set of species, chosen from amongst those with a chromosomal level assembly in Ensembl 101 (August 2020), avoiding teleost fish and other species with an additional genome duplication. The seven species (with their Ensembl acronyms) comprised: spotted gar (*Lepisosteus oculatus*, Loc), reedfish (*Erpetoichthys calabaricus*, Ecr), xenopus (*Xenopus tropicalis*, Xet), zebra finch (*Taeniopygia guttata*, Tgu), chicken (*Gallus gallus*, Gal), opossum (*Monodelphis domestica*, Mod), and human (*Homo sapiens*, Hsa). Amongst birds, I selected zebra finch as the primary species because, in comparison, the chicken karyotype appears to comprise two fusions of chromosomes; thus, Gal 1  $\approx$  Tgu 1 + Tgu 1A, and Gal 4  $\approx$  Tgu 4 + Tgu 4A. The assembly versions are listed in [supplementary methods, Supplementary Material](#) online, along with details of the curation of ohnolog families.

I chose to restrict consideration to families for which at least three ohnologs are retained in extant vertebrates. Four-member families (“quartets”) are crucial to a determination of gene phylogeny at the level of the 1R duplication, whereas the more prevalent three-member families (“trios”) can still be assigned uniquely to paralogons. Although two-member families (ohnolog pairs) are potentially useful, their placement into paralogons is often ambiguous, and they are so numerous that manual curation of the entire set is unrealistic. Many of the remaining genes, that do not form ohnolog pairs, may conceivably represent “single-member ohnologs” for which the other three presumptive copies have been lost from all extant vertebrates.

The set of curated ohnolog families for the seven species is presented in [supplementary table S1, Supplementary Material](#) online, and comprises 255 quartets plus 631 trios, making a total of 886 families encompassing 2,908 ohnologs. It is very likely that manual curation failed to locate all such families but, on the other hand, the likelihood that families have incorrectly been included should be low; that is, this set will inevitably suffer from omissions, but should exhibit a low rate of false inclusions. The total number of genes in [supplementary table S1B, Supplementary Material](#) online is 17,721, ~13% smaller than  $7 \times 2,908$  because of the loss of, or

Paralagon	A						B						C						D						CLG
	Spotted gar	Reedfish	Xenopus	Zebra finch	Opossum	Human	Spotted gar	Reedfish	Xenopus	Zebra finch	Opossum	Human	Spotted gar	Reedfish	Xenopus	Zebra finch	Opossum	Human	Spotted gar	Reedfish	Xenopus	Zebra finch	Opossum	Human	
1	12	8	9	7, 9	4, 7, 2	2, 3, 21	4	3	2	29, R59	8, (4)	12	15	14	10	27	2	17	11, 9	13, 6	6	2	8, 6, 3	7, 3, 10, 5, (9)	B*
2	22	8	2	19	2	17, 7	5	12	9	(16)	6	16	9	10	7	5	5, 8	11	20	18	1	15	3	12, 22	G*
3	16, 1	15, 3	5	3	1, 2	2, 1, 20, 6	28	1	4	R41	8, 5	11	7	16	8	5, (8)	1	14, 15	2	1, 18	8	R112, R99	4	19	a*
4	12	12	4	1A	8	22, 12	13	11	9	14	6	16, 7, 17	10	14	10	18	2	17	(26)	16	7	30	(4)	19	H*
5	8	1	3	1A, (R88)	8	12, 7, 10, 22	3	3	2	26	2	1, 6	1	11	8	R41	X	X	5	18	4	12	6	3, 9	E
6	8	1	3	1A, R88	8	12, 22, 7	3	3	2	26	2	1, 6	27	2	4	5	5	11	-	11	7	30	4	19	O
7	2, 4	7, 5	1	Z	3, 6	5, 9, 18	19	12	1	28	3	19	24	2	8	25	2	1	3	17	3	10	1	15	C
8	2, 4	7, 5	1	Z	6, 3	9, 5	19	12	1	28	3	19	10	10	4	8	2	1	6	17	3	30	3	19	L
9	21	9	8	17	1	9	-	1	8	-	2	6	10	10	4	8	2	1	6	17	3	30	3	19	M
10	(5)	2	1	R107	1	14	23	9	4	11	1	16, 19	18	10	10	20	1	20	9, 11	6, 13	6	2	3, (8)	8, 18, 6, 5, (3)	D
11	1, 16	3, 15	5	3	1, 2	6, 2, 8	6	14	2	23	4	1	18	10	10	20	1	20	9, 11	6, 13	6	2	3	8, 18, 6	J
12	1	3, 15	5	3	2	6	6	14	2	23	4	1	7	12	8	4A	X	X	3, 17, (14)	4	2	1	4, 7, 2 X, 13, 21, 11, 1, 2, (3)	K	
13	24	17	7	R100	4	19	26	9	7	24	4	11	5	16	8	25	2	1	3, 17, 14	4	2	1	4, (2)	3, 21, (1)	a
14	4, 2	5, 7	1	4, Z	5, 1	4, 2, 20, 8	6	11	3	13	1	5	1	1	3	22	1	8, 2, 7	5	2	7	6	1	10	l, Q
15	4, 2	5, 7	1	4	5	4, 8	6	11	3	13	1	5	7	12	8	4A	X	X	17, 3, 14	4	2	1	4, 7	13, 11, 2, X, 15	F
16	2	3	(3)	R90	2	17, 7	14	2	5	9	7, 2, 4	3, 2	7	12	8	4A	X	X	17, 3, (14)	4	2	1	7, 4	13, 11, X	N
17	2	3	3	R90	2	17, 7	14	2	5	9	7, 2, 4	3, 2	25	8	7	21	4, (2)	1	26	9	7	1, (1A)	8	12, 7	P

**Fig. 1.**—Summary of the chromosomes on which curated ohnologs reside. Entries are arranged into 17 rows, and into four major columns (A–D) of “quadruplicate” members, and finally by species. Each entry specifies the chromosome(s) on which ohnologs reside, for the indicated species. Each row corresponds to one paralagon, and has arisen through quadruplication of one ancestral pre-1R chromosome. The color-coding for backgrounds (chromosomal fusions) and fonts (extant chromosomes) is explained in the text. The prefix “LG” has been omitted from spotted gar chromosomes. Entries starting “R” denote unplaced scaffolds in zebra finch. [Supplementary table S2, Supplementary Material](#) online lists the numbers of curated ohnologs for each entry; brackets denote  $\leq 2$  ohnologs. The final column “CLG” lists the chordate linkage group of Simakov et al. (2020), and is considered in the Discussion.

failure to identify, genes in individual species. [Supplementary figure S1, Supplementary Material](#) online shows that the locations of these ohnologs are widely distributed throughout the genome for each species.

### Seventeen Paralogons

The principal features of this curated set are extracted in figure 1, at the chromosome level. The four major columns, labeled A–D, denote the four ohnologs (“quadruplicates”) that are possible for each row. Within each major column, just six of the seven species are listed; to avoid excessive width, chicken has been omitted here (as the entries are very similar to zebra finch), but is given in [supplementary table S1, Supplementary Material](#) online. Each of the 17 rows corresponds to a unique “signature” of chromosomes across the four ohnologs and six species. For the 886 quartets and trios in [supplementary table S1, Supplementary Material](#) online, every family was found to fall neatly into one of the 17 rows in figure 1, according to that row’s unique signature of chromosomes. The analysis in figure 1 was done independently of previous reports, yet it will be shown in the next section that these 17 rows correspond one-to-one with the 17 pre-1R ancestral chromosomes reported by Sacerdot et al. (2018). Accordingly, each row represents one paralagon, comprising four ohnologs.

The 68 cells in figure 1 (each comprising six species) will be identified using the notation  $PQ$ , where  $P$  is the paralagon (row number, 1–17), and  $Q$  is the quadruplicate (A–D); for example, the bottom row and third major column will be

identified as  $PQ = 17C$ . The number of ohnologs in each such cell is listed for each species in [supplementary table S2B, Supplementary Material](#) online. The color-coding of chromosomes and backgrounds are important features that will be explained subsequently. Note that figure 1 shows gene synteny at a macroscopic level of chromosomes (i.e., in the original sense of the term [Renwick 1971]), rather than at a microscopic level of gene positions on chromosomes; the latter information is available in [supplementary table S1B, Supplementary Material](#) online.

Initially, the allocation of families of ohnologs to groups (paralogons,  $P$ ), and of ohnologs to columns (quadruplicates,  $Q$ ) was performed entirely manually, with additional groupings being added as new patterns of chromosomes emerged. But, subsequently, it became imperative to automate the procedure, in order both 1) to confirm the manual allocation, and 2) to examine whether 17 was the correct number of groupings. The automated procedure is described in [supplementary methods, Supplementary Material](#) online; in outline, it searched for a match of chromosomes for each ohnolog to the patterns shown in figure 1, with an allowance of mismatch in a maximum of one species per family. For 854 of the 886 families, every chromosome matched, for each ohnolog; for 27 families there was a single conflict, and for 5 families there were two conflicts. No family exhibited a chromosome mismatch for more than one ohnolog. For 881 of the 886 families, the automated allocation conformed exactly to the manual allocation, when allowance was made for a chromosome mismatch in a single species. Interestingly,

for the remaining five families that exhibited two mismatches, both mismatches occurred in mammals (opossum and human; see *HOOK3*, *SS18L2*, *ELMO1*, *TMEM30B* and *PAX5* in [supplementary table S1B, Supplementary Material](#) online), suggesting the possibility that in each case a gene translocation had occurred in a stem mammal. In light of that possibility, those five families have been retained in [supplementary table S1, Supplementary Material](#) online. The relevance of these results to determining the correct number of paralogs will be considered in the Discussion.

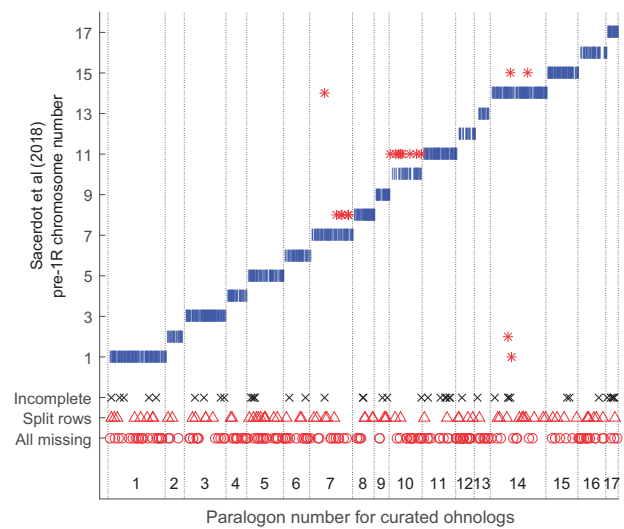
### Comparison with Seventeen “Pre-1R” Ancestral Chromosomes

It is straightforward to compare the chromosomal groupings in [figure 1](#) with the groupings reported by Sacerdot et al. (2018), by examining the human orthologs tabulated in [Supplementary Table S8](#) of that study. [Figure 2](#) plots the 886 curated quartet and trio ohnolog families from the present study, arranged into the 17 signature groupings, against the pre-1R chromosome numbers assigned previously. Note that, in the present study, I have re-arranged the order of rows so as to correspond to the Sacerdot ancestral chromosome numbering.

Each blue “|” symbol denotes a one-to-one correspondence of human ohnologs between the studies; that is, for a given family, a human gene had been assigned to the same group number in the two studies. Where multiple human genes in the same family are placed into the same grouping only a single blue “|” is visible, because the symbols superimpose; thus, even if all three or four human genes in a family coincide between the two studies, only a single blue “|” is visible in [figure 2](#).

The red and black symbols indicate discrepancies between the two studies. The red “\*” symbols plot 18 families containing mismatches of paralogon assignment between the studies, as listed in [supplementary table S3A, Supplementary Material](#) online. In addition, the red circles plot 163 families for which *none* of the human ohnologs were present in the previous study, as listed in [supplementary table S3B, Supplementary Material](#) online.

For the remaining 705 families (of the total of 886 curated ohnolog families), each human gene that was present in both studies was placed into the same grouping (the blue “|” symbols). However, in 40 of these cases (black “x”) the family in Sacerdot et al. (2018) was incomplete with respect to the curated family here; thus, it lacked at least one human ohnolog ([supplementary table S3C, Supplementary Material](#) online). And in 73 cases (red “△”) at least one of the human ohnologs had been placed in a different ohnolog *family* by Sacerdot et al. (2018), though that family was in the same *paralogon* as in this study ([supplementary table S3D, Supplementary Material](#) online). As a specific example, in

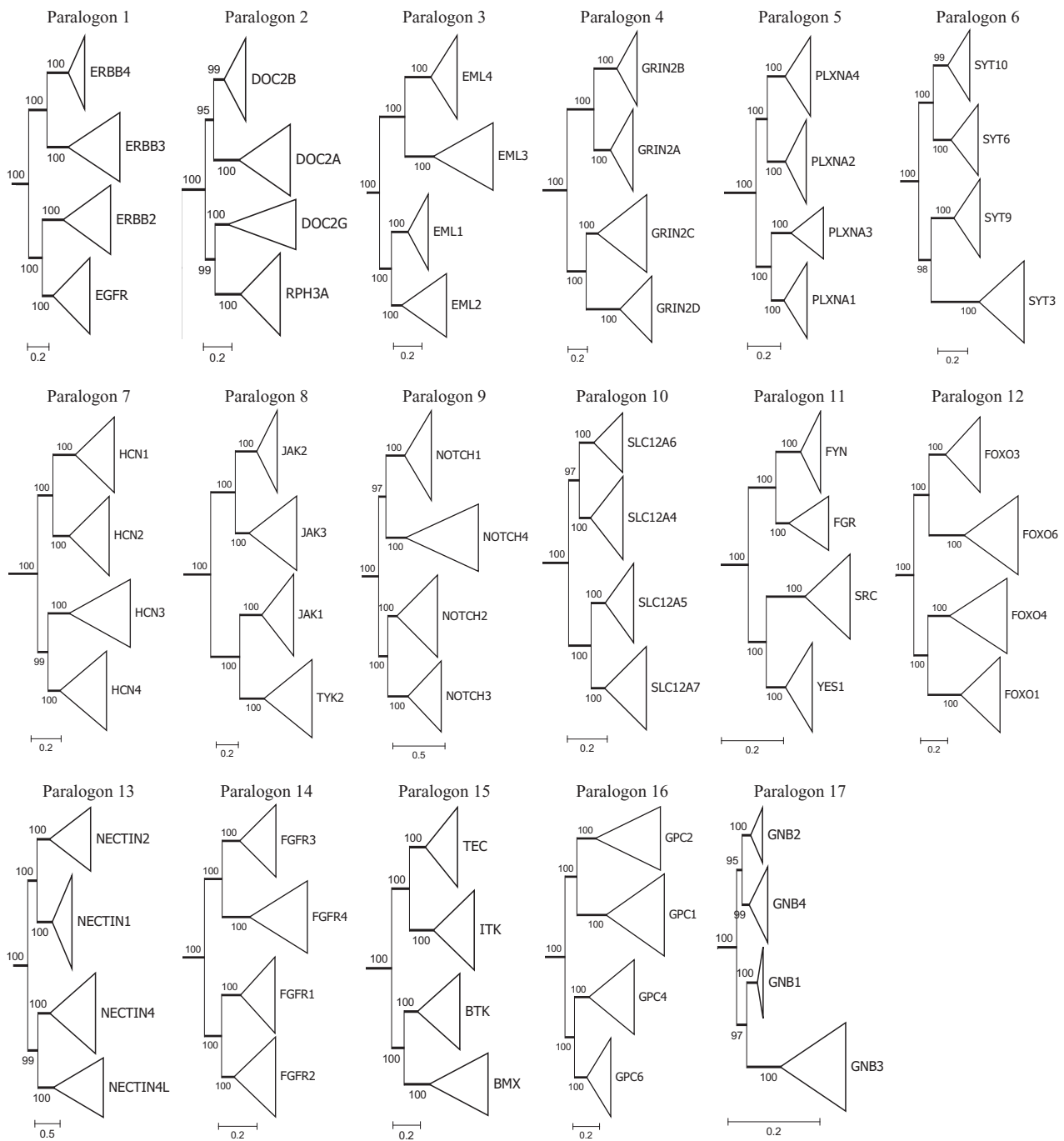


**Fig. 2.**—Allocation of human ohnologs to paralogs, compared between this study and Sacerdot et al. (2018). The blue “|” symbols denote one-to-one correspondence of allocations in the two studies. Red and black symbols indicate discrepancies in the allocations to paralogs, as listed in [supplementary table S3, Supplementary Material](#) online. As one example, the lowermost red star corresponds to genes *SLC4A4/5/9* in paralogon 14, which were incorrectly included with genes *SLC4A7/8/10* in paralogon 1.

the present study, genes *GNB1–4* are placed in a single ohnolog family (Lagman et al. 2012; Lamb 2020) in paralogon 17, whereas in Sacerdot et al. (2018) *GNB3* was placed in a separate family from the other three genes, though both those families were placed in paralogon 17.

Hence, although the blue symbols in [figure 2](#) show an overall very close correspondence between the groupings in the two studies, the red and black symbols show that there are substantial differences. The process of manual curation in the present study has identified 163 families of ohnologs that were not previously found, and has also identified a considerable number of discrepancies between the results, as listed in [supplementary table S3, Supplementary Material](#) online. Sacerdot et al. (2018) noted the similarity of their groupings to those in the CLGs of Putnam et al. (2008), and the Discussion will examine the correspondence between the paralogs here and the updated CLGs of Simakov et al. (2020).

One important feature of [figure 1](#) is that all four quadruplicates of all 17 ancestral pre-1R chromosomes can be found within extant bony vertebrates, though occasional losses are exhibited in some species (notably gar and zebra finch). This finding contrasts with the interpretation of Simakov et al. (2020), who reported that seven of the 68 quadruplicates had been lost, in their study of spotted gar, chicken, and xenopus. Further differences will be considered in the Discussion.



**FIG. 3.**—Maximum-likelihood molecular phylogenies for one example ohnolog family for each of the 17 paralogons in figure 1. The fully expanded version of each of these collapsed trees is presented in [Supplementary Material](#), along with the outgroup (which is not shown here). In each panel, the clades are arranged from top to bottom according to quadruplicates A, B, C, and D in figure 1. In every case, the ML tree conforms to the 1R branching pattern shown in figure 1 with a support level of at least 95%. Scale bar: amino acid substitutions per site.

### Pairs of Ohnologs That Diverged at the First Genome Duplication (1R)

An important extension of the present study, beyond previous reports, comprises analysis of the phylogeny of individual ohnolog families, and generalization of those results to each of the

17 rows. Thus, within the four quadruplicates (A–D), it has now been possible to assign which pair of ohnologs diverged from the other pair at 1R, for each of the 17 paralogons.

For each of the 17 paralogons, figure 3 presents one ML molecular phylogeny, in collapsed format and with its



outgroup omitted for compactness, and in each panel the four clades have been arranged from top to bottom in order of the quadruplicates A–D in figure 1. Importantly, in every panel in figure 3 the pattern of 1R branching conforms with that specified in figure 1. In ten of the panels this pattern of branching has unanimous support, and in the remaining seven panels the level of support is at least 95%. For the 68 collapsed clades, the lowest level of support is 99%. Each of these collapsed trees is shown in fully expanded format in the supplementary molecular phylogenies, [Supplementary Material](#) online.

In addition to these 17 trees, a second set of 17 rooted trees (one for each paralogon) is presented in [supplementary figure S2](#), [Supplementary Material](#) online; thus every paralogon is represented by two rooted molecular phylogenies. In this second set, each of the trees shows a branching pattern that conforms with figure 1, and the level of support is at least 95% for 16 of the 17. For the remaining tree (the *ATP2As* in paralogon 2) the level of support is 75%; when the outgroup was omitted, support for the ((A, B),(C, D)) pairing became unanimous.

The manual curation of molecular phylogenies (each typically comprising over 100 sequences) is very time-consuming, and so the decision was taken to restrict analysis to just two rooted trees per paralogon, provided that both trees showed high support for a common pattern of branching. For two of the 17 paralogs, one of the trees initially analyzed suggested a cascading (A,(B,(C, D))) pattern, as shown in Supplementary Phylogenies 2C and 9C. Accordingly, for each of paralogs 2 and 9 the examination was extended to an additional family. The resulting trees are shown in [supplementary figure S2](#) and in supplementary phylogenies 2B and 9B, [Supplementary Material](#) online; both trees exhibited the ((A, B),(C, D)) pattern shown in figure 1, with support levels of 75% and 98%, respectively.

In summary, 34 of the 36 rooted phylogenies examined conformed to the ((A, B),(C, D)) branching patterns in figure 1, and for 33 of these the level of support was at least 95%. The combination of this phylogenetic information with additional nonphylogenetic information for five paralogs (that will be presented in the next section) provides strong support for the conclusion that the branching pattern for 16 of the 17 paralogs conforms to the ((A, B),(C, D)) pattern shown in figure 1. For paralogon 2, the pattern in figure 1 is again supported, though not with as high a level of confidence.

### Chromosome Fusions during the Interval between 1R and 2R

Just as the existence of 17 pre-1R chromosomes has led to the manifestation of 17 unique signatures for chromosomes of extant species, across the four columns of ohnologs, so too the occurrence of ancient fusions of chromosomes has led to features that can be detected in figure 1. For example, the

fusion of two chromosomes during the interval between 1R and 2R would be expected to have resulted in two paralogs (rows) exhibiting common signatures across a pair of columns that originated at 1R. This phenomenon is illustrated in figure 1 by background coloring that is common, either across columns A and B or across columns C and D, between more than one row.

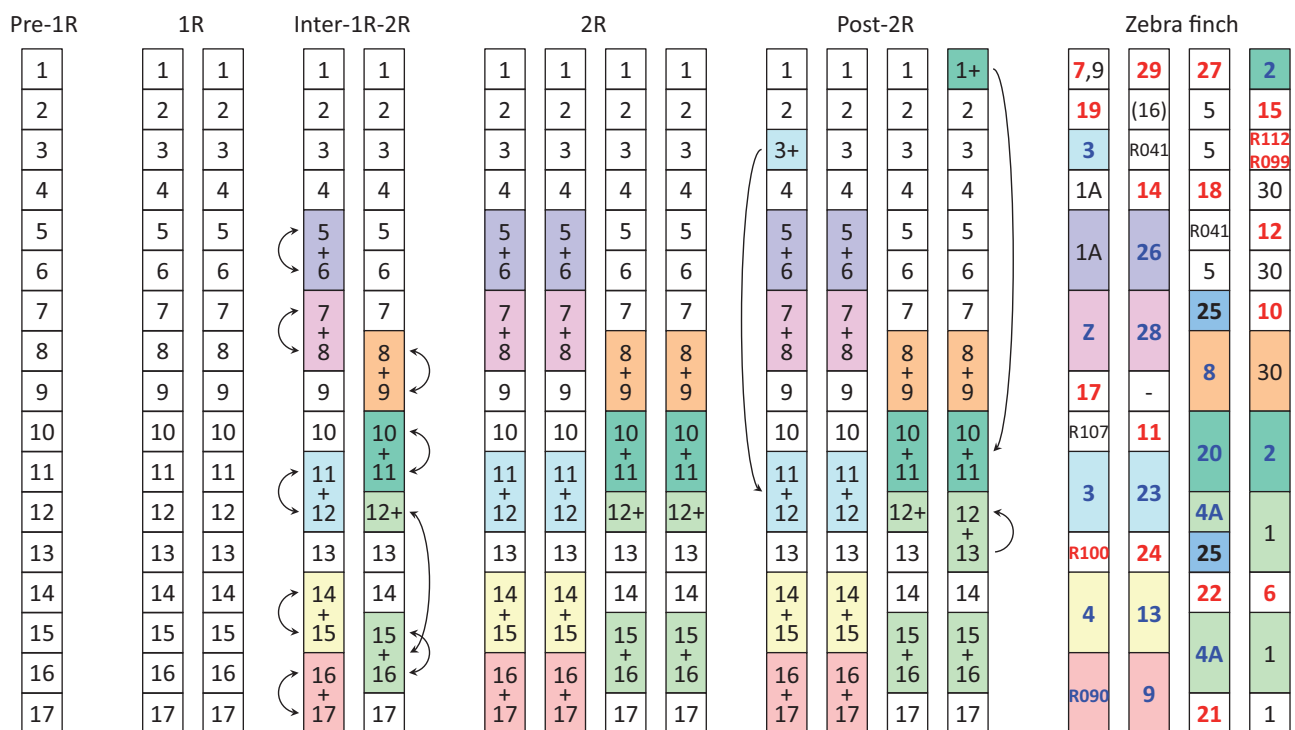
As one example, the pink coloring highlights a common signature across columns A and B of rows 7 and 8; that is for  $PQ = [7A, 7B, 8A, 8B]$ . The most parsimonious interpretation of the commonality of these two rows, in a pair of columns that originated at 1R, is that one post-1R copy of ancestral chromosome 7 fused with one post-1R copy of ancestral chromosome 8, prior to the second genome duplication (2R). Likewise, the orange coloring across columns C and D of rows 8 and 9, for  $PQ = [8C, 8D, 9C, 9D]$ , highlights a second common signature, and again the most parsimonious interpretation is that one post-1R copy of ancestral chromosome 8 fused with one post-1R copy of ancestral chromosome 9, prior to 2R.

It is interesting that row 8 exhibits two fusions: one in columns A and B and another in columns C and D. Accordingly, row 8 provides evidence, completely independent of any phylogenetic data, that in this case columns A and B must have diverged from columns C and D at 1R. Five paralogs (8, 11, 12, 15, and 16) display this feature, whereby all four quadruplicates exhibit fusions with another row, in a manner that in each case supports the divergence of quadruplicates (A, B) from quadruplicates (C, D), on the basis of evidence that is independent of phylogenetic analysis.

From examination of the patterns denoted by the colored backgrounds in figure 1, the deduced set of chromosomal fusions is presented in figure 4, with these fusions illustrated first for the interval between 1R and 2R (“Inter-1R-2R”) and, second, after 2R (“Post-2R”). According to this scheme, there were nine fusions of chromosomes in the inter-1R-2R period, indicated by the nine arrows and eight hues. All but one of those were simple fusions of two chromosomes, though one (overlapping arrows, pale green background) involved fusions between three chromosomes; however, there is no information about the temporal order in which any of the nine fusions occurred. These nine fusions would have reduced the number of chromosomes from 34 to 25, just prior to 2R.

### Chromosome Fusions after 2R

Then, following 2R, there were additional fusions, three of which apparently occurred prior to the radiation of bony vertebrates; these three fusions following 2R are indicated in the “Post-2R” column of figure 4 by the three arrows and additional background coloring. These post-2R fusions created fused cells  $PQ = [3A, 11A, 12A]$ ,  $PQ = [1D, 10D, 11D]$  and  $PQ = [12D, 13D, 15D, 16D]$ , with the latter comprising



**FIG. 4.**—Deduced fusions of chromosomes between 1R and 2R (“Inter-1R-2R”) and after 2R (“Post-2R”), and also a comparison with extant zebra finch chromosomes. The 17 ancestral chromosomes (“Pre-1R”) duplicated at the first round (“1R”) to generate 34 chromosomes. The nine fusions of chromosomes, proposed to have occurred after that time but before 2R, are indicated using the color codes from figure 1 together with nine curved double-headed arrows; the fusion of three chromosomes is shown in pale green and with overlapping arrows. The resultant 25 chromosomes duplicated to 50 at the second round (“2R”). Three additional fusions are proposed to have occurred prior to the radiation of bony vertebrates, indicated by single-headed arrows. The right-hand panel for zebra finch is extracted from figure 1.

progeny from four of the 17 ancestral pre-1R chromosomes.

**Ancient Chromosomes Preserved in Extant Organisms: “Archeochromosomes”**

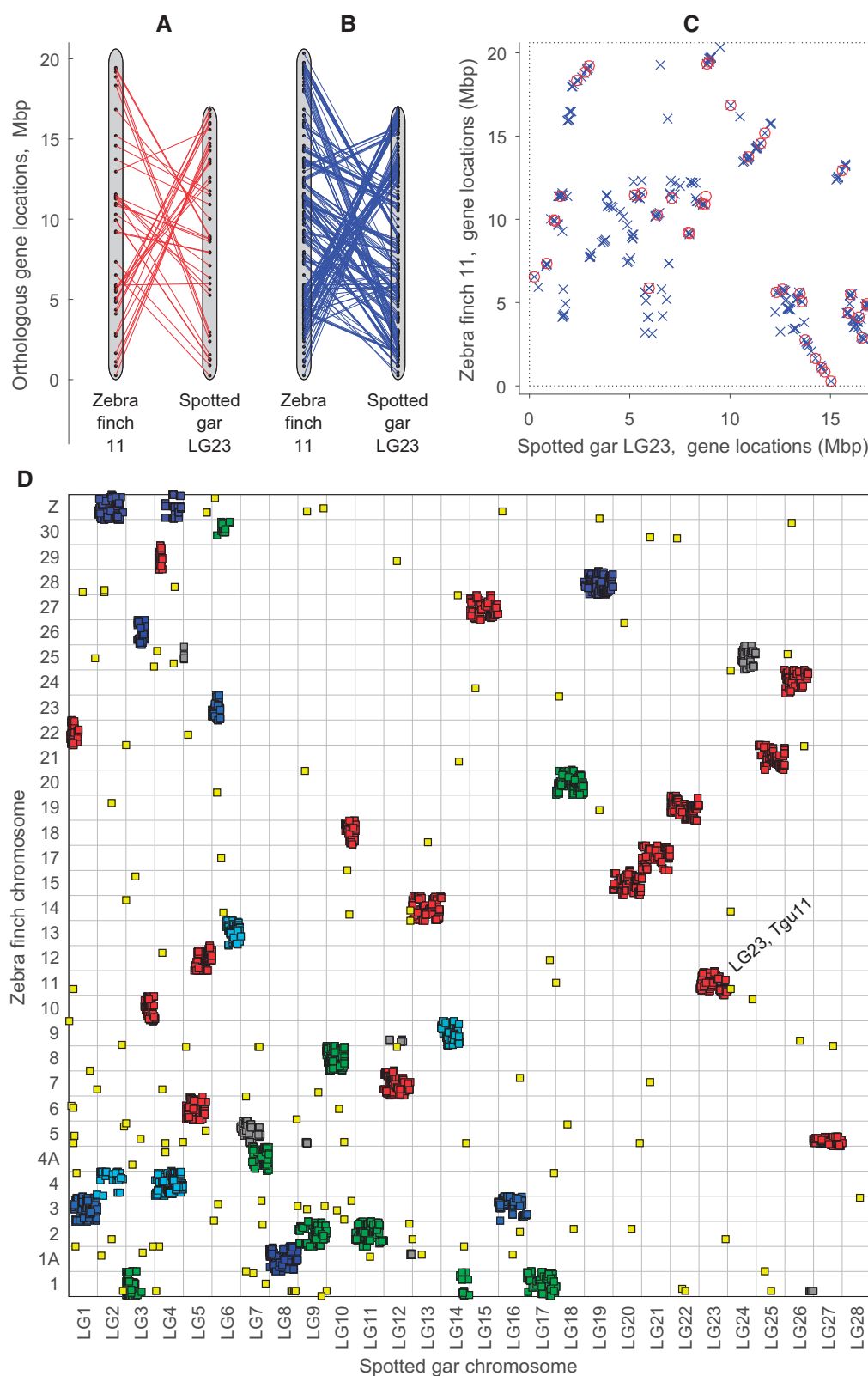
An interesting feature of the arrangement of ohnologs on the chromosomes of some extant species is indicated by the use of bold red and bold blue font in figures 1 and 4. For certain species, exemplified in figure 1 by zebra finch and spotted gar, a number of extant chromosomes appear to be “insular,” meaning that they are like islands, containing ohnologs that are descended from only one particular pre-1R chromosome.

Thus, for each of the zebra finch chromosomes shown in bold red font in figure 1 or 4, that chromosome is essentially the only location in the entire genome where ohnologs for one particular ancestral chromosome and quadruplicate are found. Supplementary table S4, Supplementary Material online provides a list of insular chromosomes for zebra finch and spotted gar; its accuracy may be confirmed by searching entries in supplementary table S1B, Supplementary Material online. Those insular chromosome that occur in a single (unfused) cell in figure 1 are marked in bold red font.

Interestingly (though with separate interpretation, see below), in all but one case that red chromosome is the only chromosome from that species found in that cell.

Consider, for example, zebra finch chromosome 11 (Tgu 11). Inspection of figure 1 shows that this chromosome is found *only* in cell PQ = 10B (row 10, quadruplicate B); hence, Tgu 11 is shown in bold red font and is referred to as an insular chromosome. Furthermore, Tgu 11 is the only chromosome found in that cell. In other words, 1) every curated ohnolog on Tgu 11 is descended from one particular ancestral bony-vertebrate chromosome (PQ = 10B), and 2) every curated ohnolog that is descended from that ancestral bony-vertebrate chromosome resides on Tgu 11. This can be verified by searching the chromosome columns of the zebra finch sheet in supplementary table S1, Supplementary Material online for occurrences of “11.” Interestingly, exactly the same situation applies for spotted gar chromosome LG23 (fig. 1). As will be described below, this correspondence with LG23 reflects just one example (here, in zebra finch) of the previously reported correlation between certain chromosomes in spotted gar and chicken.

This finding, that for those cases in figure 1 marked by bold red font, all curated ohnologs on the indicated chromosome are restricted almost exclusively to a single paralogon and



**FIG. 5.**—Mapping of orthologous gene locations between zebra finch and spotted gar chromosomes: (A–C) for one example pair of chromosomes (Tgu11, LG23), and (D) for all chromosomes. (A) Using pairs of orthologs from the curated families of quartets and trios, for  $PQ = 10B$  (see fig. 1 and [supplementary table S1, Supplementary Material](#) online). (B) Using all orthologous gene pairs between the same two chromosomes, as determined from



quadruplicate, strongly suggests that each such chromosome represents a relic, with its content of ohnologs essentially unchanged from that in the bony-vertebrate ancestor. Such chromosomes, that appear to have remained virtually unchanged since 2R-WGD will here be termed “archeochromosomes,” to denote the stability of their content of genes since ancient times.

Chromosomes in figure 1 marked in bold blue font reflect a closely similar situation, except that they correspond to ancestral chromosomes that fused during the inter-1R-2R period. For example, curated ohnologs on zebra finch chromosome 28 (Tgu 28) are restricted exclusively to the fused cell  $PQ = [7B, 8B]$  (pink background); likewise, in spotted gar, curated ohnologs on chromosome LG19 are restricted exclusively to this same fused cell. Similarly, curated ohnologs on zebra finch Tgu 20 are restricted exclusively to fused cell  $PQ = [10C, 11C]$  (blue-green background), as are those for spotted gar LG18. Accordingly, these four example chromosomes (two from zebra finch and two from spotted gar) are also “insular” chromosomes, and are shown in bold blue font.

Some of the chromosomes marked in bold blue are likely to be even more ancient than those in red, because they arose during the inter-1R-2R period, whereas the chromosomes marked in red could not have arisen until 2R, when quadruplicates first appeared. However, several of the bold blue chromosomes are found not to be alone in the ancestral cells, but appear alongside other chromosomes, so it will be important to examine these cases in more detail.

#### Example of a Pair of Presumptive “Matching Archeochromosomes” in Zebra Finch and Spotted Gar

From the first example in the previous section (for  $PQ = 10B$ ), it would appear that zebra finch chromosome 11 and spotted gar chromosome LG23 share a common set of families of curated ohnolog quartets and trios. The mapping of locations of common ohnologs on the two chromosomes is illustrated by the red lines in figure 5A; for the 40 common ohnologs, the positions of the genes are distributed along substantially the whole length of each chromosome. Figure 5B extends this analysis to all the orthologous genes on these two presumptive archeochromosomes; [supplementary table S5D](#), [Supplementary Material](#) online shows that 261 such pairs were found.

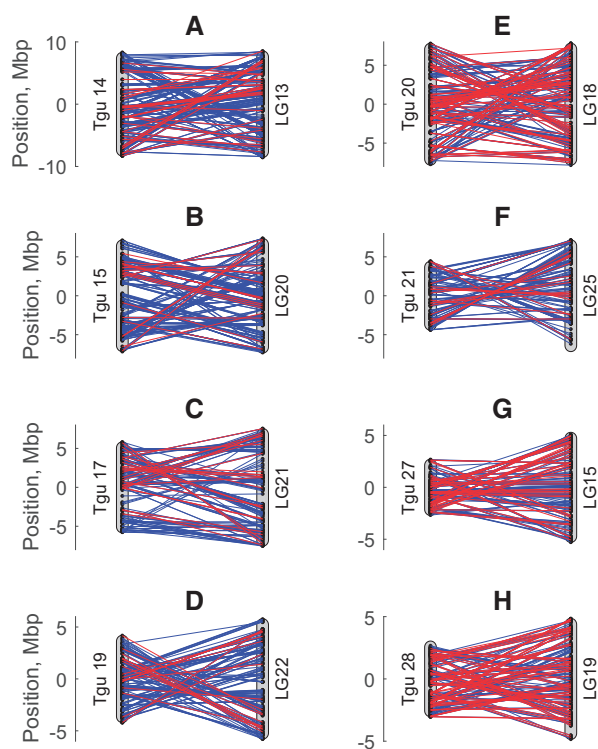
This number (261) of orthologs common to zebra finch chromosome 11 and spotted gar LG23 is >6-fold higher than the number (40) of curated ohnologs plotted in figure 5A, and includes around 60% of the coding genes (413 and 447, respectively) on the two chromosomes. However, as shown in figure 5B, the inclusion of this considerably larger number of orthologs makes little qualitative difference to the plot, which, apart from the greater density of lines, closely resembles the plot for curated ohnologs in the first panel. In particular, orthologs are found along substantially the whole length of each chromosome. Accordingly, for the pair of presumptive archeochromosomes in figure 5A and B, the pattern of the mapping between genes on the two chromosomes is very similar, whether one examines curated ohnologs or all orthologs.

The mapping of gene locations between the two chromosomes is examined as a “dot plot” in figure 5C, where the dashed rectangle indicates the extent of each chromosome, of 20.6 and 17.0 Mb, respectively. The red circles plot the 40 curated ohnolog pairs, and correspond to the red lines in figure 5A; the blue crosses plot the 261 common orthologs, and correspond to the blue lines in figure 5B. The spatial locations appear not to be completely random; instead, several local regions appear highly correlated, and represent local synteny blocks between the two chromosomes. This pair of chromosomes represents one example of the conservation, that was reported by Braasch et al. (2016), of entire chromosomes between spotted gar and some tetrapods. In this case, both are descended from a common ancestral bony-vertebrate chromosome (defined here as  $PQ = 10B$ ), in species whose ancestors diverged around 400 Ma.

#### Locations of Orthologous Genes on All Chromosomes of Zebra Finch and Spotted Gar

For the pair of presumptive matched archeochromosome described in the previous section, it is important to ascertain whether appreciable numbers of genes on either chromosome (Tgu 11 or LG23) have their orthologs located on chromosomes other than the paired chromosome. This is investigated in the dot plot of ortholog location for all zebra finch and spotted gar chromosomes in figure 5D, where yellow symbols denote “outlier” locations. Inspection of row Tgu 11 shows that only three orthologs are located on chromosomes other than LG23, and inspection of column LG23

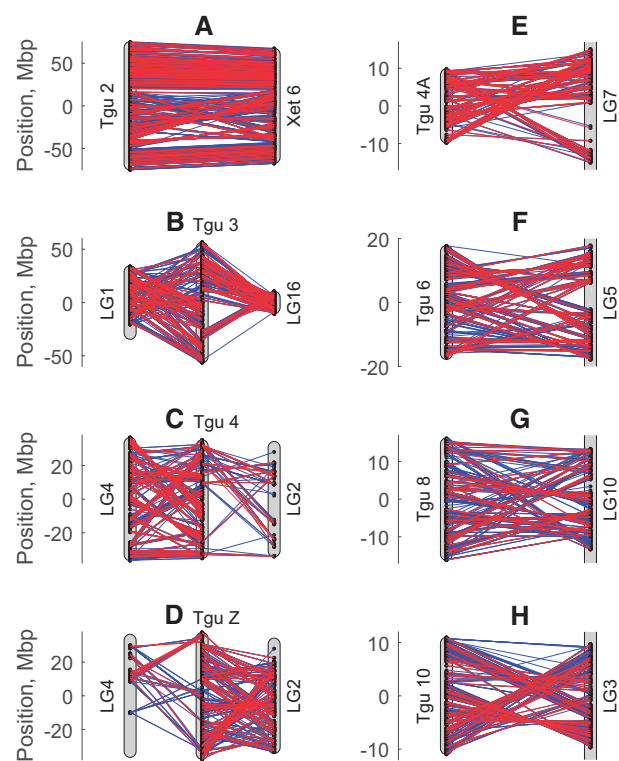
Ensembl using the criteria set out in [supplementary methods](#), [Supplementary Material](#) online. (C) Red circles as for red lines in (A); blue crosses as for blue lines in (B); unfilled symbols have been chosen so that overlap of ohnologs (red) with all orthologs (blue) can be seen. Dotted box indicates the extent of each chromosome. (D) The 10,092 orthologous pairs on all chromosomes, shown using the following color code based on  $PQ$  cells in figure 1. Cells with insular chromosomes (denoted by red font in fig. 1) are shown in red. Cells with fused ancestral chromosomes (denoted by blue font in fig. 1) are shown using shades of blue and green: the five shades of blue indicate the five fusions in columns  $Q = [A, B]$ , and the four shades of green represent the four fusions in columns  $Q = [C, D]$ . Cells that contain neither insular nor fused ancestral chromosomes are shown in gray. Yellow represents the 137 (1.4%) “outlier” combinations of chromosome pairings that do not correspond to any  $PQ$  cell in figure 1. Note that the plot of all orthologs in panel (C) is represented in panel (D) by the rectangular region marked “LG23, Tgu11.”



**FIG. 6.**—Mappings between orthologous gene locations on pairs of matching archeochromosomes in zebra finch (Tgu) and spotted gar. Blue lines apply for all identified orthologs, and red lines apply for genes from curated ohnolog quartets and triplets; because the blue lines were plotted first, they are often obscured by the red lines.

shows that only one ortholog is located on a chromosome other than Tgu 11. This compares with 261 orthologs on the tested pair, so that only 4/265 orthologs (1.5%) are outliers.

The data plotted in figure 5D are tabulated in [supplementary table S5C, Supplementary Material](#) online, and they are also re-plotted in [supplementary figure S3A, Supplementary Material](#) online in coordinates of equal Mb rather than of equal length for each chromosome; in addition a similar plot for curated ohnologs is presented in [supplementary figure S3B, Supplementary Material](#) online. Across the entire dataset, there are just 137 yellow outliers out of the total of 10,092 orthologs detected between all chromosomes of the two species, equating to just 1.4% of the total. The remaining 9,955 orthologs (98.6%) are restricted to just 44 out of the 896 rectangles ( $= 32 \times 28$ ) in figure 5D. Of those 44 rectangles, 16 contain red symbols and represent unfused ancestral chromosomes, whereas 21 contain blue/green symbols and represent chromosomes that fused during the 1R–2R interval, and the remaining seven contain gray symbols and represent chromosomes that are neither insular nor fused. The great majority of rectangles, 748 of 896 (83%), in figure 5D are empty, and another 83 (9%) contain just a single yellow outlier.



**FIG. 7.**—Mapping of orthologous gene locations from eight zebra finch chromosomes to chromosomes in spotted gar or xenopus. Zebra finch chromosomes are on the left, or (in B, C, and D) in the middle. As shown in [table 1](#), zebra finch chromosomes 2, 3, 4, 4A, 6, 8, and 10 qualify as archeochromosomes. Zebra finch chromosome Z almost qualifies, though its orthologs are not entirely restricted to a single cell of figure 1 (see [table 1](#)).

This analysis of orthologs across all chromosomes is entirely consistent with the interpretation of ohnolog locations presented in figure 1, but extends the number of pairs analyzed from  $\sim 2,200$  ohnologs to  $\sim 10,000$  orthologs. The very low incidence of outliers supports the conclusion that the occurrence of any small-scale rearrangements between chromosomes in either zebra finch or spotted gar (other than shown in fig. 1) can at most have been minimal. The small proportion of orthologs (1.4%) that were located “incorrectly” presumably reflects the combination of 1) occasional genes that had genuinely been translocated to other chromosomes, 2) errors in the identification of orthologs, and 3) assembly errors, together with any other possible artefacts.

#### Matching Pairs of Archeochromosomes between Zebra Finch and Spotted Gar

In the [Supplementary Material](#), four criteria are set out that can be applied to test whether a pair of chromosomes in distantly related species, that appear similar, genuinely qualify as “matching archeochromosomes.” The first two criteria require that orthologs for the putative archeochromosome

**Table 1.**  
Results of Testing for Archeochromosomes

Zebra Finch	1	2	3	4	Result	Gar or Other	Figure	Gar or Other	1	2	3	4	Result	Zebra Finch	Figure
1	X	✓		✓	X	LG17, LG3, LG14, LG26, (LG8)	S7A	LG1	X	X		X	X	3, 22, R041	
1A	✓	✓	X	✓	X	LG8, LG12	S7B	LG2	X	X	X	X	X	Z, 4, R090, R112	
2	✓	✓		✓	MA	Xet 6	7A	LG3	X	X		X?	X	10, 1, 26	
3	✓	✓		✓	A	LG1, LG16	7B	LG4	X	X		X	X	4, 29, Z	
4	✓	✓		✓	A	LG4, LG2	7C	LG5	X		X	X	X	6, 12, (25)	
4A	✓	✓		✓	A	LG7	7E	LG6	X		X	X?	X	13, 23, 30	S8A
5	*	*	X	✓	X	LG7, LG27, LG9	S7C	LG7	X		X	X	X	5, 4A	S8B
6	✓	✓		✓	A	LG5	7F	LG8	✓	✓		✓	A	1A, (1)	S8C
7	✓	✓?		✓	A?	LG12	S7D	LG9	X		X	✓	← -	2, (5)	S8D
8	✓	✓		✓	A	LG10	7G	LG10	X		X	✓	← +	8, 18	S8E
9	✓?	✓		✓	A?	LG14, LG12	S7E	LG11	✓		X	✓	← -	2	S8F
10	✓	✓		✓	A	LG3	7H	LG12	X		X	X?	X	7, 1A, 9	S9A
11	✓	✓	✓	✓	MA	LG23	5	LG13	✓	✓	✓	✓	MA	14	6A
12	✓	✓		✓	A	LG5	S6A	LG14	X	X	X	✓	X	9, 1	S9B
13	✓	✓		✓	A	LG6	S6B	LG15	✓	✓	✓	✓	MA	27	6G
14	✓	✓	✓	✓	MA	LG13	6A	LG16	✓	X		✓	← -	3	S9C
15	✓	✓	✓	✓	MA	LG20	6B	LG17	✓	X		✓	← -	1	S9D
(16)					—	—	—	LG18	✓	✓	✓	✓	MA	20	6E
17	✓	✓	✓	✓	MA	LG21	6C	LG19	✓	✓	✓	✓	MA	28	6H
18	✓	✓		✓	A	LG10	S6C	LG20	✓	✓	✓	✓	MA	15	6B
19	✓	✓	✓	✓	MA	LG22	6D	LG21	✓	✓	✓	✓	MA	17	6C
20	✓	✓	✓	✓	MA	LG18	6E	LG22	✓	✓	✓	✓	MA	19	6D
21	✓	✓	✓	✓	MA	LG25, (LG26)	6F	LG23	✓	✓	✓	✓	MA	11	5
22	✓	✓		✓	A	LG1	S6D	LG24	X			X	X	25, R100	S9E
23	✓	✓		✓	A	LG6	S6E	LG25	✓	✓	✓	✓	MA	21	6F
24	✓	✓		✓	A	LG26	S6F	LG26	✓?		X	X	X	24, 1	S9F
25	✓	✓		X	X	LG24, LG5	S7F	LG27	✓	✓		✓	A	5	S9G
26	✓	✓		✓	A	LG3	S6G	LG28	✓	✓		X?	A?	R041	S9H
27	✓	✓	✓	X?	MA	LG15	6G	(LG29)					—	—	
28	✓	✓	✓	✓	MA	LG19	6H								
29	✓	✓		✓	A	LG4	S6H	Xet 6	✓	✓	✓	✓	MA	2	7A, S11
30	*	*		✓	X	LG6, (LG26)	S7G	Ecr 4	✓	✓		✓	A	1	S11
Z	✓?	✓		✓	A?	LG2, LG4	7D	Hsa 4	✓	X		✓	X	4	S11

NOTE.—Left side: Zebra finch chromosomes, followed by outcomes of the four criterion tests (✓ = pass; X = fail; ✓? = borderline; \* = see below; X? = failure in KS2 test, but may not indicate failure of criterion 4). “Result” column gives the interpretation of these tests (MA = matching archeochromosome; A = archeochromosome; A? = possible archeochromosome or “near miss”; X = not an archeochromosome). Next column gives spotted gar chromosome(s) on which orthologs are found (from [supplementary table S7, Supplementary Material](#) online). “Figure” column gives the figure in which ortholog mappings are presented, with prefix “S” denoting supplementary figure, [Supplementary Material](#) online. Right side: Similar, but for spotted gar chromosomes, plus one chromosome each from xenopus, reedfish, and human. The asterisks indicate that in [supplementary table S4, Supplementary Material](#) online, chromosomes Tgu 5 and 30 were tested against combinations of cells that had not fused prior to the origin of bony vertebrates; hence they failed.

should be found in figure 1 in *only* one cell (or one set of inter-1R-2R fused cells), and should *not* be accompanied in that cell(s) by ohnologs from any other chromosome. The third criterion relates to the uniqueness of mapping of orthologs between the two chromosomes, and the fourth involves the absence of substantial gaps in the spatial coverage of orthologs along either chromosome; for details, see the Supplementary Results.

Application of those criteria showed that the pair of chromosomes in figure 5A–C meet all four criteria and thereby

qualify as matched archeochromosomes. Figure 6 plots the mappings of orthologs for another eight pairs of zebra finch and spotted gar chromosomes that likewise passed all of these tests, so that a total of nine pairs of matched archeochromosomes were identified between these two species. The results of testing are summarized in table 1.

Hence, it appears that, for all nine pairs, each is the only chromosome from that species that is associated with the relevant cell(s) in figure 1, and that there is no region on either chromosome that contains genes for which the orthologs are

reliably located elsewhere than on the matching archeochromosome. Hence, these are eighteen chromosomes (nine in each of two species), containing over 2,000 orthologous gene pairs, where it appears there has not been a single occurrence of a chromosomal fusion or fission in over 400 My. These cases extend the findings of previous studies that have shown very slow rates of genome evolution in Holostei (gar and bowfin) and sauropsids (birds and reptiles), as discussed for example in Shaffer et al. (2013) and Braasch et al. (2016).

In most species of bird, and in a number of other lineages, the majority of chromosomes are so short ( $\leq 20$  Mb) that they are classified as “microchromosomes.” This classification applies to each of the eighteen matching archeochromosomes reported here: in zebra finch, five are  $< 12$  Mb, and the longest (Tgu 11) is 20.6 Mb; in spotted gar, the lengths range from 10 to 17 Mb. The relationship between microchromosomes and archeochromosomes will be further considered in the Discussion.

### Comparison with Report of Pairs of One-to-One Chromosomes in Chicken and Spotted Gar

The paper describing the original assembly of the spotted gar genome (Braasch et al. 2016) noted the close similarity of multiple chromosomes between spotted gar and chicken, and reported that “Almost half of the gar karyotype (14/29 chromosomes) showed a nearly one-to-one relationship in gar-chicken comparisons.” Those authors listed 12 cases involving pairs of microchromosomes, together with 2 cases involving pairs of macrochromosomes; see their [Supplementary Information](#).

Nine of those 14 cases were shown in the previous section to be matching pairs of archeochromosomes. The remaining five cases are examined in [supplementary figure S5, Supplementary Material](#) online, where it is shown that none of those five combinations represents a matching pair of archeochromosomes. Accordingly, only nine of the 14 pairs of chromosomes previously reported to be in nearly one-to-one relationships qualify as *matching* archeochromosomes between spotted gar and either chicken or zebra finch. However, despite rejection of the other five nearly one-to-one pairings, it remains possible that any of those chromosomes might be an archeochromosome, though one for which a match has not been found.

### Testing for Additional Archeochromosomes in Zebra Finch and Spotted Gar

In cases where a potential matching archeochromosome in a distantly related species cannot be identified, a revised set of criteria for classification of a chromosome as an archeochromosome is developed in the [Supplementary Material](#). For each chromosome (other than those already identified as matching), in both zebra finch and spotted gar, these criteria were tested. For eight zebra finch chromosomes, the

mappings of orthologs to spotted gar chromosomes are illustrated in [figure 7](#), and for all chromosomes the results of testing are presented in [table 1](#). The column “Result” gives the interpretation of the outcomes of the four tests, with “MA” indicating a matching archeochromosome, “A” indicating an archeochromosome, “A?” indicating a possible archeochromosome or “near miss,” and “X” indicating “not an archeochromosome.”

To summarize for zebra finch: out of the total of 33 chromosomes, 24 clearly qualify as archeochromosomes. Of those, nine are found to match a spotted gar archeochromosome, whereas another (Tgu 2) is shown in the next section to match a xenopus chromosome (Xet 6). Five other chromosomes (Tgu 1, 1A, 5, 25 and 30) clearly fail to qualify as archeochromosomes, but another three (Tgu 7, 9, and Z) remain as possible or “near-miss” archeochromosomes; finally, Tgu 16 is too short to analyze.

To summarize for spotted gar: as well as the nine archeochromosomes shown to be paired with an archeochromosome in zebra finch, two other chromosomes, LG8 and LG27, also qualify as archeochromosomes, and it is possible that LG28 does too. In addition, chromosomes LG9 and LG11 have split from an archeochromosome that was equivalent to zebra finch Tgu 2; LG9 has undergone a subsequent fusion, but LG11 has not. Another two chromosomes, LG16 and LG17, are also fission products, but their pedigree is more complicated. Chromosome LG10 appears to be a recent fusion of two archeochromosomes, and LG6 appears to be a recent fusion of two archeochromosomes plus another chromosome. Finally, it seems likely that LG26 resulted from the fusion of two ancestral chromosomes after 2R WGD ([supplementary fig. S9F, Supplementary Material](#) online).

### Testing of Potential Archeochromosomes in Other Species

Examination of [figure 1](#) reveals two chromosomes in other species that are each shown in blue font, and that are each found alone within cells corresponding to a fused ancestral chromosome: namely, xenopus chromosome 6 and reedfish chromosome 4. Accordingly, those cases were tested against the criteria set out in [Supplementary Results](#) and, as shown in [table 1](#) (bottom right), both qualified as archeochromosomes. Indeed, Xet 6 (mapped in [fig. 7A](#)) passed all the tests as a matched archeochromosome with zebra finch Tgu 2. This is the only example found to date of a matched archeochromosome in a species other than zebra finch or spotted gar. The mappings of orthologs for reedfish chromosome 4 are shown in [supplementary figure S11B–E, Supplementary Material](#) online.

Human chromosome 4 is examined in the mappings of [supplementary figure S11F, Supplementary Material](#) online. Although it passed two of the criteria, it failed ([table 1](#)) because it is accompanied in cell PQ = [14A, 15A] of [figure 1](#) by other chromosomes. Human chromosome 13 is also



examined in [supplementary figure S11C, Supplementary Material](#) online, but it clearly failed several tests and is not included in [table 1](#).

## Discussion

### Curated Set of Ohnologs: 17 Paralogs

The curated set of ohnologs presented in [supplementary table S1, Supplementary Material](#) online is a valuable resource. It currently includes 255 quartet families and 631 trio families, for a total of >2,900 ohnologs, and across the seven species considered it comprises over 17,000 genes. This set is by no means exhaustive, and there is scope for finding additional quartets and trios, subject to careful checking of phylogeny.

By initially using a manual approach, the arrangement of chromosomes in [supplementary table S1B, Supplementary Material](#) online was found to divide into 17 discrete groupings. I was not able to devise a rules-based procedure that would automatically determine the appropriate number of groupings, and instead I adopted the following procedure. I began with the hypothesis that the “chromosome signatures” for the different species are correctly described across the 17 rows of four PQ cells in [figure 1](#). Based on that assumption, I implemented an automated rules-based approach (described in [supplementary methods, Supplementary Material](#) online) to determine, for each one of the 886 putative families, whether its ohnologs could be uniquely assigned to a single paralogon at distinct quadruplicate positions within that paralogon. The rules included allowance for at most a single mismatch of chromosome across the species in each family of three or four ohnologs.

Using this automated approach, 881 of the 886 families were allocated to a unique paralogon, with each of their ohnologs allocated to a separate quadruplicate column. The remaining five families each exhibited two mismatches of chromosomes (from the pattern in [fig. 1](#)); and in each case both mismatches occurred in the two mammalian species analyzed, opossum and human, and in only one ohnolog. Given that all the chromosomes matched across all the other species, across each of the three or four ohnologs, it seems plausible that a gene translocation had occurred for one ohnolog in a stem mammal. Accordingly, those five families have been retained in [supplementary table S1, Supplementary Material](#) online.

Given that this rules-based procedure successfully allocated every one of the 886 putative families (comprising >2,900 ohnologs and >17,000 genes across seven species) to a unique member of the 17 hypothesized signature groupings (subject to a single chromosome error, or two if both were in opossum and human), it seems highly unlikely that the correct number of signature patterns of chromosomes could be anything other than 17. In other words, the 17 groupings of

chromosome patterns in [figure 1](#) are both necessary and sufficient to account for every putative family of ohnologs that was encountered.

The obvious interpretation of the finding that the chromosome signatures fall into 17 discrete groupings, representing 17 paralogs within the genomes of bony vertebrates, is that these paralogs derive from 17 ancestral pre-1R chordate chromosomes (Sacerdot et al. 2018), as a result of two rounds of whole-genome duplication. An important feature of the ensuing analysis in this study is that the split of quadruplicate pairs in [figure 1](#) at the first round of genome duplication (1R) has been determined unambiguously for each paralogon. This was achieved through molecular phylogenetic analysis of ohnolog quartets within each paralogon, as well as (for a subset of five paralogs) through analysis of the fusions of ancestral chromosomes that occurred during the inter-1R–2R period. Those fusions can be seen as common signature patterns between pairs of quadruplicates in different paralogs, highlighted by background coloring in [figure 1](#).

Another notable feature is the occurrence of “insular” chromosomes, that are found *only* in a particular “island” within [figure 1](#). These are marked in bold red font when that island is a single cell, or in bold blue font when it is a collection of cells that fused between 1R and 2R. These insular chromosomes satisfy one criterion for classification as an archeochromosome; chromosomes that occur *alone* within such an island satisfy a second criterion for classification as an archeochromosome; for details of the full set of criteria, see [Supplementary Material](#).

All the features described above rely solely on analysis of synteny according to its original definition as “presence together on the same chromosome” (Renwick 1971); these conclusions do not involve any analysis of the order of genes (microsynteny). For an examination of the ancestral vertebrate karyotype, the local order of genes is not particularly informative.

### Comparison with Interpretations of Ancestral Chordate Linkage Groups

The genome of the cephalochordate amphioxus was analyzed by Putnam et al. (2008), and a new chromosome-level assembly was recently re-analyzed by Simakov et al. (2020), to extract in both cases 17 ancestral CLGs that provide estimates of the karyotype of the ancestral chordate organism at the time of the divergence of the ancestors of amphioxus and vertebrates. In their study, Sacerdot et al. (2018) noted the close one-to-one correspondence of 15 of their 17 CARTs to the CLGs of Putnam et al. (2008), though they noted that CART 14 encompassed two CLGs (6 and 7 of Putnam et al. 2008), whereas the small CART 13 could not be assigned to a CLG.

Simakov et al. (2020) were additionally able to assign the duplications during the two rounds of WGD, into what they referred to as “ $\alpha$ - $\beta$ ” pairs within their CLGs, on the basis of



similarity of gene retention fractions, and their interpretations are presented in their Fig. 3. The results of figure 1 in the present article can readily be compared with those of Fig. 3 in Simakov et al. (2020) through a line-by-line comparison. The differences are summarized in the final column of figure 1 and will now be described; these differences are set in full in [Supplementary figure S8, Supplementary Material](#) online.

First, as noted by Sacerdot et al. (2018), there are two combinatorial differences between the CLGs and paralogs: CLGA represents the combination of paralogs 3 and 13, whereas paralogon 14 is equivalent to the combination of CLGI and CLGQ. Simakov et al. (2020) noted that CLGI and CLGQ exhibited common synteny across jawed vertebrates, but they pointed out that orthologs were found in discrete *regions* on some vertebrate chromosomes: this may be seen in their Fig. 1a in the case of chicken chromosomes 4 and 6. The simplest interpretation of these comparisons is that, during the interval between the divergence of cephalochordates and the first round of WGD, CLGA underwent fission (to generate paralogs 3 and 13), whereas CLGI and CLGQ fused (to form paralogon 14).

More importantly, there are four discrepancies between the partitioning of quadruplicates between the two studies, as indicated by the asterisked letters in the final column of figure 1. For example, in paralogon 1, the B\* indicates that CLGB places LG12 and LG15 as diverging at 2R, whereas in the present study figure 1 shows this pair to have diverged at 1R. Similar discrepancies of " $\alpha$ - $\beta$ " pairings are seen for CLGG (paralogon 2), CLGA (paralogon 3), and CLGH (paralogon 4). Furthermore, the present study (analyzing seven bony vertebrate species) found every one of the 68 (= 17 × 4) PQ cells to be populated, whereas Simakov et al. (2020) (analyzing chicken, spotted gar and xenopus) reported the loss of 7 of the 68 potential quadruplicates.

In the present study, the attribution of duplications derives from the very high level (>95% in every case) of support for the position of the basal bifurcation in the molecular phylogenies for 33 families of ohnolog quartets (see the rooted trees in fig. 3A and [supplementary fig. S2, Supplementary Material](#) online, and in expanded format in [Supplementary Phylogenies](#)), together with five occurrences in figure 1 of paired sets of inter-1R-2R fusions.

#### Generality of the Paralogon Arrangement across Vertebrate Genomes

Although the >2,900 ohnologs in [supplementary table S1, Supplementary Material](#) online comprise fewer than 20% of all protein-coding vertebrate genes, there are two observations that suggest that the conclusions of this paper apply more generally. First, when the analysis of common orthologs between zebra finch and spotted gar is extended from just quartets and trios of ohnologs to a consideration of all common orthologs, the mappings in figures 5–7 do not change in

qualitative appearance, even though the number of genes examined is greatly increased. Second, figure 5D and [supplementary table S5C, Supplementary Material](#) online show that very few orthologs (<1.5%) occur as "outliers," on combinations of chromosomes other than those identified in figure 1. The total number of orthologs mapping between chromosomes in these two species was 10,092. This represents 61% of coding genes in zebra finch (16,619) and 55% of those in spotted gar (18,341). So it appears likely that the features shown in figure 1 apply broadly across the entire vertebrate genome.

The third supporting observation involves the distribution of curated ohnologs on chromosomes of each species. This is examined in [supplementary figure S1, Supplementary Material](#) online, for the >2,900 curated ohnologs. The panels for each species make it clear that the curated ohnologs are broadly distributed across the entire genome.

#### Ancestral Bony-Vertebrate Karyotype: 47 Chromosomes

The fusions of ancestral chromosomes that generated the bony-vertebrate karyotype are shown schematically in figure 4. Immediately before the first round of genome duplication, our chordate ancestor possessed 17 chromosomes, and upon duplication that complement doubled to 34. Prior to the second duplication, nine fusions occurred (indicated by arrows in column "Inter-1R-2R") reducing the number to 25. Immediately after the second genome duplication there were 50 chromosomes. Subsequently, another three fusions occurred (indicated by arrows in column "Post-2R") leaving 47 chromosomes in the last common ancestor of bony vertebrates.

This description is closely comparable with that of Sacerdot et al. (2018), despite the use of very different approaches. The allocation of genes to the 17 pre-1R chromosomes is very similar, as shown in figure 2. In addition, many of the inter-1R-2R fusions are common between the two studies, with seven of the nine fusions in figure 4 appearing in their Fig. 5. However, the present study additionally found the fusion of one copy of pre-1R chromosomes 16 and 17, and the fusion of one copy of pre-1R chromosome 12 with the already-fused 15 and 16. Those latter two fusions were instead previously placed as occurring after 2R. Of the three post-2R fusions reported in figure 4, only one (involving a copy of pre-1R chromosome 13) is present in Fig. 5 of Sacerdot et al. (2018).

#### Archeochromosomes

This study has demonstrated the existence of a substantial number of chromosomes, in certain extant species, that appear to have undergone no fusion or fission events since at least the time of the last common ancestor of bony vertebrates, or in many cases, since 2R; such chromosomes are here termed "archochromosomes." This identification is

most easily accomplished in cases where a pair of potentially matching archeochromosomes can be located in distantly related taxa, and in this study nine such pairs were identified between zebra finch and spotted gar (table 1).

Four criteria for identification were formulated, for application in cases where putative matching pairs were apparent, but in other cases where a tentative matching archeochromosome could not be found, two of the criteria had to be modified. In all cases, it was required that the curated ohnologs for the putative archeochromosome: 1) should be found in figure 1 in *only* one cell or one set of inter-1R-2R fused cells, and 2) should *not* be accompanied in that cell(s) by ohnologs from any other chromosome. The remaining criteria involved the absence of substantial gaps in the spatial distribution of genes along the putative archeochromosome.

In addition to the nine pairs of matching archeochromosomes between zebra finch and spotted gar, this study also identified another 13 archeochromosomes in zebra finch (table 1) where no obvious matching archeochromosome could be found in any of the distantly related species examined. Furthermore, another five zebra finch chromosomes remain as possible archeochromosomes or else fail to qualify in relatively minor ways. As a result, at least 22 (and possibly as many as 27) zebra finch chromosomes appear to have undergone no fusion or fission events in over 400 My. For spotted gar, in addition to the nine archeochromosomes identified as matching those in zebra finch, another two chromosomes were identified as archeochromosomes, and one more remains as a possible archeochromosome. For xenopus (*X. tropicalis*) one chromosome (Xet 6) was identified as an archeochromosome, and was found to match zebra finch chromosome Tgu 2. Together with reedfish chromosome Ecr 4, these were the only archeochromosomes identified in a species other than spotted gar or bird. Hence, of the 47 ancestral chromosomes deduced to have existed in the last common ancestor of bony vertebrates, at least 24 can still be identified in extant species as having neither fused nor split, and thereby remain in a form that is apparently little changed, apart from the loss of individual genes, or of local stretches of genes.

Why might archeochromosomes have persisted essentially unaltered for such immense periods of time in some species? First of all, genome evolution is known to have occurred very slowly in certain taxa; notably in turtles (Shaffer et al. 2013) and birds (O'Connor et al. 2019), as well as gar (Braasch et al. 2016), and the existence of archeochromosomes may simply reflect this slowness of change. Second, it was recently proposed (Huang and Rieseberg 2020) that the reason that major inter-chromosomal translocations are found in much lower abundance than chromosomal inversions is because of the lower likelihood of their establishment. Those authors pointed out that translocation heterozygotes involving different chromosomes would show mis-segregation during meiosis, and produce unbalanced and unviable gametes, and that the

resulting strong heterozygous disadvantage (under-dominance) would make it difficult for inter-chromosomal translocations to be established. On that view, the shuffling of genes within a chromosome is far more likely to occur than the fusion or fission of chromosomes.

### Archeochromosomes and Microchromosomes

Of the nine pairs of matching archeochromosomes between zebra finch and spotted gar, all are microchromosomes ( $\lesssim 20$  Mb). Of the other 15 archeochromosomes (including the matched Xet 6), 11 are no longer than 22 Mb, whereas the remaining four range in length from 32 up to 150 Mb. Hence, of the 24 ancestral chromosomes that can be identified as almost unchanged in extant species, 20 are microchromosomes and 4 are macrochromosomes. Therefore, one can assume that at least 20 of the 47 ancestral bony-vertebrate chromosomes are likely to have been microchromosomes and that at least 4 of the 47 were macrochromosomes. Thus, it seems that a mix of micro- and macrochromosomes must have been present at that stage, and that, of those that can still be identified, the great majority were microchromosomes.

The possible origin of microchromosomes as ancestral chromosomes and their apparent long-term conservation have previously been investigated in a number of studies (Burt 2002; Nakatani et al. 2007; Voss et al. 2011; Uno et al. 2012; Sacerdot et al. 2018). What has been added in the present study is the assignment of each microchromosome in zebra finch, chicken and spotted gar not only to a defined paralogon (or paralogons), but also to a distinct quadruplicate position, corresponding to a single ancestral chromosome that originated either at 2R or through an inter-1R-2R fusion.

For compactness, figure 1 did not present the analysis of chicken chromosomes, but examination of supplementary tables S1 and S7, Supplementary Material online indicates that the situation is very similar to that in zebra finch. It is well known that birds display a "signature" karyotype of  $\sim 10$  medium-sized chromosomes together with  $\sim 30$  microchromosomes that are often morphologically indistinguishable (Burt 2002; O'Connor et al. 2019). Cross-species analysis of over 70 bird species from 15 orders has shown remarkably little inter-chromosomal rearrangement between macrochromosomes (Griffin et al. 2007). Furthermore, a recent analysis of microchromosomes across 22 bird species from 10 orders showed that (except in falcons and parrots) there was no evidence of microchromosomal rearrangement (O'Connor et al. 2019), which led to the conclusion that the karyotype is remarkably stable across many of the  $\sim 10,000$  species of birds. Most recently, the chromosome-level assembly of the superb fairy-wren genome has confirmed that the macrochromosomes are largely conserved with other bird species, but has shown some fusions of microchromosomes with other chromosomes (Peñalba et al. 2020). Accordingly,

once the assignment of genes to paralogons and quadruplicates (as in fig. 1) is extended to species beyond chicken and zebra finch, it would seem likely that the occurrence of archeochromosomes would be found to be widespread across species of bird.

### Potential Importance of Archeochromosomes

The importance of archeochromosomes is likely to lie not so much in their existence per se, but in their potential utility for deciphering the different fusions and fissions of chromosomes that have occurred during the radiation of vertebrate species. Two examples of the way in which the mapping of archeochromosomes to chromosomes in other vertebrate species could assist in delineating the major chromosomal fusion and fission events that have occurred in different vertebrate lineages are illustrated in the Supplementary Results.

First, [supplementary figure S12A, Supplementary Material](#) online plots the mapping of orthologs on one zebra finch archeochromosome (Tgu 27) onto chromosomes of the five other species (apart from chicken) analyzed in this article. As would be predicted from its unique location in cell  $PQ = 1C$  of figure 1, orthologs covering essentially the entire length of Tgu 27 map to the illustrated chromosomes (LG15, Ecr 14, Xet 10, Mdo 2, and Hsa 17). Furthermore, as shown in figure 5D, in spotted gar they map to virtually no chromosomes other than LG15 (which is the paired archeochromosome).

Second, [supplementary figure S12B, Supplementary Material](#) online plots the mapping onto human chromosome 17 of orthologs from five zebra finch archeochromosomes, Tgu 14, 18, 19, 27, and R090 (scaffold RRCB01000090.1). These archeochromosomes were chosen because inspection of figure 1 shows that Hsa 17 occurs only in cells  $PQ = 4B, 4C, 2A, 1C,$  and  $[16A, 17A]$ . Examination of [supplementary figure S12B, Supplementary Material](#) online shows that the mappings from these five archeochromosomes encompass a substantial proportion of the length of Hsa 17. This leads to the conclusion that human chromosome Hsa 17 has resulted from fusions (in as-yet undetermined lineages) of the five ancestral chromosomes associated with those  $PQ$  cells. For the future, it will be of great interest to attempt to track the sequence of fusions and fissions of chromosomes that have occurred in the different major lineages, through analysis of the mapping of archeochromosomes onto the chromosomes of key vertebrate species.

### Final Remarks regarding 2R-WGD

The demonstration in this article that all four quadruplicate members are present in all 17 paralogons (fig. 1) supports the interpretation that 2R-WGD indeed corresponded to two rounds of duplication of the *entire* genome, and did not involve omission of any major blocks at either round. Furthermore, the finding (in fig. 3 and [supplementary fig.](#)

[S2, Supplementary Material](#) online) that, for multiple families of ohnologs that retain all four quadruplicates, the molecular phylogenies exhibit near-unanimous support for the common topology ((A, B),(C, D)) predicted by figure 1, provides powerful evidence for the attribution of duplications to 1R and to 2R, respectively, that is set out in figure 1. The finding that so many molecular phylogenies exhibit near-unanimous support for the 1R duplication node, together with the finding (fig. 4) of as many as nine chromosomal fusions during the inter-1R-2R period, would suggest that the time interval between 1R and 2R may have been rather longer than has commonly been assumed. And the finding of just three chromosomal fusions in the subsequent interval, between 2R and the radiation of bony vertebrates, would suggest that the latter interval might have been shorter than the interval between 1R and 2R.

## Materials and Methods

The methods used are set out below, with further details presented in [supplementary methods, Supplementary Material](#) online.

### Ensembl Biomart Data

Lists of genes and orthologs for the species examined were downloaded from Ensembl (Yates et al. 2020) using Biomart (Kinsella et al. 2011). All data appearing in [supplementary table S1, Supplementary Material](#) online were drawn from Ensembl 101 (August 2020). Details of the species and assemblies are listed in [supplementary table S9, Supplementary Material](#) online, and the criteria for acceptance of orthologs are given in [supplementary methods, Supplementary Material](#) online.

### Paralogy

For some families of ohnologs, the status of paralogs could be determined using Ensembl 101 gene phylogenies. However, as explained in the [supplementary methods, Supplementary Material](#) online, for many ohnolog families it was more appropriate to use Ensembl 93.

### Candidate Ohnologs: Singh and Isambert (2020)

Candidate ohnolog families for human were downloaded from Ohnologs v2 (ohnologs.curie.fr) (Singh et al. 2015; Singh and Isambert 2020) for the three pre-compiled conditions (strict, intermediate, and relaxed). For use with those lists, the applicable version of Ensembl human genes (release 84) was downloaded.

### Compilation of the Set of Curated Ohnologs

The set of curated ohnologs presented in [supplementary table S1, Supplementary Material](#) online was compiled using a

laborious manual process that is explained in the [supplementary methods, Supplementary Material](#) online. Briefly, a search was made for potential three- and four-member families within the list of candidate ohnologs set out in Ohnologs V2 (Singh and Isambert 2020). This searching was assisted by a set of custom Matlab (The MathWorks) scripts that are provided in the supplementary online material. For each potential family, I manually downloaded the Ensembl gene tree and examined it carefully, rejecting any tree that clearly included protostome sequences, and flagging in [supplementary table S1, Supplementary Material](#) online any that the Ensembl gene trees suggested may possibly have included either tunicate or protostome sequences. Subsequently, this search was extended using the list of human genes assembled by Sacerdot et al. (2018) in their Supplementary Table S8; that list was used solely to identify potential ohnologs, and their assignment to paralogs was ignored until after the curated list had been finalized. Their analysis used Ensembl 69 (October 2012), which is no longer available in the Ensembl archive, and so the closest currently available version of Ensembl human genes (release 67) was downloaded. In addition, further potential families were found by serendipity, through browsing of gene trees in Ensembl 93. As a result of these procedures, the great majority of three- and four-member families found in [supplementary table S1, Supplementary Material](#) online are also contained within the lists of either or both of the two earlier studies, but on the other hand a substantial number of the families in their lists have been excluded, or sometimes split into more than one family, after careful examination of recent Ensembl gene trees. Figures 2 and 5–7, and [supplementary tables S1–S7, Supplementary Material](#) online, were produced by the custom Matlab scripts, as described in [supplementary methods, Supplementary Material](#) online.

### Molecular Phylogenies

Amino acid sequences for the ohnolog quartets to be tested were downloaded from NCBI, typically for 20–24 vertebrate species ([supplementary table S10, Supplementary Material](#) online). Sequences were aligned using MAFFT v7.471 (Katoh and Standley 2013) with its L-ins-i option. The maximum-likelihood phylogeny was inferred using IQ-Tree 2.1.1 (Nguyen et al. 2015; Minh et al. 2020) with protein substitution model WAG (Whelan and Goldman 2001) and with the ultrafast bootstrap approximation (Hoang et al. 2018) using 10,000 pseudo-replicates. Collapsed trees are shown in figure 3, and fully expanded trees in the supplementary molecular phylogenies, [Supplementary Material](#) online.

### Supplementary Material

[Supplementary material](#) are available at *Genome Biology and Evolution* online.

### Acknowledgments

I am most grateful to Dr Hardip Patel, Professor Dan Larhammar, and two anonymous reviewers, for constructive comments.

### Data Availability

All custom code and data underlying this article are available in the article and its online [supplementary material](#).

### Literature Cited

- Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H. 2002. Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* 31(1):100–105.
- Braasch I, et al. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat. Genet.* 48(4):427–437.
- Burt DW. 2002. Origin and evolution of avian microchromosomes. *Cytogenet. Genome Res.* 96(1–4):97–112.
- Dehal P, Boore JL. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3(10):e314.
- Griffin DK, Robertson LBW, Tempest HG, Skinner BM. 2007. The evolution of the avian genome as revealed by comparative molecular cytogenetics. *Cytogenet. Genome Res.* 117(1–4):64–77.
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35(2):518–522.
- Hokamp K, McLysaght A, Wolfe KH. 2003. The 2R hypothesis and the human genome sequence. *J. Struct. Funct. Genomics* 3(1–4):95–110.
- Holland PW. 1999. Gene duplication: past, present and future. *Semin. Cell Dev. Biol.* 10(5):541–547.
- Huang K, Rieseberg LH. 2020. Frequency, origins, and evolutionary role of chromosomal inversions in plants. *Front. Plant Sci.* 11:296.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30(4):772–780.
- Kinsella RJ, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database* 2011:bar030.
- Kohn M, et al. 2006. Reconstruction of a 450-My-old ancestral vertebrate protokaryotype. *Trends Genet.* 22(4):203–210.
- Lagman D, Sundström G, Ocampo Daza D, Abalo XM, Larhammar D. 2012. Expansion of transducin subunit gene families in early vertebrate tetraploidizations. *Genomics* 100(4):203–211.
- Lamb TD. 2020. Evolution of the genes mediating phototransduction in rod and cone photoreceptors. *Prog. Retin. Eye Res.* 76:100823.
- Larhammar D, Lundin L-G, Hallböök F. 2002. The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Res.* 12(12):1910–1920.
- Lundin L-G, Larhammar D, Hallböök F. 2003. Numerous groups of chromosomal regional paralogies strongly indicate two genome doublings at the root of the vertebrates. *J. Struct. Funct. Genomics* 3(1–4):53–63.
- Minh BQ, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37(5):1530–1534.
- 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(9):1254–1265.
- Naz R, Tahir S, Abbasi AA. 2017. An insight into the evolutionary history of human MHC paralogon. *Mol. Phylogenet. Evol.* 110:1–6.



- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32(1):268–274.
- O'Connor RE, et al. 2019. Patterns of microchromosome organization remain highly conserved throughout avian evolution. *Chromosoma* 128(1):21–29.
- Ohno S. 1970. *Evolution by Gene Duplication*. London: Allen and Unwin.
- Pébusque MJ, Coulier F, Birnbaum D, Pontarotti P. 1998. Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. *Mol. Biol. Evol.* 15(9):1145–1159.
- Peñalba JV, et al. 2020. Genome of an iconic Australian bird: high-quality assembly and linkage map of the superb fairy-wren (*Malurus cyaneus*). *Mol. Ecol. Resour.* 20(2):560–578.
- Putnam NH, et al. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453(7198):1064–1071.
- Renwick JH. 1971. The mapping of human chromosomes. *Annu. Rev. Genet.* 5:81–120.
- Sacerdot C, Louis A, Bon C, Berthelot C, Roest Crolius H. 2018. Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol.* 19(1):166.
- Shaffer HB, et al. 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* 14(3):R28.
- Simakov O, et al. 2020. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* 4(6):820–830.
- Simakov O, et al. 2015. Hemichordate genomes and deuterostome origins. *Nature* 527(7579):459–465.
- Singh PP, Arora J, Isambert H. 2015. Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. *PLoS Comput. Biol.* 11(7):e1004394.
- Singh PP, Isambert H. 2020. OHNOLOGS v2: a comprehensive resource for the genes retained from whole genome duplication in vertebrates. *Nucleic Acids Res.* 48(D1):D724–D730.
- Smith JJ, et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* 45(4):415–421. 421e1–2.
- Smith JJ, et al. 2018. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* 50(2):270–277.
- Smith JJ, Keinath MC. 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res.* 25(8):1081–1090.
- Uno Y, et al. 2012. Inference of the protokaryotypes of amniotes and tetrapods and the evolutionary processes of microchromosomes from comparative gene mapping. *PLoS One* 7(12):e53027.
- Voss SR, et al. 2011. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res.* 21(8):1306–1312.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18(5):691–699.
- Yates AD, et al. 2020. Ensembl 2020. *Nucleic Acids Res.* 48(D1):D682–D688.

Associate editor: Kenneth Wolfe