ELSEVIER

Evolution and diversity of fish genomes

Byrappa Venkatesh

The ray-finned fishes ('fishes') vary widely in genome size, morphology and adaptations. Teleosts, which comprise ~23,600 species, constitute >99% of living fishes. The radiation of teleosts has been attributed to a genome duplication event, which is proposed to have occurred in an ancient teleost. But more evidence is required to support the genome-duplication hypothesis and to establish a causal relationship between additional genes and teleost diversity. Fish genomes seem to be 'plastic' in comparison with other vertebrate genomes because genetic changes, such as polyploidization, gene duplications, gain of spliceosomal introns and speciation, are more frequent in fishes.

Addresses
Institute of Molecular and Cell Biology 30, Medical Drive, Singapore 117609, Singapore
e-mail: mcbbv@imcb.nus.edu.sg

Current Opinion in Genetics & Development 2003, **13**:588–592

This review comes from a themed issue on
Genomes and evolution
Edited by Evan Eichler and Nipam Patel

0959-437X/$ – see front matter

DOI 10.1016/j.gde.2003.09.001

Abbreviation
My million years

Introduction

The ray-finned fishes, which comprise ~23,700 extant species [1], are the most diverse and successful group of vertebrates. They show vast differences in their morphology and adaptations. Their sister group, the lobe-finned fishes, include the rest of the bony vertebrates, such as coelacanths, lungfishes and tetrapods, and are represented by ~23,600 living species (Figure 1). The two bony vertebrate lineages diverged ~450 million years (My) ago [2]. The ray-finned fishes ('fishes') can be subdivided into the basal 'non-teleosts', represented by four major lineages: Polypteriformes (bichirs), Acipenseriformes (sturgeons and paddlefish), Semionotiformes (gar) and Amiiformes (bowfin); and the higher teleosts. Teleosts are the largest group of vertebrates and comprise ~23,600 species. The most ancient teleost fossil is ~235 My old [3], and fossils of diverse teleost species have been recorded from Jurassic and Cretaceous times. Thus, teleosts appear to have undergone a rapid radiation that is unparalleled in other vertebrate taxa.

Although traditionally fishes have been the subject of comparative studies, recently there has been an increased interest in these vertebrates as model organisms in genomics and molecular genetics. Indeed, the second vertebrate genome to be sequenced completely was that of a pufferfish (*Fugu rubripes*) [4••], the first being the human genome. The genome of another pufferfish (*Tetraodon nigroviridis*) is essentially complete, and that of the zebrafish (*Danio rerio*) is nearing completion. The genome of a fourth fish, medaka (*Oryzias latipes*), is also being sequenced.

The analyses of the fish genome sequences have provided useful information for understanding the structure, function and evolution of vertebrate genes and genomes. In this review, I discuss the insights gained from recent studies on the evolution of fish genomes.

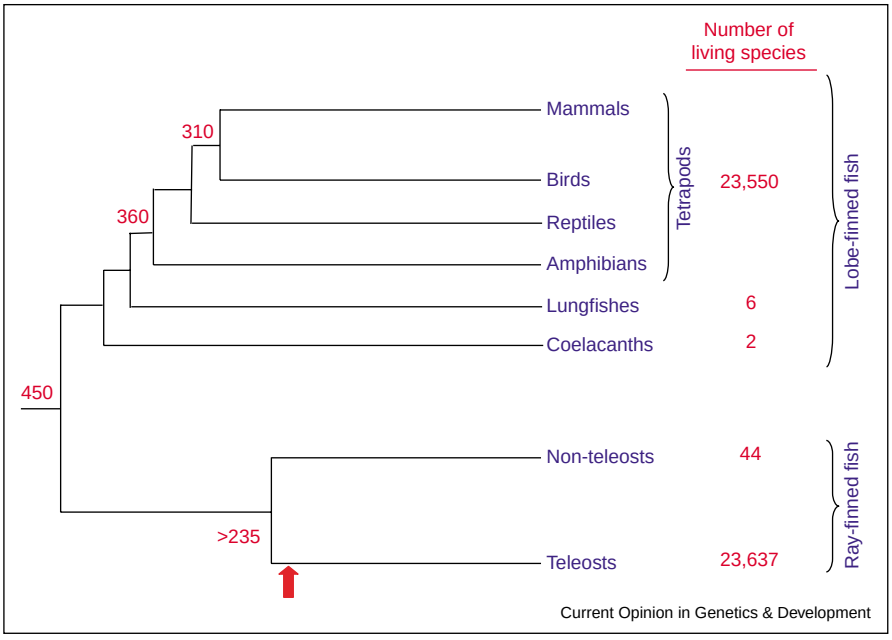
Genome size of fishes

Fish genomes vary widely in size, from 0.39 pg to >5 pg of DNA per haploid cell (Figure 2), with a modal value of ~1 pg (equivalent to ~1000 Mb). Most of the large genomes (>2 pg) are polyploids. Among vertebrates, polyploidization is common only in fishes, amphibians and reptiles. In fishes, polyploidization has occurred independently in several lineages including non-teleosts such as the paddlefish, shovelnose sturgeon and spotted gar, as well as teleosts such as cyprinids (carps), cyprinodontiformes (live bearers), catostomids (suckers) and salmonids [5,6]. In fact, all members of the families Catostomidae and Salmonidae are polyploids [5,6].

The pufferfish — family Tetraodontidae, including, for example, *Fugu* and *Tetraodon* — have the smallest genomes among vertebrates that have been characterized to date. Their genomes therefore offer an interesting model for understanding the evolutionary forces that lead to a reduction in genome size. A paucity of repetitive elements is clearly one of the factors that contributes to the compact genome size of pufferfish. Interestingly, although the repetitive sequences account for <15% of the *Fugu* genome, almost every class of transposable elements known in eukaryotes is represented in *Fugu*. Furthermore, a large number of transposable elements (40 families as compared with 6 in the human genome) seem to be of recent origin, as they have accumulated substitutions at a level of <5% [4••]. This indicates that the pufferfish genome is susceptible to transposable elements, but the propagation of these elements is somehow restricted.

Interspersed repeats of the same divergence level in *Fugu* and humans have more small internal deletions in *Fugu*

Figure 1

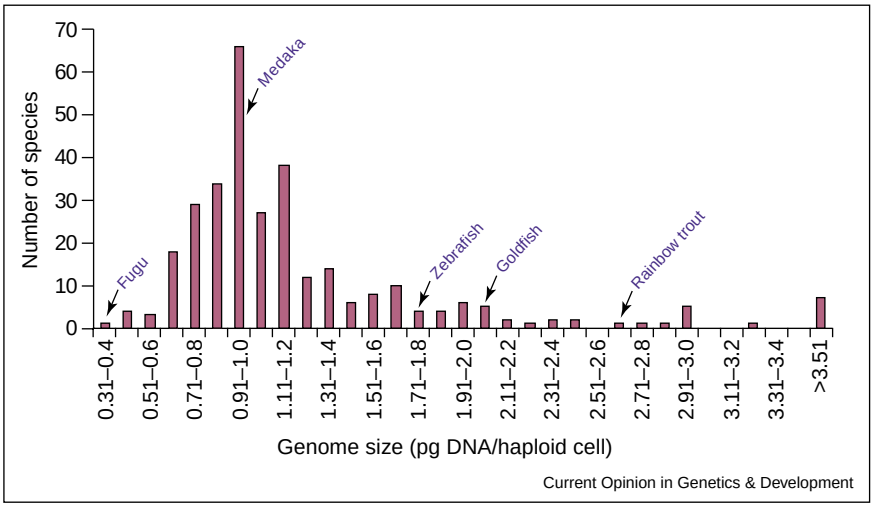


Evolution of bony vertebrates. 'Non-teleosts' is not a taxonomic group: it includes the basal groups Polypteriformes, Acipenseriformes, Semionotiformes and Amiiformes, which are not teleosts. The numbers at the nodes are the divergence time in million years [2]. There is no reliable estimate for the divergence time of teleosts, but the oldest fossil record of teleosts is ~235 My old [3]. The data on numbers of living species are from [1]. A whole-genome duplication has been proposed to have occurred in an ancient teleost (indicated by arrow).

than in humans [4••], indicating that deletions occur at a higher rate in *Fugu*. Similarly, the overall frequency of small deletions in pseudogenes has been found to be higher in another pufferfish, *Tetraodon*, than in mammals [7]. Such a bias for DNA loss provides a mechanism for inactivating and deleting transposable elements and redundant genes, and probably accounts for the lower abundance of such sequences in pufferfish than in mammals.

In addition to DNA loss, these pufferfish (referred to as 'smooth puffers') seem to be subject to other mechanisms that minimize their genomes. For example, comparisons

Figure 2



Distribution of genome size of fishes. Genome size is given in picograms of DNA per haploid cell: 1 pg of DNA is roughly equivalent to 1000 Mb. Genome sizes of 312 species [8,42,43] belonging to 30 of the known 42 orders of fishes are represented. Only one representative species for each genera is included.

of mutation profiles between the smooth puffers (Tetraodontidae) and their sister group the spiny puffers (Diodontidae), whose genome size (~800 Mb) is twice that of the smooth puffers [8], suggest that a reduction in the rate of large insertions — rather than an increase in large deletions — was the probable cause of the reduction in the genome size of the smooth puffers after they diverged from the spiny puffers [9•]. Thus, besides a high rate of DNA loss, bias against insertion of large DNA elements may be responsible for the 'smallest vertebrate' genomes of the smooth puffers.

Gene duplications in fishes

Since the first identification of additional *Hox* gene clusters in diploid teleosts such as zebrafish, *Fugu* and medaka [10–12], other additional duplicate genes have been identified in these and other teleosts [13–24,25•]. Many of the duplicate genes in zebrafish and *Fugu* map to similar pairs of chromosome segments, suggesting that they arose as a result of large-segment or whole-chromosome duplications [22,24,25•,26,27••]. Furthermore, orthologs for 22 of the 49 pairs of duplicate zebrafish genes have been identified in the *Fugu* [27••]. These observations have led to the hypothesis that a whole-genome duplication occurred in an ancestor common to the *Fugu* and zebrafish lineages [10,20,27••]. Because zebrafish and *Fugu* are phylogenetically distant (they are grouped under the subdivision Euteleostei, which includes >90% of extant teleost species), it has been proposed that the duplication event occurred before the radiation of teleosts [27••].

It has been also argued that the abundance of duplicate genes in teleosts might be due to independent gene duplications in different lineages rather than to a whole-genome duplication [14,15]. Phylogenetic analysis of 37 gene families from three or more different teleost lineages has shown that gene duplications occurred in only 18 gene families. Of these 18 families, duplications in 7 families arose in a common ancestor whereas the duplications in the remaining families occurred independently in different lineages [15]. Furthermore, gene trees for some of the zebrafish duplicate genes do not show a topology consistent with the whole-genome duplication hypothesis [20,27••]. These results suggest that some of the duplicate genes in zebrafish and other teleosts might be the result of independent gene duplications.

Tracing the history of ancient genome duplication events is rather difficult because of secondary losses of genes or whole chromosomes, chromosomal rearrangements, independent duplications and different evolutionary rates of duplicate gene copies. Even the strongest evidence for an ancient whole-genome duplication can be only a statistical argument based on the size and number of duplicated segments in different lineages, the distribution of duplication times, and the congruence between gene duplication and speciation events.

To date, gene-duplication studies in fishes are limited to a small number of gene loci in only a few species. Thus, to generate strong statistical evidence, it is necessary to investigate duplication events in a large number of loci in diverse lineages, including a basal lineage. Comparisons of the completed genome sequences of *Fugu* — the present 'draft' is in the form of 12,000 fragments and lacks chromosome coordinates — and other teleosts that are being sequenced should provide useful data for tracing the history of duplications.

Duplicate genes and teleost radiation

It has been suggested that the vast morphological and species diversity of teleosts might be related to large-scale independent gene duplications or to a whole-genome duplication in an ancient teleost [10,27••,28]. After gene duplication, either one of the duplicates is silenced and eliminated, or both of the duplicates are retained through mutations that divide the functions between the two or that confer a novel function on one of them. Contrary to previous thinking that silencing of a duplicate gene copy has no consequence for the species, the recently proposed 'reciprocal silencing' and 'divergent resolution' models show that the silencing of different copies of duplicate genes in allopatric populations can genetically isolate populations, thereby spurring speciation [29,30]. Furthermore, different subfunctionalization patterns of duplicates in different populations can also lead, like gene silencing, to genetic isolation [30].

A whole-genome duplication generates thousands of duplicate genes that can be selectively silenced in different populations or retained with partitioned function, leading to genetic isolation and speciation. The vast diversity of species in tetraploid families such as Salmonidae and Catostomidae, which underwent polyploidization between 25 and 100 My ago [5,6], is often cited as an example of species radiation that followed genome duplications [31]. A lack of species diversity among polyploid amphibians and reptiles indicates, however, that genome duplication alone is not sufficient to drive species diversity.

Whole-genome duplication also provides raw genetic material for the evolution of genes with novel functions. Nevertheless, although there are several cases of duplicate fish genes that apparently share the functions of their single ortholog in mammals [16–19,21,24,25•,32], not many examples of duplicate fish genes that have acquired novel function are known. One classic example of a duplicate gene that has acquired a novel function is the antifreeze protein gene in Antarctic fishes that evolved from a protease gene [33].

Although some duplicate zebrafish genes show expression patterns and functions that apparently differ from those of their single ortholog in mammals [25•,32], it is unclear

whether these genes have acquired new functions since the duplication. Given that only a limited number of duplicate fish genes have been investigated to date, it remains to be seen if the list of duplicate genes with novel functions will grow and whether a causal relationship between duplicate genes and the diversity of teleosts can be demonstrated.

'Plastic' genomes?

The fish genomes seem to undergo genetic changes more rapidly than do other vertebrate genomes, suggesting that fish genomes are 'plastic' as compared with the genomes of other vertebrates. Besides polyploidization, several independent gene duplications seem to have occurred in fishes [14,15]. The fastest known rate of vertebrate speciation has been recorded among fishes: the ~500 species of cichlids colonizing Lake Victoria in East Africa have been shown to have evolved from only a few ancestors within the past 100,000 years [34••]. The fish lineage has 'gained' many spliceosomal introns after it diverged from the ancestor of the mammalian lineage [35]. By contrast, intron gain is extremely rare in mammals [36]. Consistent with these findings, the *Fugu* genome has more spliceosomal introns than the human genome, although they both contain a similar number of genes encoding proteins [4••]. A comparison of the evolutionary rates of fish genes, albeit based on a small data set, has shown that fish genes may be accumulating substitutions faster than mammalian genes [37].

The wide spectrum of sex and sex determination in fishes perhaps illustrates the plasticity of fish genomes. Many fishes exhibit hermaphroditism, and some even change sex at a specific stage in their life cycle. Fishes also show a range of sex determination mechanisms, from classical male or female heterogametic sex to environmental and hormonal sex determination [38]. The search for the fish equivalent of mammalian *Sry* gene has proved fruitless for a long time. Recently, a strong candidate gene, *dmY* (also called *dmrt1Y*) was identified in medaka (*O. latipes*), in which the male is heterogametic as in mammals [39•,40•]. Notably, this gene was not found in other fishes investigated, including the closely related species *Oryzias celebensis* [41]. Thus, it seems that *dmY* has been recently recruited for sex determination in medaka. This illustrates the continued evolution of sex-determining mechanisms in fishes.

Conclusions

Fishes comprise slightly more than half of living bony vertebrates, and teleosts account for >99% of living fishes. Palaeontological evidence suggests that the radiation of teleosts occurred between 150 and 250 My ago. A whole-genome duplication in an ancestral teleost has been proposed to have provided the genetic raw material to spur the teleost radiation, but more evidence is required to confirm this hypothesis.

In addition to the proposed whole-genome duplication, independent gene duplications as well as polyploidization have occurred in different teleost lineages. The extent and contribution of independent gene duplications to the abundance of genes in diploid fish genomes remains to be ascertained. Comparisons of the genome sequences of *Fugu* and other teleosts such as *Tetraodon*, zebrafish and medaka, which are currently being sequenced, should provide useful data that can shed light on the history of gene duplications and the diversity of teleosts.

Comparative genomics of representative basal fishes such as bichirs and bowfin will be informative in validating the fish-specific whole-genome duplication hypothesis. The availability of the whole-genome sequences of the two pufferfishes *Fugu* and *Tetraodon* provides an unprecedented opportunity for understanding the genetic basis of evolutionary changes between two closely related vertebrate species.

Acknowledgements

The author thanks members of his laboratory for discussion and suggestions.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Nelson JS: *Fishes of the World*, edn 3. New York: Wiley & Sons; 1994:600.

2. Kumar S, Hedges SB: **A molecular timescale for vertebrate evolution**. *Nature* 1998, **392**:917-920.

3. Maisey JG: *Discovering Fossil Fishes*. New York: Henry Holt and Company; 1996:223.

4. •• Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A *et al.*: **Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes***. *Science* 2002, **297**:1301-1310.
This paper reports the landscape, gene complement and predicted proteome of the *Fugu* genome, the first fish genome to be sequenced to the 'draft' level. The *Fugu* genome encodes almost the same number of genes as the human genome; however, nearly 25% of human proteins have no matches to predicted *Fugu* proteins. This set of genes is thought to have diverged significantly after the split of the mammalian and fish lineages. Chromosomal segments conserved between *Fugu* and human comprise ~14% of the *Fugu* genome.

5. Allendorf FW, Thorgaard GH: **Tetraploidy and the evolution of Salmonid fishes**. In *Evolutionary Genetics of Fishes*. Edited by Turner BJ. New York: Plenum Press; 1984:1-53.

6. Ferris SD: **Tetraploidy and the evolution of the catostomid fishes**. In *Evolutionary Genetics of Fishes*. Edited by Turner BJ. New York: Plenum Press; 1984:55-93.

7. Dasilva C, Hadji H, Ozouf-Costaz C, Nicaud S, Jaillon O, Weissenbach J, Crollius HR: **Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome**. *Proc Natl Acad Sci USA* 2002, **99**:13636-13641.

8. Brainerd EL, Slutz SS, Hall EK, Phillis RW: **Patterns of genome size evolution in tetraodontiform fishes**. *Evolution* 2001, **55**:2363-2368.

9. • Neafsey DE, Palumbi SR: **Genome size evolution in pufferfish: a comparative analysis of diodontid and tetraodontid pufferfish genomes**. *Genome Res* 2003, **13**:821-830.

On the basis of a comparison of repetitive elements, deletions and insertions between *Fugu* and a spiny pufferfish whose genome is twice that of *Fugu*'s, the authors conclude that a bias against large-scale insertions in the *Fugu* is responsible for further reduction in the genome size of this family of puffers after they diverged from the spiny puffers.

10. Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho RK, Langeland J, Prince V, Wang YL *et al.*: **Zebrafish *hox* clusters and vertebrate genome evolution**. *Science* 1998, **282**:1711-1714.

11. Aparicio S, Hawker K, Cottage A, Mikawa Y, Zuo L, Venkatesh B, Chen E, Krumlauf R, Brenner S: ***Hox* clusters: evidence for continuing evolution of vertebrate *Hox* complexes**. *Nat Genet* 1997, **16**:79-83.

12. Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R *et al.*: **A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution**. *Genetics* 2000, **154**:1773-1784.

13. Woods IG, Kelly PD, Chu F, Ngo-Hazelett P, Yan YL, Huang H, Postlethwait JH, Talbot WS: **A comparative map of the zebrafish genome**. *Genome Res* 2000, **10**:1903-1914.

14. Robinson-Rechavi M, Marchand O, Escriva H, Bardet P, Zelus D, Hughes S, Laudet V: **Euteleost fish genomes are characterized by expansion of gene families**. *Genome Res* 2001, **11**:781-788.

15. Robinson-Rechavi M, Marchand O, Escriva H, Laudet V: **An ancient whole-genome duplication may not have been responsible for the abundance of duplicated fish genes**. *Curr Biol* 2001, **11**:458-459.

16. Chiang EF, Pai CI, Wyatt M, Tan YL, Postlethwait J, Chung B: **Two *sox9* genes on duplicate zebrafish chromosomes: expression of similar transcription activators in distinct sites**. *Dev Biol* 2001, **231**:149-163.

17. Chiang EF, Yan YL, Guiguen Y, Postlethwait J, Chung B: **Two *Cyp19* (P450 aromatase) genes on duplicated zebrafish chromosomes are expressed in ovary or brain**. *Mol Biol Evol* 2001, **18**:542-550.

18. Lister JA, Close J, Raible DW: **Duplicate *mitf* genes in zebrafish: complementary expression and conservation of melanogenic potential**. *Dev Biol* 2001, **237**:333-344.

19. Serluca FC, Sidow A, Mably JD, Fishman MC: **Partitioning of tissue expression accompanies multiple duplications of the Na^+/K^+ ATPase α subunit gene**. *Genome Res* 2001, **11**:1625-1631.

20. Taylor JS, Van de Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish**. *Philos Trans R Soc Lond B Biol Sci* 2001, **356**:1661-1679.

21. Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L, Volff JN, Schartl M: **Subfunctionalization of duplicate *mitf* genes associated with differential degeneration of alternative exons in fish**. *Genetics* 2002, **161**:259-267.

22. Smith SF, Snell P, Gruetzner F, Bench AJ, Haaf T, Metcalfe JA, Green AR, Elgar G: **Analyses of the extent of shared synteny and conserved gene orders between the genome of *Fugu rubripes* and human 20q**. *Genome Res* 2002, **12**:776-784.

23. Williams H, Brenner S, Venkatesh B: **Identification and analysis of additional copies of the platelet-derived growth factor receptor and colony stimulating factor1 receptor genes in fugu**. *Gene* 2002, **295**:255-264.

24. Yu WP, Brenner S, Venkatesh B: **Duplication, degeneration and subfunctionalization of the nested *synapsin-Timp* genes in *Fugu***. *Trends Genet* 2003, **19**:180-183.

25. • Winkler C, Schafer M, Duschl J, Schartl M, Volff JN: **Functional divergence of two zebrafish midkine growth factors following fish-specific gene duplication**. *Genome Res* 2003, **13**:1067-1081.
The authors demonstrate that the duplicate *midkine* genes in zebrafish, which arose as a result of fish-specific large-segmental duplications, have undergone functional divergence and show expression pattern and functions that are distinct from that of the single-copy mammalian *midkine* gene.

26. Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS: **Zebrafish comparative genomics and the origins of vertebrate chromosomes**. *Genome Res* 2000, **10**:1890-1902.

27. •• Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y: **Genome duplication, a trait shared by ~22,000 species of ray-finned fish**. *Genome Res* 2003, **13**:382-390.
The authors extend their previous work [20] by identifying *Fugu* orthologs for 22 pairs out of 58 pairs of zebrafish genes and show by phylogenetic analysis that these genes duplicated before the divergence of the *Fugu* and zebrafish. This indicates that there has been a large-scale gene duplication, or a whole-genome duplication, in the ancestor of *Fugu* and zebrafish.

28. Meyer A, Schartl M: **Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions**. *Curr Opin Cell Biol* 1999, **11**:699-704.

29. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes**. *Science* 2000, **290**:1151-1155.

30. Lynch M, Force A: **The origin of interspecific genomic incompatibility via gene duplication**. *Am Nat* 2000, **156**:590-605.

31. Taylor JS, Van de Peer Y, Meyer A: **Genome duplication, divergent resolution and speciation**. *Trends Genet* 2001, **17**:299-301.

32. McClintock JM, Carlson R, Mann DM, Prince VE: **Consequences of Hox gene duplication in the vertebrates: an investigation of the zebrafish Hox paralogue group 1 genes**. *Development* 2001, **128**:2471-2484.

33. Cheng CC, Chen L: **Evolution of an antifreeze glycoprotein**. *Nature* 1999, **401**:443-444.

34. •• Verheyen E, Salzburger W, Snoeks J, Meyer A: **Origin of the superflock of cichlid fishes from lake Victoria, East Africa**. *Science* 2003, **300**:325-329.
This study traces the origin of cichlid species in Lake Victoria in East Africa to the geologically older Lake Kivu. On the basis of mitochondrial sequence analysis and geological evidence, the authors show that >500 cichlid species colonizing Lake Victoria evolved during the past 100,000 years.

35. Venkatesh B, Ning Y, Brenner S: **Late changes in spliceosomal introns define clades in vertebrate evolution**. *Proc Natl Acad Sci USA* 1999, **96**:10267-10271.

36. Roy SW, Fedorov A, Gilbert W: **Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain**. *Proc Natl Acad Sci USA* 2003, **100**:7158-7162.

37. Robinson-Rechavi M, Laudet V: **Evolutionary rates of duplicate genes in fish and mammals**. *Mol Biol Evol* 2001, **18**:681-683.

38. Chan STH, Yeung WSB: **Sex control and sex reversal in fish under natural conditions**. In *Fish Physiology*, vol 9B. Edited by Hoar WS, Randall DJ, Donaldson EW. New York: Academic Press; 1983:171-222.

39. • Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N *et al.*: ***DMY* is a Y-specific DM-domain gene required for male development in the medaka fish**. *Nature* 2002, **417**:559-563.
See annotation [40•].

40. • Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, Shan Z, Haaf T, Shimizu N, Shima A *et al.*: **A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes***. *Proc Natl Acad Sci USA* 2002, **99**:11778-11783.
These two groups [39•,40•] independently identified the medaka gene, *dmY* (also known as *dmrt1Y*) as a strong candidate for male sex determination. This is the first sex-determining gene identified in fishes.

41. Kondo M, Nanda I, Hornung U, Asakawa S, Shimizu N, Mitani H, Schmid M, Shima A, Schartl M: **Absence of the candidate male sex-determining gene *dmrt1b(Y)* of medaka from other fish species**. *Curr Biol* 2003, **13**:416-420.

42. Hinegardner R, Rosen DE: **Cellular DNA content and the evolution of teleostean fishes**. *Am Nat* 1972, **106**:621-644.

43. Hinegardner R: **The cellular DNA content of sharks, rays and some other fishes**. *Comp Biochem Physiol* 1976, **55**:367-370.