

papers from the IHGSC<sup>1</sup> and She *et al.*<sup>4</sup> argue for a hybrid strategy in which WGS is supplemented by a modest amount of BAC cloning and mapping. This would protect draft WGS sequences from some of the 'simplification' reported by She *et al.* and provide the clones needed for finishing selected regions of special interest.

What is next for the human genome project? Even with a finished sequence in hand there is much still to do. Surprisingly, one task is to develop the definitive catalogue of protein-coding genes. In the current paper<sup>1</sup>, the number is estimated to be between 20,000 and 25,000. This wide range reflects limitations to state-of-the-art gene-prediction software that leave doubts about the validity of many predicted genes. One promising approach is to use comparative genomics to align the human genome with the genomes of other animals. Because natural selection ensures that functional regions are more highly conserved than non-functional ones, this approach highlights candidate protein-coding regions. The same approach shows promise for finding other functional elements such as gene promoters, which control the timing and level of expression of genes, and micro-RNAs, which have been implicated as regulatory agents of many developmental processes.

Much farther in the future is the task of sequencing the remaining 20% of the genome that lies within heterochromatin, the gene-poor, highly repetitive sequence that is implicated in the processes of chromosome replication and maintenance. The repetitiveness of heterochromatin means that it cannot be tackled using current sequencing methods, and new technologies will have to be developed to attack it. So don't be shocked to see another paper announcing the 'finishing' of the human genome in 2010 — it will describe how the heterochromatin problem has been cracked.

In sequencing the human genome, researchers have already climbed mountains and travelled a long and winding road. But we are only at the end of the beginning: ahead lies another mountain range that we will need to map out and explore as we seek to understand how all the parts revealed by the genome sequence work together to make life.

Lincoln D. Stein is at Cold Spring Harbor

Laboratory, 1 Bungtown Road,

Cold Spring Harbor, New York 11724, USA.

e-mail: lstein@cshl.edu

1. International Human Genome Sequencing Consortium *Nature* **431**, 931–945 (2004).

2. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).

3. Venter, J. C. *et al.* *Science* **291**, 1304–1351 (2001).

4. She, X. *et al.* *Nature* **431**, 927–930 (2004).

split into two distinct evolutionary lineages: the actinopterygians (ray-finned fish), which include teleosts such as pufferfish and zebrafish, and the sarcopterygians (lobe-fins), which include lungfish, coelacanths and ourselves (Fig. 1). By matching up the genes on each pufferfish chromosome to the related genes on human chromosomes, Jaillon *et al.* deduce that the extinct ancestor had 12 pairs of chromosomes. Work on partially completed genome sequences had suggested this number<sup>4,5</sup>, but the new analyses add fascinating detail to the picture. For example, it is now possible to say which genes were on which chromosomes, despite this unknown animal having been extinct for more than 400 million years.

One puzzling observation concerns the apparent stability of the genomes of ray-finned fish. It seems that the ancestral genome underwent as few as ten large inter-chromosomal rearrangements (exchanges, fissions or fusions) to give rise to the present-day *Tetraodon* genome. Indeed, 11 *Tetraodon* chromosomes have not experienced any such rearrangements. Only one human chromosome (14) can make the same claim, despite the timescale being identical.

Although the genomes of ray-finned fish may have been slowly evolving in terms of chromosome breakages and fusions, they have experienced a cataclysmic event in their history. Jaillon and colleagues' analyses of the complete *Tetraodon* genome sequence show clearly that a duplication of the whole genome occurred sometime within the ray-finned-fish lineage. This inference is not new, having been previously suggested from analyses of the Hox-gene clusters and other gene families in zebrafish, *Takifugu* and other teleosts<sup>4–7</sup>, but the conclusion has remained controversial<sup>8</sup>.

Two new analyses should now settle the issue, however. First, Jaillon and colleagues plotted the chromosome positions for about 750 pairs of 'ancient' duplicated genes within the *Tetraodon* genome, revealing related pairs of chromosomes or chromosomal regions. Every chromosome is involved, consistent with an ancient whole-genome duplication. In the second test, chromosome positions for more than 6,000 pufferfish genes were compared with the positions of related genes in the human genome. This revealed a striking pattern of 'double conserved synteny', meaning that one chromosomal region in humans matches two in pufferfish, across the entire genome. This is a clear echo of whole-genome duplication in the ray-finned-fish lineage. Every gene, on every chromosome, was duplicated, although there has since been a massive degree of gene loss and local gene shuffling.

When did this whole-genome duplication occur? Analysis of zebrafish genetic maps strongly suggests that this species also underwent such an event in its history<sup>4</sup>.

## Comparative genomics

# Small genome, big insights

John Mulley and Peter Holland

The genome of a second pufferfish species has been sequenced. Why is this important? Because comparing this genome with that of other animals yields a wealth of information on genome evolution.

It is still early days for the field of comparative genomics. Only around a dozen species of animal have so far had their complete DNA sequence determined, even to draft coverage. These are predominantly the widely studied model species, such as mice, fruitflies and nematode worms, or species of particular interest to humans, such as the malaria-carrying mosquito.

It may come as a surprise, therefore, to find that the list now includes not one, but two species of Tetraodontiformes, a relatively obscure group of fish also known as puffers. Following on from the publication two years ago of the genome sequence of the Japanese pufferfish *Takifugu rubripes*<sup>1</sup>, Jaillon and colleagues<sup>2</sup> report, on page 946 of this issue, the near-complete sequence of the spotted green pufferfish *Tetraodon nigroviridis*. *Takifugu* is a poisonous marine fish best known to connoisseurs of sushi restaurants, whereas *Tetraodon* is a small, brackish-water pufferfish commonly kept in aquaria. But, like all Tetraodontiformes, the

two species share a feature of great convenience for genomics: their cells possess less DNA than those of any other group of back-boned animals — about eight or nine times less than human cells.

Although the *Tetraodon* genome is small compared with that of other vertebrates, sequencing it was still a hugely formidable task. The research reported in this issue<sup>2</sup> was performed in a collaboration between Genoscope in France and the Broad Institute of the Massachusetts Institute of Technology and Harvard University in the United States. Together they have generated a genome sequence of impressive accuracy and coverage, with 64% of the DNA sequence mapped to specific chromosomes<sup>3</sup>.

By comparing the *Tetraodon* genome sequence with that of humans, Jaillon *et al.* even allow us to peer into the genome of the last common ancestor of pufferfish and humans — a primitive bony fish that lived hundreds of millions of years ago. The descendants of this long-extinct ancestor

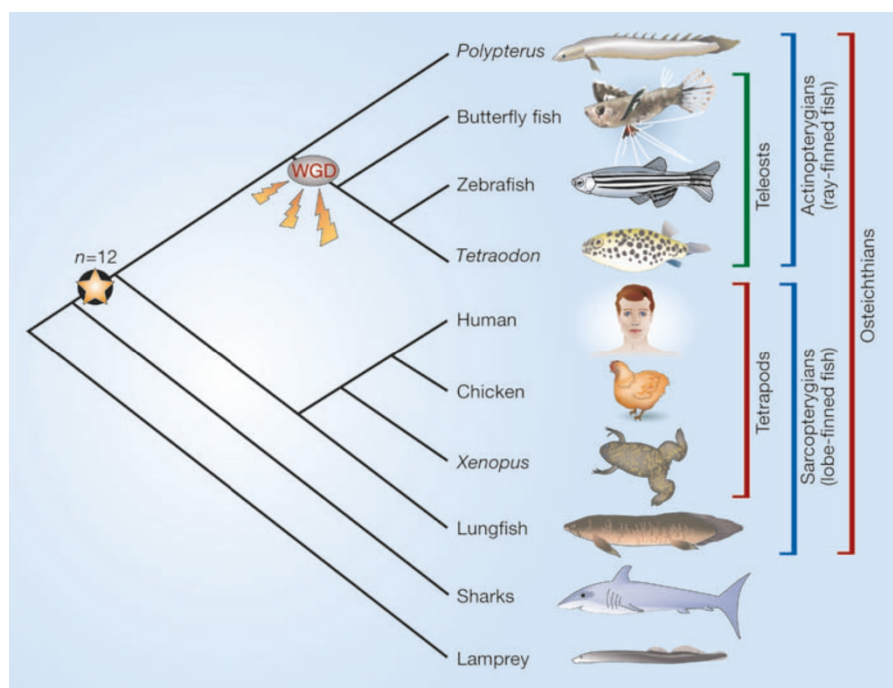


Figure 1 Evolutionary relationships between humans, pufferfish and other vertebrates. Jaillon *et al.*<sup>2</sup> have sequenced the genome of the pufferfish *Tetraodon nigroviridis*. By comparing this genome sequence with that of humans, the authors deduce that the extinct ancestor of actinopterygians (ray-finned fish, including pufferfish) and sarcopterygians (lobe-finned fish, the lineage that gave rise to humans) had 12 pairs of chromosomes ( $n=12$ ). They also show that a whole-genome duplication (WGD) occurred during the evolution of ray-finned fish.

Pufferfish and zebrafish belong to distinct taxonomic orders of fish, so the duplication must have occurred early in teleost evolution. As previously pointed out<sup>7</sup>, this implies that traces of the ancient whole-genome duplication should be found in more than 20,000 species of living teleost fish. But teleosts do not make up the whole of the ray-finned fish. Significantly, a study of one of the Hox-gene clusters of an earlier (more 'basally') branching ray-finned fish, *Polypterus* (Fig. 1), found no evidence of a genome duplication<sup>9</sup>. Together with data from other basal actinopterygians<sup>10</sup>, this suggests that the genome duplication occurred close to the origin of the teleost fish themselves, perhaps 230 million years ago.

Less clear are the biological consequences. It is tempting to suggest that the species richness of the teleosts is somehow related to the whole-genome duplication, either because natural selection has 'exploited' the extra genes, or because differential mutation of duplicate genes caused reproductive isolation, facilitating speciation<sup>11</sup>. However, much of teleost diversity is found in just one group, the acanthopterygians ('spiny fins'), which underwent a massive increase in diversity only around 55 million years ago. So if the whole-genome duplication did affect species richness, it was not immediate, and further studies of morphological and genetic evolution in teleosts will be needed to resolve the mechanisms involved.

A final lesson from the *Tetraodon* study<sup>2</sup> concerns the power of comparative genomics, both for gaining insights into mechanisms of genome evolution and for deducing genome organization in extinct species. But we have a long lineage of extinct ancestors, which means that a wide range of genomes will need to be compared if we want to look at each node in our evolutionary history (Fig. 1). Particularly useful will be complete genome sequences for a shark, a lamprey and amphioxus, as each will provide insight into yet more ancient ancestral states. We may not have long to wait: this year the Joint Genome Institute in California began sequencing the amphioxus genome, while the National Human Genome Research Institute has announced plans to sequence that of the sea lamprey.

John Mulley and Peter Holland are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. e-mail: peter.holland@zoology.oxford.ac.uk

1. Aparicio, S. *et al.* *Science* **297**, 1301–1310 (2002).
2. Jaillon, O. *et al.* *Nature* **431**, 946–957 (2004).
3. www.genoscope.cns.fr/externe/tetranew
4. Postlethwait, J. *et al.* *Genome Res.* **10**, 1890–1902 (2000).
5. Naruse, K. *et al.* *Genome Res.* **14**, 820–828 (2004).
6. Amores, A. *et al.* *Science* **282**, 1711–1714 (1998).
7. Taylor, J., Braasch, I., Frickley, T., Meyer, A. & Van de Peer, Y. *Genome Res.* **13**, 382–390 (2003).
8. Robinson-Rechavi, M., Maranchi, O., Escriva, H. & Laudet, V. *Curr. Biol.* **11**, R458–R459 (2001).
9. Chiu, C.-H. *et al.* *Genome Res.* **14**, 11–17 (2004).
10. Hoegg, S., Brinkmann, H., Taylor, J. & Meyer, A. *J. Mol. Evol.* **59**, 190–203 (2004).
11. Lynch, M. & Force, A. G. *Am. Nat.* **156**, 590–605 (2000).



#### 100 YEARS AGO

*The Cultivation of Man.* The author of this book is very much in earnest. He condemns modern civilisation in strong terms for its many vices, especially for its worship of money and the mammonite marriages that result from it, and urges that men should apply to their own species the methods of the breeder of cattle. He recommends polygamy, apparently in all seriousness, and not as a mere counsel of perfection. It would, of course, destroy the family, but to this Mr. Witchell has no objection... Certainly he speaks out fearlessly, and that is no small merit. But it is to be regretted that he did not study his subject more before writing. "Natural selection," he says, "is sometimes operative, chiefly among the poor." Considering that in England nearly fifty per cent. of the population die before the average age of marriage, this is a wonderful understatement. If we bear the facts in mind, we can hardly agree with Mr. Witchell that the business man is "the surviving type," *i.e.* apparently the type that is to survive to the exclusion of others. Business men are not a separate species. There is a continual upward movement of able men from the great underlying social stratum, and from this stratum directly or indirectly our successful men, as we call them, have emerged.

From *Nature* 20 October 1904.

#### 50 YEARS AGO

Anatomist, pathologist, epidemiologist, sanitarian and clinician, and one of the most advanced thinkers in the history of the medical sciences, Giovanni Maria Lancisi was born in Rome three hundred years ago, on October 26, 1654... It was at [Pope] Clement's request that in 1707 he wrote his monumental treatise "De subitaneis mortibus", in which he carefully records the pathological lesions of the brain and heart observed at autopsy, gives the first description of syphilis of the heart and of growths on the valves, and lists hypertrophy and dilatation of the heart as a cause of sudden death. Lancisi's book, "De motu cordis et aneurysmatibus" (1728), is another landmark in the history of heart disease, for it stresses the significance of heredity, syphilis and violent emotions as causes of aneurysm... In the tercentennial year of his birth he is gratefully remembered chiefly for having laid the foundation for a true understanding of the pathology of the heart.

From *Nature* 23 October 1954.