

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

The compact pufferfish genome

ArticleInfo		
ArticleID	:	4308
ArticleDOI	:	10.1186/gb-2002-3-11-reports0059
ArticleCitationID	:	reports0059
ArticleSequenceNumber	:	19
ArticleCategory	:	Paper report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 2002-9-19 Received : 2002-9-19 OnlineDate : 2002-10-29
ArticleCopyright	:	BioMed Central Ltd2002
ArticleGrants	:	

Jean-Nicolas Volff

Summary

Comparative genomics using the pufferfish genome will help with identifying and characterizing human genes

Significance and context

Identifying coding and regulatory sequences is a major challenge of the post-sequencing era of the human genome project. A number of prediction programs have been developed, but the results with different programs are frequently divergent, and complete DNA information is not available for many genes. Comparative genomics - the comparison of genomic sequences from different species - is another promising approach. The very simplified principle is that functional sequences are subject to evolutionary constraints, while 'junk' DNA is generally not. Hence, over evolutionary time, important sequences will be more conserved and become recognizable by comparing the genomes of different species. The ideal vertebrate model organism to be compared with human would have a genome with conserved gene structure but much less repetitive DNA and small intergenic and intronic sequences, giving it with a high proportion of coding sequences and so allowing rapid and low-cost sequencing. This led Sydney Brenner, Greg Elgar and collaborators to propose as a genomic model the tiger pufferfish (*Taki*) *Fugu rubripes*, which is separated from *Homo sapiens* by about 450 million years of evolution. The haploid genome of *Fugu* is about 365 million base pairs, only about one-ninth the size of the human genome. In contrast to its human counterpart, *Fugu*'s genome is compact, with less than 20% made up of repetitive sequences and fully one third occupied by gene loci. In a recent issue of *Science*, Aparicio *et al.* (The *Fugu* Genome Consortium) report the sequencing and initial analysis of the *Fugu* genome.

Key results

The genome of *Fugu rubripes* was sequenced to over 95% coverage using a whole-genome shotgun strategy, and assembled into 12,381 scaffolds longer than 2 kb (745 scaffolds longer than 100 kb). A total of 31,059 gene loci covering about one third of the genome and encoding 33,609 peptides/proteins were predicted from the assembled sequences, and the upper bound for gene number was estimated to be around 38,000. Hence, pufferfish and human have roughly the same number of genes. Importantly, 961 novel human putative genes were predicted on the basis of *Fugu*/human comparisons. The genome of

Fugu contains much less repetitive DNA than the human genome, but surprisingly recent activity was detected for at least 40 different families of transposable elements. The number of introns is roughly the same in *Fugu* and human, but differential gain and loss of introns were observed. In particular, some genes are intronless in *Fugu* but have multiple introns in human, and *vice versa*. Introns are generally shorter in *Fugu*, but 'giant' genes with average coding sequence lengths (1-2 kb) covering a genomic region larger than those for homologs in other organisms were detected too. Conservation of chromosomal segments was observed between the *Fugu* and human genomes but there was also considerable gene scrambling. No evidence was found for frequent recent duplications, but numerous examples of ancient duplication events were detected, which might have been generated either by an ancient fish-specific whole genome duplication or by more local events. Comparative study of the *Fugu* and human predicted proteomes revealed that about 25% of predicted human proteins (8,109) apparently do not have a homolog in pufferfish; homologs of 5% of these were detected in fruit fly, nematode or yeast, suggesting that they have been lost from *Fugu*. Conversely, about 6,000 predicted proteins from *Fugu* did not show any homology to human proteins, highlighting the dynamics of protein evolution in the fish and human lineages after their divergence about 450 million years ago.

Links

The [Fugu genome project](#) homepage provides information about the sequencing project, a site where data can be downloaded and links to annotation tools. The [Fugu genomics project](#) at the Medical Research Council UK human genome mapping project resource centre includes a blast server, as does the [Fugu rubripes site](#) at the Joint Genome Institute. The related pufferfish species [Tetraodon nigroviridis](#) has a homepage at Genoscope in France with a BLASTN server, and there is also a [Tetraodon nigroviridis database](#) hosted by the Whitehead Institute Center for Genome Research. The [Ensembl zebrafish BLAST server](#) is available at the European Bioinformatics Institute.

Reporter's comments

The goal outlined by Brenner, Elgar and colleagues has been reached: the low-cost sequencing (about one hundred times less costly than the human project) of a second vertebrate genome to allow comparative genomics studies with the human genome. This was so inexpensive that the scientific community can even afford the sequencing of the genome of *Fugu*'s cousin, the freshwater pufferfish *Tetraodon nigroviridis*, the genome of which being currently being assembled. The usefulness of vertebrate comparative genomics was demonstrated by the identification of about thousand new genes in the human genome through comparison with *Fugu*. Nevertheless, both *Fugu* and *Tetraodon* will only be genomic models: they are not laboratory animals, there are no genetic maps and no mutants available, and functional analysis is so far not possible in these fish. The larger genome of two fish species used as models for vertebrate development is also being currently sequenced: the zebrafish *Danio rerio*, the genome project of which is well advanced, and the medaka *Oryzias latipes*, object of a first [genome meeting](#) several weeks ago in Heidelberg. Functional comparisons between these different fish genomes

and the genomes of higher vertebrates will shed new light on vertebrate evolution and lead to the identification of the genes that distinguish fish and humans.

Table of links

[*Science*](#)

[*Fugu* genome project](#)

The *Fugu* genomics project

[*Fugu rubripes*](#)

[*Tetraodon nigroviridis*](#)

[*Tetraodon nigroviridis* database](#)

[Ensembl zebrafish BLAST server](#)

[First MGI Symposium on "Medaka Genomics 2002"](#)

References

1. Aparicio A, Chapman J, Stupka E, Putman 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.