



## Less Is More: Compact Genomes Pay Dividends

Misha Angrist

*Genome Res.* 1998 8: 683-685

Access the most recent version at doi:[10.1101/gr.8.7.683](https://doi.org/10.1101/gr.8.7.683)

---

**References** This article cites 27 articles, 5 of which can be accessed free at:  
<http://genome.cshlp.org/content/8/7/683.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# Less Is More: Compact Genomes Pay Dividends

Misha Angrist<sup>1</sup>

Department of Genetics, Case Western Reserve University, Cleveland, Ohio 44106-4955 USA

## Sydney and the Blowfish

In 1993, Sydney Brenner, like many others, recognized that vertebrates are distinct in their morphology and development and that access to the complete sequence of a vertebrate genome would yield valuable insights into the biology of higher species not obtainable from genome studies of yeast, fly, or even the nematode. Moreover, at that time it was not possible, through sequencing technology, to rapidly generate the huge volumes of accurate, inexpensive sequence data necessary for sequencing the human genome to succeed in a cost-effective manner. The good news was that evolution likely dictated that vertebrates would be distinguished from each other at the genomic level more by the ways in which their genes were *regulated* rather than by profound differences in gene *repertoire* (Elgar 1996a). Thus, the goal was to find a vertebrate genome of intermediate size that would be both (1) representative of higher organisms and (2) small enough and gene-dense enough for large-scale genomic studies. Brenner's organism of choice was the pufferfish (a.k.a. blowfish) *Fugu rubripes* (Brenner et al. 1993). The era of compact vertebrate genomics was born.

But why *Fugu*, an exotic Japanese delicacy, known more for its bloated appearance and potent neurotoxin (potentially lethal to diners if not prepared by a knowledgeable chef) than for its utility as a model organism? In a survey of the haploid DNA content of nearly 300 teleostean fishes, Ralph Hinegardner (1968) observed that members of the Tetraodontidae family in particular tend to have very small genomes—on the order of 400 Mb, some 7.5-fold smaller than the 3000-Mb human genome (0.4–0.5

pg/cell vs. 3 pg in the human). Whereas its size suggested that it would be more manageable than a mammalian genome, what was more telling was the discovery that <8% of *Fugu* DNA is repetitive, as compared to ~60% of human DNA. In their review, Koop and Nadeau (1996) summarized transcript quantity data in *Fugu* versus man and concluded that whereas five times more of the pufferfish genome is *transcribed*, its absolute gene *number* appears to be comparable to mammals. How is this possible? Beyond an overall paucity of repetitive DNA in *Fugu*, the pufferfish also exhibits substantial compaction of both introns and intergenic distances. Undoubtedly, such compaction is directly related to the intronic expansion that has taken place in the last ~430 million years, that is, the span of time separating teleosts from mammals, making them our most distant extant vertebrate ancestors (Powers 1991).

In addition to Koop and Nadeau (1996), at least two reviews by Greg Elgar, Sydney Brenner, and colleagues encapsulate the promise of *Fugu* as a model organism and provide a survey of comparative pufferfish genomics efforts through 1996 (Elgar 1996b; Elgar et al. 1996). A cursory survey of Medline from the last 5 years reveals that representatives of >30 distinct gene families have been cloned and genomically characterized in *Fugu*. In addition, comparative mapping studies have now examined physical distance, genetic distance, and/or, at a minimum, conserved synteny relationships between *Fugu* genes and their mammalian orthologs. Although it is difficult to synthesize these data concisely, the general tendency of *Fugu* genes to be significantly smaller than their human counterparts has been borne out. Of genes where both human and *Fugu* genomic sequences are available, to my knowledge only one—the neural cell adhesion molecule *L1*—is ac-

tually larger in *Fugu* than in man (by a factor of 1.1; Coutelle et al. 1998). For most other individual genes, the extent of compaction in pufferfish ranges from 2.5-fold for the tuberous sclerosis gene *TSC2* (Maheshwar et al. 1996) to >30-fold for complement component *C9* and the amyloid precursor protein gene *APP* (Yeo et al. 1997; Villard et al. 1998).

As stated, substantially smaller introns appear to be the rule in pufferfish and, along with reductions in intergenic distances, the primary source of *Fugu*'s reduction in genomic size. Although intron–exon boundaries, exon size, exon number, and even alternative splice site usage have been more or less conserved for >400 million years in a number of genes (e.g., Gardiner 1997; Coutelle et al. 1998; Llevadot et al. 1998; Villard et al. 1998), intron size does not appear to be under the same selective constraints. In some cases, intronic compaction may approach 50-fold (McNaughton et al. 1997; Villard et al. 1998). Similarly, studies of intergenic distances in *Fugu* also prove reduction to be the order of the day. How et al. (1996) and Schofield et al. (1997) have both examined tandemly arrayed genes that span large chromosomal regions in human and found substantial compaction in pufferfish. Aparicio et al. (1997) showed that *Fugu Hox* gene clusters are compact by virtue of a combination of gene loss and reduction in intergenic distances. To date, perhaps the most dramatic example of reduction of intergenic distance comes from Trower et al. (1996), who showed that a three-gene cluster occupied >600 kb on human chromosome 14q24.3 but only 12.4 kb in *Fugu*. Why vertebrate genome expansion has taken place over the course of evolution—the explosion of repeat families in mammals is one obvious reason—is a fascinating subject but, unfortunately, also beyond the scope of this commentary.

<sup>1</sup>E-MAIL [mha6@worldnet.att.net](mailto:mha6@worldnet.att.net); FAX (216) 368-5857.

# Insight/Outlook

## Junk Is Where You Find It

Although genomic structuralists and mutation hunters can use orthologous sequence information from pufferfish to study conservation of synteny and the workings of individual vertebrate genes, Kathleen Gardiner and her colleagues (Gardiner 1997; Villard et al. 1998) have taken a broader view by seeking to interpret *Fugu* genome data in terms of “junk versus function.” Her group’s primary interest is human chromosome 21, and they have examined not only intron compaction and conservation of synteny of specific genes residing on this chromosome, but also base composition and global chromosome organization. In particular, Gardiner (1997) questioned the validity of suggestions that the proximal half of chromosome 21q contains only 10% of the q arm’s genes. Proximal 21q is known to be AT-rich and to harbor few *NotI* sites yet contains >20 of the long arm’s 40 Mb of DNA. To test the gene-poor hypothesis, members of her laboratory carried out cDNA selection and exon trapping using YACs known to map to the region. In addition, they analyzed >100 radiation hybrid-mapped ESTs. All three methods demonstrated the presence of comparatively few genes, with small, relatively AT-rich exons, such as those of *APP*. Sequencing of *Fugu* genes from this region demonstrated 25- to 50-fold intron compaction in each (Gardiner 1997; Villard et al. 1998). The genomic sequence of the pufferfish *APP* gene was found to be >90% smaller than the human version, implying that the overwhelming majority of this DNA, with regard to *APP* function, is dispensable, or “true junk.”

Gardiner (1997, and pers. comm.) was also able to show a preliminary correlation between the extent of compression for particular genes in *Fugu* and their isochore locations. As defined by Bernardi (1993, 1995), isochores in mammalian genomes are long (typically 300 kb or more), compositionally homogeneous DNA segments that can be subdivided into “light” GC-poor (L1 and L2) families and “heavy” GC-rich (H1 and H2) and very GC-rich (H3) families. The L isochores account for some 62% of the genome, whereas the H isochores make up ~35% of the genome. The remaining 3%–4% consists of satellite and ribosomal DNAs, which are also considered isochores because of their homoge-

neous nucleotide composition. Bernardi’s analyses (for review, see Bernardi 1995) of isochores have demonstrated what is intuitively obvious, if not common knowledge, to most molecular biologists, namely that gene density varies greatly with GC level (Zoubak et al. 1997). Regions >50% GC average one gene per 5–10 kb while regions <40% GC may average as little as one gene per 100–200 kb. Intron size has also been shown to be inversely related to GC-content (Duret et al. 1995). The discovery by Gardiner and co-workers (Gardiner 1997, and pers. comm.; Villard et al. 1998) that *APP* and *GABPA*, two large AT-rich genes located in the L isochore on the northern half of 21q, are compressed as much as 50-fold in *Fugu* versus human suggests that *Fugu* orthologs may be especially valuable in the identification and analysis of large, AT-rich mammalian genes residing in L isochores in particular, and AT-rich Giemsa (G) dark bands in general. It will be interesting to determine whether analyses of additional genes on 21q and elsewhere support this trend.

### Another Fish In The Sea?

Unlike *Caenorhabditis elegans* and zebrafish, both of which were touted on the basis of the experimental accessibility to and visibility of their developmental fates, *Fugu* was chosen primarily because of its vertebrate status and relatively small genome. Although its popularity as *haute cuisine* has ensured that some amount of *Fugu* material will always be available, its restricted natural saltwater habitat around the Japanese islands may prove a stumbling block to postgenomics researchers interested in *in vivo* studies of pufferfish.

Recently, another pufferfish has swum onto the scene. The Tetraodontidae family includes both *F. rubripes* and *Tetraodon fluviatilis*. The latter also possesses an extremely compact genome (~380 Mb), slightly smaller than that of *Fugu* (Hinegardner 1968; Crnogorac-Jurcevic et al. 1997). *T. fluviatilis* is a freshwater species that has recently been proposed as an alternative pufferfish model because unlike *Fugu*, it can be bred and maintained in an aquarium, thereby ensuring an adequate supply of fresh material for studying overall pufferfish biology (Crnogorac-Jurcevic et al. 1997). Clearly, ready availability of such

material is appealing; however, it may be difficult for *T. fluviatilis* investigators to overcome *Fugu*’s 4-year head start. Of course, one need not preempt the other. Perhaps the best scenario would be for genome studies of the two pufferfishes—estimated by Crnogorac-Jurcevic et al. (1997) to have diverged some 20–30 million years ago—to overlap and complement each other in much the same way genomics initiatives in mouse and rat have done.

### Honey, I Shrunk the Chicken Genome

It is fair to ask whether teleosts represent an aberration or whether other vertebrates might also exhibit increased gene density. Although the chicken genome’s threefold reduction in size compared to the human is not as dramatic as the pufferfish’s eightfold compression, recent work by Adrian Bird and colleagues (McQueen et al. 1998) at the University of Edinburgh has revealed that for at least a subset of the chicken genome, gene density may rival that of its teleost forebears. Chickens, like other avians, possess karyotypes comprised of macrochromosomes (MACs) and microchromosomes (MICs). Of the 39 haploid chromosomes, six are MACs that harbor some two-thirds of the genome, whereas the remaining DNA is found on the 33 MICs. MICs 11–39 are particularly small and cannot be distinguished individually by cytology (Bloom et al. 1993).

Two years ago, Bird’s group showed that MICs were not only GC-rich in general, but were particularly enriched for CpG islands, strong indicators of the presence of functional genes. Conversely, MACs were shown to harbor relatively few CpG islands (McQueen et al. 1996). Nevertheless, mapping has placed far more genes on MACs (Burt et al. 1995). Because the presence of CpG islands is not perfectly correlated with the presence of genes, McQueen et al. (1998) sought to test the hypothesis that MICs are indeed gene-rich. To do this, they examined the distribution of acetylated histone H4 throughout the chicken genome. Increased acetylation of the amino terminus of histone H4 is associated with transcriptional activation, whereas hypoacetylation is known to be associated with heterochromatin (Brownell et al. 1996). McQueen et al. (1998) also investigated the timing of

MIC replication during S phase, because early replication is correlated with transcriptional activation (Holmquist 1987). Finally, they simply counted the number of CpG island-like sequences whose genomic origin was known.

All three approaches confirmed investigators' suspicions that MICs were highly gene-rich and that previous mapping of chicken genes was biased toward MACs. Chicken MICs are selectively enriched for acetylated histone H4 and do tend to replicate during the first half of S phase, both indicative of increased transcriptional activity. Moreover, McQueen et al. (1998) showed that CpG island-like fragments are enriched on MICs and that these fragments are associated with genes. So why the ascertainment bias toward genes resident on MACs? These workers believe it to be caused by difficulties in physical mapping of tiny, cytologically indistinguishable MICs and by the fact that the chicken genetic map is based largely on (CA)<sub>n</sub> microsatellites, polymorphic markers found at a much higher frequency on MACs (Primmer et al. 1997).

The total number of CpG islands from known regions available to be counted was not huge (22 islands on 11 MICs vs. 3 islands on 9 MACs). However, by assuming a reasonable 60% of genes to be associated with CpG islands, McQueen et al. (1998) calculate that if the sixfold enrichment they observed for CpG islands on MICs holds up, one should expect to find about one gene per 10 kb of chicken microchromosomal DNA, a figure comparable to the genome-wide estimate for *Fugu* (400 Mb of *Fugu* DNA/60,000 genes = one gene every 7 kb). The implication of the work of the Bird group is that the genome community may now have access to ~40,000 avian genes that, like their pufferfish progenitors, appear to take up much less space than their mammalian counterparts. Indeed, less is becoming more all the time.

#### ACKNOWLEDGMENTS

I am grateful to Kathleen Gardiner for sharing data prior to publication, to Evan Eichler for helpful discussions and comments, and to Aravinda Chakravarti for continued support.

#### REFERENCES

Aparicio, S., K. Hawker, A. Cottage, Y. Mi-

kawa, L. Zuo, B. Venkatesh, E. Chen, R. Krumlauf, and S. Brenner. 1997. *Nature Genet.* 16: 79–83.

Bernardi, G. 1993. *Mol. Biol. Evol.* 10: 186–204.

———. 1995. *Annu. Rev. Genet.* 29: 445–476.

Bloom, S.E., M.E. Delany, and D.E. Muscarella. 1993. *Manipulation of the avian genome* (ed. R.J. Etches and A.M.V. Gibbons), pp. 39–59. CRC Press, Boca Raton, FL.

Brenner, S., G. Elgar, R. Sandford, A. Macrae, B. Venkatesh, and S. Aparicio. 1993. *Nature* 366: 265–268.

Brownell, J.E., J. Zhou, T. Ranalli, R. Kobayashi, D.G. Edmondson, S.Y. Roth, and C.D. Allis. 1996. *Cell* 84: 843–851.

Burt, D.W., N. Bumstead, J.J. Bitgood, F.A. Ponce De Leon, and L.B. Crittenden. 1995. *Trends Genet.* 11: 190–194.

Coutelle, O., G. Nyakatura, S. Taudien, G. Elgar, S. Brenner, M. Platzer, B. Drescher, M. Jouet, S. Kenwick, and A. Rosenthal. 1998. *Gene* 208: 7–15.

Crnogorac-Jurcevic, T., J.R. Brown, H. Lehrach, and L.C. Schalkwyk. 1997. *Genomics* 41: 177–184.

Duret, L., D. Mouchiroud, and C. Gautier. 1995. *J. Mol. Evol.* 40: 308–317.

Elgar, G. 1996a. Fishing for genes: *Fugu* expands (but its genome doesn't). <http://Fugu.hgmp.mrc.ac.uk/Fugu/Fugu.html>.

———. 1996b. *Hum. Mol. Genet.* 5: 1437–1442.

Elgar, G., R. Sandford, S. Aparicio, A. MacRae, B. Venkatesh, and S. Brenner. 1996. *Trends Genet.* 12: 145–150.

Gardiner, K. 1997. *Gene* 205: 39–46.

Hinegardner, R. 1968. *Am. Nat.* 102: 517–523.

Holmquist, G.P. 1987. *Am. J. Hum. Genet.* 40: 151–173.

How, G.-F., B. Venkatesh, and S. Brenner. 1996. *Genome Res.* 6: 1185–1191.

Koop, B.F. and J.H. Nadeau. 1996. *Proc. Natl. Acad. Sci.* 93: 1363–1365.

Llavadot, R., X. Estivill, P. Scambler, and M. Pritchard. 1998. *Gene* 208: 279–283.

Maheshwar, M.M., R. Sandford, M. Nellist,

J.P. Cheadle, B. Sgotto, M. Vaudin, and J.R. Sampson. 1996. *Hum. Mol. Genet.* 5: 131–137.

McNaughton, J.C., G. Hughes, W.A. Jones, P.A. Stockwell, H.J. Klamut, and G.B. Petersen. 1997. *Genomics* 40: 294–304.

McQueen, H.A., J. Fantes, S.H. Cross, V.H. Clark, A.L. Archibald, and A.P. Bird. 1996. *Nature Genet.* 12: 321–324.

McQueen, H., G. Siriaco, and A.P. Bird. 1998. *Genome Res.* 8: 621–630.

Powers, D.A. 1991. *Adv. Genet.* 29: 119–228.

Primmer, C.R., T. Raudsepp, B.P. Chowdhary, A.P. Moller, and H. Ellegren. 1997. *Genome Res.* 7: 471–482.

Schofield, J.P., G. Elgar, J. Greystrom, G. Lye, R. Deadman, G. Micklem, A. King, S. Brenner, and M. Vaudin. 1997. *Genomics* 45: 158–167.

Trower, M.K., S.M. Orton, I.J. Purvis, P. Sanseau, J. Riley, C. Christodoulou, D. Burt, C.G. See, G. Elgar, R. Sherrington et al. 1996. *Proc. Natl. Acad. Sci.* 93: 1366–1369.

Villard, L., F. Tassone, T. Crnogorac-Jurcevic, K. Clancy, and K. Gardiner. 1998. *Gene* 210: 17–24.

Yeo, G.S., G. Elgar, R. Sandford, and S. Brenner. 1997. *Gene* 200: 203–211.

Zoubak, S., O. Clay, and G. Bernardi. 1997. *Gene* 174: 95–102.