

# Genome size is negatively correlated with effective population size in ray-finned fish

Soojin Yi and J. Todd Strelman

School of Biology, Georgia Institute of Technology, 310 Ferst Dr., Atlanta, GA 30332, USA

Corresponding author: Yi, Soojin (soojin.yi@biology.gatech.edu).

## Materials and Methods

### Genome Size

The haploid genome size (C-value, in pg) of 1461 ray-finned fish species (Actinopterygii) was downloaded from the Animal Genome Size Database ([www.genomesize.com](http://www.genomesize.com)). Unless otherwise stated, comparisons of genome sizes are performed on the C-values, which can be converted to number of base pairs by the following: number of base pairs = C-value  $\times$   $0.978 \times 10^9$ . Several species are represented more than once in this database, often with different estimates of genome size owing to differences in the methods of estimation and the cell types used. We collapsed such redundant cases to single data points by either taking the average (when there are two entries per species) or taking the median (if more than three entries observed for each species). Clear case of recent polyploidy (induced by hybridization or experimental manipulation) has been removed from our data set.

### Habitat assignment

We separated fishes into two categories based on the habitat in which they breed (freshwater, FW or marine, M). This information was gathered from Nelson [1] and/or from FishBase ([www.fishbase.org](http://www.fishbase.org)). In most cases, this distinction is straightforward. Anadromous fishes such as salmonids (live in oceans but breed in rivers) were assigned to the FW category, whereas catadromous fishes such as the American eel (which live in rivers but breed in the Sargasso Sea) were assigned to the M category. We removed 26 cases where the assignment was ambiguous.

### Microsatellite heterozygosity

Data Collection: Data on microsatellite heterozygosity were collected from the literature (see Supplemental Table 1). Frequencies of marker types (di-, tri- and tetra- nucleotides) did not differ between marine or freshwater groupings ( $P > 0.09$ ). We excluded marker loci if the cloned allele was (i) monomorphic in sampled individuals, (ii) short (less than six repeat units) or (iii) complex and interrupted. We confined analysis to loci cloned from focal species to avoid ascertainment bias. In cases of known ancient polyploidy (e.g. sturgeon, paddlefish), only disomic loci were included.

Population sampling: In the majority of examples ( $n = 30$ ), we report mean heterozygosity averaged over multiple loci from a single population. In *Anguilla anguilla* we used the average intra-population heterozygosity value (six loci) averaged over 11 populations, as given in the literature. For *Danio rerio*, we used the average intra-population heterozygosity value (six loci) averaged over four populations (SD = 0.036). In the case of *Thymallus thymallus*, we employed the grand mean of 17 loci from 15 populations (SD = 0.079).

Transformation: According to the stepwise mutation model [2], expected microsatellite heterozygosity ( $H_e$ ) at equilibrium is equal to  $1 - (1 + 8N_e u)^{-0.5}$ . We take a transformed value of  $H_T = (1 - H_e)^{-1}$ , which is linearly related to  $N_e$  on a log-log scale for the range of  $H_e$  in our data set.

Phylogenetically independent contrasts: We used the 'Crunch' algorithm of CAIC v. 2.0.0 [3] without branch length information, to compute phylogenetically independent contrasts of transformed data,

using a phylogeny of taxa inferred from published molecular analyses [4,5]. The inferred phylogeny is presented in the Supplementary Figure.

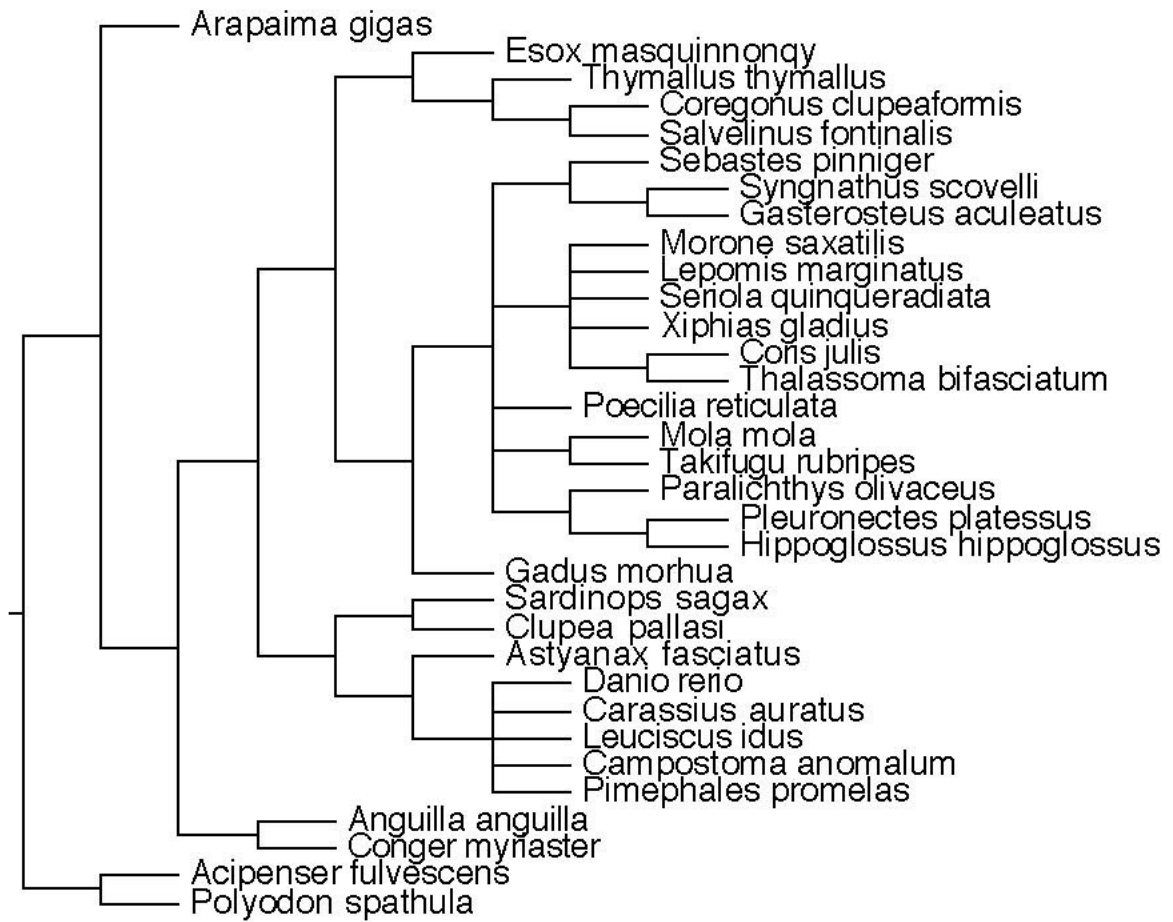
#### *Multiple regression analyses*

To address the relationship between genome size and microsatellite heterozygosity, we further took into account the effects of body size and generation time. We recorded estimates of maximum body size (in mm) and generation time (in years), for each of the 33 species in Figure 2 in the main text, from FishBase ([www.fishbase.org](http://www.fishbase.org)). All values are transformed by taking natural logarithms, to be approximately normally distributed before calculating phylogenetically independent contrasts. Qualitative conclusions are similar if the variables are not transformed or are always transformed (results not shown).

#### **References**

- 1 Nelson, J.S. 1984. *Fishes of the World*, 2<sup>nd</sup> edition. John Wiley and Sons
- 2 Ohta, T. and Kimura, M. (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* 22, 201–204
- 3 Purvis, A. and Rambaut, A. (1995) Comparative analysis by independent contrasts (CAIC): an Apple Macintosh application for analyzing comparative data. *Computer Appl. Biosciences* 11, 247–251
- 4 Inoue, J. G. *et al.* (2003) Basal actinopterygian relationships: a mitogenomic perspective on the phylogeny of the ‘ancient fish’. *Mol. Phylogen. Evol.* 26, 110–120
- 5 Miya, M. *et al.* (2003) Major patterns of higher teleost phylogenies: a new perspective based on 100 complete mitochondrial DNA sequences. *Mol. Phylogen. Evol.* 26, 121–138

Supplementary Figure 1



Species	Family/Order	He	Genome size	Max Size (mm)	Generation time (years)	Reference (He)
<i>Marlin</i>						
<i>Sardinops sagax</i>	Clupeidae/Clupeiformes	0.88	1.35	395	1.8	Pereyra et al. 2004. <i>Molecular Ecology Notes</i> 4: 739-741.
<i>Conger myriaster</i>	Congridae/Anguiliformes	0.76	1.22	1000	5.1	Kimura et al. 2003. <i>Molecular Ecology Notes</i> 3: 203-204.
<i>Cottus julis</i>	Labridae/Perciformes	0.6766	1.21	300	7	Chuttemund et al. 2000. <i>Molecular Ecology</i> 9: 631-632.
<i>Anguilla anguilla</i>	Anguillidae/Anguilliformes	0.82	1.6	1330	18.3	Mank and Avise 2003. <i>J. Heredity</i> 94: 310-314.
<i>Thalassoma bifasciatum</i>	Labridae/Perciformes	0.917	0.98	250	2.2	Williams et al. 2004. <i>Molecular Ecology Notes</i> 4: 525-527.
<i>Sebastes pinniger</i>	Scorpaenidae/Scorpaeniformes	0.788	0.97	760	6.8	Gomez-Uchida et al. 2003. <i>Molecular Ecology Notes</i> 3: 387-389.
<i>Xiphias gladius</i>	Xyphidae/Perciformes	0.736	0.88	4550	6.7	Reeb et al. 2003. <i>Molecular Ecology Notes</i> 3: 147-149.
<i>Mola mola</i>	Molidae/Tetraodontiformes	0.84	0.905	3330	4	Streehman et al. 2003. <i>Molecular Ecology Notes</i> 3: 247-249.
<i>Seriola quinqueradiata</i>	Carangidae/Perciformes	0.89	0.83	1500	2.2	Okata et al. 2003. <i>Molecular Ecology Notes</i> 3: 390-391.
<i>Clupea palasi</i>	Clupeidae/Clupeiformes	0.822	0.77	460	1.9	Olsen et al. 2002. <i>Molecular Ecology Notes</i> 2: 101-103.
<i>Hippoglossus hippoglossus</i>	Pleuronectidae/Clupeiformes	0.84	0.73	2400	36.1	McGowan and Reith 1999. <i>Molecular Ecology</i> 8: 1761-1763.
<i>Paralichthys olivaceus</i>	Bothidae/Pleuronectiformes	0.814	0.71	1030	6.4	Kim et al. 2003. <i>Molecular Ecology Notes</i> 3: 491-493.
<i>Syngnathus scovelli</i>	Syngnathidae/Gasterosteiformes	0.919	0.54	183	1.7	Jones and Avise 1997. <i>Molecular Ecology</i> 6: 203-213.
<i>Takifugu rubripes</i>	Tetraodontidae/Tetraodontiformes	0.916	0.4	700	5.9	Takagi et al. 2003. <i>Fisheries Science</i> 69: 1085-1095.
<i>Gadus morhua</i>	Gadidae/Gadiformes	0.896	0.665	2000	6	Bentzen et al. 1996. <i>Can. J. Fish. Aquat. Sci.</i> 53: 2706-2721.
<i>Pleuronectes platessius</i>	Pleuronectidae/Pleuronectiformes B	0.88	0.39	1000	14.8	Hoarwa et al. 2002. <i>Molecular Ecology Notes</i> 2: 60-61.
<i>Freshwater</i>						
<i>Acipenser fulvescens</i>	Acipenseridae/Acipenseriformes	0.51	4.45	2740	24	Welsh et al. 2003. <i>Molecular Ecology Notes</i> 3: 47-55.
<i>Coregonus clupeaformis</i>	Salmonidae/Salmoniformes A	0.498	2.94	1000	2.3	Li et al. 2001. <i>Molecular Ecology</i> 10: 965-985.
<i>Salvelinus fontinalis</i>	Salmonidae/Salmoniformes B	0.608	3.24	860	2.9	Fraser et al. 2004. <i>Molecular Ecology</i> 13: 67-80.
<i>Polyodon spathula</i>	Polyodontidae/Actinopterygiformes	0.72	1.95	2210	17	Heist et al. 2002. <i>Conservation Genetics</i> 3: 205-207.
<i>Thymallus thymallus</i>	Salmonidae/Salmoniformes C	0.4466	2.15	600	8	Koskinen et al. 2002. <i>Heredity</i> 88: 391-401.
<i>Danio rerio</i>	Cyprinidae/Cypriniformes \$	0.77	1.78	600	0.3	Gratton et al. 2004. <i>J. Syst. Evol. Research</i> 42: 54-62.
<i>Carassius auratus</i> #	Cyprinidae/Cypriniformes	0.72	1.79	590	4.9	Yue and Orban 2004. <i>Molecular Ecology Notes</i> 4: 404-405.
<i>Astronotus fasciatus</i>	Characidae/Characiformes	0.8	1.565	100	0.6	Streeker 2003. <i>Molecular Ecology Notes</i> 3: 150-151.
<i>Leuciscus idus</i>	Cyprinidae/Cypriniformes	0.66	1.4	760	26	Barilova et al. 2004. <i>Molecular Ecology Notes</i> 4: 86-88.
<i>Esox masquinomy</i>	Esocidae/Salmoniformes	0.607	1.28	1830	13.9	Reading et al. 2003. <i>Molecular Ecology Notes</i> 3: 447-449.
<i>Campostoma anomalum</i>	Cyprinidae/Cypriniformes	0.746	1.15	220	2.6	Dinoski et al. 2000. <i>Molecular Ecology</i> 9: 2187-2189.
<i>Pimephales promelas</i>	Cyprinidae/Cypriniformes	0.754	1.1	100	1.4	Beasert and Ort 2003. <i>Molecular Ecology Notes</i> 3: 532-534.
<i>Lepomis microchilus</i> ^	Centrarchidae/Cypriniformes	0.648	1.08	120	2	Schable et al. 2002. <i>Molecular Ecology Notes</i> 2: 509-511.
<i>Poecilia reticulata</i>	Cyprinodontidae/Cyprinodontiformes	0.73	0.885	350	0.2	Becher et al. 2002. <i>Molecular Ecology Notes</i> 2: 456-458.
<i>Arpaima rigus</i>	Osteoglossidae/Osteoglossiformes	0.647	0.98	4500	12.2	Farris et al. 2003. <i>Molecular Ecology Notes</i> 3: 128-130.
<i>Morone saxatilis</i>	Percichthyidae/Perciformes	0.623	0.92	2000	14.4	Brown et al. 2003. <i>Molecular Ecology Notes</i> 3: 414-416.
<i>Gasterosteus aculeatus</i> **	Gasterosteidae/Gasterosteiformes	0.6344	0.64	110	0.7	Largiadier et al. 1999. <i>Molecular Ecology</i> 8: 342-344.
** sampled from FW population (He)						
# sampled from pet store (He)						
^ observed heterozygosity						

\$ The Cyprinidae is one of the largest families of fishes (~190 genera, 2100 species). At various times, these species have been included in distinct sub-families.