J Mol Evol (2002) 55:356–363
DOI: 10.1007/s00239-002-2333-y

JOURNAL OF MOLECULAR EVOLUTION

Analysis of the Phylogenetic Distribution of Isochores in Vertebrates and a Test of the Thermal Stability Hypothesis

Elise M.S. Belle,[1] Nick Smith,[2] Adam Eyre-Walker[1]

[1] Centre for the Study of Evolution & School of Biological Sciences, University of Sussex, BN1 9QG Brighton, United Kingdom
[2] Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden

Received: 19 March 2002 / Accepted: 25 March 2002

Abstract. Warm-blooded vertebrates show large-scale variation in G + C content along their chromosomes, a pattern which appears to be largely absent from cold-blooded vertebrates. However, compositional variation in poikilotherms has generally been studied by ultracentrifugation rather than sequence analysis. In this paper, we investigate the compositional properties of coding sequences from a broad range of vertebrate poikilotherms using DNA sequence analysis. We find that on average poikilotherms have lower third-codon position GC contents (GC3) than homeotherms but that some poikilotherms have higher mean GC3 values. We find that most poikilotherms have lower variation in GC3 than homeotherms but that there is a correlation between GC12 and GC3 for some species, indicating that there is systematic variation in base composition across their genomes. We also demonstrate that the GC3 of genes in the zebrafish, *Danio rerio*, is correlated with that in humans, suggesting that vertebrates share a basic isochore structure. However, we find no correlation between either the mean GC3 or the standard deviation in GC3 and body temperature.

Key words: Genome evolution — Vertebrates — Natural selection — Isochores — Thermal stability — Compositional patterns

Correspondence to: Elise M.S. Belle; *email:* e.m.s.belle@ sussex.ac.uk

Introduction

It has been more than 30 years since Giorgio Bernardi and colleagues first showed that the G + C content varies substantially across the mammalian genome, over a scale of hundreds of kilobases, with some regions of the genome having a G + C content as low as 35% while others are at 60% (Bernardi 2001). Subsequent work showed that birds also had marked large-scale heterogeneity in composition (Bernardi et al. 1985; Kadi et al. 1993) but that most cold-blooded vertebrates showed very limited compositional heterogeneity (Bernardi et al. 1990). This phylogenetic distribution led to the suggestion that G + C-rich isochores evolved as an adaptation to homeothermy.

The vast majority of our knowledge about the phylogenetic distribution of compositional variation comes from studies in which genomic DNA is randomly sheared into large blocks (~100 kb) and then separated according to G + C content on a CsCl gradient by centrifugation. Subsequent analysis allows one to determine how much of the genome belongs to a particular G + C content class. This technique has been applied to literally hundreds of organisms. With very few exceptions, warm-blooded species have high levels of compositional heterogeneity, while cold-blooded species have low levels.

However, these centrifugation studies may mask some compositional variation. Crocodile DNA shows fairly low levels of compositional heterogeneity in centrifugation studies, yet Hughes et al. (1999)

showed that the GC3 of crocodile and turtle genes was highly correlated with the GC3 values of homologous chicken genes. Since the GC3 of a gene in chicken is strongly correlated with the G + C content of the genomic region in which the gene is situated, and chicken DNA shows marked compositional heterogeneity, the results of Hughes et al. imply that crocodiles have some significant level of compositional heterogeneity. Why this compositional heterogeneity is not apparent in centrifugation studies has not been resolved.

We therefore decided to investigate in further detail whether cold-blooded vertebrates show compositional heterogeneity by examining the pattern of compositional heterogeneity in coding sequences. It seems to be generally true in vertebrates that the GC3 content of a gene is correlated with the genomic region in which it is situated (Bernardi 2000), therefore large amounts of variation in GC3 should indicate marked compositional heterogeneity across the genome. However, selection on synonymous codon use could be a confounding factor. For example, certain codons are favored by translational selection in *Drosophila melanogaster*; the majority of these are either G or C ending (Shields et al. 1988). Hence, variation in the strength of translational selection among genes generates variation in GC3 which is largely uncorrelated with the G+C content of the genomic region in which the gene resides (Kliman and Hey 1994). Selection on synonymous codon use has never been conclusively demonstrated in any vertebrate species (but see DeBry and Marzluff 1994; Musto et al. 2001), however, the potential for a confounding influence led us to test for compositional heterogeneity using two further analyses. In the first we tested for a correlation between GC3 and GC12 (the G + C content of the first two codon positions). This correlation has been observed in all species which show significant compositional variation (D'Onofrio et al. 1991; Kadi et al. 1993). The correlation indicates that there is systematic variation in the pattern of base substitution across the genome, variation which is expected to generate general compositional heterogeneity, i.e., compositional variation outside coding regions. In the second test we investigated whether there was a correlation between the GC3 values of a fish species and the GC3 values of human homologues. If this correlation can be demonstrated, then it suggests that vertebrates could share some basic pattern of genome organization. Finally, we investigated whether there is a correlation between base composition and body temperature.

Bernardi and Bernardi (1991) have previously considered compositional variation in poikilotherms; they showed that genes from cold-blooded vertebrates had lower GC3 values than homeotherms, that GC3 values were correlated with the G + C contents of the genomes which contained the genes, and that GC12 was correlated with GC3. However, they had relatively few data so they treated all fish genes together and all amphibian and reptile genes together; in doing so they confounded variation within a species and variation between species. For example, for fish genes they found a correlation between GC12 and GC3, but this could have been due to differences in the mean G + C content between different species of fish, rather than a correlation between GC12 and GC3 within a fish genome.

Materials and Methods

We extracted coding sequences from GenBank for a broad range of cold-blooded species for which we could find at least 10 genes. Our sample includes 18 species and taxa: *Alligator mississipiensis*, *Trachemys scripta*, *Ambystoma mexicanum*, *Pleurodeles waltlii*, *Xenopus laevis*, *Rana catesbaiana*, *Anguilla anguilla*, *Salmo gardnairi*, *Fundulus heteroclitus*, *Takifugu rubripes*, the *Nototheniidae*, *Cyprinus carpio*, *Carassius auratus*, *Danio rerio*, *Raja erinacea*, *Squalus acanthias*, *Heterodontus franscisi*, and *Petromyzon marinus*. As representatives of warm-blooded vertebrates we extracted 13,348 human genes and 529 chicken genes. We also extracted the available genes from armadillo (*Chaerophractus villosus*), since it has an unusually low body temperature for a mammal.

The extracted sequences were obtained from the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov) or from the ACNUC database (http://www.pbil.univ-lyon1.fr/search/query.html). The species were chosen either because they were distantly related or because they were well represented in the databases. To eliminate paralogue genes from our analyses, we used the Clean Up program (http://bigarea.area.ba.cnr.it:8000/EmbIT/coda clean.html) or removed them by hand.

Finally, we searched for the human homologues of 100 genes of *Danio rerio* using Unigene entries from the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov/UniGene). We then investigated the correlation between the GC3 of these genes and their human homologues. We also checked that this correlation was not due to the amino acid composition by investigating the correlation for each amino acid separately.

The temperature data for cold-blooded species used to investigate the thermal stability hypothesis were obtained from the Internet (details available on request). For mammals, birds, and reptiles we used the actual body temperature; for amphibians and fish we took their body temperature to be that of the environment in which they live. Since poikilotherms can survive in a range of temperatures we searched for the minimum and maximum temperatures at which they live or, if this information could not be found, the temperature range in which they can be reared.

Results

Mean GC3

On average cold-blooded vertebrates have lower GC3 values than either of the two warm-blooded vertebrates for which we have compiled data (*t* test, $p < 0.0001$). The mean GC3 is also significantly different between all classes, with the order from the

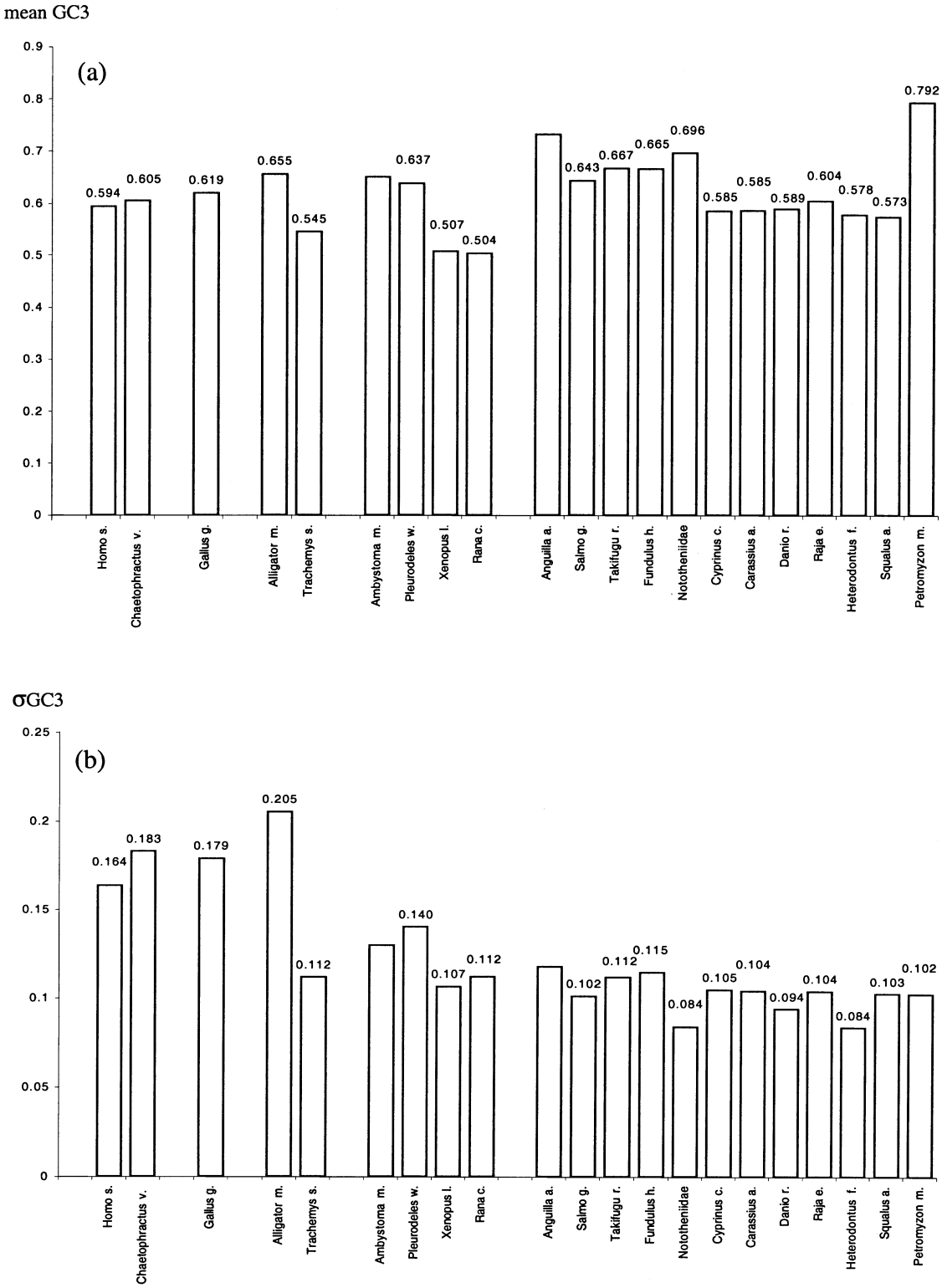

Fig. 1. (**a**) Mean GC3 level and (**b**) GC3 standard deviation in different vertebrate species and taxa.

lowest to the highest value: fish, amphibians, reptiles, mammals, and birds (Fig. 1, Tables 1 and 2). However, this picture is a little deceptive since there is great variability among species in the same class. Furthermore, some cold-blooded species, especially fish, show mean GC3 values which are significantly higher than those of either human or chicken. This remains true even when we control for multiple tests; for example, the mean GC3 of the puffer fish *Takifugu rubripes* (0.6665) is significantly higher than

Table 1. Mean GC3 levels of genes of different taxa and possible relationships between their GC12 and their GC3 level (with and without the three amino acids leucine, arginine, and serine)

Species/taxa	No. of genes	Mean GC3	σ GC3	r CG12 vs r GC3 (−Leu/Arg/Ser)
Homo sapiens	13,348	0.5942	0.1637	0.555***
Chaetophractus villosus (armadillo)	11	0.6047	0.1830	0.148
Gallus gallus (chicken)	529	0.6187	0.1790	0.548***
Alligator mississipiensis (American alligator)	17	0.6553	0.2051	0.555* (0.141)
Trachemys scripta (red-eared slider turtle)	16	0.5445	0.1120	0.816*** (0.769***)
Ambystoma mexicanum (axolotl)	35	0.6500	0.1300	0.291
Pleurodeles waltlii (Iberian ribbed newt)	27	0.6373	0.1403	0.550** (0.478*)
Xenopus laevis (African clawed frog)	1,872	0.5073	0.1067	0.155*** (0.071**)
Rana catesbaiana (bullfrog)	80	0.5040	0.1122	0.118
Anguilla anguilla (European eel)	28	0.7313	0.1182	0.290
Salmo gairdneri (trout)	415	0.6432	0.1015	0.063
Takifugu rubripes (fugu)	102	0.6665	0.1120	0.100
Fundulus heteroclitus (killifish)	35	0.6647	0.1148	0.084
Nototheniidae (Antarctic cod)	25	0.6959	0.0839	−0.510*,a
Cyprinus carpio (common carp)	210	0.5852	0.1049	0.008
Carassius auratus (goldfish)	172	0.5854	0.1042	0.200** (0.077)
Danic rerio (zebrafish)	980	0.5891	0.0940	0.392*** (0.303***)
Raja erinacea (little skate)	28	0.6044	0.1040	0.477
Heterodontus franscisi (horn shark)	63	0.5779	0.0836	0.253* (0.285*)
Squalus acanthias (spiny dogfish)	30	0.5734	0.1027	0.308
Petromyzon marinus (lamprey)	32	0.7920	0.1024	0.170

***$P \leq 0.0001$; **$P \leq 0.01$; *$P \leq 0.05$.

[a] Negative correlation only due to one gene coding for an antifreeze glycoprotein with an abnormally high GC12 value.

Table 2. Mean GC3 levels and standard deviations of the genes in cold-blooded and warm-blooded species and in the five vertebrate classes studied

	Mean GC3	Weighted σ GC3[a]	Number of genes
Warm-blooded	0.5951	0.1635	13,887
Cold-blooded	0.5674	0.1032	3,729
Human	0.5942	0.1637	13,348
Chicken	0.6187	0.1789	529
Reptiles	0.5751	0.1592	33
Amphibians	0.5058	0.1051	1,543
Fish	0.6128	0.1011	2,154

[a] $V_w = (\sum(n_i - 1)V_i)/(\sum(n_i - 1))$.

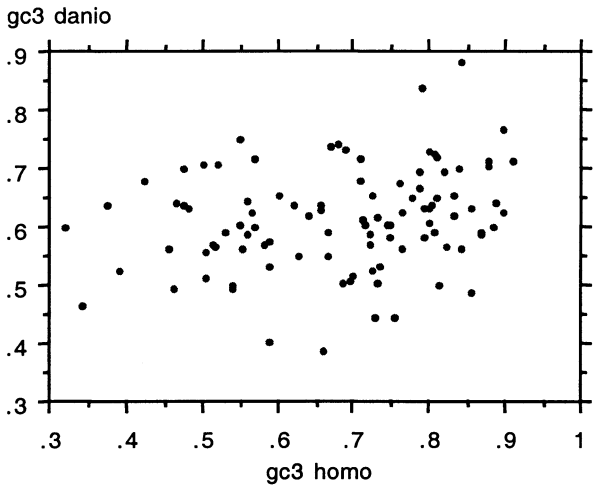

Fig. 2. GC3 values of *Danio rerio* genes versus their homologues *Homo sapiens* (r = 0.0.057, p = 0.02).

those of both human (0.5942) and chicken (0.6187) for ($p < 0.0001$ for both).

Standard Deviation in GC3

In all vertebrates, which have been studied in detail, including the amphibian *Xenopus laevis*, there is a correlation between the GC3 value of a gene and the genomic region in which it resides (Bernardi 2000); the amount of variation in GC3 is therefore likely to be a measure of compositional variation. On average, the variance in GC3 is lower in cold-blooded vertebrates than in either humans or chickens, and with only one exception this pattern is seen for all cold-blooded vertebrates. The exception is alligator, which shows a slightly, but not significantly, higher variance in GC3 than human and chicken (Fig. 1, Table 1); this could be due to its small sample size. Within classes (i.e., fish, reptiles, and amphibia) there was no evidence of variation in the level of compositional variation; all pairwise F tests for the equality of variances were nonsignificant.

Table 3. Mean GC3 and standard deviation in GC3 of the species studied and the temperature of their environment

Species/taxon	Mean GC3	σ GC3	Number of genes	Max T (°C)	Mean T (°C)
Alligator m.	0.6552	0.2050	17	33.0	20.0
Trachemys s.	0.5445	0.0458	16	28.0	24.0
Xenopus 1.	0.5073	0.1067	1872	20.0	18.0
Rana c.	0.5040	0.1122	80	24.0	21.0
Pleurodeles w.	0.6373	0.1403	27	25.0	24.0
Ambystoma m.	0.6500	0.1300	35	20.0	17.5
Petromyzon m.	0.7920	0.1024	32	20.0	19.0
Squalus a.	0.5734	0.1027	30	11.0	8.5
Raja e.	0.6044	0.1040	28	15.0	10.0
Heterodontus f.	0.5779	0.0836	63	15.0	10.0
Takifugu r.	0.6665	0.1120	102	28.0	25.0
Salmo g.	0.6432	0.1015	415	24.0	16.0
Danio r.	0.5894	0.0940	980	29.0	28.0
Anguilla a.	0.7313	0.1182	28	24.0	19.5
Carassius a.	0.5854	0.1042	172	20.0	15.0
Cyprinus c.	0.5852	0.1049	210	28.0	24.0
Fundulus h.	0.6647	0.1148	35	24.0	23.0
Nototheniidae	0.6959	0.0839	25	7.0	5.0

Correlation Between GC12 and GC3

Although we expect the variance in GC3 to be a measure of general compositional variation in vertebrates, other factors, such as translational selection on synonymous codon use, can generate variation in GC3 which is unrelated to composition. We therefore tested for compositional heterogeneity by considering the correlation between GC12 and GC3 (Table 1). As expected, and reported before, there is a strong correlation between GC12 and GC3 in humans and chickens. All the reptile species surveyed show a significant and fairly strong correlation; whereas only about half the amphibian and a quarter of the fish taxa show a significant but weak correlation. Although *Salmo gairdnerii* shows no correlation between GC12 and GC3, we found a significant positive correlation between GC1 and GC3 levels ($r = 0.295$, $p < 0.0001$).

The correlation between GC12 and GC3 could in principle be generated by selection on synonymous codon use; for example, if translational selection favored the leucine codons CUG and CUC, and particularly disfavored UUA, then variation in the strength of translational selection would generate a correlation between GC12 and GC3. To investigate whether such selection could be responsible for the correlation between GC12 and GC3, we repeated the analysis removing all codons of the sixfold degenerate amino acids leucine, arginine and serine (Table 1). Overall the correlations tended to be a little weaker, as expected. However, in only two cases did a significant correlation become nonsignificant; these were *Alligator mississipiensis* and *Carassius auratus*, in which the Pearson correlation coefficient dropped to 0.141 and 0.077, respectively. The correlation between GC1 and GC3 in *Salmo gairdnerii* remained significant ($r = 0.029$, $p < 0.001$).

Correlation Between GC3 in Humans and GC3 in Zebrafish

Hughes et al. (1999) found that the GC3 values of genes in a crocodile and a turtle were correlated with the GC3 values of homologous genes in chicken; this suggested that crocodiles and turtles have large-scale compositional variation, since the GC3 value of a gene is correlated with the G + C content of the genomic region in which the gene resides, and chickens show substantial large-scale variation in composition. We therefore performed a similar analysis on the zebrafish, *Danio rerio*. We chose *Danio rerio* because Unigene lists a human homologue for many of the zebrafish genes, and *Danio rerio* shows a significant correlation between GC12 and GC3. Surprisingly, although teleosteans and tetrapods separated more than 450 million years ago (Hedges 2001), there is a significant correlation between human and zebrafish GC3 values ($r = 0.057$, $p = 0.02$) (Fig. 2).

However, this result could have been caused by a correlation in amino acid composition between human and zebrafish genes. For example, it may be that in both species leucine is preferentially coded for by CUG; if there is variation in the frequency of leucine across genes and the frequency of leucine is correlated in humans and zebrafish, then there will be a correlation between GC3 in humans and GC3 in zebrafish. Therefore, to check that the correlation between human and zebrafish GC3 was not biased because of amino acid composition, we looked at the correlation

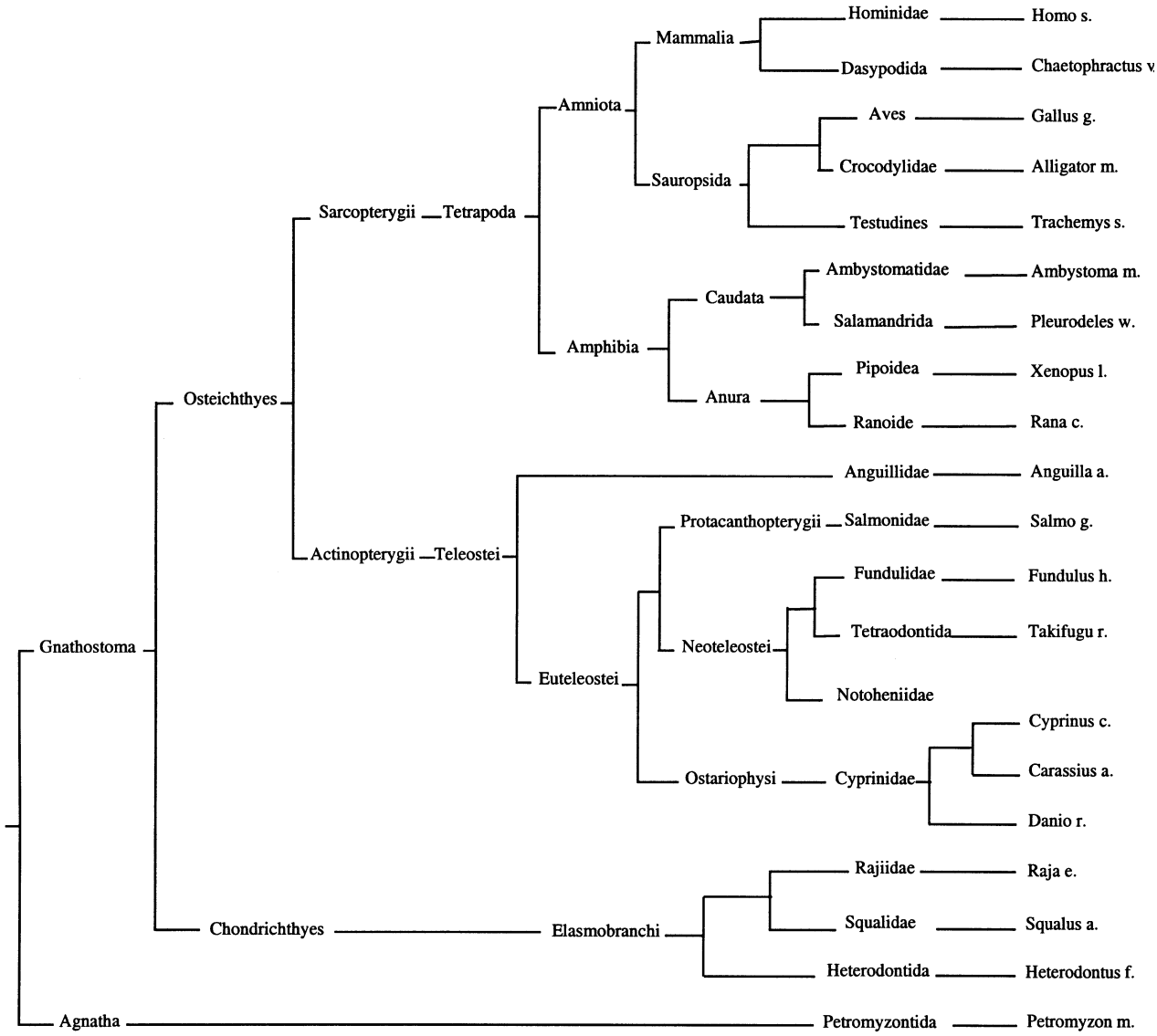

Fig. 3. Phylogeny of the different species and taxa studied [adapted from the NCBI database, Johnson and Paterson (1997), and Nelson (1994) for the fish].

coefficients between the GC3 level of human and that of zebrafish genes for each amino acid individually. These correlations were always found to be positive but never significant (results not shown): the probability of observing 18 positive, and no negative, correlations by chance is 3.8×10^{-6}. As an alternative test of overall significance, we also combined the probabilities for each correlation using the approximation that $-2\ln P$ is χ^2 distributed, with 2 degrees of freedom for each correlation, where P is the probability value from a one-tail test: the overall χ^2 was equal to 58.27, with 36 degrees of freedom, which is highly significant ($p < 0.0001$), therefore confirming that the correlations in amino acid composition were not responsible for the correlation observed between the GC3 level of human and that of zebrafish genes.

Thermal Stability

Finally, to investigate the thermal stability hypothesis, we compiled information on the body temperature of the cold-blooded species we have considered. For fish, we have assumed that their body temperature corresponds to the temperature of their environment; this can be quite variable, so we have compiled information on the maximum and minimum temperatures at which they can live. We performed four analyses: we considered the correlation between the mean GC3 and the standard deviation in the GC3 versus the maximum temperature, as well as the temperature midway between the maximum and the minimum (midtemperature) (Table 3). To remove phylogenetic nonindependence we used the method of orthogonal contrasts

Table 4. Correlation coefficients obtained with the method of orthogonal contrasts for the mean GC3 and standard deviation in GC3 versus the mean T and maximum T (°C)

	Mean GC3		σ GC3	
	Mean T	Max T	Mean T	Max T
All species	$r = 0.22$ ($p = 0.34$)	$r = 0.31$ ($p = 0.18$)	$r = 0.29$ ($p = 0.22$)	$r = 0.36$ ($p = 0.12$)
Species with No. genes > 50	$r = 0.28$ ($p = 0.50$)	$r = 0.05$ ($p = 0.20$)	$r = 0.40$ ($p = 0.32$)	$r = 0.33$ ($p = 0.42$)
Cold-blooded only	$r = 0.16$ ($p = 0.54$)	$r = 0.40$ ($p = 0.12$)	$r = 0.14$ ($p = 0.60$)	$r = 0.273$ ($p = 0.31$)
Cold-blooded with No. genes > 50	$r = 0.63$ ($p = 0.13$)	$r = 0.48$ ($p = 0.28$)	$r = 0.51$ ($p = 0.24$)	$r = 0.36$ ($p = 0.43$)

(Felsenstein 1985) assuming the phylogeny given in Fig. 3. We found no correlation between the mean GC3 or the standard deviation in GC3 and either measure of their body temperatures whether we restricted our analysis to species with more than 50 genes or included both warm-blooded and cold-blooded species (Table 4).

Discussion

We have investigated some of the compositional properties of coding sequences in a broad range of cold-blooded vertebrates and compared their properties to those of two warm-blooded species, humans and chickens. Although on average GC3 is lower in cold-blooded vertebrates than humans and chickens, there is substantial variation in the mean GC3 between cold-blooded vertebrate species, with many showing significantly greater mean GC3. In contrast, almost all cold-blooded vertebrates show substantially less variability in GC3 between genes within their genomes than warm-blooded species. However, in a number of cold-blooded vertebrates there is a significant correlation between GC12 and GC3, which suggests that there is systematic variation in the pattern of base substitution across their genome. This variation in the pattern of base substitution between genes is expected to produce variation in composition in introns and flanking DNA and, quite possibly, in intergenic DNA. The correlations are generally much weaker than are seen in homeotherms. Thus our study largely corroborates previous work—poikiliotherms have significantly less compositional heterogeneity than homeotherms—although many poikiliotherms show significant variation in their patterns of base composition.

In addition, we show that the GC3 values in the zebrafish are correlated with those in humans, which suggests that the two species share some basic pattern of genome organization. If so, it is remarkable that the compositional structure has been preserved over 450 million years of evolution.

The fact that both groups of homeotherms, birds and mammals, show much greater compositional variation than poikilotherms led to the suggestion that the increase in variation was associated with body temperature and that it might be associated with the thermal stability of DNA—GC base pairs are more thermally stable than AT base pairs. The thermal stability hypothesis was supported by the observation that two species of fish living at high temperatures (the killifish *Cyprinodon salinus* and tipalia *Oreochromis grahami*) showed a relatively high compositional heterogeneity (Bernardi 1990). Kadi (1993) also suggested that the higher G + C levels attained by avian compared to mammalian genomes might be correlated with the higher body temperature of birds (41–43°C) compared to mammals (37°C). However, we found no correlation between either the mean GC3 or the standard deviation in GC3 and the body temperature among our taxa. Thus the thermal stability hypothesis does not appear to explain the general patterns of base composition among the species we have studied.

References

Akashi H (1994) Synonymous codon usage in Drosophila melanogaster: Natural selection and translational accuracy. Genetics 136(3):927–935

Bernardi G (2000) Isochores and the evolutionary genomics of vertebrates. Gene 241:3–17

Bernardi G (1995) The human genome: Organization and evolutionary history. Annu Rev Genet 29:445–476

Bernardi G, Bernardi G (1986) Compositional constraints and genome evolution. J Mol Evol 24:1–11

Bernardi G, Bernardi G (1990a) Compositional transitions in the nuclear genomes of cold-blooded vertebrates. J Mol Evol 31:282–293

Bernardi G, Bernardi G (1990b) Compositional patterns in the nuclear genomes of cold-blooded vertebrates. J Mol Evol 31:265–281

Bernardi G, Bernardi G (1991) Compositional properties of nuclear genes from coldblooded vertebrates. J Mol Evol 33:57–67

Bernardi G, Mouchiroud D, Gautier C, Bernardi G (1988) Compositional patterns in vertebrate genomes: Conservation and change in evolution. J Mol Evol 28:7–18

Brown TC, Jiricny J (1988) Different base/base mispairs are corrected with different efficiency and specificities in monkey kidney cells. Cell 54:705–711

DeBry RW, Marzluff WF (1994) Selection on silent sites in the rodent H3 histone gene family. Genetics 138:191–202

D'Onofrio G, Mouchiroud D, Aissani B, Gautier C, Bernardi G (1991) Correlations between the compositional properties of human genes, codon usage, and amino acid composition of proteins. J Mol Evol 32:504–510

Eyre-Walker A (1993) Recombination and mammalian genome evolution. Proc Roy Soc Ser B 252:237–243

Eyre-Walker A (1999) Evidence of selection on silent site base composition in mammals: Potential implications for the evolution of isochores and junk DNA. Genetics 152:675–683

Felstenstein J (1985) Phytogenies and the comparative method. Am Nat 125:1–15

Filipski J (1987) Correlation between molecular clock ticking, codon usage, fidelity of DNA repair, chromosome banding and chromtin compactness in germline cells. FEBS Lett 217: 184–186

Filipski J, Thiery JP, Bernardi G (1973) An analysis of the bovine genome by Cs2SO4-Ag+ density centrifugation. J Mol Biol 80:177–197

Hedges SB (2001) Major events in early vertebrate evolution: Palaeontology, phylogeny, genetics and development. In: Ahlberg PE (ed) Molecular evidence for the early history of living vertebrates. Taylor & Francis, London, pp 119–134

Holmquist GP (1992) Chromosome bands, their chromatin flavors and their functional features. Am J Hum Genet 51:17–37

Hughes AL, Yeager M (1997) Comparative evolutionary rates of introns and exons in murine rodents. J Mol Evol 45:125–130

Hughes S, Zelus D, Mouchiroud D (1999) Warm-blooded isochore structure in Nile crocodile and turtle. Mol Biol Evol 16:1521–1527

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Johnson GD, Patterson C (1997) Relationships of lower euteleostean fishes. In: Stiassny MLJ, Parenti LR, Johnson, GD (eds) Interrelationships of fishes. Academic Press, San Diego, pp 251–253.

Kadi F, Mouchiroud D, Sabeur G, Bernardi G (1993) The compositional patterns of the genomes and their evolutionary implications. J Mol Evol 37:544–551

Kliman R, Hey J (1994) The effects of mutation and natural selection on codon bias in the genes of Drosophila. Genetics 137:1049–1056

Mouchiroud D, Gautier C, Bernardi G (1988) The compositional distance of coding sequences and DNA molecules in human and murids. J Mol Evol 27:311–320

Musto H, Cruveiller S, D'Onofrio G, Romero H, Bernardi G (2001) Translational selection on codon usage in *Xenopus laevis*. Mol Biol Evol 18:1703–1707

Nekrutenko A, Li W-H (2000) Assessment of compositional heterogeneity within and between eukaryotic genomes. Genome Res 10:1986–1995

Nelson JS (1994) Fishes of the world. John Wiley & Sons, New York

Shields DC, Sharp PM, Higgins DG, Wright F (1988) "Silent" sites in *Drosophila* are not neutral: Evidence of selection among synonymous codons. Mol Biol Evol 5:704–716

Suoeka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Wolfe KH, Sharp PM, Li W-H (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285