J Mol Evol (1994) 39:160–173

Journal of Molecular Evolution

Analysis of Directional Mutation Pressure and Nucleotide Content in Mitochondrial Cytochrome *b* Genes

Lars S. Jermiin, Dan Graur,* Roger M. Lowe, Ross H. Crozier

School of Genetics and Human Variation, La Trobe University, Bundoora, Victoria 3083, Australia

Received and accepted: 23 February 1994

Abstract. We present a new approach for analyzing directional mutation pressure and nucleotide content in protein-coding genes. Directional mutation pressure, the heterogeneity in the likelihood of different nucleotide substitutions, is used to explain the increasing or decreasing guanine–cytosine content (GC%) in DNA and is represented by μ_D, in agreement with Sueoka (1962, *Proc Natl Acad Sci USA* 48:582–592). The new method uses simulation to facilitate identification of significant A + T or G + C pressure as well as the comparison of directional mutation pressure among genes, even when they are translated by different genetic codes. We use the method to analyze the evolution of directional mutation pressure and nucleotide content of mitochondrial cytochrome *b* genes. Results from a survey of 110 taxa indicate that the cytochrome *b* genes of most taxa are subjected to significant directional mutation pressure and that the gene is subject to A + T pressure in most cases. Only in the anseriform bird *Cairina moschata* is the cytochrome *b* gene subject to significant G + C pressure. The GC% at nonsynonymous codon sites decreases proportionately with increasing A + T pressure, and with a slope less than one, indicating a presence of selective constraints. The cytochrome *b* genes of insects, nematodes, and eumycotes are subject to extreme A + T pressures (μ_D = 0.123, 0.224, and 0.130) and, in parallel, the GC% of the nonsynonymous codon sites has decreased from about 0.44 in organisms that are not subjected to A + T or G + C pressure to about 0.332, 0.323, and 0.367, respectively. The distribution of taxa according to the GC% at nonsynonymous codon sites and directional mutation pressure supports the notion that variation in these parameters is a phylogenetic component.

Key words: Directional mutation pressure — A + T pressure — G + C pressure — Synonymous codon sites — Nonsynonymous codon sites — Mitochondrial cytochrome *b* gene — Metazoa — Magnoliophyta — Chlorophyta — Eumycota

Introduction

More than 30 years age Freese (1962) and Sueoka (1962) formulated a quantitative theory of directional mutation pressure and used it to explain wide interspecific (Lee et al. 1956; Belozersky and Spirin 1958) and narrow intragenomic (Sueoka 1959; Sueoka et al. 1959; Rolfe and Meselson 1959) heterogeneities in the base composition of bacterial DNA. The theory is based on the assumption that all nucleotide substitutions in DNA are not equiprobable, but can have an overall directionality toward higher or lower GC%. The theory predicts that the GC% is determined by selective constraints and two antagonistic mutation rates u and v; u being the mutation rate from an A:T or a T:A nucleotide pair (α pairs) to a G:C or a C:G nucleotide pair (γ pairs), with v being the reverse mutation rate. Mathematically, directional mutation pressure is defined as occurring when $\mu_D \neq 0.5$, where $\mu_D = v/(v + u)$, and is equal to the GC% in those nucleotide sites that are selectively neutral and in equilibrium (Sueoka 1962, 1988, 1992).

* *Permanent address:* Department of Zoology, George S. Wise Faculty of Life Science, Tel Aviv University, Ramat Aviv 69978, Israel
Correspondence to: L.S. Jermiin

The first experimental study of directional mutation pressure showed that the GC% of *Escherichia coli* increases in the presence of a *mutT* mutant (Cox and Yanofsky 1967). Later, it was found that the GC% of various elements in protein-coding bacterial and mitochondrial genes responds differently to directional mutation pressure and that they are subject to different degrees of selective pressure (Jukes and Bhushan 1986; Muto and Osawa 1987). In particular, third codon positions appear more affected by directional mutation pressure than do first and second codon positions. As a result, the third codon position is considered nearly neutral to selection whereas the other two codon positions are regarded as being selectively constrained (Jukes and Bhushan 1986; Muto and Osawa 1987; Sueoka 1988, 1992).

The theory of directional mutation pressure also contains a quantitative definition of the relative contributions of directional mutation pressure and selective constraints to different regions of DNA, e.g., the first, second, and third codon positions of protein-coding genes (Sueoka 1988, 1992). Assuming selective neutrality of the third codon position, and equilibrium between directional mutation pressure and selective constraints at the first and second codon positions, the relationship between the average GC% of the first and second codon positions ($\hat{P}_{12}$) and the GC% of the third codon position ($\hat{P}_3$) is expressed as $\hat{P}_{12} = E_p + \varepsilon_{12}(\hat{P}_3 + E_p)$, where E_p is the point at which $\hat{P}_{12}$ equals $\hat{P}_3$, and ε_{12} is the degree of neutrality of the first two codon positions (Sueoka 1988). Conversely, the degree of selective constraints of the first and second codon positions is defined as $1 - \varepsilon_{12}$ (Sueoka 1988).

The theory of directional mutation pressure is supported by the presence of significant correlations between the GC% at different codon positions (D'Onofrio et al. 1991; Jukes and Bhushan 1986) and between the GC% of different genomic parts (tRNA, rRNA, protein, and spacers) and that of the total genome (Muto and Osawa 1987). Further support to the theory is offered by the existence of significant correlations between the relative abundance of particular amino acids and the GC% of the total genome (Sueoka 1961a,b), of silent sites (Collins and Jukes 1993), of first and second codon positions (D'Onofrio et al. 1991), of third codon positions (Sueoka 1992), and of codon families (Crozier and Crozier 1992, 1993; Jermiin and Crozier 1994). The theory is also supported by reports of AT or GC bias in the third codon positions of a variety of bacterial protein-coding genes (Muto and Osawa 1987; Ohama et al. 1987, 1989; Ohkubo et al. 1987; Hemmingsen et al. 1989; Ohtaka and Ishikawa 1993) and by the accumulation of A and T in the coding and noncoding regions of insect mitochondrial DNA (e.g., Crozier and Crozier 1993).

In opposition to the view above, many authors (e.g., Kagawa et al. 1984; Bernardi et al. 1985; Bernardi and Bernardi 1986; Wada and Suyama 1986; Kushino et al. 1987; Mouchiroud et al. 1988; Bernardi 1989; D'Onofrio et al. 1991) argue that variation in GC% reflects the functional significance of the GC% in the DNA. While this is true for replacement sites in protein-coding genes, it is hard to imagine that the GC% of silent sites selectively is determined, for example, on the basis of infinitesimal increments in heat stability of DNA by single mutations of A or T to G or C (Sueoka 1992). Nevertheless, the negative regression of the relative usage of AGG and TTG on the GC% of third codon position in Sueoka's (1992: Table 3) analysis of human protein-coding genes and the variation in codon usage of four insect cytochrome *b* genes (Jermiin and Crozier 1994) support the notion that codon usage is not always positively associated with the directional mutation pressure.

While some authors (e.g., Sueoka 1988, 1992; Collins and Jukes 1993) found that the directional mutation pressure on genes is quite variable within single taxa, little attempt has been made to compare the directional mutation pressure on homologous genes derived from different taxa. Analysis of homologous genes ensures that intergenic differences in selective constraints are comparatively small relative to those which exist among nonhomologous genes, and this enables a more precise evaluation of whether variation in the GC% among taxa is associated with directional mutation pressure, as previously proposed (Jukes and Bhushan 1986; Muto and Osawa 1987; Osawa et al. 1992; Jermiin and Crozier 1994), whether variation of the directional mutation pressure among taxa has a phylogenetic component, as suggested by Muto and Osawa (1987), Osawa et al. (1992), and Jermiin and Crozier (1994), and whether the gene products have responded to directional mutation pressure, as previously proposed (Jukes and Bhushan 1986; Muto and Osawa 1987; Osawa et al. 1992; Jermiin and Crozier 1994).

The present study addresses these issues. We have chosen to use the mitochondrial protein-coding cytochrome *b* gene from a variety of taxa, because (1) it encodes for an apoprotein which is well known with respect to its structure and function (Hatefi 1985; Howell and Gilbert 1988; Howell 1989; di Rago et al. 1990; Tron et al. 1991; Crozier and Crozier 1992); (2) it is a relatively large mitochondrial protein-coding gene, (3) it is the most frequently used mitochondrial gene in phylogenetic and evolutionary studies (e.g., Irwin et al. 1991; Helm-Bychowski and Cracraft 1993; Ma et al. 1993; Kornegay et al. 1993; Kusmierski et al. 1993; Martin and Palumbi 1993), and (4) details regarding directional mutation pressure on mitochondrial DNA are generally lacking. (See Jukes and Bhushan 1986 and Asakawa et al. 1991 for exceptions.) In order to compare the patterns of evolution among DNA sequences which are translated using different genetic codes, we present a new approach for calculating GC% at the synonymous and nonsynonymous codon sites and for cal-

Table 1. Mitochondrial genetic codes and codon families used in the present study[a]

Amino acid	Vertebrates	Arthropods, nematodes	Echinoderms	Plants	Euascomycetes	Yeasts
Ala (A)	GCN	GCN	GCN	GCN	GCN	GCN
Arg (R)	CGN	CGN	CGN	CGN	CGN	CGN
Arg (R)	—	—	—	AGR	AGR	AGR
Asn (N)	AAY	AAY	AAY/AAA	AAY	AAY	AAY
Asp (D)	GAY	GAY	GAY	GAY	GAY	GAY
Cys (C)	TGY	TGY	TGY	TGY	TGY	TGY
Glu (E)	GAR	GAR	GAR	GAR	GAR	GAR
Gln (Q)	CAR	CAR	CAR	CAR	CAR	CAR
Gly (G)	GGN	GGN	GGN	GGN	GGN	GGN
His (H)	CAY	CAY	CAY	CAY	CAY	CAY
Ile (I)	ATY	ATY	ATY/ATA	ATY/ATA	ATY/ATA	ATY
Leu (L)	CTN	CTN	CTN	CTN	CTN	—
Leu (L)	TTR	TTR	TTR	TTR	TTR	TTR
Lys (K)	AAR	AAR	AAG	AAR	AAR	AAR
Met (M)	ATR	ATR	ATG	ATG	ATG	ATR
Phe (F)	TTY	TTY	TTY	TTY	TTY	TTY
Pro (P)	CCN	CCN	CCN	CCN	CCN	CCN
Ser (S)	TCN	TCN	TCN	TCN	TCN	TCN
Ser (S)	AGY	AGN	AGN	AGY	AGY	AGY
Thr (T)	ACN	ACN	ACN	ACN	ACN	ACN
Thr (T)	—	—	—	—	—	CTN
Trp (W)	TGR	TGR	TGR	TGG	TGR	TGR
Tyr (Y)	TAY	TAY	TAY	TAY	TAY	TAY
Val (V)	GTN	GTN	GTN	GTN	GTN	GTN
Stop (*)	TAR	TAR	TAR	TAR	TAR	TAR
Stop (*)	AGR	—	—	TGA	—	—

[a] R = A or G, Y = C or T, N = any base. The genetic codes are from Osawa et al. (1992). They based the arthropod mitochondrial genetic code on *Drosophila* and were therefore unsure about the function of the codon AGG. Since this codon translate to Ser in the mitochondrial genome of *Apis mellifera* (Crozier and Crozier 1993), we consider the mitochondrial genetic codes of arthropods and nematodes identical

culating and evaluating the directional mutation pressure.

The Synonymous-Sites Aproach

Synonymous and Nonsynonymous Codon Sites. In a protein-coding DNA sequence a synonymous codon is a triplet which may potentially undergo nucleotide substitution without changing the amino acid (Li and Graur 1991). The mitochondrial genetic codes of vertebrates, arthropods, nematodes, echinoderms, plants, euascomycetes, and yeasts differ from one another (Osawa et al. 1992), and a general framework for analyzing the GC% of these taxonomic groups is necessary.

A simple approach involves treating each codon family as a unit. A codon family is a group of up to four synonymous codons which differ from one another at the third codon position (Li and Graur 1991). Accordingly, we define nonsynonymous codon sites as those codon positions which specify the codon family and synonymous codon sites as those codon positions which potentially may undergo substitution without changing the codon family.

This approach has three main consequences: (1) all third codon positions in the mitochondrial protein-coding genes in vertebrates, arthropods, nematodes, echinoderms (except AAG and ATG), plants (except ATG and TGG), euascomycetes (except ATG), and yeasts are synonymous sites (Table 1); (2) third codon positions in codon families consisting of more than one codon will always be occupied by A or G, by T or C, by A, T, or C, or by A, T, C, or G (Table 1); (3) all first and second codon positions as well as a few third codon positions (ATG and AAG in echinoderms, ATG and TGG in plants, and ATG in euascomycetes) are nonsynonymous codon sites (Table 1).

Thus, μ_D can be estimated by considering the G + C content at synonymous codon sites. Since most amino acids are specified by one codon family only, variation in the G + C content at nonsynonymous codon sites will reflect variation in the ratio between amino acids specified by G + C rich codon families and those specified by A + T rich codon families.

The only drawback of this approach is that the codons specifying Arg, Leu, Ser and Thr in some cases are encoded by two codon families (Table 1); the codon families specifying Ser (all six mitochondrial genetic codes) or Thr (the yeast mitochondrial genetic code) are each linked via two concurrent substitutions in the first and second codon positions and the codon families specifying Arg (non-metazoan mitochondrial genetic codes) or Leu (all but the yeast mitochondrial genetic code) via a substitution in first codon position (Table 1). While these silent substitutions between synonymous codon families are possible, they are unlikely to have much effect on conclusions which are drawn from the analysis of the GC% at nonsynonymous codon sites, and, hence, an evaluation of the effect of directional mutation pressure on the GC% at nonsynonymous codon sites (or the amino acid composition) is facilitated.

The GC% of Synonymous and Nonsynomymous Codon Sites. The above approach allows us to calculate the observed GC% of a DNA sequence (P_{obs}) as

$$P_{obs} = (1 - p)\, P_{non} + p\, P_{syn} \qquad (1)$$

where P_{syn} is the GC% at synonymous codon sites, P_{non} is the GC% at nonsynonymous codon sites, and p is the proportion of synonymous codon sites in the DNA sequence after the exclusion of the stop codon. Of special interest are two values which express the GC% of a protein-coding DNA sequence when its synonymous codon sites are saturated with either α or γ pairs. These values, P_{min} and P_{max}, are given as

$$P_{min} = (1 - p)\, P_{non} \tag{2}$$

and

$$P_{max} = (1 - p)\, P_{non} + p \tag{3}$$

where P_{min} is the GC% of a DNA sequence saturated with α pairs and P_{max} is the GC% of a DNA sequence saturated with γ pairs. In order to determine P_{min} and P_{max}, the original DNA sequence is used to generate two sequences in which the synonymous codon sites are saturated either with γ pairs or with α pairs.

The proportion of synonymous codon sites in a DNA sequence (p) is given by

$$p = P_{max} - P_{min} \tag{4}$$

whereas the GC% of its synonymous and nonsynonymous codon sites is given by

$$P_{syn} = \frac{(P_{obs} - P_{min})}{(P_{max} - P_{min})} \tag{5}$$

and

$$P_{non} = \frac{(P_{obs} - p\ P_{syn})}{1 - p} \tag{6}$$

Directional Mutation Pressure. In an analysis of the mitochondrial GC% among five species, Jukes and Bhushan (1986) found that the GC% at synonymous codon sites correlates well with that of the control region. Thus, it is reasonable to assume (1) that nucleotide substitutions at the synonymous codon sites are selectively neutral, (2) that the GC% at synonymous codon sites is governed only by the mutational processes, and (3) that P_{syn} equals 0.5 when the nucleotide substitutions at these sites are random. In agreement with Sueoka (1962), we define μ_D as being equal to P_{syn} only when the synonymous codon sites are selectively neutral and in equilibrium, i.e., when $uP_{syn} = v(1 - P_{syn})$.

If a protein-coding DNA sequence is subject to random nucleotide substitutions only, its observed GC% (P_{ran}) is given by

$$P_{ran} = (1 - p)\, P_{non} + 0.5\, p \tag{7}$$

If the same DNA sequence is subject to directional mutation pressure, then $P_{syn} \neq 0.5$. In accordance with Sueoka (1988), we say that the DNA sequence is subject to G + C pressure when $0.5 < \mu_D \leq 1.0$ (i.e., when $P_{ran} < P_{obs}$) and to A + T pressure when $0.0 \leq \mu_D < 0.5$ (i.e., when $P_{obs} < P_{ran}$). Consequently, P_{min} and P_{max} correspond to the GC% obtained under maximum A + T or G + C pressure.

Testing Directional Mutation Pressure. The distribution of the GC% at equilibrium ($\hat{P}$) follows the binomial distribution, with a variance given by

$$\sigma_b^2(P) = \frac{\hat{P}\,(1 - \hat{P})}{b} \tag{8}$$

where b is the number of base pairs per DNA molecule (Freese 1962; Sueoka 1962). This estimate of variance is appropriate to use when two values of μ_D are compared, but is not appropriate for testing whether μ_D differs significantly from 0.5. As $\hat{P}$ diverges from 0.5, the variance decreases and certain values of μ_D will therefore be considered significantly different from 0.5, if equation (8) is applied, and not significantly different from 0.5 if the variance of a randomly mutating (i.e., where $\mu_D = 0.5$) DNA sequence is used. In order to evaluate μ_D conservatively, we apply the latter approach when testing whether $\mu_D \neq 0.5$.

The statistical test described above applies directly to protein-coding genes which are translated by the vertebrate, arthropod, nematode, and yeast mitochondrial genetic codes because they only consist of two- and fourfold degenerated codon families. However, the mitochondrial genetic codes for echinoderms, plants, and euascomycetes accomodate threefold degenerated codon families (Table 1) specifying for Ile (plants, euascomycetes, and echinoderms) and Asn (echinoderm), and the asymmetries of these induce a bias in P_{ran}. For example, if an echinoderm protein-coding DNA sequence encodes for Ile or Asn exclusively and is subject to random mutations only, then $P_{syn} = 0.333$. Indeed, this would imply that the GC% at the synonymous codon sites at which there is not directional mutation pressure equals 0.333 and not 0.5, as the theory predicts (Sueoka 1962). The bias of P_{ran} depends on the relative proportion of threefold degenerated codon families and, although this proportion generally is small, it is necessary to correct for the bias they may induce.

A solution to this problem can be obtained by simulation.[1] The corresponding codon family sequence forms a template for randomly generated, synonymously coding DNA sequences (Table 2, Fig. 1a). Many simulations are performed and P_{syn} is determined for each newly generated DNA sequence, as is the arithmetic mean ($\bar{P}_{syn}$) of the emerging frequency distribution (Fig. 1b). All values of P_{syn} (original and synthesized DNA sequences) are multiplied by $1/(2\bar{P}_{syn})$ to produce an unbiased estimate of μ_D and a normalized frequency distribution. The procedure proportionally moves the density distribution to a position where its arithmetic mean equals 0.5 (Fig. 1c). Thus, the normalization facilitates comparing μ_D values from genes that are translated by different genetic codes and, most importantly, the normalization does not alter the probability of obtaining a particular value of μ_D.

Analysis of the Mitochondrial Cytochrome *b* Gene

The proposed analytical approach is used to examine the directional mutation pressure and GC% at nonsynonymous codon sites in the mitochondrial cytochrome *b* genes of 110 taxa (Table 3).

The Number of Codons and the Proportion of Synonymous Codon Sites

The size of the cytochrome *b* gene is relatively uniform among taxa (sample mean: 380.5 ± 4.264 [SD] codons; range: 365–398 codons) (Table 3), implying that the size is well-preserved despite evolutionary divergences of up to 1.2 billion years (Hori and Osawa 1987). The cytochrome *b* gene is shorter in Nematoda and the ant (*Tetraponera rufonigra*) and longer in Magnoliophyta and Eumycota, and the deletions are confined mainly to the beginning and the end of the genes (Okimoto et al. 1992; Jermiin and Crozier 1994). However, since the redox centers Q_o and Q_i (Irwin et al. 1991) are maintained in these taxa, we believe that the apoproteins' configurations and functions are well-preserved.

The proportion of synonymous codon sites (p) is always 0.333 in the mitochondrial cytochrome *b* genes of vertebrates, nematodes, arthropods, and yeasts. This is

[1] The simulation program DMP, which was developed for this purpose, can be obtained from L.S.J.

Table 2. Determination of significant A + T or G + C pressure is explained using a hypothetical protein-coding DNA sequence[a]

Hypothetical DNA sequence	ATG	GGA	TTA	CCC	CTA	ATA	TTT	ACC	ATC	TAT	TTA	TCG	GAC	GCT	AAT	$P_{obs} = 0.356$
Codon family sequence	ATR	GGN	TTR	CCN	CTN	ATR	TTY	ACN	ATY	TAY	TTR	TCN	GAY	GCN	AAY	
A + T saturated DNA	ATA	GGT	TTA	CCT	CTT	ATA	TTT	ACT	ATT	TAT	TTA	TCT	GAT	GCA	AAT	$P_{min} = 0.222$
G + C saturated DNA	ATG	GGC	TTG	CCG	CTC	ATG	TTC	ACC	ATC	TAC	TTG	TCG	GAC	GCC	AAC	$P_{max} = 0.556$
Randomly generated DNA (1)	ATG	GGA	TTG	CCA	CTA	ATA	TTC	ACC	ATC	TAT	TTA	TCG	GAT	GCG	AAC	$P_{ran}(1) = 0.400$
Randomly generated DNA (2)	ATA	GGG	TTA	CCA	CTG	ATG	TTC	ACA	ATT	TAC	TTG	TCC	GAT	GCT	AAT	$P_{ran}(2) = 0.378$
Randomly generated DNA (3)	ATG	GGA	TTA	CCG	CTA	ATA	TTT	ACT	ATT	TAT	TTA	TCA	GAC	GCA	AAC	$P_{ran}(3) = 0.331$
...								.....								
Randomly generated DNA (n)	ATG	GGG	TTG	CCC	CTA	ATA	TTT	ACG	ATC	TAC	TTG	TCG	GAT	GCG	AAC	$P_{ran}(n) = 0.467$

[a] The analytical procedure described here was used for each of the DNA sequences in Table 3. The numbers of G and C in the hypothetical DNA sequence are counted together with the total number of nucleotides, and the observed GC% is calculated (P_{obs}). Using the arthropod mitochondrial genetic code, the hypothetical DNA sequence is translated into a codon family sequence (variable codon positions are *underlined*). The variable codon positions in the latter are saturated with A and T or G and C to produce the A + T saturated and G + C saturated, synonymously coding DNA sequences, and the GC% of these is calculated (P_{min} and P_{max}, respectively). The P_{min} and P_{max} values are used to calculate p [equation (4)], P_{syn} [equation (5)], and P_{non} [equation (6)] for the hypothetical DNA sequence. A new, synonymously coding DNA sequence is generated by randomly choosing among "correct" nucleotides for each of the variable codon positions, and its GC% [P_{ran} (1)] is obtained; this process is repeated n times to produce the frequency distribution of P_{ran} that allows us to determine whether the hypothetical DNA sequence is subjected to significant A + T and G + C pressure (Fig. 1a). Alternatively, the n P_{ran} values can be converted into n P_{syn} values [equation (5)], and the frequency distribution of the latter can be used to determine whether the hypothetical DNA sequence is subjected to significant A + T or G + C pressure (Fig 1b). The unbiased estimate of μ_D is obtained by a normalization in which the arithmetic mean ($\bar{P}_{syn}$) of the frequency distribution of P_{syn} values is used to generate a factor equal to $1/(2\bar{P}_{syn})$; all the P_{syn} values (hypothetical and synthesized DNA sequences) are multiplied by this factor to yield an unbiased estimate of μ_D and a normalized frequency distribution (Fig. 1c)

due to the lack of asymmetrical codon families in their mitochondrial genetic code (Table 1). The proportion of synonymous codon sites is not much lower in the cytochrome *b* genes of echinoderms, magnoliophytes, chlorophytes, and euascomycetes (sample mean: 0.317 ± 0.004 [SD]; range: 0.310–0.324) (Table 3), implying that 93.1–97.3% of all codons in those genes belong to two-, three-, or fourfold degenerate codon families. Accordingly, we conclude that the proportion of synonymous codon sites in the echinoderms, magnoliophytes, chlorophytes, and euascomycetes is affected only marginally by the occurrence of codon families with only one codon.

The Observed GC% of the DNA Sequences

Since the proportion of synonymous codon sites (p) specifies the difference between the GC% of a DNA sequences under maximum G + C pressure and that of a synonymously coding DNA sequence under maximum A + T pressure [equation (4)], the figures above imply that the GC% of synonymously coding mitochondrial cytochrome *b* genes may differ by up to 33.3% in vertebrates, arthropods, nematodes, and yeasts and by as much as 31.0–32.4% in the echinoderms, magnoliophytes, chlorophytes, and euascomycetes listed in Table 3.

Bearing this in mind we look at the observed GC% of DNA sequences in Table 3. Among these the maximum difference in the observed GC% is 0.311 for vertebrates, insects, and nematodes (*Cairina moschata* [$P_{obs} = 0.505$]; *Apis mellifera* [$P_{obs} = 0.194$]) and 0.147 for echinoderms, magnoliophytes, chlorophytes, and euascomycetes (*Strongylocentrotus purpuratus* [$P_{obs} = 0.429$]; *Aspergillus nidulans* [$P_{obs} = 0.282$]).

Since these two differences (0.311 and 0.147) are smaller than the corresponding values of p (0.333 and 0.310–0.324), it may be argued that the 110 DNA sequences in Table 3 encode for identical apoproteins (length polymorphism excluded) and that the variation in the observed GC% is due to differences in the GC% at the synonymous codon sites. Consequently, this implies that some of the DNA sequences in Table 3 are subject to G + C pressure and others to A + T pressure. However, directional mutation pressure on mitochondrial genes has not previously been statistically confirmed, although some mitochondrial genomes have been suggested to be under directional mutation pressure (e.g., Jukes and Bushhan 1986; Crozier and Crozier 1993; Jermiin and Crozier 1994).

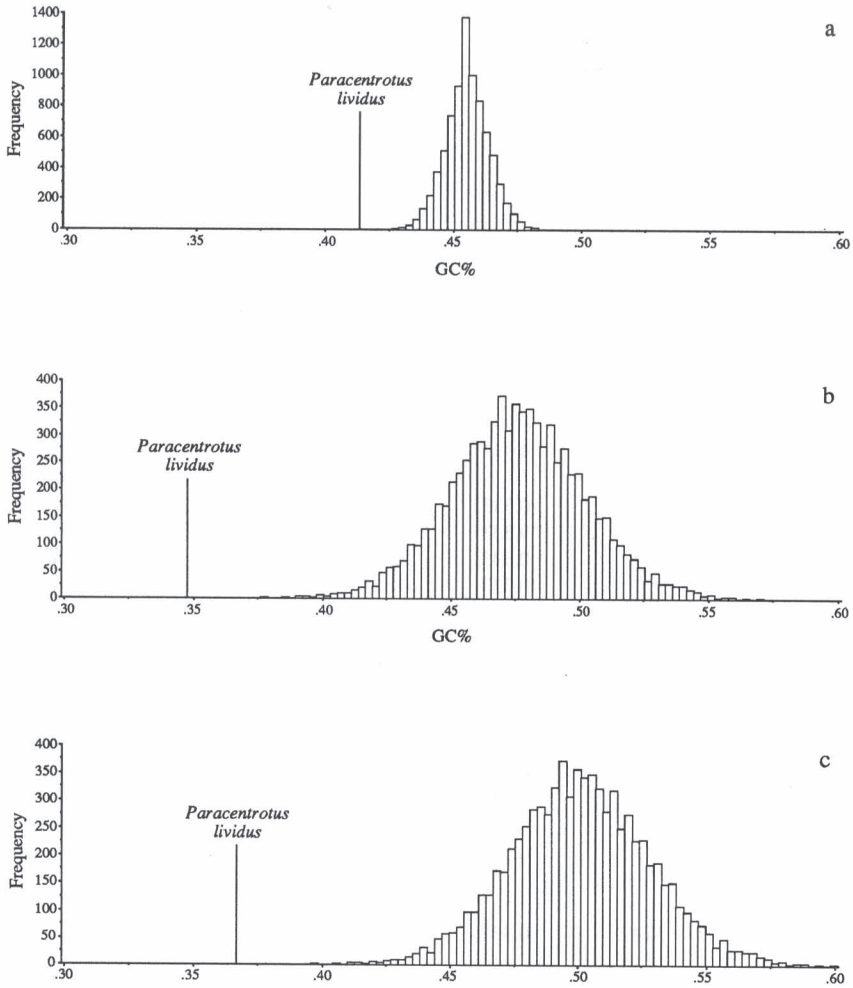

Fig. 1. Normalization of the estimate of directional mutation pressure μ_D is illustrated using the mitochondrial cytochrome *b* gene of the sea urchin *Paracentrotus lividus*. The density distributions were obtained from 8,000 simulations with random nucleotide substitutions at the synonymous codon sites. **a** The value of P_{obs} is plotted together with the frequency distribution of the simulated GC% for the whole gene (P_{ran}). **b** P_{syn} is plotted with the frequency distribution of the simulated GC% at the gene's synonymous codon sites (P_{syn} values derived from P_{ran} values). Note that the arithmetic mean of the frequency distribution equals 0.475, thus implying a bias induced by Ile and Asn. **c** μ_D is plotted with the frequency distribution of the normalized, simulated GC% at the gene's synonymous codon sites. Note that the arithmetic mean of the frequency distribution is equal to 0.5 and that normalization does not alter the probability of obtaining the nucleotide content bias of the specimen.

Alternatively, variation in the observed GC% may be due to differences in the GC% at nonsynonymous codon sites. This is supported since most DNA sequences in Table 3 encode for apoproteins that differ markedly from one another (comparison not shown).

To determine the cause of the variation in the observed GC%, it is necessary to determine whether or not the DNA sequences in Table 3 are subject to directional mutation pressure and whether the GC% at the nonsynonymous codon sites varies among taxa.

Directional Mutation Pressure and the GC% at Nonsynonymous Codon Sites

Significant directional mutation pressure is observed in 57% of the taxa and appears to be a common phenomenon since it is not confined to a single taxonomic group (Table 3). Furthermore, the majority of those taxa (83%) appear to be affected to a great extent by directional mutation pressure (group-wide 0.1% level), signifying the large magnitude of the directional mutation pressure.

Among the 63 taxa which are subject to significant directional mutation pressure only the duck (*Cairina moschata*) is subject to G + C pressure ($\mu_D = 0.595 \pm 0.025$; group-wide 2.5% level); the rest are subject to A + T pressure. The relative magnitude of the directional mutation pressure on the mitochondrial cytochrome *b* genes of eumycotes (arithmetic mean: $\mu_D = 0.130 \pm 0.053$ [SD]; range: 0.240–0.081) and insects (arithmetic mean: $\mu_D = 0.123 \pm 0.040$ [SD]; range: 0.222–0.052) is greater than those recorded for the other taxa.

The GC% at the nonsynonymous codon sites also varies but to a lesser extent than that of the directional mutation pressure (Table 3), possibly reflecting the fact that amino acid substitutions occur less frequently than nucleotide substitutions at silent sites. Among the DNA sequences in Table 3, the maximum difference in the GC% at nonsynonymous codon sites is 0.205 for vertebrates, insects, nematodes, and yeasts (*Vireo olicaceus* [$P_{non} = 0.471$]; *Apis mellifera* [$P_{non} = 0.265$]) and 0.120 for the echinoderms, magnoliophytes, chlorophytes, and euascomycetes (*Triticum aestivum* [$P_{non} = 0.479$]; *Podospora anserina* [$P_{non} = 0.369$]).

Although our approach differs from previous methods, the results obtained are in good agreement with those obtained from analyses of bacterial genomes (Jukes and Bhushan 1986; Muto and Osawa 1987;

Table 3. Mitochondrial cytochrome *b* DNA sequences from the taxa listed below were obtained from GenBank using the GenBank accession number shown[a]

Phylum Class Order Family Genus, species, subspecies	GenBank	Codons	p	P_{obs}	$\mu_D \pm SD$[b]	P_{non}
Chordata						
Mammalia						
Primates						
Hominidae						
Homo sapiens	J01415	380	0.333	0.462	0.516 ± 0.026	0.436
Rodentia						
Muridae						
Mus musculus	J01420	382	0.333	0.391	0.361 ± 0.025***	0.406
Rattus norvegicus	J01436	380	0.333	0.427	0.432 ± 0.025	0.425
Geomyidae						
Cratogeomys fumosus	L11903	379	0.333	0.392	0.319 ± 0.024***	0.429
C. gymnurus	L11905	379	0.333	0.379	0.293 ± 0.023***	0.422
C. merriami	L11906	379	0.333	0.387	0.311 ± 0.023***	0.425
C. tylorninus	L11909	379	0.333	0.384	0.303 ± 0.024***	0.425
C. castanops castanops	L11902	379	0.333	0.388	0.327 ± 0.024***	0.418
C. c. tamaulipensis	L11908	379	0.333	0.378	0.306 ± 0.024***	0.414
C. goldmani goldmani	L11904	379	0.333	0.387	0.325 ± 0.024***	0.418
C. g. rubellus	L11907	379	0.333	0.384	0.322 ± 0.024***	0.416
Pappageomys bulleri	L11900	379	0.333	0.384	0.311 ± 0.024***	0.420
Geomys bursarius juggosicularis	L11901	379	0.333	0.391	0.319 ± 0.024***	0.427
Hystricidae						
Hystrix africaeaustralis	X70674	379	0.333	0.398	0.348 ± 0.024***	0.422
Cavidae						
Cavia porcellus	—[c]	378	0.333	0.449	0.471 ± 0.026	0.438
Cetacea						
Balaenopteridae						
Balaenoptera physalus	X61145	379	0.333	0.443	0.449 ± 0.026	0.441
Delphinidae						
Stenella longirostris	X56292	379	0.333	0.421	0.393 ± 0.025**	0.435
S. attenuata	X56294	379	0.333	0.419	0.383 ± 0.025***	0.437
Carnivora						
Phocidae						
Phoca vitulina	X63726	379	0.333	0.440	0.446 ± 0.026	0.437
Perissodactyla						
Equidae						
Equus grevyi	X56282	379	0.333	0.446	0.475 ± 0.026	0.431
Rhinocerotidae						
Diceros bicornis	X56283	379	0.333	0.440	0.433 ± 0.025	0.442
Proboscidea						
Elephantidae						
Loxodonta africana	X56285	379	0.333	0.411	0.372 ± 0.025***	0.430
Artiodactyla						
Antilocapridae						
Antilocarpa americana	X56286	379	0.333	0.427	0.425 ± 0.025	0.429
Bovidae						
Bos taurus	J01394	379	0.333	0.436	0.430 ± 0.025	0.439
Capra hircus	X56289	379	0.333	0.419	0.401 ± 0.025**	0.427
Ovis aries	X56284	379	0.333	0.413	0.385 ± 0.025***	0.426
Giraffidae						
Giraffa camelopardalis	X56287	379	0.333	0.418	0.404 ± 0.025**	0.425
Cervidae						
Dama dama	X56290	379	0.333	0.404	0.367 ± 0.025***	0.422
Odocoileus hemionus	X56291	379	0.333	0.413	0.369 ± 0.025***	0.434
Tayassuidae						
Tayassa tajacu	X56296	379	0.333	0.433	0.427 ± 0.025	0.435
Tragulidae						
Tragulus napu	X56288	379	0.333	0.436	0.427 ± 0.025	0.441
Camelidae						
Camelus dromedarius	X56281	379	0.333	0.426	0.401 ± 0.025**	0.438

Table 3. Continued

Phylum Class Order Family Genus, species, subspecies	GenBank	Codons	p	P_{obs}	$\mu_D \pm SD^b$	P_{non}
Suidae						
Sus scrofa	X56295	379	0.333	0.420	0.417 ± 0.025	0.421
Marsupialia						
Didelphidae						
Didelphis virginiana	Z29573[d]	382	0.333	0.371	0.272 ± 0.023***	0.420
Monodelphis domestica	X70673	382	0.333	0.369	0.283 ± 0.023***	0.412
Aves						
Anseriformes						
Anatidae						
Cairina moschata	L08385	380	0.333	0.505	0.595 ± 0.025*	0.461
Galliformes						
Phasianidae						
Alectoris chucar	L08378	380	0.333	0.473	0.526 ± 0.026	0.446
Coturnix coturnix	L08377	380	0.333	0.461	0.484 ± 0.026	0.449
Gallus gallus	L08376	380	0.333	0.484	0.545 ± 0.026	0.454
Pavo cristatus	L08379	380	0.333	0.455	0.474 ± 0.026	0.446
Lophura nycthemera	L08380	380	0.333	0.467	0.492 ± 0.026	0.454
Meleagris gallopavo	L08381	380	0.333	0.454	0.471 ± 0.026	0.445
Numididae						
Numida meleagris	L08383	380	0.333	0.472	0.505 ± 0.026	0.455
Odontophoridae						
Lophortyx gambelii	L08382	380	0.333	0.475	0.526 ± 0.026	0.450
Cracidae						
Ortalis vetula	L08384	380	0.333	0.475	0.508 ± 0.026	0.458
Passeriformes						
Tyrannidae						
Empidonax minimus	X74251	380	0.333	0.455	0.474 ± 0.026	0.446
Turdidae						
Catharus guttatus	X74261	380	0.333	0.473	0.500 ± 0.026	0.459
Ptilonorhynchidae						
Ailuroedes melanotus	X74257	380	0.333	0.464	0.476 ± 0.026	0.458
Ptilonorhynchus violaceus	X74256	380	0.333	0.457	0.447 ± 0.026	0.462
Vireonidae						
Vireo olicaceus	X74260	378	0.333	0.463	0.447 ± 0.026	0.471
Laniidae						
Lanius ludovicianus	X74259	380	0.333	0.439	0.424 ± 0.025	0.446
Corvidae						
Cyanocitta cristata	X74258	380	0.333	0.449	0.445 ± 0.025	0.451
Paradisaeidae						
Diphyllodes magnificus	X74255	380	0.333	0.447	0.429 ± 0.025	0.457
Manucodia keraudrenii	X74252	380	0.333	0.441	0.434 ± 0.025	0.445
Ptiloris paradiseus	X74254	380	0.333	0.445	0.424 ± 0.025	0.455
Epimachus fastuosus	X74253	380	0.333	0.446	0.432 ± 0.025	0.453
Amphibia						
Anura						
Pipidae						
Xenopus laevis	M10188	379	0.333	0.380	0.311 ± 0.024***	0.414
Osteichthyes						
Acipenseriformes						
Acipenseridae						
Acipences transmontanus	X14944	380	0.333	0.466	0.487 ± 0.026	0.455
Cypriniformes						
Homalopteridae						
Crossostoma lacustre	M91245	380	0.333	0.472	0.516 ± 0.026	0.450
Cyprinidae						
Lythrusus roseipennis	X66456	379	0.333	0.456	0.464 ± 0.026	0.451
Cyprinus carpio	X61010	381	0.333	0.442	0.424 ± 0.025	0.450
Pegasiformes						
Centrarchidae						
Micropterus salmoides	L14074	380	0.333	0.483	0.518 ± 0.026	0.465

Table 3. Continued

Phylum Class Order Family Genus, species, subspecies	GenBank	Codons	p	P_{obs}	$\mu_D \pm SD^b$	P_{non}
Chondrichthyes						
Carcharhiniformes						
Carcharhinidae						
Carcharinus plumbeus	L08032	381	0.333	0.417	0.425 ± 0.025	0.413
C. porosus	L08033	381	0.333	0.409	0.412 ± 0.025*	0.407
Galeocerdo cuvier	L08034	381	0.333	0.381	0.336 ± 0.024***	0.403
Negaprion brevirostris	L08039	381	0.333	0.402	0.383 ± 0.025***	0.411
Prionace glauca	L08040	381	0.333	0.388	0.336 ± 0.024***	0.413
Sphyrnidae						
Sphyrna lewini	L08041	381	0.333	0.409	0.391 ± 0.025**	0.419
S. tiburo tiburo	L08042	381	0.333	0.409	0.404 ± 0.025**	0.411
S. t. vespertina	L08043	381	0.333	0.414	0.423 ± 0.025	0.409
Heterodontiformes						
Heterodontidae						
Heterodontus fransci	L08035	381	0.333	0.433	0.462 ± 0.025	0.419
Lamniformes						
Lamnidae						
Isurus oxyrhynchus	L08036	381	0.333	0.446	0.472 ± 0.026	0.433
I. paucus	L08037	381	0.333	0.467	0.530 ± 0.026	0.436
Lamna nasus	L08038	381	0.333	0.430	0.451 ± 0.025	0.419
Carcharhodon carcharias	L08031	381	0.333	0.433	0.457 ± 0.026	0.421
Echinodermata						
Echinoidea						
Echinoida						
Echinidae						
Paracentrotus lividus	J04815	380	0.322	0.415	0.367 ± 0.025***	0.446
Strongylocentrotidae						
Strongylocentrotus purpuratus	X12631	385	0.321	0.429	0.462 ± 0.026	0.425
Stelleroidea						
Spinulosida						
Asterinidae						
Asterina pectinifera	—[e]	380	0.324	0.388	0.339 ± 0.025***	0.420
Nematoda						
Secernentea						
Ascarida						
Ascarididae						
Ascaris suum	X54253	365	0.333	0.316	0.290 ± 0.024***	0.329
Rhabditida						
Rhabditidae						
Caenorhabditis elegans	X54252	370	0.333	0.263	0.157 ± 0.019***	0.316
Arthropoda						
Brachiopoda						
Anostraca						
Artermiidae						
Artemia franciscana	X69067	381	0.333	0.397	0.347 ± 0.024***	0.423
Insecta						
Hymenoptera						
Apidae						
Apis mellifera	L06178	383	0.333	0.194	0.052 ± 0.011***	0.265
Formicidae						
Tetraponera rufonigra	U02458	370	0.333	0.302	0.222 ± 0.022***	0.342
Diptera						
Drosophilidae						
Drosophila melanogaster	M37275	378	0.333	0.258	0.082 ± 0.014***	0.347
D. yakuba	X03240	378	0.333	0.262	0.077 ± 0.014***	0.355
D. albomicans	—[f]	378	0.333	0.276	0.111 ± 0.016***	0.360
D. angularis	—[f]	378	0.333	0.273	0.101 ± 0.015***	0.360
D. hypocausta	—[f]	378	0.333	0.275	0.109 ± 0.016***	0.359
D. kohkoa	—[f]	378	0.333	0.271	0.087 ± 0.015***	0.362
D. pallidifrons	—[f]	378	0.333	0.272	0.095 ± 0.015***	0.360

Table 3. Continued

Phylum Class Order Family Genus, species, subspecies	GenBank	Codons	p	P_{obs}	$\mu_D \pm SD^b$	P_{non}
D. virilis	—[f]	378	0.333	0.277	0.106 ± 0.016***	0.362
D. sulfurigaster albostigata	—[f]	378	0.333	0.275	0.106 ± 0.016***	0.360
D. s. bilimbata	—[f]	378	0.333	0.280	0.122 ± 0.017***	0.360
D. s. sulfurigaster	—[f]	378	0.333	0.271	0.093 ± 0.015***	0.360
Culicidae						
Anopheles gambia	L20934	378	0.333	0.276	0.103 ± 0.016***	0.362
A. quadrimaculatus	L04272	378	0.333	0.274	0.103 ± 0.016***	0.360
Magnoliophyta						
Magnoliopsida						
Fabales						
Fabaceae						
Vicia faba	X07237	392	0.315	0.415	0.292 ± 0.024***	0.475
Solanales						
Solanaceae						
Solanum tuberosum	X58437	393	0.315	0.418	0.320 ± 0.024***	0.468
Myrtales						
Onagraceae						
Oenothera berteriana	X07126	394	0.315	0.415	0.300 ± 0.024***	0.472
Liliopsida						
Cyperales						
Poaceae						
Zea mays	X00789	388	0.314	0.416	0.305 ± 0.024***	0.471
Oryza sativa	X53710	397	0.315	0.419	0.306 ± 0.024***	0.475
Triticum aestivum	X02352	398	0.315	0.423	0.313 ± 0.024***	0.478
Chlorophyta						
Chlorophyceae						
Volvocales						
Chlamydomonadaceae						
Chlamydomonas smithii	X55305	381	0.311	0.456	0.462 ± 0.026	0.458
C. reinhardtii	X52168	381	0.310	0.452	0.451 ± 0.026	0.458
Eumycota						
Hemiascomycetes						
Endomycetes						
Saccharomycetaceae						
Pichia pijperi	X66593	386	0.333	0.326	0.240 ± 0.022***	0.369
Schizosaccharomyces probe	X54421	386	0.333	0.295	0.189 ± 0.019***	0.348
Saccharomyces cereviciae	J01476	387	0.333	0.278	0.140 ± 0.016***	0.348
S. douglasii	X59280	385	0.333	0.272	0.117 ± 0.014***	0.349
Plectomycetes						
Eurotiales						
Trichocomaceae						
Aspergillus nidulans	J01388 J01389	387	0.320	0.282	0.081 ± 0.014***	0.377
Pyrenomycetes						
Sordariales						
Sordariaceae						
Podospora anserina	M61734 M30937	387	0.321	0.289	0.123 ± 0.017***	0.369

[a] Taxa are ordered roughly according to their systematic affiliation. For each taxon the following results were obtained: the number of codons, stop codons excluded (Codons); the proportion of synonymous codon sites (p); the observed GC% of the DNA sequence (P_{obs}); the G:C/A:T bias estimated as directional mutation pressure (μ_D) ± its standard deviation (SD); and the GC% of the nonsynonymous codon sites (P_{non})

[b] Test of μ_D (H_o: $\mu_D = 0.5$; H_1: $\mu_D \neq 0.5$). Deviations from the 1:1 ratio of G:C/A:T were evaluated using χ^2 tests. Control over the group-wide type-I error rate was obtained using the sequential Bonferroni technique (Rice 1989) (* = significant at a group-wide 5% level, ** = significant at a group-wide 1% level, *** = significant at a group-wide 0.1% level)

[c] Sequence from Ma et al. (1993)

[d] Sequence kindly provided by Axel Janke before its publication (Janke et al. 1994)

[e] Sequence kindly provided by Kimitsuna Watanabe before its publication (Asakawa et al. 1994)

[f] Sequences from Tamura (1992)

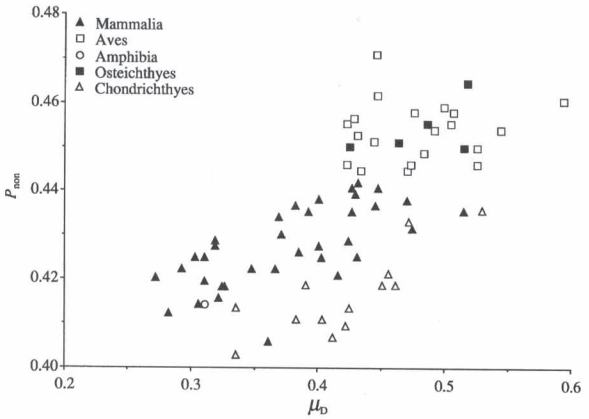

Fig. 2. The distribution of vertebrate taxa as a function of the directional mutation pressure (μ_D) and the GC% at the nonsynonymous codon sites (P_{non}). Data are from Table 3.

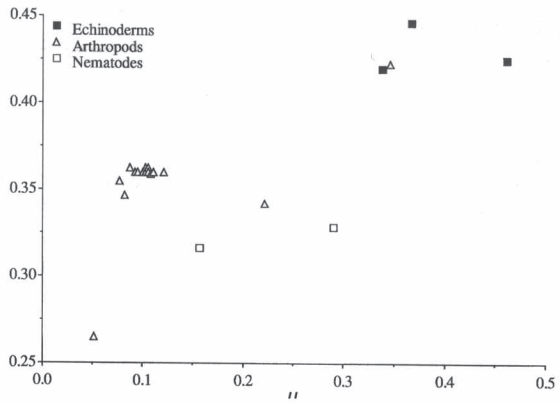

Fig. 3. The distribution of invertebrate taxa as a function of the directional mutation pressure (μ_D) and the GC% at the nonsynonymous codon sites (P_{non}). Data are from Table 3.

Ohama et al. 1987, 1989; Ohkubo et al. 1987; Ohtaka and Ishikawa 1993), protein-coding regions in the mitochondrial DNA of five metazoans (Jukes and Bhushan 1986), and nuclear genes of a variety of vertebrates (e.g., Sueoka 1988, 1992).

Covariation Between GC% at Nonsynonymous Codon Sites and the Directional Mutation Pressure

Since the GC% at a gene's nonsynonymous codon sites and the directional mutation pressure at its synonymous codon sites vary among taxa, it is possible that a functional relationship exists between these two, as suggested by previous analyses of bacterial (Jukes and Bhushan 1986; Muto and Osawa 1987) and mitochondrial genomes (Jukes and Bhushan 1986).

The data in Table 3 support the notion of a positive functional relationship between the GC% at nonsynonymous codon sites and the directional mutation pressure. The positive functional relationship is evident among vertebrates (Fig. 2) and invertebrates (Fig. 3). Among vertebrates (Fig. 2) there is good separation between the distributions of several classes; Aves and Osteichthyes overlap each other in the upper right portion of the graph but are separated from Mammalia, Amphibia, and Chondrichthyes. In relation to Aves and Osteichthyes, the latter three classes are dispersed toward the lower left corner of the graph. The distributions of the Mammalia and Amphibia overlap but are almost separated from that of the Chondrichthyes. The distributions of Chondrichthyes and Mammalia are elongated and with a positive cline, and they have much wider ranges along both axes than any of the other three classes.

Among the invertebrates (Fig. 3), there is a slight overlap between the distribution of the arthropods and those of the nematodes and echinoderms, whereas there is no overlap between distributions of the latter two. The

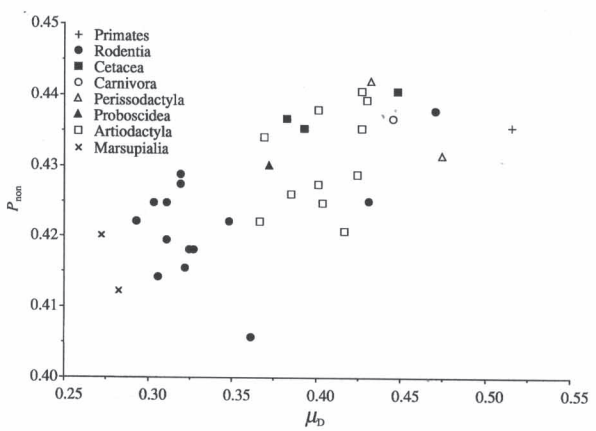

Fig. 4. The distribution of mammalian taxa as a function of the directional mutation pressure (μ_D) and the GC% at the nonsynonymous codon sites (P_{non}). Data are from Table 3.

distribution of the arthropods is much wider along both axes than those of the other two phyla.

Within Mammalia (Fig. 4) and Chondrichthyes (Fig. 5) there is also good separation between several orders. The Marsupialia, Rodentia, and Primates are well separated from the Cetacea, Carnivora, Perissodactyla, Proboscidae, and Artiodactyla (Fig. 4), whereas the latter five orders separate poorly; the Carcharhiniformes separate well from the Lamniformes and Heterodontiformes (Fig. 5), whereas the latter two groups overlap.

Although the diagrams above may be biased by the unequal representation of different taxanomic groups and by a taxonomic system which may be inconsistent with the actual phylogeny, it appears reasonable to conclude that the GC% at the nonsynonymous sites is associated positively with directional mutation pressure. Also, the clear separation among some taxonomic groups suggests that there is indeed a phylogenetic component in the distributions of the GC% at nonsynonymous codon sites and the directional mutation pressure.

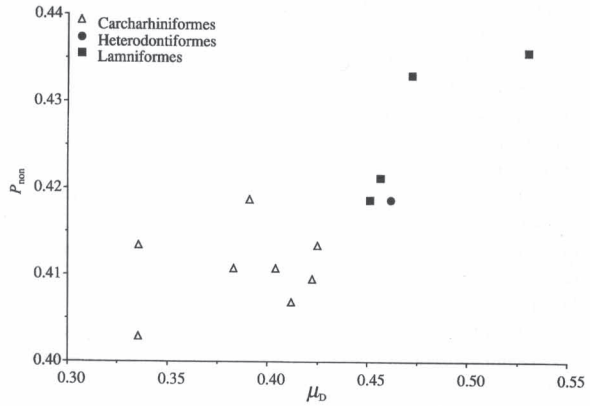

Fig. 5. The distribution of chondrichthyan taxa as a function of the directional mutation pressure (μ_D) and the GC% at the nonsynonymous codon sites (P_{non}). Data are from Table 3.

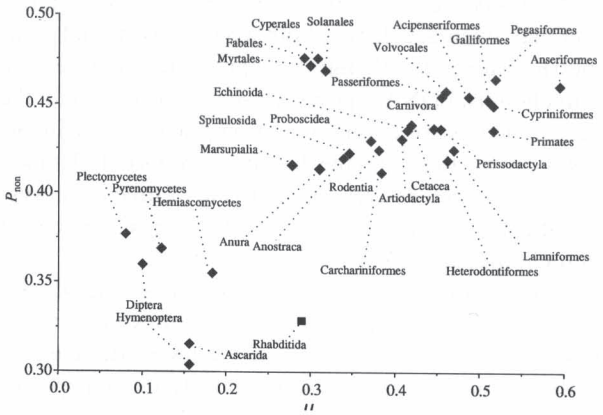

Fig. 6. The distribution of orders as a function of the directional mutation pressure (μ_D) and the GC% at the nonsynonymous codon sites (P_{non}). Arithmetic mean values for each order were obtained using a hierarchal approach. Data are from Table 3.

The Phylogenetic Component

In order to investigate the phylogenetic components among the GC% at nonsynonymous codon sites and the directional mutation pressure, while at the same time removing some of the above-mentioned bias, we calculated the arithmetic means[2] of P_{non} and μ_D for each other and then plotted these values against each other (Fig. 6). Although much of the variation from Figs. 2–5 has been removed, the GC% at nonsynonymous sites is still positively associated with the directional mutation pressure. However, it is also evident that the functional relationship between the two parameters varies among the metazoan orders.

Furthermore, the existence of a phylogenetic component is reinforced by the close grouping[3] of related orders (Fig. 6). For example, the magnoliophytic orders (Fabales, Solanales, Myrtales, and Cyperales) are distinctively separated from all other orders in the upper central portion of the diagram, whereas the insect (Diptera and Hymenoptera), nematode (Ascarida and Rhabditida), and eumycote (Hemiascomycetes, Plectomycetes, and Pyrenomycetes) orders are separated from one another, and from all other orders, in the lower left part of the diagram. Among the chordates, the chondrichthyan orders (Carchariniformes, Heterodontiformes, and Lamniformes) are located below and parallel to other chordate orders, whereas avian (Anseriformes, Passeriformes, and Galliformes) and osteichthyan (Acipenseriformes, Cypriniformes, and Pegasiformes) orders overlap one another in the upper right portion of the diagram. The six avian and osteichthyan orders are well separated from mammalian orders (Marsupialia, Proboscidea, Rodentia, Artiodactyla, Cetacea, Carnivora, Perissodactyla, and Primates), these being dispersed in a narrow ribbon parallel to the chondrichtyan orders. Embedded within the "range" of the chordate orders are brachiopod (Anostraca), echinoderm (Echinoida, Spinulosida), and amphibian (Anura) orders, whereas a chlorophytic order (Volvocales) is embedded within the "range" of avian orders.

Implications of Directional Mutation Pressure

In this analysis we present evidence on the taxonomic prevalence of directional mutation pressure in mitochondrial genes. The results support previous suggestions on the subject (Jukes and Bhushan 1986; Crozier and Crozier 1993; Jermiin and Crozier 1994). Because the GC% at nonsynonymous codon sites is correlated positively with the magnitude and direction of the mutational pressure, we argue that directional mutation pressure is responsible to a large extent for the amino acid composition of proteins encoded by the mitochondrial DNA. Indeed, our results suggest that A + T pressure has been the primary force underlying the evolution of many cytochrome *b* genes.

The abundance of cytochrome *b* genes that are subject to A + T pressure is difficult to explain because we focus on symmetrical directional mutation pressure and, thus, omit additional information about the occurrence of asymmetrical directional mutation pressure (Asakawa et al. 1991). However, the highly asymmetrical replication of mitochondrial DNA, which is reported to occur in many organisms (Clayton 1992), leaves the H strand exposed as a single strand for much longer than the L strand and effectively prevents double-stranded DNA editing by DNA endonucleases (Asakawa et al.

[2] The arithmetic means were calculated using a hierarchal approach, i.e., within species, within genus, and within family. This approach reduces, although it does not eliminate, the bias induced by unequal representation of taxa.

[3] A nested analysis of variance is not appropriate with the present data (Table 3). While the Chordata are well represented with different classes which each contain several orders, the other phyla are only poorly represented, and statistical confirmation of the phylogenetic component will have to await publication of more DNA sequence data and establishment of a well-founded phylogeny.

1991; Osawa et al. 1992). This probably facilitates, and may even accelerate, both asymmetrical and symmetrical directional mutation pressure. A defective mitochondrial DNA γ-polymerase may also explain the occurrence of directional mutation pressure. Misincorporation by this enzyme has been reported (Kunkel 1985), but whether over a long period of time it can systematically increase or decrease the G + C content of mitochondrial DNA is as yet unknown.

The fact that the variation in GC% has a significant phylogenetic component means that synonymous codon positions may only be used in phylogenetic analysis with caution. On the one hand, closely related species seem to be subject to similar A + T pressures, and, therefore, the bias in codon usage may be used in itself as an informative character in phylogeny. On the other hand, when two DNA sequences possess the same nucleotide at a homologous position, the similarity may be due to convergence rather than to shared ancestry. Therefore, our suggestion is to use overall similarity in codon usage as a qualitative support for the clustering of data but to exclude the sites that are affected most by directional mutation pressure from the detailed phylogenetic analysis.

Note Added at Proof The authors apologize for the misspelling of *Tetraponera rufoniger* [sic] in Jermiin and Crozier (1994). The correct spelling is *Tetraponera rufonigra.*

Acknowledgments. This paper constitutes a part of the research done by L.S.J. in partial fulfillment of a Ph.D. degree at La Trobe University. We are grateful to Ú. Árnason, A. Janke, K. Tamura, and K. Watanabe for providing DNA sequences which were either unpublished or not available through GenBank; to K. Watanabe, J.R. Kornegay, and T.D. Kocher for sending manuscripts before publication; to C.B. Schedvin and M.B. Schedvin for allowing us to use the computer on which the computer program DMP was created; and to A. Eyre-Walker, T.H. Jukes, R.M. Kusmierski, S. Osawa, D.R. Rowe, E.R. Smith, and N. Sueoka for valuable comments on the manuscript. This study was financially supported by the Danish Research Academy (L.S.J.), La Trobe University's Postgraduate Scholarship (L.S.J.), the Australian Government's Overseas Postgraduate Research Scholarship (L.S.J.), a CRA–La Trobe University Distinguished Visitor Fellowship (D.G.), the Australian Research Council (R.H.C.), La Trobe University (R.H.C.), and the Ian Potter Foundation (R.H.C.).

References

Asakawa S, Kumuzawa Y, Araki T, Himeno H, Miura K-I, Watanabe K (1991) Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. J Mol Evol 32:511–520

Asawaka S, Himeno H, Miura K-I, Watanabe K (1994) Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics (submitted)

Belozersky AN, Spirin AS (1958) A correlation between the compositions of deoxyribonucleic and ribonucleic acids. Nature 182:111–112

Bernardi G (1989) The isochore organization of the human genome. Annu Rev Genet 23:637–661

Bernardi G, Olofsson B, Filipski J, Zerial M, Salanis J, Cuny G, Meunier-Rotival M, Rodier F (1985) The mosaic genome of warm-blooded vertebrates. Science 228:953–958

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

Clayton DA (1992) Transcription and replication of animal mitochondrial DNAs. Int Rev Cytol 141:217–323

Collins DW, Jukes TH (1993) Relationship between G + G in silent sites of codons and amino acid composition of human proteins. J Mol Evol 36:201–213

Cox EC, Yanofsky C (1967) Altered base ratios in the DNA of an *Echerichia coli* mutator strain. Proc Natl Acad Sci USA 58:1895–1902

Crozier RH, Crozier YC (1992) The cytochrome b and ATPase genes of honeybee mitochondrial DNA. Mol Biol Evol 9:474–482

Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera:* complete sequence and genome organization. Genetics 113:97–117

di Rago J-P, Netter P, Slonimski PP (1990) Pseudo-wild type revertants from inactive apocytochrome *b* mutants as a tool for the analysis of the structure/function relationships of the mitochondrial ubiquinol-cytochrome *c* reductase of *Saccharomyces cerevisiae.* J Biol Chem 265:3332–3339

D'Onofrio GD, Mouchiroud D, Aïssani 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

Freese E (1962) On the evolution of base composition of DNA. J Theor Biol 3:82–101

Hatefi Y (1985) The mitochondrial electron transport and oxidative phosphorylation system. Annu Rev Biochem 54:1015–1069

Helm-Bychowski K, Cracraft J (1993) Recovering phylogenetic signal from DNA sequences: relationship within the Corvine assemblage (Class Aves) as inferred from complete sequences of the mtDNA cytochrome b gene. Mol Biol Evol 10:1196–1214

Hemmingsen SM, Woolford C, van der Vies SM, Tilly K, Bennis DT, Georgopoulos C, Hendrix RW, Ellis RJ (1989) Homologous plant and bacterial proteins chaperone oligomeric protein assembly. Nature 333:330–334

Hori H, Osawa S (1987) Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol Biol Evol 4:445–472

Howell N (1989) Evolutionary conservation of protein regions in the protonmotive cytochrome *b* and their possible roles in redox catalysis. J Mol Evol 29:157–169

Howell N, Gilbert K (1988) Mutational analysis of the mouse mitochondrial cytochrome *b* gene. J Mol Biol 203:607–618

Irwin DA, Kocher TD, Wilson AC (1991) Evolution of cytochrome *b* gene of mammals. J Mol Evol 32:128–144

Janke A, Feldmaier-Fuchs G, Thomas WK, von Haeseler A, Pääbo S (1994) The marsupial mitochondrial genome and the evolution of placental mammals. Genetics 137:243–256

Jermiin LS, Crozier RH (1994) The cytochrome *b* region in the mitochondrial DNA of the ant *Tetraponera rufoniger:* sequence divergence in hymenoptera may be associated with nucleotide content. J Mol Evol 38:282–294

Jukes TH, Bhushan V (1986) Silent nucleotide substitutions and G + C content of some mitochondrial and bacterial genes. J Mol Evol 24:39–44

Kagawa Y, Nojima H, Nukiwa N, Ishizuka M, Nakajima T, Yasuhara T, Tanaka T, Oshima T (1984) High guanine plus cytosine content in third letter of codons of an extreme thermophile. J Biol Chem 259:2956–2960

Kornegay JR, Kocher TD, Williams LA, Wilson AC (1993) Pathways of lysosyme evolution inferred from the sequences of cytochrome *b* in birds. J Mol Evol 37:367–379

Kushino A, Shimizu M, Tomita K-I (1987) Molecular cloning and se-

quence determination of the *tuf* gene coding for the elongation factor Tu of *Thermus thermophilus* HB8. Eur J Biochem 170:93–98

Kunkel TA (1985) The mutational specificity of DNA polymerase-α and -γ during *in vitro* DNA synthesis. J Biol Chem 135:327–351

Kusmierski R, Borgia G, Crozier RH, Chan BHY (1993) Molecular information on bower-bird phylogeny and the evolution of exaggerated male characteristics. J Evol Biol 6:737–752

Lee KY, Wahl R, Barbu E (1956) Contenu en bases puriques et pyrimidiques des acids desoxyribonucleiques des bacteries. Ann Inst Pasteur 91:212–224

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinaur Associates, Sunderland, MA, pp 1–284

Ma D-P, Zharkikh A, Graur D, VandeBerg JL, Li W-H (1993) Structure and evolution of opussum, guinea pig, and porcupine cytochrome *b* genes. J Mol Evol 36:327–334

Martin AP, Palumbi SR (1993) Protein evolution in different cellular environments: cytochrome *b* in sharks and mammals. Mol Biol Evol 10:873–891

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

Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci USA 84: 166–169

Ohama T, Yamao F, Muto A, Osawa S (1987) Organization and codon usage of the streptomycin operon in *Micrococcus luteus,* a bacterium with a high genomic G + C content. J Bacteriol 169: 7440–4777

Ohama T, Yamao F, Muto A, Osawa S (1989) Sprectinomycin operon of *Micrococcus luteus*: evolutionary implications of organization and novel codon usage. J Mol Evol 29:381–395

Ohkubo S, Muto A, Kawauchi Y, Yamao F, Osawa S (1987) The ribosomal protein gene cluster of *Mycoplasma capricolum.* Mol Gen Genet 210:314–322

Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR (1992) The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum.* Genetics 130:471–498

Osawa S, Jukes TH, Watanabe K, Muto A (1992) Recent evidence for evolution of the genetic code. Microbiol Rev 56:229–264

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43: 223–225

Rolfe R, Meselson M (1959) The relative homogeneity of microbial DNA. Proc Natl Acad Sci USA 45:1039–1043

Sueoka N (1959) A statistical analysis of deoxyribonucleic acid distribution in density gradient centrifugation. Proc Natl Acad Sci USA 45:1480–1490

Sueoka N (1961a) Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. Proc Natl Acad Sci USA 47:1141–1149

Sueoka N (1961b) Compositional correlation between deoxyribonucleic acid and protein. Cold Spring Harbor Symp Quant Biol 26: 35–43

Sueoka N (1962) On the genetic basis of variation and heterogeneity of DNA base composition. Proc Natl Acad Sci USA 48:582–592

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

Sueoka N (1992) Directional mutation pressure, selective constraints, and genetic equilibria. J Mol Evol 34:95–114

Sueoka N, Marmur J, Doty P (1959) Heterogeneity in deoxyribonucleic acids. II. Dependence of the density of deoxyribonucleic acids on the guanine-cytocine. Nature 183:1427–1431

Tamura K (1992) The rate and pattern of nucleotide substitution in *Drosophila* mtDNA. Mol Biol Evol 9:814–825

Tron T, Crimi M, Colson A-M, Degli Esposti M (1991) Structure/function relationships in mitochondrial cytochrome b revealed by the kinetic and circular dichroic properties of two yeast inhibitor-resistant mutants. Eur J Biochem 199:753–760

Wada A, Suyama A (1986) Local stability of DNA and RNA secondary structure and its relation to biological functions. Prog Biophys Mol Biol 47:113–157