

Scientific Correspondence

Dear Editor,

Moriyama and Powell (1) have recently reported a negative correlation between gene length and synonymous codon usage bias in *Drosophila melanogaster*. They suggest that selection may be acting on this genome to reduce the size of highly expressed genes: i.e. weaker selective constraints allow longer genes with lower codon bias. We have repeated their analysis with data from a gene sample that contains only genes with a well characterized intron/exon structure (2). Noteworthy, the correlation between gene length and codon bias is positive ($r = 0.459$, $P < 0.01$, $n = 31$) in the intron-lacking gene set, but negative ($r = –0.277$, $P < 0.01$, $n = 86$) in the split gene set. The length of reading frames is related to DNA GC content (3–5): gene length in prokaryotes (*v.g.*, *Escherichia coli*, *Bacillus subtilis* and *Haemophilus influenzae*), and exon length in mammalian genes are positively related to GC content (5,6). However, mammalian long genes are mostly GC-poor, because the average number of exons in GC-poor genes is greater than that in GC-rich ones (5,6). The negative correlation reported in (1) may be analogous to what occurs in mammals, since *D.melanogaster* codon bias is correlated to gene GC content ($r = 0.651$, $P < 0.001$, $n = 117$). Furthermore, gene GC content correlates negatively to the number of exons (Spearman's $r = –0.335$, $P < 0.01$, $n = 117$), and a positive correlation exists between exon length and exon GC content ($r = 0.215$, $P < 0.01$, $n = 263$). Thus, the situation in *D.melanogaster* is reminiscent to what occurs in mammals, where no selective effects on codon bias are known. In summary, the negative correlation reported (1) might be a widespread property among eukaryotes, where gene structure is associated to DNA compositional variation.

Antonio Marín*, Francisco González, Gabriel Gutiérrez and José L. Oliver[1]
Departamento de Genética
Universidad de Sevilla
Apartado 1095
E-41080 Sevilla
Spain
[1]Departamento de Genética
Universidad de Granada
Campus de Fuentenueva
E-18071 Granada
Spain
*Tel: +34 95 45 57 113; Fax: +34 95 45 57 104;
Email: anmarin@cica.es
July 20, 1998

REFERENCES

1 Moriyama,E.N. and Powell,J.R. (1998) *Nucleic Acids Res.*, **26**, 3188–3193.
2 Kliman,R.M. and Eyre-Walker,A. (1998) *J. Mol. Evol.*, **46**, 534–541.
3 Merino,E., Balbás,P., Puente,J.L. and Bolivar,F. (1994) *Nucleic Acids Res.*, **22**, 1903–1908.
4 Guigó,R. and Fickett,J.W. (1996) *J. Mol. Biol.*, **253**, 51–60.
5 Oliver,J.L. and Marín,A. (1996) *J. Mol. Evol.*, **43**, 216–223.
6 Duret,L., Mouhiroud,D. and Gautier,C. (1995) *J. Mol. Evol.*, **40**, 308–317.

Dear Editor,

Marín *et al*. rightly point out that the issue is not simply between prokaryotes and eukaryotes, but between non-segmented coding sequences (prokaryote genes and eukaryote exons) and segmented genes with introns. Non-segmented genes in bacteria, vertebrates and *Drosophila melanogaster* have a positive correlation between length and codon usage bias (or GC content), whereas the correlation is opposite in segmented genes (1–3). Both selective constraints on translational accuracy (1) and stop codon probability depending on the GC content (2) can generate a positive correlation. However, how the correlation is created is totally different between the two models. In the former model, the length itself (actually the product of length and expression level; 1) determines the degree of codon usage bias. This is particularly clear among genes with equal amount of expression (1,3). There has been evidence for selective constraints acting on synonymous codon usage in order to maintain translational accuracy both in bacteria and *Drosophila* genes (1,4), whereas it is not known in vertebrate genes. In the latter model, GC content affects the length of non-segmented genes and exons. Therefore, it is possible that the similar correlation is realized by different mechanisms. Note that the gene/exon length could also be maintained by functional constraints. With sufficiently strong selective constraints, the latter model cannot be effective. As discussed in our paper (3), the negative correlation found in segmented genes in *Drosophila* and yeast appears to require another source of selection. Interestingly, the GC-rich 'house-keeping' subgenome in vertebrates also has more shorter and more compact segmented genes than GC-poor region (2,5). Whether gene structure is directly associated with base composition variation, is the consequence of some selective forces, or is caused by both, remains an open question.

Etsuko N. Moriyama* and Jeffrey R. Powell
Department of Ecology and Evolutionary Biology
Yale University
165 Prospect Street
New Haven, CT 06520-8106
USA
*Tel: +1 203 432 3886; Fax: +1 203 432 6066;
Email: moriyama@peaplant.biology.yale.edu
August 10, 1998

REFERENCES

1 Eyre-Walker,A. (1996) *Mol. Biol. Evol.*, **13**, 864–672.
2 Oliver,J.L. and Marín,A. (1996) *J. Mol. Evol.*, **43**, 216–223.
3 Moriyama,E.N. and Powell,J.R. (1998) *Nucleic Acids Res.*, **26**, 3188–3193.
4 Akashi,H. (1994) *Genetics*, **136**, 927–935.
5 Duret,L., Mouchiroud,D. and Gautier,C. (1995) *J. Mol. Evol.*, **40**, 308–317.