Genet. Res., Camb. (1996), **68**, *pp.* 101–108 *With* 1 *text-figure*

Reduced variation at *concertina*, a heterochromatic locus in *Drosophila*

MARTA L. WAYNE* AND MARTIN KREITMAN
Department of Ecology and Evolution, 1101 E. 57th Street, University of Chicago, Chicago, IL 60637, USA

(*Received 23 October 1995 and in revised form 31 May 1996*)

Summary

In *Drosophila melanogaster* and closely related species, polymorphism has been shown to be reduced at loci located in regions of low recombination on the X chromosome and on the fourth chromosome, which does not normally recombine. This positive correlation between nucleotide polymorphism level and recombination rate is not predicted by standard neutral theory and therefore must result from natural selection and genetic hitchhiking along the chromosomes. We report here the near-complete absence of variation at *concertina* (*cta*), a locus located in the β-heterochromatic base of chromosome 2L, a region of strongly reduced recombination. A 1·2 kilobase region containing coding regions and introns was sequenced from each of nine lines of *D. melanogaster* and nine lines of *D. simulans* representing worldwide collections. Variation is significantly reduced in *cta* in both species compared with other available loci on the same chromosome. Two analyses of background selection demonstrate that the reduction in variation at *cta*, considered in combination with other loci on chromosome 2L or alone, is consistent with the background selection model.

1. Introduction

Studies of genetic variation in *Drosophila melanogaster* reveal a positive correlation between levels of DNA sequence polymorphism and the amount of recombination (Begun & Aquadro, 1992; reviewed in Aquadro *et al.*, 1994; Kreitman & Wayne, 1994). The level of within-species variation is not correlated with between-species divergence, indicating that selective constraint cannot be the cause of this relationship (Begun & Aquadro, 1992). Moreover, the correlation of reduced variation with low recombination is not limited to *D. melanogaster*. In three additional species in the *Drosophila* genus – *simulans*, *mauritiana*, and *ananassae* (Miyashita & Langley, 1988; Aguadé *et al.*, 1989, 1992; Stephan & Langley, 1989; Begun & Aquadro, 1991; Berry *et al.*, 1991; Martín-Campos *et al.*, 1992; Stephan & Mitchell, 1992; Hey & Kliman, 1993) – variation is reduced at loci in regions which have lower recombination compared with loci in regions of higher recombination. Distinct variants are fixed for each locus in each species, indicating that separate events have caused the loss of variation in the individual species. These data have led to the search for general selective mechanisms for the reduction of variation in regions of low recombination.

Selection against deleterious mutations can cause a substantial reduction in neutral polymorphism levels at linked sites in *D. melanogaster* (Charlesworth *et al.*, 1993). With a deleterious mutation rate of approximately one per haploid genome per generation, the reduction in neutral polymorphism is expected to be many-fold stronger at the base of the autosomes than at more distally located regions, where recombination is higher (Hudson & Kaplan, 1994, 1995; Charlesworth, 1996; Nordborg *et al.*, 1996).

Positive selection and genetic hitchhiking can also lead to the loss of variation, particularly in areas of low recombination (Maynard Smith & Haigh, 1974; Kaplan *et al.*, 1989; Stephan *et al.*, 1992). For certain regions, such as the fourth chromosome and the base of the X, background selection may not suffice to account for the observed reductions in variation, suggesting the additional presence of positive selection (Charlesworth *et al.*, 1993). Charlesworth (1996), however, has recently considered the additional contribution of transposable elements to background selection. With their higher concentrations in regions

* Corresponding author. Present address: Department of Genetics, Box 7614, North Carolina State University, Raleigh, NC 27695, USA. Fax: +1(919) 515-3355. e-mail: mlwayne@esssjp.ncsu.edu.

of low recombination and their weak deleterious effects on fitness, the data for the tip of the X and fourth chromosome in *D. melanogaster* can now also be potentially explained by background selection. Supporting this view, the frequency spectrum of nucleotide site variation in regions of low recombination does not conform to that predicted by a simple model of recurrent selective sweeps (Braverman *et al.*, 1995).

We have chosen to study sequence polymorphism at *concertina* (*cta*), a locus cytologically (40f(h30)) and genetically (2–54·8) localized to the most proximal region of chromosome 2L. *cta* is the alpha subunit of a G-protein (Parks & Wieschaus, 1991), a member of a large, highly conserved family of signal transduction proteins, and is required for the correct timing and location of gastrulation. Although the low-recombination centromeric regions of the autosomes comprise approximately 15% of the *Drosophila* genome (Charlesworth *et al.*, 1993), and genetic variation in these regions has been estimated at the restriction fragment length polymorphism (RFLP) level (Aquadro *et al.*, 1992, 1994), they have not previously been characterized for genetic variation at the sequence level, despite the large reduction in polymorphism predicted by both the background selection and selective sweep models.

2. Materials and methods

(i) *Fly lines and DNA preparation*

For *D. melanogaster*, a worldwide collection of nine lines isochromosomal for chromosome 2 was used. The lines originated from five population samples and were selected originally for their *Adh* allozyme phenotype (described in Kreitman, 1983; also see Kreitman & Hudson, 1991). They are *AfricaFast* (*AfF*), *AfricaSlow* (*AfS*), *FrenchFast* (*FrF*), *FrenchSlow* (*FrS*), *FloridaFast* (*FlF*), *Florida1Slow* (*Fl1S*), *Florida2Slow* (*Fl2S*), *JapanSlow* (*JpS*) and *WashingtonFast* (*WaF*). For *D. simulans*, a worldwide collection of isofemale lines was used: two lines from Mpala Ranch, Kenya; two lines from northern Australia; two lines from Valparaiso, Indiana; two lines from Trinidad; and one line from Lantana, Florida. Genomic DNA was prepared using Chelex resin from single flies (Walsh *et al.*, 1991).

(ii) *Polymerase chain reaction (PCR)*

Amplifications were carried out for 30 cycles in 100 μl volumes using *Taq* polymerase (Boehringer Mannheim) and 2 mM [Mg^{2+}]. Samples were initially heated to 95 °C for 2 min before addition of the enzyme. Reaction conditions were 25 s denaturation at 95 °C, 1 min annealing at 51 °C, and 1 min 30 s extension at 72 °C. For the 5′ region (exon 1), a 380 bp fragment was amplified from genomic DNA using 20mer oligonucleotide primers at positions 62 (5′ base of 'plus' primer, GGGAGTGTTTGACGTATTGG) to 441 (3′ base of 'minus' primer, CGCAGCCTTACTAGATATG). The numbering corresponds to the cDNA sequence of Parks & Wieschaus (1991). The amplification reaction product was purified using Promega Magic PCR Preps, and the yield quantified against a standard by computer image analysis (NIH Image 1·47) after electrophoresis on TRIS-borate EDTA agarose mini-gels and ethidium bromide staining. For the rest of the sequence, a 1·3 kb fragment was amplified from genomic DNA using 20mer oligonucleotide primers from position 625 (5′ base of the 'plus' primer, GTTGCTGGAATACCAAAGTG) to position 1690 (3′ base of the 'minus' primer, CCGAAGGATTTCACTGTAGC). Contents of the reaction tube were ethanol precipitated, dried, resuspended, and separated on a 1·0% agarose mini-gel. The band was then purified (Weichenhan, 1991) and resuspended in approximately 100 μl of distilled water. Two microlitres of this purified DNA served as template for a reamplification using the same primers, or various combinations of internal primers. Reamplifications were carried out in 100 μl volumes using Boehringer Mannheim enzyme under conditions described above for 20 cycles using 2 mM [Mg^{2+}]. The PCR products were purified using Promega Magic PCR Preps and yields were determined as described above.

(iii) *Sequencing*

Sequencing of templates was accomplished using cycle sequencing and dye-terminators (Applied Biosystems). Two hundred femtomoles of template were used with 15 pmol of primer. Cycle sequencing proceeded for 25 cycles under the following conditions: denaturation, 15 s at 95 °C; annealing, 30 s at 50 °C; extension, 1 min at 60 °C. Purified products were electrophoresed on an ABI 373A automated sequencer. Sequence read lengths averaged approximately 300 bases, ranging from 150 to 450 bases. All sequences were determined for both strands.

(iv) *Analysis*

Sequences were aligned manually or with the assistance of the SeqEd program version 1.0 supplied by ABI. The effective number of silent sites was determined as described in Kreitman & Hudson (1991).

3. Results

We sequenced 1178 bases in nine lines of *D. simulans* and nine lines of *D. melanogaster*. The two sequenced regions include codons 10–64 and 174–440 of the published cDNA sequence (966 bp; Parks & Wieschaus, 1991). A single polymorphism was found in one line of *D. simulans*; the *D. melanogaster* lines

https://doi.org/10.1017/S001667230003398X Published online by Cambridge University Press

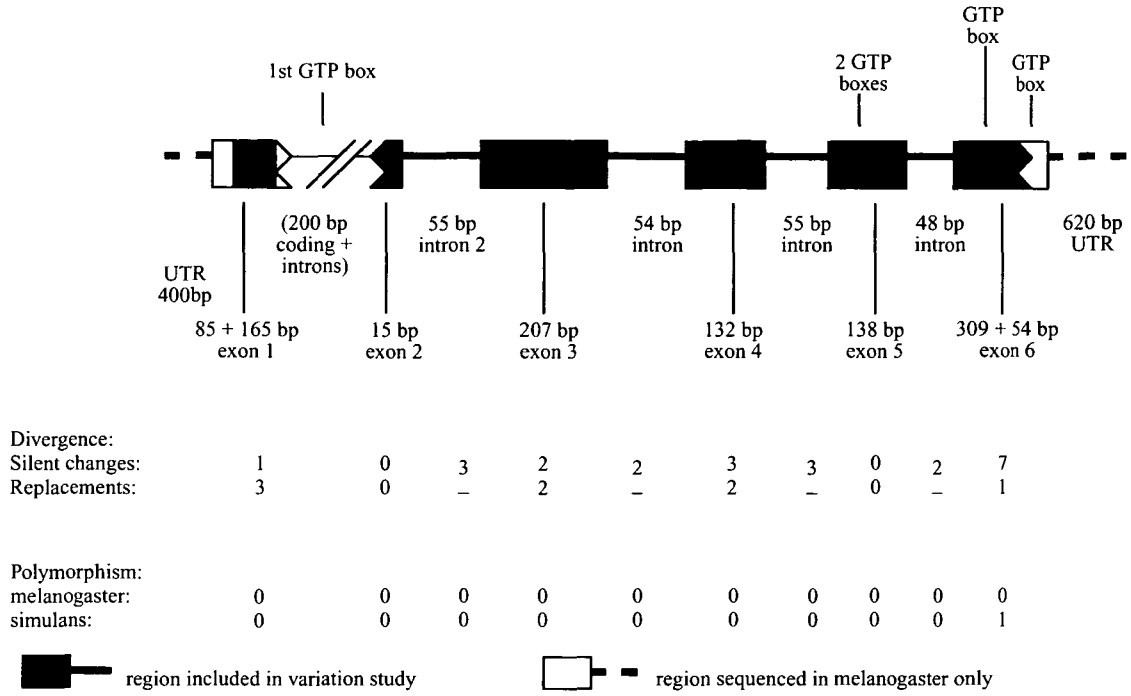

Fig. 1. A cartoon of the *concertina* locus indicating the locations of the introns and divergent and polymorphic sites.

were invariant. Comparing our genomic DNA sequences with the published cDNA sequence revealed the presence of four introns (212 bp (209 bp in *D. simulans*)), all typically present in G-proteins (Wilkie & Yokoyama, 1993). We were unable to amplify the region spanning codons 64 and 174, suggesting the presence of one or more additional large introns. A representation of the locus with intron positions is shown in Fig. 1.

An interspecies comparison revealed 13 synonymous and eight replacement substitutions in the coding regions, and 10 silent substitutions in introns. The sequences also differ by one insertion/deletion, with *D. melanogaster* having three bases starting at position 14 of intron 2 which are absent in *D. simulans*. The effective number of silent sites in introns and redundant positions of codons, calculated according to Kreitman & Hudson (1991), is 196 bp for introns and 216 bp for exons for *D. melanogaster* and 193 bp and 216 bp, respectively, for *D. simulans*. Silent divergence, calculated by dividing the number of silent changes by the effective number of silent sites, is 5·18 % for exons and 5·58 % for introns. Correcting for multiple hits (Nei & Gojobori, 1986), total divergence (k_s) is estimated to be 5·99 %.

(i) *Tests of neutrality*

To test whether the absence of silent polymorphism in *cta* is compatible with neutral molecular evolution, we used the HKA test (Hudson *et al.*, 1987) to compare *cta* polymorphism and divergence levels with those of four other loci on the left arm of chromosome 2: *Alcohol dehydrogenase* (*Adh*) (Kreitman & Hudson 1991), *topoisomerase 2* (*top2*) (Palopoli & Wu 1996), *refractaire* (*ref(2)P*) (Wayne *et al.*, 1996) and *sn-glycerol-3-phosphate dehydrogenase* (*Gpdh*) (Takano *et al.*, 1993), all of which have comparable worldwide sample data sets. For these tests, only synonymous (and/or silent) changes were used, as these variants are thought to approximate neutrality. Results of the HKA tests are presented in Table 1. The variation is significantly reduced in *cta* relative to *Adh* ($X^2 = 4{\cdot}66$, $P < 0{\cdot}05$) and *Gpdh* ($X^2 = 5{\cdot}70$, $P < 0{\cdot}05$) in *D. melanogaster*, and relative to *Adh* ($X^2 = 6{\cdot}31$, $P < 0{\cdot}01$) and *top2* ($X^2 = 4{\cdot}86$, $P < 0{\cdot}05$) in *D. simulans*. A fifth comparison, *cta* versus *ref(2)P* in *D. melanogaster*, is marginally significant ($X^2 = 4{\cdot}24$, $P < 0{\cdot}07$). These results clearly demonstrate that the level of variation in *cta* in both species is low compared with that of other loci on the same chromosome arm. It is also reduced compared with *top2* and *ref(2)P*, both of which are located in the proximal region of the euchromatin less than 1 cM from *cta*.

(ii) *Tests of background selection*

Having rejected the completely neutral model, we now present two methods for testing a background selection model. The first approach involves a straightforward modification of the HKA test of neutrality (Hudson *et al.*, 1987). Under the infinite sites model of neutral evolution, the number of segregating sites, S, is a function of the sample size, n, and the neutral parameter, θ:

$$E(S) = \theta \sum_{j=1}^{n-1} 1/j. \quad (1)$$

https://doi.org/10.1017/S001667230003398X Published online by Cambridge University Press

Table 1. *HKA tests comparing* concertina *with other loci on chromosome 2L*

Locus	Species	Sample size	Effective no. of sites	Observed S^a		$\hat{\theta}$	χ^2 (vs *cta*)
				Within sp.	Between sp.		
cta	*mel.*	9	412	0	23	0·0	—
	sim.	9	409	1		0·0009	
ref(2)P	*mel.*	10	1308	23	82	0·006	3·24†
top2	*mel.*	9	446	5	43	0·0041	1·85
	sim.	5	446	25		0·0269	4·86*
Adh	*mel.*	11	1249	37	76	0·0101	4·66*
	sim.	5	1249	64		0·025	6·31**
Gpdh	*mel.*	6	4500	108	220	0·0105	5·70*

mel., *melanogaster*; *sim.*, *simulans*.
[a] S, number of segregating sites.
† $P < 0·07$.
* $P < 0·05$; ** $P < 0·01$.

Table 2. *HKA test of background selection model of Charlesworth (1996)*

Locus	Location	π_k/π_0[a]	Contribution by locus to total HKA test statistic, χ^2		
			Poly.	Div.	Total
cta	54·8 40 het.	0·02	0·4583	0·0114	0·4697
ref(2)P	54·0 37E2-F4	0·08	4·6270	1·3937	6·0207
top2	54·0 37D2-E2	0·08	0·0036	0·0007	0·0043
Adh	50·1 35B3	0·40	0·0154	0·0179	0·0333
Gpdh	17·8–20·5 25A5-F5	0·67	0·0102	0·0879	0·0981
Total					6·6261

[a] Charlesworth (1996).

This equation can be rearranged to produce an unbiased estimator of θ:

$$\hat{\theta} = S\left(\sum_{j=1}^{n-1} 1/j\right)^{-1}. \quad (2)$$

Nucleotide diversity, π, the heterozygosity per nucleotide site, has an equally simple relationship with the neutral parameter:

$$E(\pi) = \theta. \quad (3)$$

In general, the number of segregating sites, S, provides a better estimator of the neutral parameter than π (Hudson, 1992). For this reason, $\hat{\theta}$ estimated from S has also been used as a proxy for π (Charlesworth, 1996).

Under background selection, nucleotide diversity is reduced by a factor π/π_0, where π and π_0 are the expected nucleotide diversity when background selection is present or absent, respectively. Approximate formulae for π/π_0 have been derived under background selection (Charlesworth *et al.*, 1993; Hudson & Kaplan, 1994, 1995; Charlesworth, 1994; Nordborg *et al.*, 1996). Under the assumption that neutral variation conforms to the equilibrium neutral model with effective population size reduced by this factor, eqn (1) can be modified to yield the expected number of segregating neutral sites under background selection:

$$E(S) = (\pi/\pi_0)\theta \sum_{j=1}^{n-1} 1/j. \quad (4)$$

In this manner a test of background selection can be formulated by incorporating the expected reduction in polymorphism, π/π_0, into appropriate terms in the system of equations describing polymorphism in the HKA test. Background selection is not expected to influence the nucleotide divergence under a strictly neutral model.

https://doi.org/10.1017/S001667230003398X Published online by Cambridge University Press

Charlesworth (1996) has constructed an explicit model of background selection for *D. melanogaster* which incorporates the varying recombination rates along the chromosomes, the contributions of transposable elements to background selection, and allows selection intensity against the deleterious mutations to vary. Overall he finds a reasonable correlation between the observed and predicted values of nucleotide variation, including the proximal part of chromosome 2L. The expected reduction in polymorphism under this model is given in Table 2 for *cta*, *ref(2)P*, *top2*, *Adh* and *Gpdh*. The results of a five-locus background selection test are given in Table 2, with the individual contribution of within- and between-species polymorphism at each locus to the test statistic. The background selection model cannot be rejected by the modified five-locus combined HKA test ($\chi^2 = 6{\cdot}63$, $P > 0{\cdot}2$).

One may generate a 'best guess' estimate of the expected value for θ_{cta} under background selection by considering the expected reduction in polymorphism relative to *Adh*. The predicted ratio under the Charlesworth (1996) model is $\pi_{cta}:\pi_{Adh} = 0{\cdot}02/0{\cdot}4 = 0{\cdot}05$. Given that $\hat{\theta}_{Adh}$ in *D. melanogaster* is 0·0101, the best guess for $\theta_{cta} = 0{\cdot}05(\hat{\theta}_{Adh}) = 0{\cdot}05 \times 0{\cdot}0101 = 0{\cdot}0005$. For a sample size of 412 silent sites and nine lines, the probability of observing zero segregating sites for the best guess value of θ_{cta} under background selection is 0·58 (R. R. Hudson, personal communication). This probability will underestimate the true probability if *Adh* is under balancing selection and there has been a local elevation of silent variation in the locus (Kreitman & Hudson, 1991). The same conclusion is reached using the *D. simulans* data: $\theta_{cta} = 0{\cdot}05(\hat{\theta}_{Adh}) = 0{\cdot}05 \times 0{\cdot}0245 = 0{\cdot}0012$, which given a sample size of 409 silent sites and nine lines yields a probability of 0·32 (R. R. Hudson, personal communication) that we should observe one segregating site. For both species, the observed number of segregating sites is compatible with the Charlesworth (1996) background selection model.

4. Discussion

concertina is virtually devoid of variation: no polymorphisms in 1178 bp in nine lines of *D. melanogaster*, and only one polymorphic site in one of nine lines of *D. simulans*. Silent divergence, which is 5·58%, is not correspondingly low. The neutral expectation is that intraspecific and interspecific variation will be correlated in the absence of selection. The significant pairwise HKA tests indicate that some form of selection must be acting to reduce variation in the centromeric region.

The high rate of mutation to deleterious alleles in *Drosophila* (and many other organisms), coupled with low recombination rates in some parts of the genome, virtually ensures that background selection is contributing to the observed reduction in variation. Our results are consistent with background selection as a selective force shaping the levels of variation on chromosome 2L. However, the background selection model we have chosen to test, while it may be the most sophisticated one available, nevertheless requires a number of approximations and guesses of parameter values and distribution functions in specifying the model. Further, our analyses, while compatible with the Charlesworth (1996) model of background selection as the sole selective force governing reduced variation, do not eliminate selective sweeps as a contributor to reduced variation. Indeed, the observed reduction in polymorphism at *cta* is substantially larger than that predicted under the Charlesworth model without transposable elements, and remains somewhat larger than that predicted by the model even when they are included, although it is not statistically incompatible with the model.

Three of the four loci against which *cta* polymorphism was compared – *ref(2)P*, *Adh* and *Gpdh* – are thought to have higher than average levels of silent variation as a consequence of balancing selection or positive selection. In addition, the estimates of polymorphism at *top2* in *D. melanogaster* ($\hat{\theta} = 0{\cdot}0041$) and *D. simulans* ($\hat{\theta} = 0{\cdot}0269$) are substantially different, again indicating the possibility of selection acting at or near this locus in one or both species. Therefore, the apparent fit of the data to the Charlesworth model must be viewed with caution: it is likely that additional selective forces have been operating at or near many of the loci included in this analysis.

Our statistical test of background selection is based on the most conservative tests of neutrality, which assume no recombination. For small departures under background selection of the equilibrium neutral model, our background selection test is likely to remain conservative. Nordborg *et al.* (1996) show, for example, that the equilibrium neutral model with appropriately reduced population size holds reasonably well for their simulated data. Proof that our test of background selection is indeed conservative will require either a better analytical understanding of the process or extensive simulation.

There are additional evolutionary consequences of a gene's location with respect to recombination. Selective sweeps and background selection will both contribute to an increase in the fixation of deleterious mutations due to a reduction in the effective population size. Genes residing in regions of restricted recombination will be expected to have an increased substitution rate of deleterious changes. Kliman & Hey (1993) make this argument to explain lower codon bias in genes located in regions of restricted recombination; like most loci in these regions, *cta* has low codon bias (CAI, 0·184; Kliman & Hey 1993).

The intron locations of 13 G-proteins were analysed by Wilkie & Yokoyama (1993) from *Drosophila*, mice and humans. In all the G-proteins except *concertina*,

Table 3. *Divergent positions in* concertina *between* D. melanogaster *and* D. simulans

Region	Position	Base		Amino acid		
		mel.	*sim.*	*mel.*	*sim.*	Similarity[a]
Exon 1	188	A	C	Asn	His	68
	220	A	T	Asn	Ile	149
	231	G	A	Asp	Asn	23
	281	G	C	—	—	—
Intron 2	40	G	A	—	—	—
	45	C	T	—	—	—
	46	T	A	—	—	—
Exon 3	734	A	G	—	—	—
	804[b]	C	A	Pro	Thr	38
	805[b]	C	T	Pro	Leu	98
	833	T	C	—	—	—
Intron 3	13	G	T	—	—	—
	38	A	T	—	—	—
Exon 4	887	A	T	—	—	—
	909	C	G	Gln	Glu	29
	935	G	T	—	—	—
	946	A	G	Lys	Arg	26
	947	G	A	—	—	—
Intron 4	17	A	C	—	—	—
	21	T	C	—	—	—
	42	T	A	—	—	—
Intron 5	45	A	G	—	—	—
	48	G	A	—	—	—
Exon 6	1154	C	A	—	—	—
	1247	A	T	—	—	—
	1250	C	T	—	—	—
	1268	T	A	Cys	Ser	112
	1292	C	T	—	—	—
	1325	C	T	—	—	—
	1331	T	G	—	—	—
	1415	C	G	—	—	—

[a] Similarity is the mean chemical distance as proposed by Grantham (1974). Identical amino acids have a score of zero; the average difference between two amino acids is 100.

[b] The two substitutions at positions 804 and 805 are bases 1 and 2 of a single codon, resulting in a single amino acid substitution between *D. melanogaster* and *D. simulans* of Pro-Ile (similarity = 95). However, each base change counted independently would have caused an amino acid change, hence they are counted as two replacements.

four intron locations are conserved exactly with respect to the amino acid sequence, including their positions within codons. In *cta*, these positions are shifted towards the N-terminus of the protein (22, 5, 4 or 3 bases). The positions of the introns within codons have also shifted. The fact that the introns have not simply moved by multiples of three bases indicates that complex rearrangements must have occurred to conserve the reading frame. This may have been made possible by a reduction in the efficacy of selection in this G-protein due to its position in a region of low recombination; however, it may also be due to the gene's presence in heterochromatin, where the mutation rate for rearrangements may be high.

concertina is a G-protein, a member of a family conserved from yeast to primates (Simon *et al.*, 1991). However, there is a large amount of amino acid substitution between the sibling species *D. melanogaster* and *D. simulans*. Although most of the substitutions in *cta* are conservative (Table 3), they may be mildly deleterious. Comparison of the ratios of synonymous to replacement substitution between mammalian G-proteins suggests that *Drosophila cta* is evolving unusually quickly. The question arises whether the fast evolutionary rate is restricted to *cta* or whether fly proteins, in general, are evolving faster than those of mammals. Since flies are often assumed to have larger population sizes than mammals, this might come about if most amino acid replacement changes are adaptive. Few homologous loci have been sequenced in two or more species of flies and two (or more) species of mammals, making this a difficult hypothesis to evaluate. Another possibility is that some fraction of amino acid replacement changes is completely neutral in G-proteins, so that these changes will rapidly accumulate in evolution. Appropriate

data for evaluating this hypothesis – sequences of other *Drosophila* species – do not yet exist.

In contrast to the paucity of variation at *cta*, *ref(2)P*, a locus less than one map unit away, is highly polymorphic for amino acid replacement changes (Dru *et al.*, 1993; Wayne *et al.*, 1996). This suggests that the reduction in variation at *cta* may be restricted to a relatively small region immediately around the centromere. Fine-scale surveys of additional loci near the centromere will be required to determine whether local changes in levels of variation are in accordance with background selection.

We wish to thank C. F. Aquadro and B. Charlesworth for their helpful comments, and R. Hudson for suggesting, and calculating probabilities for, the 'best guess' test of the background selection model. This work was supported by NIH grant no. GM39355 to M.K.

References

Aguadé, M., Miyashita, N. & Langley, C. H. (1989). Reduced variation in the *yellow–achaete–scute* region in natural populations of *Drosophila melanogaster*. *Genetics* **122**, 607–615.

Aguadé, M., Miyashita, N. & Langley, C. H. (1992). Polymorphism and divergence in the *Mst26A* male accessory gland gene region in *Drosophila*. *Genetics* **132**, 755–770.

Aquadro, C. F., Begun, D. J. & Kindahl, E. C. (1994). Selection, recombination, and DNA polymorphism in *Drosophila*. In *Alternatives to the Neutral Model* (ed. B. Golding), pp. 46–56. London: Chapman and Hall.

Aquadro, C. F., Jennings, R. M. Jr, Bland, M. M., Laurie, C. C. & Langley, C. H. (1992). Patterns of naturally occurring restriction map variation, dopa decarboxylase activity variation and linkage disequilibrium in the *Ddc* gene region of *Drosophila melanogaster*. *Genetics* **132**, 443–452.

Begun, D. J. & Aquadro, C. F. (1991). Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the *yellow–achaete* region. *Genetics* **129**, 1147–1158.

Begun, D. J. & Aquadro, C. F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**, 519–520.

Berry, A. J., Ajioka, J. W. & Kreitman, M. (1991). Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* **129**, 1111–1117.

Braverman, J. R., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. (1995). The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **141**, 783–796.

Charlesworth, B. (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research* **63**, 213–227.

Charlesworth, B. (1996). Background selection and patterns of genetic diversity in Drosophila. *Genetical Research*, in press.

Charlesworth, B., Morgan, M. T. & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Dru, P., Bras, F., Dezélée, S., *et al.* (1993). Unusual variability of the *Drosophila melanogaster ref(2)p* protein which controls the multiplication of sigma rhabdovirus. *Genetics* **133**, 943–954.

Grantham, R. (1974). Amino acid difference formula to help explain protein evolution. *Science* **185**, 862–864.

Hey, J. & Kliman, R. M. (1993). Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. *Molecular Biology and Evolution* **10**, 804–822.

Hudson, R. (1992). Gene genealogies and the coalescent process. In *Oxford Series in Ecology and Evolution* (ed. R. M. May & P. H. Harvey), pp. 1–44. Oxford: Oxford University Press.

Hudson, R. R. & Kaplan, N. L. (1994). Gene trees with background selection. In *Alternatives to the Neutral Model* (ed. B. Golding), pp. 140–153. London: Chapman and Hall.

Hudson, R. R. & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics* **141**, 1605–1617.

Hudson, R. R., Kreitman, M. & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. (1989). The 'hitchhiking effect' revisited. *Genetics* **123**, 887–899.

Kliman, R. & Hey, J. (1993). Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Molecular Biology and Evolution* **10**, 1239–1258.

Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. *Nature* **304**, 412–417.

Kreitman, M. & Hudson, R. R. (1991). Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**, 565–582.

Kreitman, M. & Wayne, M. L. (1994). Organization of genetic variation at the molecular level: lessons from *Drosophila*. In *Molecular Approaches to Ecology and Evolution* (ed. B. Schierwater, B. Streit, G. P. Wagner & R. DeSalle), pp. 157–184. Basel: Birkhauser-Verlag.

Martín-Campos, J. M., Comerón, J. M., Miyashita, N. & Aguadé, M. (1992). Intraspecific and interspecific variation at the *y–ac–sc* region of *Drosophila simulans* and *Drosophila melanogaster*. *Genetics* **130**, 805–816.

Maynard Smith, J. & Haigh, J. (1974). The hitchhiking effect of a favorable gene. *Genetical Research* **23**, 23–25.

Miyashita, N. & Langley, C. H. (1988). Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. *Genetics* **120**, 199–212.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution* **3**, 418–426.

Nordborg, M., Charlesworth, B. & Charlesworth, D. (1996). The effect of recombination on background selection. *Genetical Research* **67**, 159–174.

Palopoli, M. & Wu, C.-I. (1996). Rapid evolution of a coadapted gene complex: evidence from the *Segregation Distorter* (*SD*) system of meiotic drive in *Drosophila melanogaster*. *Genetics* **143**, 1675–1688.

Parks, S. & Wieschaus, E. (1991). The *Drosophila* gastrulation gene *concertina* encodes a G-alpha-like protein. *Cell* **64**, 447–458.

Simon, M. W., Strathmann, M. P. & Guatam, N. (1991). Diversity of G proteins in signal transduction. *Science* **252**, 802–808.

Stephan, W. & Langley, C. H. (1989). Molecular genetic variation in the centromeric region of the X chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. *Genetics* **121**, 89–99.

Stephan, W. & Mitchell, S. H. (1992). Reduced levels of DNA polymorphism and fixed between-population

differences in the centromeric region of *Drosophila ananassae*. *Genetics* **132**, 1039–1045.

Stephan, W., Wiehe, T. H. E. & Lenz, M. W. (1992). The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. *Theoretical Population Biology* **41**, 237–254.

Takano, T. S., Kusakabe, S. & Mukai, T. (1993). DNA polymorphism and the origin of protein polymorphism at the *Gpdh* locus of *Drosophila melanogaster*. In *Mechanisms of Molecular Evolution: Introduction to Molecular Paleopopulation Biology* (ed. N. Takahata & A. G. Clark), pp. 179–190. Sunderland, MA: Sinauer.

Walsh, S. P., Metzger, D. A. & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR based typing from forensic material. *BioTechniques* **10**, 506–513.

Wayne, M. L., Contamine, D. & Kreitman, M. (1996). Molecular population genetics of *ref(2)P*, a locus which confers viral resistance in *Drosophila*. *Molecular Biology and Evolution* **13**, 191–199.

Weichenhan, D. (1991). Fast recovery of DNA from agarose gels by centrifugation through blotting paper. *Trends in Genetics* **7**, 109.

Wilkie, T. M. & Yokoyama, S. (1993). Evolution of the G protein alpha subunit multigene family. *Society of General Physiologists Series* **49**, 249–270.