rodegeneration exceeding the spontaneous rate in a given brain region, we used an unbiased stereological disector method (*17*) to quantify the numerical density (neurons/mm^3) of normal neurons in 70-μm Nissl-stained sections, or of degenerating neurons in 70-μm sections stained by the DeOlmos silver method (*13*). A total of 8 to 10 disectors (0.05 mm by 0.05 mm, disector height 0.07 mm) were used to sample each brain region. Counts were performed in a blinded manner. To establish the absolute numbers of degenerating neurons, we identified boundaries of individual brain regions [thalamus, dentate gyrus, CA1 hippocampus, subiculum, caudate, septum, hypothalamus, amygdala, and frontoparietal, cingulate, retrosplenial, temporal, and pyriform/entorhinal cortices, according to G. Paxinos, I. Tork, L. H. Tecott, K. L. Valentino, *Atlas of the Developing Rat Brain* (Academic Press, New York, 1991)] in Nissl-stained sections. We used an image analysis system (IMAGE 1.54, National Institutes of Health) to facilitate volume determinations for each brain region. Multiplication of the volume of a given region by the numerical density of degenerating neurons in that region provided an estimate of total numbers of neurons deleted from each brain region by ethanol treatment. The regional values were summed to give a total for each brain, and from these totals, means (±SEM) were calculated separately for the brains in the ethanol- and saline-treated groups.

17. M. J. West, *Neurobiol. Aging* **14**, 275 (1993); H. J. G. Gundersen, T. F. Bendtsen, L. Korbo, M. J. West, *APMIS* **96**, 379 (1988).
18. In addition to GABAergic agents, we tested 6-nitro-7-sulfamoylbenzo[f]quinoxaline-2,3-dione (NBQX), an antagonist of non-NMDA glutamate receptors, at a dose of 20 mg/kg ip given at 0, 75, and 150 min and at 8 hours ($n = 5$); scopolamine hydrobromide, an antagonist of cholinergic muscarinic receptors, at a dose of 0.3 mg/kg ip given at 0, 4, and 8 hours ($n = 5$); haloperidol, an antagonist of dopamine receptors, at a dose of 10 mg/kg ip given at 0 and 8 hours ($n = 5$); and L-dopa, an agent that increases the availability of dopamine at dopamine receptors, administered at 50, 100, or 200 mg/kg ip together with benserazide (dopamine decarboxylase inhibitor) at 14, 28, or 56 mg/kg ip ($n = 5$ per dose). Finally, because NMDA receptor–activated ion channels are highly permeable to Ca^{2+} ions, we tested whether blockade of calcium influx via other routes (i.e., voltage-dependent Ca^{2+} channels) may also cause apoptotic neurodegeneration in the brain. For this purpose we treated P7 rats with the Ca^{2+} channel blockers nimodipine (50 mg/kg ip at 0 and 8 hours; $n = 6$) or nifedipine (10 mg/kg ip at 0 and 8 hours; $n = 6$). We examined the brains histologically 24 hours after each of these treatments. None of these agents reproduced the apoptosis-inducing action of NMDA antagonists.
19. Supported in part by Deutsche Forschungsgemeinschaft grant Ik2/2-1, Humboldt University grant 98-649, National Institute for Mental Health Research Scientist Award MH 38894 (J.W.O.), National Institute on Aging grant AG 11355, National Institute on Drug Abuse grant DA 05072, National Eye Institute grant EY 08089, and a National Alliance for Research on Schizophrenia and Depression Established Investigator Award (J.W.O.).

3 August 1999; accepted 21 December 1999

Evidence for DNA Loss as a Determinant of Genome Size

Dmitri A. Petrov,[1]* Todd A. Sangster,[2] J. Spencer Johnston,[3] Daniel L. Hartl,[2] Kerry L. Shaw[2]

Eukaryotic genome sizes range over five orders of magnitude. This variation cannot be explained by differences in organismic complexity (the *C* value paradox). To test the hypothesis that some variation in genome size can be attributed to differences in the patterns of insertion and deletion (indel) mutations among organisms, this study examines the indel spectrum in *Laupala* crickets, which have a genome size 11 times larger than that of *Drosophila*. Consistent with the hypothesis, DNA loss is more than 40 times slower in *Laupala* than in *Drosophila*.

Wide variation in eukaryotic genome size is a pervasive feature of genome evolution. Large differences in haploid DNA content (*C* value) are found within protozoa (5800-fold range), arthropods (250-fold), fish (350-fold), algae (5000-fold), and angiosperms (1000-fold) (*1*). This variation is called the *C* value paradox (*2*, *3*) because genome size is not correlated with the structural complexity of organisms or with the estimated number of genes. Despite much progress in the study of genomes, the *C* value paradox remains largely unresolved.

Drosophila species, which have small genomes, spontaneously lose DNA at a much higher rate than mammalian species, which have large genomes (*4*–*7*). Although many mechanisms can affect genome size—including polyploidy, fixation of accessory chromosomes or large duplications (*8*), and expansions of satellite DNA or transposable elements (*9*)—the *Drosophila* findings suggest that some differences in haploid genome size may result from variation in the rate of spontaneous loss of nonessential DNA (*4*). Here, we test this hypothesis by examining the indel spectrum in Hawaiian crickets (*Laupala*), which have a genome size ~11-fold larger than that of *Drosophila* (*10*). Specifically, we test the prediction of a lower rate of DNA loss in *Laupala* than in *Drosophila*, corresponding to the large difference in genome size.

Sequences unconstrained by natural selection exhibit patterns of substitution, reflecting the underlying spectra of spontaneous mutations (*11*). As pseudogene surrogates we chose nontransposing copies of non-LTR (long terminal repeat) retrotransposable elements (*4*, *12*). Transposition of non-LTR elements usually results in a 5′-truncated copy that is unable to transpose because of lack of a promoter and lack of the capacity to encode functional proteins (*13*, *14*); these "dead-on-arrival" (DOA) elements are essentially pseudogenes.

We identified a new non-LTR element in *Laupala*, here designated *Lau1*, by means of polymerase chain reaction (PCR) with degenerate primers to conserved regions of the non-LTR reverse transcriptase (*15*). Evolution of unconstrained DOA elements can be distinguished from that of the constrained, active elements via phylogenetic analysis of nucleotide sequences of individual DOA elements (*4*, *12*). Substitutions in a transpositionally active lineage are represented in multiple DOA elements generated by transposition of the active copy, whereas substitutions in each DOA lineage are unique (barring parallel mutations) because of the inability of DOA elements to transpose. This implies that, in a gene tree of non-LTR sequences from closely related species, the active lineages map to internal branches (identified through substitutions shared among elements), whereas DOA lineages map to terminal branches (identified through unique substitutions). Some DOA lineages may also map to internal branches, because elements from different species may be identical by descent (IBD) because of transmission of the same (allelic) DOA copy from a common ancestor (*7*, *12*). Nevertheless, as long as the number of active lineages is small and the sampling is dense, substitutions in the terminal branches will correspond primarily to the DOA element evolution (*12*).

If the terminal branches of the *Lau1* gene tree (Fig. 1) represent unconstrained evolution of DOA elements, we predict the absence of purifying selection operating along these branches. Confirming this prediction, point substitutions in terminal branches map with equal frequencies to all three codon positions (*G* test; $P = 0.64$). In addition, the terminal branches feature numerous element-specific indels (48 deletions and 18 insertions in 49 terminal branches). The internal branches show evidence of relaxed selection as well, which suggests that many elements in our sample are IBD through inheritance of ancestral allelic DOA copies. The substitutions in the internal branches are found at equal fre-

[1]Harvard University Society of Fellows, [2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [3]Department of Entomology, Texas A&M University, College Station, TX 77843, USA.

*To whom correspondence should be addressed. E-mail: dpetrov@oeb.harvard.edu

quencies among the three codon positions (G test; $P = 0.55$), with many indels shared among sequences. Most shared indels can be assigned to the tree without homoplasy (21 of 24 deletions and 8 of 8 insertions), which offers strong independent support for our phylogenetic inference (*16*). Although the deep internal branches are expected to show evidence of purifying selection (*4*, *12*), in this case they are too short and poorly resolved to demonstrate this feature.

Our analysis of the relative rates of indels and nucleotide substitutions is confined to the terminal branches only. The length of these branches varies widely, from 0 (in *Laupala kohalensis 514*) to 50 (in *L. kohalensis 183*), most likely because of varying time that individual elements have been accumulating independent nucleotide substitutions, from the moment of either the original DOA insertion or the most recent speciation for allelic copies. As expected on the basis of this supposition, we find a strong positive correlation between both the numbers of insertions and nucleotide substitutions (Spearman's rank correlation, $r_s = 0.58$; $P = 8 \times 10^{-6}$) and the numbers of deletions and nucleotide substitutions ($r_s = 0.79$; $P = 1.4 \times 10^{-11}$).

These positive correlations provide a basis for estimating relative rates of indels and point substitutions in *Laupala* (Table 1). The maximum likelihood (ML) estimate of the rate of deletions relative to nucleotide substitutions (*4*) is about half as great in *Laupala* as in *Drosophila* (Fig. 2), and the difference is statistically significant (ML analysis, $P = 1 \times 10^{-3}$). *Laupala* exhibits a 40% higher rate of insertions than does *Drosophila*, although this is not statistically significant (ML analysis, $P = 0.14$). The size distribution of indels also implies a lower rate of DNA loss in *Laupala*. On average, *Laupala* deletions are almost four times smaller (Wilcoxon test, $P = 0.009$), and insertions are almost two times larger (*17*) (Wilcoxon test, $P = 0.03$) than those in *Drosophila*. Most of the difference in DNA loss is due to *Laupala* having a much smaller fraction of deletions larger than 15 base pairs (bp) (Fig. 3). For deletions smaller than 15 bp, the rates of deletions per substitution are indistinguishable (ML analysis, $P = 0.34$).

The differences in indel spectra result in a 10.8-fold lower rate of DNA loss per nucleotide substitution in *Laupala* versus *Drosophila*. Our data also suggest that the rates of nucleotide substitution vary in a way that magnifies the difference in the rates of DNA loss. Figure 1 includes five clusters of sequences that share indels within each cluster and that also contain a sequence of *Lau1* from *Prolaupala kukui* (for example, the cluster of *Laupala molokoiensis 52* and *P. kukui 102*). Within each cluster the sequences are likely to be IBD from a DOA copy present in a

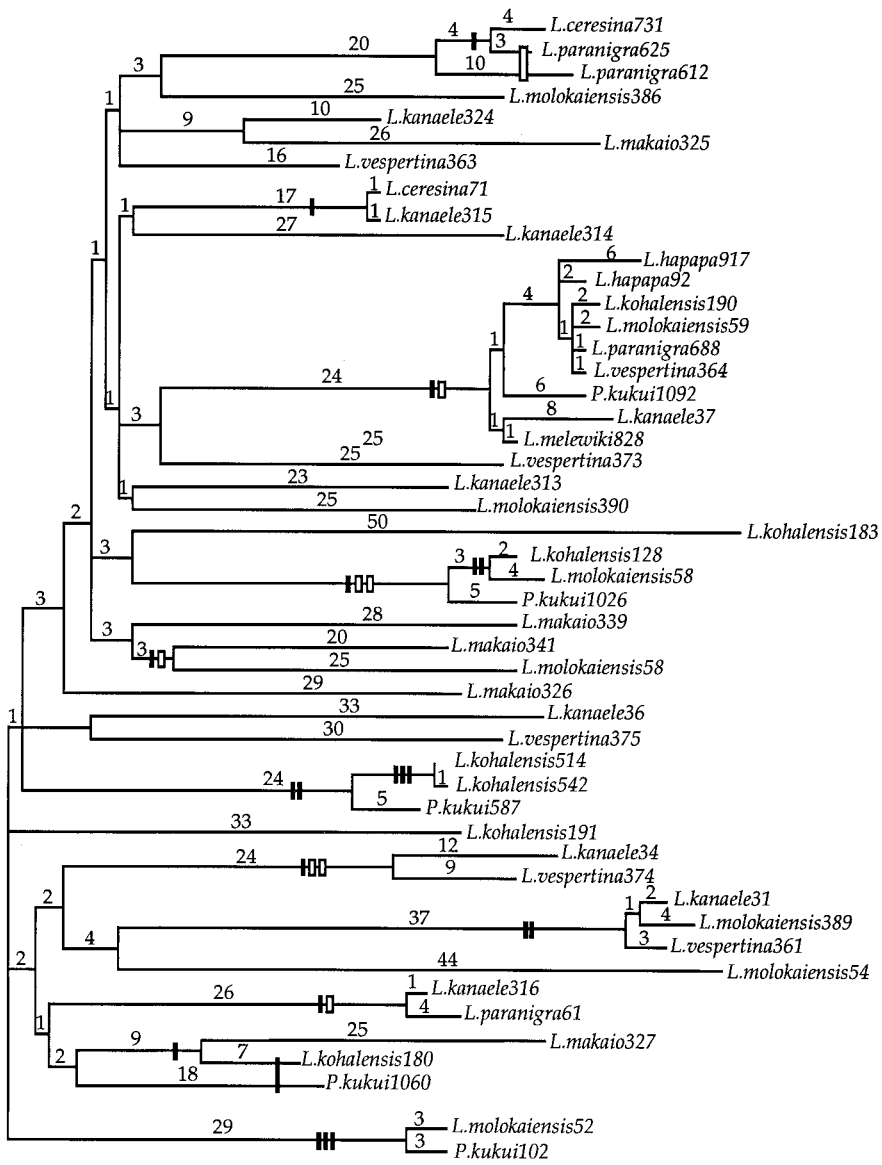

Fig. 1. Phylogenetic analysis of *Lau1*. The 50% majority-rule tree of 1000 equally parsimonious trees is shown, rooted at midpoint. Numbers of unambiguous nucleotide substitutions are shown above each branch. Indels on internal branches are shown by solid (deletions) and open (insertions) bars. The tree length is 1014 steps (confidence interval = 0.64). The *Lau1* gene trees were estimated by NJ, UPGMA, maximum-likelihood (F84 model), and maximum-parsimony methods [as implemented in PAUP4.0b (*24*)], with equal weighting of all positions, ignoring insertions, and treating deletions as missing data. Although trees differ somewhat, depending on the reconstruction method, none of the conclusions is sensitive to these differences. This is because the conclusions depend only on the changes observed in the terminal branches (the DOA lineages), whereas the alternative trees differ primarily in the deep internal branches. Sequence alignment was done by hand with the aid of Sequencher 3.0 (GeneCodes).

Table 1. Comparison of deletion and insertion profiles in *Laupala* and *Drosophila*.

	Laupala	*Drosophila**	Significance
Ratio of deletions to nucleotide substitutions	0.07 (0.05–0.10)†	0.13 (0.11–0.18)	$P \ll 0.01$
Ratio of insertions to nucleotide substitutions	0.03 (0.02–0.04)	0.01 (0.01–0.03)	NS‡
Mean deletion size (bp)	6.0 ± 1.59	24.9 ± 2.62	$P < 0.01$
Mean insertion size (bp)	5.1 ± 1.79§	3.2 ± 0.75	$P < 0.05$
Half-life of a pseudogene (Myr)¶	615	14	$P \ll 0.01$

*Data from (*4*, *9*). †Numbers in parentheses are 95% confidence intervals. ‡Not statistically significant. §Conservative estimate that excludes three large insertions in *Laupala* of 443, 705, and >900 bp. ¶Assumes neutral rate of 15×10^{-3} substitution per Myr in *Drosophila* (*19*) and 3.8×10^{-3} substitution per Myr in *Laupala* (see text).

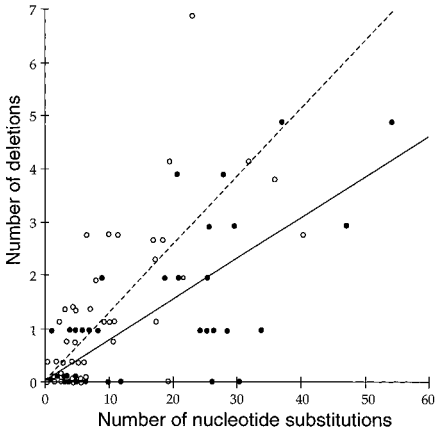

Fig. 2. Relationships between numbers of deletions and numbers of terminal-branch substitutions corrected for sequence length and multiple substitutions (Jukes-Cantor correction) in *Drosophila* (*4*, *7*) (open circles) and *Lau1* (solid circles). ML regression line for *Drosophila* is dashed and that for *Lau1* is solid.

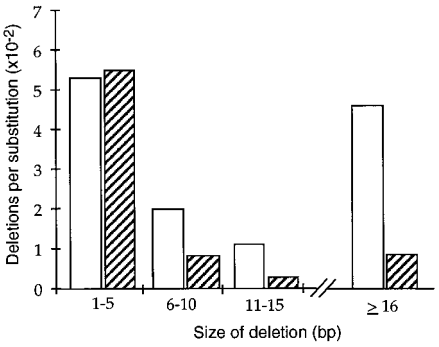

Fig. 3. Distribution of deletion sizes in *Lau1* (hatched bars) and *Drosophila* (*7*) (open bars).

common ancestor of *Laupala* and *Prolaupala*. By dividing the pairwise nucleotide distances between *Prolaupala* and *Laupala* elements within each cluster by 10 million years (Myr) (twice the divergence time between *Laupala* from *Prolaupala* estimated from biogeography) (*18*), we obtain an estimate of the absolute rate of point substitution. The average over all clusters is 3.8×10^{-3} nucleotide substitution per Myr, which is almost four times lower than the 15×10^{-3} nucleotide substitution per Myr estimated in *Drosophila* (*19*). Combined with the 10.8-fold lower rate of DNA loss per nucleotide substitution in *Laupala*, the 4 times lower rate of nucleotide substitution yields an overall rate of DNA loss per Myr that is 42-fold less in *Laupala* than in *Drosophila*. Thus, as we predicted, the rate of DNA loss in *Laupala* is substantially lower than that in *Drosophila*.

We have examined the possibility that, to optimize genome size, each individual DOA copy may be selected for length (*4*, *5*, *20*). Selection of this type should produce a correlation between the number of substitutions and the lengths of deletions in the terminal branches (*12*, *20*), but the predicted correlation is not observed in either our *Laupala* data ($r_s = -0.07$; $P = 0.62$) or the sample of DOA elements studied in *Drosophila* (*7*, *20*). These results imply that our estimates of the indel spectra are not significantly biased by natural selection for individual indels.

Our results have no bearing on the presence or absence of selective forces that may affect genome size (*1*), nor do they imply anything about the lengths of particular classes of constrained sequences, such as introns, intergenic spacers, or 5′ and 3′ untranslated regions (*21, 22*). Our data do suggest one reason why some small genomes have few pseudogenes (*21*, *23*): a high rate of DNA loss should result in a lower steady-state number of pseudogenes in small genomes.

The key question that remains is empirical and quantitative: how much of the variation in genome size can be explained by variation in the indel spectra? The relative ease with which indel patterns can be assayed using non-LTR elements should enable us to answer this question in a wide variety of eukaryotes and thus to test the mutational hypothesis for the *C*-value paradox in a comprehensive fashion.

References and Notes

1. T. Cavalier-Smith, *The Evolution of Genome Size* (Wiley, New York, 1985), p. 523.
2. A. E. Mirsky and H. Ris, *J. Gen. Physiol.* **34**, 451 (1951).
3. C. A. Thomas, *Annu. Rev. Genet.* **5**, 237 (1971).
4. D. A. Petrov, E. R. Lozovskaya, D. L. Hartl, *Nature* **384**, 346 (1996).
5. B. Charlesworth, *Nature* **384**, 315 (1996).
6. D. A. Petrov, Y.-C. Chao, E. C. Stephenson, D. L. Hartl, *Mol. Biol. Evol.* **15**, 1562 (1998).
7. D. A. Petrov and D. L. Hartl, *Mol. Biol. Evol.* **15**, 293 (1998).
8. B. John and G. L. G. Miklos, *The Eukaryotic Genome in Development and Evolution* (Allen & Unwin, London, 1988).
9. P. SanMiguel, B. S. Gaut, A. Tikhonov, Y. Nakajima, J. L. Bennetzen, *Nat. Genet.* **20**, 43 (1998).
10. Orthopteran genomes tend to be large compared with other insects [J. E. Belda, J. Cabrero, J. P. M. Camacho, J. S. Rufas, *Cytobios* **67**, 13 (1991)]; large genome sizes in Orthoptera are not known to be associated with polyploidy. We measured the haploid genome size of *Laupala cerasina* to be 1.91×10^9 bp, which is ~11 times larger than the 1.7×10^8 bp genome of *Drosophila melanogaster* [G. T. Rudkin, *Genetics* **61** (suppl. 1), 227 (1969)]. Genome size estimation was as in [H. P. Thindwa, G. L. Teetes, J. S. Johnston, *Southwest. Entomol.* **19**, 371 (1994)]. The nuclei were isolated from the specimen heads and resuspended in staining buffer containing propidium iodide at 50 parts per million. The combined insect and chicken erythrocyte nuclei were stained for 1 hour and then analyzed with an Epics Elite flow cytometer equipped with a Coherent 4-W argon laser tuned to 514 nm at 500 mW. The head contained primarily G_1 nuclei, with a G_1/G_2 nuclei ratio of about 3:1. Brain cells were used because they are almost exclusively diploid in insects, thus minimizing the possibility of somatic polyploidy biasing our estimates.
11. W.-H. Li, T. Gojobori, M. Nei, *Nature* **292**, 237 (1981).
12. D. A. Petrov and D. L. Hartl, *Gene* **205**, 279 (1997).
13. D. D. Luan, M. H. Korman, J. L. Jacubczak, T. H. Eickbush, *Cell* **72**, 595 (1993).
14. C. A. Hutchison III, S. C. Hardies, D. D. Loeb, W. R. Shehee, M. H. Edgell, in *Mobile DNA*, D. E. Berg and M. M. Howe, Eds. (American Society for Microbiology, Washington, DC, 1989), pp. 593–617.
15. Degenerate primers (*DVO144* and *DVO145*) were described in D. A. Wright *et al.* [*Genetics* **142**, 569 (1996)]. The rest of the sequences were obtained by PCR with the DNA from a single individual per species and *Lau1*-specific primers *Lau1* (5′-GTGGAATGGG-TATTTGGTTTG-3′) and *Lau2*$^+$ (5′-GAATACACAA-TTTTTGGTGGA-3′), using 30 cycles of 94°C for 30 s, 56°C for 1 min, and 72°C for 2 min. Size-unselected products of the total PCRs were TA-cloned (Invitrogen). For identical clones from the same species (<2% divergence with no length differences), we used a single representative clone. Such a high cutoff is conservative because PCR mistakes in the absence of length differences lead to an underestimate of the rate of indels. The eight eliminated sequences were from 0 to 0.7% (average, 0.2%) divergent from their "identical" counterparts. All sequences have been deposited in GenBank (accession numbers AF109009 through AF108961).
16. The *Lau1* gene tree provides evidence for diploidy in *Laupala*. Although multiple orthologous IBD copies of the same element were sampled repeatedly, in no case did we find more than two copies from a single individual. The simplest explanation for this pattern is that *Laupala* is diploid, with a maximum of two alleles at each locus.
17. To be conservative, we ignored three insertions of much greater length in the *Laupala* data (443, 705, and >900 bp).
18. K. L. Shaw, *Evolution* **50**, 237 (1996); R. C. Fleischer, C. E. McIntosh, C. L. Tarr, *Mol. Ecol.* **7**, 533 (1998).
19. P. M. Sharp and W.-H. Li, *J. Mol. Evol.* **28**, 398 (1989).
20. D. A. Petrov and D. L. Hartl, unpublished data.
21. G. Elgar *et al.*, *Trends Genet.* **12**, 145 (1996).
22. H. Ogata, W. Fujibuchi, M. Kanehisa, *FEBS Lett.* **390**, 99 (1996); E. Moriyama, D. A. Petrov, D. L. Hartl, *Mol. Biol. Evol.* **15**, 770 (1997); A. L. Hughes and M. K. Hughes, *Nature* **377**, 391 (1995).
23. A. M. Weiner, P. L. Deininger, A. Efstratiadis, *Annu. Rev. Biochem.* **55**, 631 (1986); P. Jeffs and M. Ashburner, *Proc. R. Soc. London, Ser. B* **244**, 151 (1991).
24. D. L. Swofford, *PAUP: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 4.0 beta* (Sinauer, Sunderland, MA, 1998).

25. Supported by National Science Foundation grant DEB-9729325 to K.L.S., National Institutes of Health grants to D.L.H, and the William F. Milton Fund Award to D.A.P. We thank M. Siegal, J. Morris, J. Parsch, and members of the D.L.H. and K.L.S. laboratories for helpful discussions.

1 November 1999; accepted 7 December 1999