

ORIGINAL ARTICLE

Genome size diversity in the family Drosophilidae

TR Gregory¹ and JS Johnston²¹Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada and ²Department of Entomology, Texas A&M University, College Station, TX, USA

Flies in the genus *Drosophila* have been the dominant model organisms in genetics for over a century and, with a dozen complete sequences now available, continue as such in modern comparative genomics. Surprisingly, estimates of genome size for this genus have been relatively sparse, covering less than 2% of species. Here, best practice flow cytometric genome size estimates are reported for both male and female flies from 67 species from six genera in the family Drosophilidae, including 55 species from the genus *Drosophila*. Direct and phylogenetically corrected correlation

analyses indicate that genome size is positively correlated with temperature-controlled duration of development in *Drosophila*, and there is indication that genome size may be positively related to body size and sperm length in this genus. These findings may provide some explanation for the streamlined genomes found in these insects, and complement recent work demonstrating possible selective constraints on further deletion of noncoding DNA.

Heredity (2008) **101**, 228–238; doi:10.1038/hdy.2008.49; published online 4 June 2008

Keywords: C value; developmental rate; flow cytometry; *Drosophila*; satellite DNA; transposable elements

Introduction

Through all the major transitions in genetics over the past 100 years—from early mutation and mapping studies involving countless crosses and phenotypic analyses, to karyotyping and polytene chromosome banding, to the application of allozymes in population-level surveys, to the advent of complete genome sequencing and the rise of ‘evo-devo’—the fly *Drosophila melanogaster* has maintained its uncontested status as a preeminent model organism (Brookes, 2001; Beller and Oliver, 2006). Several entire volumes have been devoted to its use in experimental genetics (Demerec and Kaufman, 1996; Powell, 1997; Sullivan *et al.*, 2000; Henderson, 2003; Ashburner *et al.*, 2005), and it is estimated that there are well over 1000 research groups worldwide who use *Drosophila* as a key model (Clark *et al.*, 2003). As Demerec and Kaufman (1996, p. 1) put it, ‘it would be not an exaggeration to say that we have learned more about the basic laws of heredity from the study of this fly than from work on all other organisms combined’.

As might be expected, the *D. melanogaster* genome was among the earliest to be sequenced for a metazoan (behind only *Caenorhabditis elegans*) and was the first animal genome to be completed using the now-standard shotgun approach (Adams *et al.*, 2000; Celniker *et al.*, 2002). Although genome sequences are available for other pairs of related species (for example, the rodents *Mus musculus* and *Rattus norvegicus* and the pufferfishes *Takifugu rubripes* and *Tetraodon nigroviridis*),

the recently completed sequence of *D. pseudoobscura* (Richards *et al.*, 2005) made *Drosophila* the first genus to be represented by multiple species. These have since been joined by 10 additional species: *D. ananassae*, *D. erecta*, *D. grimshawi*, *D. mojavensis*, *D. persimilis*, *D. sechellia*, *D. simulans*, *D. virilis*, *D. willistoni* and *D. yakuba* (Ashburner, 2007; Crosby *et al.*, 2007; *Drosophila* 12 Genomes Consortium, 2007; Gilbert, 2007).

Drosophila species continue to contribute to important discoveries with regard to the complex nature of genome form, function and evolution. Most recently, studies of *Drosophila* genomes have provided new insights into the extent of chromosomal rearrangements (Bartolomé and Charlesworth, 2006), the occurrence of heterochromatic genes (Yasuhara and Wakimoto, 2006) and the potential role of some regions of noncoding DNA in both the origin of new genes (Levine *et al.*, 2006; Casola *et al.*, 2007) and the regulation of existing ones (Halligan and Keightley, 2006; Stark *et al.*, 2007).

In light of these considerations, it is very surprising that only limited effort has been devoted to the study of genome size in *Drosophila*. Not including the data presented here, only 42 of the roughly 2600 species of *Drosophila* (that is, ~1.6%) have been assessed with regard to this fundamental aspect of genome organization (Bosco *et al.*, 2007; Gregory, 2008; see also Ashburner *et al.*, 2005, which provides a preliminary subset of the present data). Many of the previous estimates have been based on non-best-practice methods, and even some that have used modern techniques such as flow cytometry have used potentially problematic staining protocols. This scarcity and limited reliability of drosophilid genome size data is a significant concern given the growing emphasis on interspecific comparisons in the family (Markow and O’Grady, 2006), the unparalleled focus on *Drosophila* in comparative genomics (Clark *et al.*, 2003; *Drosophila* 12 Genomes

Correspondence: Dr TR Gregory, Department of Integrative Biology, University of Guelph, 50 Stone Road East, Guelph, Ontario, Canada N1G2W1.

E-mail: rgregory@uoguelph.ca

Received 23 January 2008; revised 28 April 2008; accepted 2 May 2008; published online 4 June 2008

Consortium, 2007; Stark *et al.*, 2007) and the combined theoretical and practical importance of genome size data for large-scale genome research (Gregory, 2005a, b).

As an early corrective, the present study provides new genome size data for 67 species in the family Drosophilidae using best-practice techniques involving more than 800 individual flies. The analysis includes representatives from the genera *Chymomyza*, *Drosophila*, *Hirtodrosophila*, *Samoaia*, *Scaptodrosophila* and *Zaprionus*; more than 40 of the species analyzed have never been the subject of genome quantification. In combination with previously published estimates, these genome size data are used to evaluate the phenotypic implications of bulk DNA content in the Drosophilidae. The results shed new light on the significance of genome size diversity in this family.

Materials and methods

Source of specimens

Specimens were obtained from the Tucson *Drosophila* Species Stock Center (<http://stockcenter.arl.arizona.edu/>), with the exception of the *D. melanogaster* Iso-1 strain which was provided by Jerry Rubin from the line sequenced in the *D. melanogaster* genome project. Species names and other taxonomic details follow the BioSystematic Database of World Diptera (<http://www.sel.barc.usda.gov/Diptera/biosys.htm>) and the TaxoDros Database (<http://www.taxodros.unizh.ch/>).

Flow cytometry

Genome sizes were estimated using neural tissue nuclei and a propidium iodide flow cytometry protocol described in detail elsewhere (Bennett *et al.*, 2003; DeSalle *et al.*, 2005). Briefly, this involved the grinding of heads in ice-cold Galbraith buffer using 2 ml Kontes dounce tissue grinders with a type A pestle to free individual nuclei, filtering through 20 µm nylon mesh to eliminate large debris, staining for 1–9 h in 50 µg ml⁻¹ propidium iodide and analysis using a Coulter Epics Elite flow cytometer with a laser tuned to 514 nm and 500 mW. Given potential problems with overlap between fluorescence peaks from species of *Drosophila* (Figure 1), sperm and blood of the green pufferfish, *T. fluviatilis*, were used as internal standards rather than *D. melanogaster* Iso-1 (cf. DeSalle *et al.*, 2005). However, the relative DNA content of *T. fluviatilis* was first established by comparison with female *D. melanogaster* Iso-1 strain (1C = 0.18 pg), such that in the present study *D. melanogaster* can be considered an indirect secondary standard. All runs included at least five female *D. melanogaster*, so that linearity or other machine related problems could be detected. A second measurement run was performed in any cases in which deviations of more than 0.01 pg were registered among female *D. melanogaster* from the same sample colony. Standards and unknowns were run independently and directly against female *D. melanogaster* Iso-1 as necessary to distinguish among close peaks in the output histograms (Figure 1), and all genome size determinations were ultimately made using co-stained samples as per best-practice techniques (DeSalle *et al.*, 2005).

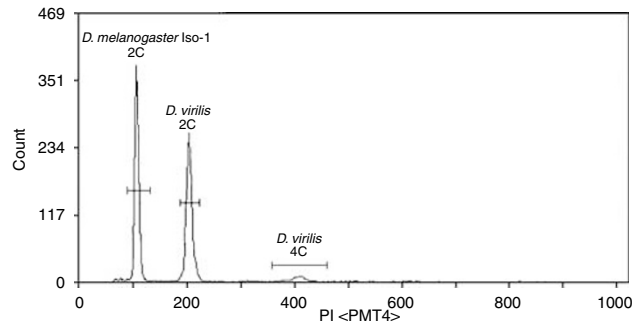


Figure 1 Sample histogram from the flow cytometric analysis of propidium iodide-stained neural nuclei of female *D. melanogaster* Iso-1 and *D. virilis*. 2C and 4C peaks were apparent for *D. virilis*, but any 4C nuclei from *D. melanogaster* were likely to have fallen under the *D. virilis* 2C peak as this species has a genome almost exactly twice as large. In cases where fly peaks were not so distantly separated, *T. fluviatilis* sperm was used as a relative internal standard, first calibrated against *D. melanogaster* Iso-1.

Comparative data

Data for various potential correlates of genome size were obtained from compiled lists available in the literature. Temperature-controlled developmental time data (duration from egg to adult in days at 18 °C) were taken from the Tucson *Drosophila* Species Stock Center, as reported in Markow and O'Grady (2006). Data for post-eclosion time to sexual maturity (in days) were also obtained from Markow and O'Grady (2006), whereas data with regard to sperm length (in mm) and body size (thorax length in mm) were derived from Pitnick *et al.* (1995), Kacmarczyk and Craddock (2000) and Moreteau *et al.* (2003). Karyotypic data were taken from Clayton and Wheeler (1975) and Clayton and Guest (1986). Total haploid chromosome number was computed independently of chromosome morphology, and a 'karyotypic index' that reflected a general measure of total chromosomal size was calculated according to the following formula:

$$\text{Index} = 4V + 3J + 2R + 1D + 2v + 1r \quad (1)$$

where the letters refer to the number of chromosomes per morphological category as designated in Clayton and Wheeler (1975): V, V-shaped; J, J-shaped; R, rod; D, dot; v, small V-shaped; r, small rod.

Phylogenetic analyses

Two tree topologies ('Tree 1: more genes' and 'Tree 2: more taxa') were assembled from phylogenetic information available in the literature. The trees and detailed descriptions of their assembly are given in the Supplementary Material. Felsenstein's (1985) phylogenetically independent contrasts (PICs), positivized and forced through the origin, were calculated using log-transformed data under both tree topologies with the PDAP module (Midford *et al.*, 2002) in Mesquite version 1.11 (Maddison and Maddison, 2006). Branch lengths were all set to 1 because the assembled trees did not contain branch length information. One degree of freedom was subtracted for soft polytomies (Purvis and Garland, 1993). In all comparative analyses (direct or phylogenetic), data from different strains and from male and female flies were averaged to give a single species value.

Results

General patterns of variation

The genome sizes of the 67 species analyzed in the present study varied nearly threefold, from 0.14 pg in male *D. mauritiana* (Rob strain) and *Hirtodrosophila pictiventris* to 0.40 pg in female *Chymomyza amoena* (Table 1). With only a few exceptions, male flies had significantly smaller estimated genome sizes than female flies in any given species (paired *t*-test, $P < 0.0001$; Table 1). The mean genome size among all species of *Drosophila* studied to date is $0.21 \text{ pg} \pm 0.005 \text{ s.e.}$ (Table 1; Gregory, 2008). Some previously published estimates suggest both smaller and larger values in the genus (Gregory, 2008), however the lowest estimates must be interpreted with caution as they were obtained using older methods such as reassociation kinetics. Likewise, the value of 0.40 pg reported for *D. nasutoides* by Zacharias *et al.* (1982) was based on a Feulgen densitometry comparison of brain tissue versus a chicken blood standard, and may have been overestimated due to differences in DNA compaction level (and thus stain uptake) between the two tissue types. Bosco *et al.* (2007) reported values as high as 0.44 pg in some drosophilids, but these were acknowledged to be overestimates based on their choice of flow cytometric protocols. The present study therefore represents the most reliable estimate of the range of genome size diversity in *Drosophila* species and their relatives.

A nested analysis of variance (ANOVA) using all available data indicated that 5% of the variation among the *Drosophila* species studied thus far occurs at the level of subgenera within the genus, 59% among groups within subgenera, 14% among subgroups within groups and 22% among species within subgroups. However, it bears noting that this analysis is sensitive to the taxa that are included, especially in terms of the representation of higher level categories. At the subgenus level, there is a significant difference in mean genome size between the subgenera *Drosophila* (mean = $0.24 \pm 0.02 \text{ s.e.}$) and *Sophophora* (mean = $0.20 \pm 0.004 \text{ s.e.}$) (*t*-test, $P < 0.04$), which may reflect an underlying difference in modal chromosome numbers between the two subgenera (*Drosophila*: $2n = 12$; *Sophophora*: $2n = 8$) (Powell, 1997). Nevertheless, across the genus as a whole there were no significant correlations between genome size and either chromosome number or karyotypic index (a rough indicator of total chromosome size) with or without log transformation (all $P > 0.56$, $n = 63$). Phylogenetically corrected correlation analysis revealed a significant negative relationship between genome size and chromosome number, although this was marginally nonsignificant with one topology (Tree 1; $r = -0.30$, $P < 0.02$; Tree 2: $r = -0.23$, $P < 0.09$). No relationships were found between genome size and karyotypic index using either tree topology (Tree 1: $P > 0.43$; Tree 2: $P > 0.18$).

Finally, species were grouped according to biotic region of habitation following the designations provided by the BioSystematic Database of World Diptera (Afrotropical, Australasian/Oceanian, Nearctic, Neotropical, Oriental and Palearctic) (Figure 2). Species from the Afrotropical region had the smallest mean genome size in absolute terms, but overall there were no clear differences among groups (ANOVA, $P > 0.05$). Species

inhabiting multiple regions had significantly smaller mean genomes than the pooled average of the other regions (*t*-test, $P < 0.05$).

Phenotypic correlates

A significant positive correlation was found between haploid genome size and development time from egg to adult at 18 °C ($r = 0.48$, $P = 0.0004$, $n = 50$), indicating that at a constant temperature, species with smaller genome sizes develop more quickly than those with larger genomes (Figure 3a). Hierarchical taxonomic correlation analyses (Gregory, 2002a) showed that this relationship becomes increasingly strong at the subgroup ($r = 0.56$, $P < 0.008$, $n = 21$) and group ($r = 0.82$, $P < 0.03$, $n = 7$) levels. The relationship between genome size and development time remained significant when controlled for phylogenetic nonindependence of species data using either of the two tree topologies (Tree 1: $r = 0.34$, $P < 0.02$; Tree 2: $r = 0.35$, $P < 0.02$; Figure 3b).

Although data were available for a relatively small number of species, preliminary relationships were found between genome size and additional biological characteristics (Figure 4). A significant positive correlation was noted between genome size and post-eclosion time to sexual maturity in male flies ($r = 0.63$, $P < 0.02$, $n = 15$), but not in female flies ($P = 0.79$), the latter of which appears highly constrained in this character. A significant positive correlation was also observed between genome size and thorax length in both male flies ($r = 0.53$, $P < 0.005$, $n = 26$) and female flies ($r = 0.47$, $P < 0.02$, $n = 26$), and a significant positive association was found between genome size and sperm length ($r = 0.62$, $P < 0.02$, $n = 15$). Unlike with developmental duration, these relationships did not remain significant following phylogenetic correction (all $P > 0.19$), but this could reflect the sensitivity of such analyses to small sample sizes (Martins *et al.*, 2002).

Insights from 12 *Drosophila* genome sequences

The results of correlation analyses using data from the 12 sequenced *Drosophila* genomes are presented in Table 2. The number of predicted protein-coding genes per genome did not correlate with genome size, nor did the number of predicted pseudogenes. Although they did not vary greatly among species, the percentages of the genome composed of coding genes and introns were both inversely correlated with genome size, indicating that noncoding DNA amount is the driving force behind genome size diversity in this genus. In keeping with this, satellite DNA and (euchromatic) transposable element content were the strongest positive correlates of genome size.

Discussion

The drosophilid genome size data set

The present study more than doubles the available genome size data set to 74 species of *Drosophila*, including representatives of several more species groups, and broadens the coverage of the family Drosophilidae by including members of five additional genera (*Chymomyza*, *Hirtodrosophila*, *Samoaia*, *Scaptodrosophila* and *Zaprionus*). Drosophilids display genome sizes much smaller than the average of all currently available insect

Table 1 New flow cytometry-based haploid genome size estimates for 67 species from six genera in the family Drosophilidae

Genus	Species	Sex	n	Genome size (pg)	s.e.	Subgenus	Group	Subgroup
<i>Chymomyza</i>	<i>amoena</i>	Male	4	0.39	0.004		Fuscimana	
		Female	5	0.40	0.004			
<i>Chymomyza</i>	<i>pararufithorax</i>	Male	8	0.25	0.001		Procnemis	
		Female	7	0.29	0.001			
<i>Chymomyza</i>	<i>procnemis</i>	Male	4	0.29	0.003		Procnemis	
		Female	5	0.26	0.002			
<i>Drosophila</i>	<i>affinis</i>	Male	10	0.18	0.004	Sophophora	Obscura	Affinis
		Female	10	0.20	0.003			
<i>Drosophila</i>	<i>ambigua</i>	Male	7	0.18	0.005	Sophophora	Obscura	Obscura
		Female	5	0.19	0.008			
<i>Drosophila</i>	<i>ananassae</i>	Male	6	0.19	0.005	Sophophora	Melanogaster	Ananassae
		Female	9	0.20	0.007			
<i>Drosophila</i>	<i>azteca</i>	Male	5	0.19	0.006	Sophophora	Obscura	Affinis
		Female	5	0.20	0.008			
<i>Drosophila</i>	<i>barbarae</i>	Male	4	0.20	0.002	Sophophora	Obscura	Montium
		Female	2	0.20	0.002			
<i>Drosophila</i>	<i>biarmipes</i> ^a	Male	6	0.20	0.005	Sophophora	Melanogaster	Suzukii
		Female	5	0.20	0.007			
<i>Drosophila</i>	<i>biauraria</i>	Male	2	0.21	0.011	Sophophora	Obscura	Montium
		Female	4	0.24	0.012			
<i>Drosophila</i>	<i>bicornuta</i>	Male	3	0.21	0.014	Sophophora	Obscura	Montium
		Female	4	0.21	0.013			
<i>Drosophila</i>	<i>bifasciata</i>	Male	6	0.20	0.003	Sophophora	Obscura	Affinis
		Female	5	0.21	0.002			
<i>Drosophila</i>	<i>bipectinata</i>	Male	5	0.19	0.005	Sophophora	Melanogaster	Ananassae
		Female	6	0.21	0.007			
<i>Drosophila</i>	<i>birchii</i>	Male	3	0.19	0.001	Sophophora	Melanogaster	Montium
		Female	3	0.20	0.001			
<i>Drosophila</i>	<i>bunnanda</i>	Male	7	0.22	0.006	Sophophora	Melanogaster	Montium
		Female	7	0.23	0.004			
<i>Drosophila</i>	<i>diplacantha</i>	Male	5	0.23	0.005	Sophophora	Melanogaster	Montium
		Female	5	0.23	0.004			
<i>Drosophila</i>	<i>elegans</i>	Male	5	0.20	0.003	Sophophora	Melanogaster	Elegans
		Female	6	0.19	0.007			
<i>Drosophila</i>	<i>equinoxialis</i>	Male	7	0.25	0.005	Sophophora	Willistoni	Willistoni
		Female	7	0.25	0.006			
<i>Drosophila</i>	<i>ercepeae</i>	Male	5	0.22	0.004	Sophophora	Melanogaster	Ananassae
		Female	7	0.23	0.006			
<i>Drosophila</i>	<i>erecta</i>	Male	5	0.16	0.005	Sophophora	Melanogaster	Melanogaster
		Female	6	0.16	0.002			
<i>Drosophila</i>	<i>eugracilis</i>	Male	5	0.23	0.002	Sophophora	Melanogaster	Eugracilis
		Female	5	0.23	0.004			
<i>Drosophila</i>	<i>ficuspila</i>	Male	6	0.18	0.002	Sophophora	Melanogaster	Ficuspila
		Female	5	0.19	0.003			
<i>Drosophila</i>	<i>greeni</i>	Male	4	0.20	0.008	Sophophora	Melanogaster	Montium
		Female	4	0.20	0.008			
<i>Drosophila</i>	<i>hydei</i>	Male	4	0.20	0.001	Drosophila	Repleta	Hydei
		Female	6	0.22	0.001			
<i>Drosophila</i>	<i>jambulina</i>	Male	5	0.21	< 0.001	Sophophora	Melanogaster	Montium
		Female	5	0.21	< 0.001			
<i>Drosophila</i>	<i>kikkawai</i>	Male	7	0.21	0.005	Sophophora	Melanogaster	Montium
		Female	5	0.21	0.001			
<i>Drosophila</i>	<i>lucipennis</i>	Male	5	0.25	0.008	Sophophora	Melanogaster	Suzukii
		Female	6	0.29	0.011			



Table 1 Continued

Genus	Species	Sex	n	Genome size (pg)	s.e.	Subgenus	Group	Subgroup
<i>Drosophila</i>	<i>lutescens</i>	Male	6	0.21	0.003	Sophophora	Melanogaster	Takahashii
		Female	5	0.22	0.005			
<i>Drosophila</i>	<i>malerkotliana</i>	Male	8	0.20	0.008	Sophophora	Melanogaster	Ananassae
		Female	5	0.21	0.010			
<i>Drosophila</i>	<i>mauritiana</i>	Male	3	0.16	0.001	Sophophora	Melanogaster	Melanogaster
		Female	2	0.16	0.005			
<i>Drosophila</i>	<i>mauritiana</i> Pet	Male	1	0.15	< 0.001	Sophophora	Melanogaster	Melanogaster
		Female	3	0.16	0.003			
<i>Drosophila</i>	<i>mauritiana</i> Rob	Male	1	0.14	< 0.001	Sophophora	Melanogaster	Melanogaster
		Female	3	0.16	0.004			
<i>Drosophila</i>	<i>mauritiana</i> wht	Male	1	0.15	< 0.001	Sophophora	Melanogaster	Melanogaster
		Female	3	0.16	0.001			
<i>Drosophila</i>	<i>melanogaster</i> CS	Male	7	0.18	0.004	Sophophora	Melanogaster	Melanogaster
		Female	22	0.18	0.002			
<i>Drosophila</i>	<i>melanogaster</i> Neth	Male	1	0.16	< 0.001	Sophophora	Melanogaster	Melanogaster
		Female	3	0.18	0.007			
<i>Drosophila</i>	<i>melanogaster</i> Rale	Male	1	0.17	< 0.001	Sophophora	Melanogaster	Melanogaster
		Female		N/A	N/A			
<i>Drosophila</i>	<i>melanogaster</i> W Af	Male	1	0.17	< 0.001	Sophophora	Melanogaster	Melanogaster
		Female	3	0.18	0.001			
<i>Drosophila</i>	<i>mimetica</i>	Male	5	0.21	0.004	Sophophora	Melanogaster	Suzukii
		Female	6	0.21	0.003			
<i>Drosophila</i>	<i>miranda</i>	Male	5	0.18	0.003	Sophophora	Obscura	Obscura
		Female	6	0.18	0.004			
<i>Drosophila</i>	<i>mojavensis</i>	Male	12	0.17	0.001	Drosophila	Repleta	Mulleri
		Female	7	0.17	0.003			
<i>Drosophila</i>	<i>narragansett</i>	Male	5	0.20	0.002	Sophophora	Obscura	Affinis
		Female	6	0.21	0.004			
<i>Drosophila</i>	<i>nebulosa</i>	Male	5	0.18	0.003	Sophophora	Willistoni	Bocainensis
		Female	5	0.19	0.013			
<i>Drosophila</i>	<i>nigrens</i> ^b	Male	6	0.23	0.003	Sophophora	Melanogaster	Ananassae
		Female	4	0.23	0.006			
<i>Drosophila</i>	<i>orena</i>	Male	5	0.28	0.003	Sophophora	Melanogaster	Melanogaster
		Female	5	0.28	0.002			
<i>Drosophila</i>	<i>pallidosa</i>	Male	5	0.19	0.005	Sophophora	Melanogaster	Ananassae
		Female	5	0.20	0.003			
<i>Drosophila</i>	<i>pallidosa</i> -like	Male	5	0.18	0.005	Sophophora	Melanogaster	Ananassae
		Female	5	0.19	0.003			
<i>Drosophila</i>	<i>parabipectinata</i>	Male	5	0.21	0.003	Sophophora	Melanogaster	Ananassae
		Female	5	0.21	0.003			
<i>Drosophila</i>	<i>paralutea</i>	Male	6	0.22	0.006	Sophophora	Melanogaster	Takahashii
		Female	6	0.23	0.014			
<i>Drosophila</i>	<i>paulistorum</i>	Male	7	0.24	0.008	Sophophora	Willistoni	Willistoni
		Female	5	0.23	0.005			
<i>Drosophila</i>	<i>persimilis</i>	Male	5	0.17	0.004	Sophophora	Obscura	Obscura
		Female	5	0.20	0.008			
<i>Drosophila</i>	<i>phaeopleura</i>	Male	5	0.19	0.001	Sophophora	Melanogaster	Ananassae
		Female	5	0.20	0.004			
<i>Drosophila</i>	<i>prostipennis</i>	Male	7	0.23	0.008	Sophophora	Melanogaster	Takahashii
		Female	5	0.23	0.007			
<i>Drosophila</i>	<i>pseudoananassae</i>	Male	5	0.22	0.003	Sophophora	Melanogaster	Ananassae
		Female	5	0.23	0.006			
<i>Drosophila</i>	<i>pseudoobscura</i>	Male	5	0.16	0.005	Sophophora	Obscura	Obscura
		Female	6	0.17	0.007			

Table 1 Continued

Genus	Species	Sex	n	Genome size (pg)	s.e.	Subgenus	Group	Subgroup
<i>Drosophila</i>	<i>pseudotakahashii</i>	Male	5	0.20	0.002	Sophophora	Melanogaster	Takahashii
		Female	6	0.21	0.002			
<i>Drosophila</i>	<i>sechellia</i>	Male	8	0.18	0.004	Sophophora	Melanogaster	Melanogaster
		Female	6	0.17	0.005			
<i>Drosophila</i>	<i>serrata</i>	Male	6	0.21	<0.001	Sophophora	Melanogaster	Montium
		Female	3	0.22	<0.001			
<i>Drosophila</i>	<i>simulans</i> Aust	Male	1	0.15	<0.001	Sophophora	Melanogaster	Melanogaster
		Female	3	0.17	0.001			
<i>Drosophila</i>	<i>simulans</i> S Fr	Male	4	0.15	0.002	Sophophora	Melanogaster	Melanogaster
		Female	3	0.15	0.003			
<i>Drosophila</i>	<i>simulans</i> wht	Male	1	0.15	<0.001	Sophophora	Melanogaster	Melanogaster
		Female	3	0.16	0.003			
<i>Drosophila</i>	<i>takahashii</i>	Male	5	0.19	0.006	Sophophora	Melanogaster	Takahashii
		Female	5	0.21	0.001			
<i>Drosophila</i>	<i>teissieri</i>	Male	5	0.16	0.001	Sophophora	Melanogaster	Melanogaster
		Female	4	0.17	0.002			
<i>Drosophila</i>	<i>tolteca</i>	Male	5	0.17	0.001	Sophophora	Obscura	Affinis
		Female	5	0.18	<0.001			
<i>Drosophila</i>	<i>varians</i>	Male	4	0.16	0.005	Sophophora	Melanogaster	Ananassae
		Female	6	0.17	0.003			
<i>Drosophila</i>	<i>virilis</i>	Male	5	0.34	0.001	<i>Drosophila</i>	Virilis	Virilis
		Female	17	0.34	0.001			
<i>Drosophila</i>	<i>yakuba</i>	Male	7	0.17	0.005	Sophophora	Melanogaster	Melanogaster
		Female	7	0.17	0.002			
<i>Hirtodrosophila</i>	<i>pictiventris</i> ^c	Male	5	0.14	0.005			
		Female	7	0.16	0.009			
<i>Samoaia</i>	<i>leonensis</i>	Male	5	0.27	<0.001			
		Female	7	0.27	0.001			
<i>Scaptodrosophila</i>	<i>brooksae</i> ^d	Male	6	0.28	0.005	Scaptodrosophila	Victoria	
		Female	5	0.27	0.005			
<i>Scaptodrosophila</i>	<i>latifasciaeformis</i> ^e	Male	5	0.21	0.001	Scaptodrosophila	Latifasciaeformis	
		Female	7	0.20	0.001			
<i>Scaptodrosophila</i>	<i>lebanonensis</i> ^f	Male	26	0.22	0.003	Scaptodrosophila	Coracina	
		Female	26	0.22	0.006			
<i>Zaprionus</i>	<i>camarounensis</i>	Male	5	0.20	0.001	Zaprionus	Armatus	Vittager
		Female	5	0.20	0.001			
<i>Zaprionus</i>	<i>davidi</i>	Male	4	0.17	0.001	Zaprionus	Armatus	Vittager
		Female	4	0.17	<0.001			
<i>Zaprionus</i>	<i>sepsoides</i>	Male	6	0.22	0.001	Zaprionus	Armatus	Tuberculatus
		Female	6	0.22	0.001			
<i>Zaprionus</i>	<i>tuberculatis</i>	Male	5	0.20	0.001	Zaprionus	Armatus	Tuberculatus
		Female	5	0.20	0.001			

Separate estimates and numbers of specimens used (*n*) are provided for males and females. Genome sizes are presented in picograms (1 pg = 10⁻¹² g = 978 Mbp).

^aInitially identified as *Drosophila rajasekari*.

^bInitially identified as *Drosophila pseudoananassae nigrens*.

^cFormerly classified as *Drosophila pictiventris*.

^dFormerly classified as *Drosophila brooksae*.

^eFormerly classified as *Drosophila latifasciaeformis*.

^fIncludes samples from populations formerly classified as *Drosophila dimorpha*, *D. lebanonensis casteeli*, *D. lebanonensis lebanonensis*, *D. stonei* and *D. pattersoni* (see Supplementary Material for details).

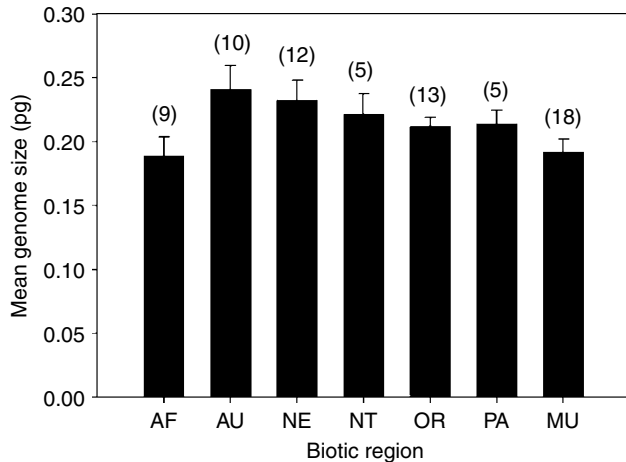


Figure 2 Mean genome sizes for species of *Drosophila* grouped according to biotic region, following the designations provided by the BioSystematic Database of World Diptera. AF, Afrotropical; AU, Australasian/Oceanian; NE, Nearctic; NT, Neotropical; OR, Oriental; PA, Palearctic; MU, multiple regions. There were no statistical differences between species inhabiting the different regions, nor between species inhabiting multiple regions and those restricted to a single region. The number of species included from each biotic region is given in parentheses. Error bars represent standard error.

data (0.22 vs 1.56 pg). Their genomes are also significantly smaller than the current average for Diptera (0.61 pg), and appear to be less than 25% as large as those of the next best-studied group of true flies, the mosquitoes (family Culicidae) (Gregory, 2008).

Since the late 1940s, genome size has been considered a (mostly) constant characteristic within species, and limited intraspecific variation is a standard assumption in measurements and comparative analyses of genome size. Modern methods of genome size estimation are sufficiently rapid and precise to allow investigations of intraspecific variation, however this is only true if best-practice methods are implemented as there are numerous sources of error that could compromise such analyses (Greilhuber, 1998).

Following the conventions of the botanical literature (in which the issue has been far more thoroughly investigated than in animals), only genome size differences within species that are based on demonstrable chromosomal or sequence-level effects are sufficiently convincing to qualify as 'orthodox' intraspecific variation (Greilhuber, 1998; Bennett and Leitch, 2005). Some species of *Drosophila* studied in this and other reports exhibit orthodox intraspecific variation on the basis of the chromosomal sex determination system(s) in the genus. Repeatable variation among highly inbred laboratory strains is also apparent in some species (Table 1; Bosco *et al.*, 2007), but it is unclear to what extent this reflects the situation in nature. Some studies have been conducted to investigate this in Diptera (Rao and Rai, 1987; Black and Rai, 1988; Kumar and Rai, 1990; Vieira *et al.*, 2002), but to date best-practice methods in combination with extensive sampling have yet to be brought to bear on the question.

Phenotypic consequences of genome size variation

On the basis of the data available at the time, Powell (1997, p. 303) argued that:

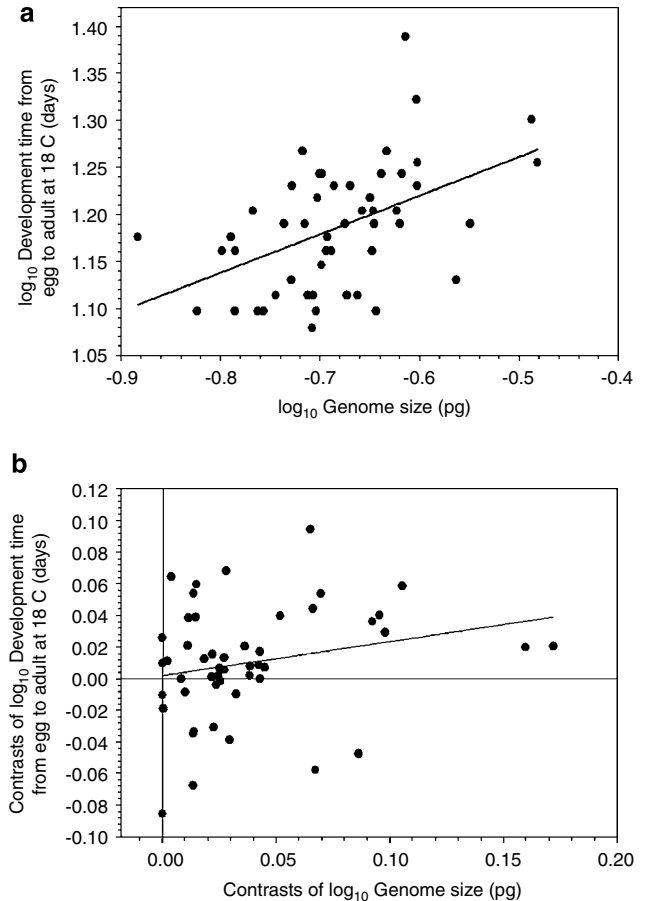


Figure 3 The relationship between genome size and duration of development from egg to adult at 18 °C in 50 species of the genus *Drosophila*. (a) The relationship is significant at the species level using Pearson's correlation ($r=0.48$, $P=0.0004$). It also persists when the analysis is conducted at the subgroup ($r=0.56$, $P<0.008$, $n=21$) and group ($r=0.82$, $P<0.03$, $n=7$) levels. (b) The relationship is also significant with phylogenetically independent contrasts (data shown are from Tree 1: $r=0.34$, $P<0.02$; results were very similar using Tree 2: $r=0.35$, $P<0.02$; see Supplementary Material for details about the alternate trees).

One of the major adaptive hypotheses about genome size is that it is related to rate of cell division and thus development time; furthermore, larger genomes are associated with larger cell volume, which often increases the size of the organism. This explanation does not seem to hold for *Drosophila*. Species with relatively large genomes like *D. virilis* and *D. funebris* are quite large and slowly developing, but *D. willistoni*, with [a] genome size nearly identical to *D. funebris*, is a rapidly developing small species. *D. hydei* and its close relatives including *D. neohydei*, are nearly as large as *D. virilis*. Finally, *D. nasutoides* is not particularly large or slowly developing, yet has a huge genome relative to other drosophilids.

In contrast to this assessment, the present study revealed a strong positive relationship between genome size and the temperature-controlled duration of development that persists when the data are analyzed at multiple taxonomic levels or corrected for phylogenetic nonindependence (Figure 3). Moreover, the pattern observed

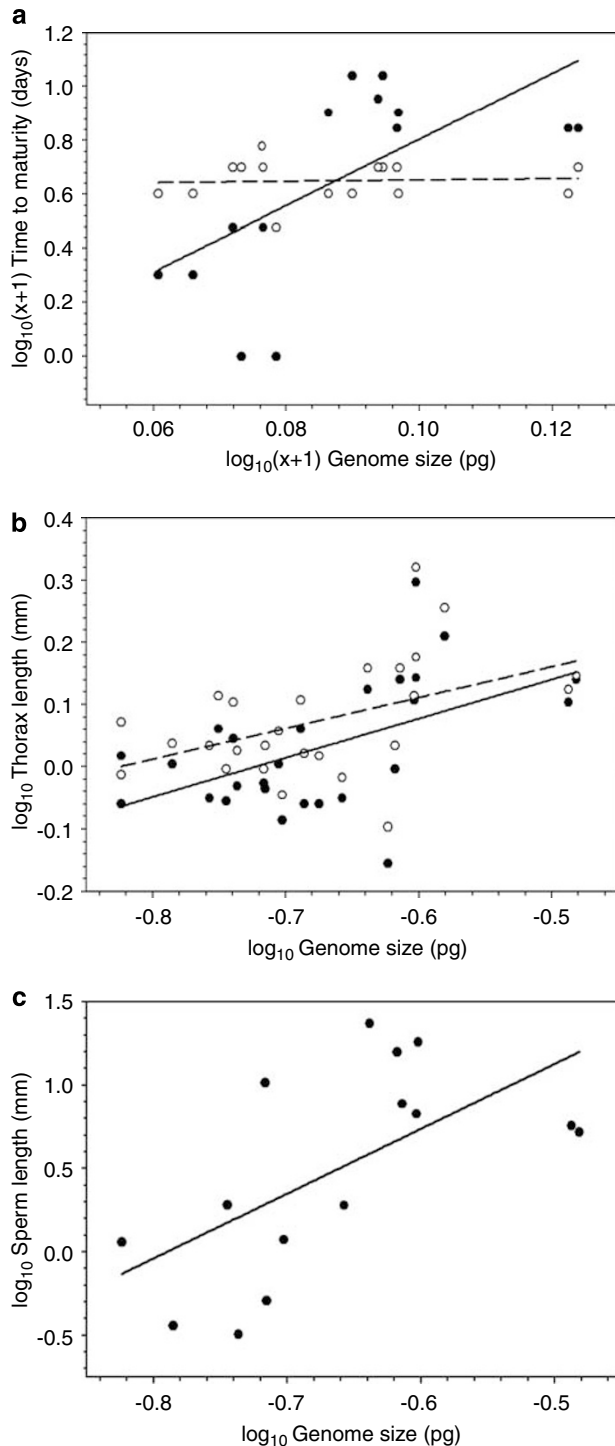


Figure 4 Relationships between genome size and various phenotypic parameters in the genus *Drosophila*. (a) Post-eclosion time to sexual maturity versus genome size in male flies (●, solid line; $r=0.63$, $P<0.02$) and female flies (○, dashed line; $P=0.79$). Eclosion time data were $\log(x+1)$ -transformed as male individuals of some species have a value of 0 for this parameter. (b) Thorax length versus genome size in male flies (●, solid line; $r=0.53$, $P<0.005$) and female flies (○, dashed line; $r=0.47$, $P<0.02$). (c) Sperm length versus genome size ($r=0.62$, $P<0.02$). Note that, unlike the correlation with development time (Figure 3), these relationships generally did not persist following phylogenetic correction.

using the present data set indicates that this relationship is in fact fully consistent with every proposed counterexample listed by Powell (1997). Thus, *D. virilis* development is indeed comparatively slow (20 days at 18 °C), and while *D. funebris* data were not available at 18 °C, its development is likewise relatively protracted when data from higher temperatures are considered (18 days at 21 °C). *D. willistoni* is not rapidly developing compared to other species in the present data set, rather it exhibits both an above-average genome size (0.24 vs 0.21 pg; Powell, 1997) and slightly slower than average development (15.5 vs 15.1 days). *D. neohydei* was not included in the present analysis as developmental data were not provided in the source data set, but *D. hydei* was not an outlier in the correlation. *D. nasutooides*, with a reported (but somewhat questionable) genome size estimate of 0.40 pg, would be expected to exhibit slow development, and although data were not available for development at 18 °C, this species has been reported to take 19 days at 20 °C—this is slower than all other species in the data set assessed at 18 °C, with the notable exception of *D. virilis*. In addition, it appears that post-eclosion time to sexual maturity may be related to genome size in male (but not female) flies.

Although sample sizes need to be expanded considerably before any reliable conclusions can be drawn, the present results are suggestive of a positive association between genome size and body size (Figure 4). This is consistent with earlier reports of positive correlations between wing cell area and measures of body size, including thorax length, wing length, whole body length, fore-tibia length and body mass in a sample of six Hawaiian and two non-Hawaiian species of *Drosophila* (Kacmarczyk and Craddock, 2000). Cell size and body size were also reported to be correlated positively with genome size in the same species (Craddock *et al.*, 2000), although details of this analysis were not provided.

At first sight it may seem remarkable that phenotypic relationships can be determined at all within the narrow absolute range of genome sizes exhibited by these insects, especially considering the additional error that is introduced by compiling phenotypic and genomic data from several different studies. However, in relative terms the reported range in *Drosophila* is at least 2.5-fold, and it should be borne in mind that correlations with cell size and metabolic rate can be found within birds, whose genome sizes vary by a smaller relative margin (~2.2-fold; Gregory, 2002b). It would seem that in organisms like *Drosophila*, which reside near the extreme of diminutive body size and complex but rapid development, even small absolute variations in DNA content may be biologically relevant.

Sources and significance of genome size diversity

Large-scale features such as chromosome number do not appear to be related to genome size diversity in the genus *Drosophila*, but there is mounting evidence that several subgenomic components do scale with total genome size. For example, it has been noted that *D. virilis* has both longer introns (Moriyama *et al.*, 1998) and lengthier stretches of microsatellites (Schlötterer and Harr, 2000) than *D. melanogaster*, in keeping with its roughly twofold larger genome. Rough comparisons between *D. melanogaster* and its close relative *D. simulans*

Table 2 Results of correlation analyses against haploid genome size (pg) using data from the 12 sequenced *Drosophila* genomes

	Pearson's		Spearman's rank		PICs	
	r	P	ρ	P	r	P
Transposable element content (%) ^a	0.363	0.25	0.522	0.084	0.649	0.02
Intron content (%) ^a	-0.977	<0.0001	-0.946	<0.0001	-0.857	0.0004
Protein-coding gene content (%) ^a	-0.817	0.001	-0.827	0.0017	-0.797	0.002
Satellite DNA content (%) ^b	0.471	0.12	0.541	0.0707	0.686	0.014
Number of pseudogenes ^c	-0.367	0.27	-0.132	0.694	0.542	0.085
Number of genes ^c	-0.239	0.45	-0.112	0.733	0.046	0.89
Number of genes ^a	-0.303	0.34	-0.347	0.266	0.199	0.53
<i>DINE-I</i> copy number ^d	0.163	0.61	0.074	0.82	0.033	0.91

Pearson's correlation analyses were performed using log-transformed data, whereas Spearman's rank correlations employed untransformed data. Phylogenetically independent contrasts (PICs) were performed on log-transformed data using the tree topology and branch lengths in Heger and Ponting (2007). Transposable element contents are based on sequence assemblies, and therefore primarily reflect euchromatic content (*Drosophila* 12 Genomes Consortium, 2007).

Genome size data as in Table 1, supplemented by estimates from Bosco *et al.* (2007).

^a*Drosophila* 12 Genomes Consortium (2007).

^bBosco *et al.* (2007).

^cHeger and Ponting (2007).

^dYang and Barbash (2008).

also suggest that larger-genomed species may contain more transposable elements (Biémont and Cizeron, 1999; Vieira *et al.*, 2002; Vieira and Biémont, 2004). Using data from the 12 *Drosophila* genomes, it appears that micro-satellite and transposable element content, at least, probably do scale with genome size across the genus (Table 2), as do the total and pseudogenized copy numbers of chemoreceptor genes (Gardiner *et al.*, 2008). However, copy number of the *DINE-I* element in particular does not scale with genome size (Table 2).

With regard to transposable elements, it has been argued that the stress inherent in the invasion of new environments triggers their spread, which generates significant genome size variation within, and later presumably among, species of *Drosophila* (Biémont *et al.*, 2001; Vieira *et al.*, 1999, 2002; Nardon *et al.*, 2005). Although this is an interesting possibility, caution is warranted when extrapolating to larger taxonomic scales. Notably, species inhabiting multiple biotic regions exhibit smaller, not larger, average genome sizes than those of species restricted to a single region. Moreover, the climatic differences among biotic regions themselves, which surely differ in their degrees of environmental 'stress', do not appear to contribute substantially to the pattern of genome size diversity in the genus (Figure 2).

Recent analyses of introns in *D. melanogaster* have suggested that these sequences are under selective constraints even though they do not encode any protein products (Andolfatto, 2005; Marais *et al.*, 2005). A subsequent comparison of the *D. melanogaster* and *D. simulans* genomes indicated that >50% of point mutations in both intronic and intergenic elements are removed by selection (Halligan and Keightley, 2006). Such high levels of selective constraint could be suggestive of function, for example relating to gene regulation, and it has been argued that 'there is now increasing evidence that the *Drosophila* genome may be highly compact and contain very little nonfunctional DNA' (Halligan and Keightley, 2006, p. 881). If so, then the maintenance of noncoding DNA in *Drosophila* genomes could be explained in terms of adaptive benefit

at the organismal level, rather than through (or perhaps in addition to) mutation pressure exerted by selfish elements. However, this does not provide an answer as to why there is a general paucity of noncoding DNA in this genus, even as compared to other Diptera.

As Powell noted in 1997 (p. 304), 'an adaptive basis of variation in genome size in *Drosophila* remains to be documented', and it is therefore perhaps not surprising that many recent discussions have focused on neutralist models of genome size change (Petrov, 2002; Boulesteix *et al.*, 2006; Halligan and Keightley, 2006). In contrast, the present study provides evidence that genome size is indeed of biological significance in *Drosophila* through its effects on developmental rate (Figure 3) and perhaps also on body size and/or sperm size (Figure 4). Thus, there may be pressures to reduce genome size due to organismal effects, which are limited by genome level constraints on the deletion of a moderate quantity of functional non-genic DNA.

Concluding remarks

The broad view of genomic diversity that is being adopted in *Drosophila* research by expanding beyond the core set of model species is a very welcome development. As a result, there is little doubt that flies in the family Drosophilidae will feature prominently in future studies of both genome sequences and genome sizes, and that the integration of these two levels of analysis will provide many mutually enlightening insights. In combination, such studies will make it possible to decipher the complex relationships that link the genome and the phenotype, which clearly extends beyond the influence of individual protein-coding genes. As a start, the role of (some) noncoding sequences in structural and regulatory capacities is becoming more widely appreciated, thanks in part to work with *Drosophila*, and the results of the present study suggest that its bulk amount is also important through effects on traits of fundamental biological importance such as development and morphology. Many challenging

questions remain with regard to the mechanisms of gain and loss, potential functions and phenotypic consequences of noncoding DNA, but as they enter their second century of undisputed preeminence in genetics research, these humble flies are poised to contribute mightily in the quest to answer each of them.

Acknowledgements

This study was supported by an NSF grant to JSJ and an NSERC Discovery Grant to TRG. We thank Sarah Adamowicz for providing invaluable assistance with the assembly of phylogenetic trees and associated statistical analyses.

References

- Adams MD, Celniker SE, Holt RA *et al.* (195 co-authors) (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Andolfatto P (2005). Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.
- Ashburner M (2007). *Drosophila* genomes by the baker's dozen. *Genetics* **177**: 1263–1268.
- Ashburner M, Golic KG, Hawley RS (2005). *Drosophila: A Laboratory Handbook*, 2nd edn. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
- Bartolomé C, Charlesworth B (2006). Rates and patterns of chromosomal evolution in *Drosophila pseudoobscura* and *D. miranda*. *Genetics* **173**: 779–791.
- Beller M, Oliver B (2006). One hundred years of high-throughput *Drosophila* research. *Chromosome Res* **14**: 349–362.
- Bennett MD, Leitch IJ (2005). Genome size evolution in plants. In: Gregory TR (ed). *The Evolution of the Genome*. Elsevier: San Diego. pp 89–162.
- Bennett MD, Leitch IJ, Price HJ, Johnston JS (2003). Comparisons with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus ~25% larger than the *Arabidopsis* Genome Initiative estimate of ~125 Mb. *Annals Bot* **91**: 547–557.
- Biémont C, Cizeron G (1999). Distribution of transposable elements in *Drosophila* species. *Genetica* **105**: 43–62.
- Biémont C, Vieira C, Borie N (2001). Éléments transposables et évolution du génome d'une espèce invasive: le cas de *Drosophila simulans*. *Genet Select Evol* **33** (Suppl 1): S107–S120.
- Black WC, Rai KS (1988). Genome evolution in mosquitoes: intraspecific and interspecific variation in repetitive DNA amounts and organization. *Genet Res* **51**: 185–196.
- Bosco G, Campbell P, Leiva-Neto JT, Markow TA (2007). Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* **177**: 1277–1290.
- Boulesteix M, Weiss M, Biémont C (2006). Differences in genome size between closely related species: the *Drosophila melanogaster* species subgroup. *Mol Biol Evol* **23**: 162–167.
- Brookes M (2001). *Fly: The Unsung Hero of 20th-Century Science*. HarperCollins: New York.
- Casola C, Lawing AM, Betrán E, Feschotte C (2007). PIF-like transposons are common in *Drosophila* and have been repeatedly domesticated to generate new host genes. *Mol Biol Evol* **24**: 1872–1888.
- Celniker SE, Wheeler DA, Kronmiller B *et al.* (32 co-authors) (2002). Finishing a whole-genome shotgun: release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biol* **3**: research0079.0071–0079.0014.
- Clark A, Gibson G, Kaufman T, Myers E, O'Grady P (2003). Proposal for *Drosophila* as a model system for comparative genomics. www.genome.gov/Pages/Research/Sequencing/SeqProposals/Drosophila.pdf.
- Clayton FE, Guest WC (1986). Overview of chromosomal evolution in the family Drosophilidae. In: Ashburner M, Carson HL, Thompson JN (eds). *The Genetics and Biology of Drosophila*, vol. 3e Academic Press: Orlando, FL. pp 1–38.
- Clayton FE, Wheeler MR (1975). A catalog of *Drosophila* metaphase chromosome configurations. In: King RC (ed). *Handbook of Genetics*, vol. 3. Plenum Press: NY. pp 471–512.
- Craddock EM, Kacmarczyk T, Dawley RM (2000). The roles of genome size and cell size in body size variation among Hawaiian *Drosophila*. Abstract, 41st Annual Drosophila Research Conference, Pittsburg, PA, 20–26 March.
- Crosby MA, Goodman JL, Strelets VB, Zhang P, Gelbart WM, The FlyBase Consortium (2007). FlyBase: genomes by the dozen. *Nucleic Acids Res* **35** (Suppl): D486–D491.
- Demerec M, Kaufman BP (1996). *Drosophila Guide*, 10th edn. Carnegie Institution of Washington: Washington, DC.
- DeSalle R, Gregory TR, Johnston JS (2005). Preparation of samples for comparative studies of arthropod chromosomes: visualization, *in situ* hybridization, and genome size estimation. *Methods Enzymol* **395**: 460–488.
- Drosophila* 12 Genomes Consortium (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.
- Felsenstein J (1985). Phylogenies and the comparative method. *Am Nat* **125**: 1–15.
- Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG (2008). *Drosophila* chemoreceptor gene evolution: selection, specialization and genome size. *Mol Ecol* **17**: 1648–1657.
- Gilbert DG (2007). DroSpeGe: rapid access database for new *Drosophila* species genomes. *Nucleic Acids Res* **35** (Suppl): D480–D485.
- Gregory TR (2002a). Genome size and developmental parameters in the homeothermic vertebrates. *Genome* **45**: 833–838.
- Gregory TR (2002b). A bird's-eye view of the C-value enigma: genome size, cell size, and metabolic rate in the class Aves. *Evolution* **56**: 121–130.
- Gregory TR (2005a). Synergy between sequence and size in large-scale genomics. *Nat Rev Genet* **6**: 699–708.
- Gregory TR (2005b). Genome size evolution in animals. In: Gregory TR (ed). *The Evolution of the Genome*. Elsevier: San Diego. pp 3–87.
- Gregory TR (2008). Animal genome size database. <http://www.genomesize.com>.
- Greilhuber J (1998). Intraspecific variation in genome size: a critical reassessment. *Annals Bot* **82** (Suppl A): 27–35.
- Halligan DL, Keightley PD (2006). Ubiquitous selective constraints in the *Drosophila* genome revealed by a genome-wide interspecies comparison. *Genome Res* **16**: 875–884.
- Heger A, Ponting CP (2007). Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. *Genome Res* **17**: 1837–1849.
- Henderson DS (2003). *Drosophila Cytogenetics Protocols*. Humana Press: Totowa, NJ.
- Kacmarczyk T, Craddock EM (2000). Cell size is a factor in body size variation among Hawaiian and nonHawaiian species of *Drosophila*. *Drosophila Info Serv* **83**: 144–148.
- Kumar A, Rai KS (1990). Intraspecific variation in nuclear DNA content among world populations of a mosquito, *Aedes albopictus* (Skuse). *Theor Appl Genet* **79**: 748–752.
- Levine MT, Jones CD, Kern AD, Lindfors HA, Begun DJ (2006). Novel genes derived from noncoding DNA in *Drosophila melanogaster* are frequently X-linked and exhibit testis-biased expression. *Proc Natl Acad Sci USA* **103**: 9935–9939.
- Maddison WP, Maddison DR (2006). Mesquite: a modular system for evolutionary analysis, version 1.11. <http://mesquiteproject.org>.
- Marais G, Nouvellet P, Keightley PD, Charlesworth B (2005). Intron size and exon evolution in *Drosophila*. *Genetics* **170**: 481–485.
- Markow TA, O'Grady PM (2006). *Drosophila: A Guide to Species Identification and Use*. Elsevier: Amsterdam.

- Martins EP, Diniz-Filho JAF, Housworth EA. (2002). Adaptive constraints and the phylogenetic comparative method: a computer simulation test. *Evolution* **56**: 1–13.
- Midford PE, Garland T, Maddison WP (2002). PDAP:PDTree package for Mesquite, version 1.00. http://mesquiteproject.org/pdap_mesquite/.
- Moreteau B, Gibert P, Pétavy G, Moreteau J-C, Huey RB, David JR (2003). Morphological evolution in a *Drosophila* clade: the *Drosophila obscura* group. *J Zool Syst Evol Res* **41**: 64–71.
- Moriyama EN, Petrov DA, Hartl DL (1998). Genome size and intron size in *Drosophila*. *Mol Biol Evol* **15**: 770–773.
- Nardon C, Deceliere G, Loevenbruck C, Weiss M, Vieira C, Biémont C (2005). Is genome size influenced by colonization of new environments in dipteran species? *Mol Ecol* **14**: 869–878.
- Petrov DA (2002). DNA loss and evolution of genome size in *Drosophila*. *Genetica* **115**: 81–91.
- Pitnick S, Markow TA, Spicer G (1995). Delayed male maturity is a cost of producing large sperm in *Drosophila*. *Proc Natl Acad Sci USA* **92**: 10614–10618.
- Powell JR (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press: New York.
- Purvis A, Garland T (1993). Polytomies in comparative analyses of continuous characters. *Syst Biol* **42**: 569–575.
- Rao PN, Rai KS (1987). Inter and intraspecific variation in nuclear DNA content in *Aedes* mosquitoes. *Heredity* **59**: 253–258.
- Richards S, Liu Y, Bettencourt BR *et al.* (52 co-authors) (2005). Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res* **15**: 1–18.
- Schlötterer C, Harr B (2000). *Drosophila virilis* has long and highly polymorphic microsatellites. *Mol Biol Evol* **17**: 1641–1646.
- Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW *et al.* (46 co-authors) (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**: 219–232.
- Sullivan W, Ashburner M, Hawley RS (2000). *Drosophila Protocols*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
- Vieira C, Lepetit D, Dumont S, Biémont C (1999). Wake up of transposable elements following *Drosophila simulans* worldwide colonization. *Mol Biol Evol* **16**: 1251–1255.
- Vieira C, Biémont C (2004). Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. *Genetica* **120**: 115–123.
- Vieira C, Nardon C, Arpin C, Lepetit D, Biémont C (2002). Evolution of genome size in *Drosophila*. Is the invader's genome being invaded by transposable elements?. *Mol Biol Evol* **19**: 1154–1161.
- Yang H-P, Barbash DA (2008). Abundant and species-specific *DINE-I* transposable elements in 12 *Drosophila* genomes. *Genome Biol* **9**: R39.31–R39.14.
- Yasuhara JC, Wakimoto BT (2006). Oxymoron no more: the expanding world of heterochromatic genes. *Trends Genet* **22**: 330–338.
- Zacharias H, Hennig W, Leoncini O (1982). Microspectrophotometric comparison of the genome sizes of *Drosophila hydei* and some related species. *Genetica* **58**: 153–157.

Supplementary Information accompanies the paper on *Heredity* website (<http://www.nature.com/hdy>)