

Mitochondrial genome size variation in New World and Old World populations of *Drosophila melanogaster*

JP Townsend¹ and DM Rand

Department of Ecology and Evolutionary Biology, Brown University, Providence, RI 02912, USA

Drosophila melanogaster originated in Africa, spread to Europe and Asia, and is believed to have colonized the New World in the past few hundred years. Levels of genetic variation are typically reduced in New World populations, consistent with a founder event following range expansion out of Africa and the Old World. We describe the patterns of mtDNA length variation within and among several populations of *Drosophila melanogaster* from the Old and New World. MtDNA length variation is due to insertion and deletion of tandem repeats in the control region (D-loop) of *D. melanogaster* mitochondrial genome. The distinct mutational dynamics of this system provide an opportunity to compare the patterns of variation in this marker to those of other markers with different mutational pressures and linkage relationships. The data show significantly more length variation in African and Asian samples than in New World

samples. New World samples also show more pronounced skew of the length distribution. Our results are distinct from an earlier study that showed significantly higher levels of length variation and heteroplasmy. The level of heteroplasmy is highly correlated with the number of years that samples have been maintained in laboratory culture, suggesting that relaxed selection in small populations permits the accumulation of mtDNA length variation and heteroplasmy. Together, the data indicate that mtDNA length variants retain a signature of founder events and selection, and suggest that further investigation into the mutation-selection dynamics of the D-loop region of mtDNA would provide a distinct and informative marker for analysis of the recent history of populations.

Heredity (2004) 93, 98–103, advance online publication, 12 May 2004; doi:10.1038/sj.hdy.6800484

Keywords: population genetics; genome size; mtDNA; genetic drift; heteroplasmy; mutation accumulation

Introduction

An extensive literature exists on the patterns of genetic variation within and among populations of *Drosophila melanogaster*. Like *Homo sapiens*, *D. melanogaster* has a molecular signature of population expansion from Africa (Hale and Singh, 1991; Aquadro *et al.*, 2001). Population samples from Africa show abundant genetic variation, while samples from non-African populations carry a subset of that variation (Begun and Aquadro, 1995b; Moriyama and Powell, 1996). Datasets on mitochondrial DNA (mtDNA) restriction site variation (Hale and Singh, 1991) and sequence polymorphism (Rand *et al.*, 1994) are consistent with an African origin for *D. melanogaster*. Comparisons between X-linked and autosomal loci reveal conflicting patterns of variation in African *vs* non-African populations, suggesting that both founder events and selection have shaped patterns of polymorphism in non-African populations observed on the X (Begun and Aquadro, 1995a; Moriyama and Powell, 1996; Andolfatto, 2001; Caracristi and Schlotterer, 2003). Clearly, the most effective means of understanding

population structure is to compare patterns of genetic variation with multiple genetic markers. A more complete picture is obtained when markers with different evolutionary dynamics can be contrasted.

Mitochondrial genome size variation has not been widely used as a marker for population structure due to its peculiar mutational dynamics. Mitochondrial genome size in *D. melanogaster* varies within populations almost exclusively by changes in the length of the control region (Solignac *et al.*, 1986), which is believed to contain the origin of transcription and translation. The control region is characterized by variable numbers of tandem repeats (Solignac *et al.*, 1986; Lewis *et al.*, 1994). Large-scale slipped-strand mispairing during replication is a likely model to account for the high rate of insertion and deletion in this region (Buroker *et al.*, 1990; Hayasaka *et al.*, 1991). Tandem repeats in the control region (D-loop) of animal mtDNAs are commonly associated with heteroplasmy for length variants (Rand, 1993; Lunt *et al.*, 1998). Because drift among germ cell lineages should act to remove heteroplasmic variation, high levels of heteroplasmy in D-loop regions imply a high mutation rate for insertion and deletion (indel) of repeats. Estimates of this indel rate are rather indirect and may vary widely across species, but appear to be higher than mutation rates for microsatellites (Schug *et al.*, 1997; Rubinsztein *et al.*, 1999). For this reason, quantification of the population structure of mtDNA length variation should provide focused information on recent population history compared to other types of molecular polymorphisms.

Correspondence: DM Rand, Department of Ecology and Evolutionary Biology, 69 Brown Street, Brown University, Providence, Rhode Island 02912 USA. E-mail: David_Rand@brown.edu

¹Current address: Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720, USA.

Received 26 November 2003; accepted 22 February 2004; published online 12 May 2004

Hale and Singh (1986) demonstrated that there is mtDNA length variation and heteroplasmy in a world-wide sample of *D. melanogaster*, but did not attribute a larger component of the variation to any geographic region. In order to assess how mtDNA length variation is distributed among populations, we assayed isofemale lines from populations in the New World and the Old World for the length variation in the control region of mtDNA. The results show that heteroplasmy is significantly less frequent than in the earlier study, and North American samples do show reduced diversity for length variation. These data are discussed in light of the distinct mutational dynamics of this molecular marker system and shed light on the balance of mutation and drift in structuring *D. melanogaster* populations.

Materials and methods

Fly stocks

The collection origin, number of isofemale lines, and time held in lab for each line studied are listed in the legend to Figure 3. Lines from Zimbabwe, China, Australia, and California were kindly provided by Dr Chip Aquadro. Connecticut fly lines were collected in 1988 by Dr Phil Barnes. Lines from Massachusetts and North Carolina were collected by the authors from the Four Town Farm fruit and vegetable stand in Seekonk, MA, and a farmer's market in Fayetteville, NC, respectively. Flies were caught overnight in a bottle traps containing banana slices, with an inverted paper cone preventing egress. All fly lines analyzed for mtDNA were descended from a single wild-caught female one generation from the wild.

DNA analysis

MtDNA size estimates were made by Southern blot analysis following standard protocols (Sambrook *et al*, 1989). Genomic DNA was isolated from samples of F1 offspring of individual females from each line. Total DNA was digested with *Hae*III, electrophoresed in 0.8% agarose gels and transferred to nitrocellulose membranes. These membranes were probed with the 4.8 kb *Hind*III fragment of *D. melanogaster* lying adjacent to the A+T rich control regions (Solignac *et al*, 1986). This probe fragment was isolated by digesting a plasmid clone containing the fragment, electrophoresing the DNA on low-melting agarose, and excising the mtDNA band from the gel (Sambrook *et al*, 1989). This DNA (~20 ng) was then labeled with d-AT³²P using the Prime-It II random priming kit from Stratagene (La Jolla, CA, USA). The probe hybridizes to unique-sequence DNA in the length-variable *Hae*III fragment and to the constant-sized 6.2 kb band in all flies (see Figure 1). The constant-sized fragment could be used to detect and account for differences in lane mobility due to variation in blotting or electrophoresis conditions (cf. Figure 1, lane 106). Washed membranes were exposed to Fuji X-Ray film with two intensifying L-Plus screens. Two exposures were obtained for each gel in most experiments.

MtDNA length heteroplasmy was detected in some samples. The proportions of the two different-sized mtDNAs was estimated from densitometry of autoradiographs using an imaging densitometer (Model GS-670, Bio-Rad) interfaced with a Macintosh computer. Relative

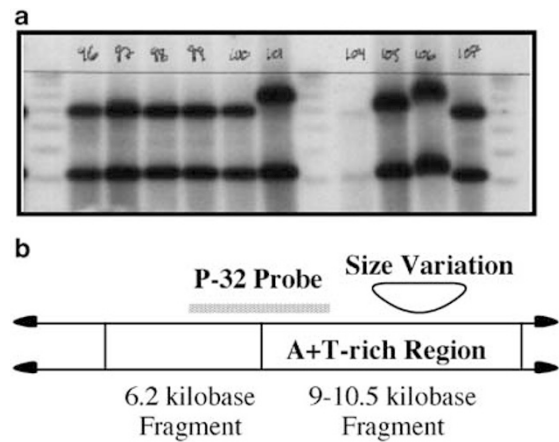


Figure 1 mtDNA length variation in *D. melanogaster*. (a). Autoradiograph of Southern blot of *Hae*III restriction digestion showing single fragment containing the control region, which varied between 9 and 10.5 kb in length (upper band) and a constant sized 6.2 kb fragment (lower band in a), which served as an internal control for each lane in cases of poor migration, for example, lane 106 in (a).

frequencies of the two length variants were estimated from the relative band areas on the scanned image as described in Kann *et al* (1998). The averages of the values from the replicate autoradiographs were used as the estimates of the frequencies of mtDNA length variants within a single fly (ie, a fly's heteroplasmic 'genotype').

Results

Length variation in different populations

MtDNA length variation was determined for 194 isofemale lines of *D. melanogaster* from seven populations (Australia, Beijing, Zimbabwe, California, Connecticut, Massachusetts, and North Carolina). The frequency distributions of length variation for these populations are shown in Figure 2. The distributions can be characterized by a common 'normal' sized mitochondrial genome length of 18.6 kb, and a scattering of larger genomes. If the samples are binned into 'normal' (<19 kb) vs larger mtDNA size classes (>19 kb), the Old World samples have more of the larger genome sizes than do the more homogeneous New World samples (Fisher's exact test, $P < 0.0001$). The colonization history described by David and Capi (1988) places Australia as New World; a Fisher's exact test remains significant when Australia or Beijing or both are excluded from the 'Old World'. These results are consistent with an African origin to *D. melanogaster* and subsequent spread across the globe, especially encompassing a founder event during its transit to the Americas. This is evident in the greater frequency skew in the New World samples, relative to Zimbabwe and Beijing (see Figure 2). When the diversity for mtDNA length is tabulated as $D = 1 - \sum x_i^2$, where x_i is the frequency of the i th length variant, it is clear that the New World samples have lower diversity, as does the pooled sample from the current study, relative to the early study by Hale and Singh (1991) (see Table 1). Again, this is consistent with a loss of variation in the New World samples.

Heteroplasmy

Table 1 shows the frequencies of large mtDNA and the incidence of heteroplasmy in the populations we have sampled. Also listed in Table 1 are comparable data from an earlier study by Hale and Singh (1991), where 144 isofemale lines of *D. melanogaster* were assayed for mtDNA length variation and heteroplasmy. The current

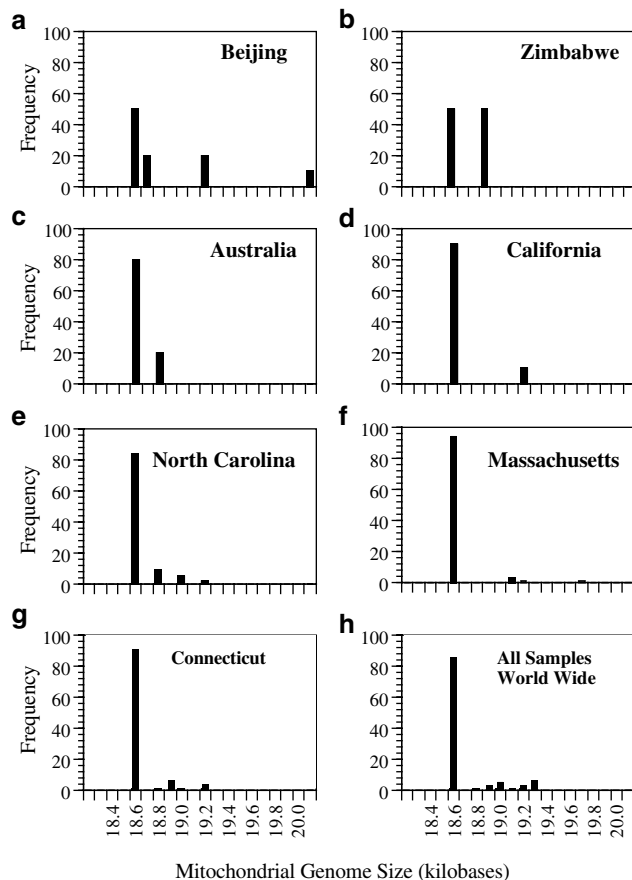


Figure 2 Histograms of mitochondrial genome size, assessed from isofemale lines isolated from several populations around the globe. (a) Beijing, $n=10$; (b) Zimbabwe, $n=9$; (c) Australia, $n=9$; (d) Arvin, CA, $n=10$; (e) Fayetteville, NC, $n=25$; (f) Massachusetts, $n=100$; (g) Connecticut, $n=31$; (h) All samples in this study, worldwide, $n=194$. In cases of heteroplasmy, densitometry was used to estimate the frequency of each size class and thus its proportional contribution to the total distribution.

study shows significantly fewer long mtDNAs and significantly lower incidence of heteroplasmy than the Hale and Singh (1991) study. As expected, the localities with the largest sample size show the only incidence of heteroplasmy (Connecticut $n=31$, and Massachusetts $n=100$).

The Hale and Singh data are a heterogeneous collection of relatively small samples ($n=5-12$) from 18 different localities around the world. The higher incidence of heteroplasmy in these samples is not due to a specific locality showing very common heteroplasmy; rather it appears at moderate levels in many samples. Hale and Singh (1991) confirmed that heteroplasmy was intra-individual, and not due to heterogeneity among flies within a female line. One notable difference between the Hale and Singh (1991) samples and those from our study is the length of time the flies had been in culture. A number of the Hale and Singh (1991) lines had been in the laboratory for as long as 9 years before analysis. The oldest lines we studied were collected about 5 years prior to analysis, and the Massachusetts and North Carolina samples were one generation from the wild (Figure 3). The relationship between frequency of heteroplasmy and age of the stock is significant (Figure 3) for our seven samples plus the Hale and Singh (1991) data. This relationship remains significant if only one of our 3-year old samples is used. These data suggest that lines maintained in small cultures in the lab may accumulate heteroplasmy for mtDNA length mutations.

Discussion

Our analyses of 194 female lines of *D. melanogaster* from around the world show that mtDNA length variation is significantly reduced in North American samples relative to Old World samples. Our results further show that the overall levels of mtDNA size variation and heteroplasmy are significantly lower than those reported in an earlier worldwide study using lines that had been in laboratory culture for a number of years (see Table 1 and Hale and Singh, 1991). However, both studies show a strong skew to the distribution of length variants in natural populations (Figure 2), a pattern that is common in mitochondrial VNTR systems (Rand, 1993; Lunt *et al.*, 1998). What are the causes of these skewed distributions?

One explanation for the skewed distribution is an equilibrium between insertion mutations that stochastically increase mitochondrial genome size, and selection for smaller, faster replicating genomes (Rand, 1993, 2001;

Table 1 Proportions of larger mtDNAs and heteroplasmy in samples of *D. melanogaster*

Population	Sample size	No. of large	Proportion	Length diversity	No. heteroplasmic	Proportion
Australia	9	2	0.22	0.320	0	0.00
Beijing	10	5	0.50	0.634	0	0.00
California	10	1	0.10	0.180	0	0.00
Connecticut	31	3	0.10	0.164	2	0.07
Fayetteville	25	5	0.20	0.268	0	0.00
Massachusetts	100	6	0.06	0.115	1	0.01
Zimbabwe	9	5	0.56	0.500	0	0.00
Combined ^a	194	27	0.14	0.272	3	0.02
Hale and Singh world	144	46	0.32	0.739	24	0.17

^aThe combined sample has significantly fewer large mtDNAs ($\chi^2=9.3$; $P<0.003$), and significantly fewer heteroplasmic lines ($\chi^2=19.7$, $P<0.0001$) than the Hale and Singh (1991) worldwide sample.

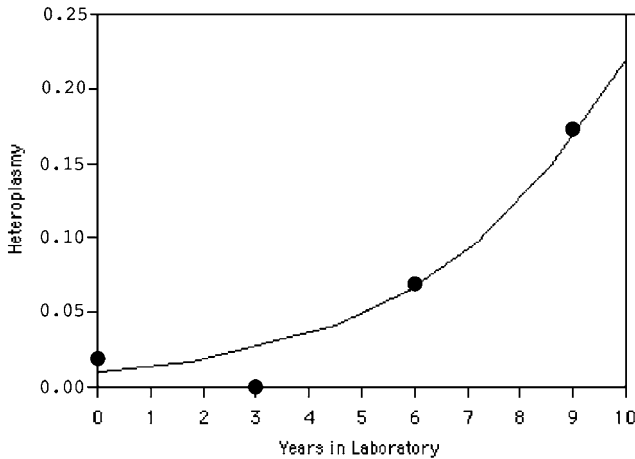


Figure 3 Logistic regression of the probability of heteroplasmy versus number of years in the laboratory. The regression was performed on the presence/absence of heteroplasmy in 275 independently maintained lineages. The four points represent the proportion found to be heteroplasmic in lines derived from different locales. Zero years (one generation) in the laboratory: Fayetteville, North Carolina (45 lineages) and Seekonk, Massachusetts (100 lineages). Three years: Australia (9), China (10), Zimbabwe (9), and California (10). Five years: Connecticut (31). Nine years: Hale and Singh world sample (144). The logistic model is that $\log \frac{\text{Pr}\{\text{heteroplasmy}\}}{1-\text{Pr}\{\text{heteroplasmy}\}} = a + bt$, where t is the time in years. The model has a highly significant fit ($\chi^2 = 19.4$, $P < 0.0001$). The estimated intercept for the log odds, a , of a heteroplasmic versus homoplasmic lineage is -4.7 (significant, $P < 0.0001$), and the estimated slope with years, b , is 0.34 (significant, $P = 0.0012$). Although it does not incorporate as data the number of samples assayed from each locale, a linear regression on the data graphed (frequency of heteroplasmy in each of the eight locales above) is also significant ($R^2 = 0.81$, $P < 0.001$).

Selosse *et al*, 2001). While a mutation-selection balance may account for the skewed distribution, for this explanation to underlie the geographic variation in the skew, selection would have to be acting differently in Eastern and Western Hemispheres. Available evidence suggests that mtDNA length variants are neutral at the level of the individual: there is no association between mtDNA length variants and growth rate in bivalves (Zouros *et al*, 1992), and mtDNA length variants behave as neutral markers in experimental populations of *Drosophila* (D Rand, unpublished observations). While differences in the intracellular replication rates of length variants could be attributed to population differentiation in nuclear genes regulating mtDNA replication, without direct evidence for this, such a speculation is difficult to credibly evaluate (eg, Feder and Hofmann, 1999).

Another explanation for the strong skew of these distributions toward smaller mitochondrial genomes is a length-biased model of neutral mutations. Consider a variable number of tandem repeats (VNTR) model assuming a minimum size of α , and a maximum size of Ω . Further assume that the indel rate is proportional to the number of tandem repeats, and mutations are equally likely to be insertions or deletions. Counterintuitively, this model produces a strong skew toward smaller lengths, such that a sequence with i repeats will be represented with frequency

$$P_i = \frac{1}{i \sum_{j=\alpha}^{\Omega} 1/j}, \quad (1)$$

(see Appendix A, cf. Falush and Iwasa, 1999). If long length classes were lost in the founder event associated with the colonization of North America, the difference in distributions between New and Old World could result from a different source pool of mutational templates from which to generate new variation.

Population structure for mtDNA length variation

Whether the low frequency of larger mitochondrial genomes is due to mutational dynamics or selection, the biogeographic differences observed here seem to have retained a signal of the presumed founder event that occurred during the colonization of the new world. This is surprising since the high indel mutation rate of mitochondrial control region VNTRs should result in a rapid recovery of equilibrium. If we assume that *D. melanogaster* colonized North America 400–500 years ago and has 10 generations per year in the wild (David and Capi, 1988; Powell, 1997), then is ~ 4000 generations sufficient time to regenerate the frequencies of the larger mtDNA that might have been lost in the founding populations?

Tallying results from this study and that of Hale and Singh (1991) yields a proportion of mtDNAs shorter than 19.0 kb of 69%. Assume that a longer mtDNA size class had existed at an equilibrium (from Equation (1) with $\alpha = 1$ and $\Omega = 2$) of $q_0 = 33\%$ in the source population that was reduced to a very low frequency in the colonization of North America. Have enough generations passed to restore this size class to its original frequency? An estimate of the indel rate in cricket mtDNA is 10^{-4} – 10^{-3} per generation (Rand and Harrison, 1989). In the absence of other data, we will assume that the rate is similar for mtDNA length variation in *D. melanogaster*. The time evolution of the frequency of an allele subject to reversible neutral mutation can be calculated as

$$q_t = \left(q_0 - \frac{v}{\mu + v} \right) (1 - \mu - v)^t + \frac{v}{\mu + v} \quad (2)$$

(Hartl and Clark, 1997), where q_t is the current frequency of genome sizes longer than 19.0 kb, we consider a mutation rate from the shorter size class to the longer size class of $\mu = 10^{-4}$, and a mutation rate to the smaller size class of $v = 2 \times 10^{-4}$, conservatively consistent with the data and consistent with Equation (1).

If the founder population in the New World had had no genomes at all of the larger sizes ($q_0 = 0$), it would have taken only about 1200 generations to reach the current $q = 0.1$. Migration, of course, would have accelerated this equilibration. Alleles of the long size class in the hypothetical founder population with an initial $q_0 = 0$ would have climbed beyond the current frequency of 0.1 to $q_{4000} = 0.23$ after the expected ~ 4000 generations. This discordance between observed and expected patterns of mtDNA length variation in North America may be due to an overestimation of the mutation rate in this model, or some selection for mtDNA length variation that has hindered the recovery of mtDNA length diversity in the founder populations.

Clearly, better estimates of indel mutation rates are needed to use this marker as a reliable population genetic tool. Coupled with better models of the mutational dynamics of mitochondrial control region VNTRs, these markers could provide a short time frame window

into the history of populations. They are the common source of mitochondrial genome size variation in many organisms (Boyce *et al*, 1989; Rand and Harrison, 1989).

Mutation accumulation and germ-line aging in mtDNA

Our samples of mtDNA length variants show significantly lower frequencies of long mtDNAs and heteroplasmic lines compared to an earlier study by Hale and Singh (1986). The composition of the samples in these studies is similar in that several localities were surveyed from different continents. The sample sizes for each locality in the Hale and Singh (1986) study ($n=5-12$) tend to be lower than in our study ($n=9-100$), but Hale and Singh (1986) surveyed more localities. We only detected heteroplasmy in two samples with relatively large sample sizes ($n=31$ and 100).

An important difference between our lines and those of Hale and Singh (1991) is the length of time that the flies had been in culture in the lab (see Table 1 and Figure 3). Long-term culture of *Drosophila* lines leads to a number of phenotypic changes (Sgro *et al*, 2000), and to the accumulation of deleterious mutations and inbreeding depression (Bijlsma *et al*, 1999). MtDNA point mutations and indels have been detected among mutation accumulation lines of *C. elegans* (Denver *et al*, 2000). Moreover, there are a large number of nuclear genes regulating mitochondrial function (Grivell *et al*, 1999; Steinmetz *et al*, 2002) that may provide a large mutational target for deleterious effects related to mtDNA replication or turnover. The reduced effective population size of long-term culture could weaken selection against longer mtDNA variants. Similarly, the accumulation of deleterious mutations in genes coding for the machinery of mitochondrial biogenesis and maintenance results in a similar reduced purification of the mtDNA population within germ cell lineages. There is an extensive literature indicating that mtDNA mutations accumulate with age in somatic tissue (eg, Chinnery *et al*, 2002). Relatively little is known about the population dynamics of mitochondria in the germ-line stem cells that give rise to each successive generation. The relatively high mutation rate for mtDNA length variants may provide an effective means of studying this phenomenon in mutation accumulation lines. As shown here, these mtDNA length variants can retain a signature of population structure into the recent evolutionary past. The unique evolutionary dynamics of mtDNA length variants provides a marker for the analysis of population genetic phenomena across multiple hierarchical levels (Rand, 2001): within and between cell lineages as well as within and between geographic populations.

References

- Andolfatto P (2001). Contrasting patterns of X-linked autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol Biol Evol* **18**: 279–290.
- Aquadro CF, Bauer DuMont V, Reed FA (2001). Genome-wide variation in the human fruitfly: a comparison. *Curr Opin Genet Dev* **11**: 627–634.
- Begun DJ, Aquadro CF (1995a). Evolution at the tip base of the X chromosome in an African population of *Drosophila melanogaster*. *Mol Biol Evol* **12**: 382–390.
- Begun DJ, Aquadro CF (1995b). Molecular variation at the vermilion locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. *Genetics* **140**: 1019–1032.
- Bijlsma R, Bundgaard J, Van Putten WF (1999). Environmental dependence of inbreeding depression purging in *Drosophila melanogaster*. *J Evol Biol* **12**: 1125–1137.
- Boyce TM, Zwick ME, Aquadro CF (1989). Mitochondrial DNA in the bark weevils: size structure heteroplasmy. *Genetics* **123**: 825–836.
- Buroker NE, Brown JR, Gilbert TA, O'Hara PJ, Beckenbach AT, Thomas WK, Smith MJ (1990). Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics* **124**: 157–163.
- Caracristi G, Schlotterer C (2003). Genetic differentiation between American European *Drosophila melanogaster* populations could be attributed to admixture of African alleles. *Mol Biol Evol* **20**: 792–799.
- Chinnery PF, Samuels DC, Elson J, Turnbull DM (2002). Accumulation of mitochondrial DNA mutations in ageing cancer, mitochondrial disease: is there a common mechanism? *Lancet* **360**: 1323–1325.
- David J, Capy P (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet* **4**: 106–111.
- Denver DR, Morris K, Lynch M, Vassilieva L, Thomas WK (2000). High direct estimate of mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**: 2342–2344.
- Falush D, Iwasa Y (1999). Size-dependent mutability microsatellite constraints. *Mol Biol Evol* **16**: 960–966.
- Feder ME, Hofmann GE (1999). Heat-shock proteins molecular chaperones, the stress response: evolutionary ecological physiology. *Annu Rev Physiol* **61**: 243–282.
- Grivell LA, Artal-Sanz M, Hakkaart G, de Jong L, Nijtmans LG, van Oosterum K *et al* (1999). Mitochondrial assembly in yeast. *FEBS Lett* **452**: 57–60.
- Hale LR, Singh RS (1986). Extensive variation heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **78**: 8813–8817.
- Hale LR, Singh RS (1991). A comprehensive study of genetic variation in natural populations of *Drosophila melanogaster*. IV. Mitochondrial DNA variation the role of history vsselection in the genetic structure of geographic populations. *Genetics* **129**: 103–117.
- Hartl DL, Clark AG (1997). *Principles of Population Genetics*. Sinauer: Sunderland, MA.
- Hayasaka K, Ishida T, Horai S (1991). Heteroplasmy polymorphism in the major noncoding region of mitochondrial DNA in Japanese monkeys: association with tandemly repeated sequences. *Mol Biol Evol* **8**: 399–415.
- Kann LM, Rosenblum EB, Rand DM (1998). Aging mating, the evolution of mtDNA heteroplasmy in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **95**: 2372–2377.
- Lewis DL, Farr CL, Farquhar AL, Kaguni LS (1994). Sequence organization, evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. *Mol Biol Evol* **11**: 523–538.
- Lunt DH, Whipple LE, Hyman BC (1998). Mitochondrial DNA variable number of tandem repeats (VNTRs): utility and problems in molecular ecology. *Mol Ecol* **7**: 1441–1455.
- Moriyama EN, Powell JR (1996). Intraspecific nuclear DNA variation in *Drosophila*. *Mol Biol Evol* **13**: 261–277.
- Powell JR (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press: New York, NY.
- Rand DM (1993). Endotherms, ectotherms, and mitochondrial genome-size variation. *J Mol Evol* **37**: 281–295.
- Rand DM (2001). The units of selection on mitochondrial DNA. *Annu Rev Ecol System* **32**: 415–448.
- Rand DM, Dorfsman ML, Kann LM (1994). Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. *Genetics* **138**: 741–756.

- Rand DM, Harrison RG (1989). Molecular population genetics of mtDNA size variation in crickets. *Genetics* **121**: 551–569.
- Rubinsztein DC, Amos B, Cooper G (1999). Microsatellite and trinucleotide-repeat evolution: evidence for mutational bias and different rates of evolution in different lineages. *Philos Trans R Soc Lond B Biol Sci* **354**: 1095–1099.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning: A laboratory Manual*. Cold Spring Harbor Laboratory: Cold Spring Harbor, New York.
- Schug MD, Mackay TF, Aquadro CF (1997). Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nat Genet* **15**: 99–102.
- Selosse M-A, Albert B, Godelle B (2001). Reducing the genome size of organelles favours gene transfer to the nucleus. *Trends Ecol Evol* **16**: 135–141.
- Sgro CM, Geddes G, Fowler K, Partridge L (2000). Selection on age at reproduction in *Drosophila melanogaster*: female mating frequency as a correlated response. *Evolution* **54**: 2152–2155.
- Solignac M, Monnerot M, Mounolou J-C (1986). Mitochondrial DNA evolution in the *melanogaster* species subgroup of *Drosophila*. *J Mol Evol* **23**: 31–40.
- Steinmetz LM, Scharfe C, Deutschbauer AM, Mokranjac D, Herman ZS, Jones T *et al* (2002). Systematic screen for human disease genes in yeast. *Nat Genet* **31**: 400–404.
- Zouros E, Pogson GH, Cook DI, Dadswell MJ (1992). Apparent selective neutrality of mitochondrial DNA size variation: a test in deep sea scallop *Placopecten magellanicus*. *Evolution* **46**: 1466–1476.

Appendix A

Here we use mathematical induction to show that an equilibrium skewed distribution described by Equation (1) in the text is a consequence of a variable number of tandem repeats (VNTR) model assuming a minimum size of $\alpha > 0$, a maximum size of Ω , an indel rate proportional to the number of tandem repeats, and mutations that when permissible are equally likely to be insertions or deletions.

Let C_j be the average count of genotypes of repeat number j , where j is a positive integer. Since insertion and deletion mutations from each size class are assumed to occur at a rate μ times the number of repeats, the average counts in each size class change each $1/\mu$ generations as

$$C'_j = C_j + (j-1)C_{j-1} - 2C_j + (j+1)C_{j+1} \quad (\text{A1})$$

At equilibrium, $C'_j = C_j$, so that

$$(j-1)C_{j-1} - 2jC_j + (j+1)C_{j+1} = 0 \quad (\text{A2})$$

Since $\alpha > 0$, C_1 is a boundary and transitions from C_1 to C_0 are prohibited. From Equation (A2) for the case

of $j = 1$, then,

$$-C_1 + 2C_2 = 0$$

$$C_2 = \frac{1}{2}C_1 \quad (\text{A3})$$

Substituting Equation (A3) into Equation (A2) for the case of $j = 2$,

$$C_1 - 4\left(\frac{1}{2}C_1\right) + 3C_3 = 0$$

$$C_3 = \frac{1}{3}C_1 \quad (\text{A4})$$

Similarly, substituting Equation (A4) into Equation (A2) for the case $j = 3$,

$$C_4 = \frac{1}{4}C_1 \quad (\text{A5})$$

a pattern is apparent. To prove that for all positive integers j , $C_n = C_1/j$, we assume that for some positive integer k ,

$$C_{k-1} = \frac{1}{k-1}C_1 \quad (\text{A6})$$

$$C_k = \frac{1}{k}C_1 \quad (\text{A7})$$

and show that when Equations (A6) and (A7) are true, it is also true that $C_{k+1} = 1/(k+1)C_1$

From Equations A2, A6, and A7,

$$(k-1)\left(\frac{1}{k-1}C_1\right) - 2k\left(\frac{1}{k}C_1\right) + (k+1)C_{k+1} = 0$$

$$C_{k+1} = \frac{1}{k+1}C_1, \text{ Q.E.D.} \quad (\text{A8})$$

The result, $C_n = C_1/j$, is demonstrated for the case $j = 1$ by identity, and for integral values of j from two to four by Equations (A3–A5). It is demonstrated for all integers $j > 4$ by the satisfaction of Equations (A6) and (A7) for the case $j = 4$, and the proof of Equation (A8).

If there is a maximum number of repeats for a VNTR of Ω , beyond which the VNTR may not grow, the equilibrium frequency of size class i , P_i , is then:

$$P_i = \frac{1/i}{\sum_{j=\alpha}^{\Omega} 1/j} = \frac{1}{i \sum_{j=\alpha}^{\Omega} 1/j} \quad (\text{A9})$$