npg

REVIEW

The recent demographic and adaptive history of *Drosophila melanogaster*☆

W Stephan and H Li
Section of Evolutionary Biology, Department of Biology II, University of Munich, Planegg-Martinsried, Germany

Population genetic analyses of the past two decades confirmed an earlier hypothesis by L Tsacas and D Lachaise that the cosmopolitan species *Drosophila melanogaster* has an Afrotropical origin, and that it colonized the rest of the world only very recently. Maximum likelihood analyses based on multilocus data suggest that the putative ancestral African population expanded its size about 60 000 years ago (ya). These demographic changes were accompanied by the fixation of numerous beneficial mutations, as revealed by signatures of positive directional selection in the genome (selective sweeps). The estimated rate of adaptive substitution on the X chromosome is in the order of 10^{-11} per nucleotide site per generation. Comparable (but not significantly higher) substitution rates are found in derived populations that colonized new habitats outside Africa, such as in a European population that branched off from the African lineage about 16 000 ya.
Heredity (2007) **98,** 65–68. doi:10.1038/sj.hdy.6800901; published online 27 September 2006

Introduction

Biogeographical and systematic studies supported the hypothesis by L Tsacas and D Lachaise (first proposed in 1974) that the cosmopolitan species *Drosophila melanogaster* has an Afrotropical origin; furthermore, they were consistent with the notion that *D. melanogaster* colonized the rest of the world only relatively recently (Lachaise *et al.*, 1988). Suggested times of the colonization events range from about 10 000 or 15 000 years for Europe (after the last glaciation) to a few hundred years for American and Australian populations (David and Capy, 1988). However, these and other early publications left a number of questions unanswered, including a definition of which *D. melanogaster* populations are ancestral and which are derived, the question of population subdivision, and also methodological questions about the inferences made that now form the basis of our understanding of the demography of this species.

We review here the progress of the past 20 years in revealing the demographic history of *D. melanogaster*. At the same time we will highlight current attempts by molecular population geneticists to detect the selective events that enabled this species to adapt to environmental changes in the recent past and/or colonize new habitats.

Inferring a species' demographic history from patterns of genetic variation is an essential prerequisite in the search for adaptive signatures in the genome. Traditionally, DNA variation is compared with the expectations of the neutral theory of molecular evolution (Kimura, 1983). While it is tempting to attribute departures from the neutral equilibrium model to the action of positive selection, caution must be taken due to the possible confounding effects of demography. For example, a strong reduction in the level of variation ('selective sweep') may result from genetic hitchhiking associated with positive directional selection (Maynard Smith and Haigh, 1974) or from a strong population size bottleneck. To disentangle demographic and selective forces, it is helpful to employ multilocus approaches (Harr *et al.*, 2002; Glinka *et al.*, 2003; Orengo and Aguadé, 2004). The rationale behind these studies is the observation that, while demography affects patterns of variation across the entire genome, positive selection acts locally near or at relatively few individual genes.

Demographic history

To infer the demographic history of a species, neutral markers should be used. In the past two decades, this became feasible for many species based on neutrality tests. For these tests, usually only a few molecular loci were available. However, in the past 5 years high-throughput sequencing and genotyping techniques allowed us to use large numbers of molecular markers (such as microsatellites and multilocus SNP data). Our understanding of the population history of *D. melanogaster* that, up to 20 years ago (ya), rested largely on studies of characters that are the potential targets of natural selection (such as biometrical traits, enzyme variation, and inversion polymorphism) benefited greatly from these developments.

Using four-cutter restriction site polymorphism data from seven X-linked loci, Begun and Aquadro (1993) showed that East African and North American

Correspondence: *Professor W Stephan, Section of Evolutionary Biology, Department of Biology II, University of Munich, 82152 Planegg-Martinsried, Germany.*
E-mail: *stephan@zi.biologie.uni-muenchen.de*

D. melanogaster populations are substantially different at the molecular level and that genetic variation in Africa is much higher than in the American populations. This result was later confirmed by large-scale microsatellite and DNA sequencing studies and additional population sampling (e.g., Glinka *et al.*, 2003; Ometto *et al.*, 2005; Schlötterer *et al.*, 2006). In addition, these studies showed that that most segregating sites in non-African populations are also polymorphic in Africa. Together, these results clearly support the hypothesis outlined in Introduction that *D. melanogaster* has its origin in Africa and subsequently expanded its range to the temperate zones of the rest of the world.

A sequencing study based on four loci that included samples from 11 populations (four African and seven non-African ones from across the world) found evidence that all non-African populations are derived from a single ancestral population (having undergone a substantial reduction in population size, probably through a bottleneck) and that the ancestral species range appears to be East Africa (Baudry *et al.*, 2004). Population structure among the African populations was small (at least among the East African populations). A subsequent microsatellite study by Schlötterer *et al.* (2006) confirmed these results, in particular the unique origin of the non-African populations.

Based on a heuristic argument, Baudry *et al.* (2004) calculated that this bottleneck occurred about 6400 ya. More recently, efforts have been made to estimate the time of the bottleneck event (or, equivalently, the splitting time between African and non-African populations), using maximum likelihood or approximate Bayesian methods (Ometto *et al.*, 2005; Thornton and Andolfatto, 2006). Ometto *et al.* (2005) considered a bottleneck model with equal pre- and postbottleneck population sizes characterized by two parameters, the time of the population size reduction (T_b) and the 'strength' of the bottleneck. Thornton and Andolfatto (2006) used the same model with three parameters T_b, D_b and R_b, where D_b is the duration of the bottleneck and R_b the ratio of the population sizes during and after the bottleneck. Based on a large data set of 105 short X-linked, noncoding loci from both a European and an (ancestral) East African population (Glinka *et al.*, 2003) and 10 loci from Haddrill *et al.* (2005), they obtained estimates of the splitting time of these populations of about 16 000 years (assuming 10 generations per year). This time estimate is roughly consistent with the earlier suggestion by David and Capy (1988), but it also indicates that the bottleneck was somewhat further in the past than the calculation of Baudry *et al.* (2004) showed. The other two parameters, D_b and R_b, were estimated by Thornton and Andolfatto as 13 000 years and 0.029, respectively, that is, recovery from this relatively long-lasting bottleneck occurred only 3000 ya. In these analyses, the ancestral African population was assumed to be in equilibrium (with constant population size, which may not be true; see below) and the data were given as summary statistics.

Relaxing the assumptions of Thornton and Andolfatto (by considering simultaneously the African and European populations and using the frequency spectrum of SNPs), we recently analyzed a data set of 260 X-linked loci (i.e., the data sets collected by Glinka *et al.* (2003), Haddrill *et al.* (2005) and Ometto *et al.* (2005)) based on a

maximum likelihood procedure (Li and Stephan, 2006). We also estimated T_b as about 16 000 years. However, we found a much deeper bottleneck that lasted only a few hundred years. Furthermore, we showed that the bottleneck inferred in our analysis fits the overall polymorphism pattern better than the scenario proposed by Thornton and Andolfatto (2006). This difference in the inferred bottleneck scenarios has important implications for detecting signatures of selection in the data (see below).

An excess of rare derived mutations in the African sample was observed across these 260 loci (Glinka *et al.*, 2003; Haddrill *et al.*, 2005; Ometto *et al.*, 2005), indicating that the African population is not in equilibrium but may have expanded in recent time. Based on these data, we estimated that the expansion occurred approximately 60 000 ya (Li and Stephan, 2006). This expansion of the African population appears to have occurred during a transition from a full glacial to an interglacial period (70–55 kya; Webb and Bartlein, 1992). After the transition, the African continent was dry and arid, leading to a reduction of rain forests in favor of a mosaic of savannas during the last glacial maximum in the late Pleistocene (18–21 kya) (De Vivo and Carmignotto, 2004). It is noteworthy, however, that a reduction of population size during the last galciation was not observed in the data. This may suggest that the recent wild-to-domestic habit shift of *D. melanogaster* postulated by Lachaise and Silvain (2004) has occurred before the last glacial maximum and has sheltered the nonforest-dwelling populations from suffering population size reductions.

Recent adaptive history

Numerous attempts have been undertaken to study adaptation in *D. melanogaster*. Latitudinal clines generally considered a consequence of an adaptive response to different climates have been analyzed for many traits (David and Capy, 1988). QTL analysis has been used to study temperature tolerance (Norry *et al.*, 2004; Morgan and Mackay, 2006) and other adaptive traits. At the molecular level, interspecific comparisons of DNA sequences yielded estimates of the frequency of adaptive changes that occurred over long periods of time (Smith and Eyre-Walker, 2002). Molecular population geneticists used neutrality tests to detect positive selection at individual loci (e.g., Hudson *et al.*, 1987; Tajima, 1989). Here, we will review multilocus approaches that are based on the hitchhiking effect in revealing positive selection in the *D. melanogaster* genome. With this approach, we can detect *very recent* selective events within a time window of roughly $0.1N_e$ generations, where N_e is the effective population size (Kim and Stephan, 2002). For derived populations, such as the European one (see above), this corresponds to a time period of 10 000–15 000 years, and for the ancestral African population of more than 60 000 years (Li and Stephan, 2006), that is, times in which major demographic processes occurred.

Several attempts have been made to find outliers in multilocus data sets, that is, fragments that show a pattern of variation that cannot be explained by demography alone (Haddrill *et al.*, 2005; Ometto *et al.*, 2005; Thornton and Andolfatto, 2006). However, these approaches may lack power in detecting genomic

regions under selection, if the fragments are relatively short (e.g., 500–600 bp, which is the average length of the loci in the multilocus approaches considered here). In contrast, likelihood ratio tests in which a demographic model (without selection) is compared with a more general model including demography, selection and hitchhiking (Li and Stephan, 2006) appear to be much more powerful.

Applying these methods to the data set of 115 X-linked loci (Glinka *et al.*, 2003; Haddrill *et al.*, 2005) described above, Thornton and Andolfatto (2006) found that the data of the European population can be explained by demography alone without invoking selection, whereas Ometto *et al.* (2005) reported eight outliers (analyzing the larger data set of 260 loci mentioned above). We investigated this discrepancy in detail (Li and Stephan, 2006). Instead of employing the standard neutral model (of constant population size) as null hypothesis, we used the demographic scenarios for the two populations described in the last section. We found that for all outlier loci identified by Ometto *et al.* (2005) and included in our analysis, the inferred demographic scenario alone cannot explain the data. The main reason for this discrepancy appears to be that the bottleneck parameters D_b and R_b we estimated are quite different from those of Thornton and Andolfatto (2006) (see above). For a bottleneck that lasts until about 3000 ya (as inferred by the latter authors), it is more difficult to find signatures of selective sweeps in the data than for bottlenecks that were over much earlier, probably because the level of nucleotide diversity expected under their model is much lower than the observed one.

Furthermore, using the same data set of 260 loci, we estimated the rate of recent adaptive substitution as 0.012×10^{-9} per X-chromosome site per generation for the African population and 0.017×10^{-9} for the European population. If we assume that all selectively driven substitutions occur in coding regions, the corresponding rates are 0.061×10^{-9} and 0.088×10^{-9} for the African and European populations, respectively. These estimates are remarkable for at least two reasons: first, they are comparable for both populations (in fact, the 95% confidence intervals overlap; Li and Stephan, 2006). This suggests that positive directional selection did not only play a critical role in the derived European population during its adaptation to temperate zones, but was also important for the African population. Although not appreciated thus far, this result is not surprising as during the past 60 000 years, that is, since this population expanded in size, dramatic climatic changes occurred in Africa (see above). Consistent with our findings, signatures of positive selection in the African population have also been reported elsewhere (Andolfatto, 2005). Second, although several arguments suggest that we underestimated the rate of adaptive substitution, in particular since our method can detect only strong selective events (Li and Stephan, 2006), our estimates are only slightly lower than the published rate (0.092×10^{-9} per site per generation; Smith and Eyre-Walker, 2002), which was obtained by averaging over a long time period. Considering the fact that the published rate also takes relatively weak selection into account and that our method considers only very recent events, our results indicate an accelerate rate of adaptation in both populations due to recent environmental changes.

Future prospects

For chromosomal regions that deviate from the predictions of the demographic model (as indicated by the methods described above), it can be examined whether the observed reduction of variation is due to positive directional selection. Several statistical tests are available to detect such local signatures of selection in the genome (Kim and Stephan, 2002; Jensen *et al.*, 2005; Li and Stephan, 2006). The overall goal of these studies is to identify the target of selection and to find the genes and associated phenotypes involved in adaptation. It is expected that in the near future, considerable efforts will be made to reach these goals.

Thus far, work on only a few such candidate regions of selective sweeps has begun. One of them is the *wapl* region. Beisswanger *et al.* (2006) found clear evidence of a selective sweep in the European population that is likely to have originated in Africa. In both the European and African populations, variation is significantly reduced. The genomic segment of reduced variation in the African population is much narrower than that of the European one (which is about 60 kb). While the reduction of variation in both populations is consistent with selective sweeps (rather than demographic processes), tests show that the African sweep is older. The selection coefficients of the selected mutation estimated for both populations by the method of Kim and Stephan (2002) are very similar (approximately 1.3×10^{-3}).

The next goals are to narrow down the genomic segment in which the target of selection is located, identify the functionally important nucleotide polymorphisms, and understand the population genetic process underlying this sweep. The genomic region of reduced variation in the European sample harbors several genes coding for products with metabolic functions, including *Cyp4d1*, a cytochrome *P450* gene putatively involved in steroid metabolism, and *Pgd* coding for a metabolic enzyme involved in the pentose-phosphate-shunt. Genes encoding metabolic enzymes have frequently been suggested as targets of positive selection (Eanes, 1999). Interestingly, of the two most likely genes involved in adaptation of *D. melanogaster* to temperate zones (*Cyp4d1* and *Pgd*), only the 5′ regulatory region of *Pgd* is located within the sweep region of the African population, which makes this region the prime candidate for future work.

Glinka *et al.* (2006) chose a second region for analysis of selective sweeps (around the *unc-119* gene). Using a combination of the Kim and Stephan (2002) test and the approach by Jensen *et al.* (2005), they found strong evidence that variation in this region was shaped by a very recent selective sweep in the European population. The region of reduced polymorphism extended over a distance of about 45 kb. The selection coefficient was estimated as 2.9×10^{-3} and the target of selection localized near a gene cluster composed of the genes *CG1677*, *CG2059* and *unc-119*. Of these genes, *CG2059* may be most interesting, as its product has hydrolase activity involved in lipid metabolism. Except for these three genes, there is only one other gene (*CG1958*) in the valley of reduced variation (14.2 kb away from the gene cluster). As it is more distant from the estimated position of the target of selection and its product has a general house-keeping function (a putative nucleic acid binding

activity), it seems to be most likely that the target of selection is located within the gene cluster *CG1677*, *CG2059* and *unc-119*.

However, as J Jensen *et al.* (personal communication) have pointed out, it is difficult to identify the exact targets of selection and hence the causative DNA changes, if the candidate regions of selective sweeps are only partially sequenced (which is the case for the *wapl* and *unc-119* regions). For this reason it would be more promising to concentrate in the future on smaller regions of reduced variation and sequence them completely. Pool *et al.* (2006) have recently published such an example. They were able to localize the sweep region to a 361-bp window within the 5′ regulatory region of the *roughest* gene in a population from Zimbabwe, with one nucleotide substitution representing the best candidate for the target of selection.

The identification of functionally relevant variation will be further improved by combining large-scale sequencing and population genetic analyses with other high-throughput methods, such as microarray expression techniques. Such methods will help to unravel whether single genes or gene networks are the determinants of adaptation and whether the DNA changes under selection are structural or regulatory. Finally, it is important to link genotypic and phenotypic variation by combining the method of selection mapping (described here for identifying regions of selective sweeps) with quantitative genetic analyses of particular traits associated with the genes under selection.

Acknowledgements

References

Andolfatto P (2005). Adaptive evolution of non-coding DNA in *Drosophila*. *Nature* **437**: 1149–1152.

Baudry E, Viginier B, Veuille M (2004). Non-African populations of *Drosophila melanogaster* have a unique origin. *Mol Biol Evol* **21**: 1482–1491.

Begun DJ, Aquadro CF (1993). African and North American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**: 548–550.

Beisswanger S, Stephan W, De Lorenzo D (2006). Evidence for a selective sweep in the *wapl* region of *Drosophila melanogaster*. *Genetics* **172**: 265–274.

David JR, Capy P (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet* **4**: 106–111.

De Vivo M, Carmignotto AP (2004). Holocene vegetation change and the mammal faunas of South America and Africa. *J Biogeography* **31**: 943–957.

Eanes WF (1999). Analysis of selection on enzyme polymorphisms. *Annu Rev Ecol Syst* **30**: 301–326.

Glinka S, De Lorenzo D, Stephan W (2006). Evidence of gene conversion associated with a selective sweep in *Drosophila melanogaster*. *Mol Biol Evol* (in press).

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D (2003). Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics* **165**: 1269–1278.

Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005). Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* **15**: 790–799.

Harr B, Kauer M, Schlötterer C (2002). Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **99**: 12949–12954.

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

Jensen JD, Kim Y, Bauer DuMont V, Aquadro CF, Bustamante CD (2005). Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* **170**: 1401–1410.

Kim Y, Stephan W (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.

Kimura M (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press: Cambridge.

Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L *et al* (1988). Historical biogeography of the *Drosophila melanogaster* species subgroup. In: Hecht MK, Wallace B, Prance GT (eds) *Evolutionary Biology*, vol. 22. Plenum Press: New York. pp 159–225.

Lachaise D, Silvain J-F (2004). How two Afrotropical endemics made two cosmopolitan human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle. *Genetica* **120**: 17–39.

Li H, Stephan W (2006). Inferring the demographic history and rate of adaptive substitution in *Drosophila*. *PLoS Genet* **2**: e166.

Maynard Smith J, Haigh J (1974). The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35.

Morgan TJ, Mackay TFC (2006). Quantitative trait loci for thermotolerance phenotypes in *Drosophila melanogaster*. *Heredity* **96**: 232–242.

Norry FM, Dahlgaard J, Loeschcke V (2004). Quantitative trait loci affecting knockdown resistance to high temperature in *Drosophila melanogaster*. *Mol Ecol* **13**: 3585–3594.

Ometto L, Glinka S, De Lorenzo D, Stephan W (2005). Inferring the effects of demography and selection on *Drosophila melanogaster* populations from a chromosome-wide scan of DNA variation. *Mol Biol Evol* **22**: 2119–2130.

Orengo DJ, Aguadé M (2004). Detecting the footprint of positive selection in a European population of *Drosophila melanogaster*: multi-locus pattern of variation and distance to coding regions. *Genetics* **167**: 1759–1766.

Pool JE, Bauer DuMont V, Mueller JL, Aquadro, CF (2006). A scan of molecular variation leads to the narrow localization of a selective sweep affecting both Afrotropical and cosmopolitan populations of *Drosophila melanogaster*. *Genetics* **172**: 1093–1105.

Schlötterer C, Neumeier H, Sousa C, Nolte V (2006). Highly structured Asian *D. melanogaster* populations – a new tool for hitchhiking mapping? *Genetics* **172**: 275–286.

Smith NGC, Eyre-Walker A (2002). Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Thornton K, Andolfatto P (2006). Approximate Bayesian inference reveals evidence for a recent, severe, bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* **172**: 1607–1619.

Webb III T, Bartlein PJ (1992). Global changes during the last 3 million years: climatic controls and biotic responses. *Annu Rev Ecol Syst* **23**: 141–173.