View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex

Bettina Harr*, Steven Weiss*, Jean R. David†, Gottfried Brem* and Christian Schlötterer*

Uncovering the genealogy of closely related species remains a major challenge for phylogenetic reconstruction. It is unlikely that the phylogeny of a single gene will represent the phylogeny of a species as a whole [1], but DNA sequence data across a large number of loci can be combined in order to obtain a consensus tree [2]. Long sequences are needed, however, to minimize the effect of (infrequent) base substitutions, and sufficient individuals must be sequenced per species to account for intraspecific polymorphisms, an overwhelming task using current DNA sequencing technology. By contrast, microsatellites are easy to type [3], allowing the analysis of many loci in multiple individuals. Despite their successful use in mapping [4,5], behavioural ecology [6] and population genetics [7], their usefulness for the phylogenetic reconstruction of closely related taxa has never been demonstrated, even though microsatellites are often conserved across species [8–10]. One drawback to microsatellite use is their high mutation rate (10^{-4}–10^{-2}), combined with an incomplete understanding of their mutation patterns. Many microsatellites are available for *Drosophila melanogaster*, and they are distributed throughout the genome [11]. Most can be amplified in the *D. melanogaster* species complex [12,13] and have low mutation rates [14,15]. We show that microsatellite-specific distance measurements [16] correlate with other multilocus distances, such as those obtained from DNA–DNA hybridization data. Thus microsatellites may provide an ideal tool for building multilocus phylogenies. Our phylogenetic reconstruction of the *D. melanogaster* complex provides strong evidence that *D. sechellia* arose first, followed by a split between *D. simulans* and *D. mauritiana*.

Addresses: *Institut für Tierzucht und Genetik, Josef Baumann Gasse 1, A-1210 Wien, Austria. †CNRS, Laboratoire de Biologie et Genetique Evolutives, Avenue de la Terrasse, F-91198 Gif-sur-Yvette Cedex, France.

Correspondence: Christian Schlötterer
E-mail: christian.schloetterer@vu-wien.ac.at

Received: 4 August 1998
Revised: 14 September 1998
Accepted: 14 September 1998

Published: 12 October 1998

Current Biology 1998, 8:1183–1186
http://biomednet.com/elecref/0960982200801183

© Current Biology Ltd ISSN 0960-9822

Results and discussion

Under the stepwise mutation model, the squared average difference in mean repeat number, $(\delta\mu)^2$, is linearly correlated with time [16]. Kimmel *et al.* [17] noted that the linearity of stepwise distances is independent of the assumptions of both single repeat-unit step sizes and symmetry in mutation rates. Hence, the greatest concerns for the use of microsatellites in phylogenetic reconstruction are potential constraints on allele size and whether or not the mutational properties of loci are maintained across species [18]. Allele size constraints would result in an underestimate of genetic divergence between species [19,20]. As constraints are expected to be more pronounced the more diverged the species are, a nonlinear relationship between microsatellite-based distances and other multilocus-based estimates would result. Our genetic divergence estimates (see Supplementary material published with this paper on the internet) based on $(\delta\mu)^2$ are, however, highly correlated both with DNA–DNA hybridization data [21] ($r = 0.918$, $p = 0.036$) and with allozyme data [22] ($r = 0.939$, $p = 0.020$). Hence, microsatellite evolution appears to be relatively unconstrained across species within the divergence time of *D. melanogaster* and *D. simulans*, which is estimated to be 2.5–3.5 million years [23]. Recently, we demonstrated [12] that the mutational properties of microsatellite loci are conserved between *D. melanogaster* and *D. simulans*. Thus, the two greatest concerns for phylogenetic reconstruction based on microsatellites — size constraints and differences in mutational properties — appear to be of minimal concern within the *D. melanogaster* species complex.

The distance $(\delta\mu)^2$ can be used to estimate times of divergence if the average mutation rate of microsatellites is known. Two recent studies obtained an average microsatellite mutation rate of 6.3×10^{-6} per generation in *D. melanogaster*, which is more than one order of magnitude lower than in mammals [14,15]. Using this average mutation rate, the estimated divergence time between *D. melanogaster* and *D. simulans* is 130,000 years (Table 1), a result that is clearly incompatible with previous divergence estimates of 2.5 – 3.5 million years [23]. Several compounding factors may contribute to this discrepancy, including small violations in the assumptions required to satisfy the model. For example, there may be slight constraints because of an increased rate of back mutations for long alleles [24]. Furthermore, although variances in repeat number between *D. simulans* and *D. melanogaster* are significantly correlated, only 36% of the

variation was explained by the regression equation [12]. An additional source of error is the estimated number of generations per year, which may be inaccurate. The most important assumption, however, is the mutation rate itself, which may be overestimated as a result of the experimental design of the studies measuring mutation rates in *D. melanogaster*; both studies used a set of lines with identical alleles, which could have resulted in an over-representation of hypervariable alleles causing a higher mutation rate estimate [14]. Given all these uncertainties, divergence times based on $(\delta\mu)^2$ should be viewed with caution.

Genealogy of the *D. melanogaster* complex

Genetic distances between species were calculated by various methods including $(\delta\mu)^2$ [16], Nei's distance's [25], and the proportion of shared alleles. Irrespective of the distance measurement used, all UPGMA (unweighted pair-group method using an arithmetic average) and neighbor-joining trees supported the same grouping, with *D. melanogaster* depicted as the most distantly related species. In the remaining clade, *D. sechellia* arose first, followed by the split between *D. simulans* and *D. mauritiana*. To test the consistency of this result, we constructed an allele-sharing tree of individuals, a method which has been successfully used for the reconstruction of the phylogenetic relationships of human populations based on microsatellites [26]. The UPGMA tree in Figure 1 shows that all individuals from the same species cluster together. The major difference between the UPGMA and a neighbor-joining tree is that a single *D. simulans* individual is not clustering with the other 31 *D. simulans* individuals but instead forms a sister group to *D. mauritiana* and the remaining *D. simulans* individuals. The bootstrap support for each species ranges from 84 to 100%, with *D. simulans* having the lowest bootstrap support (Figure 1).

Table 1

Expected time of divergence, in millions of years, for various mutation rates.

Pairwise comparison of species	$(\delta\mu)^2$	Mutation rate					
		10^{-5}	6.3×10^{-6}	10^{-6}	5×10^{-7}	10^{-7}	10^{-8}
mel/sec	21.574	0.11	0.17	1.08	2.2	10.8	107.9
mel/mau	17.936	0.09	0.14	0.90	1.8	9.0	89.7
mel/sim	15.9795	0.08	0.13	0.80	1.6	8.0	79.9
mau/sec	12.5221	0.06	0.01	0.63	1.3	6.3	62.6
sim/sec	11.6995	0.06	0.09	0.58	1.2	5.8	58.5
sim/mau	5.1765	0.03	0.04	0.26	0.5	2.6	25.9

The time of divergence was calculated on the basis of an assumption of 10 generations per year. The species names are abbreviated to their first three letters.

Variation in the DNA region flanking microsatellites is well described for cross-species comparisons [9,27,28]. Insertions or deletions in flanking regions of one species can influence the estimated number of repeats if the DNA sequence is known only for another species. To test whether length variation in the flanking regions affects the topology of the tree of individuals, we used the PCR-product length rather than the number of repeats for phylogenetic reconstruction; interestingly, this resulted in a similar grouping of individuals, with a comparable bootstrap support. This is encouraging because it suggests that the microsatellite-based phylogeny is sufficiently robust to mask the phylogenetic noise introduced by variation in the flanking regions. Further investigations will reveal whether phylogenies with a weaker phylogenetic signal could also be reliably reconstructed if PCR-product length is used rather than repeat number.

The species phylogeny of the *D. melanogaster* complex is far from being resolved. Recently, a set of 12 single-copy genes and a ribosomal spacer sequence (ITS) were used to investigate discrepancies between individual gene trees in the *D. melanogaster* species complex [29]. Although all genes demonstrated the sister-group status of *D. melanogaster*, all three possible groupings of the *D. simulans* clade were supported by different genes. The combined data set from all 13 chromosomal regions provided strong evidence that *D. simulans* arose first, followed by a split between *D. mauritiana* and *D. sechellia*. Similarly, DNA–DNA hybridization data [21] support the same topology. Most studies did not use multiple individuals from each species, and studies which did so reported a different pattern: a series of papers using six individuals for each species showed for some genes that *D. sechellia* arose first, followed by the split between *D. simulans* and *D. mauritiana* [30–33]. Furthermore, Kliman and Hey [32] demonstrated that some alleles of the *period* gene are shared between *D. simulans* and *D. mauritiana*, a result which also favors the closer phylogenetic relationship of these species. Solignac and Monnerot [34] showed with restriction fragment length polymorphism analysis of mitochondrial DNA, that *D. simulans* and *D. mauritiana* have multiple mitochondrial DNA haplotypes and that the haplotypes of both these species cluster together. Allozyme data also show that *D. simulans* and *D. mauritiana* are closer to each other than either is to *D. sechellia* [35].

Much thought has been given to the hypothesis that *D. simulans* represents a large population with very old lineages which are still segregating. Our data set, however, provides very little support for this hypothesis. In a neighbor-joining tree of individuals, only a single individual of *D. simulans* split before the branch leading to *D. mauritiana*. The average genetic distance (based on the proportion of shared alleles) between individuals is 0.80 for *D. simulans* and 0.75 for *D. melanogaster*. The unimodal

distribution of the pairwise distances of *D. simulans* individuals suggests that the greater average genetic distance of *D. simulans* can not be explained by the presence of two different lineages, as this would have resulted in a bimodal distribution. Hence, a larger effective population size of *D. simulans* is a more likely explanation for the higher average genetic distance in our data set. As we included individuals from five different populations, it is unlikely that our results can be ascribed to a non-representative sampling of *D. simulans*.

While our results indicate that microsatellites are adequate for phylogenetic reconstruction, it should be mentioned that the obtained tree topologies differed between the various multilocus distances. The topologies reported for allozymes and DNA–DNA hybridization were not statistically supported, however, [21,22]. Furthermore, out of 33 allozyme loci surveyed, 16 showed no variation in the *D. melanogaster* species complex, whereas all of the 39 microsatellite loci did so. Hence, the larger number of informative loci probably explains why our study found a robust branching pattern and the author of the allozyme study had to conclude that the chronology of the speciation events remains unresolved [22].

The great benefit of microsatellites for the reconstruction of phylogenies of closely related species is their mutation rate. Although base substitutions are highly likely to be shared between two closely related species, such as *D. simulans* and *D. mauritiana* [36], microsatellite alleles are less likely to be shared between species because of their higher mutation rate. New mutations would be expected to have occurred before the lineage sorting of DNA sequences is completed. Hence, a less contradictory signal is to be expected when microsatellite data combined over several genomic regions are used for phylogenetic reconstruction of closely related species.

A general difference between the present study and others using microsatellites to reconstruct phylogenies [26] is that we used microsatellites with low mutation rates, so fewer mutational events are likely to have occurred since the split of two species. If the mutational behavior of some of the loci studied deviates from the assumed pattern, then microsatellites with high mutation rates are more likely to result in an inaccurate phylogenetic reconstruction. Our conclusions about the appropriateness of microsatellites for phylogenetic reconstructions

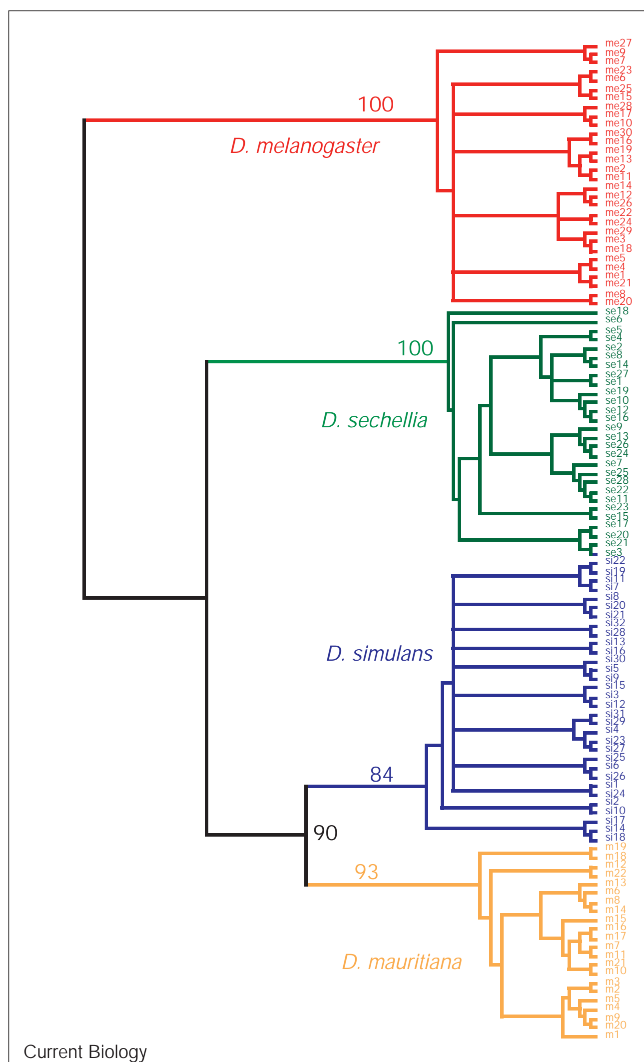

Figure 1

A tree of individuals based on microsatellites. For *D. melanogaster*, F1 individuals from freshly collected females were typed and both alleles analyzed; 10 lines were from France, 10 from Russia, and 10 from Austria. For all other species, isofemale lines were maintained in the laboratory and so a single allele was randomly selected from each individual. *D. simulans* were provided by M. Turelli: United States, 6 lines; Mexico, 6 lines; New Caledonia, 7 lines; Columbia, 7 lines; and Zimbabwe, 6 lines. *D. mauritiana* lines (22) and *D. sechellia* (5) lines were obtained from Bowling Green stock center. The remaining *D. sechellia* lines were collected on various islands of the Seychelles, with most samples originating from the major island, Mahé. Thirty-nine dinucleotide microsatellite loci were typed in all four species. Thirty-two loci were developed from *D. melanogaster* and 7 from *D. sechellia* (B.H., B. Zangerl, M. Imhof, G.B., C.S., unpublished observations). Radioactive microsatellite typing essentially followed procedures given by Schlötterer [3]. After completing 30 PCR cycles, the products were incubated for 50 min at 72°C to assure completion of the terminal transferase activity of the *Taq* polymerase. Electrophoresis was carried out on 7% polyacrylamide gels with 32% formamide and 5.6 M urea to assure complete denaturation of the PCR products. DNA fragments were sized by using a $(GT/CA)_n$ slippage ladder, which produced a band every second base-pair covering a size range from 50 to 230 base pairs [37]. Absolute sizes were determined by running a size reference alongside products. The repeat number for all loci was inferred separately for each species either by using sequences available from GenBank or by sequencing a single allele. If DNA sequencing detected a point mutation in the microsatellite, only the number of uninterrupted repeats in the longest contiguous stretch was counted. Genetic distances were determined using Microsat software [38]. UPGMA and neighbor-joining trees were reconstructed with PHYLIP [39] and tree files were graphically represented using TREEVIEW [40].

may be strongly influenced by the low microsatellite mutation rate of *Drosophila*. As repeat number is a good predictor of a microsatellite's mutation rate, we suggest that the use of microsatellites with a small repeat number should be a successful strategy for phylogenetic reconstruction in other species.

Acknowledgements

We thank D. Goldstein for helpful comments on the manuscript. This study was supported by grants from the Fonds zur Förderung der Wissenschaften to C.S.

Supplementary material

A table showing mean pairwise distances is published with this paper on the internet.

References

1. Tateno Y, Nei M, Tajima F: **Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species.** *J Mol Evol* 1982, **18:**387-404.
2. Pamilo P, Nei M: **Relationships between gene trees and species trees.** *Mol Biol Evol* 1988, **5:**568-583.
3. Schlötterer C: **Microsatellites.** In *Molecular Genetic Analysis of Populations: a Practical Approach. Second edition.* Edited by Hoelzel RA. Oxford: Oxford University Press; 1998:237-261.
4. Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, *et al.*: **A comprehensive genetic map of the human genome based on 5,264 microsatellites.** *Nature* 1996, **380:**152-154.
5. Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, *et al.*: **A comprehensive genetic map of the mouse genome.** *Nature* 1996, **380:**149-152.
6. Schlötterer C, Pemberton J: **The use of microsatellites for genetic analysis of natural populations.** In *Molecular Ecology and Evolution: Approaches and Applications.* Edited by Schierwater B, Streit B, Wagner GP, DeSalle R. Basel: Birkhäuser; 1994:203-214.
7. Bruford MW, Wayne RK: **Microsatellites and their application to population genetic studies.** *Curr Opin Genet Dev* 1993, **3:**939-943.
8. FitzSimmons NN, Mority C, Moore SS: **Conservation and dynamics of microsatellite loci over 300 million years of marine turtle evolution.** *Mol Biol Evol* 1995, **12:**432-440.
9. Schlötterer C, Amos B, Tautz D: **Conservation of polymorphic simple sequence loci in cetacean species.** *Nature* 1991, **354:**63-65.
10. Rico C, Rico I, Hewitt G: **470 million years of conservation of microsatellite loci among fish species.** *Proc R Soc Lond B Biol Sci* 1996, **263:**549-557.
11. Schug MD, Wetterstrand KA, Gaudette MS, Lim RH, Hutter CM, Aquadro CF: **The distribution and frequency of microsatellites in the genome of *Drosophila melanogaster*.** *Mol Ecol* 1998, **7:**57-69.
12. Harr B, Zangerl B, Brem G, Schlötterer C: **Conservation of locus specific microsatellite variability across species: a comparison of two *Drosophila* sibling species *D. melanogaster* and *D. simulans*.** *Mol Biol Evol* 1998, **15:**176-184.
13. Goldstein DB, Clark AG: **Microsatellite variation in North American populations of *Drosophila melanogaster*.** *Nucleic Acids Res* 1995, **23:**3882-3886.
14. Schlötterer C, Ritter R, Harr B, Brem G: **High mutation rates of long microsatellite alleles in *Drosophila melanogaster* provide evidence for allele specific mutation rates.** *Mol Biol Evol* 1998, in press.
15. Schug MD, Mackay TFC, Aquadro CF: **Low mutation rates of microsatellite loci in *Drosophila melanogaster*.** *Nat Genet* 1997, **15:**99-102.
16. Goldstein DB, Ruiz Lineares A, Cavalli-Sforza LL, Feldman MW: **Genetic absolute dating based on microsatellites and the origin of modern humans.** *Proc Natl Acad Sci USA* 1995, **92:**6723-6727.
17. Kimmel M, Chakraborty R, Stivers DN, Deka R: **Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci.** *Genetics* 1996, **143:**549-555.
18. Goldstein DB, Pollock DD: **Launching microsatellites: a review of mutation processes and methods of phylogenetic inference.** *J Heredity* 1997, **88:**335-342.
19. Feldman MW, Bergman A, Pollock DD, Goldstein DB: **Microsatellite genetic distances with range constraints: analytic description and problems of estimation.** *Genetics* 1997, **145:**207-216.
20. Nauta MJ, Weissing FJ: **Constraints on allele size at microsatellite loci: implications for genetic differentiation.** *Genetics* 1996, **143:**1021-1032.
21. Caccone A, Amato GD, Powell JR: **Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup.** *Genetics* 1988, **118:**671-683.
22. Cariou M-L: **Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*.** *Genet Res* 1987, **50:**181-185.
23. Bodmer M, Ashburner M: **Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*.** *Nature* 1984, **309:**425-430.
24. Schlötterer C: **Are microsatellites really simple sequences?** *Curr Biol* 1998, **8:**R132-R134.
25. Nei M: **Genetic distance between populations.** *Am Nat* 1972, **106:**283-292.
26. Bowcock AM, Ruiz-Lineares A, Tonfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL: **High resolution of human evolutionary trees with polymorphic microsatellites.** *Nature* 1994, **368:**455-457.
27. Grimaldi M-C, Crouau-Roy B: **Microsatellite allelic homoplasy due to variable flanking sequence.** *J Mol Evol* 1997, **44:**336-340.
28. Garza JC, Slatkin M, Freimer NB: **Microsatellite allele frequencies in humans and chimpanzees with implications for constraints on allele size.** *Mol Biol Evol* 1995, **12:**594-603.
29. Caccone A, Moriyama EN, Gleason JM, Nigro L, Powell JR: **A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism data.** *Mol Biol Evol* 1996, **19:**1224-1232.
30. Hilton H, Kliman RM, Hey J: **Using hitchhiking genes to study adaptation and divergence during speciation within the *Drosophila melanogaster* complex.** *Evolution* 1994, **48:**1900-1913.
31. Hey J, Kliman RM: **Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex.** *Mol Biol Evol* 1993, **10:**804-822.
32. Kliman RM, Hey J: **DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex.** *Genetics* 1993, **133:**375-381.
33. Hey J, Kliman RM: **Genealogical portraits of speciation in the *Drosophila melanogaster* complex.** In *Non-neutral Evolution.* Edited by Golding B. New York: Chapman and Hall; 1994:208-216.
34. Solignac M, Monnerot M: **Race formation, speciation, and introgression within *Drosophila simulans*, *D. mauritiana* and *D. sechellia* inferred from mitochondrial DNA analysis.** *Evolution* 1986, **40:**531-539.
35. Singh RS: **Population genetics and evolution of species related to *Drosophila melanogaster*.** *Annu Rev Genet* 1989, **23:**425-453.
36. Clark AG: **Neutral behavior of shared polymorphism.** *Proc Natl Acad Sci USA* 1997, **94:**7730-7734.
37. Schlötterer C, Tautz D: **Slippage synthesis of simple sequence DNA.** *Nucleic Acids Res* 1992, **20:**211-215.
38. Minch E, Ruiz-Linares A, Goldstein D, Feldman M, Cavalli-Sforza LL: **Microsat (version 1.4d): a computer program for calculating various statistics on microsatellite allele data.** Stanford, California: University of Stanford; 1995.
39. Felsenstein J: **PHYLIP, version 3.57c.** Seattle, Washington: University of Washington; 1991.
40. Page RD: **TREEVIEW: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12:**357-358.

Supplementary material

A microsatellite-based multilocus phylogeny of the *Drosophila melanogaster* species complex

Bettina Harr, Steven Weiss, Jean R. David, Gottfried Brem and Christian Schlötterer

Current Biology 12 October 1998, **8**:1183–1186

Table S1

Mean pairwise distances (determined by 100 bootstrap replicates).

$(\delta\mu)^2$			
	mau	*sec*	*sim*
sec	54.171		
sim	21.538	48.961	
mel	72.630	86.606	66.087
Nei's distance			
	mau	*sec*	*sim*
sec	1.393		
sim	0.593	1.153	
mel	2.350	2.100	1.982
Proportion of shared alleles			
	mau	*sec*	*sim*
sec	1.675		
sim	0.919	1.484	
mel	2.484	2.333	2.068

The species names are abbreviated to their first three letters.