Accepted on 12 July 2006
J Zool Syst Evol Res doi: 10.1111/j.1439-0469.2006.00389.x

[1]*UPR 9034 Evolution, génomes et spéciation, CNRS, Gif sur Yvette cedex, France;* [2]*INRA, Centre de Biologie et de Gestion des Populations, Montferrier-sur-Lez, France;* [3]*Institut Curie, Orsay, France;* [4]*IRD, UR R072, c/o CNRS UPR 9034, Gif sur Yvette cedex, France*

# A phylogeny of Drosophilidae using the *Amyrel* gene: questioning the *Drosophila melanogaster* species group boundaries

J.-L. Da Lage[1], G. J. Kergoat[2], F. Maczkowiak[3], J.-F. Silvain[4], M.-L. Cariou[1] and D. Lachaise[1]*

## Abstract

In this study, the phylogenetic relationships of 164 species of the family Drosophilidae are discussed, using the *Amyrel* gene, a member of the α-amylase multigene family. This study focuses on numerous species groups in the subgenera *Sophophora* and *Drosophila* of the genus *Drosophila* but also includes other closely related genera. Nucleotide data were analysed by several methods: maximum parsimony, neighbour joining, maximum likelihood and Bayesian inference. Heterogeneity of base composition (mainly low GC contents in the species groups *willistoni* and *saltans*) has been addressed. In all analyses, the genus *Drosophila* appeared paraphyletic. The subgenus *Sophophora* clearly appeared to be a monophyletic group, showing well-resolved clades, with the Neotropical groups arising in a basal position. Here, it is proposed to raise the species subgroups *ananassae* and *montium* to the rank of species group, and to restrict the *melanogaster* species group to the *melanogaster* subgroup plus the 'Oriental' subgroups, among which the *suzukii* subgroup is polyphyletic. Some related genera such as *Zaprionus*, *Liodrosophila*, *Scaptomyza* and *Hirtodrosophila* are clustered with, or inside the subgenus *Drosophila*, which is therefore paraphyletic and should be reviewed.

**Key words:** Molecular phylogeny – *Zaprionus* – *ananassae* species group – *montium* species group – base composition bias – *Sophophora*

## Introduction

In a few years, 100 years of using *Drosophila melanogaster* as a laboratory model will be celebrated. This little fly has been widely used in all fields of biology, including genetics, cell biology, ecology, physiology, developmental biology and evolution. Among other obvious, convenient and scientific features of *D. melanogaster*, which make it suitable for this purpose, it has the advantage of belonging to a genus comprising more than 3500 species distributed worldwide in a variety of climates and ecological niches (Powell 1997; Ashburner et al. 2005). This has made it a gold mine for research into evolutionary genetics and ecology. The occurrence of a number of pairs or complexes of sibling species among the genus *Drosophila* has led some researchers to focus on speciation and its mechanisms (Powell 1997, p. 213–266; Wu 2001; Coyne and Orr 2004).

More recently, in developmental genetics for instance, it has become necessary to compare between more and less closely related species (DeSalle and Grimaldi 1993; Gompell and Carroll 2003). Thus, apart from the interest of phylogenetic relationships *per se* between drosophilid species, it has become even more crucial to have accurate and extensive phylogenetic trees, covering as many species as possible.

Following the pioneer works of Throckmorton (1975) or Grimaldi (1990), several attempts to reconstruct species trees of *Drosophila* have been published since the early 1990s. Molecular data have been used to produce numerous trees, using rDNA (Pélandakis et al. 1991; Pélandakis and Solignac 1993), mtDNA (DeSalle 1992; Gleason and Powell 1997; Goto and Kimura 2001; Kastanis et al. 2003), nuclear genes or combined data (e.g. Powell 1997; Durando et al. 2000; Kopp and True 2002; O'Grady and Kidwell 2002; Remsen and O'Grady 2002; Robe et al. 2005), sometimes including morphology (Remsen and O'Grady 2002; Schawaroch 2002). These works have usually focused on particular taxonomic groups, or inversely, at a larger scale, on relationships between *Drosophila* subgenera and closely related genera. Molecular studies have often been hindered by the varied evolutionary patterns displayed by the genes used, such as composition bias and lineage-specific substitution rates (Tarrio et al. 2001).

Some consistent data are now available, such as the monophyly of the subgenus *Sophophora* (O'Grady and Kidwell 2002, but see Powell 1997; Katoh et al. 2000), relationships between some species groups and the existence of deep (ancient) radiations within the subgenus *Drosophila*. There is still no large-scale phylogeny of Drosophilidae, including distant species, but also showing more details in some species groups. An attempt was made to produce such a phylogeny using sequence data from the *Amyrel* gene.

*Amyrel* is a nuclear gene, a paralogue of the α-amylase (*Amy*) genes. The usefulness of *Amy* genes in molecular phylogenetics is usually limited by the fact that they are mostly found in multiple copies, unless each copy has been identified in each species (see Inomata et al. 1997; Zhang et al. 2003; also Kopp 2006). In contrast, *Amyrel* is a single-copy gene (Da Lage et al. 1998). It has 40% divergence with *Amy* in amino acids and is easily distinguished from *Amy*. It was first identified in species of the subgenus *Sophophora*, and was originally thought to be restricted to this subgenus (Da Lage et al. 1998). Further investigations have shown that *Amyrel* is also present in other families of the Muscomorpha clade (Da Lage et al. 2002; Maczkowiak and Da Lage, 2006). In *Drosophila*, the coding sequence is about 1470–1485 bp in length, with a single short intron located near the middle of the gene (position 655 in *D. melanogaster*). *Amyrel* is an attractive candidate for use in producing molecular phylogenies because it can be entirely amplified by PCR using external primers. Conserved regions exist in 5′ and 3′, which make it possible to design suitable primers (Da Lage et al. 1998). *Amyrel* has been already used to resolve phylogenetic relationships (Cariou

*Dr Daniel Lachaise died on 2 July 2006, suddenly and unexpectedly. He has contributed very much to this article.

et al. 2001; Kopp 2006). It is also informative at the intraspecific level in several species (M.L. Cariou, unpublished data).

A single-gene phylogeny with a large amount of original data is presented here. This single-gene approach has been chosen to be retained in order to increase the number of species covered by this study, rather than a multicharacter strategy, with a lesser number of taxa. Recent works on various organisms (Gontcharov et al. 2004; May-Collado and Agnarsson 2006) have suggested that dense taxon sampling on a single gene is relevant for improving phylogenetic accuracy. In addition, recently, Kopp (2006) used *Amyrel* in a multigene approach of the *D. melanogaster* species group. Clearly, his *Amyrel* tree was the most congruent with the retained consensus tree. Despite the drawbacks of using only one gene, this work has the advantage of covering a large number of species, some of which were never included in previously published molecular phylogenies. The complete sequences for 146 species have been obtained. Partial sequences have also been obtained from 18 other species.

## Materials and Methods

### Flies and DNA extraction

Live specimens of Drosophilidae species were available from the Gif laboratory collection, which also contained frozen and alcohol-preserved specimens of other species. The species are listed in Table 1. Genomic DNAs from single individuals were prepared as described previously (Gloor and Engels 1992).

### *Amyrel* amplification

External primers were designed for the subgenus *Sophophora*, using the conserved regions described in Da Lage et al. (1998): RELZONE2 (forward) TCGTAAATTGGACCCAAGCG; RELAVALREV (reverse) CATACATTATGTGCGTTCG. Using this pair of primers, it has been found that reducing the elongation temperature yielded more products, perhaps because of the low $T_m$ of the reverse primer. The cycles were as follows: denaturation 94°C 30 s; annealing 53°C 1 min; extension 65°C 2 min; 40 cycles. The size of amplified DNA was ca. 1.7 kb. For the non-*Sophophora* species, no data were available pertaining to the conservation of external regions. Amplifications with internal primers previously used in sequencing were obtained thus: 1U (forward) GTTCACCTCTTCGAGTGG, REV1230 (reverse) TTGCTGCCRTTRTCCCACC. The *Amyrel* gene of *Drosophila virilis* was then cloned from a genomic minilibrary; and the genes of *Drosophila funebris* and *Hirtodrosophila confusa* were completed using the Universal Genome Walker Kit (Clontech, Mountain View, CA). An alignment of these three sequences enabled to design external primers that could be used for the non-*Sophophora* species: ZONE2BIS (forward) GTAAATNGGNNCCACGCGAAG; RELAVBIS (reverse) GCATTTGTACCGTTT GTGTCGTTATCG. However, in many cases, it was more convenient to obtain two shorter overlapping fragments, ca. 1 kb each, between ZONE2BIS and RELREV + (GTTCCCCAGCTCTGCAGCC) for the left part and between RELUDIR (TGGATGCNGCCAAGCACATGGC) and RELAVBIS for the right part. The positioning of the primers used in this study is shown in Fig. 1.

When possible, PCR products were directly sequenced on an ABI373 sequencer, otherwise they were cloned in the pGEM-T plasmid vector (Promega, Madison, WI). In this case, several clones were sequenced and one has been retained in the data set.

The *Amyrel* gene from the medfly *Ceratitis capitata* (Tephritidae) was cloned from a minilibrary using a fragment of the *Amy* gene of this species as a probe. All the sequences were deposited in GenBank. The accession numbers are shown in Table 1.

### Sequence alignment

The DNA sequences, excluding introns, were translated, and then the protein sequences were aligned using CLUSTALW (Thompson et al. 1994). This alignment served as a guide to manually align the DNA sequences. Ambiguities putatively occurred only in the signal peptide, about 20 amino acids in length, which is highly variable between distant species. A total of 164 taxa were aligned (1515 positions). Using a chi-square test implemented in PAUP* version 4.0b10 (Swofford 2002), significant base composition heterogeneity was detected between taxa.

### Phylogenetic reconstruction methods

Several methods were used to reconstruct the phylogenetic relationships of the species sampled. PAUP* version 4.0b10 (Swofford 2002) was used to carry out both neighbour joining (NJ) and maximum parsimony (MP) analyses. Bayesian inference (BI) analyses were conducted using MrBayes version 3.1.1 (Huelsenbeck and Ronquist 2001), whereas PHYML version 2.1b (Guindon and Gascuel 2003) was used for maximum likelihood (ML) inference. All these analyses used two outgroups: the close relative *Leucophenga maculata* (Drosophilidae, Steganinae) (accession number DQ021938) and the more distant *C. capitata* (Tephritidae) (accession number AF146758).

Neighbour joining analyses were conducted as preliminary analyses with both Kimura two-parameters (K2P) and LogDet (LD) corrections. The latter correction was used in an attempt to better take into account the significant heterogeneity in base composition between taxa (Galtier and Gouy 1998).

Given the large number of taxa in the data set, all parsimony analyses were performed using heuristic search option (tree bisection and reconnection (TBR), random sequence addition). To better explore the tree space when working with a large number of taxa, 1000 random addition replicates were used and TBR branch swapping was limited to a maximum of 500 trees (i.e. MaxTrees set to 500), as proposed by Sorenson (1999). Both unweighted and weighted analyses were conducted. Weighted analyses were carried out, in which the third codon positions were underweighted, to minimize the effect of transitions that may accumulate at high frequency at the third codon position because of the degeneracy of the genetic code. To determine the sensitivity of the phylogenetic reconstructions to weighting, the weighting of the first and second codon positions has been progressively increased over the third codon positions using three differential weightings of the three codon positions (2 : 2 : 1, 4 : 4 : 1 and 10 : 10 : 1). Given the fact that transitions are more frequent than transversions (Brown et al. 1982), a weighted analysis of the data set was also performed following Hillis et al. (1994), who recommended increasing the resolution of parsimony analyses by downweighting transitions. Since similar results are generally obtained when using different numbers to downweight transitions versus transversions (Barker and Lanyon 2000), a single weighting scheme was used for downweighting transitions (transversions were weighted twice over transitions). The robustness of topologies was assessed by bootstrap procedures (1000 replicates) and by estimating Bremer support (BS) values (Bremer 1994) using TreeRot version 2.0 (Sorenson 1999).

Before carrying out the BI and ML analyses, Modeltest version 3.06 (Posada and Crandall 1998) was used to determine the best-fit substitution model for the data under a likelihood framework. The Akaike information criterion selected the general time reversible (GTR) model including the proportion of invariable sites and the gamma distribution of rate variation among sites (GTR + I + $\Gamma$: Lanave et al. 1984; Yang 1994; Gu et al. 1995) as the best-fit evolutionary model. For BI analyses, four distinct 2 000 000 generation runs were conducted using the Markov Chain Monte Carlo (MCMC) algorithm implemented in MrBayes (four incrementally heated chains were used with a GTR + I + $\Gamma$ model and trees were saved to a file every 100 generations). Identical topologies were recovered from all four runs. A burn-in period of 100 000 generations was identified graphically by plotting likelihood values for each generation. The results were presented in the form of a 50% majority-rule consensus tree (in which trees corresponding to the burn-in period were discarded) and the support for the nodes of the tree were given by posterior probability estimates of clades. The ML analysis was performed using an input tree generated by an NJ heuristic search (BIONJ option in PHYML), the GTR + I + $\Gamma$ model and all parameters estimated from the data. Branch support was assessed by 1000 bootstrap replicates using ML distances as implemented in PAUP*.

Table 1. List of *Drosophila* taxa used in this study. Asterisks indicate partial *Amyrel* sequences. The new terminology proposed in this article was used in the former *melanogaster* species group. Group and subgroup taxonomy follows that proposed by Bächli (1999)

| Subgenus (or genus) | Group | Subgroup | Species | Origin | Accession number |
|---|---|---|---|---|---|
| *Drosophila* | annulimana | | *aracataca* (Vilela and Val, 1983) | Costa Rica | AY733052 |
| | | | *ararama* (Pavan and Cunha, 1947) | | AY733048 |
| | | | *gibberosa* (Patterson and Mainland, 1943) | Mexico | AY733041 |
| | | | *talamancana* (Wheeler, 1968) | El Salvador | AY736509 |
| | bromeliae | | *bromeliae* (Sturtevant, 1921) | Guadeloupe | AY733049 |
| | cardini | cardini | *cardini* (Sturtevant, 1916) | Martinique | AF462599 |
| | | cardini | *neomorpha* (Heed and Wheeler, 1957) | | AY736481 |
| | | cardini | *polymorpha* (Dobzhansky and Pavan, 1943) | Florida | AY736495 |
| | | dunni | *arawakana* (Heed, 1962) | Guadeloupe | AF491630 |
| | | dunni | *caribiana* (Heed, 1962) | Martinique | AY733050 |
| | dreyfusi | | *camargoi* (Pavan, 1950) | British Guyana | AF462598 |
| | funebris | funebris | *funebris* (Fabricius, 1787) | France | AF335557 |
| | guarani | guarani | *guaru* (Dobzhansky and Pavan, 1943) | | AF491631 |
| | histrio | | *sternopleuralis* (Okada and Kurokawa, 1957) | Japan | AY736505 |
| | immigrans | hypocausta | *hypocausta* (Osten Sacken, 1882) | | AY733043 |
| | | hypocausta | *rubida (Mather*,1960) | New Ireland | AY736502 |
| | | hypocausta | *siamana* (Ikeda et al. 1983) | Thailand | AY736504 |
| | | immigrans | *formosana* (Duda, 1926) | | AF462601 |
| | | immigrans | *immigrans* (Sturtevant, 1921) | New Caledonia | AF491632 |
| | | immigrans | *ruberrima* (de Meijere, 1911) | | AY736501 |
| | | immigrans | *ustulata* (de Meijere, 1908) | Indonesia | AY736541* |
| | | nasuta | *albomicans* (Duda, 1923) | Thailand | AF462595 |
| | | nasuta | *kepulauana* (Wheeler, 1969) | | AY733044 |
| | | nasuta | *nasuta* (Lamb, 1914) | | AY733059 |
| | | nasuta | *pallidifrons* (Wheeler, 1969) | New Caledonia | AY736486 |
| | | nasuta | *sulfurigaster* (Duda, 1923) | New Caledonia | AY736508 |
| | | | *trilimbata* (Bezzi, 1928) | | AY736512 |
| | melanica | | *melanica* (Sturtevant, 1916) | Arizona | AY733056 |
| | | | *tsigana* (Burla and Gloor, 1952) | | AY736513 |
| | mesophragmatica | | *pavani* (Brncic, 1957) | Argentina | AY736490 |
| | Modified mouthparts | mimica | *mimica* (Hardy, 1965) | Hawaii | AY736537* |
| | pallidipennis | | *pallidipennis* (Patterson and Mainland, 1943) | Panama | AY736487 |
| | Picture wing | grimshawi | *grimshawi* (Oldenberg, 1914) | Hawaii | AY736533* |
| | polychaeta | | *daruma* (Okada, 1956) | Okinawa | AY736532* |
| | | | *iri (hirtipes)* (Burla, 1954) | Congo | AF491633 |
| | | | *latifshahi* (Gupta and Ray-Chaudhury, 1970) | India | AY736536* |
| | | | *polychaeta* (Patterson and Wheeler, 1942) | Guadeloupe | AY736494 |
| | quinaria | | *angularis* (Okada, 1956) | Japan | AY736530* |
| | | | *kuntzei* (Duda, 1924) | | AF491634 |
| | | | *limbata* (von Roser, 1840) | France | AF491636 |
| | | | *nigromaculata* (Kikkawa and Peng, 1938) | Japan | AY736483 |
| | | | *palustris* (Spencer, 1942) | | AY736538* |
| | | | *phalerata* (Meigen, 1830) | | AY736492 |
| | | | *subpalustris* (Spencer, 1942) | | AY736539* |
| | | | *transversa* (Fallen, 1823) | | AY736511 |
| | repleta | hydei | *hydei* (Sturtevant, 1921) | | AY733042 |
| | | hydei | *hydeoides* (Patterson and Wheeler, 1942) | | AY736534* |
| | | mulleri | *nigrodumosa* (Wasserman and Fontdevila, 1990) | Venezuela | AY736482 |
| | | mulleri | *wheeleri* (Patterson and Alexander, 1952) | | AY736514 |
| | | repleta | *limensis* (Pavan and Patterson, 1947) | Peru | AY736479 |
| | | repleta | *repleta* (Wollaston, 1858) | France | AY736496 |
| | testacea | | *testacea* (von Roser, 1840) | Caucase | AY736510 |
| | tripunctata | iii | *mediopictoides* (Heed and Wheeler, 1957) | Panama | AY733055 |
| | | iv | *albirostris* (Sturtevant, 1921) | | AY733047 |
| | | iv | *metzii* (Sturtevant, 1921) | | AY744447 |
| | tumiditarsus | | *repletoides* (Hsu, 1943) | China | AY736500 |
| | virilis | americana | *americana* (Spencer, 1938) | | AY736529* |
| | | americana | *americana texana* (Patterson, 1940) | | AY736540* |
| | | virilis | *virilis* (Sturtevant, 1916) | Spain | AF136603 |
| | | | *kanekoi* (Watabe and Higuchi, 1979) | Japan | AY736535* |
| | | | *littoralis* (Meigen, 1830) | | AY733045 |
| | | | *lummei* (Hackman, 1972) | Finland | AY733046 |
| | | | *novamexicana* (Patterson, 1941) | | AY736484 |
| | Ungrouped | | *adamsi* (Wheeler, 1959) | | AY736528* |
| | Ungrouped | | *aracea* (Heed and Wheeler, 1957) | Mexico | AY736531* |
| | Ungrouped | | *pruinosa* (Duda, 1940) | Madagascar | AY756177 |
| *Sophophora* | ananassae | | *ananassae* (Doleschall, 1858) | Ivory Coast | AF024691 |

Table 1. (Continued)

| Subgenus (or genus) | Group | Subgroup | Species | Origin | Accession number |
|---|---|---|---|---|---|
| | | | *atripex* (Bock and Wheeler, 1972) | Thailand | U96154 |
| | | | *bipectinata* (Duda, 1923) | Thailand | AF136936 |
| | | | *ercepeae* (Tsacas and David, 1975) | Réunion | U96155 |
| | | | *lachaisei* (Tsacas, 1984) | Sao Tomé | AY736478 |
| | | | *malerkotliana pallens* (Bock and Wheeler, 1972) | Borneo | AY733054 |
| | | | *malerkotliana* (Parshad and Paika, 1964) | Madagascar | AY733053 |
| | | | *merina* (Tsacas, 1997) | Madagascar | AY733057 |
| | | | *monieri* (McEvey et al. 1987) | Moorea | AF250052 |
| | | | *ochrogaster* (Chassagnard, 1992) | New Caledonia | AY736485 |
| | | | *pallidosa* (Bock and Wheeler, 1972) | Samoa | AF136931 |
| | | | *papuensis*-like | New Guinea | AY736488 |
| | | | *parabipectinata* (Bock, 1971) | Mauritius | AY736489 |
| | | | *phaeopleura* (Bock and Wheeler, 1972) | Fiji | AY736491 |
| | | | *pseudoananassae pseudoananassae* (Bock, 1971) | Thailand | AY736498 |
| | | | *pseudoananassae nigrens* (Bock and Wheeler, 1972) | Thailand | AY736497 |
| | | | *vallismaia* (Tsacas, 1984) | Seychelles | AY744446 |
| | | | *varians* (Bock and Wheeler, 1972) | Philippines | AF136937 |
| | melanogaster | elegans | *elegans* (Bock and Wheeler, 1972) | Philippines | AF136930 |
| | | elegans | *subelegans* (Okada, 1988) | | AY736507 |
| | | eugracilis | *eugracilis* (Bock and Wheeler, 1972) | Thailand | AF250055 |
| | | ficusphila | *ficusphila* (Kikkawa and Peng, 1938) | Taiwan | AF462600 |
| | | ficusphila | *levii* (Tsacas, 1988) | New Caledonia | AF491635 |
| | | flavohirta | *flavohirta* (Malloch, 1924) | Australia | AY733051 |
| | | melanogaster | *erecta* (Tsacas and Lachaise, 1974) | Ivory Coast | AF039562 |
| | | melanogaster | *mauritiana* (Tsacas and David, 1974) | Mauritius | U96157 |
| | | melanogaster | *melanogaster* (Meigen, 1830) | Canton S | AF022713 |
| | | melanogaster | *orena* (Tsacas and David, 1978) | Cameroon | U96158 |
| | | melanogaster | *santomea* (Lachaise and Harry, 2000) | Sao Tomé | AY736503 |
| | | melanogaster | *sechellia* (Tsacas and Bächli, 1981) | Seychelles | AF039558 |
| | | melanogaster | *simulans* (Sturtevant, 1919) | Réunion | U96159 |
| | | melanogaster | *teissieri* (Tsacas, 1971) | Congo | AF039557 |
| | | melanogaster | *yakuba* (Burla, 1954) | Cameroon | AF039561 |
| | | suzukii | *biarmipes* (Malloch, 1924) | India | AF462597 |
| | | suzukii | *lucipennis* (Lin, 1972) | Taiwan | AF251138 |
| | | suzukii | *mimetica* (Bock and Wheeler, 1972) | Malaysia | AY733058 |
| | | takahashii | *lutescens* (Okada, 1975) | Japan | AF491637 |
| | | takahashii | *pseudotakahashii* (Mather, 1957) | Australia | AY736499 |
| | | takahashii | *takahashii* (Sturtevant, 1927) | Thailand | U96161 |
| | montium | | *asahinai* (Okada, 1964) | Japan | AF250051 |
| | | | *auraria* (Peng, 1937) | China | U96163 |
| | | | *bakoue* (Tsacas and Lachaise, 1974) | Bénin | U96162 |
| | | | *barbarae* (Bock and Wheeler, 1972) | Thailand | AF250053 |
| | | | *biauraria* (Bock and Wheeler, 1972) | Japan | AF136932 |
| | | | *bicornuta* (Bock and Wheeler, 1972) | Taiwan | AF136933 |
| | | | *bocqueti* (Tsacas and Lachaise, 1974) | Ivory Coast | AF049092 |
| | | | *burlai* (Tsacas and Lachaise, 1974) | Ivory Coast | AF250059 |
| | | | *cauverii* (Muniyappa et al. 1982) | India | AF251137 |
| | | | *chauvacae* (Tsacas, 1984) | Madagascar | AF250056 |
| | | | *davidi* (Tsacas, 1975) | Congo | AF251139 |
| | | | *diplacantha* (Tsacas and David, 1977) | Cameroon | AF251142 |
| | | | *dossoui* (Chassagnard, 1991) | Bénin | U96164 |
| | | | *greeni* (Bock and Wheeler, 1972) | Ivory Coast | AF462602 |
| | | | *jambulina* (Parshad and Paika, 1964) | India | AF174489 |
| | | | *kikkawai* (Burla, 1954) | Madagascar | U96156 |
| | | | *leontia* (Tsacas and David, 1977) | India | AF250058 |
| | | | *lini* (Bock and Wheeler, 1972) | Taiwan | AF039559 |
| | | | *malagassya* (Tsacas and Rafael, 1982) | Madagascar | AF250057 |
| | | | *nagarholensis* (Prakash and Reddy, 1980) | India | AF250054 |
| | | | *nikananu* (Burla, 1954) | Cameroon | AF251136 |
| | | | nsp. *montium* (Lachaise, unpublished data) | Príncipe Island | |
| | | | *punjabiensis* (Parshad and Paika, 1964) | Thailand | U96165 |
| | | | *quadraria* (Bock and Wheeler, 1972) | Taiwan | AF136934 |
| | | | *rufa* (Kikkawa and Peng, 1938) | Japan | AF136935 |
| | | | *serrata* (Malloch, 1927) | Australia | AF069756 |
| | | | *triauraria* (Bock and Wheeler, 1972) | Japan | AF251141 |
| | | | *tsacasi* (Bock and Wheeler, 1972) | Ivory Coast | AF251134 |
| | | | *bocqueti*-like (Lachaise, unpublished data) | Congo | AF251131 |

Table 1. (Continued)

| Subgenus (or genus) | Group | Subgroup | Species | Origin | Accession number |
|---|---|---|---|---|---|
| | | | *vulcana* (Graber, 1957) | Zimbabwe | AF251132 |
| | obscura | obscura | *bifasciata* (Pomini, 1940) | Finland | AF251135 |
| | | obscura | *imaii* (Moriwaki and Okada, 1967) | Japan | AF251133 |
| | | obscura | *subobscura* (Collin, 1936) | France | U79724 |
| | | pseudoobscura | *pseudoobscura* (Frolova, 1929) | Arizona | U82556 |
| | | microlabis | *kitumensis* (Tsacas, 1985) | Kenya | AF306718 |
| | | affinis | *affinis* (Sturtevant, 1916) | Georgia (USA) | AF037353 |
| | saltans | | *neocordata* (Magalhaes, 1956) | Brazil | AY736480 |
| | | | *sturtevanti* (Duda, 1927) | Guadeloupe | AY736506 |
| | willistoni | | *nebulosa* (Sturtevant, 1916) | Guadeloupe | AY733060 |
| | | | *tropicalis* (Burla and Cunha, 1949) | El Salvador | AF251140 |
| | | | *willistoni* (Sturtevant, 1916) | Guadeloupe | AF039560 |
| *Chymomyza* | | | *amoena* (Loew, 1862) | Michigan | AY736544* |
| *Hirtodrosophila* | | | *confusa* (Staeger, 1844) | France | AF335559 |
| *Hirtodrosophila* | | | *pictiventris* (Duda, 1925) | | AY736493 |
| *Liodrosophila* | | | *aerea* (Okada, 1956) | Japan | AY736477 |
| *Mycodrosophila* | | | sp. | Russia (Sotchi) | AY736543* |
| *Scaptodrosophila* | | | *finitima* (Lamb, 1914) | Madagascar | AY736527 |
| *Scaptodrosophila* | | | *bryani* (Malloch, 1934) | Moorea | AY756178* |
| *Scaptomyza* | | | *pallida* (Zettzerstedt, 1847) | France | AY736542 |
| *Anaprionus* | | | *bogoriensis* (Mainx, 1958) | Thailand | AY736516 |
| *Anaprionus* | | | *lineosus* (Walker, 1860) | India | AY736521 |
| *Zaprionus* | | | *badyi* (Burla, 1954) | Ivory Coast | AY736515 |
| *Zaprionus* | | | *cercus* (Chassagnard and McEvey, 1992) | Madagascar | AY736517 |
| *Zaprionus* | | | *ghesquierei* (Collart, 1937) | Cameroon | AY736518 |
| *Zaprionus* | | | *inermis* (Collart, 1937) | Cameroon | AY736519 |
| *Zaprionus* | | | *kolodkinae* (Chassagnard and Tsacas, 1987) | Madagascar | AY736520 |
| *Zaprionus* | | | *mascariensis* (Tsacas and David, 1975) | Madagascar | AY736522 |
| *Zaprionus* | | | *sepsoides* (Duda, 1939) | Central Africa | AY736523 |
| *Zaprionus* | | | *sg vittiger* | Madagascar | AY736526 |
| *Zaprionus* | | | *tuberculatus* (Malloch, 1932) | Ivory Coast | AY736524 |
| *Zaprionus* | | | *verruca* (Chassagnard and McEvey, 1992) | Madagascar | AY736525 |

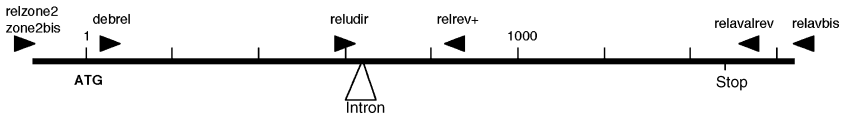



Fig. 1. Diagram of the *Amyrel* gene showing the positions of the primers used for amplification and sequencing; the intron position is also shown

## Results

### The data set

Complete sequences were obtained for 146 species and partial sequences for 18 species. All drosophilid species have an intron at the same position. This intron is almost always short, around 60 bp (minimum size 46 bp, maximum 284 bp, median 57 bp, mean 60 bp). The *Amyrel* sequence of *C. capitata* has two introns: the first one is located at the same position as the intron of classical *Amy* genes of *Drosophila* (Da Lage et al. 1996) and the second intron is homologous to the single intron of *Amyrel* genes of *Drosophila*. In *C. capitata*, the reading frame is much longer than that in *Drosophila* species. It has been truncated in the study. Many species were found to be polymorphic for *Amyrel*, that is, two haplotypes were found after cloning. High variability was found in both the length and the sequence of the signal peptide. An additional codon, in the last third of the sequence, was shared by two species of the *obscura* group, *Drosophila imaii* and *Drosophila bifasciata*.

The base composition of *Amyrel* is generally GC rich with a mean of 56%. The GC content at the third position of codons (GC3) is more variable between species groups, with a mean of 72%, but with extremes as low as 38.7% (*Drosophila sturtevanti*) and as high as 87.3% (*Drosophila lucipennis*) (Fig. 2). Indeed, the Neotropical groups of the subgenus *Sophophora*, i.e. the *willistoni* and *saltans* species groups, are strikingly GC poor, as has already been reported for several other genes (Rodriguez-Trelles et al. 2000). It is important to keep this in mind when looking at the phylogenetic trees below.

### General topology

The trees obtained by the different tree-building methods have been rooted with *L. maculata* and *C. capitata*. Two topologies, produced by MP 4 : 4 : 1 and BI, respectively, have been kept for illustration; but for better clarity only the BS values for MP 4 : 4 : 1 have been chosen for discussion (note that these values can be very high as they reflect the 4 : 4 : 1 weighted matrix). The results of the other weighting schemes or other reconstruction methods used will be discussed in the text. Two main clades emerged from both methods (Fig. 3). Clade 1

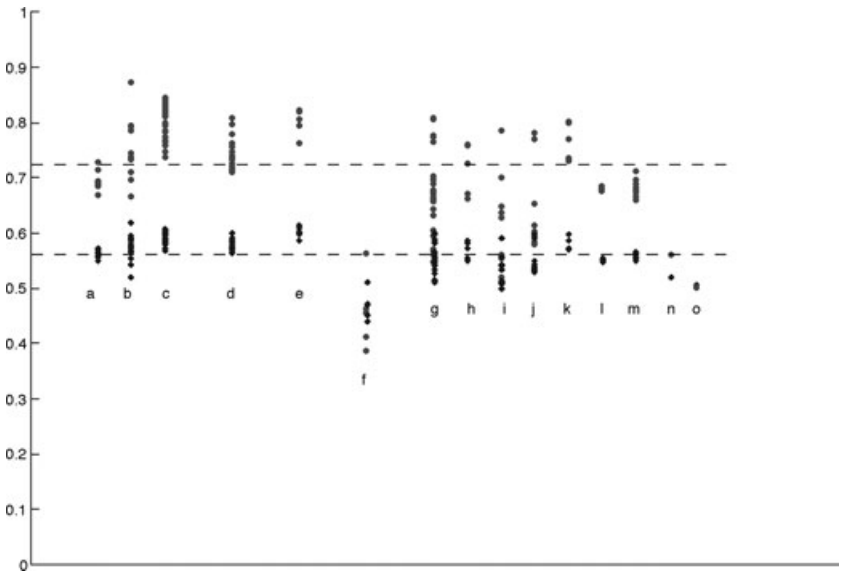



Fig. 2. Base composition of the *Amyrel* genes (complete coding sequences) of the species studied, grouped by taxonomic affinities. Lozenges are global GC contents; circles are GC content at the third codon positions. Dashed lines show the means for each parameter. a: *melanogaster* subgroup, b: Oriental subgroups, c: *montium* subgroup (new species group), d: *ananassae* subgroup (new species group), e: *obscura* group, f: Neotropical groups, g: various species in the subgenus *Drosophila*, h: *quinaria* group, i: *immigrans* group, j: *Zaprionus*, k: *repleta* group, l: *annulimana* group, m: *virilis* group, n: *S. finitima*, o: *C. capitata*

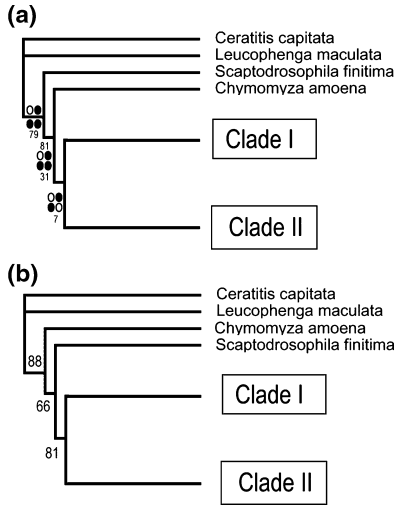
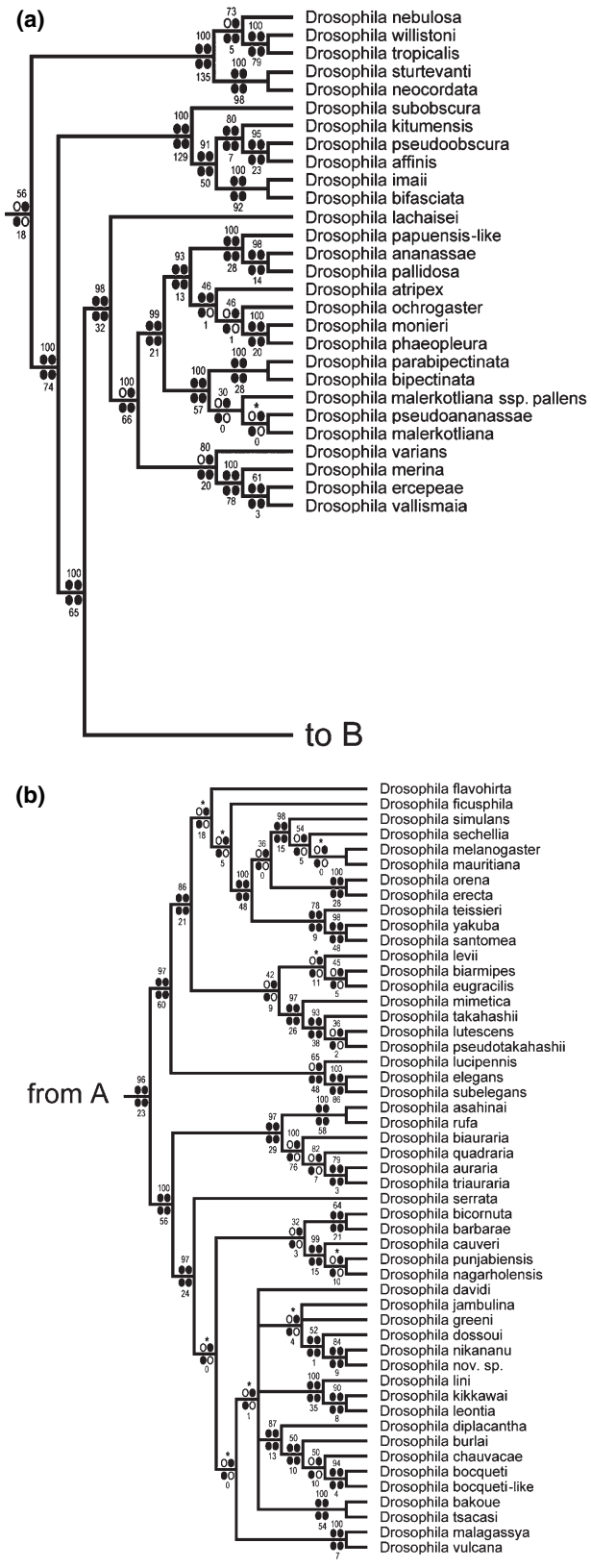



Fig. 3. Basal phylogenetic relationships obtained through parsimony (a) and Bayesian inference (b) analyses. Both inference methods recovered two identical large clades (labelled clade 1 and clade 2), the phylogenetic relationships of which are shown in detail in the corresponding figures (see Figs 4 and 5 for parsimony analyses and Figs 6 and 7 for Bayesian inference analysis). Legends are as in Figs 4 and 6

corresponds to the subgenus *Sophophora*, whereas clade 2 corresponds to the subgenus *Drosophila* and several taxa considered as distinct genera. As a consequence, the genus *Drosophila* clearly appears to be paraphyletic. More detailed results of MP analyses are shown in Figs 4 and 5. A tree obtained with 4 : 4 : 1 weighting is drawn, with an indication of its congruence with alternative MP analyses The unweighted MP (MP uw) analysis shows a paraphyletic subgenus *Sophophora*, with the Neotropical species groups *willistoni* and *saltans* branching off at the base of the tree. More over, the tree shows that *Sophophora* is included in a nested succession of clades belonging to the subgenus *Drosophila*. For instance, the *polychaeta* group is the sister group of a clade composed of the members of species groups *melanogaster* + *obscura*, which in turn is a sister of the *virilis–repleta* clade. Most of the deep nodes are poorly supported, except a highly supported *melanogaster* + *obscura* node (bootstrap value of 99%). This general topology is clearly incongruent with phylogenetic relationships already known to exist, especially for deeper nodes. MP 4 : 4 : 1 (Figs 3a, 4 and 5) and the two other codon-position-based weighted MP analyses (i.e. 2 : 2 : 1 and 10 : 10 : 1) recover two main clades: (i) the subgenus *Sophophora*, including the Neotropical groups, with a bootstrap value of 56% and BS of 18 (clade 1) and (ii) the subgenus *Drosophila* + the genera *Zaprionus*, *Hirtodrosophila*, *Liodrosophila*, *Scaptomyza*, *Mycodrosophila*, clustered together with a bootstrap support of 96% and BS of 87 (clade 2). In a more basal position, *Chymomyza amoena* and *Scaptodrosophila finitima* were recovered in a similar way to the MP uw analysis. The BI tree (Figs 3b, 6 and 7) also shows two main clades, with *Scaptodrosophila* appearing more closely related to clades 1 and 2 compared with *Chymomyza*. Neighbour joining analyses yield essentially the same general topology as weighted MP, except that the Neotropical species groups are located at the base of the tree, with *Ch. amoena* and *S. finitima* clustered together (bootstrap value of 65%). The results of K2P and LD analyses only differ by the placement of *Drosophila nebulosa* (see below) and by their somewhat distinct bootstrap support values. The ML analysis yields a general topology, which is mostly congruent with the result of the MP uw analysis. The two latter topologies only differ by the alternative placement of the Neotropical species groups, which are clustered with the subgenus *Sophophora* and related to *S. finitima* and *Ch. amoena* in the ML analysis. Thus, it can be summarized that a number of methods place the Neotropical groups *willistoni* and *saltans* just above the root of the tree (see below). The clade 2, composed by members of the subgenus *Drosophila* and some taxa hitherto classified as distinct genera, is recovered in most analyses, with a high support for both MP 4 : 4 : 1 and BI analyses. Interestingly, in MP reconstructions, the same phylogenetic hypothesis was found using all three distinct weighting schemes, in which the third position of codon was downweighted. This surprising result is consistent with previous observations made by Barker and Lanyon (2000), and suggests, for the data set pertaining to this study at least, that the exact value used for downweighting third codon




position is of little importance. In the following sections, the topologies of the different clades established by the different methods separately are presented.

### The subgenus *Sophophora*

The following species groups represented in this study are usually considered to be the members of the subgenus *Sophophora*: *melanogaster*, *obscura*, *saltans* and *willistoni*. Although some of the reconstruction methods (i.e. MP uw, NJ) suggest a very basal position for the Neotropical species groups *willistoni* and *saltans*, weighted MP, BI and ML analyses suggest a less basal position in which these two related species groups are clustered with other members of the subgenus *Sophophora* (thus forming clade 1). The remaining species groups, i.e. species groups *melanogaster* and *obscura*, form a monophyletic cluster. The nucleotide composition of *Amyrel* in the Neotropical species (Fig. 2), with an unusually low GC content, is likely to be responsible for the alternative placement of the species groups *willistoni* and *saltans* in both MP uw and NJ analyses. The latter hypothesis is well corroborated by the fact that a monophyletic subgenus *Sophophora* is recovered when the third codon positions are downweighted in MP analyses.

### The Neotropical species groups

Regardless of their position in the global tree, the relationships within the species groups *willistoni* and *saltans* vary slightly depending on the method used. *Drosophila willistoni* and *Drosophila tropicalis*, from the species group *willistoni*, and *D. sturtevanti* and *Drosophila neocordata*, from the species group *saltans*, are invariably grouped together, whereas the position of *D. nebulosa*, a member of the species group *willistoni*, is variable in this analyses. Weighted MP analyses cluster the latter species with the *willistoni–tropicalis* pair of species. Other methods show *D. nebulosa* in a more basal position, with a moderate-to-high support: bootstrap values of 71% (ML), 75% (NJ-K2P), 66% (MP uw) and posterior probabilities of 98% (BI). Interestingly, the GC3% of *D. nebulosa* is clearly higher (56.3%) than the GC3% of the other four species (42.9% on average). It is also worth mentioning that the NJ-LD method, which is less sensitive to compositional bias, regroups *D. nebulosa* with the species group *willistoni*.

### The *obscura* species group

In the data set pertaining to the study, the *obscura* species group appears to be the closest relative of the *melanogaster*

Fig. 4. Excerpt (corresponding to the clade 1 in Fig. 3) of the 50% majority-rule consensus tree (37 most equiparsimonious trees of 20 607 steps; CI = 0.21, RI = 0.76) from the 4 : 4 : 1 weighted parsimony analysis (heuristic search of 1000 random addition replicates). First and second codon positions were given a weight of four, whereas third positions were given a weight of one. The recovery of each node using four alternative weighting parameters is also given by two rows of circles (open circles figure the nodes that were not recovered under a specific weighting scheme). The top row corresponds to the results of the unweighted analysis (left) and the 2 : 2 : 1 weighted analysis (right). The bottom row corresponds to the results of the 10 : 10 : 1 weighted analysis (left) and the transition versus transversion weighted analysis (right). Numbers above the circle rows are bootstrap support values, whereas numbers below circle rows are Bremer support values (both bootstrap and Bremer support values were estimated under a 4 : 4 : 1 weighting scheme). The nodes that were not recovered by bootstrap replicates are indicated by an asterisk. In addition, to improve the clarity of the figure, the recovery and the support values of some terminal nodes (indicated by labels) are shown at the bottom of the tree

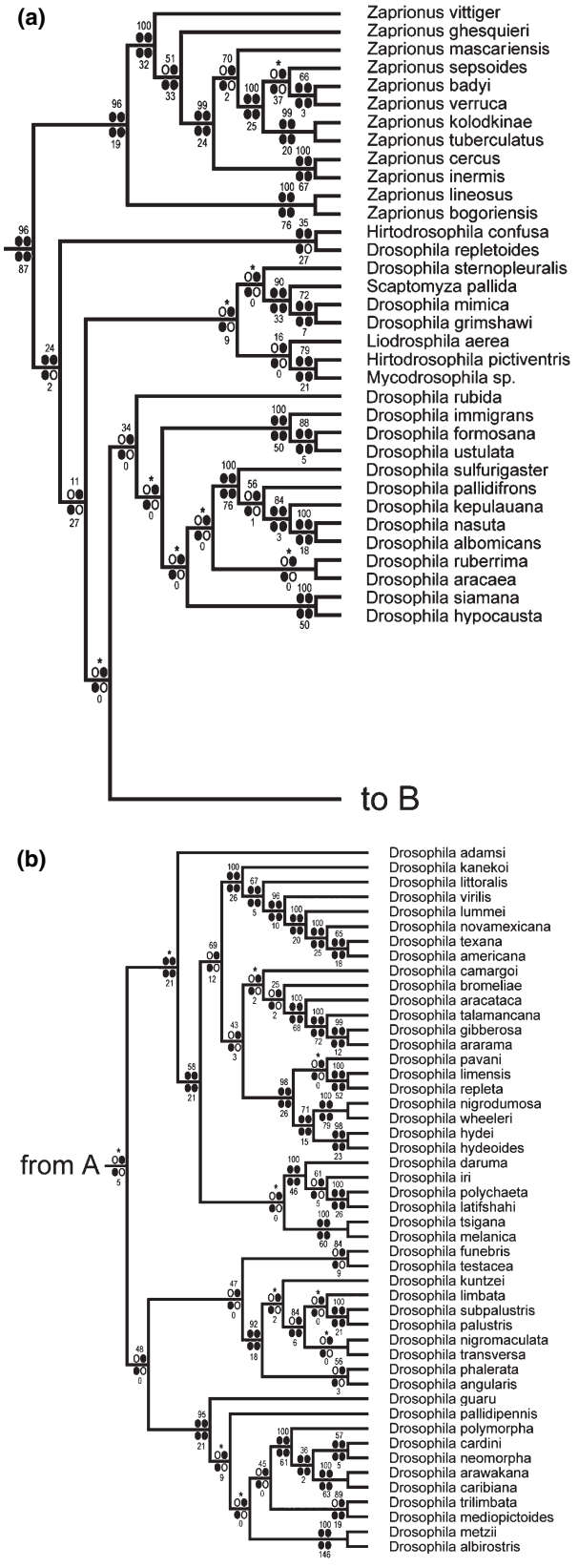



Fig. 5. Excerpt (corresponding to the clade 2 in Fig. 3.) of the 50% majority-rule consensus tree (37 most equiparsimonious trees of 20 607 steps; CI = 0.21, RI = 0.76) from the 4 : 4 : 1 weighted parsimony analysis (heuristic search of 1000 replicates of 100 random addition replicates each). Legend is as in Fig. 4

species group. The topologies for this species group were similar for all the tree-building methods, with high support.

### The Drosophila melanogaster species group

The *D. melanogaster* species group consists of 174 species (Bock and Wheeler 1972; Bock 1980; Lemeunier et al. 1986; Toda 1991). Twelve species subgroups have been described: *melanogaster*, *montium*, *ananassae*, *takahashii*, *suzukii*, *eugracilis*, *ficusphila*, *elegans*, *rhopaloa*, *denticulata*, *flavohirta* and *longissima*. Sixty-eight species or subspecies were studied: 30 from the *montium* subgroup (88 species described); all 9 known species in the *melanogaster* subgroup; 17 from the *ananassae* subgroup and several species from the so-called 'Oriental subgroups' –*takahashii*, *suzukii*, *elegans*, *ficusphila*, *eugracilis* and additionally, *Drosophila flavohirta*. All the reconstruction methods clearly show three main lineages with nested positions: (i) *melanogaster* + Oriental subgroups; (ii) *montium* subgroup; (iii) *ananassae* subgroup.

The *melanogaster* subgroup is well supported and is clustered with the Oriental subgroups. Within the *melanogaster* subgroup, the *melanogaster* complex is clearly apparent with *D. melanogaster* generally branching off from the triplet of species *mauritiana–simulans–sechellia* (best bootstrap support value of 87% in NJ-K2P), but not in weighted MP reconstructions, which result in a clearly unusual topology for the subgroup, in which *Drosophila simulans* is further from *mauritiana* + *sechellia* than *D. melanogaster* (Fig. 4). BI, MP uw and Ts/Tv, ML and NJ methods suggest a cluster *mauritiana–simulans*, but with a low support. The other part of the *melanogaster* subgroup shows another two clusters (*erecta* + *orena*) and (*yakuba* + *teissieri* + *santomea*). *Drosophila santomea* is sister to *yakuba*, as already shown with *Amy* + *Amyrel* (Cariou et al. 2001). Interestingly, these two clusters are found grouped together by several methods: BI and ML with high support; NJ, MP uw and Ts/Tv with an average support. Once again, MP 4 : 4 : 1, 2 : 2 : 1 and 10 : 10 : 1 show an inconsistent result, as the *yakuba* complex is found to branch off first. It is noteworthy that if the *melanogaster* and *yakuba* species complexes are reliably based on consistent morphological and molecular evidence, the *erecta* + *orena* grouping is anything but a deep-rooted clustering. Although appearing as sister species in the tree reconstructions, these two species actually differ markedly from one another with regard to their morphology, and it has previously been suggested that they should not be included formally into a species complex *sensu stricto* (Lachaise and Silvain 2004).

The relationships of the originally Afrotropical *melanogaster* subgroup with the Oriental subgroups, and between these subgroups, remain unclear. Their hierarchical position depends on the reconstruction methods used. In all the methods, *D. lucipennis* diverges first from the clade (*melanogaster* subgroup + Oriental subgroups), and in all methods apart from MP uw, this species is clustered with the *elegans* subgroup, with high support (over 75%). Then *Drosophila ficusphila* diverges in most methods. The relationships within the ingroup are less congruent between the different trees. However, the *suzukii* subgroup clearly appears to be polyphyletic, with *Drosophila mimetica* strongly attached to the *takahashii* subgroup, then to *Drosophila biarmipes*, whereas *D. lucipennis*, another member assigned to the *suzukii* subgroup, exhibits a close relationship with the *elegans* subgroup

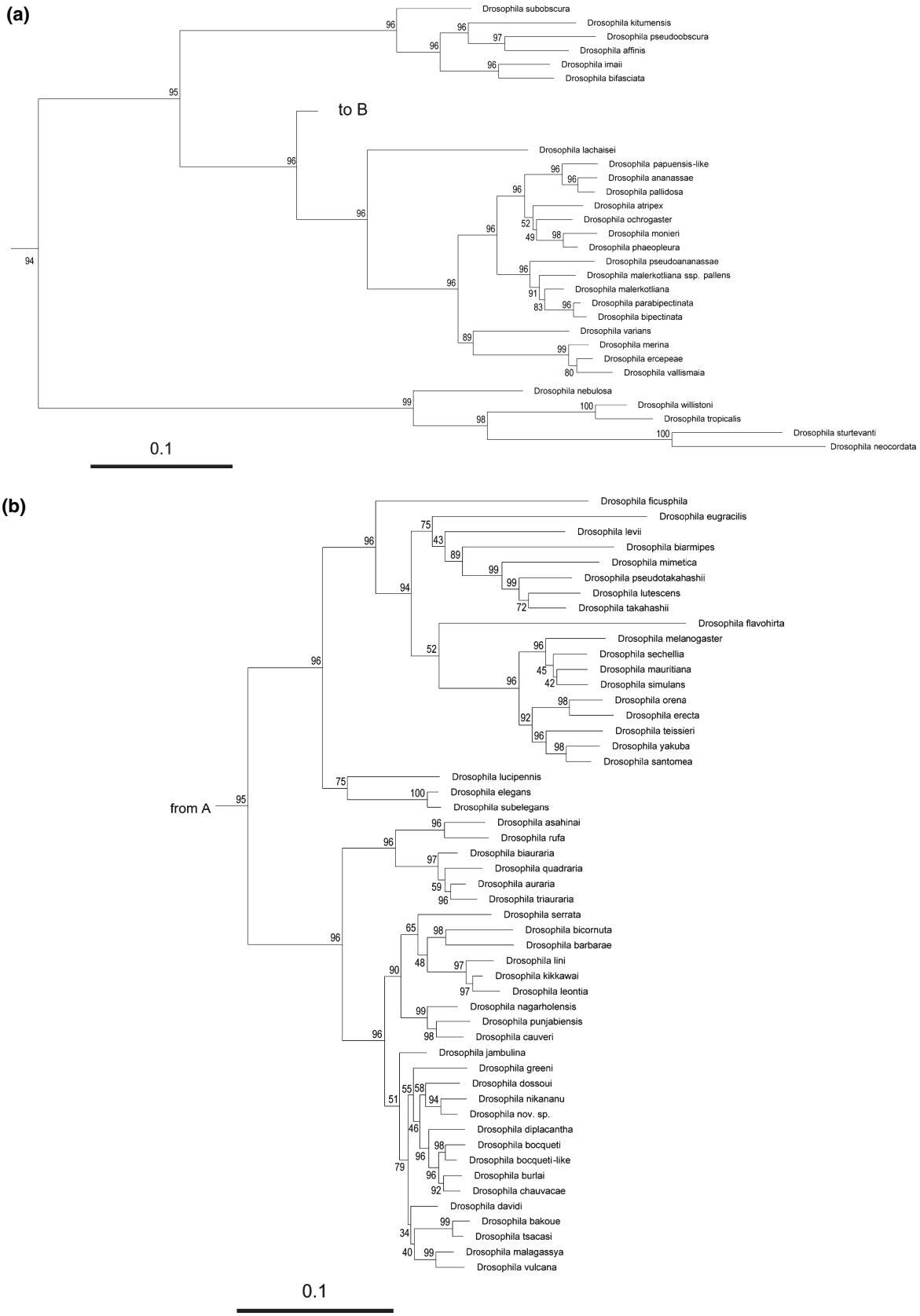



Fig. 6. Excerpt (corresponding to clade 1 in Fig. 3.) of the 50% majority-rule consensus tree from the Bayesian inference analysis (2 000 000 generations). Identical topologies were recovered in four distinct runs of MrBayes. Numbers at the nodes indicate the clade posterior probability estimates for each node

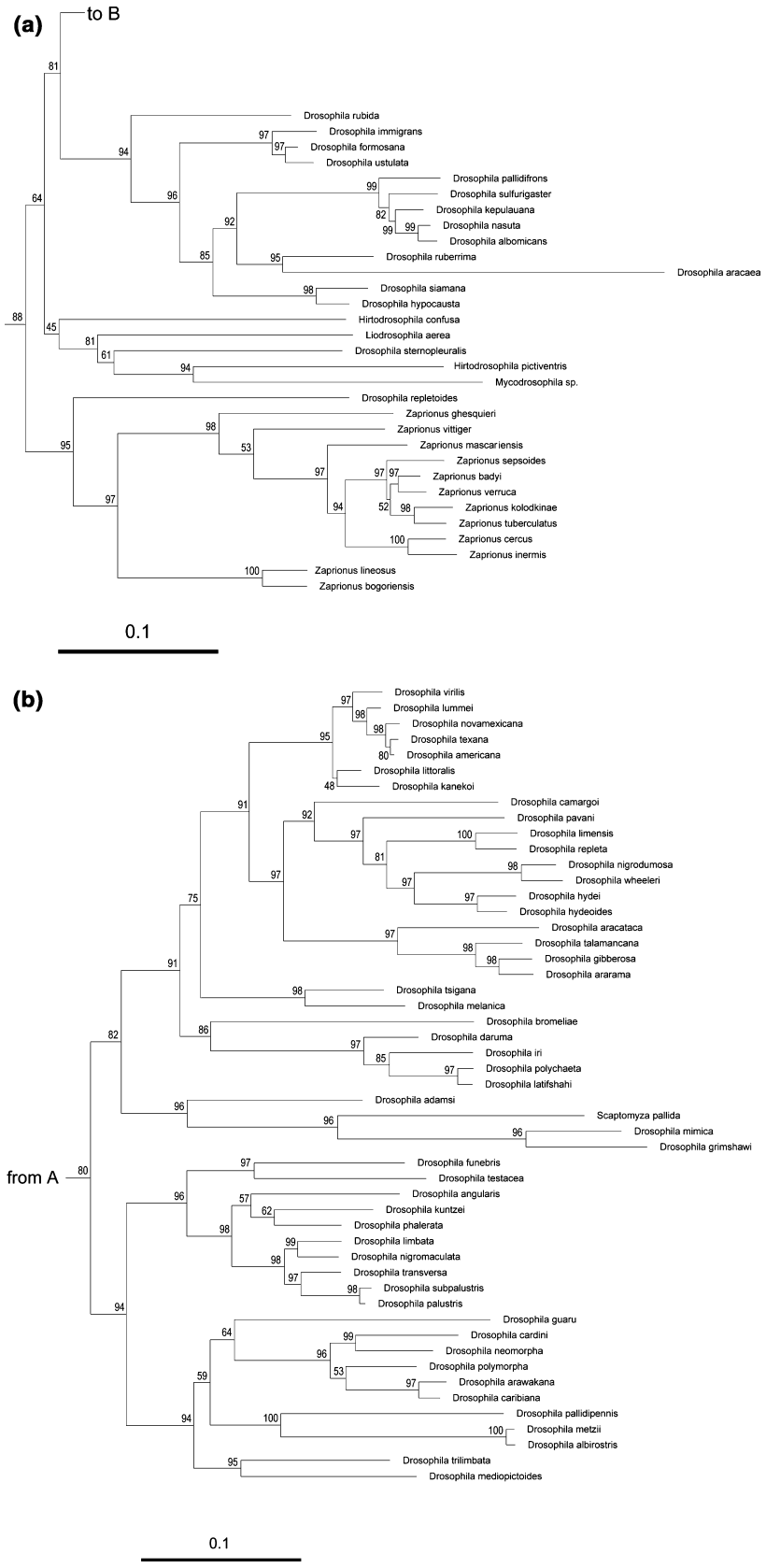



Fig. 7. Excerpt (corresponding to clade 2 in Fig. 3.) of the 50% majority-rule consensus tree from the Bayesian inference analysis (2 000 000 generations). Identical topologies were recovered in four distinct runs of MrBayes. Legend is as in Fig. 6

with high confidence, as already mentioned. In addition, the monophyly of the *ficusphila* subgroup (*D. ficusphila*, *Drosophila levii*) is not supported. *Drosophila flavohirta* clearly falls inside the *melanogaster* subgroup–Oriental subgroup clade, but its exact position is unclear, as is that of *D. levii*. Accordingly, the species or species subgroup closest to the *melanogaster* subgroup is not clearly defined.

### The *montium* subgroup

The *montium* subgroup invariably appears monophyletic. Two main lineages emerge from this major tropical subgroup: an *auraria* lineage and a *kikkawai* lineage. The *auraria* lineage is comprised of *Drosophila asahinai* and *Drosophila rufa*, clustered together and, basal to the (*Drosophila auraria*, *Drosophila biauraria*, *Drosophila triauraria*, *Drosophila quadraria*) cluster, of which *D. auraria* and *triauraria* are always grouped together. The *kikkawai* lineage includes the other species. In all reconstructions, the following species were clearly clustered together: (*bicornuta*, *barbarae*), (*punjabiensis*, *cauveri*, *nagarholensis*), *serrata*, (*kikkawai*, *lini*, *leontia*). Another cluster, less well supported, groups the remaining species of the *montium* subgroup, with *Drosophila jambulina* diverging first, except for weighted MP. Within this cluster, an Afrotropical unit, comprising *bocqueti*, *bocqueti*-like, *burlai*, *chauvacae*, *diplacantha*, displays stable topology. The position of a few species, *Drosophila serrata*, *Drosophila davidi*, *Drosophila vulcana* and *Drosophila greeni*, appears very sensitive to the phylogenetic method used. The new species from Príncipe Island in the Gulf of Guinea, assumed to be close to *Drosophila nikananu* on the basis of its morphology, is confirmed here to be *nikananu*-like.

### The *ananassae* subgroup

The *ananassae* subgroup consists of ca. 22 species and at least 2 subspecies (Tobari 1993; Lemeunier et al. 1997), divided into three species complexes – *ananassae*, *bipectinata* and *ercepeae*, and a few ungrouped species. Seventeen taxa were available in this study. All methods give essentially the same topology, with the three complexes well supported and hierarchized. Interestingly, *Drosophila varians* and *Drosophila lachaisei*, which were once included in the *ananassae* complex (Lemeunier et al. 1986) are in fact found to be the most divergent species. *Drosophila lachaisei* is always basal to the subgroup. The branching of *D. varians* with the *ercepeae* complex is not well supported in MP nw but is highly supported when data are weighted; the BI also shows a high posterior probability value. *Drosophila varians* is proposed to be more basal by both ML (with 86% value) and NJ (64%) but is still ingroup relative to *D. lachaisei*. The *bipectinata* complex consists of closely related species, which may give rise to interspecific hybrids (Bock 1978). *Drosophila pseudoananassae*, which emerges basal in this complex (except by weighted MP), is divided into two subspecies –*D. p. pseudoananassae* and *D. p. nigrens*, which differ by their mitotic karyotype (Lemeunier et al. 1986). In this study, it was found that their *Amyrel* sequences were identical (only one is shown in the trees). Perhaps, shared ancestral polymorphism is obscuring the relationships between *Drosophila malerkotliana*, *Drosophila bipectinata* and *Drosophila parabipectinata* (not shown, but see the position of the subspecies *malerkotliana pallens*). *Drosophila pallidosa* is confirmed by almost all the trees as the sister species of *Drosophila ananassae*, and is in fact the only species to produce viable hybrids with *D. ananassae* (Bock 1984). However, ML suggests that there may be a relationship between *D. pallidosa* and *Drosophila papuensis*-like.

### The subgenus *Drosophila* and related genera

A common root for the subgenus *Drosophila* and distinct genera is found in several trees, with medium-to-high support (97% and BS of 60 in MP 4 : 4 : 1, 88% in BI, 63% in NJ-LD, 58% in MP 2 : 2 : 1). Some taxa differing markedly morphologically, and considered to be distinct genera are always positioned inside the subgenus: *Zaprionus*, *Scaptomyza*, *Hirtodrosophila*, *Liodrosophila*, *Mycodrosophila*, thus making the subgenus *Drosophila* paraphyletic. However, within this large cluster, the relative positions of species groups or other genera are poorly defined, although some species groups are well supported.

### The well-supported clades

The genus *Zaprionus* forms a monophyletic cluster, with the subgenus *Anaprionus* clearly differentiated. *Zaprionus (Anaprionus) lineosus* harbours a 13-bp deletion in the coding sequence, resulting in a frameshift. This has been confirmed by two independent amplifications on two individuals from the same stock. The relationships between this genus and other groups of Drosophilidae are not clearly resolved by all the methods. It is clearly inside clade 2, and is basal to this clade in most MP trees and by the BI method.

The *immigrans* species group is well supported in all methods and is monophyletic. However, species members of the two taxonomic subgroups *hypocausta* and *immigrans* are scattered within the *immigrans* clade. *Drosophila rubida* (a member of the *hypocausta* subgroup) diverges first and appears very distant from the two other species of the *hypocausta* subgroup, *Drosophila siamana* and *Drosophila hypocausta*. Similarly, in the *immigrans* subgroup, *Drosophila ruberrima* is not clustered with the rest of its subgroup, but with the *nasuta* subgroup. The different reconstruction methods are inconsistent with regard to the clustering of other species groups with the *immigrans* group. Both the BI and weighted MP analyses isolated the *immigrans* group in a basal position, whereas the NJ methods and MP–Ts/Tv cluster this group with a well-supported (*tripunctata*, *quinaria*, *funebris*, *testacea*, *guarani*, *cardini*) clade, but with low support. *Drosophila trilimbata*, which belongs to the *immigrans* group, is clustered with the *cardini* group or with *Drosophila mediopictoides* (*tripunctata* group), which is itself clustered with the *cardini* group. The *cardini* group is connected to the *tripunctata* group, to *D. pallidipennis* and to *Drosophila guaru* (*guarani* group). The bootstrap values for this clade range from 54% (MP uw) to 95% and a BS of 21 is recovered (MP 4 : 4 : 1). *Drosophila mediopictoides* is separated from *Drosophila metzii* and *Drosophila albirostris*, suggesting a paraphyletic status of the *tripunctata* group. The *quinaria* group is clearly identified and is clustered with the *testacea* and *funebris* groups. Interestingly, this clade harbours, in each species group, a number of mushroom-feeding species abundant in temperate and boreal forest (Prigent et al. 2003). In MP 4 : 4 : 1, the *quinaria–funebris–testacea* clade is clustered with the *cardini* clade with a low support (48% bootstrap and null BS) but is also found by the BI tree, with a 94% posterior probability.

Another clade, well supported by most methods, contains the following groups: *virilis*, *repleta*, *annulimana*, *mesophragmatica*, *dreyfusi*, *melanica*, *polychaeta* and *bromeliae*. MP 4:4:1, MP 2:2:1, ML, NJ and BI show *polychaeta* branching off first (the results from MP uw and ML have been excluded, according to which the *polychaeta* group clusters with the subgenus *Sophophora*). The *melanica* group diverges from a (*virilis* + *annulimana* + *mesophragmatica* + *repleta* + *dreyfusi*) cluster in BI and NJ. In contrast, MP 4 : 4 : 1, MP 2 : 2 : 1 and MP 10 : 10 : 1 suggest that *melanica* is clustered with *polychaeta*. NJ-K2P indicates a *virilis–annulimana* clade (bootstrap support of 89%) connected to a *dreyfusi–repleta* clade, whereas MP 4 : 4 : 1, MP 2 : 2 : 1, MP 10 : 10 : 1 and BI tended to suggest that *annulimana* diverges from *repleta*, with *camargoi* (*dreyfusi* group) linked to either *repleta* or *annulimana*, and that the *virilis* group branches off from this triad. The position of *Drosophila bromeliae* is not clear: NJ trees show that this species are branching off just above *melanica*, whereas BI suggests a connection to *polychaeta* and MP a connection to *annulimana*. It should be noted that *Drosophila pavani*, a member of the *mesophragmatica* group, is found either inside or outside the *repleta* species group.

### The other groups or genera

There is no clear resolution for positioning the other taxa studied here, some of which being ungrouped by taxonomists (Bächli 1999–2005). *Scaptomyza pallida* is clearly closely related to the Hawaiian *Drosophila mimica* and *Drosophila grimshawi*, but the position of this cluster is uncertain. It may either be close to *Drosophila adamsi* and connected to the *virilis–repleta* clade (under BI) or occupy an external position in the subgenus *Drosophila* (MP, but low support). The genus *Hirtodrosophila* is represented in this study by *H. confusa* and *Hirtodrosophila pictiventris*. These two species are not obviously connected to each other. In fact, *H. pictiventris* is closer to the *Mycodrosophila* sample (unidentified species). *Liodrosophila aerea* is branched with *Mycodrosophila* + *H. pictiventris* in weighted MP and BI trees, with low support in MP, and with *Drosophila sternopleuralis* (*histrio* group) in ML and NJ. Indeed, the support is invariably low. *Drosophila repletoides* (*tumiditarsus* group) is branched with *H. confusa* in MP, with *Zaprionus* in NJ, BI and ML. *Drosophila adamsi* is basal to the 'supergroup' *virilis* in MP, and also in the BI method, but in this case linked to the Hawaiian flies. The other methods also show this species to be linked to the Hawaiian–*Sc. pallida* clade. *Drosophila aracea* is shown to be a member of the *immigrans* group in the BI tree, with a very long branch. This species is similarly located in the MP tree (weighted reconstructions only). Alternatively, this species is found with the *quinaria* group and *Mycodrosophila* in NJ trees. Another ungrouped species, *Drosophila pruinosa*, has not been included in the trees presented here. However, preliminary results (NJ, MP) suggest a clustering with *D. sternopleuralis*.

## Discussion

Using several reconstruction methods, and for most of them, incongruences or weaknesses at some levels have been revealed. However, it appeared that both weighted MP (with downweighting of the third codon positions) and BI analyses were more satisfactory for dealing with base composition bias and gave more robust trees. Overall, it appears that the use of weighting schemes (especially through the downweighting of the third codon position) provides results that match the morphological data for deeper nodes better, whereas unweighted analyses perform better for terminal nodes. The loss of phylogenetic information resulting from the downweighting of the third codon position probably accounts for the latter observation and this highlights the difficulty of defining optimal parsimony weighting schemes. In contrast, the results of the BI analyses appear to be more homogeneous, yielding well-supported results at all levels of the tree. This could probably be attributed to the supposedly better performance of likelihood-based methods when one analyses data sets exhibiting significant base composition heterogeneity between taxa (Galtier and Gouy 1995; Galtier and Gouy 1998).

The general topology of the phylogeny, rooted with *L. maculata*, a Drosophilidae (Steganinae) and a much less closely related species, the Tephritidae *C. capitata*, always shows *Scaptodrosophila* and *Chymomyza*, as distinct genera. Two clades constitute the genus '*Drosophila*', which is paraphyletic according to the current classification. The first clade corresponds to the subgenus *Sophophora*, which appears to be monophyletic; however, species from the *Lordiphosa* genus was not included, which was shown itself to be polyphyletic (Katoh et al. 2000). Indeed, earlier studies have suggested that some *Lordiphosa* species could be related to *Sophophora* as a sister clade to the Neotropical groups *willistoni* and *saltans* (Katoh et al. 2000). The second clade (roughly corresponding to the subgenus *Drosophila*) is clearly paraphyletic, as demonstrated by many investigators since Throckmorton (1975). As a consequence, there is ever-growing evidence, including the present data set, which calls into question both the genus and subgenus *Drosophila*. However, it is merely pointed out in passing the need to revisit the classification of Drosophilidae above the species group level and focus here solely on the species group level.

### Questioning the traditional *melanogaster* species group

On the basis of the data presented in this study, it is proposed to question the traditional boundaries of the *melanogaster* species group, which is a quantitatively important taxonomical unit in the subgenus *Sophophora*. However, this species group is such a mix of contrasted morphological patterns that it has long been seen as a confused taxonomical unit. The historical vicissitudes of the definition of the *melanogaster* group, as new species have accumulated (discussed in Bock and Wheeler 1972, pp 1–8; Lemeunier et al. 1986, pp 1156–157), reflects the difficulties met with in such an extended species group.

The concepts of species groups and subgroups are nearly always used in relation to Drosophilidae taxonomy, and they are not commonly used in other insect families. As a result, they are of only relative practical value. Can these taxonomical levels be recognized independently of evidence of phylogenetic relatedness? How old would have a monophyletic cluster to be, to be called a 'species group'? Should the size of the cluster in the definition be considered? In fact, no rule and no definition exist and this is certainly one of the weaknesses of the Drosophilidae taxonomical classification, and more especially for all the hierarchical levels between species and genus.

The traditional *melanogaster* species group is a questionable taxon, and the question raised here is: Is such a large and composite taxonomical unit needed? The *melanogaster* species group does not suggest any clearly identifiable diagnostic

pattern. Although a definition of the *melanogaster* group has been proposed by Bock and Wheeler (1972), and refined in Lemeunier et al. (1986) by including functional characteristics of the male genitalia, it includes a number of characteristics that do not apply to all species implicated. In contrast, all taxonomists have long accepted a clear immediate and empirical picture of the *montium*-like, *ananassae*-like and *melanogaster* (*sensu stricto*)-like morphological patterns. From the earliest days of *Drosophila* classification (Hsu 1949), the *melanogaster* group was not given any definition but species subgroups, notably the *ananassae*, *melanogaster* and *montium* subgroups imposed themselves as clear-cut homogeneous morphological units. In view of the clear data from the *Amyrel*-based phylogenetic reconstruction obtained here, which is consistent with other data in the literature (Goto and Kimura 2001; Kastanis et al. 2003), it is proposed to elevate both the *ananassae* and *montium* species subgroups to the rank of species group. An updated definition of the neo *ananassae*, *melanogaster* and *montium* species groups will be given elsewhere (Lachaise et al. in preparation). Henceforth, in the discussion, a new terminology is used, in which the *ananassae* and *montium* species subgroups are viewed as species groups, and the *melanogaster* species group is considerably reduced to include only *melanogaster* and the 'Oriental' species subgroups, the status of which remains unchanged.

### The *ananassae* species group, new status

The *ananassae* species group is well sampled in this study. Its location outside the *montium*–(*melanogaster* + Oriental) clade is clear, with a 100% support in all methods. This is consistent with other studies: Goto and Kimura (2001), Kastanis et al. (2003), Lewis et al. (2005). Tamura et al. (2004) have estimated divergence times between *Drosophila* species and suggest a longer divergence time between *D. melanogaster* and *D. ananassae* than between *D. melanogaster* and the *montium* species group. Other studies have suggested various branching topologies, but the supports are generally low in these cases: The *ananassae* group is sister to the *Drosophila obscura* group in Pélandakis and Solignac (1993), sister to *montium* in Schawaroch (2002) and sister to the *melanogaster* subgroup in Yang et al. (2004), with *montium* located further outside. Within the *D. ananassae* group, the three taxonomic complexes are clearly isolated and nested. New data concern the position of the ungrouped species *D. varians* and *D. lachaisei*. The case of the latter species is very interesting in view of the history of this species group, which is considered to be of Oriental origin (Lemeunier et al. 1986). *Drosophila lachaisei* has been found as a rare species (a total of only some two dozen individuals have been caught) from Western, Central and Eastern Africa. The fly used in this work came from São Tomé Island. Interestingly, *D. lachaisei*, which is typically a Palaeoendemic species with a broad fragmented and scattered historical home range, is basal to the entire subgroup, suggesting a very ancient colonization of the Afrotropical region by a member of the *ananassae* group, of which this species is probably a relic.

### The *montium* species group, new status

The *montium* species subgroup has so far been viewed as the most species-rich taxonomical unit within the *melanogaster* species group, and one third of the known species were studied here. The data unequivocally show that it forms a well-defined monophyletic cluster with strong support, thus confirming the recent work of Schawaroch (2002) and Zhang et al. (2003). In view of this consistency plus the marked morphological similarity, and despite its wide geographic distribution covering the Oriental, Afrotropical, Australasian and East-Palaearctic regions, it is suggested that the *montium* subgroup should be raised to species group rank. Hereafter, this will be referred to as the *montium* species group.

The *montium* species group is clearly sister to the *melanogaster* subgroup + Oriental subgroups. This result conflicts with the results reported by Schawaroch (2002), who identified the *montium* lineage as a sister (albeit with weak statistical support) of the *ananassae* lineage, and with the unlikely trichotomy *D. lucipennis* (*suzukii* subgroup)–*montium* subgroup–*elegans* subgroup (Lewis et al. 2005), but it is fully consistent with previous studies (Bock 1980; Ashburner et al. 1984) and with most gene data (Pélandakis and Solignac 1993; Inomata et al. 1997; Goto and Kimura 2001). The mitochondrial sequences were not taken into account, which are generally considered to be less reliable for phylogenetic reconstructions in the genus *Drosophila* (discussed e.g. in Goto and Kimura 2001).

Separating the *montium* group from other species groups is easy, whereas diagnostic identification is often critical within the *montium* species group because of a high level of within-group similarity regarding both male terminalia and sex combs, and to morphological homoplasy (convergence). As a result, the relationships between species and taxonomic complexes are particularly unclear. *Amyrel* provides valuable information by identifying some well-supported groupings consistent with the classical taxonomy and reveals some taxonomic inconsistencies. Data on *Amyrel*, like previous molecular analyses (Schawaroch 2002; Zhang et al. 2003), do not for instance support the Afrotropical *D. greeni* as a member of the *bakoue* complex (Rafael 1984; Lemeunier et al. 1986). In addition, the Afrotropical *Drosophila diplacantha* initially ascribed to the *kikkawai* complex (Tsacas and David 1977) clearly belongs to the *bocqueti* complex (sister of the *nikananu* complex), a position that was previously indicated by *Amy* genes (Zhang et al. 2003), the morphological affinity with the *kikkawai* complex being a homoplasy. Furthermore, the results of this study support the hypothesis that sex comb morphology similarity between the *nikananu* complex and the *melanogaster* subgroup also results from convergence, as proposed by Schawaroch (2002) rather than being indicative of an evolutionary link between the *montium* and *melanogaster* subgroups (Tsacas and Chassagnard 1992).

Compared with previous analyses, a much clearer picture of the evolutionary history of the *montium* group seems to emerge from the present study. The primeval split between a strictly Oriental *auraria–rufa* lineage and an Oriental–Afrotropical lineage (from *D. serrata* to *D. vulcana* in Figs 4 and 6), coupled with the present geographic distribution of the species, support the Oriental origin of the species group. This conclusion is consistent with the two most comprehensive studies in terms of taxon sampling, that is those of Schawaroch (2002) and Zhang et al. (2003). In these studies, based on the nuclear *Amy*, *Adh* and *hb* genes and the mitochondrial *COII*, the clade that includes *auraria–rufa* was basal, although a few Asian species, namely *Drosophila kanapiae* and *Drosophila parvula*, which were not included in the present study, were located outside.

The reconstruction shown in Fig. 6 suggests a second-level split within the *montium* species group separating the primarily

Oriental *kikkawai–nagarholensis–barbarae* lineage and the Afrotropical *greeni–nikananu–bocqueti–bakoue* lineage. The former lineage is actually comprised of Oriental species and species that have wider geographical distribution, such as the Australasian *D. serrata* or the circumtropical *Drosophila kikkawai*, the geographical range of which partially overlaps the Palaearctic region. The second lineage consists exclusively of taxa from the Afrotropical region, and it is therefore postulated that all the Afrotropical species of the *montium* group have a common origin. According to the constructions used (not shown), either *D. greeni* or the pair *Drosophila vulcana–Drosophila malagassya* might be the most basal species, which agrees with the *Amy* genes (Zhang et al. 2003). The position of *D. jambulina*, a species widely distributed from Taiwan to India, invariably clustered with the Afrotropical lineage, is interesting in that it appears to be a possible link to the Oriental species within this highly diversified *Drosophila* species group which has undergone extensive geographic dispersal.

### The *melanogaster* species group, new status

Once the *ananassae* and *montium* clades have been removed from the large traditional *melanogaster* species group and raised to their own species group status, the remaining species subgroups, namely the *melanogaster* and the 'Oriental' species subgroups become the only two components of the novel restricted *melanogaster* species group.

Classically, the *melanogaster* species subgroup is divided into three complexes, nested as [*erecta*(*yakuba*(*melanogaster*))] (Lachaise et al. 1988). The data do not solve the problem of the trifurcation of the triad *D. simulans*, *Drosophila mauritiana* and *Drosophila sechellia*, inside the *melanogaster* complex, which reflects speciation events that took place within a short period (Kliman et al. 2000). The support values are low, and the topologies vary depending on the reconstruction method. The main point deserving discussion here is the clustering of the *yakuba* species complex with the *erecta–orena* clade, rather than with the *melanogaster* species complex. This is proposed by most methods applied to *Amyrel*. Although this conflicts with the classical view that the *erecta–orena* clade diverged first [Cariou 1987; Lachaise et al. 2004; but see the two possibilities in Lachaise et al. 1988, and note that when *Drosophila orena* was used as an outgroup in both *Amyrel* and *Amy* phylogenetic trees (Cariou et al. 2001), the topology reflected the classical view], it is consistent with molecular phylogenies obtained using other markers. While the *Adh*-based phylogenies of Russo et al. (1995), Katoh et al. (2000) (minimum evolution method for the latter) and Lachaise et al. (2000) (*period* and allozymes) agree with the classical view with good support, the trees proposed by Kastanis et al. (2003) on mtDNA, Ko et al. (2003) on *Adh* + *Adhr* + *Gld* + *Xdh*, Tatarenkov et al. (1999) on *Ddc*, Katoh et al. (2000) on *Adh* (with MP method), Lewis et al. (2005) on COI + COII show a clade (*erecta–orena*) + (*yakuba–santomea–teissieri*) with good support values. In a recent review (Lachaise and Silvain 2004), it was argued that this latter five-species clade includes three species, either strictly endemic (*D. santomea* and *D. orena*) or partly endemic (*D. erecta*) confined to the Cameroon volcanic line. Although the two other allied species (*Drosophila teissieri* and *Drosophila yakuba*) have extended their distribution to Eastern and South-eastern Africa since their speciation, they were also thought to have originated at different times from that mountain region, the uplift of which dates back some 13–15 Myr ago. A clear-cut split between a monophyletic Western clade (*erecta–orena*) + (*yakuba–santomea–teissieri*) and a monophyletic Eastern clade (*melanogaster*) + (*simulans*, *sechellia*, *mauritiana*) would satisfactorily match the palaeobiogeographic pattern proposed in the aforementioned paper. In particular, this would provide further support for the nascent hypothesis of a major West–East separation of the African mainland that might predate the formation of the Rift.

Turning to the ever-puzzling question of the affinities of the Oriental species subgroups, the findings of this study do not clarify which of the Oriental subgroups is the closest to the *melanogaster* subgroup. The relationships between the Oriental subgroups are not well solved, except that *Drosophila elegans* does indeed seem to have diverged first, as already proposed by Goto and Kimura (2001). As shown by earlier studies (Pélandakis and Solignac 1993; Goto and Kimura 2001; Kopp and True 2002; Schawaroch 2002; Lewis et al. 2005), the *suzukii* subgroup is polyphyletic. Some species are related to the *takahashii* subgroup, and *D. lucipennis* is close to *D. elegans*. *Drosophila flavohirta*, which was previously ungrouped within the traditional *D. melanogaster* group, is clearly included in the Oriental cluster, although not clearly related to any other species. Several methods propose *D. flavohirta* as the closest relative of the *melanogaster* subgroup, although with low support.

### Difficulties with the Neotropical groups and biased composition

The conflictual position of the Neotropical groups in the reconstructions has been highlighted earlier, with their widely accepted membership to the subgenus *Sophophora* (O'Grady et al. 1998; O'Grady and Kidwell 2002). The particularly low GC content of *Amyrel* in these species has been pointed out. This unusual compositional bias has been reported for several genes in species of the *willistoni* and *saltans* groups. It has been shown that it is a derived state and that the ancestral base composition in *Drosophila* had a high GC content (Rodriguez-Trelles et al. 2000). It has been suggested that this conflicting base composition was mainly responsible for the misplacing of the Neotropical groups in the phylogeny. Indeed, the weighted MP methods, which lower the influence of GC3%, give a clustering at the base of the subgenus *Sophophora*. This problem has been addressed by Tarrio et al. (2001), using sophisticated models (Galtier and Gouy 1998) accounting for the non-uniformity and non-stationarity of substitutions. The results clearly showed the branching of the Neotropical groups basal to *Sophophora*, with high bootstrap values. This has also been found by other workers (Russo et al. 1995; Kwiatowski and Ayala 1999; Tatarenkov et al. 1999), albeit with lower support. However, Tatarenkov et al. (1999) obtained good support when using only codon positions 1 + 2. Katoh et al. (2000) placed these groups with *Sophophora* (with low support) using parsimony methods, but outside the *Drosophila* genus using minimum evolution method.

### The deep nodes in the subgenus Drosophila are not well resolved

Interestingly, for the external branches of the tree, Tarrio et al. (2001) had found that using *C. capitata* as the outgroup, the first diverging taxon was *Scaptodrosophila*, followed by *Chymomyza*, as was found with MP (Fig. 3a). These authors have pointed out that *C. capitata* itself has a low GC content

(Fig. 2), which could be detrimental to the reconstruction if not accounted for. Their data provide strong support for the possibility that the genera *Zaprionus* and *Hirtodrosophila* should be included in a clade involving the subgenus *Drosophila*, as was found (96% bootstrap and BS of 87 in MP 4 : 4 : 1). The position of *Zaprionus* and of taxa considered to belong to distinct genera (*Scaptomyza*, *Hirtodrosophila*) within the 'subgenus' *Drosophila* has been consistently proposed earlier, among many others, notably by Throckmorton (1975), Pélandakis and Solignac (1993), Tatarenkov et al. (1999), Kwiatowski and Ayala (1999), and Katoh et al. (2000) and should now be accepted. As a consequence, the current classification in the subgenus *Drosophila* appears unquestionably paraphyletic suggesting that the generic and subgeneric classification must be revisited. The deep nodes within the current subgenus *Drosophila* are unfortunately not clearly resolved by the data, and this remains a shortcoming. Indeed, Tatarenkov et al. (1999), Katoh et al. (2000) and Remsen and O'Grady (2002), who investigated various species groups and closely related genera, found generally low support for the deep nodes. Kwiatowski and Ayala (1999), working with a limited number of species and using combined data from *Adh* and *Sod*, found high support for a relationship between the Hawaiian-*Scaptomyza* and the *virilis–repleta* clade, on one hand, and between *Hirtodrosophila* and *Drosophila immigrans*, on the other hand. The branching of Hawaiian species with the *virilis–repleta* clade was also found by Amador and Juan (1999) using *Adh*. These conclusions could not be clearly confirmed. Data using only the same (or similar) species as those in Kwiatowski and Ayala (1999) in an MP 4 : 4 : 1 assay (100 bootstraps of 100 heuristic replicates) and NJ-K2P or LD have been reanalysed. The relationships reported by these authors (not shown) were yet to be found. It is however worth pointing out that the Neotropical species groups were found strongly branched at the base of the subgenus *Sophophora* (bootstrap value of 90% with MP).

The topology of the *virilis* group is consistent with the phylogeny proposed by Spicer and Bell (2002). A relevant sampling for several of the species groups included in the *virilis–repleta* 'supergroup' [remembering the *virilis–repleta* radiation of Throckmorton (1975)] was had: six species for the *repleta* group, seven species for the *virilis* group, four species for the *polychaeta* group, four species for the *annulimana* group, two species for the *melanica* group, which may improve the reliability of the phylogeny (Zwickl and Hillis 2002). Several topologies are proposed by the different methods, but all of them place *polychaeta* as a basal group, followed by *melanica*. The hierarchy proposed by the BI tree in this study is the same as that shown by Robe et al. (2005), for the *Amd* gene. The MP 4 : 4 : 1 tree shows a topology similar to that proposed by Tatarenkov and Ayala (2001) from combined *Amd* and *Ddc* gene sequences. These authors also found a good support for placing *D. bromeliae* ingroup relative to the *annulimana* group, whereas none of the trees confidently shows branching of *D. bromeliae*. As discussed above, the Hawaiian species have been proposed to be the closest sister group to the *virilis–repleta* clade (e.g. Amador and Juan 1999; Tatarenkov and Ayala 2001), but no such significant branching was found in this study, with the noticeable exception of the BI tree. Only two Hawaiian species were included here, while the Hawaiian radiation gave hundreds of species. More species could help clarifying the real topology. More detailed studies within the Hawaiian species have been conducted (Baker and DeSalle 1997; DeSalle and Brower 1997; Remsen and DeSalle 1998; Bonacum et al. 2005).

The *immigrans* group is clearly monophyletic, although some discrepancies regarding the classical taxonomy of two of the subgroups sampled here were found. It has been suggested that the *immigrans* group may be a member of a *immigrans–tripunctata* radiation, well separated from the *virilis–repleta* radiation, with the *tripunctata* group linked to *quinaria*, *funebris* and *cardini* (see Bächli 1999–2005). This view had been supported by previous studies (Pélandakis and Solignac 1993; Katoh et al. 2000; Remsen and O'Grady 2002); however, none of the trees supports this cluster with confidence. Although some methods, such as NJ, do suggest this classical view, the level of support is very low. The best trees show the *immigrans* group branching off first and the *tripunctata* clade tends to be linked to the *virilis–repleta* clade. The other groups mentioned are clearly clustered together by most methods. However, the hierarchy within this clade is not exactly the same as that found by other workers (Robe et al. 2005). Interestingly, we have also been found that the *tripunctata* group is probably paraphyletic, as has been shown by other authors (e.g. Robe et al. 2005).

The phylogenetic trees presented in this study are gene trees, constructed using a single gene, *Amyrel*. Interestingly, many results in this study are consistent with previous findings and phylogenetic analyses. In addition, the recent work by Kopp (2006) shows that *Amyrel* gives a useful phylogenetic information. This gives confidence regarding the analyses of this study reflecting the phylogeny of Drosophilidae rather than the evolution of a specific gene merely. It is likely that combined analyses from several additional genes would lead to more reliable phylogenies, but available sequences of other genes in the literature or databases correspond to only a limited fraction of the sampling in this study. The problem of using more taxa or more characters for improving phylogenies has been debated under several aspects (e.g. Zwickl and Hillis 2002; Hillis et al. 2003; Gontcharov et al. 2004; May-Collado and Agnarsson 2006). Delsuc et al. (2005) stated that 'A long-standing debate in phylogenetics is whether the greatest improvement in accuracy results from an increased number of characters (in this case, genes) or species. Evidence from computer simulations has been equivocal, whereas empirical studies tend to support the importance of extensive species sampling.' The phylogeny presented here may be therefore improved by additional taxa as well as addition of other genes. Finally, *Amyrel*, which allows easy amplification of the full coding sequence all at once, offers a powerful tool to investigate *Drosophila* phylogeny. It is hoped that *Amyrel* will be used by researchers attempting to clarify the phylogenetic relationships of their Drosophila species of interest, thus enlarging the trees proposed here.

## Acknowledgements

We thank H. Hachani and S. Prébot for technical help, E. Renard for providing sequence data on the *obscura* species group and Kim van der Linde for fruitful discussion. The Gif laboratory collection is part of the Centre de ressources Drosophiles, granted from the CNRS and the Ministère de la Recherche (n° 03G138).

## Résumé

*Une phylogénie des Drosophilidae avec le gène Amyrel: remise en question des limites du groupe d'espèces Drosophila melanogaster*

Cette étude analyse les relations phylogénétiques chez les Drosophilidae. Nous avons utilisé le gène *Amyrel*, un paralogue éloigné de la famille multigénique des alpha-amylases, sur 164 espèces, en essayant d'avoir un bon échantillon sur de nombreux groupes d'espèces de *Drosophila* et quelques genres voisins. Les séquences nucléiques ont été analysées en maximum de parcimonie, neighbor joining, et avec des méthodes probabilistes. Nous avons tenu compte de l'hétérogénéité de composition (surtout la faible teneur en GC dans les groupes *willistoni* et *saltans*). Nos résultats montrent que le genre *Drosophila* est paraphylétique; et si le sous-genre *Sophophora* apparaît monophylétique, le sous-genre *Drosophila* est paraphylétique, certains genres distincts y étant inclus. Nous proposons d'élever au rang de groupes d'espèces les sous-groupes *ananassae* et *montium*, et de limiter le groupe *melanogaster* au sous-groupe *melanogaster* et aux sous-groupes « orientaux ».

## References


Amador A, Juan E (1999) Nonfixed duplication containing the *Adh* gene and a truncated form of the *Adhr* gene in the *Drosophila funebris* species group: different modes of evolution of *Adh* relative to *Adhr* in *Drosophila*. Mol Biol Evol **16:**1439–1456.

Ashburner M, Bodmer M, Lemeunier F (1984) On the evolutionary relationships of *Drosophila melanogaster*. Dev Genet **4:**295–312.

Ashburner M, Golic KG, Hawley RS (2005) Drosophila: A Laboratory Handbook, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bächli G (1999–2005) The database on taxonomy of Drosophilidae. Available at: http://taxodros.unizh.ch/.

Baker RH, DeSalle R (1997) Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst Biol **46:**654–673.

Barker K, Lanyon SM (2000) The impact of parsimony weighting schemes on inferred relationships among toucans and neotropical barbets (Aves: Piciformes). Mol Phylogenet Evol **15:**215–234.

Bock IR (1978) The *bipectinata* complex: a study in interspecific hybridization in the genus *Drosophila* (Insecta: Diptera). Aust J Biol Sci **31:**197–208.

Bock IR (1980) Current status of the *Drosophila melanogaster* species group (Diptera). Syst Entomol **5:**341–356.

Bock IR (1984) Interspecific hybridization in the genus *Drosophila*. Evol Biol **18:**41–67.

Bock IR, Wheeler MR (1972) The *Drosophila melanogaster* species group. Univ Tex Publ **7213:**1–102.

Bonacum J, O'Grady PM, Kambysellis M, DeSalle R (2005) Phylogeny and age of diversification of the *planitibia* species group of the Hawaiian Drosophila. Mol Phylogenet Evol **37:**73–82.

Bremer K (1994) Branch support and tree stability. Cladistics **10:**295–304.

Brown NM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol **18:**839–843.

Cariou M-L (1987) Biochemical phylogeny of the eight species in the *Drosophila melanogaster* subgroup, including *D. sechellia* and *D. orena*. Gen Res Camb **50:**181–185.

Cariou M-L, Silvain J-F, Daubin V, Da Lage J-L, Lachaise D (2001) Divergence between *Drosophila santomea* and allopatric or sympatric populations of *D. yakuba* using paralogous amylase genes and migration scenarios along the volcanic line. Mol Ecol **10:**649–660.

Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland.

Da Lage J-L, Wegnez M, Cariou M-L (1996) Distribution and evolution of introns in Drosophila amylase genes. J Mol Evol **43:**334–347.

Da Lage J-L, Renard E, Chartois F, Lemeunier F, Cariou M-L (1998) *Amyrel*, a paralogous gene of the amylase gene family in *Drosophila melanogaster* and the *Sophophora* subgenus. Proc Natl Acad Sci USA **95:**6848–6853.

Da Lage J-L, Van Wormhoudt A, Cariou M-L (2002) Diversity and evolution of the alpha-amylase genes in animals. Biologia (Bratisl) **57:**181–189.

Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. Nature Rev Genet **6:**361–375.

DeSalle R (1992) The phylogenetic relationships of flies in the family Drosophilidae deduced from mtDNA sequences. Mol Phylogenet Evol **1:**31–40.

DeSalle R, Brower AV (1997) Process partitions, congruence, and the independence of characters: inferring relationships among closely related Hawaiian drosophila from multiple gene regions. Syst Biol **46:**751–764.

DeSalle R, Grimaldi DA (1993) Phylogenetic pattern and developmental process in drosophila. Syst Biol **42:**458–475.

Durando CM, Baker RH, Etges WJ, Heed WB, Wesserman M, DeSalle R (2000) Phylogenetic analysis of the *repleta* species group of the genus *Drosophila* using multiple sources of characters. Mol Phylogenet Evol **16:**296–307.

Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base composition. Proc Natl Acad Sci USA **92:**11317–11321.

Galtier N, Gouy M (1998) Inferring pattern and process: maximum-likelihood implementation of a nonhomogeneous model of DNA sequence evolution for phylogenetic analysis. Mol Biol Evol **15:**871–879.

Gleason JM, Powell JR (1997) Interspecific and intraspecific comparisons of the period locus in the *Drosophila willistoni* sibling species. Mol Biol Evol **14:**741–753.

Gloor G, Engels W (1992) Single fly DNA preps for PCR. Drosoph Inf Serv **71:**148–149.

Gompell N, Carroll SB (2003) Genetic mechanisms and constraints governing the evolution of correlated traits in drosophilid flies. Nature **424:**931–935.

Gontcharov AA, Marin B, Melkonian M (2004) Are combined analyses better than single gene phylogenies? A case study using SSU rDNA and *rbc*L sequence comparisons in the Zygnematophyceae (Streptophyta). Mol Biol Evol **21:**612–624.

Goto SG, Kimura MT (2001) Phylogenetic utility of mitochondrial *COI* and nuclear *Gpdh* genes in Drosophila. Mol Phylogenet Evol **18:**404–422.

Grimaldi DA (1990) A phylogenetic, revised classification of the genera in the Drosophilidae (Diptera). Bull Am Mus Nat Hist **197:**1–139.

Gu X, Fu Y-X, Li W-H (1995) Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol Biol Evol **12:**546–557.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol **52:**696–704.

Hillis DM, Huelsenbeck JP, Cunningham CW (1994) Application and accuracy of molecular phylogenies. Science **264:**671–677.

Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? Syst Biol **52:**124–126.

Hsu TC (1949) The external genital apparatus of male Drosophilidae in relation to systematics. Univ Tex Publ **4920:**80–142.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. Bioinformatics **17:**754–755.

Inomata N, Tachida H, Yamazaki T (1997) Molecular evolution of the *Amy* multigenes in the subgenus Sophophora of Drosophila. Mol Biol Evol **14:**942–950.

Kastanis P, Eliopoulos E, Goulielmos GN, Tsakas S, Loukas M (2003) Macroevolutionary relationships of species of *Drosophila melanogaster* group based on mtDNA sequences. Mol Phylogenet Evol **28:**518–528.

Katoh T, Tamura K, Aotsuka T (2000) Phylogenetic position of the subgenus *Lordiphosa* of the genus *Drosophila* (Diptera: Drosophilidae) inferred from alcohol dehydrogenase (*Adh*) gene sequences. J Mol Evol **51:**122–130.

Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J (2000) The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics **156:**1913–1931.

Ko W, David R, Akashi H (2003) Molecular phylogeny of the *Drosophila melanogaster* subgroup. J Mol Evol **57:**562–573.

Kopp A (2006) Basal relationships in the *Drosophila melanogaster* species group. Mol Phylogenet Evol **39:**787–798.

Kopp A, True JR (2002) Phylogeny of the oriental *Drosophila melanogaster* species group: a multilocus reconstruction. Syst Biol **51:**786–805.

Kwiatowski J, Ayala FJ (1999) Phylogeny of drosophila and related genera: conflict between molecular and anatomical analyses. Mol Phylogenet Evol **13:**319–328.
Lachaise D, Silvain J-F (2004) How two Afrotropical endemics made two cosmopolitan human commensals from Africa: the *Drosophila melanogaster-D. simulans* riddle. Genetica **120:**17–39.
Lachaise D, Cariou M-L, David JR, Lemeunier F, Tsacas L, Ashburner M (1988) Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol Biol **22:**159–225.
Lachaise D, Harry M, Solignac M, Lemeunier F, Bénassi V, Cariou M-L (2000) Evolutionary novelties in islands: *Drosophila santomea*, a new *melanogaster* sister species from São Tomé. Proc Roy Soc London **267:**1487–1495.
Lachaise D, Capy P, Cariou M-L, Joly D, Lemeunier F, David JR (2004) Nine relatives from one African ancestor: the population biology of the *Drosophila melanogaster* subgroup species. In: Singh RS, Uyenoyama MK (eds), The Evolution of Population Biology. Cambridge University Press, Cambridge, pp. 315–343.
Lanave C, Preparata GCS, Serio G (1984) A new method for calculating evolutionary substitution rates. J Mol Evol **20:**86–93.
Lemeunier F, Tsacas L, David J, Ashburner M (1986) The *melanogaster* species group. In: Thomson JR, Carson HL (eds), The Genetics and Biology of Drosophila, Vol. 3e. Academic Press, New York, pp. 147–256.
Lemeunier F, Aulard S, Arienti M, Jallon J-M, Cariou M-L, Tsacas L (1997) The *ercepeae* complex: new cases of insular speciation within the *Drosophila ananassae* species subgroup (*melanogaster* group) and descriptions of two new species (Diptera: Drosophilidae). Ann Entomol Soc Am **90:**28–42.
Lewis RL, Beckenbach AT, Mooers AO (2005) The phylogeny of the subgroups within the *melanogaster* species group: likelihood tests on COI and COII sequences and a Bayesian estimate of phylogeny. Mol Phylogenet Evol **37:**15–24.
Maczkowiak F, Da Lage J-L (2006) Origin and evolution of the *Amyrel* gene in the α-amylase family of Diptera. Genetica **128:** 145–158.
May-Collado L, Agnarsson I (2006) Cytochrome b and Bayesian inference of whale phylogeny. Mol Phylogenet Evol **38:** 344–354.
O'Grady PM, Kidwell MG (2002) Phylogeny of the subgenus *Sophophora* (Diptera: Drosophilidae) based on combined analysis of nuclear and mitochondrial sequences. Mol Phylogenet Evol **22:**442–453.
O'Grady PM, Clark JB, Kidwell MG (1998) Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. Mol Biol Evol **15:**656–664.
Pélandakis M, Solignac M (1993) Molecular phylogeny of Drosophila based on ribosomal RNA sequences. J Mol Evol **37:**525–543.
Pélandakis M, Higgins DG, Solignac M (1991) Molecular phylogeny of the subgenus Sophophora of Drosophila derived from large subunit of ribosomal RNA sequences. Genetica **84:**87–94.
Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics **14:**817–818.
Powell JR (1997) Progress and Prospects in Evolutionary Biology. The Drosophila Model. Oxford University Press, New York.
Prigent S, Renard E, Cariou M-L (2003) Electrophoretic mobility of amylase in Drosophilids indicates adaptation to ecological diversity. Genetica **119:**133–145.
Rafael V (1984) Relations interspécifiques dans le nouveau complexe africain de *Drosophila bakoue* du groupe *melanogaster*, sous-groupe *montium* (Diptera, Drosophilidae). Bull Soc Zool Fr **109:**179–189.
Remsen J, DeSalle R (1998) Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. Mol Phylogenet Evol **9:**225–235.
Remsen J, O'Grady P (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. Mol Phylogenet Evol **24:**249–264.
Robe LJ, Valente VLS, Budnik M, Loreto ELS (2005) Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical species and groups: a nuclear versus mitochondrial approach. Mol Phylogenet Evol **36:**623–640.
Rodriguez-Trelles F, Tarrio R, Ayala FJ (2000) Evidence for a high ancestral GC content in Drosophila. Mol Biol Evol **17:**1710–1717.
Russo CAM, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of Drosophilid species. Mol Biol Evol **12:**391–404.
Schawaroch V (2002) Phylogeny of a paradigm lineage: the *Drosophila melanogaster* species group (Diptera: Drosophilidae). Biol J Linn Soc **76:**21–37.
Sorenson MD (1999) TreeRot, version 2. Boston University, Boston.
Spicer GS, Bell CD (2002) Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. Ann Entomol Soc Am **95:**156–161.
Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony* and Other Methods. Version 4.0b10. Sinauer Associates, Sunderland, MA.
Tamura K, Subramanian S, Kumar S (2004) Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. Mol Biol Evol **21:**36–44.
Tarrio R, Rodriguez-Trelles F, Ayala FJ (2001) Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. Mol Biol Evol **18:**1464–1473.
Tatarenkov A, Ayala FJ (2001) Phylogenetic relationships among species groups of the *virilis-repleta* radiation of Drosophila. Mol Phylogenet Evol **21:**327–331.
Tatarenkov A, Kwiatowski J, Skarecky D, Barrio E, Ayala FJ (1999) On the evolution of *Dopa decarboxylase* (*Ddc*) and Drosophila systematics. J Mol Evol **48:**445–462.
Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673–4680.
Throckmorton LH (1975) The phylogeny, ecology and geography of Drosophila. In: King RC (ed), Handbook of Genetics. Plenum Press, New York, pp. 421–469.
Tobari YN (1993) *Drosophila ananassae*, Genetical and Biological Aspects. Japan Scientific Societies Press, Tokyo, 289 p.
Toda MJ (1991) Drosophilidae (Diptera) in Myanmar (Burma) VII. The *Drosophila melanogaster* species-group, excepting the *D. montium* species-subgroup. Orient Insects **25:**69–94.
Tsacas L, Chassagnard M-T (1992) L'identité de *Drosophila* (*Sophophora*) *vulcana* Graber (Diptera, Drosophilidae). Bull Soc Entomol Fr **96:**427–432.
Tsacas L, David J (1977) Systematics and biogeography of the *Drosophila kikkawai* complex, with description of new species (Diptera, Drosophilidae). Ann Soc Entomol Fr (N.S.) **13:**675–693.
Wu C-I (2001) The genic view of the process of speciation. J Evol Biol **14:**851–865.
Yang Z (1994) Estimating the pattern of nucleotide substitution. J Mol Evol **39:**105–111.
Yang Y, Zhang Y-P, Qian Y-H, Zeng Q-T (2004) Phylogenetic relationships of *Drosophila melanogaster* species group deduced from spacer regions of *histone* gene H2A-H2B. Mol Phylogenet Evol **30:**336–343.
Zhang Z, Inomata N, Cariou M-L, Da Lage J-L, Yamazaki T (2003) Phylogeny and evolution of the amylase multigenes in *Drosophila montium* species subgroup. J Mol Evol **56:**121–130.
Zwickl DJ, Hillis DM (2002) Increased taxon sampling greatly reduces phylogenetic error. Syst Biol **51:**588–598.

*Authors' addresses*: Jean-Luc Da Lage (for correspondence), Marie-Louise Cariou and Daniel Lachaise (deceased), UPR 9034 Evolution, génomes et spéciation, CNRS, 91198 Gif sur Yvette cedex, France. E-mail: jldl@legs.cnrs-gif.fr, marie-louise.cariou@legs.cnrs-gif.fr; Gaël Joseph Kergoat, INRA, Centre de Biologie et de Gestion des Populations, Campus International de Baillarguet, CS 30016, 34988 Montferrier-sur-Lez, France. E-mail: kergoat@ensam.inra.fr; Frédérique Maczkowiak, Institut Curie, 91400 Orsay, France. E-mail: maczkowiak@curie.u-psud.fr; Jean-François Silvain, IRD, UR R072, c/o CNRS UPR 9034, 91198 Gif sur Yvette cedex, France. E-mail: jean-francois.silvain@legs.cnrs-gif.fr