

## Molecular Population Genetics of the *rp49* Gene Region in Different Chromosomal Inversions of *Drosophila subobscura*

Julio Rozas,\* Carmen Segarra,\* Griselda Ribó<sup>†</sup> and Montserrat Aguadé\*

\*Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, 08071 Barcelona, Spain and <sup>†</sup>Centre d'Investigació i Desenvolupament, Centro Superior de Investigaciones Científicas, 08036 Barcelona, Spain

Manuscript received April 13, 1998

Accepted for publication September 21, 1998

### ABSTRACT

Nucleotide variation at the ribosomal protein 49 (*rp49*) gene region has been studied in 75 lines of *Drosophila subobscura* belonging to four chromosomal arrangements ( $O_{st}$ ,  $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$ ). The location of the *rp49* gene region within the inversion loop differs among heterokaryotypes: it is very close to one of the breakpoints in heterozygotes involving  $O_{st}$  chromosomes, while it is in a more central position in all other heterokaryotypes. The distribution of nucleotide polymorphism in the different arrangements is consistent with a monophyletic origin of the inversions. The data also provide evidence that gene conversion and possibly double crossover are involved in shuffling nucleotide variation among gene arrangements. The analyses reveal that the level of genetic exchange is higher when the region is located in a more central position of the inverted fragment than when it is close to the breakpoints. The pairwise difference distributions as well as the negative values of Tajima's and Fu and Li's statistics further support the hypothesis that nucleotide variation within chromosomal arrangements still reflects expansion after the origin of the inversions. Under the expansion model, we have estimated the time of origin of the studied inversions.

**I**NVERSION polymorphism is widespread in the genus *Drosophila*. In *Drosophila subobscura*, for instance, >80 different chromosomal gene arrangements have been described (reviewed in Krimbas and Loukas 1980; Krimbas 1992; Powell 1997). In this species, inversion polymorphism is distributed among its five acrocentric chromosomes, in contrast to other *Drosophila* species (like *D. pseudoobscura*). In *D. subobscura* there is also ample evidence for the adaptive value of chromosomal inversion polymorphism. Studies of the distribution of chromosomal polymorphism in different localities have shown that the frequencies of some chromosomal arrangements exhibit clines correlated with latitude (Krimbas 1992). Although parallel gradual variation does not imply a causal relationship, and therefore is not unambiguous evidence of selection, in *D. subobscura* the correlation between latitude and the frequency of certain chromosomal arrangements was found both in the Old World and in the recently colonized areas of North and South America. The presence of parallel clines in different geographic areas clearly supports the adaptive value of the inversion polymorphism (Prevosti *et al.* 1988).

It is generally accepted that the adaptive character of inversion polymorphism is due to the gene content of the different gene arrangements (see Powell 1997).

Each gene arrangement could contain a particular combination of genes that, because of the suppression of recombination in heterokaryotypes, would be maintained as a block. Little is known, however, about the relative importance of recombination (*i.e.*, gene flow) and mutation in shaping nucleotide variation of different gene arrangements. If recombination were totally suppressed in inversion heterozygotes, DNA regions included in a particular inversion would evolve independently in inverted and noninverted chromosomes; on the other hand, recombination as a homogenizing factor can prevent their independent evolution.

To date, there are few empirical studies relating nucleotide and chromosomal variation, and most of them have been performed in three species of *Drosophila*: *D. melanogaster* (Aguadé 1988a; Bénassi *et al.* 1993; Wesley and Eanes 1994; Hasson and Eanes 1996), *D. pseudoobscura* (Aquadro *et al.* 1991; Popadić and Anderson 1994, 1995; Popadić *et al.* 1995; Babcock and Anderson 1996), and *D. subobscura* (Rozas and Aguadé 1990, 1993, 1994; Rozas *et al.* 1995). However, these studies have provided useful information about the evolution of inversions. For example, at the *rp49* (ribosomal protein 49) gene region of *D. subobscura*, Rozas and Aguadé (1990, 1994) detected genetic exchange between different chromosomal gene arrangements, and they proposed that this transfer was accomplished by gene conversion.

Here, we have studied nucleotide variation at the *rp49* gene region of *D. subobscura* in the four major chromosomal arrangements of the *O* chromosome with

Corresponding author: Julio Rozas, Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Diagonal 645, 08071 Barcelona, Spain. E-mail: julio@porthos.bio.ub.es

inversions that include the *rp49* gene ( $O_{st}$ ,  $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$ ). As depicted in Figure 1, arrangements  $O_{st}$  and  $O_{3+4}$  arose independently from the  $O_3$  arrangement (Ramos-Onsins *et al.* 1998). Arrangement  $O_3$  is not present in extant populations of *D. subobscura*, but it is present in the closely related species *D. madeirensis* and *D. guanche*. On the other hand,  $O_{3+4+8}$  and  $O_{3+4+23}$  originated from the  $O_{3+4}$  arrangement, and in contrast these three arrangements coexist in some populations. In the arrangements studied, the location of the gene within the inversion loop differs in the different heterokaryotypes: between  $O_{st}$  and the other arrangements the *rp49* gene is located near one of the breakpoints, while in heterokaryotypes between  $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$ , the gene is in a more central position in the inversion loop. The different location of the gene in the heterokaryotypes is expected to affect the level of genetic exchange between arrangements because near breakpoints it can be due exclusively to gene conversion; however, in more internal positions of the inversion loop both gene conversion and double crossover may contribute to the genetic exchange. The present study therefore tries to investigate the level of gene flow between arrangements. Moreover, the study aims to assess if nucleotide variation in the different gene arrangements is at stationary equilibrium and to estimate the age of the inversions.

#### MATERIALS AND METHODS

**Fly samples:** Seventy-five *D. subobscura* lines from El Pedroso (Spain) and from Bizerte (Tunisia) were used. Strains (and DNA sequences) from El Pedroso have been reported previously (Rozas and Aguadé 1994). The 41 strains newly reported were collected at Bizerte (Tunisia) in May 1996. The *Va/Ba* balancer stock (Sperlich *et al.* 1977) was used to obtain isochromosomal lines. For each line the chromosomal gene arrangement was determined as previously described in Rozas and Aguadé (1990). The location of the *rp49* gene on the *O* chromosome defines four major chromosomal classes:  $O_{st}$ ,  $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$ . As shown in Figure 1, these chromosomal classes differ by at least one inversion that includes the *rp49* gene, and therefore recombination by a single crossover

would be prevented in the different heterokaryotypes. Chromosomes with both inversions 3 and 4 (*i.e.*,  $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$  chromosomes) are generically cited as  $O_{3+4+x}$ . Although the sequenced lines are not a random sample of the population, for a given chromosomal class they were randomly chosen (Table 1).

**DNA sequencing:** The *rp49* nucleotide sequences for the 34 lines of El Pedroso have been previously reported (Rozas and Aguadé 1994; EMBL/GenBank accession numbers X80076–X80109). Genomic DNA from the Tunisian isochromosomal lines was extracted using a modification of protocol 48 from Ashburner (1989). An  $\sim 1.6$ -kb region including the *rp49* coding region (402 bp) and its 5' and 3' flanking regions was amplified by PCR (Saiki *et al.* 1988). Several oligonucleotides, designed at intervals of  $\sim 300$  nucleotides, were used as primers for sequencing. The amplified fragments were cycle-sequenced and separated on a Perkin-Elmer (Norwalk, CT) ABI PRISM 377 automated DNA sequencer following the manufacturer's instructions. For each line, the sequence of both strands was determined. The newly reported nucleotide sequences will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession numbers AJ228881–AJ228921.

**Data analysis:** The *rp49* sequence of the H27 line of *D. subobscura* (Aguadé 1988b; accession number M21333) was used as the reference sequence. Sequences were multiply aligned using the Clustal W program (Thompson *et al.* 1994) and edited with the MacClade version 3.06 program (Maddison and Maddison 1992). Phylogenetic analysis was performed using the neighbor-joining algorithm (Saitou and Nei 1987) implemented in the MEGA version 1.01 program (Kumar *et al.* 1994); bootstrap values were based on 500 replicates. The *rp49* nucleotide sequences of *D. guanche* and *D. madeirensis* (Ramos-Onsins *et al.* 1998; accession numbers Y09707–Y09708) were used as outgroups.

The DnaSP version 2.82 program (Rozas and Rozas 1997) was used to estimate nucleotide diversity, genetic distances, and genetic differentiation between populations and to detect gene conversion tracts. Nucleotide diversity,  $\pi$ , or the average number of nucleotide differences per site, was estimated using Equation 10.5 of Nei (1987). The silent nucleotide diversity (including both noncoding positions and synonymous sites in the coding region) was estimated as the average number of silent nucleotide differences per silent site (Nei and Gojobori 1986; Nei 1987). DNA divergence between arrangements was estimated as  $D_{xy}$ , the average number of nucleotide substitutions between populations (Equation 10.20 in Nei 1987), and as  $D_a$ , the net number of nucleotide substitutions (Equation

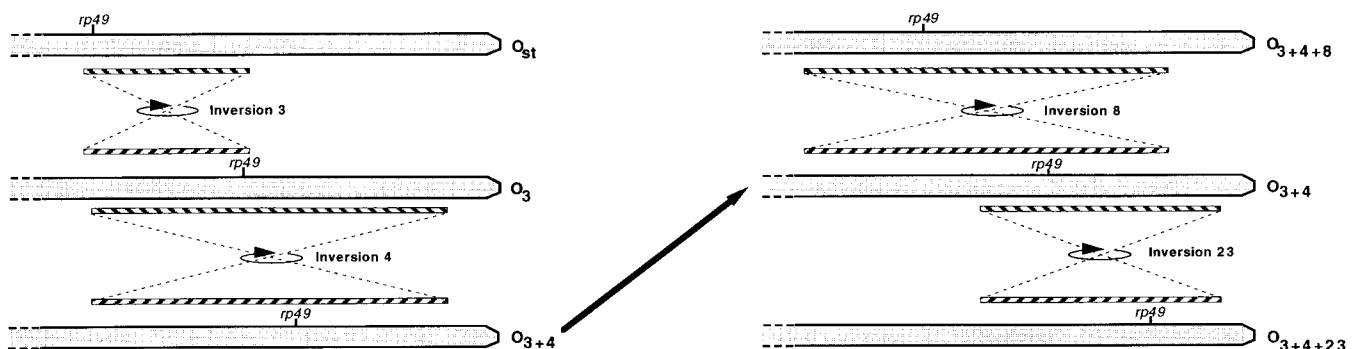


Figure 1.—Schematic representation of the location of the *rp49* gene in different chromosomal arrangements of *D. subobscura*. Only the distal part of the *O* chromosome is depicted. Shaded bars indicate the regions affected by inversions.  $O_3$  gene arrangement, present in *D. madeirensis* and *D. guanche*, is not present in extant populations of *D. subobscura*.

**TABLE 1**  
**Chromosomal gene arrangement frequencies**

	$O_{st}$		$O_{3+4}$		$O_{3+4+8}$		$O_{3+4+23}^a$	
	$n$	$F$	$n$	$F$	$n$	$F$	$n$	$F$
El Pedroso	16	0.13	18	0.79	0	0.08	0	0.00
Bizerte	0	0.00	0	0.11	21	0.58	20	0.31
Total	16		18		21		20	

$n$ , number of chromosomes analyzed in the present study;  $F$ , frequency of the chromosomal arrangement in the population.

<sup>a</sup> All  $O_{3+4+23}$  chromosomes studied also carried the inversion 6 (see Krimbas 1992), although this inversion (located in segment II of the  $O$  chromosome) does not include the *rp49* gene.

10.21 in Nei 1987).  $F_{ST}$ , the proportion of nucleotide diversity that is attributable to variation between populations, was estimated using Equation 3 in Hudson *et al.* (1992a). The average level of gene flow, measured as  $Nm$ , was estimated from  $F_{ST}$ , assuming that populations are structured in an island model and are at migration-drift equilibrium (Wright 1951; Hudson *et al.* 1992a).

The DnaSP program (Rozas and Rozas 1997) was also used to perform some neutrality tests. Tajima's test (Tajima 1989) and Fu and Li's tests (Fu and Li 1993) were performed using  $\eta$ , the total number of mutations (also called the minimum number of mutations), instead of  $S$ , the number of segregating sites (*i.e.*, considering the additional mutations present in those sites segregating for three nucleotides). The outgroup used for Fu and Li's tests was the *D. guanche rp49* sequence (Ramos-Onsins *et al.* 1998).

The critical values of Tajima's  $D$  statistic (Tajima 1989) and of the raggedness statistic  $r$  (Harpending *et al.* 1993; Equation 1 in Harpending 1994) were obtained by computer simulations (10,000 replicates) using the coalescent algorithm (for no recombination) described in Hudson (1990). The coalescent approach (Hudson 1990) was also used to estimate the confidence intervals of Tajima's  $D$  and  $r$  statistics for different levels of recombination.

We performed computer simulations to estimate the probability of obtaining  $k$  shared polymorphic sites segregating among all gene arrangements. The main assumption of the model is that the mutation rate is the same for all silent sites. In each replicate we generated four sets of sequences (one for each chromosomal class) of  $n$  sites (the number of silent sites), and we randomly spread in each set as many mutations as we had observed in each chromosomal class (each mutation would generate a polymorphic site). The probability of obtaining a number of shared polymorphic sites equal to or higher than the observed number  $k$ ,  $P(K \geq k)$ , was estimated as the proportion of the simulated replicates, where the number of shared polymorphic sites segregating in all four sets of sequences was  $\geq k$ . The null hypothesis that shared polymorphic sites are due to independent accumulation of mutations is rejected if that probability is  $<5\%$  critical value.

## RESULTS

**Nucleotide variation:** A total of 137 polymorphic sites (representing 147 mutations) were detected in the *rp49* region among the 75 lines of *D. subobscura* (Figure 2); all polymorphisms were silent. The average length of the DNA region studied was 1515 bp, while the total number of sites compared was 1467 (excluding sites

with alignment gaps), which represented a total of 1161 silent sites (including both synonymous sites and non-coding positions). Tables 2 and 3 give a summary of nucleotide polymorphism for the complete region, for the different functional regions, and for the different chromosomal classes. For the total data set, the average nucleotide diversity was  $\pi = 0.010$ , although the corresponding estimates for each chromosomal class were slightly lower (0.006–0.008). Figure 3 shows a plot of the estimates of polymorphism and divergence across the region studied. In the 5' flanking region there were some peaks of high polymorphism ( $\pi$ ) corresponding to regions of high divergence ( $K$ ).

**Genetic differentiation between gene arrangements:** Table 4 gives a summary of the genetic differentiation between gene arrangements. We found the highest values of genetic differentiation, as measured by  $D_{xy}$  and  $D_a$ , in all comparisons including the  $O_{st}$  gene arrangement. In fact, in comparisons between  $O_{3+4+X}$  chromosomes, the  $D_a$  values were  $\sim 0.001$ , while in comparisons between  $O_{st}$  and  $O_{3+4+X}$  chromosomes the  $D_a$  values were  $\sim 0.008$ . In addition, in comparisons including  $O_{st}$ , the  $D_{xy}$  values (and also  $D_a$  values) were rather similar regardless of the coexistence of arrangements in the same population (as, for instance,  $O_{st}$  and  $O_{3+4}$ ) or not (as, for instance,  $O_{st}$  and  $O_{3+4+23}$ ). Thus, geographic origin has a weak effect on genetic differentiation estimates, which is consistent with the lack of geographic differentiation within arrangement detected in Europe (Rozas *et al.* 1995). As in  $D_{xy}$  and  $D_a$  estimates, the largest  $F_{ST}$  values were also found in comparisons involving the  $O_{st}$  arrangement, and consequently these comparisons gave the lowest  $Nm$  values. We tested whether the different gene arrangements were genetically differentiated using the  $K_s^*$  test statistic (Hudson *et al.* 1992b). All pairwise comparisons showed highly significant probability values (Table 4), indicating that, despite the lower genetic differentiation of  $O_{3+4+X}$  chromosomes, all chromosomal classes were genetically differentiated. The differentiation between gene arrangements can be attributed to the presence of several fixed (or nearly fixed) nucleotide differences.



27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

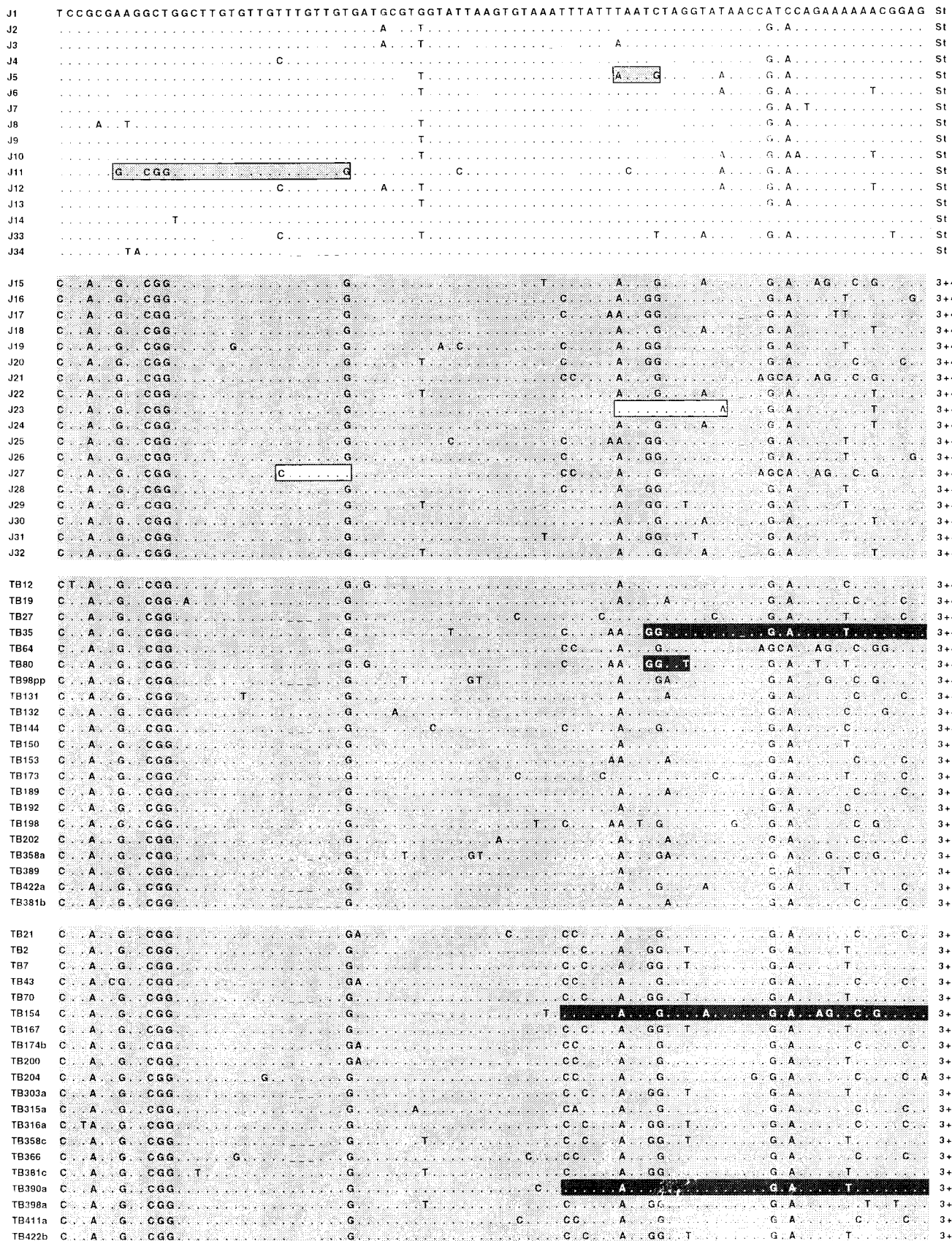


Figure 2.—Nucleotide polymorphism at the *rp49* gene region of *D. subobscura*. The distribution of the polymorphic sites in the different functional regions is as follows: polymorphic sites 27 to 850, 5' flanking region; 902 to 944, exon 1 (black); 981 to 1018, intron (gray); 1111 to 1207, (black); 1380 to 1540, 3' flanking region. Nucleotide numbering is according to Aguadé (1988b). Nucleotides identical to the first sequence are indicated by a dot, and deletions by a dash. Blocks of nucleotides transferred from one gene arrangement to the other are depicted as shaded boxes: white and gray colors indicate genetic information of  $O_{St}$  and  $O_{3+4+X}$ , respectively. Transfer of genetic information among  $O_{3+4+X}$  gene arrangements is depicted as black boxes.

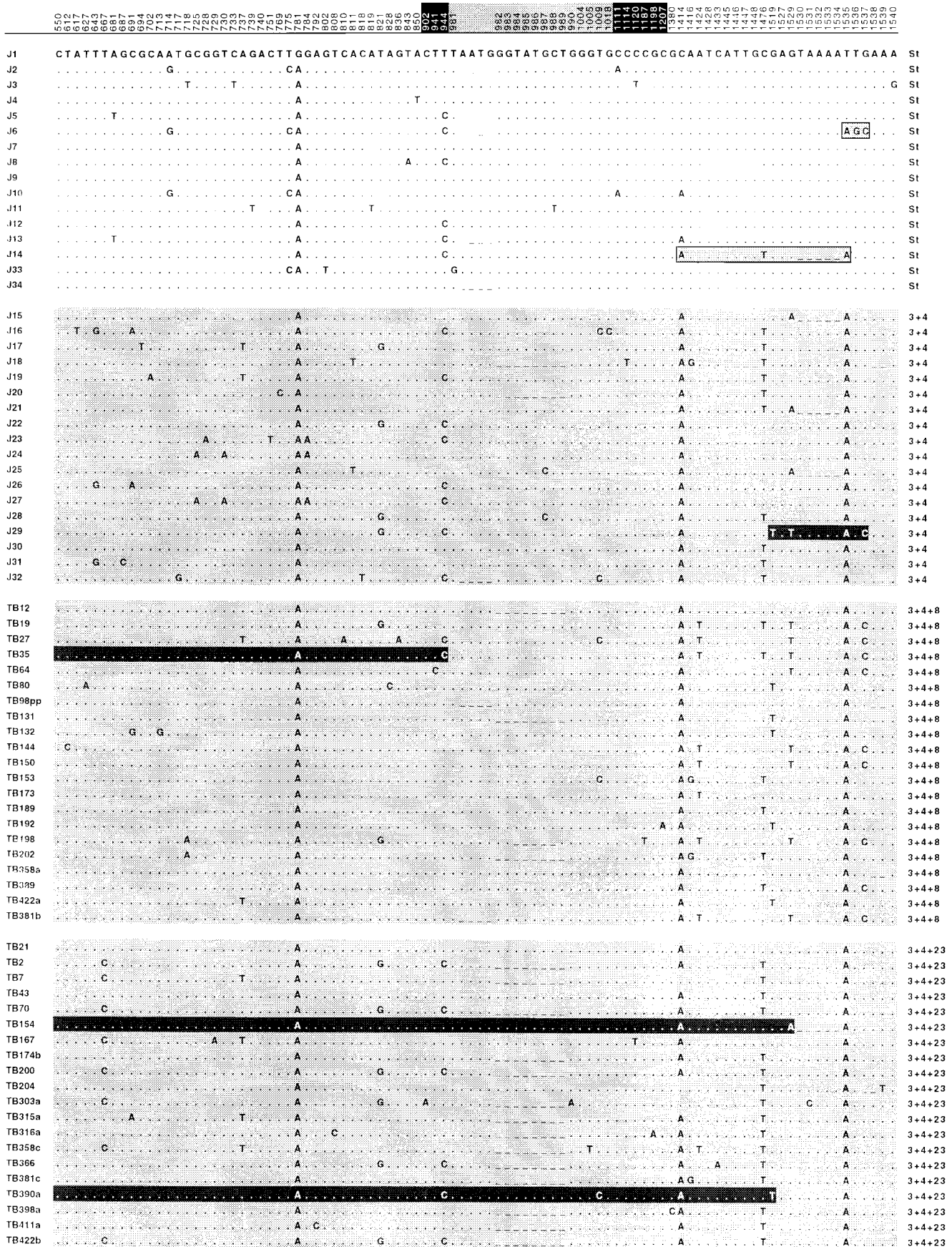


Figure 2.—(Continued)

**TABLE 2**  
**Nucleotide polymorphism in functional regions**

	5'	Exon 1	Intron	Exon 2	3'	Total
Number of sites <sup>a</sup>	817	93	48	309	200	1467
Silent sites	817	22	48	74.2	200	1161.2
$S(\eta)$	112 (121)	3 (3)	5 (5)	6 (6)	11 (12)	137 (147)
$\pi$	0.0149	0.0051	0.0048	0.0007	0.0102	0.0103
$\pi_s$	0.0149	0.0215	0.0048	0.0029	0.0102	0.0130

$S$ , number of segregating sites;  $\eta$ , total number of mutations;  $\pi$ , nucleotide diversity;  $\pi_s$ , silent nucleotide diversity.

<sup>a</sup>Total number of nucleotide sites excluding those sites with alignment gaps.

Comparison of the polymorphisms present in the different gene arrangements revealed several polymorphic sites segregating for the same variants in different arrangements, that is, shared polymorphic sites. Assuming that inversions have a unique origin, no shared polymorphism is expected at the origin of a particular inversion. Therefore, the presence of shared polymorphic sites can be due to: (i) mutations arisen independently in both gene arrangements (parallel mutations) or (ii) the transfer of genetic information between gene arrangements. Under the assumption that mutations arise randomly along the *rp49* region, the probability of obtaining  $k$  polymorphic sites shared by two chromosomal classes was estimated from the hypergeometric distribution (Rozas and Aguadé 1994) (Table 4), while the probability of obtaining  $k$  polymorphic sites shared among all chromosomal classes was estimated by computer simulations (see materials and methods). Probability values were obtained considering only the silent sites (*i.e.*, noncoding sites plus synonymous sites in the coding region,  $n = 1161$ ) and assuming that each site can have only two states (*i.e.*, two different nucleotides). This assumption is more conservative than that pre-

viously used in Rozas and Aguadé (1994), considering that each site had four possible states. The computer simulation analysis was performed according to the number of polymorphic sites observed in each gene arrangement (46, 54, 56, and 53); in this case, the probability of obtaining four (the observed number of polymorphic sites shared by all chromosomal classes) or more shared polymorphic sites segregating in all four gene arrangements is very low ( $P = 0.0000$ ). In all cases the null hypothesis that shared polymorphic sites are due to independent accumulation of mutations was rejected.

Genetic differentiation is a prerequisite to identify the footprint left by gene conversion in DNA sequences. Betrán *et al.* (1997) described an algorithm to detect gene conversion tracts from DNA sequence data from different populations. That method depends on the parameter  $\psi$  (Equation A4 in Betrán *et al.* 1997), which measures the probability of detecting a converted site; the higher the value of  $\psi$ , the more accurate will be the estimate of the number and length of the gene conversion tracts. In our sample, and in agreement with the  $D_{xy}$  and  $D_a$  values, the  $\psi$  values were relatively high

**TABLE 3**  
**Nucleotide polymorphism in different gene arrangements**

	$O_{st}$	$O_{3+4}$	$O_{3+4+8}$	$O_{3+4+23}$	$O_{3+4+X}$	Total
Sample size	16	18	21	20	59	75
$S(\eta)$	46 (47)	54 (56)	56 (58)	53 (54)	107 (115)	137 (147)
$\pi$	0.00635	0.00801	0.00789	0.00701	0.00822	0.01030
$\pi_s$	0.00799	0.01009	0.00994	0.00882	0.01037	0.01301
Tajima's $D$	-1.414	-1.114	-1.087	-1.263	-1.783*	-1.706*
$P$ value <sup>a</sup>	0.061	0.130	0.138	0.093	0.016	0.017
Fu and Li's $D^a$	-1.684	-1.103	-1.281	-2.773**	-3.638**	-3.636**
Fu and Li's $F^a$	-1.958*	-1.358	-1.489	-2.761*	-3.476**	-3.384**
Raggedness, $r$	0.0122**	0.0103*	0.0115*	0.0235	0.0053**	0.0024***
$P$ value <sup>a</sup>	0.008	0.011	0.028	0.215	0.011	0.001

$S$ , number of segregating sites;  $\eta$ , total number of mutations;  $\pi$ , nucleotide diversity;  $\pi_s$ , silent nucleotide diversity. \*  $0.01 < P < 0.05$ ; \*\*  $0.001 < P < 0.01$ ; \*\*\*  $P < 0.001$ .

<sup>a</sup> $P$  values for one-tailed test.



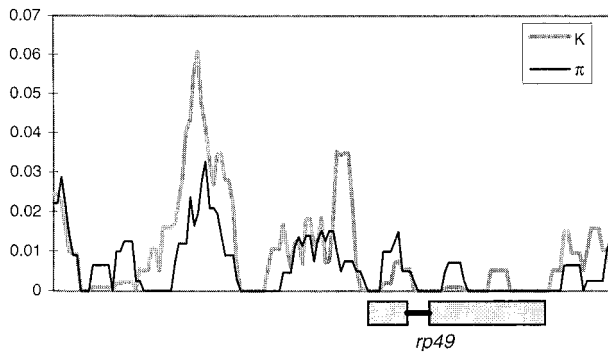


Figure 3.—Sliding window plot of nucleotide diversity,  $\pi$ , and of nucleotide divergence,  $K$ , across the *rp49* region for sequences of the  $O_{st}$  gene arrangement. Nucleotide divergence was estimated as the average number of nucleotide substitutions per site between *D. subobscura* (16  $O_{st}$  sequences) and *D. guanche*; values of  $K$  are divided by  $T + 1 = 3.883$ , where  $T$  is the divergence time (in  $2N$  generations, where  $N$  is the effective population size) between *D. guanche* and the  $O_{st}$  gene arrangement of *D. subobscura*. The  $T$  value was estimated as described in Hudson *et al.* (1987). Window size, 50 sites (excluding sites with alignment gaps); window displacement, 10 sites.

between  $O_{st}$  and  $O_{3+4+X}$  sequences and rather low among  $O_{3+4+X}$  sequences (Table 4). Using the algorithm of Betrán *et al.* (1997) we identified several gene conversion tracts (Figure 2). This algorithm can detect gene conversion tracts only if they contain at least two informative sites (Equation A1 in Betrán *et al.* 1997), and the tract length is underestimated by the nucleotide distance between them. Nearly all putative gene conversion tracts detected were transferred between  $O_{st}$  and  $O_{3+4+X}$  sequences. Although no distinction can be made between gene conversion and double crossover events if the tract includes the outermost informative nucleotides, genetic exchange by double crossover is expected

to affect longer chromosomal regions than gene conversion. In this sense, two of the detected recombination events (in lines TB154 and TB390a with the  $O_{3+4+23}$  arrangement) could correspond to double crossovers instead of to gene conversion events as the length of these tracts (between  $O_{3+4+X}$  arrangements) is unusually large. Tracts in lines TB154 and TB390a are longer than 1 kb, while the mean length of the gene conversion tracts observed between  $O_{st}$  and  $O_{3+4}$  arrangements is 51 bp. In addition, these long tracts include the outermost informative nucleotides of the data set, and, consequently, it cannot be dismissed that their length was still considerably large.

Figure 4 shows the neighbor-joining tree (Saitou and Nei 1987) of the *rp49* sequences. In the gene tree, sequences corresponding to the  $O_{st}$  gene arrangement form a monophyletic group, while those corresponding to the other gene arrangements ( $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$ ) are mixed. Under the assumption that inversions are monophyletic, the topology of the tree would indicate that gene flow is higher among  $O_{3+4+X}$  sequences than between  $O_{3+4+X}$  and  $O_{st}$  sequences.

**Neutrality tests:** The observed distribution of mutations within species was contrasted with that predicted by the neutral theory. The Tajima (1989) and the Fu and Li (1993) tests contrast different measures of  $\theta$  ( $\theta = 4Nu$ , where  $N$  is the effective population size and  $u$  is the mutation rate): those based on the number of segregating sites, on the number of pairwise differences, and on the number of mutations in external branches of the genealogy. Although the neutral model in general was not rejected, the probability values were very close to the 5% critical value (Table 3; see also Table 5).

We have also analyzed the pairwise nucleotide difference distribution, or mismatch distribution (Slatkin and Hudson 1991; Rogers and Harpending 1992), for

TABLE 4

Genetic differentiation between gene arrangements

	$O_{st}$ $O_{3+4}$	$O_{st}$ $O_{3+4+8}$	$O_{st}$ $O_{3+4+23}$	$O_{3+4}$ $O_{3+4+8}$	$O_{3+4}$ $O_{3+4+23}$	$O_{3+4+8}$ $O_{3+4+23}$
$\eta$	92	100	92	89	84	89
Shared mutations	12	6	8	25	24	21
$P$ value <sup>a</sup>	0.000	0.018	0.000	0.000	0.000	0.000
$D_{xy}$	0.01459	0.01445	0.01521	0.00907	0.00820	0.00877
$D_a$	0.00734	0.00727	0.00855	0.00103	0.00068	0.00144
$P$ value <sup>b</sup>	0.000	0.000	0.000	0.000	0.001	0.000
$\psi$	0.00874	0.00905	0.01046	0.00182	0.00134	0.00233
$F_{ST}$	0.501	0.501	0.560	0.113	0.083	0.164
$Nm$	0.25	0.25	0.20	1.97	2.77	1.28

$\eta$ , total number of mutations;  $\psi$ , parameter that measures the probability of detecting a converted site (Betrán *et al.* 1997).

<sup>a</sup>  $P$  values based on the hypergeometric distribution.

<sup>b</sup>  $P$  values for the  $K_s^*$  statistic.

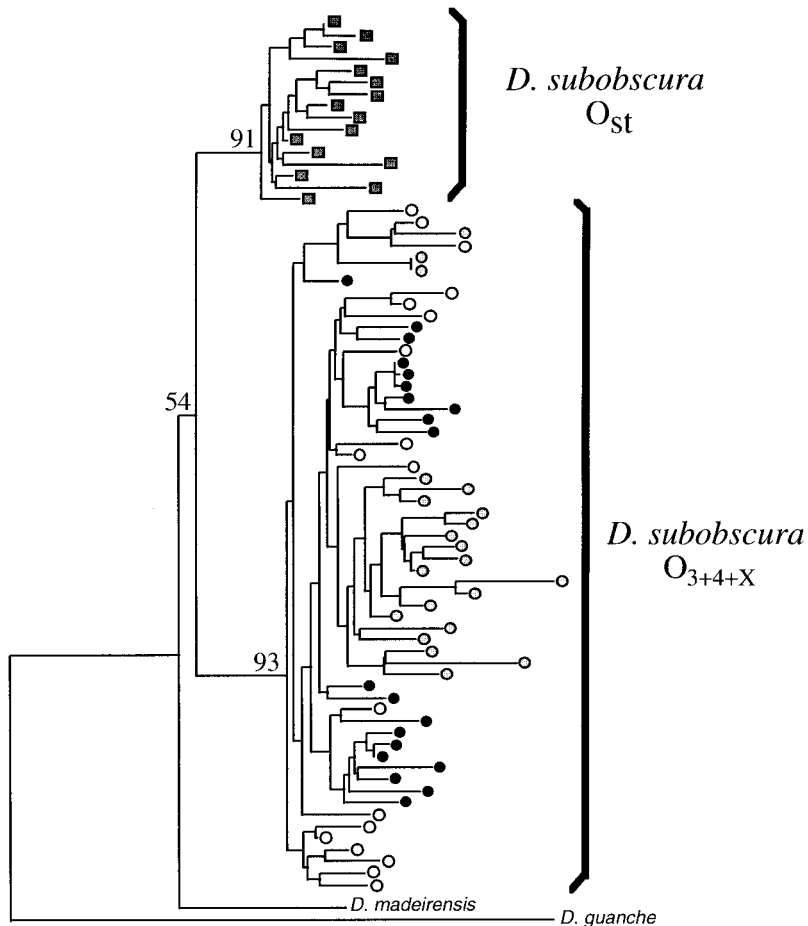


Figure 4.—Genealogy of the 75 sequences of the *rp49* gene region of *D. subobscura*. Informative nucleotide variants with evidence of genetic exchange (shaded boxes in Figure 2) have been subtracted. The tree was built by the neighbor-joining method and rooted using as outgroups the sequences of *D. madeirensis* and *D. guanche*. Numbers indicate bootstrap percentages based on 500 replicates; only bootstrap values for the main nodes are shown on the tree.  $O_{St}$  sequences are depicted as squares;  $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$  sequences are depicted as open, shaded, and solid circles, respectively.

each gene arrangement (Figure 5). These authors have shown that the expected pairwise difference distribution in a growing population resembles a Poisson distribution. In all gene arrangements, the observed distribution was close to the Poisson distribution and was fit poorly by the geometric distribution, which is expected in populations of constant size. Although the observed shape of the distribution is consistent with a process of population (in this case, gene arrangement) expansion, it does not constitute any evidence for expansion. In fact, under the neutral model with no intragenic recombination, the expected geometric distribution in populations of constant size is the sum of several quite different distributions (including bimodal and trimodal distributions; Watterson 1975; Slatkin and Hudson 1991; Rogers and Harpending 1992). However, the expected values of the raggedness statistic  $r$ , which quantifies the smoothness of the pairwise difference distribution, are lower in expanding populations than in populations of constant size (Harpending *et al.* 1993; Harpending 1994). To test whether our data are compatible with the constant population size model (Eller and Harpending 1996), we performed simulations on the basis of the neutral coalescent process without recombination (Hudson 1990). In the simulations the

sample size (the studied sample size in each gene arrangement) and  $\theta$  (the  $\theta$  estimated from the average number of nucleotide differences) were fixed for all replicates. The parameter  $r$  was estimated in each replicate, and the proportion of cases (replicates) where the  $r$  value was lower than the observed value was computed (Table 3). The constant population size model was rejected in all cases except for the  $O_{3+4+23}$  gene arrangement ( $P = 0.215$ ); however, when nucleotide variants included in gene conversion tracts were subtracted, that model was also rejected for this arrangement ( $P = 0.04$ ; Table 5).

On the other hand, in populations of constant size the expected shape of the pairwise difference distribution is a function of the intragenic recombination parameter  $C = 4Nc$  (where  $N$  is the effective population size, and  $c$  is the recombination rate per generation between the most distant sites). For high values of  $C$  the distribution of the pairwise differences might resemble a Poisson distribution, and thus the raggedness  $r$  statistic is expected to be lower. Additionally, the variance of the distribution of Tajima's  $D$  statistic is expected to be reduced with increasing values of the recombination parameter. Because there is recombination in homokaryotypes, we have conducted computer simulations



TABLE 5  
Tajima's *D* and raggedness *r* statistics in different gene arrangements

	$O_{st}$	$O_{3+4}$	$O_{3+4+8}$	$O_{3+4+23}$	$O_{3+4+X}$	Total
Sample size	16	18	21	20	59	75
Tajima's <i>D</i>	-1.262	-0.863	-1.074	-1.129	-1.798*	-1.722*
<i>P</i> value <sup>a</sup>	0.091	0.200	0.136	0.122	0.012	0.017
Raggedness, <i>r</i>	0.0137*	0.0130*	0.0092*	0.0147*	0.0049**	0.0020***
<i>P</i> value <sup>a</sup>	0.012	0.026	0.010	0.040	0.004	0.000

Estimates of Tajima's *D* and of the raggedness *r* statistics were performed using all nucleotide polymorphisms after subtracting all informative nucleotide variants with evidence of genetic exchange (shaded boxes in Figure 2). \*  $0.01 < P < 0.05$ ; \*\*  $0.001 < P < 0.01$ ; \*\*\*  $P < 0.001$ .

<sup>a</sup> *P* values for one-tailed test.

to determine the critical values of the raggedness *r* and Tajima's *D* statistics under different values of the recombination parameter (Table 6). In general, either the raggedness *r* statistic or Tajima's *D* statistic estimated from our sequences was significantly different from the expected values.

#### DISCUSSION

**Origin and evolution of inversions:** It has been classically considered that inversions have a unique origin, that is, that they are monophyletic (Powell 1997 and

references therein), due to the low probability of generating two simultaneous breaks at exactly the same positions independently in different chromosomes. The monophyletic character has been questioned, however, by the observation in some species of hot spots of inversion breakpoints and by the discovery of transposable elements. It has been argued that transposable elements, which can generate chromosomal rearrangements in the laboratory, could also have generated the naturally occurring inversions of *Drosophila*. However, the nucleotide sequence of regions covering the breakpoints of two such inversions did not reveal the

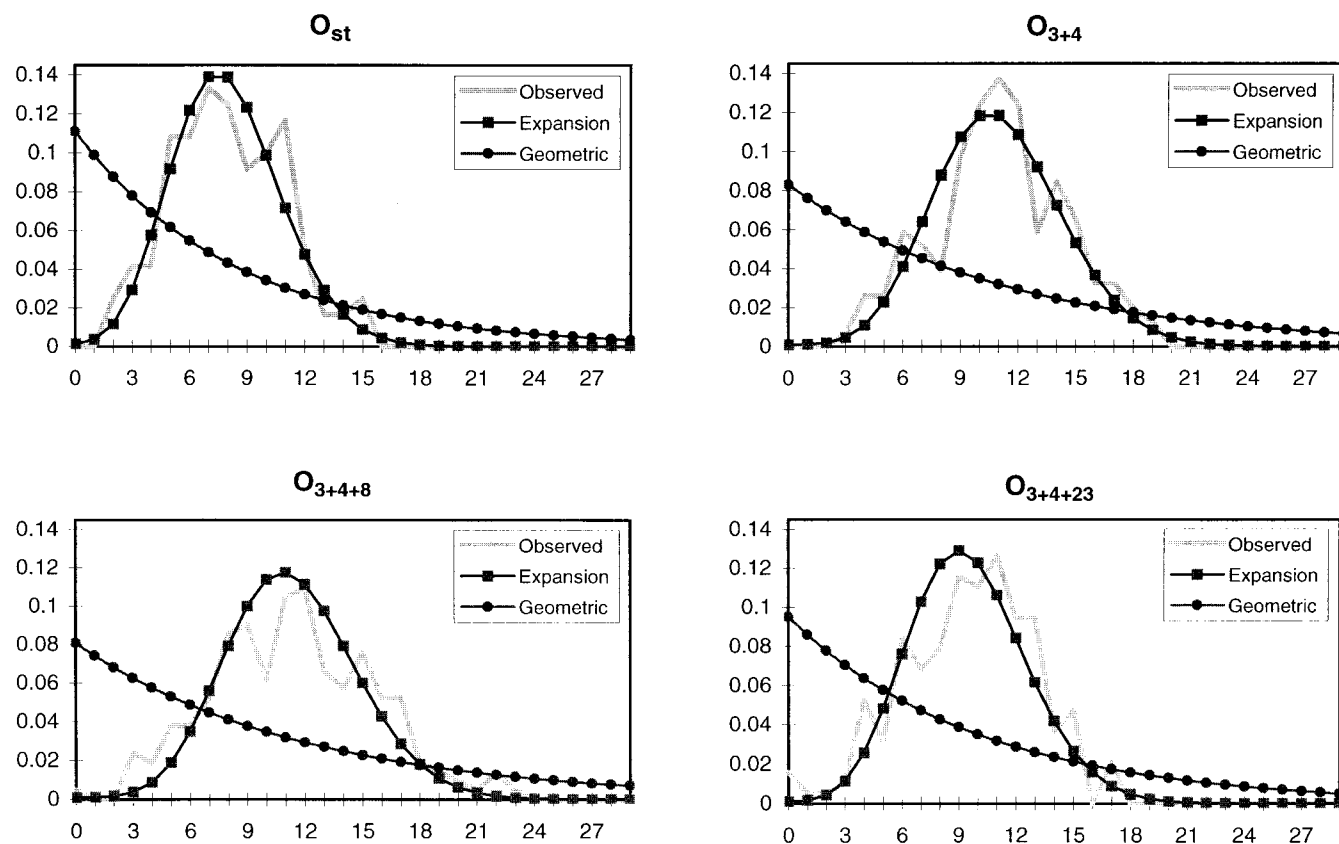


Figure 5.—Distribution of the pairwise nucleotide differences in the different gene arrangements studied. The expected values were obtained from Equation 3 in Rogers (1995) and considering  $\theta_0 = 0$ .

**TABLE 6**  
**Tajima's  $D$  and raggedness  $r$  statistics under different recombination values**

	$P(D < D_{\text{obs}})$	95% CI of $D$	$P(R < r_{\text{obs}})$	95% CI of $r$
		$O_{\text{st}} (n = 16; \theta = 9.425)$		
$C = 0$	0.0612	(-1.708, 1.690)	0.0085**	(0.0153, 0.2025)
$C = 10$	0.0170*	(-1.305, 1.248)	0.0273*	(0.0119, 0.1421)
$C = 25$	0.0052**	(-1.108, 1.063)	0.0511	(0.0106, 0.1075)
$C = 50$	0.0013**	(-0.965, 0.930)	0.0800	(0.0095, 0.0855)
		$O_{3+4} (n = 18; \theta = 11.895)$		
$C = 0$	0.1297	(-1.753, 1.701)	0.0109*	(0.0122, 0.1474)
$C = 10$	0.0503	(-1.321, 1.212)	0.0306*	(0.0099, 0.1052)
$C = 25$	0.0216*	(-1.083, 1.035)	0.0566	(0.0086, 0.0793)
$C = 50$	0.0075**	(-0.903, 0.919)	0.0922	(0.0078, 0.0637)
		$O_{3+4+8} (n = 21; \theta = 11.748)$		
$C = 0$	0.1377	(-1.711, 1.685)	0.0280*	(0.0112, 0.1262)
$C = 10$	0.0572	(-1.325, 1.235)	0.0830	(0.0084, 0.0924)
$C = 25$	0.0238*	(-1.077, 1.014)	0.1680	(0.0072, 0.0669)
$C = 50$	0.0101*	(-0.927, 0.894)	0.2521	(0.0064, 0.0527)
		$O_{3+4+23} (n = 20; \theta = 10.458)$		
$C = 0$	0.0927	(-1.785, 1.705)	0.2152	(0.0122, 0.1514)
$C = 10$	0.0309*	(-1.336, 1.243)	0.3941	(0.0093, 0.1035)
$C = 25$	0.0133*	(-1.108, 1.041)	0.5592	(0.0079, 0.0775)
$C = 50$	0.0040**	(-0.939, 0.927)	0.7006	(0.0071, 0.0587)

$C$ , recombination parameter;  $n$ , sample size;  $P(D < D_{\text{obs}})$ , probability of obtaining values of the Tajima's  $D$  lower than the observed;  $P(R < r_{\text{obs}})$ , probability of obtaining values of  $r$  (raggedness statistic) lower than the observed; CI, confidence intervals. \*  $0.01 < P < 0.05$ ; \*\*  $0.001 < P < 0.01$ .

presence of any transposable element (Wesley and Eanes 1994; Cirera *et al.* 1995).

Studies of nucleotide variation in inverted and noninverted chromosomes have revealed that the distribution of variation within and between gene arrangements varies according to the location of the region studied in the inversion heterokaryotypes. In those surveys where the region studied was close to an inversion breakpoint, all sequences of the inverted chromosomes formed a unique cluster in the gene tree. This was the case for the amylase (*Amy*) locus and the most common gene arrangements of the third chromosome of *D. pseudoobscura* (Aquadro *et al.* 1991; Popadić and Anderson 1994), the esterase-5 (*Est-5*) gene region, and the *sex-ratio* and *standard* chromosomes of *D. pseudoobscura* (Babcock and Anderson 1996), the *rp49* gene region and the  $O_{\text{st}}$  and  $O_{3+4}$  gene arrangements of *D. subobscura* (Rozas and Aguadé 1993, 1994), and both the breakpoint regions and the heat shock 83 (*Hsp83*) gene and the *In(3L)Payne* and *standard* chromosomes of *D. melanogaster* (Wesley and Eanes 1994; Hasson and Eanes 1996). These studies clearly support the monophyly of inversions. However, in those studies of nucleotide regions located in a more central position of the inversion loop, the sequences did not cluster in the gene tree according to their gene arrangements. This was the case for the *Amy* region and gene arrangements Tree Line, Olympic, Estes Park, and Hidalgo of *D. pseudoobscura* (Popadić *et al.* 1995), for the *rp49* region and gene arrangements  $O_{3+4}$  and  $O_{3+4+8}$  of *D. subobscura*

(Rozas and Aguadé 1993), for the *Est-6* gene region and *In(3L)Payne* and *standard* chromosomes of *D. melanogaster* (Hasson and Eanes 1996), and for the *P6* (also named *Fbp2*) region and *In(2L)t* and *standard* chromosomes of *D. melanogaster* (Bénassi *et al.* 1993). The results of these latter studies would be consistent with: (i) a multiple origin of the inversions, (ii) a unique origin of inversions with transfer of genetic information between gene arrangements (by double crossover or gene conversion), or (iii) a unique origin of inversions with an important number of parallel mutations.

Present data on nucleotide polymorphism at the *rp49* gene region of *D. subobscura* are compatible with a monophyletic origin of the studied inversions. The topology of the *rp49* gene tree (Figure 4), including sequences of the four major gene arrangements, clearly supports the monophyly of the  $O_{\text{st}}$  gene arrangement. However, for the other gene arrangements ( $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$ ) the sequences do not form unique clusters in the gene tree. As we have shown that parallel mutation cannot account for the high number of observed shared polymorphic sites, the topology of the tree for  $O_{3+4+X}$  arrangements would be compatible either with a multiple origin of the corresponding inversions (Figure 1) or with a unique origin with genetic exchange between arrangements. It should be noted, however, that the location of the *rp49* gene in the inversion loop is different in  $O_{3+4+X}$  heterozygotes than in heterozygotes between  $O_{\text{st}}$  and  $O_{3+4+X}$ . In the latter, the *rp49* region is very close to one of the breakpoints and within the

inversion, while in heterozygotes between  $O_{3+4+X}$  chromosomes the region is located in a more central position of the inversion loop. The observed pattern of nucleotide variation in the different arrangements and the location of the DNA region in the heterokaryotypes support a differential rate of genetic transfer between arrangements: no (or reduced) genetic exchange if the gene is at (or near) the breakpoint and extensive genetic exchange if the gene is in a more central position in the inversion. Therefore, the distribution of the  $O_{3+4+X}$  sequences in the gene tree would not provide evidence for a multiple origin of either inversion 8 or inversion 23 (see Figure 1).

#### Recombination and gene flow between inversions:

Here we have detected several short gene conversion tracts and no evidence of double crossovers when the *rp49* region is close to the inversion breakpoint (*i.e.*, in heterozygotes between  $O_{st}$  and  $O_{3+4+X}$  arrangements), and few short gene conversion tracts and some long tracts that might correspond to double crossovers when the gene is in a more central position in the inversion loop (*i.e.*, in heterozygotes between  $O_{3+4+X}$  arrangements). Our results are in agreement with some predictions of the genetic exchange between gene arrangements: a major role of gene conversion in regions around the inversion breakpoints as opposed to a major role of crossing over in the central regions of inversions (Navarro *et al.* 1997). These authors showed that exchange in inversion heterokaryotypes due to double crossing over would be negligible near the breakpoints and maximum in the central regions of the inverted fragment. Assuming that the gene conversion rate is constant across the inverted region, gene flow (both by double crossing over and gene conversion) would therefore be lower near the breakpoints and maximum in the central region of inversions. According to these predictions, the lower genetic differentiation detected between  $O_{3+4+X}$  chromosomes than between  $O_{st}$  and  $O_{3+4+X}$  gene arrangements is caused by the higher genetic exchange expected in the central part of inversions.

Alternatively, it could be argued that the observed pattern of genetic differentiation between arrangements was due to their geographic differentiation or to the differential frequency of  $O_{st}/O_{3+4+X}$  heterokaryotypes relative to heterokaryotypes between  $O_{3+4+X}$  arrangements. However, genetic differentiation between  $O_{st}$  and  $O_{3+4+X}$  arrangements is similar regardless of the geographic origin of the  $O_{3+4+X}$  chromosomes, which is consistent with previous studies at the *rp49* region showing that European populations are not genetically differentiated within each of these arrangements (Rozas *et al.* 1995). Also,  $O_{st}$  and  $O_{3+4}$  are the most frequent gene arrangements in these populations, and therefore genetic exchange between these arrangements could occur in the rather frequent  $O_{st}/O_{3+4}$  heterokaryotypes. These observations support the idea that the strong

genetic differentiation between  $O_{st}$  and  $O_{3+4+X}$  would not be caused by the geographic distribution of these arrangements, but by the reduced genetic exchange at the *rp49* region in these heterokaryotypes.

It has been shown that the rate of decay of locus-inversion disequilibria only depends on the rate of transfer of genetic information between gene arrangements (Ishii and Charlesworth 1977; Nei and Li 1980) and that the decay half-life of the association is of the order of the reciprocal of the genetic exchange rate. Assuming that all gene flow detected between  $O_{st}$  and  $O_{3+4}$  was due to gene conversion, Rozas and Aguadé (1994) estimated that the gene conversion rate per base pair and per generation is  $\sim 3.5 \times 10^{-7}$ . It is also possible to estimate the gene conversion rate from the total number of gene conversion tracts (those observed plus those inferred but not observed); for the *rp49* sequences of  $O_{st}$  and  $O_{3+4}$ , this rate has been estimated as  $2.7 \times 10^{-7}$  (Betrán *et al.* 1997).

The adaptive character of chromosomal polymorphism is well documented (see, *e.g.*, Prevosti *et al.* 1988), as well as the involvement in this adaptation of the differential gene content of inversions (Powell 1997). Coadaptation among loci within inversions has been proposed to maintain chromosomal polymorphism (Prakash and Lewontin 1968; see also Powell 1997). Our results (together with previous observations) provide valuable information on the possible distribution along the inversion of the members of the co-adapted complex. In fact, the reduction of recombination in inversion heterokaryotypes would play a fundamental role in maintaining blocks of genes together. However, nucleotide variation studies have shown that recombination is not totally suppressed and can be an important homogenizing factor, especially for genes located in the central part of the inversion. Therefore, selection could more easily maintain co-adapted complexes if the corresponding loci were in regions with low rates of genetic exchange between arrangements, that is, close to the inversion breakpoints. In the case of genes in a more central location in the inversion, selection should be stronger to successfully counteract the increased exchange of genetic information due to crossing over. Therefore, if selection were maintaining blocks of genes together, the target genes would be more likely located near the breakpoints.

**Nucleotide variation distribution:** The topology of the gene tree for all  $O_{st}$  and for all  $O_{3+4+X}$  sequences clearly resembles a star phylogeny (Figure 4), that is, a phylogeny where the tree is stretched near the terminal nodes and compressed near the root. This is the topology expected for populations that have recently expanded from a very small size and are therefore in the transient phase to equilibrium. During this phase, the specific footprint left by the expansion should be detected in the pattern of nucleotide variation. No footprint would be detected, however, if the elapsed time since the



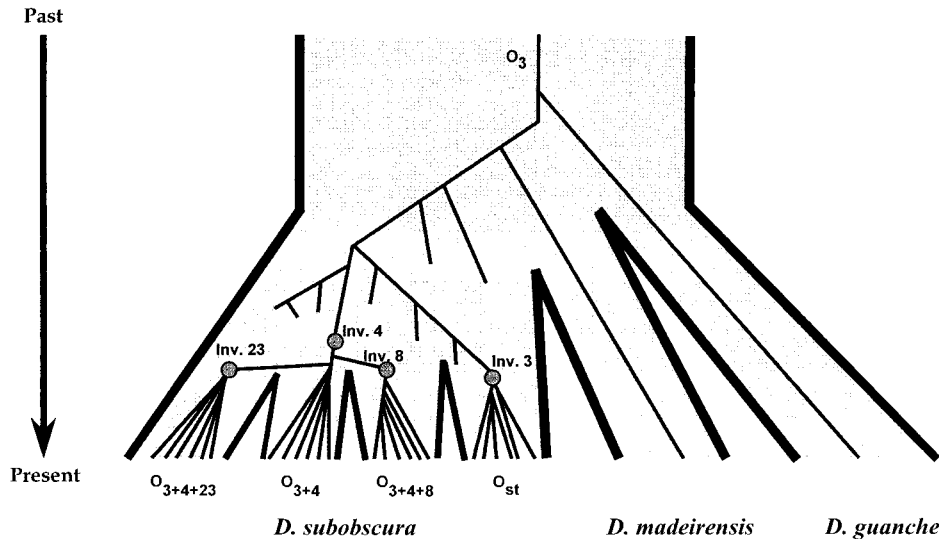


Figure 6.—Schematic representation of the history of the studied inversions and of the *rp49* genealogy. The genealogy, which is stretched near the terminal nodes and compressed near the root (star phylogeny), tries to reflect the expansion of the extant gene arrangements studied. Circles indicate the origin of inversions.

expansion were long enough (*e.g.*, more than  $4N$  generations).

Due to the unique origin of inversions, a particular gene arrangement increases in frequency (and therefore expands) from one copy to its current frequency in the population. In our case, the expansion of at least some extant arrangements was probably associated with the extinction of the ancestral  $O_3$  arrangement. The observed pattern of nucleotide variation would indicate, therefore, that the time since the origin of the particular inversion has not been long enough to reach equilibrium. The negative values of Tajima's  $D$  and of Fu and Li's  $D$  and  $F$  statistics and the Poisson shape of the pairwise difference distribution (and the corresponding small values of the raggedness statistic) might also support this interpretation. Nevertheless, in populations of constant size, Tajima's  $D$  and the raggedness  $r$  statistics are a function of the intragenic recombination level. We have shown that the higher the recombination parameter, the lower the raggedness  $r$  values and the lower the variance of Tajima's  $D$  statistic (Table 6). Although the actual value of the recombination parameter is not known for the region studied, in general, either the raggedness  $r$  statistic or Tajima's  $D$  statistic was significantly different from the expected values. This would allow us to conclude, therefore, that the observed pattern of nucleotide variation within gene arrangement still reflects the expansion of the corresponding inversion since its origin (Figure 6).

Because of the geographic distribution of the gene arrangements, some of them with a clinal distribution, the expansion hypothesis should be more appropriately contrasted with a stepping-stone model. However, both the estimated population size of *D. subobscura* ( $10^7$ , Comeron 1997) and the estimated times for the origin of the different inversions (see below) support the hypothesis that variation within gene arrangement has not yet attained equilibrium (*i.e.*, the time of origin should be

much lower than  $4N$  generations). On the other hand, the observed pattern of nucleotide variation could also reflect an expansion of the whole species, but in this case all loci in the genome would show the same pattern of variation. Nucleotide variation at the region encompassing the two *Acp70A* genes of *D. subobscura* (Cirera and Aguadé 1998), which is located in a region not affected by inversions, did not show, however, negative values of Tajima's  $D$  statistic. The hypothesis of a recent expansion of the whole species can be discarded on this basis, although additional data on variation at loci not associated with inversions would be needed to further support this conclusion.

**Age of inversions:** Assuming that the pattern of nucleotide variation within a particular arrangement reflects its expansion, we can estimate the time of origin of that arrangement from the coalescent time for its nucleotide sequences. The sudden expansion model basically depends on three parameters:  $\theta_0$ , or initial theta;  $\theta_1$ , or final theta; and  $\tau = 2ut$  (units of mutational time, where  $u$  is the mutation rate and  $t$  the time in generations; Rogers and Harpending 1992). We can estimate  $\tau$  (Rogers 1995, Equation 3) from the observed values of  $k$  (nucleotide diversity per sequence, that is, the average number of nucleotide differences between two sequences), considering that  $\theta_0 = 0$  (due to the unique origin of inversions). To estimate coalescent times, only variation that originated independently in each gene arrangement should be considered. For this reason, we have subtracted all nucleotide variants for which there is evidence of genetic transfer from another gene arrangement: each informative nucleotide variant within a detected gene conversion or double crossover tract has been replaced by the most frequent nucleotide variant in the recipient chromosomal class.

To estimate coalescent times, an estimate of the neutral mutation rate is needed. For a time of divergence between *D. guanche* and *D. subobscura* (Ramos-Onsins

TABLE 7

## Time of the origin of the different gene arrangements

	$O_{st}$	$O_{3+4}$	$O_{3+4+8}$	$O_{3+4+23}$	$O_{3+4+X}$
$\pi_S^a$	0.00678	0.00935	0.00963	0.00789	0.01001
Time <sup>b</sup>	0.24	0.33	0.34	0.28	0.35
Time <sup>c</sup>	0.37	0.51	0.53	0.43	0.55

<sup>a</sup>  $\pi_S$ , silent nucleotide diversity per site after subtracting all informative nucleotide variants with evidence of genetic exchange (shaded boxes in Figure 2).

<sup>b</sup> Time measured in millions of years for  $\lambda = 14.1 \times 10^{-9}$ .

<sup>c</sup> Time measured in millions of years for  $\lambda = 9.1 \times 10^{-9}$ .

*et al.* 1998) of 1.8 myr (if the Sophophoran radiation was 30 mya; Throckmorton 1975) or 2.8 myr (if the Sophophoran radiation was 46 mya; Beverley and Wilson 1984), the estimated rate of silent nucleotide substitution per site and per year at the *rp49* region would be  $\lambda = 14.1 \times 10^{-9}$  or  $\lambda = 9.1 \times 10^{-9}$ , respectively.

Table 7 shows the estimated times for the origins of the different inversions. These coalescent times are slightly lower than our previous estimates for some of the inversions (Rozas and Aguadé 1994). The differences in the estimates are mainly due to the different methods used for estimation: here we have assumed that variation within gene arrangement is not at equilibrium, while in our previous work the time estimates were based on the maximum value of the number of pairwise differences. In both studies, the time estimates have been obtained under the assumption that all nucleotide variants observed in a particular chromosomal arrangement originated in that arrangement. However, some additional variants not included in the detected gene conversion or double crossover tracts (*e.g.*, some shared polymorphisms) could have been transferred from other gene arrangements; in that case, most probably for the  $O_{3+4}$ ,  $O_{3+4+8}$ , and  $O_{3+4+23}$  arrangements, the coalescent time would be overestimated. If gene flow among these gene arrangements were actually quite important, the coalescent time for the  $O_{3+4+X}$  chromosomes would represent the coalescent time for inversion 4 (Figure 6).

We thank Serveis Científico-Tècnics, Universitat de Barcelona, for automated sequencing facilities. This work was supported by grants PB94-923 from Comisión Interdepartamental de Ciencia y Tecnología, Spain, and 1995SGR-577 from Comissió Interdepartamental de Recerca i Tecnologia, Catalonia, Spain, to M.A.

## LITERATURE CITED

- Aguadé, M., 1988a Restriction map variation at the *Adh* locus of *Drosophila melanogaster* in inverted and noninverted chromosomes. *Genetics* **119**: 135–140.
- Aguadé, M., 1988b Nucleotide sequence comparison of the *rp49* gene region between *Drosophila subobscura* and *D. melanogaster*. *Mol. Biol. Evol.* **5**: 433–441.
- Aquadro, C. F., A. L. Weaver, S. W. Schaeffer and W. W. Anderson, 1991 Molecular evolution of inversions in *Drosophila pseudo-*

- obscura*: the amylase gene region. *Proc. Natl. Acad. Sci. USA.* **88**: 305–309.
- Ashburner, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Babcock, C. S., and W. W. Anderson, 1996 Molecular evolution of the Sex-Ratio inversion complex in *Drosophila pseudoobscura*: analysis of the *Esterase-5* gene region. *Mol. Biol. Evol.* **13**: 297–308.
- Bénassi, V., S. Aulard, S. Mazeau and M. Veuille, 1993 Molecular variation of *Adh* and *P6* genes in an African population of *Drosophila melanogaster* and its relation to chromosomal inversions. *Genetics* **134**: 789–799.
- Betrán, E., J. Rozas, A. Navarro and A. Barbadilla, 1997 The estimation of the number and the length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146**: 89–99.
- Beverley, S. M., and A. C. Wilson, 1984 Molecular evolution in *Drosophila* and the higher Diptera. II. A time scale for fly evolution. *J. Mol. Evol.* **21**: 1–13.
- Cirera, S., and M. Aguadé, 1998 Molecular evolution of a duplication: the Sex-Peptide (*Acp70A*) gene region of *Drosophila subobscura* and *D. madeirensis*. *Mol. Biol. Evol.* **15**: 988–996.
- Cirera, S., J. M. Martín-Campos, C. Segarra and M. Aguadé, 1995 Molecular characterization of the breakpoints of an inversion fixed between *Drosophila melanogaster* and *D. subobscura*. *Genetics* **139**: 321–326.
- Comeron, J. M., 1997 Estudi de la variabilitat nucleotídica a *Drosophila*: Regió *Xdh* a *D. subobscura*. Ph.D. Thesis. Universitat de Barcelona.
- Eller, E., and H. C. Harpending, 1996 Simulations show that neither population expansion nor population stationarity in a West African population can be rejected. *Mol. Biol. Evol.* **13**: 1155–1157.
- Fu, Y.-X., and W.-H. Li, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- Harpending, H. C., 1994 Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Hum. Biol.* **66**: 591–600.
- Harpending, H. C., S. T. Sherry, A. R. Rogers and M. Stoneking, 1993 The genetic structure of ancient human populations. *Curr. Anthropol.* **34**: 483–496.
- Hasson, E., and W. F. Eanes, 1996 Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. *Genetics* **144**: 1565–1575.
- Hudson, R. R., 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**: 1–44.
- Hudson, R. R., M. Kreitman and M. Aguadé, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- Hudson, R. R., M. Slatkin and W. P. Maddison, 1992a Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.
- Hudson, R. R., D. D. Boos and N. L. Kaplan, 1992b A statistical test for detecting population subdivision. *Mol. Biol. Evol.* **9**: 138–151.
- Ishii, K., and B. Charlesworth, 1977 Associations between allozyme loci and gene arrangements due to hitch-hiking effects of new inversions. *Genet. Res.* **30**: 93–106.
- Krimbas, C. B., 1992 The inversion polymorphism of *Drosophila subobscura*, pp. 127–220 in *Drosophila Inversion Polymorphism*, edited by C. B. Krimbas and J. R. Powell. CRC Press, Boca Raton, FL.
- Krimbas, C. B., and M. Loukas, 1980 The inversion polymorphism of *Drosophila subobscura*. *Evol. Biol.* **12**: 163–234.
- Kumar, S., K. Tamura and M. Nei, 1994 MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput. Appl. Biosci.* **10**: 189–191.
- Maddison, W. P., and D. R. Maddison, 1992 MacClade: analysis of phylogeny and character evolution, version 3.0. Sinauer Associates, Sunderland, MA.
- Navarro, A., E. Betrán, A. Barbadilla and A. Ruiz, 1997 Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* **146**: 695–709.
- Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- Nei, M., and W.-H. Li, 1980 Non-random association between elec-

- tomorphs and inversion chromosomes in finite populations. *Genet. Res.* **35**: 65–83.
- Popadić, A., and W. W. Anderson, 1994 The history of a genetic system. *Proc. Natl. Acad. Sci. USA.* **91**: 6819–6823.
- Popadić, A., and W. W. Anderson, 1995 Evidence for gene conversion in the amylase multigene family of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **12**: 564–572.
- Popadić, A., D. Popadić and W. W. Anderson, 1995 Interchromosomal exchange of genetic information between gene arrangements of the third chromosome of *Drosophila pseudoobscura*. *Mol. Biol. Evol.* **12**: 938–943.
- Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology*. Oxford University Press, New York.
- Prakash, S., and R. C. Lewontin, 1968 A molecular approach to the study of genic heterozygosity III. Direct evidence of coadaptation in gene arrangements of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **59**: 398–405.
- Prevosti, A., G. Ribó, L. Serra, M. Aguadé, J. Balañá *et al.*, 1988 Colonization of America by *Drosophila subobscura*: experiment in natural populations that supports the adaptive role of chromosomal-inversion polymorphism. *Proc. Natl. Acad. Sci. USA* **85**: 5597–5600.
- Ramos-Onsins, S., C. Segarra, J. Rozas and M. Aguadé, 1998 Molecular and chromosomal phylogeny in the *obscura* group of *Drosophila* inferred from sequences of the *rp49* gene region. *Mol. Phylogenet. Evol.* **9**: 33–41.
- Rogers, A. R., 1995 Genetic evidence for a pleistocene population explosion. *Evolution* **49**: 608–615.
- Rogers, A. R., and H. Harpending, 1992 Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**: 552–569.
- Rozas, J., and M. Aguadé, 1990 Evidence of extensive genetic exchange in the *rp49* region among polymorphic chromosome inversions in *Drosophila subobscura*. *Genetics* **126**: 417–426.
- Rozas, J., and M. Aguadé, 1993 Transfer of genetic information in the *rp49* region of *Drosophila subobscura* between different chromosomal gene arrangements. *Proc. Natl. Acad. Sci. USA.* **90**: 8083–8087.
- Rozas, J., and M. Aguadé, 1994 Gene conversion is involved in the transfer of genetic information between naturally occurring inversions of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **91**: 11517–11521.
- Rozas, J., and R. Rozas, 1997 DnaSP version 2: a novel software package for extensive molecular population genetics analysis. *Comput. Appl. Biosci.* **13**: 307–311.
- Rozas, J., C. Segarra, C. Zapata, G. Álvarez and M. Aguadé, 1995 Nucleotide polymorphism at the *rp49* region of *Drosophila subobscura*: lack of geographic subdivision within chromosomal arrangements in Europe. *J. Evol. Biol.* **8**: 355–367.
- Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi *et al.*, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
- Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Slatkin, M., and R. R. Hudson, 1991 Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555–562.
- Sperlich, D., H. Feuerbach-Mravlag, P. Lange, A. Michaelidis and A. Pentzos-Daponte, 1977 Genetic load and viability distribution in central and marginal populations of *Drosophila subobscura*. *Genetics* **86**: 835–848.
- Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Throckmorton, L. H., 1975 The phylogeny, ecology and geography of *Drosophila*, pp. 421–469 in *Handbook of Genetics*, Vol. 3, edited by R. C. King. Plenum Press, New York.
- Watterson, G. A., 1975 On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- Wesley, C. S., and W. F. Eanes, 1994 Isolation and analysis of the breakpoint sequences of chromosome inversion *In(3L)Payne* in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **91**: 3132–3136.
- Wright, S., 1951 The genetical structure of populations. *Ann. Eugen.* **15**: 323–354.

Communicating editor: N. Takahata