Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5000–5004, July 1989
Evolution

Polymorphism for the number of tandemly multiplicated glycerol-3-phosphate dehydrogenase genes in *Drosophila melanogaster*

(population genetics/evolution/multigene family/gene duplication/restriction polymorphism)

TOSHIYUKI TAKANO, SHINICHI KUSAKABE, AKIHIKO KOGA, AND TERUMI MUKAI

Department of Biology, Faculty of Science, Kyushu University 33, Fukuoka 812, Japan

Communicated by Motoo Kimura, April 10, 1989 (received for review October 14, 1988)

ABSTRACT **A 26-kilobase-pair region encompassing the *sn*-glycerol-3-phosphate dehydrogenase (*sn*-glycerol-3-phosphate:NAD$^+$ 2-oxidoreductase, EC 1.1.1.8) locus in *Drosophila melanogaster* from two natural populations in Japan was surveyed by restriction mapping. Both tandem duplications and triplications in this region were found in both populations. Detailed analysis of 86 chromosome 2 lines revealed restriction site and allozyme polymorphisms in the transcriptional unit: two restriction sites and the allozymes [fast (F) or slow (S)] were polymorphic among both duplication-bearing chromosomes and those carrying the standard sequence. This finding suggests recurrent recombination and/or gene conversion in this 5-kilobase-pair region. The differences observed for restriction site and allozyme haplotypes among the triplicated sequence both within and between populations, together with the distribution in natural populations, suggest a relatively recent ancestry of the triplication events and an independent origin in respective populations. Such events may represent the process of the formation of multigene families [compare Ohta, T. (1987) *Genetics* 115, 207–213]. Finally, the evolution of this type of polymorphism is discussed.**

Gene duplication has evidently played a very important role in evolution (1, 2). The multigene families such as ribosomal RNA genes, histone genes, or immunoglobulin genes represent the extreme cases of gene multiplication. Gene duplication also allows each duplicate segment to evolve independently, resulting in the creation of functionally diverged genes. Recent advances of molecular biology have clarified not only the genetic organization of many repeated gene families but also the commonness of gene duplication for ordinary gene loci. In fact, a variety of studies for gene duplication have been reported for various gene loci in *Drosophila*, including *Bar* (3) and loci for alcohol dehydrogenase (4–6), α-amylase (7, 8), rosy (9), maroon-like and rosy (10), white (11, 12), and metallothionein (13, 14).

In a study of the evolution of multigene families, there are two major problems. One is to clarify the mechanisms of gene amplification, and the other is to understand the mutant dynamics of gene families in a population. To analyze the second problem, mutation, selection, random genetic drift, gene conversion, and the continuous processes of duplication and deletion should be taken into account (2).

Tandem duplications have recently been reported to be polymorphic in the cytogenetic region 26A, which includes the gene *Gpdh* encoding glycerol-3-phosphate dehydrogenase (GPDH; *sn*-glycerol-3-phosphate:NAD$^+$ 2-oxidoreductase, EC 1.1.1.8; map position 2-20.5) (15). However, this analysis was performed by using the previously cloned plasmid *pDm60(a)c* (16) as a probe, which itself carries a duplicated fragment of *Gpdh* (ref. 17; G. C. Bewley, J. L. Cook, S.K., T.M., G. Chambers, and D. Rigby, unpublished data). Therefore, it was difficult to compare the intensities of hybridization of the tandem duplicated fragment or fragments with that of other fragments in the same lane. We have subsequently constructed probes with no duplication and examined the restriction map variation in the same region in two natural populations of *Drosophila melanogaster*. In addition to tandem duplications, triplications were observed in both populations.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Drosophila Stocks. Fly collections were made from the following two natural populations of *D. melanogaster*: Chichijima (Bonin) Island, Ogasawara (southern Japan) in 1982, and two wineries near Hirosaki city, Aomori prefecture (northern Japan) in 1983 (18). After isofemale lines were established, they were maintained in the 18°C laboratory. With the *Cy-Pm* method (19), about 100 lines from the Ogasawara population and 300 from the Aomori population were established that were homozygous for chromosome 2 from independent isofemale lines. Before the experiments, cytological examination of the salivary gland chromosomes was conducted for each chromosome line after a cross was made to the *cn bw* line with the standard karyotype. To examine restriction map variation, 43 second chromosomes with the standard karyotype, one from each isofemale line, were chosen at random from each population. For each chromosome line, the α-GPDH allozyme was assayed by starch gel and polyacrylamide gel electrophoresis.

Probe Preparation. A genomic library was constructed in phage λ Charon 35 (20) from an isogenic second chromosome line of *D. melanogaster*, which was established from a fly caught in Raleigh, NC. This second chromosome carries no duplication in the *Gpdh* region (15). From this library, genomic clones containing *Gpdh* were isolated by using the previously cloned plasmid, *pDm60a(c)* (16), as a probe. All plasmid subclones used as the hybridization probes were constructed by using the vector pUC13 (21) and are shown in Fig. 1 together with its restriction map.

Restriction Map Analysis. Genomic DNA from each line was prepared as described (22, 23). The following eight endonucleases were used in the analysis: *Eco*RI, *Hin*dIII, *Bam*HI, *Pst* I, *Sal* I, *Sac* I, *Xba* I, and *Xho* I. Genomic DNA (5–8 μg) was digested with single or double enzymes, and the DNA digests were electrophoresed on 0.5–1.2% agarose gel. The DNA fragments were transferred to a nylon filter (Biodyne, commercially available from Pall) as described (24). The DNA probes were labeled with [α-^{32}P]dCTP by using the kit and protocol from Amersham (Multiprime DNA labeling system) (25, 26). The conditions for hybridization, wash, and autoradiography have been described (27). Modifications are as follows: hybridization was carried out for 2 or 3 days at

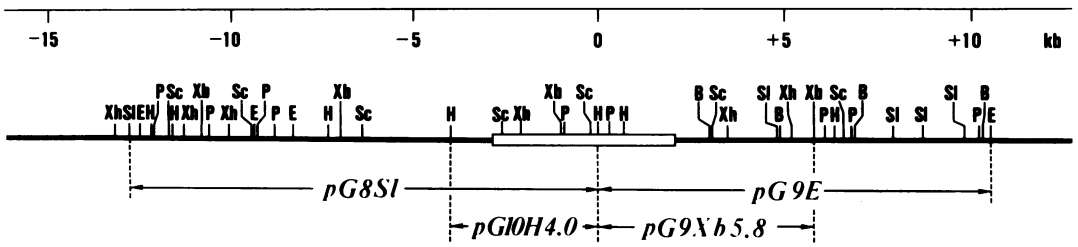

FIG. 1. Restriction map of the *Gpdh* region. This map was constructed by using the clones from an isogenic second chromosome line of *D. melanogaster* (see text) and does not represent the most common arrangement. Restriction sites that could not be detected in the present experiment were ignored in this figure. The open box represents the transcriptional unit, and its direction is left to right (G. C. Bewley, J. L. Cook, S.K., T.M., and G. Chambers, unpublished data). Restriction sites are indicated above the bar (E = *Eco*RI, H = *Hin*dIII, B = *Bam*HI, P = *Pst* I, Sl = *Sal* I, Sc = *Sac* I, Xh = *Xho* I, and Xb = *Xba* I). Four subcloned DNA fragments used as probes are indicated below the map (*pG8S1*, *pG9E*, *pG10H4.0*, and *pG9Xb5.8*). The scale in kilobase pairs (kb) has zero at one of the *Hin*dIII sites in the *Gpdh* gene.

62°C in hybridization solution (6× SSC/10× Denhardt's solution/0.1% sodium dodecyl sulfate/0.5 mg of denatured salmon sperm DNA per ml, in which 1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7, and 1× Denhardt's solution = 0.02% polyvinylpyrrolidone/0.02% Ficoll/0.02% bovine serum albumin). After hybridization, the filters were washed sequentially in 6× SSC, 3× SSC, 1× SSC, and 0.5× SSC containing 0.1% sodium dodecyl sulfate at 62°C for 1 hr each, and the final wash was in 0.5× SSC at room temperature for 15 min.

RESULTS AND DISCUSSION

Evidence for the Presence of Triplications. Sequence analysis of a duplicated fragment in the region 26A (S.K., H. Baba, A.K., G. C. Bewley, and T.M., unpublished data) has revealed the following organization for a tandem duplication of *Gpdh*: a 4.5-kb-long duplicated fragment that includes most of *Gpdh* except for about a 0.5-kb segment 5′ to the gene that includes all of exons 1 and 2. This duplicated fragment is located just downstream of the functional and complete *Gpdh* transcriptional unit. The following considerations are based on the above findings. Fig. 2 presents the autoradiographs of genomic Southern blots and restriction maps for some typical lines. *Eco*RI/*Bam*HI double digests (Fig. 2*A*) clearly show that line D contains a 9.0-kb insertion in comparison with the standard arrangement (line A). This insertion fragment can be cut into two fragments of equal size by an *Xba* I digest (lines F and D in Fig. 2*B*). Indeed, the hybridization intensities of the 4.5-kb fragment in lines F and D were compared with that of duplication-bearing lines B, E, and C by densitometry of this autoradiograph. By using the 7.0-kb fragment in this autoradiograph as an internal control, the ratios of the intensities of the 4.5-kb fragment to that of the 7.0-kb fragment are as follows: 1.02 for line B, 0.79 for line E, 0.98 for line C, 2.20 for line F, and 1.77 for line D. The same results were obtained from *Sac* I and *Xho* I digests. The transcriptional unit has only one recognition sequence for *Sac* I, *Xho* I, and *Xba* I.

In the case of *Hin*dIII and *Pst* I digests, no extra band was found in lines F and D in comparison with line B, a typical duplication-bearing line (Fig. 2*C* and Fig. 2*D*), except for a 6.6-kb *Pst* I fragment in line D caused by a loss of the 5′ flanking site. The hybridization intensity for the 3.9-kb and 0.7-kb *Hin*dIII fragments and the 2.0-kb, 1.3-kb, and 1.2-kb *Pst* I fragments in lines F and D increased when compared with line B (data not shown).

All of the above findings consistently suggest that lines F and D carry triplet elements of *Gpdh* and that both of the two downstream duplicated fragments perhaps lack the 0.5-kb sequence at the 5′ side of the gene. According to this criterion, lines were classified into three groups—i.e., standard, duplication, and triplication. The numbers of chromosomes assigned to respective classes are 27, 10, and 6 in the Ogasawara population and 30, 11, and 2 in the Aomori population.

Restriction Site Polymorphism. We have mapped and scored 54 restriction sites in the 26-kb region, 18 of which were polymorphic. Table 1 shows the distribution of polymorphic restriction sites in the transcriptional unit, together with the α-GPDH allozyme [fast (F) or slow (S)]. Also five monomorphic restriction sites were scored in the same region and in each of the duplicated and triplicated fragments. *Pst* I (position −2.2), *Xho* I (position −2.0), and *Sac* I (position −0.2) sites were polymorphic only in the transcriptional unit, but not in the additional replicated fragments. On the other hand, the precise position of the polymorphic *Hin*dIII site could not be identified in the duplication class (Fig. 2*C*). However, there was no haplotype that lacked both of the *Hin*dIII sites in the transcriptional unit and in the duplicated fragment. In addition, the α-GPDH allozymes were polymorphic in all of the three classes. The losses of the *Hin*dIII (0.0) and *Pst* I (−2.2) sites were observed not only in the standard class but also in the duplication class. Furthermore, of 17 polymorphic sites observed in the 26-kb region among the standard chromosomes, 4 sites in the Ogasawara duplications, 6 sites in the Aomori duplications, and 5 sites in the Ogasawara triplications turned out to be polymorphic. These findings suggest continuous occurrence of recombination and/or gene conversion in this region.

Unfortunately, the present data were not enough to compare the polymorphism between the transcriptional unit and duplicated copies. Direct DNA sequence analysis in this region should provide us a valuable piece of information for understanding the evolution of gene families.

It should be noted here that the haplotypes of the triplication class are similar to each other within each population, but the triplications from the two populations can be distinguished by the α-GPDH allozyme—that is, F in the Ogasawara triplication and S in the Aomori triplication. The mean number of restriction site differences (28) was estimated based on the data set of 54 restriction sites scored in the 26-kb region and the allozyme variation. The result is summarized in Table 2. Also given is the number of net restriction-site differences between populations, which are obtained by subtracting the mean of differences within populations from those between populations (28). The result for the triplication class is in striking contrast to that for the other two classes, although the standard errors of the estimates were very large. Polymorphic restriction sites observed among the Ogasawara triplication chromosomes are also polymorphic for the standard chromosomes, which suggests that this polymorphism arose through recombination.

Duplication of *Gpdh* has been observed in three natural populations, two from the United States and one from Japan (15). Using the same probes as that in this study, we reexamined 22 duplication-bearing chromosomes previously examined with *pDm60a(c)* (15). However, we found no tripli-

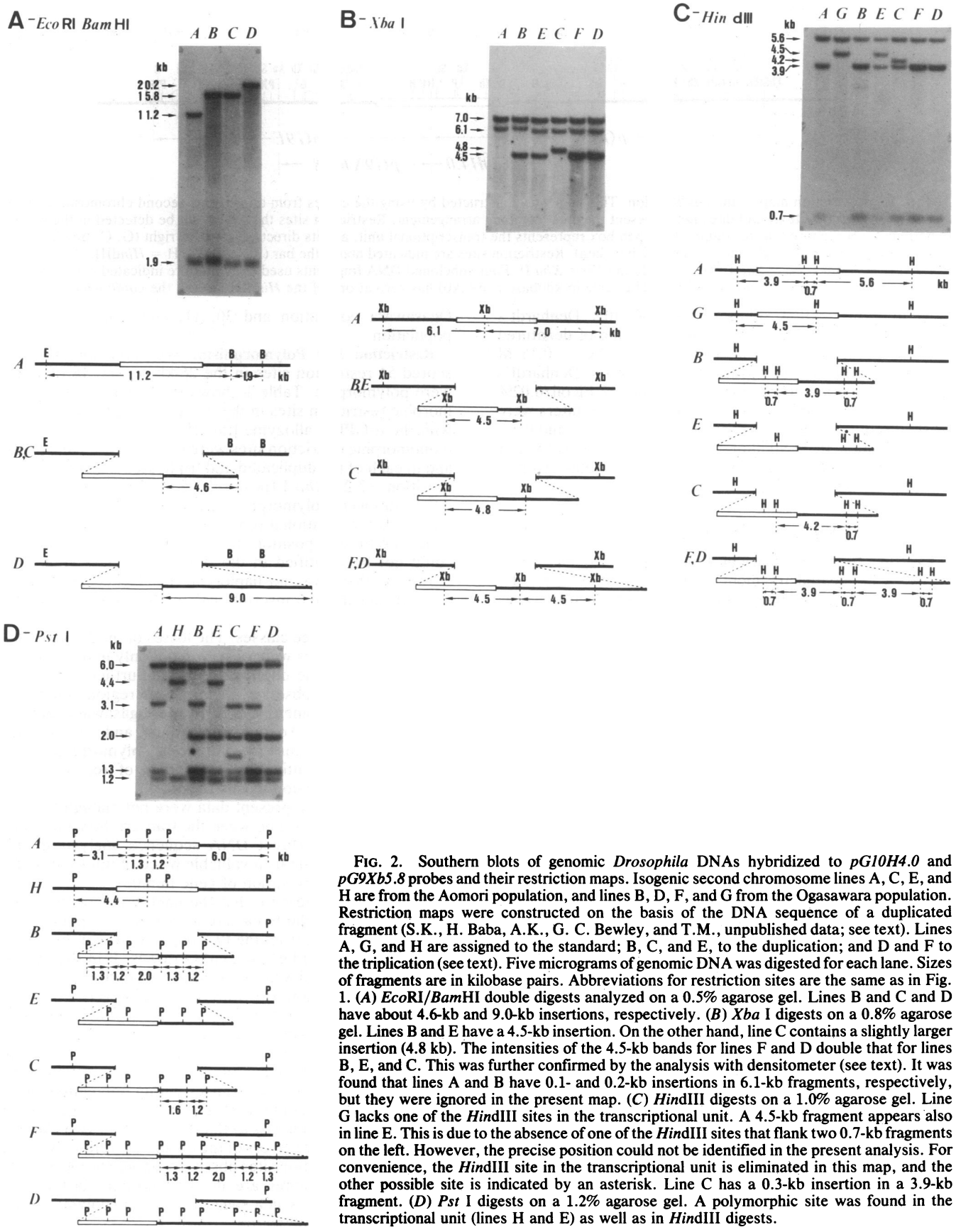

FIG. 2. Southern blots of genomic *Drosophila* DNAs hybridized to *pG10H4.0* and *pG9Xb5.8* probes and their restriction maps. Isogenic second chromosome lines A, C, E, and H are from the Aomori population, and lines B, D, F, and G from the Ogasawara population. Restriction maps were constructed on the basis of the DNA sequence of a duplicated fragment (S.K., H. Baba, A.K., G. C. Bewley, and T.M., unpublished data; see text). Lines A, G, and H are assigned to the standard; B, C, and E, to the duplication; and D and F to the triplication (see text). Five micrograms of genomic DNA was digested for each lane. Sizes of fragments are in kilobase pairs. Abbreviations for restriction sites are the same as in Fig. 1. (*A*) *Eco*RI/*Bam*HI double digests analyzed on a 0.5% agarose gel. Lines B and C and D have about 4.6-kb and 9.0-kb insertions, respectively. (*B*) *Xba* I digests on a 0.8% agarose gel. Lines B and E have a 4.5-kb insertion. On the other hand, line C contains a slightly larger insertion (4.8 kb). The intensities of the 4.5-kb bands for lines F and D double that for lines B, E, and C. This was further confirmed by the analysis with densitometer (see text). It was found that lines A and B have 0.1- and 0.2-kb insertions in 6.1-kb fragments, respectively, but they were ignored in the present map. (*C*) *Hin*dIII digests on a 1.0% agarose gel. Line G lacks one of the *Hin*dIII sites in the transcriptional unit. A 4.5-kb fragment appears also in line E. This is due to the absence of one of the *Hin*dIII sites that flank two 0.7-kb fragments on the left. However, the precise position could not be identified in the present analysis. For convenience, the *Hin*dIII site in the transcriptional unit is eliminated in this map, and the other possible site is indicated by an asterisk. Line C has a 0.3-kb insertion in a 3.9-kb fragment. (*D*) *Pst* I digests on a 1.2% agarose gel. A polymorphic site was found in the transcriptional unit (lines H and E) as well as in *Hin*dIII digests.

cations among them. The distribution of triplications appears to be restricted to local populations at present. These findings suggest relatively recent ancestry of the triplications and their independent origins in respective populations. Of course, we cannot completely exclude, with the present data, other explanations such as founder effects. Multiple independent occurrences of the triplications may be supported indirectly from the following observations: the duplication rate is

Table 1. Polymorphic restriction sites in the transcriptional unit and α-GPDH allozyme [fast (F) or slow (S)] for 86 chromosome 2 lines from the Ogasawara and the Aomori population

Class	Allozyme	Restriction sites *Pst* I (−2.2)	*Xho* I (−2.0)	*Sac* I (−0.2)	*Hin*dIII (0.0)	Frequency Ogasawara	Aomori
Standard	F	+	+	+	+	6	13
	F	+	+	+	−	10	0
	F	−	+	+	+	5	3
	F	−	+	+	−	0	4
	F	−	ns	+	−	4	0
	F	+	+	−	+	0	2
	F	+	−	+	+	0	1
	F	−	+	−	+	1	0
	S	+	+	+	+	0	6
	S	−	+	+	+	1	1
Duplication	F	+	+	+	+	1	6
	F	+	+	+	−*	4	1
	S	+	+	+	+	5	2
	S	−	+	+	−*	0	2
Triplication	F	+	+	+	+	6	0
	S	+	+	+	+	0	2

The presence or absence of a restriction site is indicated by + or −, respectively. For some lines, the absence of *Xho* I (−2.0) site could not be ascertained because of a deletion. This is indicated by ns (not scored). Two of six Ogasawara triplication chromosomes showed a slightly slower band than common F alleles, but this finding was ignored in the present analysis.
*The precise position of the polymorphic *Hin*dIII site could not be identified in the duplication class (see the text and Fig. 2*C*).

estimated to be of the order of 10^{-6} for the maroon-like locus (10) and 10^{-4} to 10^{-5} for the rosy locus (9, 10) in *D. melanogaster*. Because of tandemly repeated 4.5-kb homologous sequences in the *Gpdh* region, it is natural to assume that the duplication rate in this region is greater than in the above estimates.

In addition to restriction site polymorphisms, sequence-length variation other than duplications was also observed over most of the 26-kb region studied. This type of variation, together with restriction site polymorphisms, will be reported in detail in our following paper.

Neutral Evolution of *Gpdh* Duplications and Duplication-Deletion Through Unequal Crossing-Over As a Source of Genetic Variation. As mentioned above, the duplication contains all but the 5′-end of *Gpdh* (S.K., H. Baba, A.K., G. C. Bewley, and T.M., unpublished data). Thus, the duplications and triplications appear to have no selective advantage. In addition, the shared restriction site polymorphisms within both the standard and the duplication-bearing chromosomes suggest recurrent recombination in this region including the transcriptional unit. This finding excludes the possibility of a hitch-hike effect on *Gpdh* duplications caused by other linked selective genes for the increase in their frequency, although this might have acted at the early stage of the spreading. Also, the observations of widespread duplications in natural populations (15) suggest the lack of negative selection against this type of polymorphism. It may be concluded that random genetic drift is responsible for the spreading of this duplication (29, 30). In other words, variation in gene copy number, at least as much as that observed in the present experiments, can be brought about through continuous occurrence of unequal crossing-over and random genetic drift. In addition, once two tandem gene copies are available in a region, additional gene copies appear to accumulate more frequently through unequal crossing-over. This may provide a genetic mechanism for the formation of multigene families (31). Indeed, the present phenomenon is only a transient phase of evolution, and selection pressure is essential for the retention of gene function (31). It should be noted that further duplications and gene shuffling may result in the acquirement of two complete copies of *Gpdh* and/or internally duplicated gene with new function (e.g., see refs. 32 and 33).

As mentioned earlier, unequal crossing-over occurs fairly frequently. Gene duplication and deletion by unequal crossing-over can be an important source of genetic variation. From the standpoint of evolutionary and population genetics, Ohta (2, 31, 34) has advocated the significance of gene duplication for the progressive evolution of higher organisms. All of the above new findings must be incorporated in the study of genetic variation and its maintenance mechanisms (34).

Table 2. Means ± SEM of restriction site differences

Class	Within population Ogasawara	Aomori	Between populations*
Standard	4.62 ± 0.40	4.38 ± 0.44	4.80 ± 0.55 (0.29 ± 0.73)
Duplication	2.02 ± 0.36	3.20 ± 0.32	4.25 ± 0.93 (1.63 ± 1.15)
Triplication	2.67 ± 0.86	0.00 ± 0.00	5.00 ± 3.11 (3.67 ± 3.17)

Estimation is based on the data set of 54 restriction sites in the 26-kb region and α-GPDH allozyme variation. SEMs given are only due to sampling, and they do not include a stochastic component.
*Values in parentheses are the number of net restriction site differences (see the text).

We are grateful to Prof. Glenn C. Bewley for kindly providing the *Gpdh* clone *pDm60a(c)* and for improving the manuscript. This work was supported by research grants from the Ministry of Education, Science, and Culture of Japan. This is paper no. 53 from the Laboratory of Population Genetics, Department of Biology, Kyushu University.

1. Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, Berlin).
2. Ohta, T. (1980) *Lecture Notes in Biomathematics* (Springer, Berlin), Vol. 37, pp. 1–131.
3. Sturtevant, A. H. (1925) *Genetics* **10**, 117–147.
4. Batterham, P., Chambers, G. K., Starmer, W. T. & Sullivan, D. T. (1984) *Evolution* **38**, 644–657.

5. Schaeffer, S. W. & Aquadro, C. F. (1987) *Genetics* **117,** 61–73.
6. Chia, W., Savakis, C., Karp, R., Pelhman, H. & Ashburner, M. (1985) *J. Mol. Biol.* **186,** 679–688.
7. Gemmill, R. M., Levy, J. N. & Doane, W. W. (1985) *Genetics* **110,** 299–312.
8. Levy, J. N., Gemmill, R. M. & Doane, W. W. (1985) *Genetics* **110,** 313–324.
9. Gelbart, W. M. & Chovnick, A. (1979) *Genetics* **92,** 849–859.
10. Shapira, S. K. & Finnerty, V. G. (1986) *J. Mol. Evol.* **23,** 159–167.
11. Goldberg, M. L., Sheen, J., Gehring, W. J. & Green, M. M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5017–5021.
12. Davis, P. S., Shen, M. W. & Judd, B. H. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 174–178.
13. Otto, E., Young, J. E. & Maroni, G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 6025–6029.
14. Maroni, G., Wise, J., Young, J. E. & Otto, E. (1987) *Genetics* **117,** 739–744.
15. Koga, A., Kusakabe, S., Tajima, F., Harada, K., Bewley, G. C. & Mukai, T. (1988) *Proc. Jpn. Acad.* **64,** 9–12.
16. Cook, J. L., Shaffer, J. B., Bewley, G. C., MacIntyre, R. J. & Wright, D. A. (1986) *J. Biol. Chem.* **261,** 11751–11755.
17. Cook, J. L., Bewley, G. C. & Shaffer, J. B. (1988) *J. Biol. Chem.* **263,** 10858–10864.
18. Takano, T., Kusakabe, S. & Mukai, T. (1987) *Genetics* **117,** 245–254.
19. Wallace, B. (1956) *J. Genet.* **54,** 280–293.
20. Loenen, W. A. M. & Blattner, F. R. (1983) *Gene* **26,** 171–179.
21. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
22. Bender, W., Spierer, P. & Hogness, D. S. (1983) *J. Mol. Biol.* **168,** 17–33.
23. Kreitman, M. & Aguade, M. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3562–3566.
24. Southern, E. M. (1975) *J. Mol. Biol.* **98,** 503–517.
25. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
26. Feinberg, A. P. & Vogelstein, B. (1984) *Anal. Biochem.* **137,** 266–267.
27. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
28. Nei, M. & Tajima, F. (1981) *Genetics* **97,** 145–163.
29. Kimura, M. (1968) *Nature (London)* **217,** 624–626.
30. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, London).
31. Ohta, T. (1987) *Genetics* **115,** 207–213.
32. Kitamura, N., Kitagawa, H., Fukushima, D., Takagaki, Y., Miyata, T. & Nakanishi, S. (1985) *J. Biol. Chem.* **260,** 8610–8617.
33. McLean, J. W., Tomlinson, J. E., Kuang, W., Eaton, D. L., Chen, E. Y., Fless, G. M., Scanu, A. M. & Lawn, R. M. (1987) *Nature (London)* **300,** 132–137.
34. Ohta, T. (1983) *Theor. Pop. Biol.* **23,** 216–240.