Copyright © 1997 by the Genetics Society of America

DNA Polymorphism at the *Pgi* Locus of a Wild Yam, *Dioscorea tokoro*

Ryohei Terauchi,*,† Toru Terachi‡ and Naohiko T. Miyashita§

**Laboratory of Plant Systematics, Department of Botany and §Laboratory of Genetics, Faculty of Agriculture, Kyoto University, Kyoto 606-01, Japan, †Plant Molecular Biology Group, Biocenter, University of Frankfurt, Frankfurt am Main D-60439, Germany and ‡Department of Biotechnology, Kyoto Sangyo University, Kyoto 603, Japan*

Manuscript received December 5, 1996
Accepted for publication August 27, 1997

ABSTRACT

To study the origin and maintenance mechanisms of the PGI allozyme polymorphism of a wild plant, *Dioscorea tokoro*, DNA sequences of the entire coding region (1701 bp) and two intronic regions (total 2049 bp) of the *Pgi* gene as well as a part of the *Adh* gene (590 bp) were analyzed. Two replacement substitutions were revealed to be responsible for the differentiation of three allozymes alleles (*Pgi-a*, *Pgi-b* and *Pgi-c*) that occur in natural population in intermediate frequencies. Interspecific comparison of DNA sequences identified *Pgi-b* as the oldest allele, from which two other alleles were derived probably within the last 150,000 years. The level of DNA polymorphism at *D. tokoro Pgi* locus was low. No elevated level of DNA polymorphism was detected in the close vicinity of the two replacement sites differentiating the three allozymes. Departures from the neutral mutation hypothesis were detected by Fu and Li's and MK tests. The observed patterns of DNA polymorphism are explainable by both (1) the neutral mutation hypothesis with an assumption of small effective size of *D. tokoro* population, and (2) the positive selection hypothesis that the allele frequencies of *Pgi-a* and *Pgi-c* have increased in a short time by their selective advantages.

PHOSPHOGLUCOSE isomerase (PGI: E.C. 5.3.1.9) is a dimeric enzyme that catalyzes the reversible isomerization of glucose-6-phosphate and fructose-6-phosphate at the second step of the glycolytic pathway. PGI has a high catalytic activity and the reaction is thought to be in equilibrium (NOLTMANN 1972). It is believed that PGI provides a coupling function in glycolysis and gluconeogenesis at the glucose-6-phosphate metabolic branch point. In contrast to other glycolytic enzymes, no regulatory function mediated by the allosteric modulation was suggested for PGI. From these observations, it is proposed that the total flux of glycolysis may be determined by the coupling effectiveness of the PGI enzymes (WATT 1977; GILLESPIE 1991).

PGI has been shown to be highly polymorphic in many organisms in terms of electrophoretic mobility (HALL 1985; RIDDOCH 1993). Possible roles of natural selection in the maintenance of PGI polymorphism have been frequently indicated (GILLESPIE 1991). In most cases, temperature-related differences in kinetics or stability among PGI allozymes are thought to be subject to natural selection. For example, overdominance at the *Pgi* locus under certain temperature conditions was indicated for a butterfly, Colias (WATT 1977), and an amphipod, Gammarus (PATARNELLO and BATTAGLIA 1992). Directional selection on a particular allele under high-temperature conditions was suggested in a sea anemone, Metridium (HOFFMANN 1981), and a bivalve, *Mytilus edulis* (HALL 1985), both of which show latitudinal clines in the *Pgi* allele frequency. In a reviewing article of PGI allozyme studies on various organisms, RIDDOCH (1993) reported an association between the relative electrophoretic mobility of PGI enzymes and environmental conditions, such that the more anodal allozyme/isozyme is favored under hot, dry or low-oxygen environment.

Plants have at least two sets of PGI, one in the cytosol and the other in the chloroplast (GOTTLIEB 1982). DNA sequences for the plant cytosolic PGI have been determined in *Clarkia lewisii* (THOMAS *et al.* 1992), *Arabidopsis thaliana* (THOMAS *et al.* 1993), *Zea mays* (LAL and SACHS 1995) and *Oryza sativa* (NOZUE *et al.* 1996). Clarkia *Pgi* has 23 exons, spanning over a 6-kb region, the largest number of exons in plant genes ever reported. Its exons (1707 bp) coded for 569 amino acids. Despite the general interest in the mechanisms maintaining polymorphisms at the *Pgi* locus, so far only one plant species, *C. lewisii* (THOMAS *et al.* 1993) and three Drosophila species (*Drosophila melanogaster*, *D. simulans*, and *D. Yakuba*: J. MCDONALD and M. KREITMAN, unpublished data; MORIYAMA and POWELL 1996) have been studied for their DNA polymorphisms at the *Pgi* locus.

Dioscorea tokoro Makino is a dioecious, diploid ($2n = 2x = 20$) plant species that belongs to the family Dioscoreaceae (the family of yam) of monocotyledons. It is distributed in temperate East Asia including the entire Japan Islands. In the course of an allozyme study of natural populations of *D. tokoro*, we observed a high

Corresponding author: Ryohei Terauchi, AK Kahl, Plant Molecular Biology Group, Biocenter, University of Frankfurt, Marie-Curie-Str. 9/N200, D-60439 Frankfurt am Main, Germany.
E-mail: terauchi@em.uni-frankfurt.de

level of genetic variation at the locus encoding cytosolic PGI (Terauchi 1990). In central Japan, where 1128 individuals were sampled from 24 subpopulations, three major alleles (*Pgi-a*, *Pgi-b* and *Pgi-c*) occurred at intermediate frequencies (0.31, 0.35 and 0.34, respectively). Watterson's homozygosity test (Watterson 1978) applied to the allele frequencies rejected the neutral mutation hypothesis at the locus ($F = 0.327$, $P < 0.01$), indicating too many alleles with high frequencies. The total gene diversity (H_T: Nei 1973) was 0.67, the maximum value possible for three alleles, and gene diversity within subpopulations (H_S) was 0.59, indicating that most of the gene diversity of the *Pgi* locus resides within each subpopulation. However, there was no indication of heterozygote excess compared to the Hardy-Weinberg expectations in any subpopulations. Within the rather restricted geographical area where the study was carried out, allele frequencies of *Pgi-a* and *Pgi-c* showed latitudinal clines, such that the frequency of the most anodal allele *Pgi-a* increased toward the north and that of *Pgi-c* increased toward the south. This pattern is opposite to what is reported in Riddoch (1993). We have no data whether these latitudinal clines are present over the entire geographical range of the species distribution.

We are interested in (1) how and when these allozyme polymorphisms at *Pgi* came into exist in *D. tokoro* populations, and (2) why these polymorphisms are maintained in the population. To address these questions, we have been studying DNA polymorphism at the locus. In the previous article, Terauchi *et al.* (1996) made a preliminary report on the DNA polymorphism among cDNAs of *D. tokoro Pgi*. There, they described (1) the two amino acid replacements possibly responsible for the differentiation of the three allozyme alleles, (2) the identification of *Pgi-b* allele as the oldest allele among the three allelic lineages, (3) significantly larger intraspecific replacement polymorphism as revealed by McDonald and Kreitman's test (McDonald and Kreitman 1991), and (4) the low nucleotide diversity in the *Pgi* coding region. However, no data accompanied the article to substantiate their description.

In the present article, we supplement Terauchi *et al.* (1996) by giving the data on *D. tokoro Pgi* DNA polymorphism for the first time and report new findings obtained by the studies on *Pgi* introns as well as *Adh* gene of *D. tokoro* and a closely related species *D. tenuipes*. We analyze the observed patterns of DNA polymorphism by the statistical methods of molecular population genetics to see whether they are compatible with the neutral mutation hypothesis (Kimura 1983).

MATERIALS AND METHODS

Plant materials: For the study of DNA polymorphisms, all the plant materials except one (DT6536) were collected from a 200 × 200 km area of the Kinki region in the center of Honshu Island of Japan (Table 1). Judging from the geographical distribution of alleles at six allozyme loci, there was no apparent population subdivision within the area (Terauchi 1990). For the sequencing analysis of the *Pgi* coding region (cDNA), 10 individuals of *D. tokoro* with the known *Pgi* allozyme genotypes were used so that the frequencies of *Pgi-a*, *Pgi-b* and *Pgi-c* alleles in the sample (0.25, 0.35 and 0.40, respectively) were close to their frequencies in natural populations (0.31, 0.34 and 0.34, respectively). Although it is not a random sample, we consider this sample as reflective of the pattern of polymorphism in natural populations (see Hudson *et al.* 1994 for a similar sampling method). Additionally, genomic DNA containing a region between intron 4 and intron 10 was sequenced for 15 DNA clones obtained from 13 individuals of the known *Pgi* allozyme genotypes. The frequency of each allozyme allele in the sample (0.27, 0.4 and 0.33 for *Pgi-a*, *Pgi-b* and *Pgi-c*, respectively) is close to that in natural populations. The *Adh* sequences were obtained from eight individuals of *D. tokoro*. Plant materials used for the analyses of the *Pgi* coding region (cDNA), *Pgi* intronic regions (genomic DNA), and the *Adh* region are not identical. However, the samples for all the analyses came from the same geographical regions, so that we regard all these three subsets of samples represent the entire population. For the interspecific comparison, DNA sequences of an individual of a closely related Dioscorea species, *D. tenuipes* (DTE1) were studied. One individual of another related species *D. quinqueloba* (DQ1), was additionally studied for the analysis of the *Pgi* coding region.

Protein electrophoresis: The *Pgi* genotype of *D. tokoro* individuals were determined by starch gel electrophoresis and histochemical staining of the PGI enzyme as described (Terauchi 1990). As convention, the allele coding for the anodally fastest migrating band is named *Pgi-a*, followed by *Pgi-b* and *Pgi-c*.

Isolation of *Pgi* cDNA from *D. tokoro*: Total RNA was extracted from leaves of a single plant of *D. tokoro* (DT37; homozygote for *Pgi-b*) using the method of Chomczynski and Sacchi (1987). First-strand cDNA was synthesized from the total RNA using a kit (SuperScript Preamplification System, GIBCO BRL). This cDNA was used as template for PCR amplification of a partial *Pgi* coding region (750 bp) using degenerate primers (93.4: 5′-TTYGCNTTYTGGGAYTGGGT-3′ and 93.9H: 5′-TCIACICCCCAITGRTCTAAIGARTTIAT-3′, Figure 1) that were designed according to the conserved regions of the published *Pgi* sequences of *C. lewisii* (Thomas *et al.* 1992), *A. thaliana* (Thomas *et al.* 1993), pig (Chaput *et al.* 1988) and yeast (Green *et al.* 1988). Newly synthesized primers in this amplified region were further used to determine the 5′ and 3′ ends of the cDNA using the RACE technique (Frohman *et al.* 1988). On the basis of sequence information at the 5′ and 3′ ends of *D. tokoro Pgi* cDNA, a primer pair (94.22: 5′-GATCGTCCTTGCCAAACGAAATC-3′ and 94.24: 5′-TTGTTACTGAAAAACTGGATGAA-3′, Figure 1) was synthesized to amplify a fragment (1990 bp) containing the entire coding region of *D. tokoro Pgi*.

DNA polymorphism in the *Pgi* coding region: DNA sequences for the entire coding region were obtained by the analysis of cDNA. Total RNAs extracted from the 10 individuals in Table 1 were reverse-transcribed, and the obtained cDNA was used as template for PCR amplification of *Pgi* by the primers (94.22 and 94.24) as described above. PCR products were purified by agarose gel electrophoresis followed by glass powder elution (GeneClean II, BIO101) and used as templates for direct-sequencing by an autosequencer (ABI 373A). Both strands were sequenced using primers designed in *ca.* 200-bp intervals. Heterozygous positions in an individual were identified as those sites where two different nucleotide signals appeared in the sequences of both DNA strands. The

TABLE 1

Plant materials used for the DNA sequencing analysis of *Pgi* and *Adh* in Dioscorea

Analyzed region Plant code	Species	Allozyme genotype	Collection site	DNA clone code[a]
Pgi				
Coding region (cDNA)				
DT1	*D. tokoro*	*b/c*	Hirase, Wakayama	—
DT12	*D. tokoro*	*a/a*	Wakayama	—
DT27	*D. tokoro*	*a/a*	Kamiyu, Wakayama	—
DT33	*D. tokoro*	*c/c*	Kushimoto, Wakayama	—
DT34	*D. tokoro*	*a/c*	Ohara, Kyoto	—
DT37	*D. tokoro*	*b/b*	Wakayama	—
DT46	*D. tokoro*	*b/b*	Daihisen, Kyoto	—
DT48	*D. tokoro*	*b/b*	Ohara, Kyoto	—
DT51	*D. tokoro*	*c/c*	Horao, Wakayama	—
DT56	*D. tokoro*	*c/c'*	Nachi, Wakayama	—
DTE1	*D. tenuipes*	—	Takinohai, Wakayama	—
DQ1	*D. quinqueloba*	—	Chikatsuyu, Wakayama	—
Intron 4 to intron 10 (genomic DNA)				
DT1	*D. tokoro*	*b/c*	Hirase, Wakayama	DT1-7 (*Pgi-b*)
DT5	*D. tokoro*	*b/c*	Kuki, Wakayama	DT5-15 (*Pgi-b*)
DT7	*D. tokoro*	*b/c*	Ichikano, Wakayama	DT7-1-10 (*Pgi-b*)
DT12	*D. tokoro*	*a/a*	Wakayama	DT12-1 (*Pgi-a*)
DT18	*D. tokoro*	*c/c*	Ookuwa, Wakayama	DT18-6 (*Pgi-c*)
DT24	*D. tokoro*	*b/c*	Ohara, Kyoto	DT24-2 (*Pgi-c*)
DT25	*D. tokoro*	*a/a*	Kamiyu, Wakayama	DT25-4 (*Pgi-a*)
DT34	*D. tokoro*	*a/c*	Ohara, Kyoto	DT34-2 (*Pgi-a*)
DT37	*D. tokoro*	*b/b*	Wakayama	DT37-5/5 (*Pgi-b*)
DT49	*D. tokoro*	*c/c*	Tahara, Wakayama	DT49-3 (*Pgi-c*)
DT51	*D. tokoro*	*c/c*	Horao, Wakayama	DT51-9 (*Pgi-c*)
DT54	*D. tokoro*	*a/c*	Horao, Wakayama	DT54-3 (*Pgi-a*) DT54-10 (*Pgi-c*)
DT6536	*D. tokoro*	*b/b*	Iide, Niigata	DT6536-4 (*Pgi-b*) DT6536-6 (*Pgi-b*)
DTE1	*D. tenuipes*	—	Takinohai, Wakayama	DTE1-Pgi
Adh				
DT1	*D. tokoro*	—	Hirase, Wakayama	—
DT8	*D. tokoro*	—	Wakayama	—
DT12	*D. tokoro*	—	Wakayama	—
DT27	*D. tokoro*	—	Kamiyu, Wakayama	—
DT49	*D. tokoro*	—	Tahara, Wakayama	—
DT51	*D. tokoro*	—	Horao, Wakayama	—
DT54	*D. tokoro*	—	Horao, Wakayama	—
DT56	*D. tokoro*	—	Nachi, Wakayama	—
DTE1	*D. tenuipes*	—	Takinohai, Wakayama	DTE1-Adh

[a] *Pgi* allozyme type in parentheses.

Pgi cDNA (1331 bp) of *D. tenuipes* and *D. quinqueloba* were amplified by PCR using a primer pair, 95.1 (5′-AACTTGCTG AGGTGGCTTG-3′) and 93.9H (Figure 1) and sequenced with the same primers as used for *D. tokoro*. The nucleotide sequences are deposited in DDBJ, EMBL and GenBank DNA databases under the accession numbers, D88921–D88930 (*D. tokoro*), D88920 (*D. tenuipes*) and AB006088 (*D. quinqueloba*).

DNA polymorphism in the *Pgi* intron sequences: To obtain the sequence information of introns, a part (5.5 kb) of the *Pgi* gene containing the region between exon 3 and exon 13 was cloned. The analysis of the entire *Pgi* gene was difficult because of its large size (> ~9 kb, data not shown) in *D. tokoro*. Total DNAs were extracted from the 13 individuals of *D. tokoro* and one individual of *D. tenuipes* listed in Table 1 using CTAB (Terauchi and Konuma 1994). These DNAs served as templates for PCR amplification of a 5.5-kb fragment using a primer pair, INT-1 (5′-GCTTGTCTAAAGCAGAAG 3′) and 95.5 (5′-AATGGAGTGGAATGGAAATG-3′, Figure 1), and a DNA polymerase with a high fidelity (Expand Taq, Boehringer Mannheim). PCR products were purified by agarose gel electrophoresis and glass powder elution. PCR products were blunt-ended by treating with Klenow enzyme (NEB) and T4 polynucleotide kinase (NEB) in the presence of dNTPs and ATP and ligated into pUC19, which was cleaved with *Sma*I and dephosphorylated. The ligates were used for the transformation of competent cells of *Escherichia coli* strain DH5α. Two DNA regions corresponding to introns 4–6 and introns 9–10 were sequenced for both the strands using prim-

```
                                               gatcgtccttgccaaacgaaatcgatccgcaggcctggaaacc
                                               --------------------->
                                                  primer 94.22

     Met Ala Thr Ser Thr Leu Ile Cys Glu Thr Pro Gln Trp Lys Asp Leu Asn Asp His Val     20
   1 ATG GCT ACG TCC ACG CTT ATC TGC GAA ACG CCG CAG TGG AAG GAC CTC AAT GAC CAT GTC

     Glu Glu Ile Lys Lys Thr His Leu Arg Asp Leu Met Gln Asp Ser Asp Arg Cys Lys Ser     40
  61 GAA GAA ATC AAG AAG ACG CAT CTC CGC GAT CTC ATG CAG GAC TCC GAC CGC TGC AAG TCC

     Met Ile Thr Glu Phe Asp Gly Ile Ile Leu Asp Tyr Ser Arg Gln Arg Val Leu Pro Ala     60
 121 ATG ATC ACG GAG TTT GAT GGG ATC ATA TTG GAC TAT TCC CGG CAA CGA GTG CTT CCT GCC

     Thr Val Glu Lys Leu Phe Gln Leu Ala Glu Val Ala Cys Leu Lys Gln Lys Ile Asp Arg     80
 181 ACT GTG GAG AAG CTT TTC CAA CTT GCT GAG GTG GCT TGT CTA AAG CAG AAG ATC GAT AGG
D.te     primer 94.20<-----------------      ----------|. ..C ... ... ...-> primer INT-1
D.q               primer 95.1 -------------------------->|. ... ... ... ... ... ... ...
     Met Tyr Asn Gly Glu Lys Ile Asn Cys Thr Glu Asn Arg Ser Val Leu [His] Ile Ala Leu   100
 242 ATG TAC AAT GGA GAG AAG ATT AAT TGC ACA GAG AAT AGG TCT GTG TTG [CAT] ATA GCT CTT
D.te ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ...
     Arg Ala Ala Arg Asp Lys Ala Ile Lys Ser Asp Gly Lys Asn Val Val Pro Asp Val Trp    120
 301 CGA GCT GCA AGA GAT AAG GCC ATC AAA AGT GAT GGA AAG AAT GTG GTA CCA GAC GTG TGG
D.te ... ... ... ... ... ... .T(Val) ... ... ... ... ... ... ... ... ... ..T ... ...
D.q  ... ... ... ... ... ... .T(Val) G(Glu). ... ... ... ... ... ... ... ... ..T ...
     His Val Leu Asp Lys Ile Lys Glu Phe Ser Glu Arg Ile Arg Ser Gly Ser Trp Val Gly    140
 361 CAT GTC CTA GAC AAA ATC AAG GAG TTC TCA GAG AGA ATT CGT AGT GGC TCT TGG GTT GGA
D.te ... .C(Ala) ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
     Ala Thr Gly Lys Ala Leu Thr Asp Val Val Ala Val Gly Ile Gly Gly Ser Phe Leu Gly    160
 421 GCA ACT GGG AAG GCA TTG ACG GAT GTT GTG GCR GTA GGC ATC GGG GGT AGT TTT TTA GGA
D.te ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ...
D.q  ... ..C ... ..A ... ... ... ... ... ... ..A ..G ... ..T ... ... ... ... ... ...
     Pro Leu Phe Val His Thr Ala Leu Gln Thr Asp Pro Glu Ala Ala Glu Cys Ala Lys Gly    180
 481 CCT TTA TTT GTG CAT ACT GCT CTT CAA ACA GAT CCA GAG GCA GCA GAA TGT GCT AAA GGA
D.te ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ..C ... ... ...
D.q  ... ... ... ... ... ..C ... ... ... ..T ... ... ... ... ... ... ..C ... ... ...
     Arg Gln Leu Arg Phe Leu Ala Asn Val Asp Pro Ile Asp Val Ala Arg Ser Ile Thr Gly    200
 541 CGG CAA TTG CGA TTC CTT GCA AAT GTT GAT CCA ATT GAT GTT GCT CGA AGC ATC ACT GGT
D.te ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... G(Ala).
D.q  ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... G(Ala).
     Leu Asn Pro Glu Thr Thr Leu Val Val Val Val Ser Lys Thr Phe Thr Thr Ala Glu Thr    220
 601 TTG AAT CCT GAA ACC ACA TTA GTT GTG GTT GTT TCA AAG ACT TTT ACG ACT GCT GAA ACT
D.te ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ...
     Met Leu Asn Ala Arg Thr Leu Arg Glu Trp Ile Ser Ala Ala Leu Gly Pro Gln Ala Val    240
 661 ATG CTG AAT GCT AGA ACA CTC AGA GAA TGG ATT TCA GCT GCT CTT GGG CCC CAG GCA GTT
D.te ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... ... ... .G(Gly) ... ... ... ... ... ...
     Ser Lys His Met Val Ala Val Ser Thr Asn Leu Thr Leu Val Glu Lys Phe Gly Ile Asp    260
 721 TCA AAA CAT ATG GTT GCT GTC AGC ACA AAT CTT ACG CTT GTA GAG AAG TTT GGT ATT GAC
D.te G(Ala). ... ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ..A ... ...
D.q  G(Ala). ... ... ... ... ... ..T ... ..C ... ... ... ... ... ... ... ... ... ...
     Pro Ala Asn Ala Phe Ala Phe Trp Asp Trp Val Gly Gly Arg Tyr Ser Val Cys Ser Ala    280
 781 CCT GCT AAT GCT TTT GCA TTC TGG GAC TGG GTA GGA GGC CGA TAT AGT GTT TGC AGT GCA
D.te ... ...primer 93.4------------------------>... ... ... ... ... ... ... ... ... ..G
D.q  ... ... ..C ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ...
     Val Gly Val Leu Pro Leu Ser Leu Gln Tyr Gly Phe Ser Val Val His Lys Phe Leu Asn    300
 841 GTT GGT GTG CTT CCT CTG TCG CTC CAA TAT GGT TTT TCA GTT GTT CAT AAG TTC CTG AAT
D.te ... ... ... ... ... T.. ..T ... ... ... ... ... ... ... ... ..A ... ... ... ...
D.q  ... ... ... ... ... T.. ..T ... ... ... ... ... ... ... ... ..A ... ... ... ..A(Lys)
     Gly Ala Ala Ser Ile Asp Asp His Phe His Ser Thr Pro Phe Glu Lys Asn Ile Pro Val    320
 901 GGA GCT GCA AGC ATT GAT GAC CAT TTC CAT TCC ACT CCA TTT GAG AAA AAT ATT CCT GTA
D.te ..G ..A ... ... ... ... ... <-------------------------- primer 95.5 ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ...
```

```
     Leu Leu Gly Leu Leu Ser Val Trp Asn Val Ser Phe Leu Arg Tyr Pro Ala Arg Ala Ile    340
 961 CTT TTA GGT CTA TTG AGT GTA TGG AAT GTT TCG TTT CTT CGG TAT CCT GCA AGA GCT ATA
D.te ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... ... ... ... G(Gly). ... ... ... ... ... ...
     Leu Pro Tyr Ser Gln Ala Leu Glu Lys Phe Ala Pro His Ile Gln Gln Val Ser Met Glu    360
1021 TTA CCT TAT TCT CAA GCA CTT GAG AAA TTT GCC CCA CAT ATT CAA CAA GTT AGC ATG GAG
D.te ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ...
     Ser Asn Gly Lys Gly Val Ser Ile Asp Gly Val Pro Leu Pro Phe Glu Thr Gly Glu Ile    380
1081 AGT AAT GGG AAG GGT GTC TCA ATT GAT GGC GTC CCT CTT CCA TTT GAG ACT GGT GAA ATT
D.te ... ... ... ... ... ... A(Thr). ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... A(Ile). ... ... ... ... ... ... ... ... ...
     Asp Phe Gly Glu Pro Gly Thr Asn Gly Gln His Ser Phe Tyr Gln Leu Ile His Gln Gly    400
1141 GAT TTT GGG GAG CCT GGA ACA AAT GGT CAG CAT AGC TTC TAT CAA CTT ATT CAC CAG GGA
D.te ... ... ... ... T(Ser). ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
     Arg Val Ile Pro Cys Asp Phe Ile Gly Val Met Arg Ser Gln Gln Pro Ile Tyr Leu Lys    420
1201 AGG GTT ATT CCT TGT GAT TTT ATT GGT GTA ATG AGG AGC CAG CAA CCA ATT TAC TTG AAA
D.te ... ... ... ... ... ... ... ... ... A(Ile). ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
     Gly Glu Val Val Ser Asn His Asp Glu Leu Met Ser Asn Phe Phe Ala Gln Pro Asp Ala    440
1261 GGA GAA GTT GTA AGC AAC CAT GAT GAG CTT ATG TCA AAT TTC TTT GCG CAG CCA GAT GCC
D.te ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
     Leu Ala Tyr Gly Lys Val Ala Glu Gln Leu Leu Asn Glu Lys Val Pro Asp His Leu Ile    460
1321 CTT GCC TAT GGA AAG GTT GCT GAA CAA TTG CTT AAT GAG AAA GTC CCT GAC CAT CTT ATT
D.te ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ..C(Asp)... ... ... ... ... ... ... ... ... ... ...
     Pro His Lys Thr Phe Pro Gly Asn Arg Pro Ser Leu Ser Leu Leu Leu Pro Ser Leu Asn    480
1381 CCT CAC AAG ACC TTT CCG GGC AAT CGA CCA TCA TTG AGT CTT CTA TTA CCT TCA TTA AAT
D.te ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ...
     Ala Tyr Asn Val Gly Gln Leu Leu Ala Ile Tyr Glu His Arg Ile Ala Val Glu Gly Phe    500
1441 GCT TAC AAC GTT GGA CAG TTA CTA GCA ATA TAT GAG CAC AGG ATT GCT GTT GAA GGC TTT
D.te ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...
D.q  ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ...
     Val Trp Gly Ile Asn Ser Phe Asp Gln Trp Gly Val Glu Leu Gly Lys [Ser] Leu Ala Ser    520
1501 GTA TGG GGG ATC AAT TCT TTT GAT CAA TGG GGA GTA GAG TTG GGC AAG [TCT] CTG GCT TCA
D.te ... ... ..|<------------------------------------------
D.q  ... ... ..|          primer 93.9H
     Gln Val Arg Lys Gln Leu His Leu Ser Arg Thr Lys Ala Glu Pro Val Glu Gly Phe Asn    540
1561 CAA GTT AGG AAG CAA CTG CAT CTA TCC CGT ACA AAA GCT GAG CCT GTC GAG GGA TTT AAC

     Phe Ser Thr Thr Thr Leu Leu Thr Arg Tyr Leu Glu Ala Glu Thr Gly Val Pro Ser Asp    560
1621 TTC AGC ACC ACC ACC TTG TTA ACA AGA TAT CTA GAG GCA GAA ACA GGG GTT CCA TCT GAT

     Gln Ser Gln Leu Pro Lys Leu Stop                                                    567
1681 CAG AGT CAA CTA CCA AAA CTT TAG tctaccaaaacaccatgagcgaaaaacatcgtttcatccagttttc
                                                                      <-----------
                                                                      primer 94.24
     agtaacaaaataaagctcatttagagttcaatgtttatgagcttgcatttctgtcacagaagcatgggtttgaccttt
     ----

     tctaaagaaaattgtaataactgttttggttctgtcctctggtagtttaagacgtccaggacggttgtggaaaaatatg

     aaaaagagcaattgttggtgagctatctgta
```

FIGURE 1.—Nucleotide sequences of *Pgi* cDNA and a deduced amino acid sequence of Dioscorea: the entire coding region for *D. tokoro* (DT37) and the partial coding region (site 219 to site 1509) for *D. tenuipes* (D. te) and *D. quinqueloba* (D. q). Amino acid and nucleotide sites with bold face letters indicate the polymorphic sites in *D. tokoro*. Boxed amino acid residues are possible catalytic sites (ACHARI *et al.* 1981). Positions where *D. tenuipes* and *D. quinqueloba* share the same nucleotides as *D. tokoro* are indicated by dots. DNA primers used for PCR are indicated by arrows.

ers located in *ca.* 200-bp intervals. The nucleotide sequences are deposited in DDBJ, EMBL and GenBank DNA databases under the accession numbers, D88931–D88975 (*D. tokoro*) and AB006003–AB006004 (*D. tenuipes*).

PCR-based identification of the nucleotides at sites 335/6 and 713: Sequencing analysis of the *Pgi* cDNA suggested a possibility that the two polymorphic sites, 335 and 713, are responsible for the differentiation of the three allozymes, *Pgi-a*, *Pgi-b* and *Pgi-c* (see RESULTS). To further confirm this result, we tested the association between particular nucleotides at these sites and allozyme types among the individuals randomly chosen from a natural population. *Pgi* allozyme genotypes were determined by protein electrophoresis. PCR method was used to determine the nucleotides at sites 335 and 713. For the identification of nucleotides at sites 335/6, we prepared two allele-specific oligonucleotides, ASO1 (5′-ATCAAAAGTGATGGA-3′) and ASO2 (5′-ATCAAAAGTGAT GAT-3′), which have GA and AT at their 3′ ends, respectively. In combination with an antisense primer, AS1 (5′-CCTTGA TTTTGTCTAGGA-3′), ASO1 amplified a fragment if the template sequence had GA at sites 335/6, whereas no amplification occurred with the primer ASO2. The situation was *vice versa* if the template sequence had AT at these sites. A region containing the sites was first amplified by the primer INT-1 (5′-GCTTGTCTAAAGCAGAAG-3′) and AS1 using the total genomic DNA as templates. This primary PCR products served as templates for the secondary PCR using the primer AS1 and each of the allele specific primers to see whether amplification occurred or not. To identify the site 713, we combined PCR and restriction enzyme digestion. The site 713 resides in a sequence context 5′-CC(A/G)GG-3′. The sequence 5′-CC AGG-3′ is recognized by the restriction enzyme *Bst*NI. Therefore, the nucleotide at site 713 could be identified by the *Bst*NI digestion of the PCR product that was amplified by a primer pair, S2 (5′-TACGACTGCTGAAACTATGC-3′) and AS2 (5′-AGCATTAGCAGGGTCAAT-3′). If the site 713 was adenine, the fragment was digested. If it was guanine, it was not digested.

PCR amplification and sequencing of the *Adh*: We used the DNA sequences of alcohol dehydrogenase gene (*Adh*) from *D. tokoro* and *D. tenuipes* to serve as references for the comparison with *Pgi*. Two primer pairs in nested locations were used to obtain discrete PCR products from genomic DNA of Dioscorea. A primer pair, ADH-5′ (5′-GGNMARGTS ATYAARTGCARAG-3′) and ADH-I7 (5′-CCTKKGGATTCA CAAACTCATT-3′), designed according to the consensus *Adh* sequence of 12 plant species, amplified the partial *Adh* fragment (exon 1 to intron 7) as well as many nonspecific PCR products. These primary PCR products served as template for the secondary PCR using primer pair, ADH-AG22 (5′-CCGCTGTTTCCTCGTATCT-3′) and ADH-I7, to obtain a discrete fragment (*ca.* 700 bp) corresponding to the exon 4 to intron 7 of *Adh*. The primer ADH-AG22 was designed according to the *Adh* sequence of *A. thaliana* (CHANG and MAYEROWITZ 1986). PCR products were directly sequenced for both strands. These nucleotide sequences are deposited at the DDBJ, EMBL and GenBank DNA database under the accession numbers, D88913–D88919 and AB006001 (*D. tokoro*) and AB006002 (*D. tenuipes*).

DNA sequence analysis: Multiple alignment of the DNA sequences and the calculation of the number of replacement and silent substitutions between the sequences were carried out by ODEN sequence analysis package (INA 1992). The phylogenetic analysis of the sequences was done by using PHYLIP (Phylogeny Inference Package; FELSENSTEIN 1993).

RESULTS

The *Pgi* gene in *D. tokoro*: A single fragment was amplified from *D. tokoro* (DT37) cDNA by PCR using primer pair 94.22 and 94.24. The DNA sequence of this PCR product is shown in Figure 1. Only one site (site 453) was heterozygous (A and G), and all other sites were homozygous in this individual. The largest open reading frame (ORF) contained in this fragment has 1701 nucleotides that correspond to 567 amino acids. The length of this ORF is similar to that of the cytosolic *Pgi* of *C. lewisii* (569 amino acids; THOMAS *et al.* 1992), and these two sequences have a high homology: 75% in DNA and 83% in deduced amino acid sequences. This high-sequence homology indicates that we have correctly amplified the cytosolic *Pgi* cDNA from *D. tokoro*. All the 61 codons corresponding to 20 amino acids are used in the *D. tokoro Pgi* gene. The codon bias index (χ^2; SHIELDS *et al.* 1988) is 0.25, a typical value for nonbiased genes. Southern hybridization of a 5′-end fragment (253 bp) of *Pgi* cDNA to the restriction digests of *D. tokoro* genomic DNA revealed one or two fragments depending on the restriction enzymes used (data not shown), indicating that the *Pgi* gene coding for the cytosolic PGI is single copy in *D. tokoro*. This is consistent with the observation that the cytosolic PGI allozyme exhibits the pattern of single-locus inheritance (TERAUCHI 1990). Thus, PCR amplification and nucleotide sequencing of the PCR product are not complicated with the multiplicity of this gene in the genome.

DNA polymorphism in the *Pgi* coding region: DNA polymorphisms found among the *Pgi* coding region (cDNA) of 10 *D. tokoro* individuals are summarized in Table 2. Among the entire 1701-bp coding region of *D. tokoro Pgi*, only eight sites were polymorphic. There were doublet mutations in adjacent sites, 335 and 336. Two sites (sites nos. 336 and 453) out of eight were silent and six were replacement base changes. Amino acid changes caused by the replacement substitutions at sites, 335, 385, 713 and 1508 were the nonconservative ones that alter either the size, charge or hydrophobicity of the molecule (MIYATA *et al.* 1979), whereas amino acid replacements at sites 908 and 1369 were conservative. The estimates of nucleotide diversity, π (NEI 1987) and θ (WATTERSON 1975) were $\pi_T = 1.1 \times 10^{-3}$ and $\theta_T = 1.3 \times 10^{-3}$ for the total sites, and $\pi_S = 1.5 \times 10^{-3}$ and $\theta_S = 1.5 \times 10^{-3}$ for the silent sites (Table 3).

Among the six replacement base changes, only two replacement changes at sites 335 and 713 were segregating among the allozyme alleles, *Pgi-a*, *Pgi-b* and *Pgi-c*, and these changes were supposed to be responsible for the allozyme differentiation (TERAUCHI *et al.* 1996). The nucleotide for site 335 was adenine for *Pgi-a*, whereas guanine for *Pgi-b* and *Pgi-c*. The nucleotide for site 713 was adenine for *Pgi-a* and *Pgi-b*, whereas guanine for *Pgi-c*. Deduced amino acids at positions 112 and 238 corresponding to these two nucleotide sites are Asp-Gln for *Pgi-a*, Gly-Gln for *Pgi-b* and Gly-Arg for *Pgi-c*, respectively. Aspartic acid is negatively charged and arginine is positively charged, whereas glycine and gluta-

TABLE 2

Summary of DNA polymorphism in the coding region of *Pgi* among 10 individuals of *Dioscorea tokoro*

Plant code	Allozyme genotype	Nucleotide site no./Amino acid residue no./Amino acid change[a]						
		335/336 112 Gly/Asp	385 129 Phe/Val	453 151 Silent	713 238 Gln/Arg	908 303 Ala/Gly	1369 457 Asp/Asn	1508 503 Gly/Val
DT1	*b/c*	GA	T	A	A[b]	C	G	G
		GA	T	A	G	C	G	G
DT12	*a/a*	AT	T	A	A	C	G	G
		AT	T	A	A	C	G	G
DT27	*a/a*	AT	T	A	A	C	G	G
		AT	T	A	A	C	G	G
DT33	*c/c*	GA	T	A	G	C	G	G
		GA	T	A	G	C	G	G
DT34	*a/c*	AT	T	A	A	C	G	G
		GA	G	A	G	C	G	G
DT37	*b/b*	GA	T	A	A	C	G	G
		GA	T	G	A	C	G	G
DT46	*b/b*	GA	T	A	A	C	G	G
		GA	T	A	A	C	G	T
DT48	*b/b*	GA	T	A	A	C	G	G
		GA	T	G	A	C	G	G
DT51	*c/c*	GA	T	A	G	C	G	G
		GA	T	A	G	C	G	G
DT56	*c/c'*	GA	T	A	G	C	G	G
		GA	T	A	A	G	A	G
DTE1	—	GA	T	A	A	C	G	G
DQ1	—	GA	T	A	A	C	G	G

[a] Nucleotide site numbers (given starting with number one corresponding to the first nucleotide of the start codon), amino acid residue numbers, and amino acid change correspond to the top, middle, and bottom rows of the column heads, respectively.

[b] As the two alleles of the locus were not separated before sequencing, the allocation of segregating nucleotides to the two rows of each individual is arbitrary.

mine are uncharged. The predicted negative net charge for PGI-A and a positive net charge for PGI-C, in comparison to PGI-B, was congruent with the observed relative electrophoretic mobility of the alleles, that is, PGI-A migrates fastest in the anodal direction followed by PGI-B and PGI-C.

Further evidence that the allozyme differentiation is caused by the base changes at the two sites, 335 and

TABLE 3

Summary of DNA variation in *Pgi* and *Adh* genes of *D. tokoro*

	Pgi-coding region			*Pgi*-introns			*Adh*			
	Replacement sites	Silent sites	Total	Introns 4–6	Introns 9–10	Total	Coding	Introns 4–6	Silent site total[a]	Total
Sample size (n)	20	20	20	15	15	15	16	16	16	16
Size (bp)	1319	379	1698	446	1603	2049	357	233	323	590
No. of segregating sites (S)	6	2	8	3	27	30	2	3	4	5
Average no. of pairwise difference (k)	1.3	0.6	1.9	0.4	5.5	5.9	1.0	1.4	1.9	2.4
π ($\times 10^{-3}$)	1.0 ± 0.7	1.5 ± 1.4	1.1 ± 0.7	0.7 ± 0.8	3.4 ± 2.0	2.9 ± 1.6	2.8 ± 2.2	5.9 ± 4.3	5.9 ± 4.0	4.0 ± 2.6
θ ($\times 10^{-3}$)	1.3 ± 0.7	1.5 ± 1.2	1.3 ± 0.7	2.1 ± 1.4	5.2 ± 2.2	4.5 ± 1.9	1.7 ± 1.4	3.9 ± 2.7	3.7 ± 2.3	2.6 ± 1.5
Tajima's D[b]	−0.787 (NS)	0.085 (NS)	−0.593 (NS)	−1.685 (NS)	−1.410 (NS)	−1.515 (NS)	1.616 (NS)	1.510 (NS)	1.802 (NS)	1.862 (NS)
Fu's and Li's D^*[b]	−1.558 (NS)	0.866 (NS)	−0.920 (NS)	−2.150 (<0.05)	−1.819 (NS)	−1.979 (<0.05)	0.907 (NS)	1.044 (NS)	1.141 (NS)	1.214 (NS)
Divergence (d)[c] between *D. tokoro* and *D. tenuipes* ($\times 10^{-3}$)	7.1 ± 2.7	55.9 ± 14.1	18.1 ± 3.8	20.5 ± 7.3	48.9 ± 5.9	42.7 ± 4.9	22.7 ± 8.1	26.2 ± 10.7	34.9 ± 10.6	25.9 ± 6.7

[a] Sites at introns 4–6 and silent sites of exons 4–6 were combined.

[b] Values in parentheses are P values; NS, not significant.

[c] Calculated by JUKES and CANTOR (1969).

713, came from two experiments. By the PCR-based identification methods of the polymorphic sites, we examined the association between allozyme phenotypes and nucleotides at sites 335/6 and 713 among 15 individuals randomly sampled from a natural population (data not shown). The results are as follows: (1) for sites 335/6, the primer ASO1 with GA at the 3′ end, amplified a fragment only when the individual had at least one non-PGI-A band. On the other hand, ASO2 with AT at the 3′ end amplified a fragment only when the individual had at least one PGI-A band. Therefore, the association between PGI-A band and the dinucleotides AT at sites 335/6 was complete; and (2) for site 713, with one exception, the individual was heterozygous for *Bst*NI digestability if it had one PGI-C band together with another band and homozygous for *Bst*NI digestability if it had no PGI-C band. This indicate a strong association of guanine at site 713 and PGI-C phenotype. The probabilities to observe these associations by chance seem extremely low. DNA sequencing of the exceptional individual (DT 56 exhibiting only PGI-C band) revealed that this individual was heterozygous for *Pgi-c* and *Pgi-c′* alleles (Table 2). The *Pgi-c′* allele, although the product of which is electrophoretically similar to that of *Pgi-c* allele, should be grouped with *Pgi-b* in terms of nucleotides at sites 335/6 and 713. We have no information on the frequency of *Pgi-c′* in the natural population. However, judging from the occurrence of this allele in the sample used for sequencing and PCR experiments, the frequency of *Pgi-c′* (1/17) among the alleles corresponding to PGI-C phenotype seems to be far lower than that of *Pgi-c* (16/17), so that our previous estimate of frequency of *Pgi-c* allele by the observation of PGI-C phenotype does not seem to be much misleading. These results confirm that the nucleotide difference at site 335 (corresponding to the amino acid change at site 112) determines the differentiation of PGI-A from PGI-B and PGI-C, and the nucleotide difference at site 713 (amino acid site 238) determines the differentiation of PGI-C from PGI-A and PGI-B in predominant cases.

A part of the *Pgi* coding region (1331 bp) was amplified by PCR and sequenced for two Dioscorea species, *D. tenuipes* and *D. quinqueloba* that are closely related to *D. tokoro* (Figure 1). These two species had GA at sites 335/6 and adenine at site 713 (Table 2). This suggests that GA at sites 335/6 and adenine at site 713 are the ancestral character states among Dioscorea species. Therefore we conclude that *Pgi-b* is the oldest allelic lineage in *D. tokoro*, from which *Pgi-a* and *Pgi-c* were derived by one amino acid replacement each.

DNA polymorphism in introns: We studied two intronic regions each flanking the polymorphic site responsible for the allozyme differentiation, *i.e.*, introns 4–6, flanking the polymorphic site 335 of cDNA (site 77 of exon 5), and introns 9–10, flanking the polymorphic site 713 of cDNA (site 6 of exon 10; Table 4). These two regions are separated by ~500 bp. Among the 15 cloned haplotypes studied, there were three base substitutions and one length polymorphism in the total 446 bp of introns 4–6, and 27 base substitutions and three length polymorphisms in the total 1603 bp of introns 9 and 10. One of the length variations in intron 9 is caused by repeat number changes in a CpT dinucleotide microsatellite.

The levels of DNA polymorphism of *Pgi* introns are summarized in Table 3. Although the nucleotide diversity of introns 9–10 ($\pi = 3.4 \times 10^{-3}$) is five times as large as that of introns 4–6 ($\pi = 0.7 \times 10^{-3}$), a test of heterogeneity in the number of segregating sites between the two regions was not significant ($\chi^2_C = 1.1$; d.f. = 1; $P > 0.3$, $\chi^2_L = 2.4$; d.f. = 1; $P > 0.1$; Kreitman and Hudson 1991). Therefore, we pooled these data to represent the nucleotide diversity in the total introns ($\pi = 2.9 \times 10^{-3}$). These values are only three times as large as those of the coding region ($\pi = 1.1 \times 10^{-3}$).

The interspecific divergence between *D. tokoro* and *D. tenuipes* was lower for introns 4–6 ($d = 20.5 \times 10^{-3}$) than for introns 9–10 ($d = 48.9 \times 10^{-3}$; Table 3). This is congruent with the lower intraspecific polymorphism of *D. tokoro* in the former region than in the latter.

Estimates of the nucleotide diversity, π, of the total intronic regions separately calculated for the three allelic lineages, *Pgi-a*, *Pgi-b* and *Pgi-c* are, 1.9×10^{-3}, 2.5×10^{-3} and 2.0×10^{-3}, respectively. It reveals that *Pgi-b* has the largest nucleotide diversity among the three allelic lineages. But if we consider the possible recombinant origin of DT6536-6 and DT5-15 (see below) and exclude them from the calculation of nucleotide diversity, the three allelic lineages show similar values of nucleotide diversity ($\pi = 2.0 \times 10^{-3}$).

Theory (Strobeck 1983; Hudson and Kaplan 1988) suggests that nucleotide polymorphism that has been maintained for a long time by balancing selection can result in the accumulation of variation at the tightly linked sites. This theory has been applied to explain the experimental result in *D. melanogaster* (Kreitman 1983; Kreitman and Hudson 1991), in which an elevated level of DNA polymorphism was observed around the nucleotide site differentiating ADH-F/S allozymes. To see whether the similar phenomenon is observable around the two segregating replacement sites of *D. tokoro Pgi* gene, a sliding-window analysis was carried out for the two regions, introns 4–6, and introns 9–10 (Figure 2). It is apparent from the plot that no high peak of nucleotide polymorphism coincides with the segregating replacement sites that are responsible for the allozyme differentiation. The patterns of sliding-window plots for polymorphism in *D. tokoro* (Figure 2A) and divergence between *D. tokoro* and *D. tenuipes* (Figure 2B) did not show any obvious correlation, and we could not detect any small-scale heterogeneity in evolutionary forces (*i.e.*, functional constraints and mutation rate)

TABLE 4

Summary of DNA polymorphisms detected in the two regions (introns 4–6 and introns 9–10) of *Pgi* of *D. tokoro*

	Intron 4 (286 bp)				Exon 5 (156 bp)		Intron 5 (72 bp)[a]	Exon 6 (97 bp)	Intron 6 (88 bp)[a]	Intron 9 (926 bp)								
Nucleotide site:[b]	18	43	56	82/106	77[c]/78	119		39		19	24	32	63	84	105	118	391/400	434
Type of change:				del			None		None								del	
Consensus:	T	A	A	AGTTTATTCTTTA CTTGTCTCTAA	GA	T		A		T	T	A	A	A	T	C	TGCTA AATTT	T
DNA clone code[d]																		
DT34-2 (*Pgi-a*)	*	*	*	*	AT	*		*		*	*	*	*	*	*	*	*	*
DT54-3 (*Pgi-a*)	*	G	*	*	AT	C		*		*	*	*	*	*	C	*	*	*
DT12-1 (*Pgi-a*)	*	*	*	*	AT	*		*		*	*	*	*	*	*	*	*	*
DT25-4 (*Pgi-a*)	*	*	*	*	AT	*		*		*	*	*	*	G	*	*	*	*
DT7-1-10 (*Pgi-b*)	*	*	*	*	**	*		G		*	*	*	*	*	*	G	*	*
DT1-7 (*Pgi-b*)	*	*	*	*	**	*		G		*	*	*	*	*	*	G	*	*
DT6536-4 (*Pgi-b*)	*	*	*	*	**	*		G		*	*	*	*	*	*	*	*	*
DT37-5/5 (*Pgi-b*)	*	*	C	—	**	*		*		*	*	*	*	*	*	*	*	C
DT6536-6 (*Pgi-b*)	*	*	*	*	**	*		*		*	*	*	*	*	*	*	—	*
DT5-15 (*Pgi-b*)	*	*	*	*	**	*		*		*	*	*	*	*	*	*	—	*
DT24-2 (*Pgi-c*)	*	*	*	*	**	*		*		*	*	G	T	*	*	*	—	*
DT51-9 (*Pgi-c*)	C	*	*	*	**	*		*		*	*	*	T	*	*	*	—	*
DT54-10 (*Pgi-c*)	*	*	*	*	**	*		*		*	C	*	T	*	*	*	—	*
DT49-3 (*Pgi-c*)	*	*	*	*	**	*		*		C	*	*	T	*	*	*	—	*
DT18-6 (*Pgi-c*)	*	*	*	*	**	*		*		*	*	*	T	*	*	*	—	*
DT1-Pgi (*D. tenuipes*)	*	*	*	*	**	*		*		*	*	*	*	*	—	*	*	*

	Intron 9 (926 bp)											Exon 10 (49 bp)	Intron 10[e] (867 bp)							
Nucleotide site:[b]	497	505	535	545	556	574	746	753/790	866	878	914	6[c]	58	116	165	243	592	703/704	715	732
Type of change:								Microsatellite												
Consensus:	T	T	T	T	—	A	C		T	T	G	A	T	T	T	T	T	CC	A	T
DNA clone code[d]																				
DT34-2 (*Pgi-a*)	*	*	A	*	*	*	*	$(CT)_{18}$	*	*	*	*	*	*	*	*	*	**	*	*
DT54-3 (*Pgi-a*)	*	*	*	*	*	T	*	$(CT)_{16}$	*	*	*	*	*	*	C	*	*	**	*	*
DT12-1 (*Pgi-a*)	*	*	A	*	*	*	*	$(CT)_{19}$	*	*	*	*	*	*	*	*	*	**	*	*
DT25-4 (*Pgi-a*)	*	*	*	*	*	T	*	$(CT)_{17}$	*	*	*	*	*	*	*	*	C	**	*	*
DT7-1-10 (*Pgi-b*)	*	*	*	*	*	*	*	$(CT)_{11}$	*	*	*	*	*	*	*	*	*	**	*	*
DT1-7 (*Pgi-b*)	*	*	*	*	*	*	*	$(CT)_{11}$	*	C	*	*	*	C	*	*	*	TT	*	*
DT6536-4 (*Pgi-b*)	*	*	*	*	*	*	*	$(CT)_{12}$	*	*	*	*	*	*	*	*	*	TT	*	*
DT37-5/5 (*Pgi-b*)	*	*	*	C	*	*	*	$(CT)_{11}$	*	*	*	*	*	*	*	*	*	TT	*	*
DT6536-6 (*Pgi-b*)	C	*	*	*	T	*	*	$(CT)_{13}GT(CT)_3$	*	*	A	*	*	*	*	*	*	**	*	*
DT5-15 (*Pgi-b*)	C	*	*	*	*	*	*	$(CT)_{11}GT(CT)_3$	*	*	*	*	*	*	*	*	*	**	*	G
DT24-2 (*Pgi-c*)	C	C	*	*	*	*	*	$(CT)_{12}GT(CT)_3$	*	*	*	G	*	*	*	*	*	**	*	*
DT51-9 (*Pgi-c*)	C	*	*	*	*	*	*	$(CT)_{12}GT(CT)_3$	C	*	*	G	*	*	*	*	*	**	*	*
DT54-10 (*Pgi-c*)	C	*	*	*	*	*	T	$(CT)_{12}GT(CT)_3$	*	*	*	G	C	*	*	*	*	**	*	*
DT49-3 (*Pgi-c*)	C	*	*	*	*	*	*	$(CT)_{12}GT(CT)_3$	*	*	*	G	*	*	*	*	*	**	*	*
DT18-6 (*Pgi-c*)	C	*	*	*	*	*	*	$(CT)_{12}GT(CT)_3$	*	*	*	G	*	*	*	A	*	**	C	*
DTE1-Pgi (*D. tenuipes*)	*	*	—	*	*	—	*	$(CT)_8CA(CT)_4CA(CT)_4CA$	*	*	*	*	*	*	*	*	*	*T	*	*

[a] There was no variation in intron 5 and 6.
[b] Nucleotide site numbers are given starting with number one corresponding to the first nucleotide of each exon and intron on the basis of aligned sequences of the 15 haplotypes.
[c] Replacement sites discriminating the three allozymes.
[d] *Pgi*-allozyme type in parentheses.
[e] Sequence information of the region between sites 304 and 493 is lacking. Size of the studied region of intron 10 is 677 bp.

along the sequence that affect the levels of polymorphism and divergence in a similar manner.

Genealogical relationships among allozyme lineages: The distribution of mutations among the 15 *D. tokoro* *Pgi* haplotypes (Table 4) reveals no clear evidence for recombinational events. There is no single pair of polymorphic sites exhibiting all four possible gamete states attributable to recombination (HUDSON and KAPLAN 1985). However, there is a cluster of polymorphism that could be caused by recombinations (site 391 to site 790 of intron 9 for the two haplotypes DT6536-6 and DT5-15). To explain this pattern by recombination, two recombinations (which occurred at two sites located between sites 63 and 391 and between sites 790 and 926

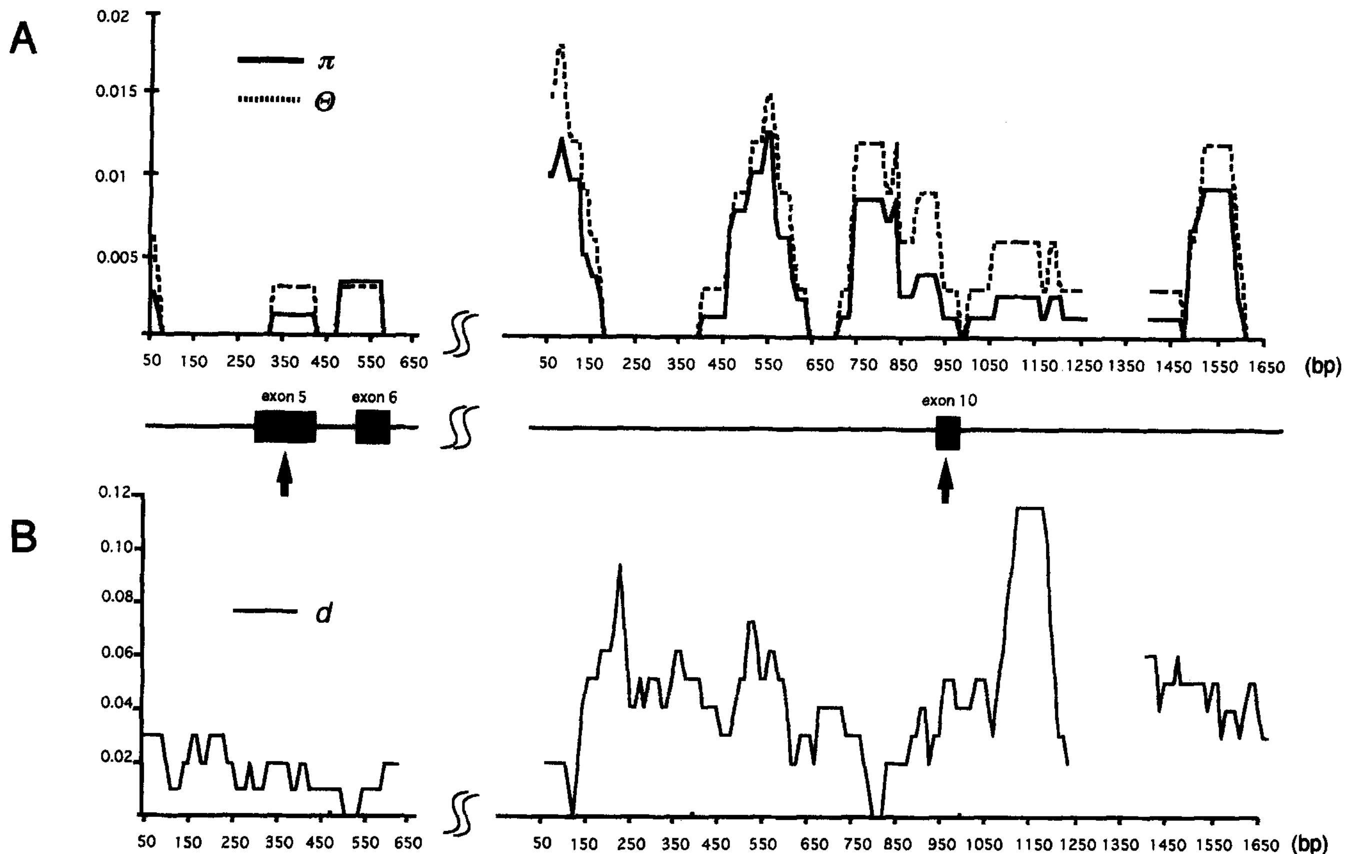

FIGURE 2.—(A) Sliding window plot of π (NEI 1987; ——) and θ (WATTERSON 1975; -----). Replacement polymorphism at the two sites (site 77 of exon 5 and site 6 of exon 10; positions indicated by arrows) responsible for the allozyme differentiation were excluded from the calculations of π and θ. (B) Sliding window plot of d (estimated number of nucleotide substitutions per site; JUKES and CANTOR 1969) between *D. tokoro* and *D. tenuipes Pgi*. From 5′ to 3′ direction, a window of 100 bp was moved in 10-bp increments.

of intron 9, respectively) have to be postulated. Two haplotypes, DT6536-6 and DT5-15, may be the recombinants with *Pgi-c* type sequence for the sites between 391 and 790 of intron 9 and *Pgi-b* type sequence for the rest of the region. Although this clustering of mutations is not statistically significant [$P = 0.23$ by the test of Stephens (STEPHENS 1985)], we cannot discard the possibility of their recombination origin. It is difficult from the present data to say whether the mutations of DT6538-6 and DT5-15 common to *Pgi-c* haplotypes are the outcome of recombinations or reflecting their shared phylogeny.

Genealogical relationships among the three major allozyme lineages, *Pgi-a*, *Pgi-b* and *Pgi-c* of *D. tokoro* in relation to *D. tenuipes Pgi* was inferred from the DNA sequences of the 16 genomic clones. Pairwise genetic distances between haplotypes were calculated (JUKES and CANTOR 1969) using all the sites (2351 bp, including exons 5, 6 and 10), which was then used to construct a tree (Figure 3) by the neighbor-joining method (SAITOU and NEI 1987). If we consider the possible recombination origin of the two haplotypes, DT6536-6 and DT5-15, and exclude them from the analysis, we get an identical tree as Figure 3 except for their absence from the corresponding positions. The haplotypes belonging to *Pgi-a* and *Pgi-c* form monophyletic groups, respectively. The star-like phylogeny within the *Pgi-c* lineage reflects that all the mutations in *Pgi-c* haplotypes are singletons. The *Pgi-b* haplotypes are separated into two groups, one of which (consisting of DT6536-6 and DT5-15) clusters with the haplotypes of *Pgi-c*. Although the statistical support is not high, the root of the tree is located on the base of the larger *Pgi-b* cluster. This genealogical relationship suggests the following two alternative hypothetical scenarios of *Pgi* evolution in *D. tokoro*: (1) among the ancestral *Pgi-b* allelic lineage, *Pgi-a* allelic lineage has first diverged by a replacement substitution at site 335, and later in a *Pgi-b* lineage of DT6536-6 and DT5-15 a substitution at site 713 gave rise to the *Pgi-c* allelic lineage; or (2) the divergence of *Pgi-a* allelic lineage from *Pgi-b* was followed by the divergence of *Pgi-c* from *Pgi-b*, but a recombination between *Pgi-b* and *Pgi-c* gave rise to the two haplotypes, DT6536-6 and DT5-15.

To infer the time of diversification of the three allozyme lineages, genetic distances among the three lineages were calculated by using the intronic sequences of the 15 genomic haplotypes taking the within-allele diversity into consideration (NEI 1987). The average

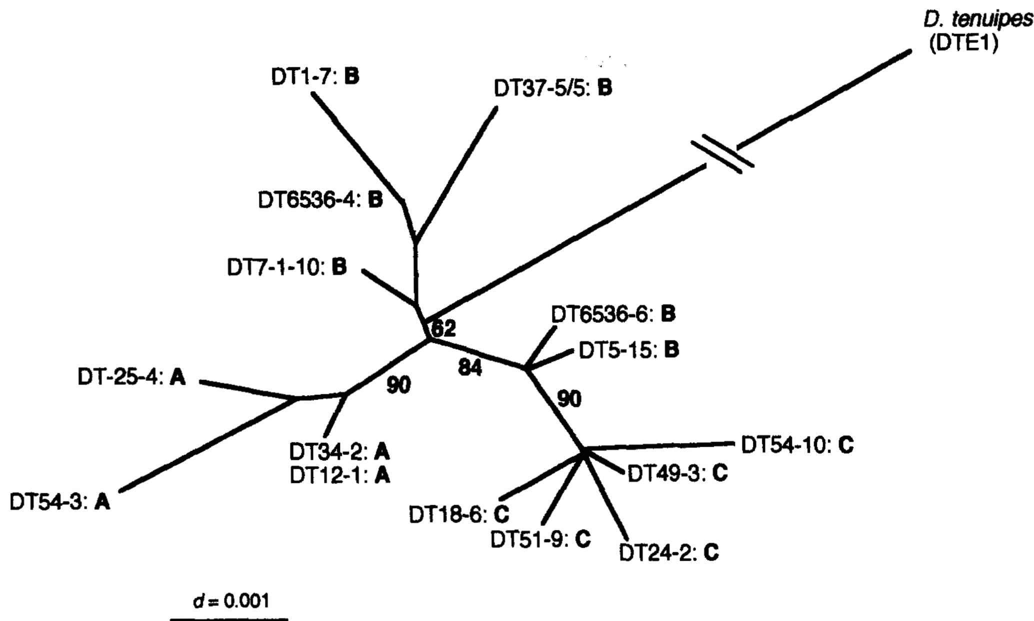

FIGURE 3.—NJ tree constructed from the DNA sequence data of 16 haplotypes (genomic clones) of *Pgi* of *D. tokoro* and *D. tenuipes*. *Pgi* allele types are also indicated (A, *Pgi-a*; B, *Pgi-b*; and C, *Pgi-c*). The numbers on branches are bootstrapping probabilities after 100 replications. The unit *d* indicates the number of substitutions per site. The exclusion of the two haplotypes, DT6536-6 and DT5-15 does not change the topology of the tree.

numbers of nucleotide substitution per site were as follows: $d_{ab} = 4.7 \times 10^{-4}$, $d_{bc} = 11.4 \times 10^{-4}$ and $d_{ca} = 16.7 \times 10^{-4}$ ($d_{ab} = 7.5 \times 10^{-4}$, $d_{bc} = 20.9 \times 10^{-4}$ and $d_{ca} = 16.7 \times 10^{-4}$ when DT6536-6 and DT5-15 were excluded from the calculation), where the subscripts *a*, *b* and *c* indicate *Pgi-a*, *Pgi-b* and *Pgi-c*, respectively. If we apply the average silent substitution rate, 7×10^{-9} per site per year, calculated for *Pgi* of gramineae (see below) to our results, the time after allozyme divergence is estimated to be $T_{ab} = (4.7 \times 10^{-4})/(2 \times 7 \times 10^{-9}) =$ 34,000 years, T_{bc} = 81,000 years and T_{ca} = 120,000 years (T_{ab} = 54,000 years, T_{bc} = 150,000 years and T_{ca} = 120,000 years when DT6536-6 and DT5-15 are excluded). All of these dates correspond to the last glaciation period (Wisconsin-Wurm period), when population sizes of many organisms were supposed to be much smaller than the present ones (NEI 1987).

Comparison with the *Adh* gene: The alcohol dehydrogenase gene (*Adh*) is the best studied gene in plant molecular population genetics (GAUT and CLEGG 1991, 1993a,b; HANFSTINGL *et al.* 1994; INNAN *et al.* 1996; MIYASHITA *et al.* 1996). To serve as a reference gene for the comparison with *Pgi*, *Adh* of *D. tokoro* and *D. tenuipes* were sequenced. A total of 590-bp region, spanning over exon 4 to exon 7, was sequenced for eight *D. tokoro* individuals corresponding to 16 alleles and one *D. tenuipes* clone. Among *D. tokoro* DNAs, five polymorphic sites (two in exons and three in introns) were observed (Table 5). All the polymorphic sites were segregating between two major haplotypes. An individual DT49 seems to be a heterozygote of these two haplotypes. The levels of polymorphism in *D. tokoro Adh* were $\pi = 2.8 \times 10^{-3}$ and $\theta = 1.7 \times 10^{-3}$ for the coding region and $\pi = 5.9 \times 10^{-3}$ and $\theta = 3.9 \times 10^{-3}$ for introns (Table 3). These levels of polymorphism are larger than that of silent sites of *Adh* of pearl millet ($\theta = 2.0 \times 10^{-3}$: GAUT and CLEGG 1993a), but smaller than the values obtained for *Adh* in other plant species studied to date: the total coding region of *Adh* of *Arabis gemmifera* ($\pi = 5.5 \times 10^{-3}$: MIYASHITA *et al.* 1996), *A. thaliana* ($\pi = 5.6 \times 10^{-3}$: INNAN *et al.* 1996) and silent site of maize ($\theta = 2.1 \times 10^{-2}$: GAUT and CLEGG 1993b).

Comparison of the levels of polymorphisms within *D. tokoro* between *Pgi* and *Adh* (Table 3) reveals more than twice larger values of π for *Adh* ($\pi_C = 2.8 \times 10^{-3}$ for the coding region and $\pi_I = 5.9 \times 10^{-3}$ for introns) than for *Pgi* ($\pi_C = 1.1 \times 10^{-3}$ and $\pi_I = 2.9 \times 10^{-3}$), but similar values of θ for *Adh* ($\theta_C = 1.7 \times 10^{-3}$ for coding region and $\theta_I = 3.9 \times 10^{-3}$ for introns) and *Pgi* ($\theta_C = 1.3 \times 10^{-3}$ and $\theta_I = 4.5 \times 10^{-3}$). This discrepancy between the comparison of π and θ is dictated in the large difference in Tajima's *D* values for *Adh* and *Pgi*. There was excess amount of rare variants in *Pgi* in contrast to excess amount of high-frequency variants in *Adh*.

Base substitution rate of a gene is a function of mutation rate and selective constraint on the gene (KIMURA 1983). Therefore, we can compare the magnitude of mutation rate and selective constraint between genes by comparing their base substitution rates. For that purpose, the sequence divergences between the pairs of three plant species, *D. tokoro*, *O. sativa* and *Z. mays*, were estimated for *Pgi* and *Adh* (Table 6). For both the replacement and silent base substitutions, substitution rates in *Pgi* were not significantly different from those of *Adh* in all species pairs. This is also true for the comparison of divergences between *D. tokoro* and *D. tenuipes* for *Pgi* and *Adh* (Table 3). The base substitution rate of *Adh* in gramineae plants is reported by GAUT and CLEGG (1991) to be 2.5×10^{-10} substitutions per site per year for replacement sites and 7.9×10^{-9} substitutions per site per year for silent sites. From these

TABLE 5

Polymorphic sites in the region between exon 4 and exon 7 of *Adh* of *D. tokoro* and sites that have diverged between *D. tokoro* and *D. tenuipes*

	Exon 4 (181 bp)				Intron 4 (72 bp)		Exon 5 (83 bp)	Intron 5 (85 bp)				Exon 6 (76 bp)			Intron 6 (76 bp)		Exon 7 (17 bp)	
Nucleotide site[a]:	1	6	81	118	13	37	32	35	63	66	71	17	32	60	21	44	2	8
R or S:	R	S	S	R			S					R	R	S			S	S
Consensus:	G	G	A	G	G	C	A	A	G	G	G	G	T	T	T	G	G	A
Plant code:																		
DT1	*	*	*	*	*	*	*	*	A	*	*	*	C	*	G	C	A	*
DT8	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
DT12	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
DT27	*	*	*	*	*	*	*	*	A	*	*	*	C	*	G	C	A	*
DT49	*	*	*	*	*	*	*	*	R	*	*	*	Y	*	K	S	R	*
DT51	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
DT54	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
DT56	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	A	*
DTE1-Adh[b]	A	A	C	T	A	T	G	T	A	A	T	A	C	C	*	*	*	G

R, replacement; S, silent.
[a] Nucleotide site numbers are given starting with number one corresponding to the first nucleotide of each exon and intron.
[b] *D. tenuipes.*

values, we can estimate the base substitution rate of gramineae *Pgi* as 2.7×10^{-10} substitutions per site per year for replacement sites and 7.0×10^{-9} substitutions per site per year for silent sites, by comparing the nucleotide divergences between *O. sativa* and *Z. mays* for *Adh* and *Pgi*.

Test of neutrality: *Tajima's test:* Tajima's test (Table 3) did not reject the neutral mutation hypothesis throughout the *Pgi* region as indicate the nonsignificant *D* values (Tajima 1989a). However, it is notable that most of the *D* values for *Pgi* are negative. This can be seen in the sliding window plot (Figure 2) as the larger values of θ than π for the most part of the intronic region. This situation is in contrast to *Adh* region where the *D* values are positive, although not significant. These results may indicate that *Pgi* and *Adh* are under different evolutionary forces.

Fu's and Li's test: Fu's and Li's test rejected the neutrality for *Pgi*-introns 4-6 ($D^* = -2.150$, $P < 0.05$) and the total introns ($D^* = -1.979$, $P < 0.05$; Table 3) (Fu and Li 1993). This test compares the two estimates of $4N_e\mu$ that are separately calculated from the numbers of total mutations and singleton mutations in the DNA sample. Too many singleton mutations in the *D. tokoro* *Pgi* intronic region seem to be responsible for the significantly negative D^* values. In the total intronic region examined (2049 bp) there were 30 segregating sites, and 22 sites were singletons, which is far more than the number (10) expected under the neutral mutation model (Tajima 1989a).

MK-test: As described in Terauchi *et al.* (1996), within the 1289 bp of *Pgi* coding region, the numbers of replacement and silent substitutions were 7 and 16 between *D. tokoro* and *D. tenuipes*, and 9 and 23 between *D. tokoro* and *D. quinqueloba*, respectively. On the other hand, there were six replacement and two silent polymorphisms within *D. tokoro*. McDonald and Kreitman's test applied to these data (McDonald and Kreitman 1991) rejected the neutral mutation hypothesis at the locus (Fisher's exact test: $P = 0.037$ for the comparison

TABLE 6

The numbers of replacement and silent substitutions per site for *Pgi* and *Adh* between the three species: *D. tokoro*, *Z. mays* and *O. sativa*

	Pgi		*Adh*	
	Replacement site (1265 bp)	Silent site (388 bp)	Replacement site (263 bp)	Silent site (91 bp)
D. tokoro/*O. sativa*[a]	0.125 ± 0.011	1.398 ± 0.158	0.102 ± 0.021	1.232 ± 0.265
D. tokoro/*Z. mays*[a]	0.123 ± 0.010	1.298 ± 0.139	0.087 ± 0.019	2.017 ± 0.709
O. sativa[a]/*Z. mays*[a]	0.052 ± 0.007	0.562 ± 0.053	0.047 ± 0.014	0.727 ± 0.138

Values in this table were calculated by Nei and Gojobori (1986).
[a] DNA sequences were obtained from the DNA database, GenBank. Accession numbers of the sequences are as follows: *O. sativa Pgi*, D45218; *O. sativa Adh1*, X16296; *Z. mays Pgi*, U17225; and *Z. mays Adh1*, M32984.

TABLE 7

HKA test for silent site differences in four regions

		Within *D. tokoro*			Between *D. tokoro* and *D. tenuipes*						
			No. of segregating sites			No. of different sites					
Region	Sample size	No. of sites	Obs.	Exp.	No. of sites	Obs.	Exp.	θ	T	χ^2	P
Pgi introns 4–6	15	415	3	2.9	396	8	8.1	0.0022	8.28	0.001	>0.10
vs. Pgi introns 9–10	15	1572	26	26.1	1458	69	68.9	0.0051			
Pgi introns total	15	1987	29	26.5	1854	77	79.6	0.0041	9.48	1.436	>0.10
vs. Pgi coding region	20	296.7	2	4.5	296.7	16	13.4	0.0043			
Pgi introns total	15	1987	29	29.0	1854	77	77.0	0.0045	8.24	0.000	>0.10
vs. Adh silent sites	16	323.3	4	4.0	323.3	11	11.0	0.0037			
Pgi coding region	20	296.7	2	3.4	296.7	16	14.7	0.0032	14.49	1.145	>0.10
vs. Adh silent sites	16	323.3	4	2.7	323.3	11	12.4	0.0025			

Obs., observed; Exp., expected.

between *D. tokoro* and *D. tenuipes*, $P = 0.022$ for *D. tokoro* and *D. quinqueloba*), in the direction that the level of replacement polymorphism within *D. tokoro* is too high compared to that of silent polymorphism.

HKA-test: We tested whether the levels of intraspecific polymorphisms of *D. tokoro* are consistent with the levels of interspecific divergence between *D. tokoro* and *D. tenuipes* (HKA-test; HUDSON *et al.* 1987) by comparing two different regions at a time. We made four different comparisons: (1) *Pgi* introns 4–6 *vs. Pgi* introns 9–10, (2) *Pgi* total introns *vs. Pgi* coding region, (3) *Pgi* total introns *vs. Adh* silent sites, and (4) *Pgi* coding region *vs. Adh* silent sites (Table 7). In all the cases the test was not significant.

DISCUSSION

Location of amino acid sites responsible for allozyme variation: The crystallographic structure of PGI protein is known at 3.5 Å resolution using pig as a source organism (SHAW and MUIRHEAD 1977). It is composed of two identical subunits, each of which consists of a large and a small domain. For each dimer, two catalytic sites are present at the interface between the large and small domains (ACHARI *et al.* 1981). Assuming that the PGI structure of *D. tokoro* is similar to that of pig, position of the polymorphic amino acid sites responsible for the allozyme differentiation can be located on the PGI three-dimensional structure. Thus, amino acid site 112 (corresponding to the DNA site 335) may be located on a beta-sheet in the larger domain, which corresponds to the interface of two subunits close to the catalytic site (site 98). The amino acid 238 (DNA site 713) is situated in the bottom of the larger domain. This site is far from the catalytic site, but represents the region where subunit association occurs.

No indication of long time balancing selection: The co-existence of the three allozymes with intermediate frequencies in *D. tokoro* population suggests a possibility of balancing selection working at the *Pgi* locus. A long time balancing selection at a DNA site is known to result in the accumulation of neutral DNA polymorphism in its close vicinity (HUDSON and KAPLAN 1988). To test this, the level of DNA polymorphism around the sites discriminating the allozymes was studied. Sliding-window analysis (Figure 2) detected no elevated DNA polymorphism in the close vicinity of the two sites responsible for the allozyme differentiation. This result is not compatible with the idea that a long time balancing selection has been operating at the two sites that are responsible for the differentiation of allozymes. Tajima's *D* and Fu's and Li's D^* values were negative throughout the intronic region examined (Table 3), which is also incompatible with the long time balancing selection scenario.

Low level of DNA polymorphism in the *D. tokoro* Pgi: The level of DNA polymorphism at the *Pgi* coding region of *D. tokoro* ($\pi = 1.1 \times 10^{-3}$; Table 3) was lower than those of *Adh* of *D. tokoro* ($\pi = 2.8 \times 10^{-3}$) as well as *Adh* of most other plant species studied to date. The silent sites of *Pgi* coding region and introns also showed low levels of polymorphism ($\pi = 1.5 \times 10^{-3}$ and 2.9×10^{-3}, respectively). The low levels of DNA polymorphism at *D. tokoro Pgi* can be caused by (1) a strong selective constraint on the *Pgi* (purifying selection), (2) a low mutation rate in the *Pgi* region, (3) the genetic hitchhiking of the *Pgi* region with a tightly linked site that is under selection (selective sweep and background selection), and (4) a small effective size of *D. tokoro* population. The comparison of base substitution rates (Table 6) indicated that *Pgi* is evolving as fast as *Adh* for both replacement and silent sites. Therefore, it is

difficult to assume that the lower polymorphism in *D. tokoro Pgi* coding region in comparison to *Adh* is because of a stronger purifying selection and/or a low mutation rate at the *Pgi* locus. Codon usage of *D. tokoro Pgi* is nonbiased, and it is improbable that the low level of polymorphism at silent sites in coding region is because of the selective constraint imposed on silent sites (SHARP and LI 1989; SHARP and MATASSI 1994) by the codon preference. Selective sweep (HUDSON 1994) and background selection (CHARLESWORTH 1994) are the mechanisms with which polymorphism is eliminated by genetic hitchhiking of the region with a tightly linked selected mutation. In view of the recent origins of *Pgi-a* and *Pgi-c* alleles from *Pgi-b*, there is a possibility that the rapid increases in frequency of these former two allelic lineages have reduced the amount of neutal variation from the *Pgi* region. It is also possible that the low levels of polymorphism at the *Pgi* locus are reflective of the small effective size of *D. tokoro* population, and that the higher level of polymorphism at the *D. tokoro Adh* than *Pgi* is maintained by nonneutral mechanisms such as balancing selection.

The lower levels of polymorphism in *Pgi* compared with other loci are also observable in Drosophila [J. MCDONALD and M. KREITMAN, unpublished data; MORIYAMA and POWELL (1996); *D. melanogaster*: $\pi = 0.8 \times 10^{-3}$ for *Pgi* coding region and $\pi = 4.4 \times 10^{-3}$ for the average of coding region of 16 autosomal genes; *D. simulans*: $\pi = 4.0 \times 10^{-3}$ for *Pgi* and 9.6×10^{-3} for six autosomal genes]. In this case also, the level of divergence between *D. melanogaster* and *D. simulans* was similar for *Pgi* ($d = 1.6 \times 10^{-2}$) and *Adh* ($d = 1.4 \times 10^{-2}$). It is of interest to see whether the lower levels of polymorphism at *Pgi* in comparison to other loci are prevailing in other organisms, especially in those with high levels of allozyme polymorphisms.

Negative values of Tajima's *D* and Fu's and Li's *D for the *Pgi* region:** The negative values of Tajima's *D* and Fu's and Li's *D** (Table 3) suggest the three possibilities: (1) the mutations are deleterious and are maintained with low frequency (purifying selection), (2) the population has recently experienced a bottleneck (TAJIMA 1989b; FU and LI 1993), and (3) a newly arisen advantageous mutation linked to the investigated region had rendered the site frequency spectrum skewed toward excess of rare variants (BRAVERMAN *et al.* 1995). The positive *D* and *D** values observed for the *Adh* region seem incompatible with the possibility (2), because the population bottleneck should affect the entire genome. Both the possibilities (1) and (3) remain as plausible reasons for the negative *D* and *D** values.

Excess of replacement polymorphism at *Pgi*: Comparison of intraspecific polymorphism and interspecific divergence for silent and replacement mutations indicated a significant excess of intraspecific replacement polymorphism in *D. tokoro* [McDonald and Kreitman's test (MCDONALD and KREITMAN 1991); $P < 0.05$]. This is similar to the results obtained for *Adh* of *Arabis gemmifera* (MIYASHITA *et al.* 1996) but opposite to what is observed in nuclear and mitochondrial genes of Drosophila (MCDONALD and KREITMAN 1991; EANES *et al.* 1993; BALLARD and KREITMAN 1994; RAND *et al.* 1994). The larger replacement polymorphism relative to the divergence is expected when (1) the effective size of population is small (OHTA 1993) and (2) balancing selection is working on amino acid replacements within the species. When the effective population size is small, behavior of slightly deleterious mutations is shown to be similar to the neutral mutations (OHTA 1976). If the number of effectively neutral replacement sites is larger than that of silent sites, we can expect more polymorphisms at replacement sites than at silent sites. This mechanism is proposed to explain the generally greater proportion of replacement polymorphism in *Drosophila melanogaster* in comparison to *D. simulans* genes (MORIYAMA and POWELL 1996). The mechanism involving balancing selection is invoked to explain the excess of replacement polymorphism in the human MHC genes (HUGHES and NEI 1988). With the data in hand, it is difficult to say which of these two mechanisms explain the results better.

Maintenance mechanism of allozyme and DNA polymorphisms: A higher allozyme heterozygosity in contrast to a low nucleotide diversity as observed in *D. tokoro* had been reported in *Drosophila melanogaster* in comparison to *D. simulans* (AQUADRO *et al.* 1988) and in *Pennisetum glaucum* in comparison to *Z. mays* (GAUT and CLEGG 1993b). AQUADRO *et al.* (1988) suggested that their observation could be explained by a smaller effective population size of *D. melanogaster* than *D. simulans*. A similar argument seems applicable to *D. tokoro*. The assumption of small effective size of *D. tokoro* fits with the observed low nucleotide diversity at the *Pgi* locus, although in this case we have to assume a nonneutral mechanism for the maintenance of the higher level of *Adh* polymorphism. The small effective size of the population is also compatible with an excess of replacement polymorphism in *D. tokoro Pgi*. Therefore, with the assumption of a small effective size of *D. tokoro* population, we can explain all the observed patterns of DNA polymorphism at *Pgi* locus with the neutral mutation hypothesis without invoking any positive selection. On the other hand, it is also possible to explain the observed pattern of *Pgi* polymorphism with positive selection. *Pgi-a* and *Pgi-c* alleles may be advantageous and had increased their frequencies to intermediate levels in a short time. The rapid increase in their frequencies had sweeped the polymorphism from the region and brought about the negative Fu's and Li's *D** values. Replacement polymorphism present in excess might be because they are selectively advantageous. The result of Watterson's test may indicate that balancing selection is maintaining the allozyme polymorphism (WATTERSON 1978). At present it is difficult to say which of the neu-

tralist and selectionist interpretations explain the observations better. However, we need more evidence in favor of positive selection scenario if we are to reject the neutral hypothesis regarding it as the null hypothesis in the molecular evolution study.

In conclusion, the combined observation of the allozyme and DNA variations of *D. tokoro Pgi* revealed that the allele frequencies of *Pgi-a* and *Pgi-c* have increased to intermediate level in a very short time after their emergence in the population. Why this rapid increase in allele frequencies have occurred is not clear. To answer this question, we are now experimentally comparing the kinetics and thermostability of *D. tokoro* PGI enzymes coded by different genotypes to see whether the differences are detectable among them. The causes of molecular evolution should be understood by the integrated studies on history and function of the molecules.

We thank Prof. L. D. GOTTLIEB, University of California, Davis, for the provision of a clone and a DNA sequence of *Pgi* of *A. thaliana* before they were published. We also thank for invaluable suggestions provided by F. TAJIMA and H. INNAN, University of Tokyo, and K. WEISING, R. ROMPF, and J. RAMSER, University of Frankfurt, to improve the manuscripts. R.T. specially thanks Prof. G. KAHL, University of Frankfurt, for his support to carry out the study at his laboratory. This study was supported by grant no. 94-103-1005 of Sumitomo Foundation, Japan, and grant no. 05740521 of Ministry of Education and Culture of Japan, and by the fellowship no. IV-1-7121-1028559 to R.T. from the Alexander-von-Humboldt Foundation, Germany, for which we express our cordial gratitude.

LITERATURE CITED

ACHARI, A., S. E. MARCHALL, H. MUIRHEAD, R. H. PALMIERI and E. A. NOLTMANN, 1981 Glucose-6-phosphate isomerase. Phil. Trans. R. Soc. Lond. B **293:** 145–157.

AQUADRO, C. F., K. M. LADO and W. A. NOON, 1988 The *rosy* region of *Drosophila melanogaster* and *Drosophila simulans*. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. Genetics **119:** 875–888.

BALLARD, J. W. O., and M. KREITMAN, 1994 Unraveling selection in the mitochondrial genome of Drosophila. Genetics **138:** 757–772.

BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphism. Genetics **140:** 783–796.

CHANG, C., and E. M. MEYEROWITZ, 1986 Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. USA **83:** 1408–1412.

CHAPUT, M., V. CLAES, D. PORTETELLE, I. CLUDTS, A. CRAVADOR *et al.*, 1988 The neurotrophic factor neuroleukin is 90% homologous with phosphohexose isomerase. Nature **332:** 454–455.

CHARLESWORTH, B., 1994 The effect of background selection against deleterious mutations on weakely selected, linked variants. Genet. Res. **63:** 213–227.

CHOMCZYNSKI, P., and N. SACCHI, 1987 Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162:** 156.

EANES, W. F., M. KIRCHNER and J. YOON, 1993 Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90:** 7475–7479.

FELSENSTEIN, J., 1993 *PHYLIP* (Phylogeny Inference Package) Version 3.57c, Department of Genetics, University of Washington, Seattle.

FROHMAN, M. A., M. K. DUSH and G. R. MARTIN, 1988 Rapid production of full length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. USA **85:** 8998–9002.

FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. Genetics **133:** 693–709.

GAUT, B. S., and M. T. CLEGG, 1991 Molecular evolution of alcohol dehydrogenase 1 in members of the grass family. Proc. Natl. Acad. Sci. USA **88:** 2060–2064.

GAUT, B. S., and M. T. CLEGG, 1993a Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (Poaceae). Genetics **135:** 1091–1097.

GAUT, B. S., and M. T. CLEGG, 1993b Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc. Natl. Acad. Sci. USA **90:** 5095–5099.

GILLESPIE, J. F., 1991 *The Causes of Molecular Evolution*. Oxford University Press, London.

GOTTLIEB, L. D., 1982 Conservation and duplication of isozymes in plants. Science **216:** 373–380.

GREEN, J. B. A., A. P. H. WRIGHT, W. Y. CHEUNG, W. E. LANCASHIRE and B. S. HARTLEY, 1988 The structure and regulation of phosphoglucose isomerase in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **215:** 100–106.

HALL, J. G., 1985 Temperature-related kinetic differentiation of glucosephosphate isomerase alleloenzymes isolated from the blue mussel, *Mytilus edulis*. Biochem. Genet. **23:** 705–728.

HANFSTINGL, U., A. BERRY, E. A. KELLOG, J. T. COSTA III, W. RUDIGER *et al.*, 1994 Haplotype divergence coupled with lack of diversity at the *Arabidopsis thaliana* alcohol dehydrogenase locus: role for both balancing and directional selection? Genetics **138:** 811–828.

HOFFMANN, R. J., 1981 Evolutionary genetics of *Metridium senile*: I. Kinetic differences in phosphoglucose isomerase allozymes. Biochem. Genet. **19:** 129–144.

HUDSON, R. R., 1994 How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination rates be explained? Proc. Natl. Acad. Sci. USA **91:** 6815–6818.

HUDSON, R. R., and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

HUDSON, R. R., and N. L. KAPLAN, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

HUDSON, R. R., M. KREITMAN and M. AGUADE, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

HUDSON, R. R., K. BAILEY, D. SKARECKY, J. KWIATOWSKI and F. J. AYALA, 1994 Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. Genetics **136:** 1329–1340.

HUGHES, A. L., and M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

INA, Y., 1992 *Oden*. National Institute of Genetics, Mishima, Japan.

INNAN, H., F. TAJIMA, R. TERAUCHI and N. T. MIYASHITA, 1996 Intragenic recombination in the *Adh* locus of a wild plant *Arabidopsis thaliana*. Genetics **143:** 1761–1770.

JUKES, T. H., and C. R. CANTOR, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism*, edited by H. MUNRO. Academic Press, New York.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

KREITMAN, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*. Nature **304:** 412–417.

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and adh-dup loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

LAL, S. K., and SACHS, M. M., 1995 Cloning and characterization of an anaerobically induced cDNA encoding glucose-6-phosphate isomerase from maize. Plant Physiol. **108:** 1295–1296.

MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

MIYASHITA, N., H. INNAN and R. TERAUCHI, 1996 Intra- and interspecific variation in the alcohol dehydrogenase locus region of wild plants *Arabis gemmifera* and *Arabidopsis thaliana*. Mol. Biol. Evol. **13:** 433–436.

Miyata, T., S. Miyazawa and T. Yasunaga, 1979 Two types of amino acid substitutions in protein evolution. J. Mol. Evol. **12:** 219–236.

Moriyama, E., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in *Drosophila*. Mol. Biol. Evol. **13:** 261–277.

Nei, M., 1973 Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA **70:** 3321–3323.

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.

Noltmann, E. A., 1972 Aldose-ketose isomerase, pp. 271–354 in *The Enzymes,* Vol. VI., edited by P. D. Boyer. Academic Press, New York.

Nozue, F., M. Umeda, Y. Nagamura, Y. Minobe and H. Uchimiya, 1996 Characterization of cDNA encoding for phosphoglucose isomerase of rice (*Oryza sativa* L.). DNA Sequence **6:** 127–135.

Ohta, T., 1976 Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Popul. Biol. **10:** 254–275.

Ohta, T., 1993 Amino acid substitution at the *Adh* locus of *Drosophila* is facilitated by small population size. Proc. Natl. Acad. Sci. USA **90:** 4548–4551.

Patarnello, T., and B. Battaglia, 1992 Glucosephosphate isomerase and fitness: effects of temperature on genotype dependent mortality and enzyme activity in two species of the genus *Gammarus* (Crustacea: Amphipoda). Evolution **46:** 1568–1573.

Rand, D. M., M. Dorfsman and L. M. Kann, 1994 Neutral and nonneutral evolution of Drosophila mitochondrial DNA. Genetics **138:** 741–756.

Riddoch, B. J., 1993 The adaptive significance of electrophoretic mobility in phosphoglucose isomerase (PGI). Biol. J. Linn. Soc. **50:** 1–17.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sharp, P. M., and W-H. Li, 1989 On the rate of DNA sequence evolution in *Drosophila.* J. Mol. Evol. **28:** 398–402.

Sharp, P. M., and G. Matassi, 1994 Codon usage and genome evolution. Curr. Opin. Genet. Dev. **4:** 851–860.

Shaw, P. J., and H. Muirhead, 1977 Crystallographic structure analysis of glucose-6-phosphate isomerase at 3.5 Å resolution. J. Mol. Biol. **109:** 475–485.

Shields, D. C., P. M. Sharp, D. G. Higgins and F. Wright, 1988 'Silent' sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5:** 704–716.

Stephens, J. C., 1985 Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. Mol. Biol. Evol. **2:** 539–556.

Strobeck, C., 1983 Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. Genetics **103:** 545–555.

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

Tajima, F., 1989b The effect of change in population size on DNA polymorphism. Genetics **123:** 597–601.

Terauchi, R., 1990 Genetic diversity and population structure of *Dioscorea tokoro* Makino, a dioecious climber. Plant Species Biol. **5:** 243–253.

Terauchi, R., and A. Konuma, 1994 Microsatellite polymorphism in *Dioscorea tokoro,* a wild yam species. Genome **37:** 794–801.

Terauchi, R., T. Terachi and N. T. Miyashita, 1996 DNA polymorphism and evolutionary history of the *Pgi* locus in *Dioscorea tokoro,* a wild yam species, pp. 245–248 in *Current Topics on Molecular Evolution,* edited by M. Nei and N. Takahata. Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, State College, PA, and Graduate School for Advanced Studies, Hayama, Japan.

Thomas, B. R., D. Laudencia-Chingguanco and L. D. Gottlieb, 1992 Molecular analysis of the plant gene encoding cytosolic phosphoglucose isomerase. Plant Mol. Biol. **19:** 745–757.

Thomas, B. R., V. S. Ford, E. Pichersky and L. D. Gottlieb, 1993 Molecular characterization of duplicate cytosolic phosphoglucose isomerase genes in Clarkia and comparison to the single gene in Arabidopsis. Genetics **135:** 895–905.

Watt, W. B., 1977 Adaptation at specific loci. I. Natural selection on phosphoglucose isomerase of Colias butterflies: biochemical and population aspects. Genetics **87:** 177–194.

Watterson, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

Watterson, G. A., 1978 The homozygosity test of neutrality. Genetics **88:** 405–417.

Communicating editor: G. B. Golding