Genet. Res., Camb. (1991), **58**, *pp.* 145–156 *With 7 text-figures Printed in Great Britain*

A method for detecting effect of beneficial mutations in natural populations of *Drosophila melanogaster*

A. KOGA*[1], S. KUSAKABE[1], F. TAJIMA[2], T. TAKANO, K. HARADA AND T. MUKAI
Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

(*Received 17 August 1990 and in revised form 12 March 1991*)

Summary

An experimental method is proposed for detecting the effects of positive natural selection on DNA polymorphisms. Since beneficial mutations are expected to increase in frequency faster than neutral mutations, variants which have reached high frequencies in a relatively short period could be linked to some beneficial mutation. *D. melanogaster* has a cosmopolitan polymorphic inversion -*In(2L)t* - whose age in some local populations has been estimated. Setting the age of *In(2L)t* as the upper limit for the age of variants, we searched for variants whose frequencies were possibly influenced by positive natural selection. We detected a single candidate whose frequency and distribution met the requirements imposed by our method.

1. Introduction

There is still a vast gap between current knowledge of molecular evolution and phenotypic adaptation. The study of beneficial mutations has been a key to understanding the relationship between molecular evolution and phenotypic adaptation. There are many examples of beneficial mutations: insecticide resistence (Weiner & Crow, 1951), allosteric proteins (Perutz, 1983), immunosystems (Hughes & Nei, 1988), G-protein-coupled receptors (Yokoyama, Isenberg & Wright, 1989), and industrial melanism (Kettlewell, 1965). Of these examples, only the last provides really clear evidence for *rapid increase* in gene frequency under *natural circumstances*. Obtaining other examples will further our understanding of the molecular basis of phenotypic adaptation. In this paper, we describe an experimental method to detect variants showing specific features in their distribution that indicate rapid increase in frequency. Although variants that display such features are not always due to beneficial mutations, the ratio of beneficial variants to neutral or deleterious variants will be much higher than the ratio in randomly collected samples. Our method allows us, on some empirically acceptable assumptions, to detect effects of potentially beneficial mutations. Because we observe genetic variation directly at the DNA level, we can find the effect of beneficial mutations irrespective of their phenotypes, i.e. all types of positive natural selection are objects of the method. We describe the procedure and the first candidate whose distribution has possibly been influenced by a beneficial mutation.

* Corresponding author.
Present addresses:
[1] Department of Biology, College of Arts and Sciences, Hiroshima University, Hiroshima 730, Japan.
[2] National Institute of Genetics, Mishima 411, Japan.

2. Strategy

(i) *Assumptions*

We make two assumptions with respect to the history of the natural population of *D. melanogaster* analysed in the present study (Raleigh, NC): (1) The effective population size is large (of the order of 10^5 or more), (2) The age of *In(2L)t* (number of generations after its invasion) is small (on the order of 10^3). We suppose that these two assumptions are acceptable based on our previous investigations of the Raleigh and other natural populations. Detailed evaluation of the assumptions will be given in section 5.

(ii) *Basic criteria*

It has been shown theoretically that it takes a long time for neutral mutations to reach high frequencies in a large population (Kimura, 1983*a*). Beneficial mutations should increase in frequency over a time-span much shorter than that for neutral mutations. Therefore, we establish the following two criteria: (1) high frequency of a variant in a population, and (2) evidence that the high frequency was attained rapidly.

When a variant in a large population is found to meet the two criteria, it can be taken as a candidate for a positively selected variant.

(iii) *Application to* In(2L)t

It is easy to check results of a restriction survey against the first criterion because the survey brings us the frequency of each variant. In order to meet the second criterion, we utilized *In(2L)t*, whose age in some local populations is known. The estimated age of the inversion is about 4100 generations in the Raleigh population (Mukai & Voelker, 1977), and about 1000 generations in the Okinawa (Ishigakijima), Japan, population (Mukai, Tachida & Ichinose, 1980). Because an inversion will suppress single recombination in its inner and neighbouring regions when it is heterozygous with an inversion-free chromosome, alleles at loci in such regions should remain tightly linked to the inversion for a long period. We conducted a survey of genetic variation in the region surrounding the *Adh* gene. The *Adh* gene (35A4-B1, Chia *et al.* 1985) is located very close to the proximal breakpoint of *In(2L)t* (22D3-E1; 34A8-9, Lindsley & Grell, 1968), and their recombination distance is practically zero.

Suppose that a variant is found to be common in *In(2L)t*-carrying chromosomes and to be non-existent in *In(2L)t*-free chromosomes. The variant must have arisen on an *In(2L)t*-carrying chromosome. Therefore, the age of the variant cannot be longer than the age of *In(2L)t* in the same population. Since the estimated age of *In(2L)t* in the Raleigh population (on the order of 10^3) is not long enough for a neutral mutation to reach a high frequency in a large population (see section 5 for quantitative consideration), it is likely that the increase in frequency was caused by positive natural selection.

(iv) *Possibility of founder effect*

Even if a variant meets the two criteria stated above, another explanation is possible: the variant may have arisen when the frequency of *In(2L)t* was still low. That is, the frequency of the variant in *In(2L)t*-carrying chromosomes may have been high ever since it arose. Information about the extent of variation in neighbouring populations would aid in determining the contribution of founder effect in producing the present-day distribution. If the variant occurred when the frequency of *In(2L)t* was low, i.e. if the age of the variant is as great as the age of *In(2L)t*, and there is constant migration of *In(2L)t* within the whole population, the variant would have migrated to neighbouring populations. Therefore, the absence of the variant in neighbouring populations suggests that the age of the variant is less than the age of *In(2L)t*. An inversion that becomes polymorphic tends to increase rather rapidly in frequency and to reach a plateau (cf. Nei, Kojima & Schaffer, 1977; Yamaguchi *et al.* 1980). The greater the difference between the ages of *In(2L)t* and the variant, the smaller the possible influence of the founder effect on their relative frequencies.

(v) *Practical criteria*

Following the arguments stated above, we used three practical criteria to identify a variant whose frequency could be influenced by a beneficial mutation: (1) the variant is observed at a high frequency in *In(2L)t*-carrying chromosomes, (2) the variant cannot be found in *In(2L)t*-free chromosomes, and (3) the variant is observed only in a local population (a geographically restricted subset of the entire species).

3. Materials and methods

(i) *Populations*

Adult flies totalling 300–500 were collected from the following locations: Raleigh, NC, in 1984 and 1987; Austin, TX, in 1985; Okinawa, Japan, in 1984; and Osaka, Japan, in 1984. Thus, we prepared five sets of samples from four local populations.

(ii) *Extraction of chromosomes*

From isofemale lines, single second chromosomes were extracted using the marker chromosomes *In(2LR)SM1* and *In(2LR)bw*^{V1}. The mating scheme for the extraction is shown in Fig. 1. These second chromosomes were maintained as chromosome lines balanced with the *Cy* (*In(2LR)SM1*) chromosome so as to maintain the less viable or lethal chromosomes.

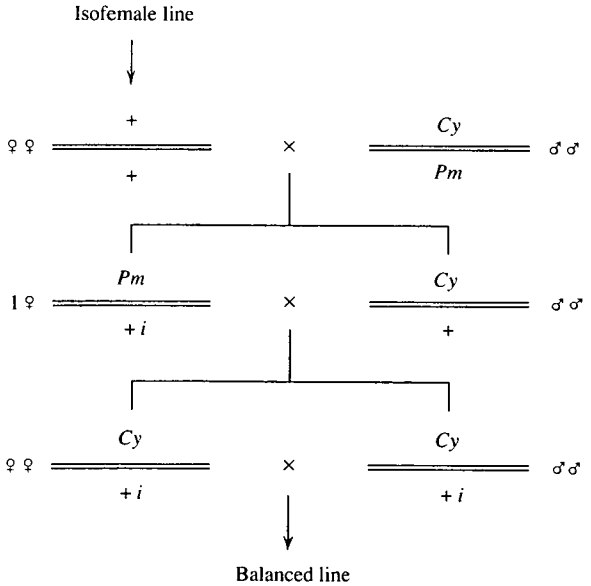

Fig. 1. Mating scheme for the extraction of second chromosomes. The symbol '$+_i$' represents a single second chromosome that was duplicated in the ith balanced line. The symbol '+' stands for any second chromosomes included in the ith isofemale line.

https://doi.org/10.1017/S0016672300029803 Published online by Cambridge University Press

Salivary gland chromosomes of each line were examined for chromosomal rearrangements (cf. Mukai & Yamaguchi, 1974). The chromosome lines were classified into two groups based on the presence of *In(2L)t*. The group without *In(2L)t* is abbreviated as *St*, and those with *In(2L)t* as *In*. The numbers of chromosome lines selected for analyses (shown in Table 7) do not reflect the frequency of *In(2L)t* in their original populations. However, the *St* and the *In* chromosomes used are random samples from *In(2L)t*-free and *In(2L)t*-carrying chromosomes, respectively, in natural populations. The electrophoretic mobility (*F* or *S*) of the *Adh* protein was also determined using 0·8 % agarose gels.

(iii) *Probes*

We isolated two recombinant bacteriophages carrying the *Adh* region from a genomic library of a laboratory strain (C160) of *D. melanogaster*. The probe for this isolation was a 2·7 kb *Cla* I-*Sal* I DNA fragment containing the coding region of the *Adh* gene (Kreitman, 1983). The isolated fragments were partitioned into four overlapping parts and subcloned into pUC13. These subclones were used as probe in the Southern hybridizations and cover the region shown in Fig. 2.

(iv) *Southern hybridization*

Genomic DNA was extracted from whole bodies of adult flies with a method outlined by Bingham, Levis & Rubin (1981). The DNA was digested completely with hexanucleotide-specific restriction endonucleases. Eight enzymes (*Bam*H I, *Eco*R I, *Hind* III, *Pst* I, *Sac* I, *Sal* I, *Xba* I and *Xho* I) were used for two sets of samples (1984 Raleigh and Okinawa). The remaining three sets of samples (1987 Raleigh, Austin and Osaka) were treated only with *Hind* III and *Sal* I. Southern blotting to nylon membranes and hybridization with the ^{32}P-labelled probe DNA were carried out (cf. Southern, 1975; Maniatis, Fritsch & Sambrook, 1982; Feinberg & Vogelstein, 1983). Restriction patterns obtained for each chromosome line were analysed for insertion or deletion of DNA fragments and for gain or loss of restriction sites.

(v) *Cloning, sequencing and* in situ *hybridization*

The following procedures were applied to a candidate obtained by the restriction mapping: cloning and subcloning into the bacteriphage λCharon 35 and the plasmid pUC13 (Maniatis, *et al.* 1982), dideoxy DNA sequencing using ^{32}P (Sanger, Nicklen & Coulson, 1977; Hattori & Sakaki, 1986), and *in situ* hybridization to salivary gland chromosomes with biotinylated probe DNA (Pardue & Gall, 1975; Rigby, Dieckmann & Rhodes, 1977).

4. Results

(i) *General features of the distribution of variation*

Samples from the 1984 Raleigh population and from the Okinawa population were used in a detailed survey of restriction fragment variation. The variation

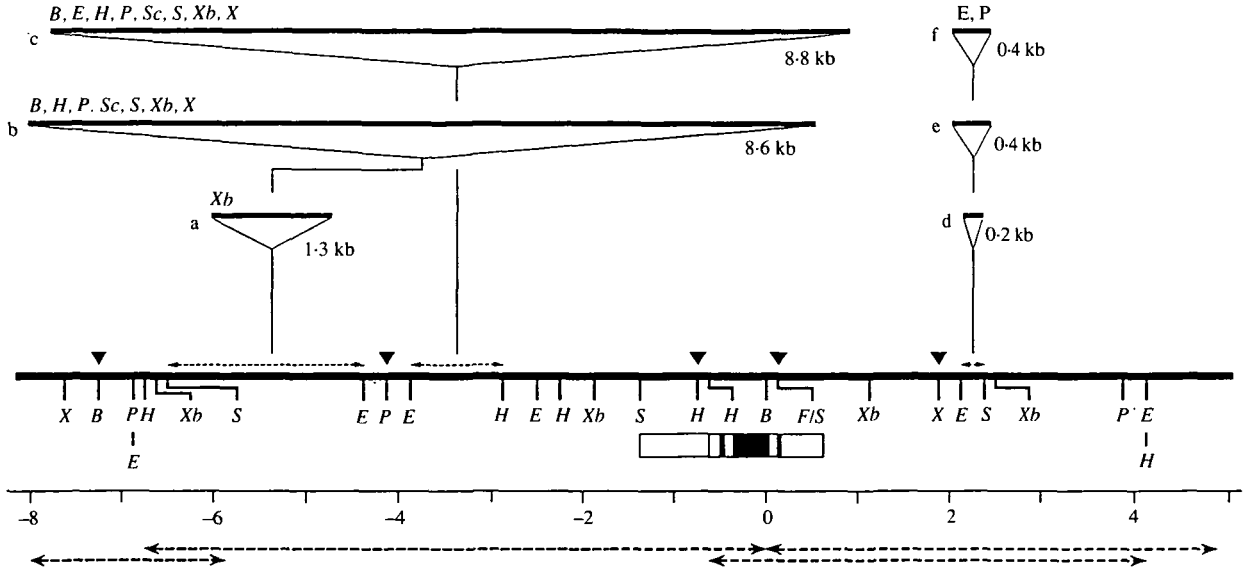

Fig. 2. Variation detected in 1984 Raleigh *St* samples. The symbols for restriction endonucleases are as follows: *B*, *Bam*H I; *E*, *Eco*R I; *H*, *Hind* III; *P*, *Pst* I; *Sc*, *Sac* I, *S*, *Sal* I; *Xb*, *Xba* I and *X*, *Xho* I. Polymorphic restriction sites and the *Adh* isoalleles are indicated by closed downward triangles. Insertions and deletions are shown by downward and upward triangles, respectively, with their names (a, b, etc.) and lengths. Symbols for restriction sites on insertions mean that they appear more than once in the insertions. Double-headed arrows along the map show restriction fragments in which insertions or deletions are included. The coding region of the *Adh* gene and the position of the nucleotide substitution leading to the *Adh* allozymes are based on data by Kreitman (1983). Double-headed arrows under the map show regions covered by the probes for the survey.

https://doi.org/10.1017/S0016672300029803 Published online by Cambridge University Press

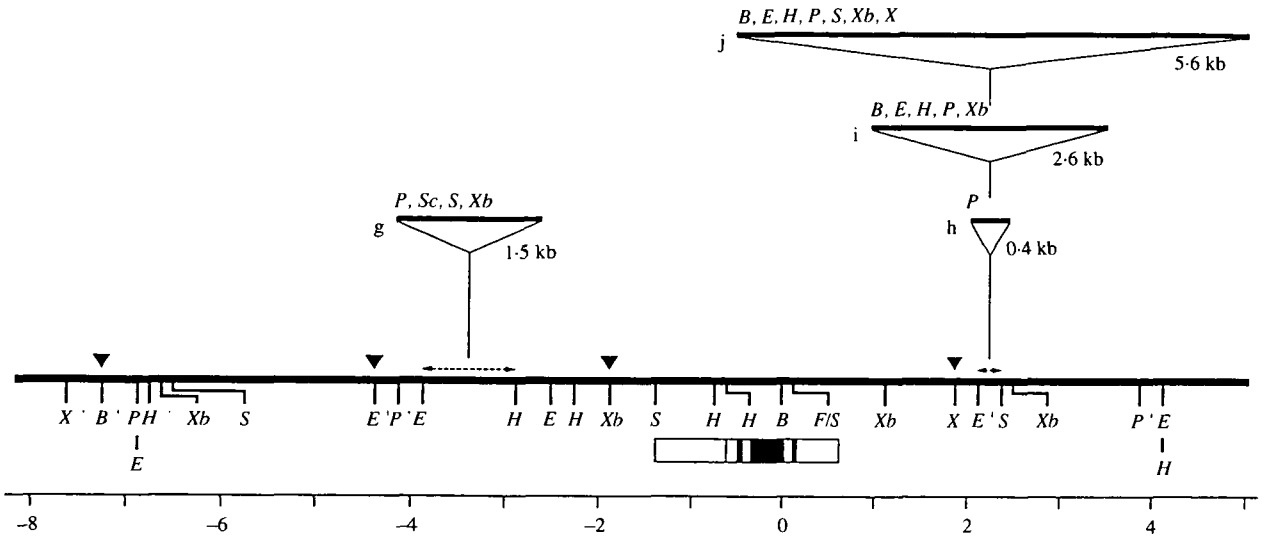

Fig. 3. Variation detected in 1984 Raleigh *In* samples. See legend for Fig. 2.

detected is schematically summarized in Figs 2–5, and the distribution of variation among the chromosome lines is shown in Tables 1–4. The eight restriction enzymes produced 26 restriction sites in the *Adh* region. The nucleotide diversity (Nei & Tajima, 1981) in the *Adh* region was estimated for the four groups separately (Table 5). Only the restriction site variation was used in the calculation for the estimates. In both populations, estimates for *In* are lower than estimates for *St*. The *Bam*H I site at −7·2 kb apparently contributes to most (1984 Raleigh) or all (Okinawa) of the nucleotide diversity in *In*. The high polymorphism at this site is a phenomenon common to the two populations (and also to the 1987 Raleigh, Austin and Osaka populations, data not shown). We infer that the *Bam*H I site or its alternative which had not been linked to *In(2L)t* was originally introduced into *In(2L)t*-carrying chromosomes by recombination soon

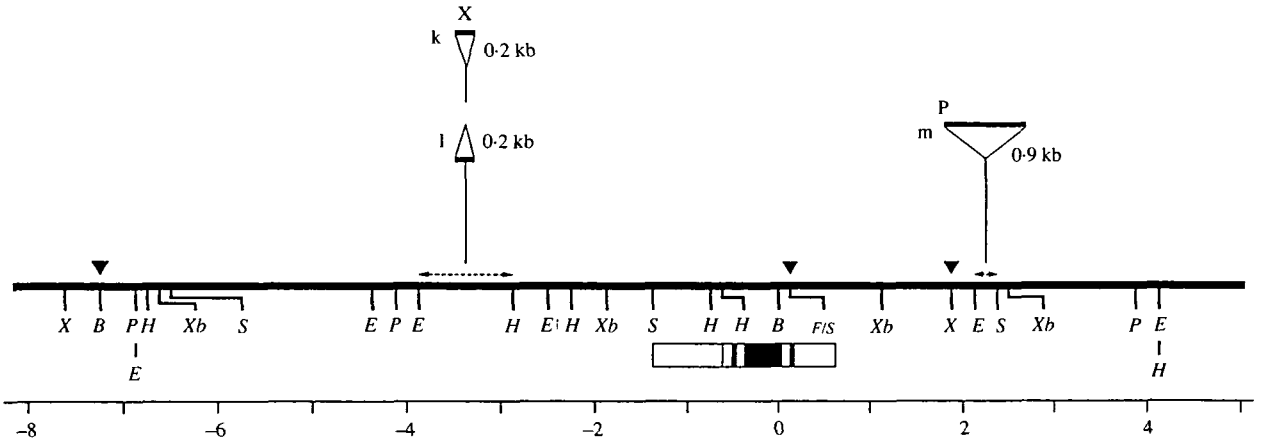

Fig. 4. Variation detected in Okinawa *St* samples. See legend for Fig. 2.

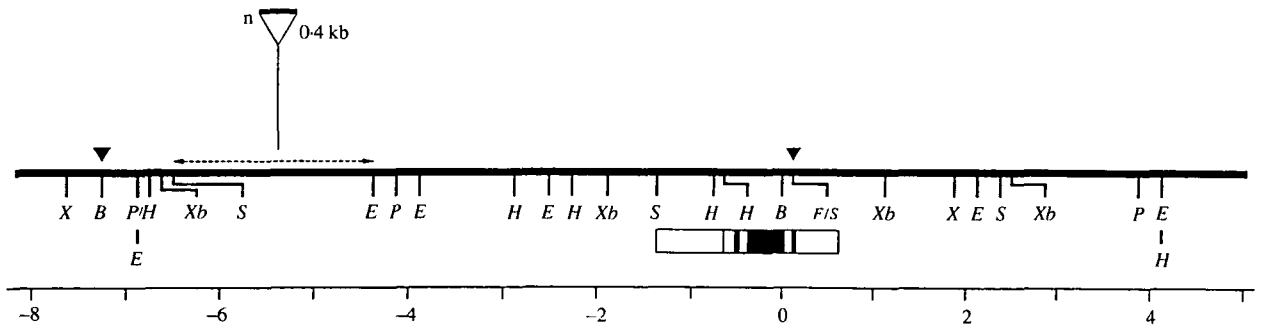

Fig. 5. Variation detected in Okinawa *In* samples. See legend for Fig. 2.

https://doi.org/10.1017/S0016672300029803 Published online by Cambridge University Press

Table 1. *Restriction map variants in 1984 Raleigh* St *samples*

Line	−7·2 B	−5·3 a	−5·3 b	−5·3 n	−4·3 E	−4·1 P	−3·4 c	−3·4 g	−3·4 k	−3·4 l	−1·9 Xb	−0·7 H	0·1 *Adh*	1·9 X	2·2 d	2·2 e	2·2 f	2·2 h	2·2 i	2·2 j	2·2 m	Haplo-type
R21	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R62	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R64	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R71	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R73	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R74	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R75	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R285	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R290	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R293	+	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	1*
R36	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R49	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R50	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R61	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R68	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R77	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R80	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R33	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R280	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R292	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R296	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R78	−	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4
R275	−	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4
R291	−	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4
R295	−	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4
R31	+	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	5*
R32	+	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	5*
R276	+	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	5*
R48	+	a	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	6*
R286	+	a	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	6*
R47	−	·	·	·	+	+	c	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	7*
R56	−	·	·	·	+	+	c	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	7*
R287	+	·	·	·	+	+	·	·	·	·	+	−	*S*	+	·	·	·	·	·	·	·	8*
R51	+	·	·	·	+	−	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	9*
R35	+	·	b	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	10*
R30	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	f	·	·	·	·	11*
R282	−	·	·	·	+	+	·	·	·	·	+	+	*F*	−	·	·	·	·	·	·	·	12*
R45	−	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	e	·	·	·	·	·	13*
R28	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	d	·	·	·	·	·	·	14*

* Indicates haplotypes that are not observed in other groups.

after *In(2L)t* emerged (sequence information supports this conclusion, M. Baba and T. Mukai, in preparation.). If the *Bam*H I site is excluded, then the nucleotide diversity becomes 0·0008 for 1984 Raleigh *In*.

We classified the variation into three categories: site variation, the *Adh* isoalleles and length variation. We estimated the nucleon diversity (Nei & Tajima, 1981) in the *Adh* region, using different subsets of variation so that it can be compared with the estimates reported by other authors (Table 6). Again, the estimates for *In* are lower than those for *St* in all the subsets and in both populations. The estimates for *St* agree with previous estimates for the same region (Langley, Montgomery & Quattlebaum, 1982; Birley, 1984; Cross & Birley, 1986; Aquadro *et al.* 1986; Kreitman & Aguadé, 1986; Aguadé, 1988).

(ii) *Haplotype 15 as a candidate*

Haplotype 15 meets two of the three criteria described in the Strategy section: (1) its frequency in *In* is high (6/33, or 18%) in the 1984 Raleigh population, and (2) it does not appear in *St* in the same population. To determine whether it also meets the last criterion, we examined its existence in other samples (1987 Raleigh, Austin and Osaka). Two enzymes (*Hin*d III and *Sal* I) were applied to these samples. For chromosome lines showing restriction patterns identical to haplotype 15, the other six enzymes were also used. Table 7 shows the observed numbers of haplotype 15 as well as the numbers of chromosome lines examined. Haplotype 15 was present only in the *In* samples of the 1987 Raleigh population. Since it met all three criteria, we took this haplotype as a prime candidate for a positively selected variant. Some additional data

Table 2. *Restriction map variants in 1984 Raleigh* In *samples*

Line	−7·2 B	−5·3 a	−5·3 b	−5·3 n	−4·3 E	−4·1 P	−3·4 c	−3·4 g	−3·4 k	−3·4 l	−1·9 Xb	−0·7 H	0·1 *Adh*	1·9 X	2·2 d	2·2 e	2·2 f	2·2 h	2·2 i	2·2 j	2·2 m	Haplo type
R22	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R95	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R115	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R120	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R124	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R244	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R268	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R301	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R353	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R373	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R381	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R393	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R429	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R437	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R544	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R585	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
R79	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	i	·	·	15*
R144	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	i	·	·	15*
R303	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	i	·	·	15*
R339	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	i	·	·	15*
R341	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	i	·	·	15*
R375	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	i	·	·	15*
R7	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R57	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R231	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R235	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R394	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
R335	−	·	·	·	+	+	·	g	·	·	+	+	*S*	+	·	·	·	·	·	·	·	16*
R336	−	·	·	·	+	+	·	g	·	·	+	+	*S*	+	·	·	·	·	·	·	·	16*
R510	−	·	·	·	+	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	17
R148	−	·	·	·	−	+	·	·	·	·	+	+	*S*	−	·	·	·	·	·	·	·	18*
R142	−	·	·	·	+	+	·	·	·	·	−	+	*S*	+	·	·	·	h	·	·	·	19*
R408	−	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	j	·	20*

* Indicates haplotypes that are not observed in other groups.

eliminate the possibilities that our findings result from biased sampling or that haplotype 15 has multiple origins.

(iii) *Allelism of recessive lethal genes*

Of the six chromosomes containing haplotype 15, two (R339 and R341) carry recessive lethals. Crosses were made between *Cy*/R339 and *Cy*/R341. Their offspring were scored with respect to the phenotype of the *Cy* locus. About one third of the adult flies had wild-type wings in both of the reciprocal crosses (the *Cy* chromosome carries a recessive lethal allele). Therefore, the recessive lethals of R339 and R341 are nonallelic. Dysgenic crosses in the sense of P–M system are not included in the extraction procedure (see Fig. 1) because the isofemale lines from the Raleigh population had been shown to have the P cytotype (K. Harada, S. Kusakabe and T. Mukai, unpublished data). The two lethal mutations must have occurred independently during the period in which the chromosomes of haplotype 15 were diverging. This is an evidence against the possibility that we may have collected closely related chromosomes by chance.

(iv) *Distribution of* copia *elements*

The distribution of *copia* elements among the six chromosomes was examined by *in situ* hybridization. The probe was the plasmid clone cDm2055 (Saigo, Millstein & Thomas, 1981). Table 8 presents the location of the bands that showed positive hybridization signals. None of the pairs of the six chromosomes show similar patterns in the distribution of the *copia* elements. This is further evidence against the possibility of biased sampling.

(v) *Origin of the haplotype*

Haplotype 15 as characterized has some site variation, the *Adh* isoallele and insertion 'i'. Of these three kinds of variation, the insertion element has the highest possibility for the cause for multiple origin of the haplotype, because some transposable elements are known to have insertional site specificity. Insertion 'i'

https://doi.org/10.1017/S0016672300029803 Published online by Cambridge University Press

Table 3. *Restriction map variants in Okinawa* St *samples*

Line	−7·2 B	−5·3 a	−5·3 b	−5·3 n	−4·3 E	−4·1 P	−3·4 c	−3·4 g	−3·4 k	−3·4 l	−1·9 Xb	−0·7 H	0·1 *Adh*	1·9 X	2·2 d	2·2 e	2·2 f	2·2 h	2·2 i	2·2 j	2·2 m	Haplo-type
I483	–	·	·	·	+	+	·	·	·	l	+	+	*F*	+	·	·	·	·	·	·	·	21*
I510	–	·	·	·	+	+	·	·	·	l	+	+	*F*	+	·	·	·	·	·	·	·	21*
I513	–	·	·	·	+	+	·	·	·	l	+	+	*F*	+	·	·	·	·	·	·	·	21*
I517	–	·	·	·	+	+	·	·	·	l	+	+	*F*	+	·	·	·	·	·	·	·	21*
I521	–	·	·	·	+	+	·	·	·	l	+	+	*F*	+	·	·	·	·	·	·	·	21*
I481	+	·	·	·	+	+	·	·	k	·	+	+	*S*	–	·	·	·	·	·	·	·	22*
I507	+	·	·	·	+	+	·	·	k	·	+	+	*S*	–	·	·	·	·	·	·	·	22*
I515	+	·	·	·	+	+	·	·	k	·	+	+	*S*	–	·	·	·	·	·	·	·	22*
I482	–	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4
I496	–	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4
I497	–	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4
I495	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I509	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I487	–	·	·	·	+	+	·	·	·	·	+	+	*S*	–	·	·	·	·	·	·	·	17
I520	–	·	·	·	+	+	·	·	·	·	+	+	*S*	–	·	·	·	·	·	·	·	17
I511	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	m	23*

* Indicates haplotypes that are not observed in other groups.

Table 4. *Restriction map variants in Okinawa* In *samples*

Line	−7·2 B	−5·3 a	−5·3 b	−5·3 n	−4·3 E	−4·1 P	−3·4 c	−3·4 g	−3·4 k	−3·4 l	−1·9 Xb	−0·7 H	0·1 *Adh*	1·9 X	2·2 d	2·2 e	2·2 f	2·2 h	2·2 i	2·2 j	2·2 m	Haplo-type
I485	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I493	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I530	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I541	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I583	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I597	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I604	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I614	+	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	2
I489	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I531	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I550	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I551	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I582	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I607	–	·	·	·	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	3
I601	–	·	·	n	+	+	·	·	·	·	+	+	*S*	+	·	·	·	·	·	·	·	24*
I501	–	·	·	·	+	+	·	·	·	·	+	+	*F*	+	·	·	·	·	·	·	·	4

* Indicates haplotypes that are not observed in other groups.

is located at position 2·2 kb (1·7 kb downstream from the proximal end of the *Adh* coding region). This is a member of the transposable element family called *BN* (Harada *et al.* 1988). The *BN* element shows no marked target sequence specificity (Harada *et al.* 1988). Taking this into account, we examined whether insertion 'i' is located at an identical position among the 'i'-carrying chromosomes. As mentioned above, two of the six chromosomes carry recessive lethals. Because lethal chromosomes cannot be made homo-

Table 5. *Nucleotide diversity in the* Adh *region*

1984 Raleigh		Okinawa	
St	*In*	*St*	*In*
0·0035 (39)	0·0025 (33)	0·0029 (16)	0·0018 (16)

Numbers of chromosome lines are shown in parentheses.

Table 6. *Nucleon diversity in the* Adh *region*

	1984 Raleigh		Okinawa	
Subset*	*St*	*In*	*St*	*In*
S and I	0·816	0·600	0·725	0·592
S, I and L	0·888 (39)	0·723 (33)	0·850 (16)	−0·642 (16)

* S, site variation; I, the *Adh* isoalleles; L, length variation. Numbers of chromosomes are shown in parentheses.

https://doi.org/10.1017/S0016672300029803 Published online by Cambridge University Press

Table 7. *Frequency of haplotype 15 in the samples from the natural populations*

	1984 Raleigh	1987 Raleigh	Austin	Okinawa	Osaka
St	0/39	0/30	0/20	0/21	0/29
In	6/33	6/32	0/9	0/17	0/22

Numerators and denominators represent the numbers of chromosome lines carrying haplotype 15 and the total examined, respectively.

zygous, gene libraries were constructed using individuals heterozygous with the *Cy* chromosomes. Only one of the two lethal chromosomes (R341) was analysed in this study. From each of the five genomic libraries, a DNA fragment including insertion 'i' was cloned. Fig. 6 shows the detailed restriction maps of these clones and that of the probe DNA (originating from the 'i'-free strain C160). The restriction maps are identical among the five 'i'-carrying chromosomes. We examined the base sequences of the regions indicated by the arrows in Fig. 6. The results are shown in Fig. 7. Comparing the base sequences of C160 with that of R375, we determined the exact insertion point and a 12 bp-long duplicated sequence which originated from the host chromosome. The insertion points are also identical among the five 'i'-carrying chromosomes, supporting a single origin for the insertion carried by the different second chromosomes.

5. Discussion

Our method for detecting effects of positive natural selection relies upon two assumptions; (1) the effective population size is large, and (2) the age of *In(2L)t* is small. We first describe evidence that these assumptions are acceptable, and then show a statistical test using these estimates. Finally, we consider alternative hypotheses.

(i) *Effective population size*

As our null hypothesis, haplotype 15 is assumed to be neutral. The effective size for neutral genes of the Raleigh population can be estimated as follows: electromorph mutation rate (excluding null mutations) was estimated to be $7{\cdot}57 \times 10^{-7}$ per locus per generation after accumulating 5285615 allele-generations (Mukai *et al.* 1990). Assuming that the detection rate in electrophoresis is roughly $\frac{1}{3}$ (Shaw, 1965) and that the proportion of selectively neutral mutations among electromorph mutations, excluding null mutations, is 0·14 (Kimura, 1983*b*), the neutral mutation rate is

Table 8. *Distribution of* copia *elements on the second chromosomes carrying haplotype 15 in the 1984 Raleigh* In *samples*

Line	22E	22F	26C	26F	27C	30B	32C	33B	35B	38B	38C	39B	40A	41F	42B	42D	47C	48B	48C	49B	51B	55A	55C	56A	56B	59D
R79								+	+			+			+		+							+		
R144			+					+	+	+						+	+					+	+			
R303		+			+						+		+	+			+			+			+			
R339						+																				
R341							+											+			+				+	+
R375		+		+					+						+	+			+						+	

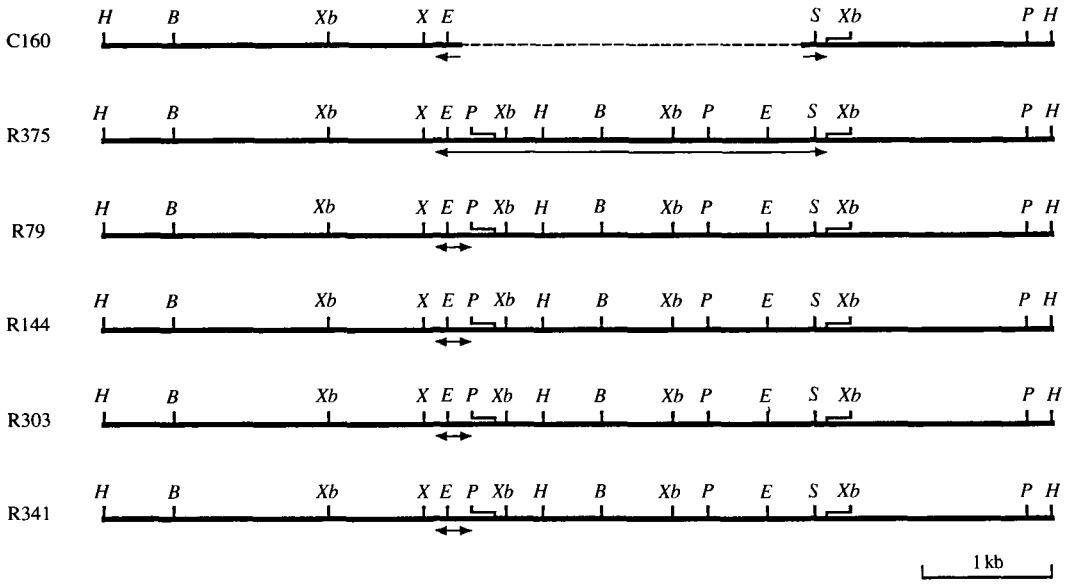

Fig. 6. Detailed restriction maps of the clones from 'i'-free and 'i'-carrying chromosomes. C160 is an 'i'-free laboratory strain. Others are 'i'-carrying chromosome lines in 1984 Raleigh *In* samples. The arrows under the maps indicate the regions sequenced.

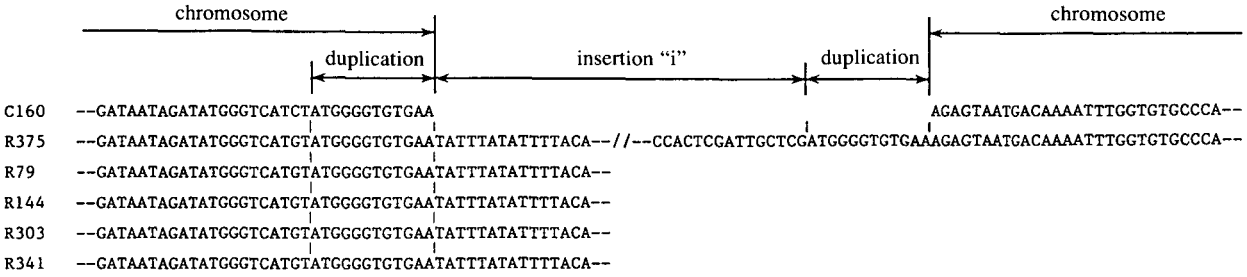

Fig. 7. Base sequences of the regions around the ends of insertion 'i'. See legend for Fig. 6.

calculated to be $3{\cdot}18 \times 10^{-7}$ per locus per generation. The average number of base pairs in exons of the genes studied is approximately 1000 (cf. Mukai & Cockerham, 1977), resulting in a neutral mutation rate per base pair of $3{\cdot}18 \times 10^{-10}$ per generation. In the present experiment, nucleotide diversity (π of Nei & Tajima, 1981) was estimated to be 0·0035 (the Raleigh population), which should be equal to $4N_e\mu$ for selectively neutral mutations where N_e is the effective population size of the population and μ is the neutral mutation rate per base pair. From this relationship, N_e can be estimated to be $2{\cdot}75 \times 10^6$.

The effective population size for genes linked to *In(2L)t* may be roughly estimated by multiplying the above N_e by the frequency of *In(2L)t*. The estimated frequency of *In(2L)t* is $0{\cdot}068 \pm 0{\cdot}005$ ($n = 2367$), which was obtained by pooling the data of 1968, 1969, 1970, 1974 and 1984 Raleigh population (Mukai, Watanabe & Yamaguchi, 1974; Mettler, Voelker & Mukai, 1977; T. Mukai and S. Kusakabe, unpublished). Thus, the effective size in question is estimated to be $1{\cdot}87 \times 10^5$.

(ii) *Age of* In(2L)t

Mukai & Voelker (1977) estimated the age of *In(2L)t* in the Raleigh population to be 4100 generations, using the decay of linkage disequilibrium (cf. Nei & Li, 1980) between *In(2L)t* and the *Gpdh* genes. In addition, Mukai *et al.* (1980) estimated the age of *In(2L)t* in the Okinawa population to be 1000 generations in exactly the same way. For the Okinawa population, we have more information on genetic variation, which supports the reliability of the estimate: if we accumulate mutations affecting sternopleural and abdominal bristle numbers under natural selection pressure, it takes 1000–2000 generations until the population reaches equilibrium (Durrant & Mather, 1954; Clayton & Robertson, 1955; Mukai *et al.* 1980). In the case of viability, it takes 100–200 generations before reaching genetic equilibrium (Mukai *et al.* 1972). Mukai *et al.* (1980) estimated genetic variance for sternopleural and abdominal bristle numbers and for viability, with the samples from the Okinawa population. The former was significantly smaller in *In(2L)t*-carrying chromosomes

https://doi.org/10.1017/S0016672300029803 Published online by Cambridge University Press

than in standard chromosomes, while there was no significant difference in the latter. Thus, a value of 1000 for the age of *In(2L)t* appears to reflect the actual situation. The same conclusion may be applied to the Raleigh population since the same recombination value ($r = 0{\cdot}00022$) was employed for the estimation of the age of *In(2L)t*.

It is possible that *In(2L)t* passed through a bottleneck after it invaded the Raleigh population, but there is no evidence for past fluctuations in population size of *In(2L)t*-carrying chromosomes. Even if *In(2L)t* experienced such a situation, the estimated age reflects the time after the last expansion in population size because it was estimated by the decay of linkage disequilibrium.

(iii) *Statistical test*

The null hypothesis is that haplotype 15 was selectively neutral and its high frequency was attained only by random genetic drift. Formula 8·37 in Kimura (1983*a*) is suitable for this test, giving the probability density function of the age of a neutral allele at a fixed present frequency. The age is defined as the number of generations for which the allele has persisted in a population after it arose by mutation (Kimura, 1983*a*). Let x, t and N_e be the present frequency of the allele, the age of the allele and the effective population size, respectively. The formula is:

$$f_x(\lambda) = 2x \sum_{i=1}^{\infty} (2i+1)\, T_{i-1}^1(z)\, e^{-i(i+1)\lambda},$$

where $\lambda = t/(4N_e)$, $z = 1-2x$ and $T_{i-1}^1(\cdot)$ is the Gegenbauer polynomial [see Kimura (1983*a*) for definition]. The formula measures time in units of $4N_e$ generations. Because the formula is a probability density function, we can obtain the probability (P) that the age of an allele is smaller than a given number of generations (t_1) by integrating the formula from 0 to $t_1/(4N_e)$ as follows:

$$P = \int_0^{t_1/(4N_e)} f_x(\lambda)\, d\lambda.$$

In the actual case, the present frequency of haplotype 15 in *In* was 0·18 (in 1984) and 0·19 (in 1987) and its average value is 0·18. Thus, we use a fixed value of $x = 0{\cdot}18$. It is reasonable to assign t_1 the age of *In(2L)t* because the age of haplotype 15 cannot be greater than the age of *In(2L)t*. We substituted different values around the estimate: t_1 = 1000, 2000, 4100, 10000, 20000 and 40000. Also for N_e, we used several values around the estimate: $N_e = 10^3$, 10^4, 10^5 and 10^6. Substituting these numbers and integrating the formula, we calculated the probability (P) that the age of haplotype 15 with its present frequenty (x) is less than the number of generations (t_1) at the effective population size (N_e) on the assumption that haplotype 15 is selectively neutral. The results are tabulated in Table 9. As expected, the probability decreases for smaller t_1 and for larger N_e. When the estimated values are employed ($t_1 = 4100$ and $N_e = 10^5$), P is very small as shown in Table 9. Thus, given that our estimates are correct and that the present high frequency is not caused by the founder effect, the probability that haplotype 15 is selectively neutral is very small.

Table 9. *Probability* (P) *at which the age of a neutral allele is smaller than the number of generations* (t_1) *at the effective population size* (N_e)

	N_e			
t_1	10^3	10^4	10^5	10^6
1000	0·52	< 0·001	< 0·001	< 0·001
2000	0·76	0·025	< 0·001	< 0·001
4100	0·93	0·175	< 0·001	< 0·001
10000	> 0·99	0·520	< 0·001	< 0·001
20000	> 0·99	0·76	0·025	< 0·001
40000	> 0·99	0·92	0·167	< 0·001

The present frequency (x) is fixed at $x = 0{\cdot}18$.

(iv) *Alternative hypotheses*

Our results suggest that haplotype 15 was under the influence of some beneficial mutation. Alternative explanations against this particular candidate are considered below. The first is, of course, the case where our assumptions (see section 2) are not correct. We have given our estimates of N_e and t_1 just before the statistical test. It is known that the expected amount of DNA polymorphism tends to be close to that of the overall population where migration occurs (Slatkin, 1987; Strobeck, 1987; Tajima, 1990). Therefore, our estimate of N_e may be an overestimate. The calculation for t_1 is based on our previous estimates of the recombination rate (r) and the linkage disequilibrium (D) between *In(2L)t* and the *Gpdh* genes (Mukai & Voelker, 1977). Therefore, there should be a large error for the estimate t_1. The above conclusion derived from the statistical test totally depends on the assumptions (see section 2).

The second alternative hypothesis is that the high frequency of haplotype 15 results from bias in our sampling, namely, we may have sampled closely related flies such as sister flies by chance. This hypothesis could be rejected by the following three findings: (1) the distribution of *copia* elements on the six second chromosomes of haplotype 15 are not very similar, (2) two of the six chromosomes carry recessive lethal mutations of independent origin, and (3) haplotype 15 was detected at similar frequencies in both 1984 and 1987.

The third is that haplotype 15 has multiple origins. In spite of the fact that the *BN* element shows little or no target sequence specificity (Harada *et al.* 1988), the

https://doi.org/10.1017/S0016672300029803 Published online by Cambridge University Press

insertion point of insertion 'i' is identical among the five chromosomes examined out of the six. Therefore, this hypothesis is difficult to support.

Another possibility is the founder effect. The third criterion (absence in neighbouring populations where migration occurs from the population in question) was necessary to reduce the possibility of the founder effect. Haplotype 15 has a high frequency in the Raleigh population, and is not observed in other populations examined (Austin, Okinawa and Osaka). In addition to our data, no insertions identical to insertion 'i' were detected in *In(2L)t*-carrying chromosomes from the Rhode Island population (cf. Aquadro *et al.* 1986). If we could show clear evidence for constant migration among these populations, the possibility of the founder effect would be ruled out. Unfortunately there is little detailed quantitative data about migration. Therefore, the possibility of the founder effect cannot be excluded. However, there is indirect evidence about the extent of migration. First, the upstream *Bam*H I site at position −7·2 kb is highly polymorphic in all populations examined for both *St* and *In* (also in Rhode Island, cf. Aquadro *et al.* 1986). Second, Kreitman & Aguadé (1986) obtained no evidence of genetic differentiation between populations in Raleigh, NC and Putah Creek, CA, with high resolution restriction mapping of the *Adh* region. Third, it has been suggeted that *P* elements have attained worldwide distribution within 50 years in this species (cf. Kidwell, 1983).

We obtained a genetic variant that possibly carries a beneficial mutation. Further analysis of certain parameters, particularly the population size and the migration rate, are needed to support the view that positive natural selection has acted directly on this variant. At the same time, additional studies of genetic variation using the method described here will identify other candidates that may represent beneficial mutations.

We thank C. C. Laurie, N. Miyashita and Y. Hiraizumi for providing the materials from the 1984 Raleigh, the 1987 Raleigh and the Austin population, respectively. We are also grateful to D. L. Hartl, C. H. Langley, H. Ochman, H. Tachida, and N. Takahata for their criticisms and suggestions. This work was supported by research grants from the Ministry of Education, Science and Culture of Japan (to T.M.), the Toray Science Foundation (to T.M.), and the Fellowship of the Japan Society for the Promotion of Science (to A.K.). This is paper no. 54 from the Laboratory of Population Genetics, Department of Biology, Kyushu University.

References

Aguadé, M. (1988). Restriction map variation of the *Adh* locus of *Drosophila melanogaster* in inverted and non-inverted chromosaomes. *Genetics* **119**, 135–140.

Aquadro, C. F., Desse, S. F., Bland, M. M., Langley, C. H. & Laurie-Ahlberg, C. C. (1986). Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster*. *Genetics* **114**, 1165–1190.

Bingham, P. M., Levis, R. & Rubin, G. M. (1981). Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. *Cell* **25**, 693–704.

Birley, A. J. (1984). Restriction endonuclease map variation and gene diversity in the *Adh* region in a population of *Drosophila melanogaster*. *Heredity* **52**, 103–112.

Clayton, G. & Robertson, A. (1955). Mutation and quantitative variation. *The American Naturalist* **186**, 151–158.

Chia, W., Karp, R., McGill, S. & Ashburner, M. (1985). Molecular analysis of the *Adh* region of the genome of *Drosophila melanogaster*. *Journal of Molecular Biology* **186**, 689–704.

Cross, S. R. H. & Birley, A. J. (1986). Restriction endonuclease map variation in the *Adh* region in populations of *Drosophila melanogaster*. *Biochemical Genetics* **24**, 415–433.

Durrant, A. & Mather, K. (1954). Heritable variation in a long inbred line of Drosophila. *Genetica* **27**, 97–199.

Feinberg, A. P. & Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry* **132**, 6–13.

Harada, K., Koga, A., Kusakabe, S. & Mukai, T. (1988). A new family of mobile dispersed middle repetitive elements in *Drosophila melanogaster*. *Proceedings of the Japan Academy* **64**, 193–196.

Hattori, M. & Sakaki, Y. (1986). Dideoxy sequencing method using denatured plasmid templates. *Analytical Biochemistry* **153**, 232–238.

Hughes, A. L. & Nei, M. (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167–170.

Kettlewell, H. B. D. (1965). Insect survival and selection for pattern. *Science* **148**, 1290–1296.

Kidwell, M. G. (1983). Evolution of hybrid dysgenesis determined in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 1655–1659.

Kimura, M. (1983*a*). *The Neutral Theory of Molecular Evolution*. Cambridge University Press. Cambridge, U.K.

Kimura, M. (1983*b*). Rare variant alleles in the light of the neutral theory. *Molecular Biology and Evolution* **1**, 84–93.

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

Kreitman, M. & Aguadé, M. (1986). Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. *Proceedings of the National Academy of Sciences, U.S.A.* **83**, 3562–3566.

Langley, C. H., Montgomery, E. & Quattlebaum, W. F. (1982). Restriction map variation in the *Adh* region of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U.S.A.* **79**, 5631–5635.

Lindsley, D. L. & Grell, E. H. (1968). *Genetic Variations of* Drosophila melanogaster. Carnegie Inst. Wash. Publ.

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Mettler, L. E., Voelker, R. A. & Mukai, T. (1977). Inversion clines in populations of *Drosophila melanogaster*. *Genetics* **87**, 169–176.

Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972). Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 335–355.

Mukai, T. & Cockerham, C. C. (1977). Spontaneous mutation rates at enzyme loci in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 2514–2517.

Mukai, T., Harada, K., Kusakabe, S. & Yamazaki, T. (1990). Estimation of spontaneous mutation rates at enzyme loci in *Drosophila melanogaster*. *Proceedings of the Japan Academy* **66**, 29–32.

Mukai, T., Tachida, H. & Ichinose, M. (1980). Selection for viability at loci controlling protein polymorphisms in *Drosophila melanogaster* is very weak at most. *Proceedings of the National Academy of Sciences, U.S.A.* **77**, 4857–4860.

Mukai, T. & Voelker, R. A. (1977). The genetic structure of natural populations of *Drosophila melanogaster*. XIII. Further studies on linkage disequilibrium. *Genetics* **86**, 175–185.

Mukai, T. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster*. XI. Genetic variability in a local population. *Genetics* **76**, 339–366.

Mukai, T., Watanabe, T. K. & Yamaguchi, O. (1974). The genetic structure of natural populations of *Drosophila melanogaster*. XII. Linkage disequilibrium in a large local population. *Genetics* **77**, 771–793.

Nei, M., Kojima, K. & Schaffer, H. E. (1967). Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics* **57**, 741–750.

Nei, M. & Li, W.-H. (1980). Non-random association between electromorphs and inversion chromosomes in finite populations. *Genetical Research* **35**, 65–83.

Nei, M. & Tajima, F. (1981). DNA polymorphism detected by restriction endonucleases. *Genetics* **97**, 145–163.

Pardue, M. L. & Gall, J. G. (1975). Nucleic acid hybridization to the DNA of cytological preparations. *Methods in Cell Biology* **10**, 1–16.

Perutz, M. F. (1983). Species adaptation in a protein molecule. *Molecular Biology and Evolution* **1**, 1–28.

Rigby, P. W. J., Dieckmann, M. & Rhodes, C. (1977). Labelling deoxyribonucleic acid to high specific activity *in vivo* by nick translation with DNA polymerase I. *Journal of Molecular Biology* **113**, 237–251.

Saigo, K., Millstein, L. & Thomas, C. A., Jr. (1981). The organization of *Drosophila melanogaster* histone genes. *Cold Spring Harbor Symposium, Quantitative Biology* **45**, 815–827.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, U.S.A.* **74**, 5463–5467.

Shaw, C. R. (1965). Electrophoretic variation in enzymes. *Science* **149**, 936–943.

Slatkin, M. (1987). The average number of sites separating DNA sequences drawn from a subdivided population. *Theoretical Population Biology* **32**, 42–49.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503–517.

Strobeck, C. (1987). Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* **117**, 149–153.

Tajima, F. (1990). Relationship between migration and DNA polymorphism in a local population. *Genetics* **126**, 231–234.

Yamaguchi, O., Ichinose, M., Matsuda, M. & Mukai, T. (1980). Linkage disequilibrium in isolated populations of *Drosophila melanogaster*. *Genetics* **96**, 507–552.

Yokoyama, S., Isenberg, K. E. & Wright, A. F. (1989). Adaptive evolution of G-protein coupled receptor genes. *Molecular Biology and Evolution* **6**, 342–353.

Weiner, R. & Crow, J. F. (1951). The resistance of DDT-resistant *Drosophila* to other insecticides. *Science* **113**, 403–404.