Proc. Natl. Acad. Sci. USA
Vol 74, No. 6, pp. 2514–2517, June 1977
Genetics

Spontaneous mutation rates at enzyme loci in *Drosophila melanogaster**

(viability polygenes/structural genes/controlling regions of chromosomes/protein polymorphism/evolution)

TERUMI MUKAI† AND C. CLARK COCKERHAM

Departments of Genetics and Statistics, North Carolina State University, Raleigh, North Carolina 27607

Contributed by C. Clark Cockerham, April 4, 1977

ABSTRACT **In a marked-inversion-balanced lethal system, mutations were accumulated at a minimum pressure of natural selection on 2000 second chromosomes of *Drosophila melanogaster* that originated from 4 stem chromosomes. Five enzyme loci were tested: α-glycerol-3-phosphate dehydrogenase (EC 1.1.1.8), malate dehydrogenase (*Mdh*, EC 1.1.1.37), alcohol dehydrogenase (EC 1.1.1.1), hexokinase-C (*Hex-C* tentative name), and α-amylase (*Amy*, EC 3.2.1.1). Three band-morph mutants, one at the *Mdh* locus, one at the *Hex-C* locus, and one at the *Amy* locus, were detected out of 1,658,308 allele replications. In addition, 17 null mutants were found. Accepting that the number of structural genes is the same as that of bands in the salivary gland chromosomes, the total mutation rate per generation for all the structural genes in the second chromosomes is estimated to be 0.008–0.040, which is much smaller than that estimated for viability polygenes (0.12–0.17). Thus, it is speculated that most viability and other fitness polygenes are located in controlling regions outside the structural genes.**

Large numbers of protein polymorphisms have been reported in many species since Lewontin and Hubby (1) published data for *Drosophila pseudoobscura*. Although several mechanisms for the maintenance of these polymorphisms have been proposed, a decisive conclusion on this matter has not been reached. One of the reasons for this is that a reliable estimate for the spontaneous mutation rate has not been obtained.

In *Drosophila melanogaster*, there are two reports (2, 3) of direct estimates of band-morph mutation rates for enzyme loci determined by electrophoresis, but the results are not very reliable because no system of checking for contamination was used.

Extensive studies were made of mutation rates of viability polygenes in *D. melanogaster* (4, 5). The results indicated that the viability polygenic mutation rate is approximately 20 times higher than the recessive lethal mutation rate on a chromosome basis. Unfortunately, these estimates were obtained by indirect statistical methods and the nature of these mutations was not known. From developments in molecular genetics and other evidence, it may be possible to say something about the location and basis of polygenic mutations.

The purposes of the present work are twofold: (*i*) to estimate the rate of occurrence of band-morph mutations, from which the mutation rate per base-pair site may be estimated; (*ii*) to focus this and other evidence onto the nature of polygenic mutations.

MATERIALS AND METHODS

Four stem chromosomes were used: two *SM1*(*Cy*) chromosomes and two unrelated lethal-carrying chromosomes, *ℓ*(*AW*) and *ℓ*(*JH*), which were derived in 1967 from a cage population (W-1) (6). A single male, which was a heterozygote for *SM1*(*Cy*) and *ℓ*(*AW*), was mated to a single C-160 [*SM1*(*Cy*)/*In*(2*LR*)-*bw*^{V1}, which is abbreviated *Cy*/*Pm*] female. To establish the chromosome lines from the progenies, *Cy*/*ℓ* males and females were collected and many single-pair matings were made between them. In the progenies, generation one and later generations, only *Cy*/*ℓ* heterozygotes survive because *Cy*/*Cy* and *ℓ*/*ℓ* are lethal. The number of single-pair matings was increased and, by generation three, 500 lines were established (AW-1, AW-2, . . . , AW-500). Each line was maintained by a single-pair mating and a five-pair mating. Whenever the single-pair mating was successful, its offspring were used to make a single-pair mating and a five-pair mating for the next generation. When the single-pair mating was not successful, the five-pair mating was used as a substitute source of flies for the next generation. It is possible to uncover any contamination from external sources using this method because, if it occurs, phenotypically wild-type flies appear. Following the above procedure, 500 lines of the other group (JH-1, JH-2, . . . , JH-500) were established using a single *SM1*(*Cy*) chromosome and a single lethal-carrying chromosome, *ℓ*(*JH*).

When lines appeared too weak to maintain they were strengthened by backcrossing to the *Cy*/*Pm* stock and single pair matings were continued. An occasional line lost in later generations was replaced by a sister line. Consequently, a small amount of error with respect to the number of generations during which mutations were accumulated could have been introduced.

The following five enzyme loci, known to be located in the second chromosome (refs. 7 and 8; T. Mukai and R. A. Voelker, unpublished), were studied: α-glycerol-3-phosphate dehydrogenase [*αGpdh* (EC 1.1.1.8)—map position, 2-17.8]; malate dehydrogenase [*Mdh* (EC 1.1.1.37)—map position, 2-35.3]; alcohol dehydrogenase [*Adh* (EC 1.1.1.1)—map position, 2-50.1]; hexokinase-C (tentative name assigned by F. M. Johnson, unpublished—map position, 2-74.5); and α-amylase [*Amy* (EC 3.2.1.1)—map position, 2-77.3] (8). The three dehydrogenases and hexokinase-C were assayed by starch gel electrophoresis (9), while α-amylase was assayed by acrylamide gel electrophoresis (10).

The genotypes of these five loci in the stem chromosomes are given in Table 1. To assay mutations for heterozygous lines, *Cy*/*ℓ*(*AW*) at the *αGpdh* and *Adh* loci and *Cy*/*ℓ*(*JH*) at the *αGpdh* locus, a single *Cy*/*ℓ* male in each line was crossed to five C-160 [*SM1*(*Cy*)/*Pm*] females in order to save the original second chromosomes (*Cy*-carrying and *ℓ*-carrying chromosomes). After that the original male fly was homogenized and electrophoretically examined. If a new form of band-morph was seen, it was interpreted as a mutation that had significantly changed the electrical charge. Then the original chromosome that carried the mutant was isolated from the offspring of the

* Paper no. 5152 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, NC.

† Present address: Department of Biology, Kyushu University, Fukuoka 812, Japan.

Table 1. Summary of the results of the experiment for estimating enzyme mutation rate

	Locus on stem chromosome AW					Locus on stem chromosome JH				
	αGpdh	*Mdh*	*Adh*	*Hex-C*	*Amy*	*αGpdh*	*Mdh*	*Adh*	*Hex-C*	*Amy*
Genotype of *Cy* chromosome	*F*	*S*	*F*	*S*	Amy^1	*F*	*S*	*F*	*S*	Amy^1
Original genotype of stem chromosome	*S*	*S*	*S*	*S*	$Amy^{1,6}$	*S*	*S*	*F*	*S*	Amy^1
Number of band-morph mutants	0	1*	0	1†	1‡	0	0	0	0	0
Number of null mutants§	1, 2	0, 2	1, 0	0, 0	0, 0	2, 2	2, 2	0, 1	2, 0	0, 0
Total number of allele generations	151,534	154,322	151,426	174,590	234,610	151,186	155,228	155,794	173,608	156,010

S indicates a slowly migrating enzyme; *F*, fast.
* *S* (or Mdh^4) changed to Mdh^6.
† *S* (or *Hex-C*^{4}) changed to *Hex-C*$^{2.5}$.
‡ $Amy^{1,6}$ changed to $Amy^{2,6}$.
§ *x*, *y* indicates that counts of *x* null mutants were observed on *Cy* and *y* on the *ℓ* chromosomes.

Cy/Pm females that had been crossed to the original male, and the chromosome was tested again.

At all the other loci, *Cy/ℓ* individuals are homozygous except for the hemizygous *6* portion of $Amy^{1,6}$ in the AW group. In these cases, a single male, *Cy/ℓ*, in each line was first crossed to an individual homozygous for the alternative allele, and then electrophoretically examined. The two types of offspring (one carrying the *Cy* chromosome and the other carrying the *ℓ* chromosome) were electrophoretically examined. Without a mutation the offspring are heterozygous at the enzyme locus in question, so that the change of the band-morph or the disappearance of the band could be detected easily.

For the test of the *Amy* alleles, a cross was made between five females of the $Amy^{4,6}$ line (which was kindly sent by Zenichi Ogita of the University of Osaka, Japan) and a single male of each line, and the parental male fly was subsequently electrophoretically examined. Using the two types of offspring (*Cy* and non-*Cy* flies), we tested whether mutations occurred at the *1* portion of the locus. In order to detect the mutations at the *6* portion of the *ℓ*-carrying chromosome in the AW group, the *Cy* offspring of the cross between a single *Cy/ℓ* male of each line and five *Cy/Pm* females were employed. Parental male flies were also examined electrophoretically.

RESULTS

A complete screening of the two sets of 500 chromosomal lines (AW and JH) was completed by generation 175. The results are summarized in Table 1. For the *Amy* locus, one chromosomal replication was counted as two in *ℓ(AW)*-carrying chromosomes, because it has been shown that the parts of *1* and *6* can be recombined (11). (While not definitely known, it was assumed that there was no counterpart to *6* in *Cy-* and *ℓ(JH)*-carrying chromosomes.) Three new band-morph alleles, one each at *Mdh* (in generation 157), at *Amy* (in generation 154), and at *Hex-C* (in generation 175), have been isolated out of 1,658,308 allele replications.

Although the accumulation method checks for contamination, further examination of contamination was made, and it was concluded that the changes of the band-morphs were not due to contamination.

Identification of the Mutants. An offspring with wild-type phenotype from the cross between Mdh^F/Mdh^F × AW-438 [$Cy(Mdh^S)/ℓ(Mdh^S)$]—the expected genotype of the offspring is $+(Mdh^F)/ℓ(Mdh^S)$—showed an abnormal band-morph. It appeared that the *ℓ*-carrying chromosome acquired a new allele at the *Mdh* locus, the product of which moves slower than that of the Mdh^S alleles in electrophoresis. In order to confirm this, a cross was made between *Cy/Pm* females (both chromosomes carry Mdh^S alleles) and a *Cy/ℓ* male from AW-438. In the offspring, phenotypically *Cy* flies and *Cy/Pm* flies segregate. The former genotype showed a band-morph which is heterozygous between $Mdh^4(S)$ and an allele carrying a slower band-morph than the Mdh^S allele. Thus, the above finding was confirmed. According to F. M. Johnson's terminology, this new allele appears to show the same banding pattern as the Mdh^6 allele, which is found in very low frequency in the northeastern population of the U.S.A. (12). Therefore, a cross was made between AW-438 and a line carrying Mdh^6 alleles that was extracted from a cage population initiated from a Maine population by F. M. Johnson. The offspring showed the band-morph of a homozygote for Mdh^6. It should be noted that the flies carrying the Mdh^6 alleles were not introduced to our laboratory before the detection of the mutant. Thus, we conclude that a band-morph mutant (from Mdh^4 to Mdh^6) occurred on an *ℓ*-carrying second chromosome of the AW group.

An offspring with the curly phenotype from the cross between *Hex-C*^{2}/*Hex-C*^{2} × AW 399 [*Cy*(*Hex-C*^{4})/*ℓ*(*Hex-C*^{4})] appeared to carry a null mutant. However, the width of the band was larger than that for a *Hex-C*^{2} homozygote. The time of electrophoresis was extended and the new band *Hex-C*$^{2.5}$ (the position is much closer to *Hex-C*^{2} than to *Hex-C*^{4}) was clearly detected. If the pH of the gel buffer is changed from 7.3 to 8.5, the separation is easily accomplished. A counterpart to this mutant in nature is not known.

A single male of AW-471 was tested for α-amylase at generation 154. A new banding pattern (*1,2,6*) appeared in contrast to the expected banding pattern [(*1,6*) or *Cy*(Amy^1)/*ℓ*($Amy^{1,6}$)]. It was necessary to decide whether the genotype of the *Cy* chromosome changed or that of the *ℓ*-carrying chromosome changed. In the offspring of a cross between $Amy^{4,6}/Amy^{4,6}$ and a phenotypically *Cy* male fly from AW-471 [*Cy*(Amy^1)/*ℓ*($Amy^{1,6}$)], phenotypically *Cy* flies showed a banding pattern of (*1,4,6*) and phenotypically wild-type flies showed (*2,4,6*). Thus, it was concluded that $Amy^{1,6}$ in the *ℓ*-carrying chromosome changed to $Amy^{2,6}$.

Calculation of Mutation Rates. The total number of allele generations tested is 1,658,308. As shown in Table 1, the number of band-morph mutants is three. Hence, the estimate of the average band-morph mutation rate (excluding null mutations) ($\hat{\mu}_B$) is:

$$\hat{\mu}_B = 3/1{,}658{,}308 = 1.81 \times 10^{-6}/\text{locus per generation.}$$

The 95% confidence interval is 3.73×10^{-7} to 5.29×10^{-6}/locus per generation from Fisher and Yates' Table (13).

The total number of null mutations is 17, providing an average null mutation rate ($\hat{\mu}_N$) of

$$\hat{\mu}_N = \frac{17}{1{,}658{,}308} = 1.03 \times 10^{-5}/\text{locus per generation.}$$

Various estimates of the detection rate of single base substitutions by electrophoresis have been made from theoretical considerations (14–16) and from empirical considerations (17, 18). We use a generous range of 0.1–0.5 that covers all of these estimates. Then, the estimate of the single base substitution rate (μ) is in the range $\hat{\mu} = (2\text{–}10)\ 1.81 \times 10^{-6} = (3.62\text{–}18.1) \times 10^{-6}$/structural gene per generation.

To convert this rate to that for base pairs, we estimate the number of base pairs for the four enzymes studied for which molecular weights have been estimated and use the average number. A formula to accomplish this purpose is developed as follows. The subunit molecular weight (M) of an enzyme is equal to the number (n) of amino acids times the average formula weight ($\overline{m}$) of the amino acids minus the $n - 1$ water molecules (formula weight = 18.02) lost during peptide bonding, i.e.,

$$M = n\overline{m} - (n - 1)\,18.02.$$

On rearranging,

$$n = \frac{M - 18.02}{\overline{m} - 18.02}$$

and $3n$ is the number of base pairs. The average amino acid weight for 17,941 residues in 160 sequenced proteins (19) was 127.84, which is close to the estimate of 128.47 found by King and Jukes (20). On substitution of 127.84 for $\overline{m}$ the estimator of the number of base pairs becomes

$$3\hat{n} = 3(\hat{M} - 18.02)/109.82.$$

Estimates of molecular weights and numbers of base pairs are given in Table 2. The base pair substitution rate, using the average number of base pairs per structural gene in Table 2, is estimated to be in the range

$$\hat{\mu}_{bp} = (3.72\text{–}18.6) \times 10^{-9}/\text{base pair per generation.}$$

DISCUSSION

Nature of Viability Polygenes. The mutation rate (M_v) of viability polygenes has been estimated to be $\hat{M}_v = 0.12\text{–}0.17$/second chromosome per generation (4, 5). Recently, evidence was presented (25) that the number of bands (chromomeres) in the salivary gland chromosomes is approximately equal to the number of cistrons. If this is so, the number of cistrons in the second chromosome is about 2200. Then, assuming a single structural element (polypeptide) per cistron, the estimate of the total mutation rate (M_s) in structural genes of the second chromosome is:

$$\hat{M}_s = (3.62\text{–}18.1) \times 10^{-6} \times 2.2 \times 10^3$$

$$= 0.008\text{–}0.040/\text{second chromosome per generation.}$$

This rate is considerably less than that for $\hat{M}_v$. Also, $\hat{M}_v$ is for the mutation of polygenes affecting viability only. It is reasonable to assume many polygenes affecting other fitness components. Evidence was presented (26) that homozygosity for second chromosomes has a much greater effect on fertility than on viability. Consequently, the total polygenic mutation rate is probably many times higher than that for structural genes. Thus, it may be speculated that most of the viability and other fitness polygenes are located outside the structural genes in the chromosome. This place may be called the "controlling region" (27).

Table 2. Estimates of molecular weights and numbers of base pairs

Enzyme	Molecular weight	Ref.	Subunit weight, $\hat{M}$	No. of base pairs, $3\hat{n}$
α-Glycerol-3-phosphate dehydrogenase	66,000*	21	33,000	901
Malate dehydrogenase	60,000*	22	30,000	819
Alcohol dehydrogenase	50,000*	23	25,000	682
α-Amylase	54,500	24	54,500	1488
Average			35,625	973

* These are considered to be dimeric weights, although in the case of malate dehydrogenase it has not been rigorously shown that it is a dimer with subunits of identical molecular weight.

The number of base pairs in a *Drosophila melanogaster* genome is reported to be about 10^8 (28). The physical length of the second chromosome is approximately $\frac{2}{5}$ of the whole genome, so that the number of base pairs in the second chromosome is about 4×10^7. Approximately 20% of the DNA is highly repeated DNA (29). Then, the estimate of the total mutation rate (M_T) in the other part (moderately repetitive and unique DNA) is

$$\hat{M}_T = (3.72\text{–}18.6) \times 10^{-9} \times 4 \times 10^7 \times 0.8$$

$$= 0.119\text{–}0.595/\text{second chromosome per generation}$$

which includes the mutation rate of viability polygenes (0.12–0.17) and other fitness polygenes. The proportion of neutral mutations, i.e., involving the DNA coding for proteins, might well be small and not add much to the total mutation rate, although there is the possibility of an appreciable frequency of neutral or near-neutral mutations occurring both in structural genes and in controlling regions.

Limitations. We have boldly interpreted the band-morph changes as single base substitutions. Other mechanisms such as frame shift mutations and deletions and additions of bases could lead to electrophoretically detectable mutants, but these mechanisms usually lead to chain terminations and probably account for an insignificant amount of band-morph mutants. Our interpretation is also supported by the fact that most of the mutants encountered in well-analyzed proteins are explainable by single base substitutions.

Our balanced lethal system with the balancer chromosome *SM1(Cy)*, while useful in controlling contamination, prevents normal recombination. Consequently, we have ruled out the generation of new variation through recombination within a cistron. A different type of experiment with crossing-over in heterozygotes at the loci studied is required to include this kind of new variation.

We have simply reported the null mutation rate and made no further use of it. Again, several mechanisms can lead to null mutations, including intragenic and larger deletions. Some kind of mutator factor that breaks chromosomes and chromatids was found (6) in many of the mutation lines, both AW and JH. The frequency distribution of the break points on the chromosome can be seen in Fig. 3 of Yamaguchi and Mukai's paper (6) which is roughly proportional to the frequencies of the null mutants. Thus, deletions could well account for many of the null mutations; however, the nature of the null mutations is yet to be determined.

Mutations and Adaptation. It could well be that random genetic drift has played a significant role in the substitution of base pairs in DNA coding for proteins (30), and that many mutant alleles that replaced preexisting alleles are effectively neutral. It is also likely, however, that some mutations that destroy the integrity of the protein are selected against. However, the importance of variation in proteins for the evolution and adaptation of organisms is being questioned. King and Wilson (31) surveyed the differences of amino acids between humans and chimpanzees for various proteins, and concluded that the sequences of human and chimpanzee polypeptides examined to date are, on the average, more than 99% identical. For example, the two species appear to carry identical fibrinopeptides, cytochrome c, and α, β, and γ hemoglobin chains. In myoglobin and the δ chain of hemoglobin, the human polypeptide chain differs from that of the chimpanzee by a single amino acid replacement. Nevertheless, they concluded that the two species are quite different from the viewpoint of adaptation and that biological adaptations occur mainly in the controlling regions in the chromosomes. Although changes in conformation of proteins and their amino acid sequences may be important for the evolution of organisms, controlling regions appear to be more important in the differentiation of species.

This paper is dedicated to the memory of the late Prof. Harry C. Kelly, who died Feb. 2, 1976, shortly after his retirement with 12 years of service as Provost and Vice Chancellor at North Carolina State University.

Thanks are due Dr. Peter M. Burrows for the method of estimating the number of base pairs, and Drs. F. M. Johnson and R. A. Voelker, who lent effective technical help. This investigation was supported in part by National Institutes of Health Research Grant GM 11546 from the National Institute of General Medical Sciences.

The costs of publication of this article were defrayed in part by the payment of page charges from funds made available to support the research which is the subject of the article. This article must therefore be hereby marked "advertisement" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

1. Lewontin, R. C. & Hubby, J. L. (1966) *Genetics* **54**, 595–609.
2. Mukai, T. (1970) Drosophila *Inform. Serv.* **45**, 99.
3. Tobari, Y. N. & Kojima, K. (1972) *Genetics* **70**, 397–403.
4. Mukai, T. (1964) *Genetics* **50**, 1–19.
5. Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. (1972) *Genetics* **72**, 335–355.
6. Yamaguchi, O. & Mukai, T. (1974) *Genetics* **78**, 1209–1221.
7. O'Brien, S. J. & MacIntyre, R. J. (1969) *Am. Nat.* **103**, 97–113.
8. Lindsley, D. L. & Grell, E. H. (1967) *Genetic Variations of* Drosophila melanogaster (Carnegie Institution of Washington Publication No. 627).
9. Stone, W. S., Johnson, F. M., Kojima, K. & Wheeler, M. R. (1968) *Studies in Genetics IV* (Univ. Texas Publ. 6818), pp. 157–170.
10. Prakash, S., Lewontin, R. C. & Hubby, J. L. (1969) *Genetics* **61**, 841–858.
11. Bahn, E. (1967) *Hereditas* **58**, 1–12.
12. Johnson, F. M. & Schaffer, H. E. (1973) *Biochem. Genet.* **10**, 149–163.
13. Fisher, R. A. & Yates, F. Y. (1970) *Statistical Tables for Biological, Agricultural, and Medical Research*, (Hafner Publishing Co., Darien, CT).
14. Shaw, C. R. (1965) *Science* **149**, 936–943.
15. Marshall, D. R. & Brown, A. H. D. (1975) *J. Mol. Evol.* **6**, 149–163.
16. Nei, M. & Chakraborty, R. (1973) *J. Mol. Evol.* **2**, 323–328.
17. Boyer, S. H., Noyes, A. N., Timmons, C. F. & Young, R. A. (1972) *J. Hum. Evol.* **1**, 515–543.
18. Hunt, L. T., Sochard, M. R. & Dayhoff, M. O. (1972) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (The National Biomedical Research Foundation, Washington, DC), Vol. 5, pp. 67–87.
19. Dayhoff, M. O., ed. (1972) *Atlas of Protein Sequence and Structure* Vol. 5, and (1973) Vol. 5, Supplement I, (The National Biomedical Research Foundation, Washington, DC).
20. King, J. L. & Jukes, T. H. (1969) *Science* **164**, 788–798.
21. Bewley, G. C., Rawls, J. M. & Lucchesi, J. C. (1974) *J. Insect Physiol.* **20**, 153–165.
22. O'Brien, S. J. (1973) *Biochem. Genet.* **10**, 191–205.
23. Schwartz, M., Gerace, L., O'Donnell, J. & Sofer, W. (1975) in *Isozymes I: Molecular Structure*, ed. Markert, C. L. (Academic Press, New York), pp. 725–751.
24. Doane, W. W., Abraham, I., Kolar, M. M., Martenson, R. E. & Deibler, G. E. (1975) in *Isozymes IV: Genetics and Evolution*, ed. Markert, C. L. (Academic Press, New York), pp. 585–607.
25. Judd, B. H., Shen, M. W. & Kaufman, T. C. (1972) *Genetics* **71**, 139–156.
26. Tracey, M. L. & Ayala, F. J. (1974) *Genetics* **77**, 569–589.
27. Georgiev, G. P. (1969) *J. Theor. Biol.* **25**, 473–490.
28. Laird, C. D. & McCarthy, B. J. (1969) *Genetics* **63**, 865–882.
29. Wu, J.-R., Hurn, J. & Bonner, J. (1972) *J. Mol. Biol.* **64**, 211–219.
30. Kimura, M. & Ohta, T. (1971) *J. Mol. Evol.* **1**, 1–17.
31. King, M. C. & Wilson, A. C. (1975) *Science* **188**, 107–116.