

The Molecular Mechanism Underlying the “Rare Allele Phenomenon” in a Subspecific Hybrid Zone of the California Field Mouse, *Peromyscus californicus*

S.M.G. Hoffman,* W.M. Brown

Department of Biology, University of Michigan, Ann Arbor, MI 48109, USA

Received: 9 January 1995 / Accepted: 30 March 1995

Abstract. Natural hybrid zones are known to have unusually high levels of novel or otherwise rare electrophoretic variants (the “rare allele phenomenon”). These variant alleles are most likely the result either of high levels of unique mutations in hybrids or of intragenic recombination between divergent alleles from the parental populations. This study uses DNA sequence comparisons to determine which process has produced a rare allele of the 6-phosphogluconate dehydrogenase (6-PGD) gene in a subspecific hybrid zone of the California field mouse (*Peromyscus californicus*).

About 70% of the coding sequence of 6-PGD was cloned and sequenced from three alleles, including two widespread alleles and one rare allele unique to hybrid populations. Sequence comparisons among the three alleles reveal no patterns that would indicate that the variant was formed by intragenic recombination. Instead, the unique allele of 6-PGD studied seems to have developed by the accumulation of base substitutions, which supports the hypothesis of increased mutation rates in hybrids.

Key words: Hybrid zone — Hybridzyme — Genetic variation — *Peromyscus* — 6-PGD

Introduction

Natural hybrid zones have been extensively studied by evolutionary biologists, since they are intriguing sites of genetic interactions between divergent, but still compatible, populations. Hybrid zones may be especially conducive to the generation of new genetic variants, and thus may play a significant part in the creation of the raw material on which selection can act (Woodruff et al. 1983). The interest in gene flow and selection across hybrid zones has led to numerous surveys of hybridizing taxa using protein electrophoresis. In 1985, Barton and Hewitt counted more than 120 hybrid zone studies in the literature; they considered 23 of these to include “thorough” electrophoretic data. Many other such studies have been published in recent years (e.g., Dowling and Moore 1985; Nelson et al. 1987; Schilthuizen and Gittenberger 1994).

In most of the hybrid zones studied, populations within the zones have unusually high levels of genetic variation, a situation known as the “rare allele phenomenon.” Hybrid populations manifest this phenomenon by having (1) increased frequencies of some alleles that are rare outside the zone and/or (2) high numbers of allelic variants that are unique to the zone (Sage and Selander, 1979). Such alleles have been estimated to be 3–20 times more common in hybrids than in pure populations for several taxa (Smith 1979; Barton et al. 1983). These two patterns may in fact represent different stages of the same process; variants arising within a hybrid zone as unique alleles may first increase in frequency within the zone and subsequently spread out into the surrounding popu-

* Present address and correspondence to: Biology and Biotechnology Research Program, L-452, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA

lations. Woodruff (1989) has proposed the term "hybrizyme" for these variant alleles; "rare allele" is somewhat inaccurate, since the variant alleles are not necessarily rare within the hybrid zones.

The hybrizyme phenomenon is both widespread and significant. It has been detected in most electrophoretic studies of animal hybrid zones, including those of species of crickets, toads, lizards, fish, birds, rodents, salamanders, and clams (references in Barton and Hewitt 1985; Woodruff 1989), and a wide range of proteins from a variety of functional classes are affected. The frequencies of these variants can significantly raise the overall levels of polymorphism within hybrid populations (Woodruff 1989).

It has been proposed that hybrid zones, as venues where the generation of new allelic variants are especially likely, are important in species evolution (Woodruff et al. 1983; Woodruff 1989), but appropriate data are lacking. These authors suggest that the new genetic combinations which form in hybrid zones sometimes provide useful new variants, even though most of the combinations are deleterious. Barton et al. (1983) believe that hybrizymes are unlikely to represent such useful variants; a new, advantageous allele should quickly spread out of a hybrid zone into one or both parental populations, and would have to be detected in the early stages of its spread to be perceived as rare. There is, however, another possible significance to hybrizymes—as selectively neutral *indicators* of a heightened production of genetic variants. Assuming that new variants arise randomly with respect to their selection potential, an increased frequency of neutral variants means that more deleterious variants have also arisen and been selected against, and that more advantageous variants have arisen and spread beyond the confines of the zone.

Before we can assess the evolutionary importance of genetic variants in hybrid zones, two questions raised by the hybrizyme phenomenon must be answered: (1) by what mechanisms are these variant alleles formed, and (2) why are they largely restricted to hybrid zones? The two possibilities most often discussed as answers to the first question are that the variants were formed by increased mutation rate or by intragenic recombination (Barton et al. 1983; Barton and Hewitt 1985; Woodruff 1989), though gene conversion and transposable elements (Bradley et al. 1993) have also been suggested as possibilities.

There have been attempts to determine the origins of hybrizymes from indirect evidence, but they have not yielded a convincing answer. Only by comparing the DNA sequence of a variant allele with the corresponding "parental" alleles can its evolutionary origins be deduced (Woodruff 1989). The hypotheses just listed make different predictions about the pattern of sequence divergence between the various allelic forms. The first hypothesis (increased rate of mutation) predicts that a variant allele will have unique sequence changes that do not

occur in either parent. The second hypothesis (intragenic recombination) predicts that a variant allele will be a physical mosaic of the two parental alleles. Gene conversion would produce a characteristic "patchy" pattern in the variant allele, while transposable elements would make the variant longer than the parental alleles. The present study tests these hypotheses directly by comparing the sequences of alleles from a natural hybrid zone with those from adjacent populations. A similar study has been published recently (Bradley et al. 1993), in which cDNA sequences of a hybrizyme from a hybrid zone of the pocket gopher *Geomys bursarius* were examined.

In the present study, we sought to infer the origins of a hybrizyme in a hybrid zone of the California field mouse, *Peromyscus californicus*. An extensive morphological and electrophoretic survey of this species was conducted by Smith (1979), who identified two subspecies, the northern *P. c. californicus* and the southern *P. c. insignis*, with a hybrid zone between them in the San Gabriel and Santa Monica mountains. She assigned the hybrids to *P. c. insignis* based on their greater similarity to that subspecies, but the hybrids are intermediate for both genetic and morphological traits (Smith 1978, 1979). Hybrizymes of several proteins are present in the hybrid zone of *P. californicus* (Smith 1979). We have investigated one of these, 6-phosphogluconate dehydrogenase (6-PGD, E.C. 1.1.1.44), a basic housekeeping enzyme of the pentose phosphate pathway that is encoded by a single-copy gene (Miksicek and Towle 1983). The hybrizyme for 6-PGD was found at a frequency of 6% in the hybrid zone by Smith (1979).

Materials and Methods

Seventy-six mice were collected in July and August of 1986 using Sherman live traps. To ensure correct identifications, only animals in adult pelage (which is acquired at the age of 3 months [McCabe and Blanchard 1950]) were kept. Pure *Peromyscus californicus californicus* were collected about 250 km northwest of the hybrid zone, in San Luis Obispo County, and pure *P. c. insignis* were collected from localities 100–160 km south of the zone, in Riverside, Orange, and San Diego counties. Blood samples were drawn from each individual; separated red blood cells were run on horizontal starch gels as described in Selander et al. (1971), and gels were stained and scored for 6-PGD.

Three alleles of 6-PGD were detected by starch gel electrophoresis; the three migrated with the same relative mobilities as the three 6-PGD alleles described by Smith (1979). After the 6-PGD phenotype of each wild-caught mouse was determined, laboratory crosses were made to produce individuals homozygous for each of the three alleles. All copies of the rare allele of 6-PGD (designated "allele 118" by Smith) originated from a single female, the only animal captured who carried this allele. Blood samples were taken from all lab-born animals when they were 3–4 months old and typed for 6-PGD as just described. The laboratory crosses demonstrated that the three forms of 6-PGD observed on starch gels are indeed allelic and are inherited in simple Mendelian fashion (data not shown). No hybrid disadvantage was observed; both the hybrid-zone animals and the lab-generated, mixed-subspecies animals had the same lifespans and litter sizes as mice from pure-subspecies populations (Hoffman 1991).

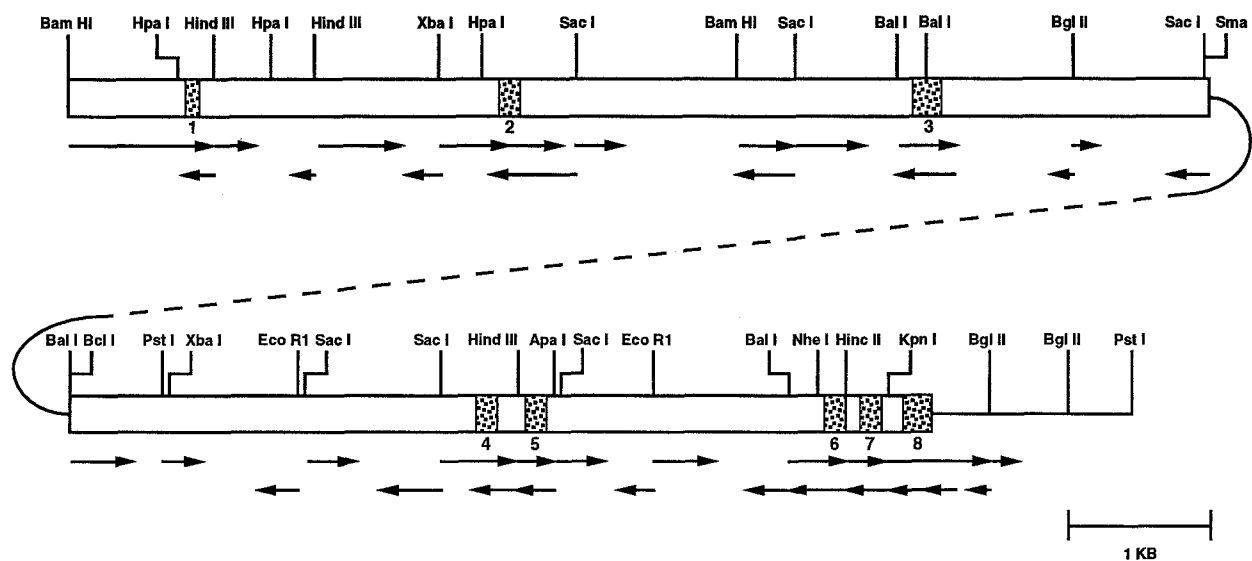


Fig. 1. Structure of the 18-kb genomic clone (PA11) containing most of the 6-PGD gene from *Peromyscus californicus*, with restriction enzyme sites indicated. Exons are represented by stippled boxes, introns by open boxes, and 3' noncoding sequence by a line. Exons are numbered from the 5' end. The length and direction of sequence determinations are indicated by arrows.

Three homozygous lab-bred mice were selected for characterization at the DNA sequence level on the basis of their 6-PGD phenotypes. Type A (a pure *P. c. californicus*) had the fast allele, type B (pure *P. c. insignis*) the slow allele, and type C (hybrid zone) the intermediate (rare) allele. A second type A mouse was later included in the study.

Total DNA was prepared from whole spleens of freshly killed mice using standard techniques (Blin and Stafford 1976; Maniatis et al. 1982). Repeated attempts were also made to prepare total DNA from liver (the most commonly used tissue for mammalian total DNA preparations), but only highly degraded DNA was recovered from these samples. This liver-specific degradation was later found to be unique to this species and to be age-dependent (Kass and Hoffman 1993).

Making and Screening Phage Libraries. A cDNA clone of lab rat (*Rattus norvegicus*) 6-PGD was obtained from H. Towles (University of Minnesota); the clone was an 880-nucleotide fragment reverse-transcribed from mRNA. Lambda phage libraries of DNA from types A, B, and C mice were made, using the supplier's protocol for a commercially prepared lambda vector (Stratagene #EM10, 1987). Total mouse DNA was incompletely digested with the restriction enzyme Sau3A using empirically determined digestion conditions that maximized the production of 15–25-kb fragments. The digested DNA was loaded on a 5–30% NaCl gradient (in 10 mM Tris-HCl, 25 mM EDTA) and spun for 20 h at 36,000 rpm in a Beckman ultracentrifuge (Sw60Ti rotor) at 20°C. Fractions of 0.5 ml were collected by bottom puncture, and the size range of the DNA in each fraction was estimated after electrophoresis on a 1% agarose gel. The fractions containing fragments of about 20 kb were pooled and the DNA was ligated onto commercially prepared arms of BamHI-digested lambda phage EMBL3 and packaged using standard conditions (Stratagene #GP10-G).

Amplification of the type B and C libraries was uneven due to poor growth of the host bacteria. Approximately 400,000 plaques were initially screened from each library. Probes for 6-PGD were prepared from the entire rat 6-PGD cDNA clone (880 bp), as well as from two subcloned fragments of 215 and 665 bp. Plaques isolated from the first screening were rescreened twice under the same conditions until individual positive plaques could be confidently identified. Screening the library for the type A allele initially yielded three strongly hybridizing phage clones, but repeated screening of the low-titer type B and C libraries never yielded probe-positive clones.

Cloning and Sequencing. DNA preparations were made from the three positive type A clones. The largest insert was about 18 kb, and inserts of the other two clones were contained within it. The 18-kb clone (designated PA11) was digested with various restriction enzymes and the restriction sites were mapped. The PA11 clone was subcloned into a plasmid vector, and most of each subclone was sequenced. All DNA sequencing was done using the dideoxy chain termination method of Sanger et al. (1977) and the Sequenase 2.0 enzyme system. Both double- and single-stranded DNA were used as templates, and subclones containing exons were completely sequenced from both strands (Fig. 1).

Synthetic oligonucleotide primers were designed based on the sequences from the A mouse DNA. These primers were used to amplify 6-PGD sequences directly from the total DNA preparations of types B and C mice, using the polymerase chain reaction (PCR). Primer pairs were chosen to be positioned on either side of an exon or exon cluster so that the entire exon as well as some flanking intron sequence was amplified (Fig. 1). PCR products were cloned into the plasmid vector pBlueScript II (Stratagene #212301, 1990) and sequenced. The PCR amplifications and sequencing were done at least twice for each primer pair for the types B and C mice and for a second type A mouse (A2), plus once for the original type A mouse (A1), to guard against errors generated by amplification.

Results

We sequenced about 70% of the coding region of the 6-PGD gene (as predicted from the 482 amino acids in sheep 6-PGD) for the type A allele in *P. californicus*. The 333 known codons are distributed into eight exons embedded in approximately 15 kb of noncoding DNA (Fig. 2). Exons were initially identified by comparison to a published amino acid sequence for 6-PGD in sheep (Carne and Walker 1983). About 10.2 kb of the 14 kb contained in the introns were also sequenced.

The sequence differences observed between the three 6-PGD alleles of *P. californicus* are summarized in Fig.

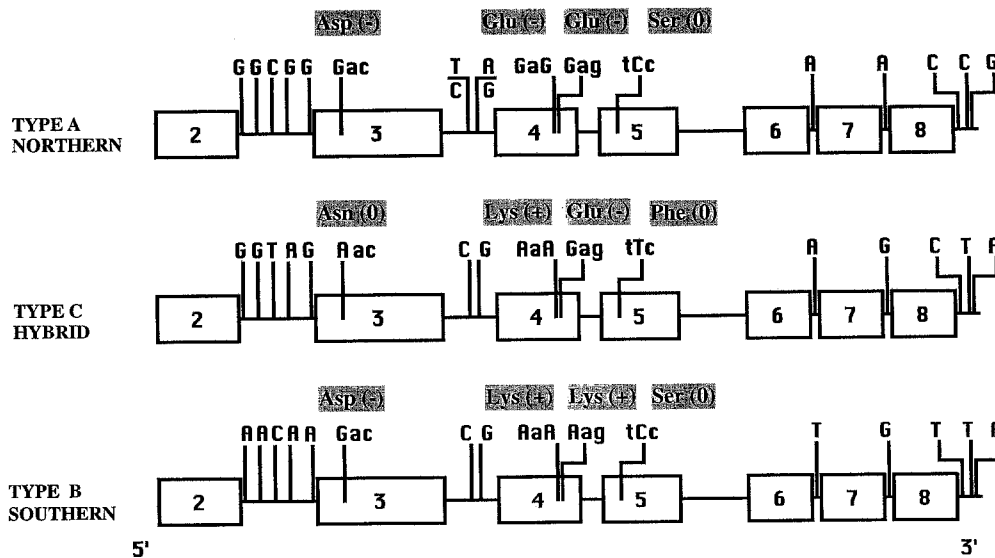


Fig. 2. Schematic comparison of the alleles of the 6-PGD gene in *Peromyscus californicus*. Base substitutions are indicated by capital letters at their approximate positions. (Introns have been telescoped to save space.) The amino acids encoded by variable codons, plus their relative charges, are indicated in the shaded boxes above the codons. Exons are numbered as in Fig. 1.

2. Exons 2 through 8 have been completely sequenced for all three alleles, but exon 1 has not been successfully amplified for the B and C alleles. A total of 17 substitutions were observed among the three alleles. All but one of these is a transition, and 11 are A \leftrightarrow G transitions (Fig. 2). Five of the substitutions are within exons, affecting a total of four codons. Two of the substitutions, in the intron between exons 3 and 4, are unique to one of the type A mice; these represent variation within the A electromorph (Fig. 2).

Codon changes caused by the substitutions appear to account for the observed differences in electrophoretic mobility of the alleles, though there may be more charge differences in the 30% of the coding sequence that remains unknown. If the B (slow) allele is taken as a standard, then the C (medium) allele is relatively more negative in charge; a negative aspartic acid changes to neutral asparagine (Fig. 2), but a positive lysine changes to negative glutamic acid, for a net change of -1 . The A (fast) allele is even more negative, and the two positive lysines in B have changed to negative glutamic acids, for a net change of -4 . Thus, the relative order of allelic migration can be explained on the basis of the detected charge differences alone.

Discussion

To reasonably infer the origin of the C allele, the distribution of sequence differences between it and the A and B alleles must be analyzed. At the protein level, the rare allele phenomenon is recognized by differences in exons alone, since differences in electrophoretic migration are

due to amino acid replacements. At the DNA level, however, differences within both exons and introns are important, since a mosaic pattern from intragenic recombination should be evident from each. The recombination hypothesis calls for a physical shuffling of DNA, which would rearrange introns as well as exons. Also, if only exons are considered, there might be too few differences between alleles to distinguish recombination from point mutation.

Whether exon sequences are examined alone or together with intron sequences, no unambiguous mosaic pattern is apparent from the substituted positions among the three alleles. For example, when one examines the five positions at which there are substitutions in exon 2 and 3, the A and C alleles are identical at three, the B and C alleles are identical at a fourth, and the A and B alleles are identical at the fifth. Similarly, within exon 4 the C allele shares one substitution with B and another, just nine nucleotides away, with A (Fig. 2). Overall, six of the substitutions unite the A and C alleles, eight unite B and C, and three unite A and B and are thus unique to the C allele.

The lack of an observed pattern provides no support for the intragenic recombination and gene conversion hypotheses, while the lack of additional DNA excludes the transposable element hypothesis. The 6-PGD hybridzyme seems to have arisen by the accumulation of base substitutions, which favors the hypothesis of increased mutation rates in hybrids. It is theoretically possible that this hybridzyme could have been formed by recombination of parental alleles unsampled in this study, either because they are electrophoretically indistinguishable from the sampled alleles or because they are no longer extant. Such a situation is highly improbable,

however; neither the A nor the B allele appears suitable as a parent allele for C and, hence, two cryptic parental alleles must be hypothesized to account for the origin of C by recombination. It is also possible that the hybridzyme was formed by a mixture of recombination and mutation events, with any recombinational pattern obscured by subsequent mutations.

We therefore conclude that, within the limitations of this study, the evidence indicates that the hybridzyme arose via a series of mutations. These results are consistent with those obtained by Bradley et al. (1993), who studied a hybridzyme of the pocket gopher *Geomys bur-sarius*. Though their study included only exons and, therefore, compared only a few differences between alleles, no pattern easily explainable by recombination was evident. Also, in their study the gene for the hybridzyme was very similar to one of the potential parental alleles and could clearly have arisen from it by simple base substitution.

This conclusion is somewhat surprising, since the recombination hypothesis is in many respects the most easily visualized of the alternatives and has generally been cited as the most likely. An explanation for why variant alleles should be more common in hybrid zones is inherent in the recombination hypothesis: hybrid zones are sites of contact between populations that have differentiated at many loci and therefore provide increased opportunities for recombination in F_1 heterozygotes. In contrast, the mutation hypothesis does not offer an obvious explanation of why substitutions should occur more often in hybrids. It should be noted that in the case of the *P. californicus* hybrid zone, the 6-PGD hybridzyme is limited to hybrid populations even though heterozygotes for the type A and type B alleles (and thus opportunities for recombination) are present throughout the southern subspecies.

Our results address only the first of the two fundamental questions raised by hybridzymes—that of how a variant allele arises. We have also examined only a single example of the hybridzyme phenomenon; additional hybridzymes must be studied to provide a more general assessment of the relative importance of the hypothesized mechanisms in hybridzyme formation and in the generation of allelic variants. If additional examples indicate that an increased rate of mutation underlies the formation of hybridzymes, then the focus should move to the second, more difficult question that asks why hybridzymes are so characteristic of hybrid zones.

Acknowledgments. We would like to thank L. Lynne Szura and Philip Myers for their advice and technical assistance. S.M.G.H. was supported in part by an NIH genetics training grant, the American

Society of Mammalogists, Sigma XI, and the U.S. Department of Energy.

References

- Barton NH, Halliday RB, Hewitt GM (1983) Rare electrophoretic variants in a hybrid zone. *Heredity* 50:139–146
- Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annu Rev Ecol Syst* 16:113–148
- Blin N, Stafford DW (1976) A general method for the isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Res* 3: 2303–2308
- Bradley RD, Bull JJ, Johnson AD, Hillis DM (1993) Origin of a novel allele in a mammalian hybrid zone. *Proc Natl Acad Sci USA* 90: 8939–8941
- Carne A, Walker JE (1983) Amino acid sequence of ovine 6-phosphogluconate dehydrogenase. *J Biol Chem* 258:12895–12906
- Dowling TE, Moore WS (1985) Genetic variation and divergence of the sibling pair of cyprinid fishes *Notropis cornutus* and *N. chrysocephalus*. *Biochem Syst Ecol* 13(4):471–476
- Hoffman SMG (1991) The molecular mechanisms underlying the “rare allele phenomenon” in a subspecific hybrid zone of the California mouse, *Peromyscus californicus*. Unpublished PhD thesis, University of Michigan, Ann Arbor
- Kass DK, Hoffman SMG (1993) Unusual patterns of susceptibility to degradation of DNA isolated from tissues in *Peromyscus californicus*. *Lab Anim* 27:273–277
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- McCabe T, Blanchard B (1950) Three species of *Peromyscus*. *Rood Assoc*, Santa Barbara
- Miksicek RJ, Towle HC (1983) Use of a cloned cDNA sequence to measure changes in 6-phosphogluconate dehydrogenase mRNA levels caused by thyroid hormone and dietary carbohydrate. *J Biol Chem* 258:9575–9579
- Nelson K, Baker RJ, Honeycutt RL (1987) Mitochondrial DNA and protein differentiation between hybridizing cytotypes of the white-footed mouse *Peromyscus leucopus*. *Evolution* 41:864–872
- Sage RD, Selander RK (1979) Hybridization between species of the *Rana pipiens* complex in Central Texas. *Evolution* 33:1069–1088
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-termination inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467
- Schilthuisen M, Gittenberger E (1994) Parallel evolution of an sAat-‘hybridzyme’ in hybrid zones in *Albinaria hippolyti* (Boettger). *Heredity* 73:244–248
- Selander RK, Smith MH, Yang SY, Johnson WE, Gentry JB (1971) Biochemical polymorphism and systematics in the genus *Peromyscus*. I. Variation in the old-field mouse (*Peromyscus polionotus*). *Stud Genet VI*, Univ Texas Publ 7103:49–90
- Smith MF (1978) Relationships between genetic variability and niche dimensions among coexisting species of *Peromyscus*. Unpublished PhD thesis, University of California, Berkeley
- Smith MF (1979) Geographic variation in genic and morphological characters in *Peromyscus californicus*. *J Mamm* 60:705–722
- Woodruff DS (1989) Genetic anomalies associated with *Cerion* hybrid zones: the origin and maintenance of new electromorphic variants called hybridzymes. *Biol J Linn Soc* 36(3):281–294
- Woodruff RC, Slatko BE, Thompson JN (1983) Factors affecting mutation rates in natural populations. In: Ashburner M, Carson HL, Thompson JN (eds) *Genetics and biology of Drosophila*: 3C. Academic Press, New York