Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 8939–8941, October 1993
Evolution

Origin of a novel allele in a mammalian hybrid zone

(hybridization/*Geomys*/DNA sequencing/alcohol dehydrogenase)

ROBERT D. BRADLEY*, JAMES J. BULL, ANDREW D. JOHNSON, AND DAVID M. HILLIS

Department of Zoology, University of Texas at Austin, TX 78712

Communicated by John C. Avise, July 6, 1993 (received for review May 10, 1993)

ABSTRACT **The occurrence of rare or novel alleles has been documented in at least 23 different hybrid zones spanning vertebrate and invertebrate taxa. As most novel alleles either occur at high frequencies in hybrid populations or are exclusively restricted to hybrids, it has seemed probable that hybridization has a role in their origin; however, the molecular nature of these novel alleles and the mechanisms responsible for their origin remain obscure. We examined the complete coding sequences of six alleles of alcohol dehydrogenase in a mammalian hybrid zone between two species of pocket gophers (*Geomys*). One of these sequences encodes a novel electromorph that had been identified in earlier allozyme studies; this novel allele differs from one of the parental alleles by a single base substitution. This substitution generates an amino acid replacement that affects the net charge of the translated protein. This resultant charge change is congruent with the observed allozyme mobility patterns. Our data thus provide evidence for simple DNA substitution as a mechanism for the origin of this novel hybrid-zone allele.**

Hybrid zones have attracted the attention of evolutionary biologists because the interbreeding of closely related species may yield unique insights into the processes of speciation (1, 2). Allozyme electrophoresis studies have revealed numerous instances of novel alleles that were found in a hybrid zone but were absent from the populations of both parental types (2–7). The phenomenon of novel hybrid-zone alleles (also called hybrizymes) is taxonomically broad, and it has attracted attention as a possible indicator of unusual genetic processes maintaining the genetic integrity of species. Speculations as to how novel hybrid-zone alleles arise have consequently been diverse: intracistronic recombination (8–10), gene conversion (11), transposable element activity (12, 13), and posttranslational modification of the protein product (3, 14) have all been suggested as mechanisms for the generation of novel alleles. As all reported novel alleles have been identified by standard allozyme electrophoresis, it would appear that the differences among the novel and common alleles (electromorphs) stem from changes that have accrued in coding or exon regions of the gene. Consequently, for any of the proposed mechanisms of novel allele origin to remain viable, they would have to exert their effect in coding regions. To date, nucleic acid sequences have not been obtained from any of the novel alleles, so it has been difficult to support or eliminate any of these hypotheses for specific cases.

In this study, we examined a novel allele at the alcohol dehydrogenase (*Adh-1*) locus from a hybrid zone between two species of pocket gophers, *Geomys bursarius major* and *Geomys knoxjonesi* (2). Previous allozyme studies indicated the presence of two parental-type electromorphs (M and K), as well as a novel electromorph (N) that is restricted to the hybrid zone (2). The frequency of the novel electromorph within the hybrid zone is ≈0.15 both in hybrids (6 of 41) and in parental types (5 of 34), where hybrid/parental status was based on four other nuclear markers (chromosomes, ribosomal DNA, peptidase, and lactate dehydrogenase) and mitochondrial DNA markers. To date, the novel electromorph has been found only in heterozygotes with either parental electromorph (MN and KN); the N electromorph has never been found outside the hybrid zone in extensive allozyme surveys (2).

We cloned and sequenced† the cDNAs of both alleles of the *Adh-1* gene from each of four individuals. Based on allozyme patterns, two of these individuals were regarded as parental types (and were collected from outside the contact zone), whereas the other two individuals (hybrids within the contact zone) were heterozygous for the novel electromorph and one of the parental electromorphs (2).

MATERIALS AND METHODS

RNA was isolated from liver tissue of four individuals (30715, K electromorph; 30511, KN heterozygote; 30682, MN heterozygote; and 30723, M electromorph), and oligo(dT) complementary to the poly(A) tail was used as a primer with reverse transcriptase to generate cDNA. *Adh-1* sequences from *Rattus*, *Mus*, and *Homo* were aligned, and conserved regions were used to design oligonucleotide primers incorporating the 5′ and 3′ ends of the *Adh-1* gene for amplification by polymerase chain reaction (PCR) techniques (15, 16) with mRNA as the initial template. The two primer sequences used were 5′-ATGAGCAC(A/T)GC(T/G)GGAAAAG-TAA-3′ and 5′-ATCCGTACCGTCCTGAC(T/G)TTCTGA-3′. PCR products were cloned into pBluescript (Stratagene) and sequenced by dideoxy DNA sequencing techniques (17). Multiple clones (Fig. 1) were sequenced to distinguish allelic differences from PCR artifacts. The *Adh-1* sequences were aligned visually (there were no insertion/deletion events within the coding sequences) and substitutions, amino acid replacement, and charge changes were identified (Figs. 1 and 2).

RESULTS AND DISCUSSION

All three electromorphic classes were found to correspond to distinct cDNA sequences. The M-class electromorphs included three distinguishable cDNA sequences (designated M, M′, and M″), and the K-class electromorphs included two distinct cDNA sequences (designated K and K′). The two sequenced copies of the N-class (novel) allele from different individuals (30511 and 30682) were identical. All alleles had identical lengths for the coding region (1128 bp, including the start and stop codons; Fig. 1). These alleles vary by base substitutions at nine positions (Fig. 2); five are missense (sites

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

*Present address: Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX 77843.

†The sequences reported in this paper have been deposited in the GenBank data base (accession nos. L15461–L15466).

```
                                                                                                  A
   1 ATGAGCACTGCGGGAAAAGTAATAAAGTGC AGAGCAGCTGTGTTGTGGGAGAAAAACAAA CCCTTCTCCATCGAGGAGGTGGAAGTTGTG CCCCCCAAAGCCTATGAAGTTCGTATAAAG  120
     MetSerThrAlaGlyLysValIleLysCys ArgAlaAlaValLeuTrpGluLysAsnLys ProPheSerIleGluGluValGluValVal ProProLysAlaTyrGluValArgIleLys

 121 ATAGTGGCCACGGGAATCTGTCGCTCAGAT GACCACGTGGTGAATGGAAGCATAATCACG CCTCTTCCTGCGATTCTAGGCCACGAGGCA GGCGGCATCGTGGAGAGCATTGGAGAAGGA  240
     IleValAlaThrGlyIleCysArgSerAsp AspHisValValAsnGlySerIleIleThr ProLeuProAlaIleLeuGlyHisGluAla GlyGlyIleValGluSerIleGlyGluGly

                                                                                                          B     C
 241 GTGACTACAGTGAAACCAGGTGATAAAGTC ATCCCTCTCTTCGTGCCGCAGTGTGGAAAG TGCAGGGCTTGCAAACACCCGGAGAGCAAC TTATGCACGCATGGCGATCTGGGGAGGGCC  360
     ValThrThrValLysProGlyAspLysVal IleProLeuPheValProGlnCysGlyLys CysArgAlaCysLysHisProGluSerAsn LeuCysThrHisGlyAspLeuGlyArgAla

                        D                    E                                                       F
 361 CAGGGCACCCTGATGGATGGCACCAGCAGA TTCACCTGCAAGGGGAAGCCTATTCACCAC TTCCTGGGCGTCACGACCTTCAGTGAGTAC ACGGTGGTGAGTGAGATCTCCGTGACCAAG  480
     GlnGlyThrLeuMetAspGlyThrSerArg PheThrCysLysGlyLysProIleHisHis PheLeuGlyValThrThrPheSerGluTyr ThrValValSerGluIleSerValThrLys

        G                                                                H        I
 481 ATCGACGCCGCGTCGCCCCTGGAGAAAGTC TGCCTCATCGGCTGCGGGTTTTCCACCGGC TACGGCTCTGCGGTCAAAGTAGGCAAGGTC GCTCGGGGTTCCATCTGTTCCTGTGTTTGG  600
     IleAspAlaAlaSerProLeuGluLysVal CysLeuIleGlyCysGlyPheSerThrGly TyrGlySerAlaValLysValGlyLysVal AlaArgGlySerIleCysSerCysValTrp

 601 TCTGGGAGGGTTGGTCTGTCTGCTATCATT GGCTGTAAAGCAGCAGGAGCCGCCAGAATC ATTGCAGTGGACATCAACAAGGACAAATTT GCAAAAGCCAAAGAGTTGGGTGCCACTGAG  720
     SerGlyArgValGlyLeuSerAlaIleIle GlyCysLysAlaAlaGlyAlaAlaArgIle IleAlaValAspIleAsnLysAspLysPhe AlaLysAlaLysGluLeuGlyAlaThrGlu

 721 TGTGTCAATCCCCAAGACTATGACAAGCCC ATCTATCAGGTGCTGCAGGAAATGACTGAT GGAGGTGTGGATTTCTCCTTTGAAGTCATT GGTCGACTTGACACCATGGTTTCTGCCCTA  840
     CysValAsnProGlnAspTyrAspLysPro IleTyrGlnValLeuGlnGluMetThrAsp GlyGlyValAspPheSerPheGluValIle GlyArgLeuAspThrMetValSerAlaLeu

 841 ATGTGCTGCCAAGAGTCCCATGGCGTCAGC GTCATTGTTGGAGTCCCTCCCAACGCTCAA AGCCTCACTATGGATCCCATGGTGCTGCTG AGTGGGCGCTCCTGGAAAGGCGCTGTGTTT  960
     MetCysCysGlnGluSerHisGlyValSer ValIleValGlyValProProAsnAlaGln SerLeuThrMetAspProMetValLeuLeu SerGlyArgSerTrpLysGlyAlaValPhe

 961 GGTGGTTATAAGGGTAAAGATGATGTCCCC AAACTGGTGGCTGATTTTATGGCAAAGAAG TTCCCATTGGAGCCATTAATAACCAATGTT TTCCCCTTTGCAAAAATAAATGAAGGATTT 1080
     GlyGlyTyrLysGlyLysAspAspValPro LysLeuValAlaAspPheMetAlaLysLys PheProLeuGluProLeuIleThrAsnVal PheProPheAlaLysIleAsnGluGlyPhe

1080 GACCTCCTTCGCGCTGGAAAGAGCATCCGT ACCGTCCTGACTTTCTGA 1128
     AspLeuLeuArgAlaGlyLysSerIleArg ThrValLeuThrPheEnd
```

FIG. 1. The complete coding sequence and inferred amino acid sequence for allele K of *Adh-1* in *G. knoxjonesi*. The nine sites that show sequence variation among the six sequenced *Adh-1* alleles are labeled A-I; variation at these sites is detailed in Fig. 2.

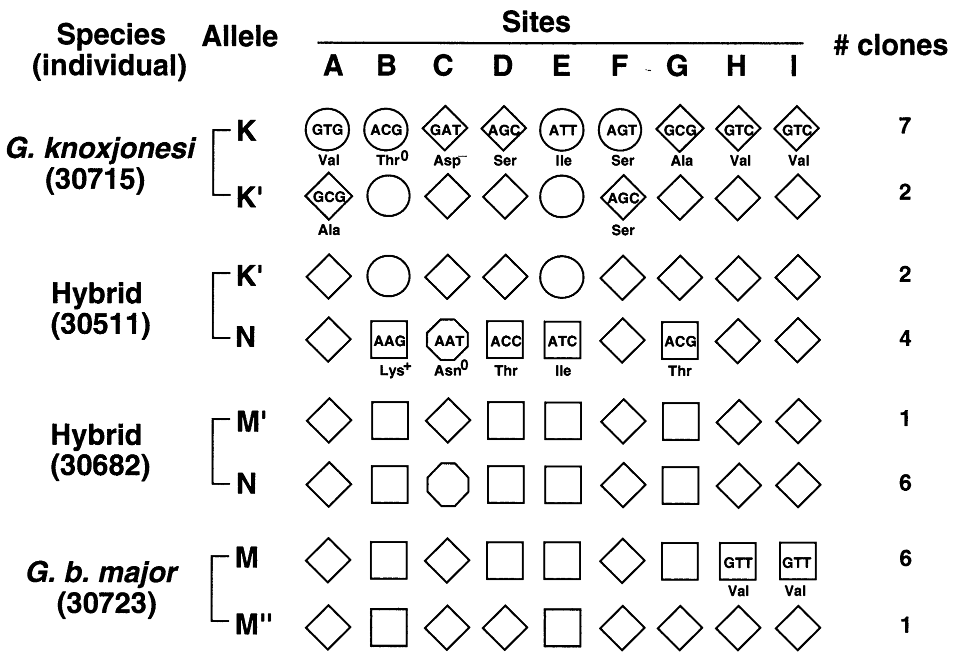

FIG. 2. Nucleotide and amino acid sequence variation among six alleles of *Adh-1* in *Geomys*. Circles represent *G. knoxjonesi* codons, squares represent *G. b. major* codons, octagons represent the codon found only in the novel allele, and diamonds represent codons shared between both species. DNA sequences corresponding to amino acids are included inside the symbols, with the amino acid identified below that symbol. Sites A, D, E, F, G, H, and I (locations shown in Fig. 1) do not generate any net change in the charge of the alleles, whereas B and C produce charge changes. These two charge differences result in the M, M′, M″, and N alleles possessing the amino acid lysine (positive charge) at site B, whereas K and K′ possess threonine (neutral) at this site; at site C, alleles M, M′, M″, K, and K′ possess aspartic acid (negative charge) and N has asparagine (neutral).

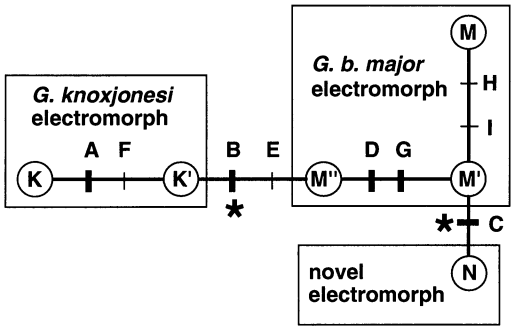

FIG. 3. Inferred evolutionary relationships among six *Adh-1* alleles identified in this study. Placement of bars indicates the substitutions necessary to generate each respective allele. Thick bars (A, B, C, D, and G) indicate missense substitutions; thin bars (E, F, H, and I) represent silent substitutions. Asterisks indicate substitutions (B and C) that led to charge changes in the protein. Boxes circumscribe the three mobility classes (K, M, and N). Note that the novel allele (N) is best explained by a single base substitution from the *G. b. major* electromorphic class.

A, B, C, D, and G) and four are silent (sites E, F, H, and I). All sequence differences within an electromorph class comprise either silent substitutions or missense mutations that do not change the charge of the inferred protein. Therefore, it is not surprising that the polymorphisms in the K and M class were not detected in the original allozyme study (2). An amino acid replacement at site B (Fig. 2) changing threonine (neutral) to lysine (positive charge) alters the net charge of the inferred protein molecule that distinguishes electromorph K from both M and N. An additional replacement further distinguishes electromorph N from both K and M (at site C). Site C generates a change in the net charge as aspartic acid (negative charge) in the parental alleles (K and M) is replaced by asparagine (neutral) in the novel allele (N). All three electromorphs have a net positive charge, but the replacements at sites B and C produce the strongest positive charge in the N electromorph, followed by the M electromorph, followed lastly by the K electromorph. These changes in total charge are congruent with the mobility pattern seen in allozyme electrophoresis of these three alleles.

Construction of an "allele tree" of these six alleles using the nine substitutions provides a model of the relationships and ancestry among the three mobility classes M, K, and N (Fig. 3). All nine substitutions are consistent with this tree. The novel allele (N) is only one step (nucleotide substitution C) away from the *G. b. major* electromorph class (allele M′) and is five steps away from the *G. knoxjonesi* electromorph class (allele K′). The simplest explanation for the origin of the novel allele (N) is through a single substitution from the M′ allele.

These data allow us to reject hypotheses concerning transposable-element activity, intracistronic recombination and/or gene conversion, and posttranslational modification as plausible mechanisms for the origin of the novel allele N. First, there is no evidence for transposition of any DNA sequence into the coding region, as the sequences of all six alleles are the same length and differ only by the nine base substitutions. Second, recombination and/or gene conversion cannot explain the differences seen at site C (the site that results in the distinct mobility of the N allele), as all parental alleles are identical and N is unique at this site. Third, results from the DNA sequence data suggest that the changes at sites B and C that are responsible for charge differences are adequate to generate the mobility patterns observed in these alleles, so posttranslational modification is unnecessary to explain the observations.

We conclude that the novel allele at the *Adh-1* locus in this hybrid zone is a result of one DNA base substitution that generated a single amino acid replacement, which in turn produced different mobility rates between the parental and novel alleles. Data from this study also suggest a higher diversity of alleles present within natural populations than were detected by standard allozyme techniques. Additional hybrid zones should be examined to determine whether simple DNA base substitution is the primary mechanism responsible for the generation of novel alleles. If so, mechanisms that maintain (rather than generate) novel alleles in hybrid zones (see refs. 1 and 18) should be studied more extensively as a possible explanation for the high frequency of novel alleles in hybrid zones.

We thank Robert J. Baker (The Museum, Texas Tech University) for providing tissue samples and Paul Krieg, Tod Reeder, Cliff Cunningham, and Marty Badgett for advice and assistance. This work was supported in part by funds from the National Science Foundation (to D.M.H.).

1. Barton, N. H. & Hewitt, G. M. (1989) *Nature (London)* **341,** 497–503.
2. Baker, R. J., Davis, S. K., Bradley, R. D., Hamilton, M. J. & Van Den Bussche, R. A. (1989) *Evolution* **43,** 63–75.
3. Woodruff, D. S. (1989) *Biol. J. Linn. Soc.* **36,** 281–294.
4. Smith, M. F. (1979) *J. Mammal.* **60,** 705–722.
5. Greenbaum, I. F. (1981) *Evolution* **36,** 306–321.
6. Barton, N. H., Halliday, R. B. & Hewitt, G. M. (1983) *Heredity* **50,** 139–146.
7. Nelson, K., Baker, R. J. & Honeycutt, R. L. (1987) *Evolution* **41,** 864–872.
8. Ohno, S., Stenius, C., Christian, L. & Schipmann, G. (1969) *Biochem. Gen.* **3,** 417–428.
9. Watt, W. B. (1972) *Am. Nat.* **106,** 737–753.
10. Sage, R. D. & Selander, R. K. (1979) *Evolution* **33,** 1069–1088.
11. Hillis, D. M., Moritz, C., Porter, C. A. & Baker, R. J. (1990) *Science* **251,** 308–310.
12. Bregliano, T. C. & Kidwell, M. G. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 363–410.
13. Kidwell, M. G. (1990) *Can. J. Zool.* **68,** 1716–1726.
14. Murphy, R. W., McCollum, F. C., Gorman, G. C. & Thomas, R. (1984) *J. Herpetol.* **18,** 93–105.
15. Mullis, K. B. & Faloona, F. A. (1987) *Methods Enzymol.* **155,** 335–350.
16. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239,** 487–491.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Hunt, W. G. & Selander, R. K. (1973) *Heredity* **31,** 11–33.