Letter to the Editor

Parallel Molecular Evolution of Deletions and Nonsense Mutations in Bacteriophage T7

C. W. Cunningham, K. Jeng,† J. Husti,† M. Badgett,† I. J. Molineux,‡ D. M. Hillis,† and J. J. Bull†*

*Zoology Department, Duke University; †Department of Zoology, University of Texas; and ‡Department of Microbiology, University of Texas

The extent to which identical evolutionary changes appear in closely related, independent lineages, a phenomenon known as parallel evolution, is a central question in evolutionary biology (Kimura 1983; Gould 1989). Inferring parallel evolution from extant lineages is instructive (e.g., Swanson et al. 1991; Doyle, Doyle, and Palmer 1995; Shyue et al. 1995), but may be misleading if rates are high enough to bias either the reconstruction of ancestral character states or reconstruction of the phylogeny itself (Felsenstein 1978, 1985). The surest way to detect parallel evolution is with experimental studies of microorganisms. Although previous studies of experimental evolution have documented parallel phenotypic changes, these changes have not generally been documented at the DNA sequence level (e.g. Dykhuizen 1990; Travisano et al. 1995).

Phylogenies of bacteriophage T7 grown in the presence of a mutagen have proven to be an unusually flexible experimental system to evaluate phylogenetic methods (Hillis et al. 1992; Hillis, Huelsenbeck, and Cunningham 1994), and to study molecular evolution (Studier 1980; Bull et al. 1993). This system not only displays very rapid DNA sequence evolution, but allows the direct observation of ancestral states during the evolution of each lineage. Here we document a complex pattern of parallel evolution at the DNA sequence level. Our results suggest caution when reconstructing ancestral states of characters under directional selection, as well as caution against giving undue phylogenetic weight to insertion-deletion events.

We serially propagated six bifurcating lineages of bacteriophage T7 in the presence of a mutagen according to an earlier protocol (fig. 1; Hillis et al. 1992). The lineages were bottlenecked to a single individual at each node shown in figure 1. This approach allowed the ancestral condition to be known with certainty at several points in the history of each lineage. Our analysis focused on DNA sequences from a region that, in wild-type phage, spans just over 2 kb and contains five genes 0.3, 0.4, 0.5, 0.6, and 0.7 (Dunn and Studier 1983: DNA alignments deposited in EMBL. Aligned 0.3 sequences as DS 27143, aligned 0.7 sequences as DS 27142).

Although our wild-type ancestral stock contained the entire 0.3–0.7 region, every lineage evolved a ~1.5-kb deletion that fused the 0.3 and 0.7 genes. Nine independent deletions were observed, but seven of them had breakpoints identical to the previously characterized H1 deletion (fig. 1–2; Studier 1973; Studier et al. 1979). Great care was taken to recognize cases of contamination. DNA substitutions appearing early in each lineage were followed throughout propagation to ensure the integrity of each lineage (see alignments submitted to EMBL). To the same end, *Mbo* I restriction sites were also mapped for approximately half of the genome (unpublished data). The frequent appearance of the H1 deletion in this study is consistent with its appearance in numerous "wild-type" T7 stocks maintained independently (Studier 1973; Studier et al. 1979).

The H1 deletion removes all of genes 0.4–0.6 and creates an in-frame fusion of the 0.3 and 0.7 genes (fig. 2A). This novel open reading frame is expressed as a fusion protein *in vivo* (Studier 1973). Of the two genes in the fusion protein, only the function of the 0.3 gene—overcoming host Type I restriction enzymes—is necessary for growth under our experimental conditions (Moffatt and Studier 1988). There is no known function for genes 0.4–0.6, and no known phenotype associated with their loss. There is also no known cost associated with the loss of the known functions of the 0.7 protein (see discussion below).

Every lineage which experienced an H1 deletion also evolved nonsense mutations in the remaining portion of the 0.7 gene (figs. 1 and 2). A total of 14 nonsense mutations arose independently—confined to just seven nucleotide positions in seven amino acid sites–resulting in additional parallel evolution among the lineages (figs. 1 and 2).

Parallel appearances of deletions and nonsense mutations likely result from a combination of directional selection for a particular genetic or phenotypic change, combined with constraints imposed by a limited spectrum of available mutations that can respond to this selection. It is clear that deletions are advantageous from their frequently observed, rapid evolution during growth in rich media (Studier 1973; Studier et al. 1979; Hillis et al. 1992). Significantly, the only lineages we have observed which never experienced a deletion event were those in which the action of selection was dramatically reduced by bottlenecking to a single individual after every lytic cycle (White et al. 1991).

A large variety of potential deletions should be equally beneficial, since the main constraints appear to be to preserve the functions of genes 0.3 and 1, the latter lying immediately downstream of 0.7 and being essential for phage growth. However, there may be a limited array of deletions that can appear, since they typically occur during recombination between or misreplication across direct repeat sequences (Studier et al. 1979; Moffatt and Studier 1988; Kong and Masker 1994). The frequent appearance of the H1 deletion is likely the result of an especially long (13 bp) repeated sequence at its

Key words: parallel evolution, experimental evolution, deletions, nonsense mutations, fusion proteins, bacteriophage T7.

Address for correspondence and reprints: Cliff Cunningham, Zoology Department, Duke University, Durham, North Carolina 27708. E-mail: cliff@acpub.duke.edu.

Mol. Biol. Evol. 14(1):113–116. 1997

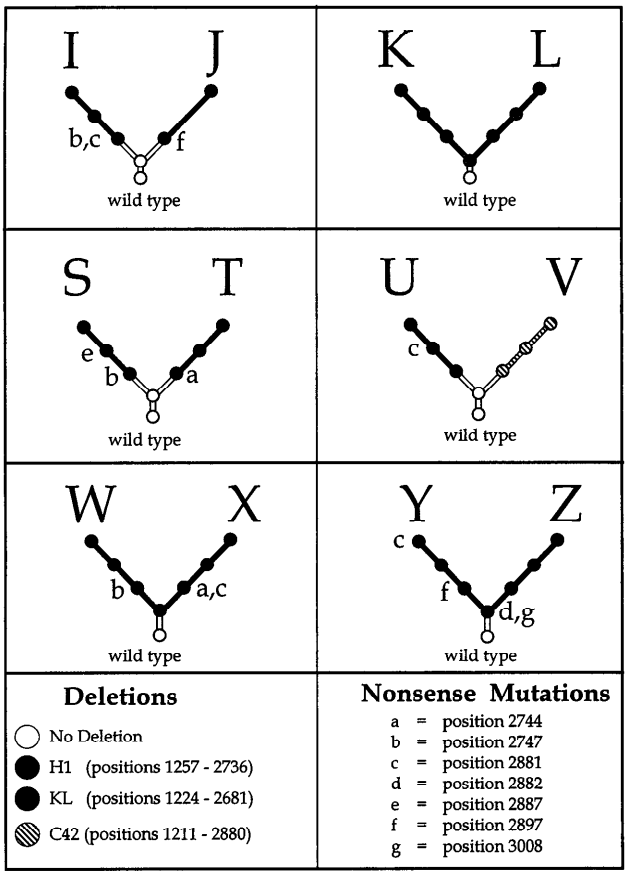

FIG. 1.—Parallel evolution of deletions and nonsense mutations in six bifurcating phylogenies of bacteriophage T7 propagated in the presence of the mutagen nitrosoguanidine. The number of phage passaged from lysate to lysate was in excess of 100,000, except at each of the nodes (designated by a circle) where lineages were bottlenecked to a single plaque. DNA sequences were obtained at every node (designated by circles) except for one node in the J lineage which was lost. In every lineage, a deletion removed between 3.6% and 4.2% of the genome. Each deletion created a fusion protein joining the 0.3 and 0.7 genes. Two of the deletions (KL and C42) are out of frame, and the frameshifts introduce stop codons soon after the breakpoint (3 and 10 codons after, respectively). The most common deletion (H1) joins the 0.3 and 0.7 genes in frame, forming a fully translated fusion protein (see fig. 2*A*). Although nonsense mutations were never observed in the 0.3 portion, the 0.7 portion of the H1 fusion protein was always rapidly truncated by nonsense mutations (a–g above). Only nonsense mutations appearing in the H1 lineages are shown. All position numbers are given as in the published sequence of bacteriophage T7 (Dunn and Studier 1983). The number of lytic cycles between each node is 50, except for the branch between the wild-type ancestor and the bifurcation (10 lytic cycles for IJ and KL, 20 for ST and UV, 30 for WX and YZ).

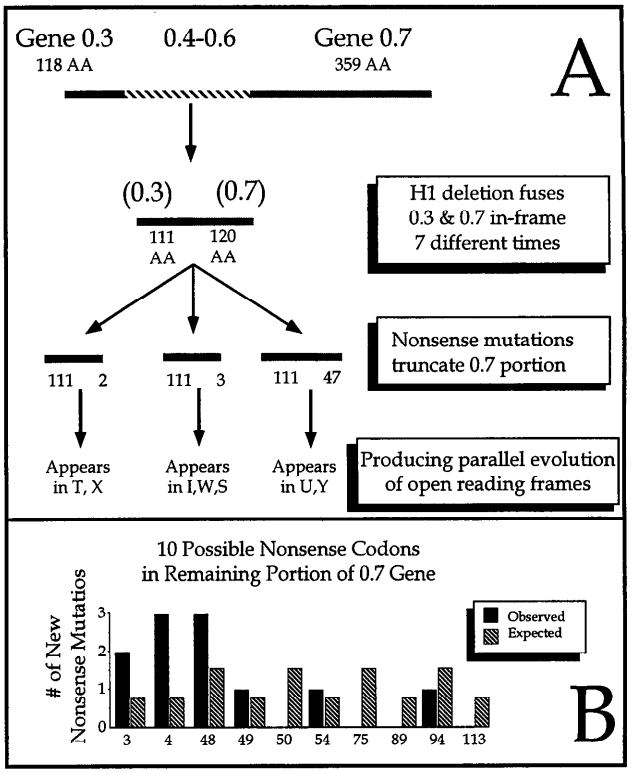

FIG. 2.—A two-fold pattern of deletions and nonsense mutations produces three identical open reading frames in independent lineages (*A*). Two other open reading frames were observed once each in the J and Z lineages. These identical open reading frames are produced by a significant 5′ bias in the appearance of nonsense mutations (*B*). The GC→AT mutation spectrum of the mutagen (Bull et al. 1993) constrains the appearance of nonsense mutations to only 14 sites in 10 codons, designated by their codon number in the 0.7 portion of the open reading frame. The 5′ positional bias was significant in a 3 × 2 *G*-test ($P < 0.02$). The test was carried out by dividing into three groups the 14 sites which could become a nonsense mutation with a GC→AT change (positions 3–4, 48–54, 75–113). Expected values were calculated assuming all 14 possible sites had an equal probability of substitution.

endpoints, since other characterized deletions in bacteriophage T7 (Studier et al. 1979), including the C42 and KL deletions described here, have repeated sequences of less than 10 bp.

Like the deletions themselves, parallel appearances of nonsense mutations in the remaining portion of the 0.7 gene also appear to result from a combination of constraints and directional selection. These nonsense mutations produced identical independent open reading frames in independent lineages (fig. 2*A*). In any DNA sequence, only a few codons can be transformed into a stop codon by a single nonsense mutation. In our system, only 14 sites (10 codons) in the residual portion of 0.7 can mutate to a stop codon with the GC→AT changes favored by our mutagen (Studier 1980; Bull et al. 1993). Two lines of evidence indicate that selection favors nonsense mutations that truncate the 0.7 portion of the H1 open reading frame.

First, the ratio of nonsense to other mutations in the period up to and including the first nonsense mutation in 0.7 is significantly greater than expected under a null model of equal change across sites (11/24 = 0.46 observed vs. 0.08 expected; $P < 0.0001$). This ratio declined significantly after the first nonsense mutation appeared (3/63 = 0.05; $P < 0.0001$). Significance in both tests was determined by simulating the null distribution of the statistics, assuming only GC→AT changes, and allowing all G/C sites in the ancestral sequence to have an equal probability of substitution. The expectations after the appearance of the first nonsense were adjusted to take into account mutations that had already appeared. Second, the distribution of nonsense codons that evolved was more 5′ in 0.7 than would be expected from a random choice among the 14 possible sites (fig. 2*B*,

$P < 0.02$, G-test). This selection for greater truncation further enhances parallel evolution of nonsense mutations (fig. 2*A*).

Modern comparative methods are designed to discover correlations between characters which appear together in parallel lineages (e.g., Maddison 1990; Harvey and Pagel 1991). Because we never observed nonsense mutations in the 0.7 gene prior to the appearance of the H1 deletion (fig. 1), it is possible that selection for nonsense mutations depends on the prior existence of the H1 deletion. The null hypothesis of independence between H1 deletions and nonsense mutations was rejected by applying Maddison's (1990) concentrated changes test ($P < 0.04$: cumulative probability from the product of probabilities calculated independently for each lineage experiencing an H1 deletion).

The selective advantage of deletions removing all or part of genes 0.4–0.7 may be explained if these nonessential genes impose a cost on the phage. This cost may result either from an energetic burden imposed by replication or translation of these genes, or by physiological disruption associated with their expression (Lenski and Nguyen 1985). Work in other experimental systems suggests that energetic burden may not be sufficient to explain the rapid appearance of these deletions (Dykhuizen 1978; Lenski and Nguyen 1985). The selective advantage of nonsense mutations truncating the fusion protein caused by the H1 deletion is more complex. The C-terminal host transcription shut-off activity of 0.7 is known to remain active in the H1 fusion protein (Rothman-Denes et al. 1973; Rahmsdorf et al. 1974). If this activity is deleterious, nonsense mutations preventing translation of this domain should be advantageous. Alternatively, the additional amino acids from 0.7 may impair (but not entirely disrupt) the function of the important 0.3 protein. This impairment might also favor nonsense mutations truncating the 0.7 portion of the fusion protein (fig. 2*A*). Interestingly, work in *E. coli* has shown that progressively truncating a fusion protein can enhance the activity of the remaining intact protein (Bassford, Silvahy, and Beckwith 1979).

Our observations of widespread parallel evolution at the molecular level have two general lessons for the practice of inferring evolutionary history. First, large insertion-deletion events such as the ones we observed are often considered to be strong evidence for common ancestry and have been weighted heavily in phylogenetic analyses (Lavin, Doyle, and Palmer 1990; Rivera and Lake 1992; Kwiatowski et al. 1994; but see Doyle, Doyle and Palmer 1995). This study provides a compelling reason to avoid the assumption that parallel evolution of deletions is rare until the mechanisms underlying insertions and deletions are better understood. This requires not only nucleotide sequence information but also an appreciation of the biology of the organism and of the biochemical or physiological consequences of the deletion. For example, the mechanism producing the parallel deletions in our system—recombination across direct repeats—is not unique to bacteriophage and has been widely reported in other systems, ranging from *E. coli* to mitochondrial lineages in aging vertebrates (Cai et al. 1995; Oldfors et al. 1995; Van Tuyle et al. 1996).

Second, the appearance of the H1 deletion in multiple lineages demonstrates how parallel evolution caused by directional selection can mislead the reconstruction of ancestral character states (fig. 1). The presence of the 0.3–0.7 region in the common ancestor of all our lineages would be impossible to reconstruct from the terminal isolates (figs. 1 and 2*A*). Ancestral reconstruction would be especially misleading in the IJ and ST lineages, where H1 deletions were acquired independently in sister taxa (fig. 1). For example, if we relied on ancestral reconstructions, the possible lack of independence described above between the H1 deletion and subsequent nonsense mutations would not have been detected by the concentrated changes test ($P = 1.0$; Maddison 1990). Studies of character evolution are often based on reconstructing ancestral characters thought to be under directional selection, and therefore prone to parallel evolution (Brooks and McLennan 1991; Harvey and Pagel 1991; Maddison and Maddison 1992). More information about systems with known ancestors is necessary to improve our understanding of comparative methods.

Acknowledgments

We thank J. Greenway for restriction mapping our lineages. We also thank J. Coyne, D. Emlen, C. Laurie, R. Lenski, J. Thorne, and two anonymous reviewers for comments on the manuscript. This research was funded with the support of an NSF grant (DEB 9508987) and an NRSA grant to C. Cunningham (SF 32GM15074-02).

LITERATURE CITED

BASSFORD, P. J., T. J. SILVAHY, and J. R. BECKWITH. 1979. Use of gene fusion to study secretion of maltose-binding protein in *Escherichia coli* periplasm. J. Bacteriol. **139**:19–31.

BROOKS, D. R., and D. A. MCLENNAN. 1991. Phylogeny, ecology, and behavior. The University of Chicago Press, Chicago and London.

BULL, J. J., C. W. CUNNINGHAM, I. J. MOLINEUX, M. R. BADGETT, and D. M. HILLIS. 1993. Experimental molecular evolution of bacteriophage T7. Evolution **47**:993–1007.

CAI, H., H. Yu, K. MCENTEE, T. A. KUNKEL, and M. F. GOODMAN. 1995. Purification and properties of wild-type and exonuclease-deficient DNA polymerase II from *Escherichia coli*. J. Biol. Chem. **270**:15327–15335.

DOYLE, J. J., J. L. DOYLE, and J. D. PALMER. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. Syst. Bot. **20**:272–294.

DUNN, J. J., and F. W. STUDIER. 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. **166**:477–535.

DYKHUIZEN, D. 1978. Selection for tryptophan auxotrophs of *Escherichia coli* in glucose-limited chemostats as a test of the energy conservation hypothesis of evolution. Evolution **32**:125–150.

———. 1990. Experimental studies of natural selection in bacteria. Annu. Rev. Ecol. Syst. **21**:373–398.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**: 401–410.

———. 1985. Phylogenies and the comparative method. Am. Nat. **125**:1–15.

GOULD, S. J. 1989. Wonderful life. Norton and Company, New York.

HARVEY, P. H., and M. D. PAGEL. 1991. The comparative method in evolutionary biology. Oxford University Press, Oxford.

HILLIS, D. M., J. J. BULL, M. E. WHITE, M. R. BADGETT, and I. J. MOLINEUX. 1992. Experimental phylogenetics: generation of a known phylogeny. Science **255**:589–592.

HILLIS, D. M., J. HUELSENBECK, and C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenies. Science **264**:671–676.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KONG, D., and W. MASKER. 1994. Deletion between direct repeats in T7 DNA stimulated by double-strand breaks. J. Bacteriol. **176**:5904–5911.

KWIATOWSKI, J., D. SKARECKY, K. BAILEY, and F. J. AYALA. 1994. Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu, Zn *Sod* gene. J. Mol. Evol. **38**:443–454.

LAVIN, M., J. J. DOYLE, and J. D. PALMER. 1990. Systematic and evolutionary significance of the loss of the large chloroplast DNA inverted repeat in the family Leguminosae. Evolution **44**:390–402.

LENSKI, R. E., and T. T. NGUYEN. 1988. Stability of recombinant DNA and its effects on fitness. TREE **3**:S18–S20.

MADDISON, W. P. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? Evolution **44**:539–557.

MADDISON, W. P., and D. R. MADDISON. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, Mass.

MOFFATT, B. A., and F. W. STUDIER. 1988. Entry of bacteriophage T7 DNA into the cell and escape from host restriction. J. Bacteriol. **170**:2095–2105.

OLDFORS, A., A.-R. MOSLEMI, I.-M. FYHR, E. HOLME, N.-G. LARSSON, and C. LINDBERG. 1995. Mitochondrial DNA deletions in muscle fibers in inclusion body myositis. J. Neuropathol. Exp. Neurol. **54**:581–587.

RAHMSDORF, H. J., S. H. PAI, H. PONTA, P. HERRLICH, R. ROSKOWSKI, M. SCHWEIGER, and F. W. STUDIER. 1974. Protein kinase induction in *E. coli* bacteriophage T7. Proc. Natl. Acad. Sci. USA **71**:586–589.

RIVERA, M. C., and J. A. LAKE. 1992. Evidence that eukaryotes and eocyte prokaryotes are immediate relatives. Science **257**:74–76.

ROTHMAN-DENES, L. B., S. MUTHUKRISHNAN, R. HASELKORN, and F. W. STUDIER. 1973. A T7 gene function required for shut-off of host and early T7 transcription. Pp. 227–239 *in* C. F. FOX and W. S. ROBINSON, eds. Virus research. Academic Press, New York.

SHYUE, S.-K., D. HEWETT-EMMETT, H. G. SPERLING, D. M. HUNT, J. K. BOWMAKER, J. D. MOLLON, and W.-H. LI. 1995. Adaptive evolution of color vision genes in higher primates. Science **269**:1265–1267.

STUDIER, W. F. 1973. Genetic analysis of non-essential bacteriophage T7 genes. J. Mol. Biol. **79**:227–236.

———. 1980. The last of the T phages. Pp. 72–78 *in* N. H. HOROWITZ and E. J. HUTCHINGS, eds. Genes, cells, and behavior: a view of biology fifty years later. W. H. Freeman, San Francisco.

STUDIER, W. F., A. H. ROSENBERG, M. N. SIMON, and J. J. DUNN. 1979. Genetic and physical mapping in the early region of bacteriophage T7 DNA. J. Mol. Biol. **135**:917–937.

SWANSON, K. W., D. M. IRWIN, and A. C. WILSON. 1991. Stomach lysozyme gene of the Langur monkey tests for convergence and positive selection. J. Mol. Evol. **33**:418–425.

TRAVISANO, M., J. A. MONGOLD, A. F. BENNETT, and R. E. LENSKI. 1995. Experimental tests of the roles of adaptation, chance, and history in evolution. Science **267**:87–90.

VAN TUYLE, G. C., J. P. GUDIKOTE, V. R. HURT, B. B. MILLER, and C. A. MOORE. 1996. Multiple, large deletions in rat mitochondrial DNA: evidence for a major hot spot. Mutat. Res. **349**:95–107.

WHITE, M. E., J. J. BULL, I. J. MOLINEUX, and D. M. HILLIS. 1991. Experimental phylogenies from T7 bacteriophage. Pages 935–943 *in* E. DUDLEY, ed. The unity of evolutionary biology: proceedings of the fourth international congress of systematic and evolutionary biology. Dioscorides Press, Portland, Oregon.

JULIAN P. ADAMS, reviewing editor

Accepted October 1, 1996