26. Two independent clones, RKOAS451 and RKOAS4513, were isolated after transfection with pCMVas45, which expresses *GADD45* complementary DNA in the antisense orientation (*29*), in combination with a pSV2neo selectable marker. Clones were screened for high-level expression of the antisense RNA by ribonuclease protection assay. Gadd45 expression was markedly reduced in these cell lines as determined by immunoblotting. The antisense cell lines and the parent RKO cells were exposed to 254-nm UV irradiation at either 5 or 20 Jm^{-2} and then grown for 7 days, at which time surviving colonies were scored after staining with crystal violet. Survival frequencies were determined by comparison with multiple dilutions of cells of each respective cell line, which were not irradiated but were otherwise identically maintained. Control cell lines consisted of the parent RKO cells, a clonal derivative of RKO, and RKO cells stably transfected with pCMV.3 vector lacking an insert. Growth rates and plating efficiencies of these different RKO lines were equivalent without irradiation. Survival frequencies were similar for each control cell line and were consistently greater than the two antisense *GADD45* cell lines at the higher UV dose, whereas survival was not appreciably different at the lower, less toxic, dose.
27. The average of the pooled absolute survival for all experiments was 0.394% for the parent line and 0.076% (reduced 5.2-fold relative to parent line) for RKOAS451 and 0.080% (reduced 4.9-fold) for RKOAS4513; absolute survival represents colony yield adjusted for plating efficiency, which was >50%. The mean and median values of cell survival relative to the parent line (determined for each experiment and then pooled) for RKOAS451 were 0.26 and 0.32 (with two-tail $P = 0.016$ by exact Wilcoxon rank-sum test for the null hypothesis of equal median relative survivals). The corresponding values for RKOAS4513 were 0.34 and 0.43 ($P = 0.032$). If the data for the two experimental lines were pooled, the values for survival relative to control were 0.30 and 0.35 ($P = 0.0047$).
28. W. S. El-Deiry *et al.*, *Cell* **75**, 817 (1993).
29. Q. Zhan *et al.*, *Mol. Cell. Biol.* **14**, 2361 (1994).
30. Increasing amounts of bacterially expressed Gadd45, ranging from 0 to 3.6 μg, were added to cyclin E–Cdk complexes, which were then tested in in vitro kinase assays with histone H1 as substrate.
31. Gadd45 that had been boiled before addition did not displace PCNA, nor did Mdm2 that was prepared in the same way.
32. N. H. Waseem and D. P. Lane, *J. Cell Sci.* **96**, 121 (1990).
33. K. Yoon, M. A. Thiede, G. A. Rodan, *Gene* **66**, 11 (1988).
34. We thank B. Stillman for recombinant PCNA, K. E. Kilpatrick for her valuable contributions in the development of the mAbs, R. Cunningham for *E. coli Nth* protein, J. N. Weinstein and T. Myers for statistical analysis, A. J. Levine for the histidine-tagged *MDM2* expression plasmid, and J. Robbins for the XP cell lines. Supported in part by grants to M.B.K. from NIH (ES05777) and the Council for Tobacco Research (3187). M.B.K. is the Steven Birnbaum Scholar of the Leukemia Society of America.

26 May 1994; accepted 29 September 1994

Clonal Divergence in *Escherichia coli* as a Result of Recombination, Not Mutation

David S. Guttman* and Daniel E. Dykhuizen

Nucleotide sequence analysis was performed on 12 natural isolates of *Escherichia coli* in four loci located in close proximity on the chromosome. A comparison of gene genealogies indicated that three recombination events have occurred in a subset of the strains (ECOR group A) in the time since their divergence from a common ancestor, while during the same time, no mutational divergence has occurred. The common ancestor of this subset existed no more than 2400 years ago, and recombination was shown to occur at a rate of 5.0×10^{-9} changes per nucleotide per generation—50-fold higher than the mutation rate. Thus, recombination has been the dominant force driving the clonal divergence of the ECOR group A strains and must be considered a significant factor in structuring *E. coli* populations.

Public policy and health decisions concerning the lateral transfer both of recombinant genes from genetically engineered microorganisms and of virulence genes from pathogens often depend on assumptions about recombination. The frequency and significance of recombination in microorganisms, however, is currently the subject of intense debate. In the *E. coli* model system, there is controversy over the extent of clonality (*1*, *2*). Clonality can be thought of as the vertical, asexual transmission of genetic material from parent to offspring, where all evolutionary change through time is strictly the result of the mutational process. In contrast, nonclonal or sexual evolution occurs by the horizontal transfer of genetic material between unrelated individuals through the recombinational process. From the perspective of the population, there is not a strict dichotomy between these two evolutionary modes. A group of strains that share a recent common ancestor and are diverging by means of the mutational process can be considered a clone. Over time, genetic material from individuals that do not share the same most recent common ancestor will be introduced into the individual members of a clone by recombination. As this occurs, the clonal background of these strains will become progressively more obscured by foreign DNA, and the clonal nature of this group will begin to deteriorate (*3*, *4*). The prevailing view of *E. coli* is that it is a primarily clonal organism. Recombination is believed to occur at such a low frequency, relative to mutation, that it is considered evolutionarily insignificant (*2*).

A misconception about the effect of recombination on the evolutionary divergence of microorganisms has further confused the issue of clonality. Recombination has often been considered incapable of breaking down clonal structure because its effect on variation is usually considered in the context of a panmictic species. Recombination is therefore considered a homogenizing force. But, recombination can also be diversifying. In a structured population, recombination within a clone will result in genetic homogenization, but recombination between individuals of different clones will cause genetic diversification within the clone. This study is concerned with the diversification of a clone when there is transfer between clones. Under these conditions, recombination can be thought of as a mechanism, complementing mutation, that generates variability within clones and drives clonal divergence.

The relative contribution of recombination and mutation to clonal divergence has been difficult to assess because of problems associated with obtaining an accurate estimate of the rate of recombination. Uncertainty exists because of the obscuring of older exchange events by more recent events and because the methods used to estimate recombination rely heavily on assumptions based on a particular population structure, even though the population structure of microorganisms is poorly understood (*5*, *6*). Milkman and Stoltzfus (*3*) and Milkman and Bridges (*4*) made the first estimates of the recombination rate in *E. coli* from nucleotide sequence data. Their calculations used estimates of the average clonal segment length (the length of a segment of DNA that recently shared ancestry with the equivalent segment from another individual) and the number of generations since their most recent common ancestor. They estimated the rate of recombination in *E. coli* to be 5×10^{-12} transfers per nucleotide per generation (*4*). This approach determines the probability that two adjacent nucleotides originated in different lineages and were brought together by a recombination event (a parameter similar to that used in transmission genetics). Thus, the recombination rate is expressed in different units than the mutation rate. If the recombination rate is to be comparable to the mutation rate, then both rates must be expressed in terms of the rate of nucleotide change per generation (with the understanding that a single recombination event will change multiple nucleotides in batch).

Department of Ecology and Evolution, Division of Biological Sciences, State University of New York, Stony Brook, NY 11794, USA.

*To whom correspondence should be addressed.

We present a method, built on the conceptual foundation of Wilson and co-workers (*7*) and Dykhuizen and Green (*1*) for identifying recombination events from nucleotide sequence data, that facilitates the assessment of the relative influence of recombination and mutation on clonal divergence. Our estimation of the rate of recombination avoids the problems associated with the obscuring of past events by those occurring more recently and is independent of population dynamics and structure.

Twelve strains were chosen for this study from the ECOR standard reference collection of natural *E. coli* isolates (*8*). Five of these strains (ECOR4, ECOR8, ECOR10, ECOR16, and K12) have been classified through multilocus enzyme electrophoresis as a well-defined and phylogenetically consistent cluster known as group A strains (*9, 10*). Their low amount of divergence reduces the likelihood of multiple mutations at the same nucleotide position and of recombination events that overlap and mask previous events.

Approximately 1 kb from each of four loci, *sppA*, *gapA*, *pabB*, and *zwf*, were sequenced in the 12 *E. coli* strains. Sequences of *gapA* and *pabB* have been reported by Guttman and Dykhuizen (*11*). Both *zwf* (coding for glucose-6-phosphate dehydrogenase, E.C. 1.1.1.49) and *sppA* (coding for protease IV) were sequenced in the same manner. The four loci, *sppA*, *gapA*, *pabB*, and *zwf*, are located at minutes 38.5, 39.3, 39.9, and 40.8, respectively, on the *E. coli* chromosome (*12*). One minute is equivalent to approximately 45 kb.

Gene genealogies were constructed for each locus by both neighbor-joining (*13, 14*) and parsimony methods (*15*). Midpoint-rooted neighbor-joining trees based on Jukes-Cantor distances (*13*) for the four loci are presented in Fig. 1. The parsimony analysis gave identical results with high consistency indices (0.826, 1.00, 0.927, and 0.821 for the *sppA*, *gapA*, *pabB*, and *zwf* genes, respectively). A bootstrap analysis, with 1000 replicates, was performed on each of the genealogies to determine the statistical stability of each node (*16*).

We identified specific recombination events on the basis of the phylogenetic congruency test of Wilson and co-workers (*7*), which identifies gene transfer events by comparing gene genealogies from a number of regions. Congruent genealogies indicate that the regions have had similar histories. Differences in genealogies can be explained most parsimoniously by gene transfer (*1, 17, 18*).

The gene genealogies of the four loci show considerable topological uniformity. Three clusters of taxa are found consistently in all loci. The ECOR38, 39, and 40 cluster and ECOR49 and 50 cluster from group D have identical or nearly identical nucleotide sequences at all loci. The group A strains of ECOR4, 8, 10, and 16 and K12 were found to be either isosequential (having identical sequences), or quite different. Only three of these strains show genealogical histories at points inconsistent with the isosequential clustering (Fig. 2). The sequence of ECOR16 at the *sppA* gene is different from the rest of the group A strains, as is the sequence of ECOR10 at the *zwf* locus. The sequence of ECOR4 is identical to the other group A strains at *sppA*, but different at the other three loci.

The divergence of these three group A strains would be difficult to account for by a stochastic mutational process, especially given that all other group A strains consistently share the identical nucleotide sequence at these sites. On the other hand, this pattern would be expected if recombination was responsible for driving the divergence of these strains. Genetic exchange with a distantly related donor strain would introduce a large number of phylogenetically inconsistent polymorphisms into the recipient. This would result in apparent massive divergence of the strain from those with which it shared a recent common ancestor. This pattern would only be found in the region introduced by recombination; the rest of the genome would continue to display phylogenetic clustering consistent with clonal descent.

Our analysis indicates that recombination has introduced a foreign DNA fragment into ECOR10 with a recombination breakpoint between genes *zwf* and *pabB*. Another event has introduced a recombinant fragment into ECOR16 with a breakpoint between genes *sppA* and *gapA*. Because multilocus enzyme electrophoresis identified ECOR4 as a group A strain, it should be isosequential with the rest of the group A strains under the null hypothesis of no recombination. Our observations reveal that a recombination event has introduced foreign DNA into ECOR4 from a donor similar to ECOR68 (Fig. 1). This recombinant fragment stretches through the *zwf*, *pabB*, and *gapA* genes and ends somewhere between *gapA* and *sppA*. This assertion can be tested by sequencing upstream of *sppA*. If ECOR4 remains tightly linked to the rest of the group A cluster, the polarity of this recombination event will be verified.

The phylogenetic congruency test shows that there have been three recombination events in the *sppA-gapA-pabB-zwf* region of five group A strains (K12, ECOR4, 8, 10, and 16) since their divergence from a common ancestor. The identical nucleotide sequences of the nonrecombinant group A strains indicate that during this same period of time these strains have undergone no mutational divergence. Regardless of the

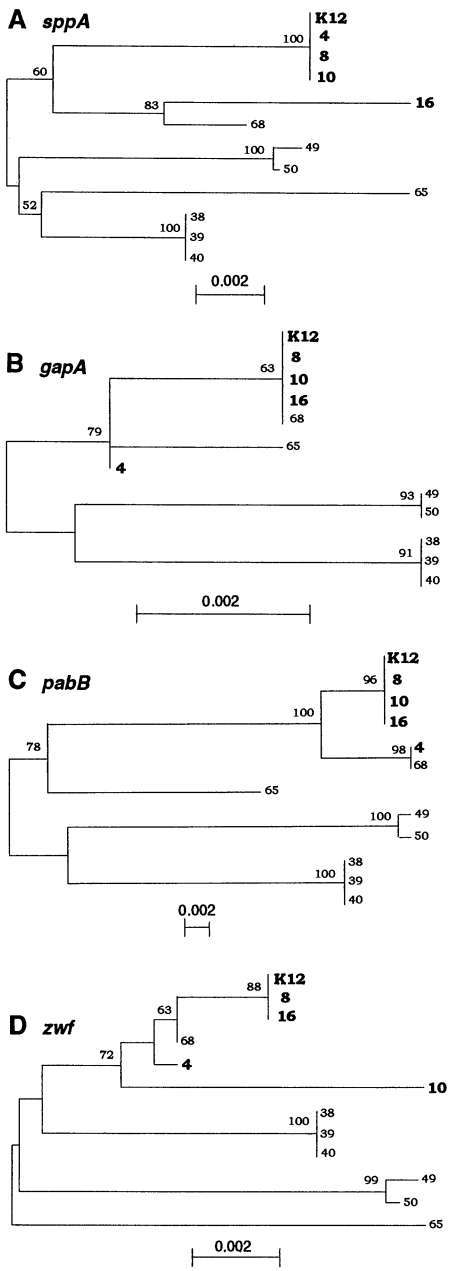

Fig. 1. Midpoint-rooted neighbor-joining trees, calculated on the basis of Jukes-Cantor distances, for the (**A**) *sppA*, (**B**) *gapA*, (**C**) *pabB*, and (**D**) *zwf* loci. Both strands of an approximately 1-kb region were sequenced for each gene in the manner described in Guttman and Dykhuizen (*11*). Construction and bootstrapping of the trees were carried out with NJBoot2 (*14*). The numbers at the base of each node are bootstrap scores (representing the statistical reliability of nodes). The numbers at the end of each line are the strains. Only those scores greater than 50% are presented. One thousand bootstrap replicates were performed for each analysis. Those strains that abut a common vertical line (meaning they have zero branch lengths, such as strains K12, ECOR4, 8, and 10 in the *sppA* locus) have identical sequences. The group A strains (K12, ECOR4, 8, 10, and 16) are denoted in a larger, bold typeface. A bar representing a Jukes-Cantor distance of 0.002 is presented below each tree for scale.

absolute values of population genetic parameters such as the rate of recombination, the effective population size, the mutation rate, or the population structure of *E. coli*, recombination has been the dominant driving force behind the divergence of these strains.

A rate of recombination can also be calculated from these data (*19*). For this calculation, the time since the most recent common ancestor of group A must first be determined. This can be done through the use of the number of segregating sites (S) in the calculation of θ, the per nucleotide heterozygosity. The expectation (E) of S is

$$E(S) = \theta \sum_{i-1}^{n-1} 1/i$$

where n is the number of strains, $\theta \equiv 2N_e\mu$ (*20*), and μ is the per generation neutral mutation rate. Because N_e (the effective population size) is scaled in units of generations, we used θ to determine the time since most recent common ancestry. We used a mutation rate of 10^{-10} changes per nucleotide per generation (*21*) and a value of 300 generations per year for *E. coli*, to calculate that the expected time since most recent common ancestry for the group A strains is less than 2400 years.

The rate of recombination can be scaled in a number of different ways. One measure would be based on the number of events per genome per unit time. We have sampled approximately 2.25 min (2.25%) of the genome and found three recombination events. Thus, the rate could be expressed as three recombinations over five strains divided by the proportion of the genome surveyed to estimate the transfers per genome, which is about 27 transfers per genome since the most recent common ancestor (*22*). Dividing by the time since the most recent common ancestor (~2400 years) translates to a maximum rate of ~3.78×10^{-5} transfers per genome per generation. If this is converted to transfers per nucleotide per generation, it becomes 3.78×10^{-5} transfers per genome per generation divided by 4.5×10^{6} nucleotides per genome, which is 8.4×10^{-12} transfers per nucleotide per generation. This is in good agreement with the rate of 5×10^{-12} transfers per nucleotide per generation estimated by Milkman and Bridges (*4*).

A more useful number for an estimation of the rate of recombination would be an expression based on the amount of change in the nucleotide sequence. The units should be the same as estimates of the mutation rate to allow direct comparisons of the relative impact of mutation and recombination. Assumptions must first be made concerning the length of the average inserted recombination fragment and the degree of sequence divergence between two randomly selected strains of *E. coli*. We assume the length of the average recombinant fragment to be approximately 1 min (roughly 45 kb of DNA) and the average pairwise difference between two randomly chosen alleles of the same gene in *E. coli* to be about 1.32%. This results in a rate of recombination equal to 5.0×10^{-9} changes per nucleotide per generation or about 50-fold higher than the mutation rate (*23*). Thus, recombination in this situation was 50 times more efficient than mutation in driving strain divergence. Although recombination has repeatedly been discarded as a significant mechanism for driving the divergence of *E. coli* clones (*2*, *3*, *24*), it alone is responsible for the divergence of the group A strains. The mutational process, in contrast, has had no impact on this divergence.

The perspective of this paper has been restricted to examining the effect of recombination on clonal divergence. This has allowed us to estimate the relative importance of recombination and mutation with respect to the population structure of *E. coli*. Although this analysis focused on clonal groups, the estimated relative effect of recombination and mutation should apply to the species as a whole. Unlike mutation, which will always increase diversity, recombination will homogenize the species gene pool, even as it creates different genotypes and diversifies clones. This homogenizing effect of recombination must be strong enough to prevent unlimited divergence of the gene pool if the integrity of the species is to be maintained. Because recombination between species is rare (*1*), it will not be capable of preventing unbounded divergence between species. Thus, an understanding of the population genetics of bacterial species will require an understanding of the interaction of mutation and recombination. Mutation will be the more important factor influencing divergence between species, but for *E. coli* our data indicate that recombination will dominate within species.

The mechanisms of recombination in *E. coli* can be investigatedby the methodology used in this paper. The precise location of recombination breakpoints can be found by sequencing the regions between *sppA* and *gapA* in ECOR16 and ECOR4 and between *pabB* and *zwf* in ECOR10. The structure of these breakpoints will help determine the natural mechanism for genetic exchange. The size of the piece inserted can be obtained by sequencing regions distal to the set of four genes presented in this study. If transduction is common, the size of the recombinant piece should be on the order of 30 to 100 kb. If conjugation is common, larger pieces would be expected.

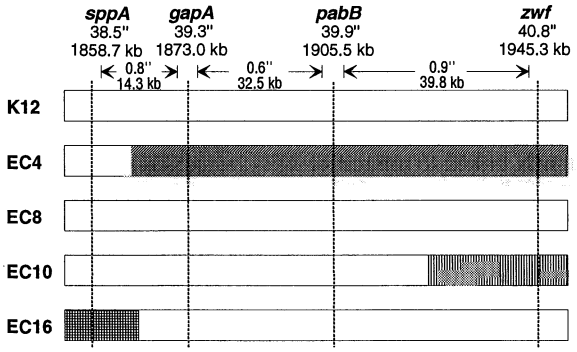

Fig. 2. A schematic representation of the *E. coli* chromosomal region between the *sppA* locus and the *zwf* locus for the five group A strains. The locations of the loci and the approximate distances from their neighboring loci are presented in both minutes (1 min of the *E. coli* chromosome is approximately equal to 45 kb) and genomic address in kilobase pairs (*12*). A clear bar represents the common clonal background. A darkened bar represents regions introduced into the respective strains by recombination. The locations of the breakpoints are approximate.

REFERENCES AND NOTES

1. D. E. Dykhuizen and L. Green, *J. Bacteriol.* **173**, 7257 (1991).
2. J. Maynard Smith, N. H. Smith, M. O'Rourke, B. G. Spratt, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4384 (1993).
3. R. Milkman and A. Stoltzfus, *Genetics* **120**, 359 (1988).
4. R. Milkman and M. M. Bridges, *ibid.* **126**, 505 (1990).
5. R. R. Hudson and N. L. Kaplin, *ibid.* **111**, 147 (1985).
6. R. R. Hudson, *Genet. Res. (Cambridge)* **50**, 245 (1987).
7. A. C. Wilson, S. S. Carlson, T. J. White, *Annu. Rev. Genet.* **46**, 573 (1977).
8. H. Ochman and R. K. Selander, *J. Bacteriol.* **157**, 690 (1984). The ECOR strains used were ECOR4, 8, 10, 16, 38, 39, 40, 49, 50, 65, and 68 in addition to the standard laboratory reference strain K12, which is electrophoretically identical to ECOR2.
9. R. K. Selander, D. A. Caugant, T. S. Whittam, in Escherichia coli *and* Salmonella typhimurium*: Cellular and Molecular Biology*, F. C. Neidhardt *et al.*, Eds. (American Society for Microbiology, Washington, DC, 1987), vol. 2, pp. 1625–1648.
10. P. J. Herzer, S. Inouye, M. Inouye, T. S. Whittam, *J. Bacteriol.* **172**, 6175 (1990).

11. D. S. Guttman and D. E. Dykhuizen, *Genetics*, in press.
12. K. E. Rudd, in *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for* Escherichia coli *and Related Bacteria*, J. H. Miller, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992), pp. 2.3–2.43.
13. N. Saitou and M. Nei, *Mol. Biol. Evol.* **4**, 406 (1987).
14. K. Tamura, Pennsylvania State University, University Park, PA (1993).
15. D. L. Swofford, *Illinois Natural History Survey* (Champaign, IL, 1993).
16. J. Felsenstein, *Evolution* **39**, 783 (1985); C. R. Woese, J. Gibson, G. E. Fox, *Nature* **283**, 212 (1980).
17. J. G. Lawrence and D. L. Hartl, *Genetics* **131**, 753 (1992).
18. A. M. Valdes and D. Piñero, *Evolution* **46**, 641 (1992).
19. The true neutral mutation rate is difficult to measure, but because it is linearly related to the time to common ancestry, discrepancy of a few fold will not change the overall conclusion. We converted the mutation rate used in the calculation into a per nucleotide per year estimate using a 300 generation per year conversion factor and then into a per gene per year estimate by dividing by the approximate length of the sequenced regions (1 kb per gene) to give a rate of $\mu = 3 \times 10^{-5}$ changes per gene per year. The summation of $1/i$, which takes into account the length of the genealogy, was determined for each gene separately because one of the five group A strains sequenced at the *sppA*, *gapA*, and *pabB* genes is a recombinant and two of the five strains are recombinants at *zwf*. Because we observed that all nonrecombinant group A strains were isosequential (have no segregating sites), the estimated number of segregating sites was set at a maximum of one to permit the computation.
20. R. R. Hudson, in *Oxford Surveys in Evolutionary Biology*, D. Futuyma and J. Antonovics, Eds. (Oxford Univ. Press, Oxford, 1990), vol. 7, pp. 1–44.
21. J. W. Drake, *The Molecular Basis of Mutation* (Holden-Day, San Francisco, 1970).
22. Strains 38, 39, and 40 (from ECOR group D) also form a recently diverged group. However, because they have identical sequences for all four genes, we could not use this cluster to distinguish the effects of mutation from recombination. Strains 49 and 50 (also from ECOR group D) additionally form a consistent cluster. They were selected for this study because they share insertion sequence bands (D. E. Dykhuizen, unpublished data), suggesting a recent common ancestor [normally, only electrophoretically identical strains share insertion sequence bands (*25*), although these strains are not electrophoretically identical]. They show no recombinational differences but do have a single segregating site in each of the *sppA*, *pabB*, and *zwf* genes. If the analysis is carried out with the data from these three clusters (the group A and group D strains) as if the clusters were part of the same clonal group (with 10 strains, three segregating sites, and three recombination events), the time to common ancestry doubles to ~4650 years, and the rate of recombination drops to 1.4×10^{-9} changes per base pair per generation, 14 times greater than the mutation rate. This minor change in the recombination rate does not change the overall conclusion that recombination has been more effective than mutation in driving clonal divergence.
23. There are no good estimates of the size of recombination fragments in natural populations of *E. coli*, but if the major means of transfer is transduction-mediated by a phage similar in size to P1, the average size of transferred fragments would be approximately 1 min or roughly 45 kb of DNA (*26*). This assumption agrees with the finding of Whittam and co-workers (*27*), who observed no correlation between map distance of loci in *E. coli* and linkage disequilibrium for electrophoretic markers. This suggests that most of the fragments are not much more than 1 or 2 min of the chromosome. Milkman and Bridges (*4*, *28*) have suggested that the length of recombination fragments could be much shorter. But, even if a length as short as 1 kb is used in the calculation, the overall picture does not change because the rate of recombination remains greater than the rate of mutation. The average pairwise difference was calculated from all nongroup A strains to the nonrecombinant group A cluster for the *sppA* and *zwf* loci. The average pairwise nucleotide difference between group A and nongroup A strains was 1.53% and 1.12% for *sppA* and *zwf*, respectively (mean = 1.32%). This results in ~594 base pairs changed during every lateral transfer event. The rate of change per nucleotide per generation will then be $(3.78 \times 10^{-5}$ transfers per genome per generation) $\times$ (594 base pairs changes per transfer) divided by $(4.5 \times 10^{6}$ nucleotides per genome), which is equal to 5.0×10^{-9} changes per nucleotide per generation.
24. T. S. Whittam and S. E. Ake, in *Mechanisms of Molecular Evolution*, N. Takahata and A. G. Clark, Eds. (Sinauer, Sunderland, MA, 1993), pp. 223–246.
25. J. G. Lawrence, H. Ochman, D. L. Hartl, *Genetics* **131**, 9 (1992).
26. P. Margolin, in Escherichia coli *and* Salmonella typhimurium*: Cellular and Molecular Biology*, F. C. Neidhardt *et al.*, Eds. (American Society for Microbiology, Washington, DC, 1987), vol. 2, pp. 1154–1168.
27. T. S. Whittam, H. Ochman, R. K. Selander, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1751 (1983).
28. R. Milkman and M. M. Bridges, *Genetics* **133**, 455 (1993).
29. We thank L. Chao, W. Eanes, D. Futuyma, A. Bely, M. McCartney, I.-N. Wang, and two anonymous reviewers for their reading and suggestions to improve the manuscript, and J. Kim and T. Morton for discussions. All nucleotide sequences are available through GenBank–European Molecular Biology Laboratory. Accession numbers for *sppA* are U13772 through U13780, U13782, U13833, and U13834. Accession numbers for *zwf* are U13783 through U13794. Supported in part by PHS grants GM30201 and AI32454 from the NIH. This is contribution number 899 from the Department of Ecology and Evolution at the State University of New York at Stony Brook.

11 May 1994; accepted 1 September 1994

Catalytic Site Components Common to Both Splicing Steps of a Group II Intron

Guillaume Chanfreau and Alain Jacquier*

The splicing of group II introns occurs in two steps involving substrates with different chemical configurations. The question of whether these two steps are catalyzed by a single or two separate active sites is a matter of debate. Here, certain bases and phosphate oxygen atoms at conserved positions in domain V of a group II self-splicing intron are shown to be required for catalysis of both splicing steps. These results show that the active sites catalyzing the two steps must, at least, share common components, ruling out the existence of two completely distinct active sites in group II introns.

Group II introns are found in the genomes of organelles of lower eukaryotes and plants and in bacteria (*1*). Some group II introns are able to catalyze their own excision in vitro in the presence of magnesium (*2*). Their excision follows a pathway similar to that of nuclear precursor mRNA (pre-mRNA) splicing, the intron being released in a branched lariat form (*2*). In this pathway, the substrates for the two steps are different: The first transesterification results from the attack at the 5′ splice site by a 2′ hydroxyl, and the second transesterification is initiated by a 3′ hydroxyl. The sequences at the 5′ and 3′ junctions are also different (*1*). Because in nuclear pre-mRNA introns the same phosphorothioate diastereomer (Rp) inhibits both steps when introduced at the junctions (*3*), the two transesterifications cannot be considered as the forward and reverse of the same reaction, as in group I introns (*4*). Whether distinct active sites catalyze the two chemical steps (*3*) or whether a structural rearrangement between the two steps accommodates the different substrates in a single active site (*5*) is a matter of conjecture. To address this question, we have searched for active site components involved in each step of group II self-splicing.

Group II introns are folded into six structural domains (*1*). Among these, domain V is the most conserved element in primary sequence, and deletion experiments have shown that domain V is essential for splicing (*6*, *7*). Moreover, it is able to catalyze 5′ junction hydrolysis when added in trans (*8*). For these reasons, domain V is a good candidate for carrying active site elements. We have adapted the modification interference technique (*9*) to investigate the role of structural elements of domain V in the catalysis of each splicing step (Fig. 1). To investigate the effects of modifications on each of the two splicing steps, we used two assays (Fig. 1A). (i) In a cis-splicing assay the CX fragment is annealed to a subunit composed of exon 1 and domains I to III (E1-XC fragment); under these conditions lariat formation is rate-limiting, allowing the study of the effects of modifications on

Unité de Génétique Moléculaire des Levures, URA 1149 du CNRS, Département de Biologie Moléculaire, Institut Pasteur, 25, rue du Dr. Roux, F-75724 Paris Cedex 15, France.

*To whom correspondence should be addressed.