In this issue Belfort and coworkers (1) report the isolation and characterization of a thermostable endonuclease from the archaeabacterium Desulfurococcus mobilis. Remarkably, the enzyme is encoded in an intron in the single gene for 23S rRNA and is expressed only after the corresponding RNA segment is excised from the newly made rRNA. Archaeabacterial introns are unique in that their excision is catalyzed by an enzyme (2). The endonuclease has a number of features characteristic of the homing endonucleases found in fungi and their organisms, and these have led the authors to wonder whether trans-kingdome gene transfer may have been involved. As the authors suggest, all of these topics—introns, archaeabacteria, homing endonucleases, and horizontal transfers—are lightning rods for debate.

Homing endonucleases are a class of enzymes that can cut double-stranded DNA with such great specificity that there may be only a single target sequence in an entire genome. The genes that encode them are widely spread, but they are especially prevalent in the group I introns of fungal mitochondria (3). Group I introns themselves are found in many different organisms, although so far they have not been found among the metazoan or archaeabacteria (3). Not all group I introns contain such genes, of course, and it is the sporadic occurrence of the homing endonucleases that is so intriguing. The endonuclease makes it possible for the intron encompassing the gene to be moved into intronless strains in a unidirectional gene-conversion event.

There appear to be several families of these endonucleases, the most common being recognizable at the level of primary structure by the occurrence of characteristic signature sequences. The best known of these is typified by a conserved decapetide sequence denoted by the standard single-letter amino acid sequence LAGLI-DADG (4). The introns themselves commonly occur in the genes for cytochrome oxidase and NAD dehydrogenase subunits, as well as in the genes for rRNA. Although they are found in all kinds of fungi, many strains or even entire species lack either the intron or the open reading frame (ORF) encoding the endonuclease. The possibility of horizontal gene transfers between strains or between species has been raised in the past (3), partly as an explanation for the apparently idiosyncratic distribution of both the introns and the endonucleases themselves.

Meanwhile, Marlene Belfort and colleagues had found that the bacteriophage T4 also contained group I introns in several of its genes (5, 6). Because these were the first introns identified in a prokaryotic system, there was immediate speculation that some kind of horizontal transfer from eukaryotes might be involved (7). These introns contained ORFs that differed from one intron to another, and it was realized that two different kinds of mobility must be involved, the intron on the one hand and the agents encoded within them on the other. As in the case of the ORFs from organellar introns, the latter were found to encode endonucleases. Although the strategy appeared to be the same in the phage and eukaryotes, there were a number of differences in detail, enough to suggest that the endonucleases evolved independently (8). For example, the phage endonucleases were less specific in their targeting, and they also cut the DNA at quite a distance from the actual insertion site. Thus, even though the group I intron structures themselves were similar in the two settings, the invaders appeared to be of different stock. Similar introns and ORFs were subsequently found in other T-even phages, but the overall distribution was sporadic and recked of promiscuous mobility (9).

The discovery by Shub and coworkers (10) that the bacteriophage SPO1, which infects Gram-positive bacteria, also contains a group I intron with an unrecognizable ORF implied that the introns themselves were in these viruses before the divergence of the major groups of bacteria. It is fair to say, however, that the existence of the introns in different genes in the two bacterial groups could just as well have been attributed to a more recent invasion. The SPO1 intron occurs in its DNA polymerase gene, whereas the T4 introns and those of other T-even phages occur in genes associated with the biosynthesis of small-molecule components of DNA.

There were other good reasons to suppose that the group I introns and the endonuclease genes that often reside within them had independent origins, even though as a functional pair their mobilities ought to be greater (29). The endonuclease finds genomic targets into which not only its gene but also the intron around it can be inserted. The self-splicing intron provides the wherewithal for preventing interference with the host genome. In this regard, it should be noted that some of the ORFs actually encode a maturase that assists in the splicing event. Remarkably, the maturases can be evolved from the endonucleases, as is evidenced by two proteins that are 60% identical encoded by ORFs in yeast mitochondria, one of which has been shown by experiment to be a maturase and the other an endonuclease (11, 12).

Recent developments have made the situation even more intriguing. Several cases of protein splicing have been found to involve endonuclease action, and the endonucleases are all of the sort involving the LAGLI-DADG motif. One of these involves a nuclear gene in yeast that encodes the vacuolar H+-ATPase (13, 14). Another involves the DNA polymerase (is this a coincidence?) of an archaeabacterium, Thermococcus litoralis (15), and a third has been found in the RecA protein of a eubacterium, Mycobacterium tuberculosis (16). In each of these cases an extended protein is made and an internal segment removed. In all three the internal segment corresponds to a protein of the LAGLI-DADG type, and in the first two the protein has been shown to encode an endonuclease (17).

The finding of this same kind of endonuclease in D. mobilis was not altogether unexpected. Dalgard and Garrett (18) had already identified ORFs in archaeabacterial introns in both D. mobilis and Pyrobaculum organotrophicum and noted the LAGLI-DADG motif. As in the case of other endonuclease–intronic affiliations, the tie to the host is not obligatory. There are closely related archaeabacterial species that lack both the endonuclease and the intron. Again, the patchy occurrence raises the possibility of broken lines of descent and unconventional gene transfers.

The full characterization of the D. mobilis endonuclease confirms that it is much more similar to the eukaryotic kind
and distinct from that found in phages (1). Given that *D. mobili* is a hyperthermophili, the resistance of the endonuclease to heat is not unexpected, but it does present a problem with regard to its evolutionary origins, since presumably any ordinary protein imported from a distantly related mesophile would not, as the authors point out, endure at elevated temperatures. Another complication—and volatile discussion point—has to do with the nomenclature of the major biological groups. Until the mid-1970s, the conventional wisdom was to divide the cellular world into the prokaryotes and eukaryotes. Studies on the sequences of rRNA revealed that a large number of bacterial species formed a separate grouping that was as different from the other bacteria as it was from the eukaryotes (19). Woese called this group, which was composed of thermophiles, hyperthermophiles, and halophiles, the archaebacteria (archae = ancient). Subsequent sequence studies on 55 RNA (20) and many proteins (21–23) have shown, however, that the archaebacteria are actually less “ancient” (read “early-diverging”) than the eubacteria. Even so, most workers in the field were content to live with a three-domained world made up of the eubacteria, the archaebacteria, and the eukaryotes (24). It was a surprise to many, then, when Woese and colleagues (25) suggested that these three groups be renamed the bacteria, the archaebacteria, and the eucarya, respectively, since the misleading “archaeon” lingered on.

That this nomenclature bears heavily on the current problem is underscored by the fact that Dalgaard et al. (1) raised the possibility of horizontal transfer to explain the fact that the archaebacterial endonuclease is so much more like the eukaryotic kinds than like the prokaryotic sorts found in the T-even bacteriophages. If, as is believed by most, the archaebacteria are actually later-diverging than the eubacteria, then there is no need to appeal to special mechanisms, whether or not eubacteria have endonucleases of the LAGLIDAG type. As it happens, however, eubacteria do have such endonucleases, as was first pointed out by Shub and Goodrich-Blair (17) in their mini-review of protein splicing in the bacterial RecA protein. Moreover, endonucleases of the sort found in the T-even phages have also been found in organellar introns (26).

How can these matters be sorted out? There are two aspects to the problem, one functional, the other structural. First, as is noted by Dalgaard et al. (1), the distribution and occurrence of the various kinds of endonuclease must be carefully surveyed. Second, we must depend on structural studies to reconstruct the evolutionary comings and goings of the genes for these enzymes. The forbidding fact is that the endonuclease sequences are highly divergent, to the point where reliable alignments and phylogenetic reconstructions are obtained only with difficulty. Three-dimensional struc-

![FIG. 1. Signature sequences for homing endonucleases of the LAGLIDAG type.](image)

![FIG. 2. Phylogenetic tree constructed from the alignment shown in Fig. 1 by the PAPA program (28).](image)
ture comparisons may ultimately offer a solution to this problem.

In the meantime, however, we ought to examine existing sequences to see whether sense can be made of matters. As a first step in this direction, then, I selected a set of well-known members of this family and cut out their characteristic branches of LAGLI-DAG sequences. I also searched the Protein Identification Resource database for the LAGLI-DAG motif and was able to add a few more interesting candidates, including a transposase from the eubacterium Lactococcus lactis, and another, directly from the recent literature (27), for an intron-embedded ORF from the archaeobacterium Pyrobaculum aerophilum. I then constructed a phylogeny based upon this abbreviated alignment (Figs. 1 and 2).

Several things emerged from this simple and preliminary inquiry. First, not only does the recently reported ORF in an intron in the single gene for 16S rRNA in P. aerophilum (27) contain the major endonuclease motif, but it is most closely related to one of those that occurs in the intron in the gene for 23S rRNA in Pyrobaculum organotrophum. Interestingly, the sequence from the L. lactis transposase falls into the same cluster. The sequence from the intron in D. mobilis clusters with the endonuclease from the protein intron in T. litoralis and the second ORF from the intron in the 23S rRNA gene of P. organotrophum. Thus, all the archaeobacterial sequences clung together in two groups. Beyond that, the groupings that emerge from the simple character-based tree-growing program are remarkably consistent with the biological distribution in that all the endonucleases from fungal mitochondrial introns cluster together, as do most of the endonuclease sequences from the protein introns (Fig. 2). The arrangement is consistent with a degree of mobility and exchange for both the host introns and the invading endonucleases. But it does not require horizontal transfers between eukaryotes and archaeabacteria. Some readers will doubtless be critical of an attempt to build a phylogeny on a simple alignment involving only 28 amino acid locations; certainly this is a preliminary effort. The point remains that these matters can be analyzed by sequence analysis. The major kinds of endonucleases are clearly present in both eubacteria and eukaryotes. The LAGLI-DAG is obviously present in eubacteria, archaeabacteria, and eukaryotic organelles and nuclear gene products. The question posed by the authors (1) as to whether or not long-range horizontal gene transfers were involved in the events leading to the present-day distribution will be resolved by further analysis and more data.