MINIREVIEW

Group I Introns and Inteins: Disparate Origins but Convergent Parasitic Strategies[▽]

Rahul Raghavan† and Michael F. Minnick*

Division of Biological Sciences, The University of Montana, Missoula, Montana 59812

Genomes of most organisms harbor DNA of foreign origin that has no known function. Since these elements may not contribute to a host's fitness but utilize host resources for their perpetuation, it is appropriate to consider them genetic parasites (4). With the advent of sequencing technologies, a wide variety of parasitic elements have been discovered in bacteria from all environments, including obligate intracellular pathogens (100), which were thought to be shielded from horizontal gene transfer (HGT). Detection of a parasitic genetic element in a genome represents only a snapshot of the continuing and dynamic interplay between the host's attempts to purge the element and the element's ability to persist. These adaptable genetic parasites have evolved mechanisms to overcome defenses (75) of the cellular machinery to ultimately invade, colonize, and replicate within the host. Their success is very evident in the human genome, which consists mostly of such apparently superfluous DNA (65). Even compact bacterial genomes packed with functional genes contain mobile genetic elements (100), underscoring their universality in nature.

A number of parasitic genetic elements are found in bacterial genomes, including transposons, insertion sequences, prophages, introns, inteins, and intervening sequences. While bacteria, especially pathogenic bacteria, are well studied, their parasitic genetic elements have not received as much attention. In the past few years, while studying the obligate intracellular pathogen *Coxiella burnetii*, we came to appreciate the intimate relationship between bacterial hosts and parasitic elements (92, 93). In addition, interesting new studies have shed light on the evolutionary histories of group I introns, inteins, and homing endonucleases (HEs) (9, 109) and infused excitement into the field. This minireview, which focuses on the biology and evolution of group I introns and inteins found in bacteria, is an attempt to catalyze interest among bacteriologists in these fascinating genetic parasites.

GROUP I INTRONS

Introns are noncoding, intragenic regions that are removed from precursor RNA to form the mature RNA by splicing the

exons (coding regions that flank introns) together. They are much more common in eukaryotes than in bacteria (50). Introns are classified into four groups based on splicing mechanisms (47): group I, group II/group III, spliceosomal, and tRNA/archaeal introns. Spliceosomal introns are found in eukaryotes and utilize spliceosomes (large protein-RNA complexes) for splicing (70), whereas tRNA introns splice with the help of specialized enzymes (74). Group I and group II introns are able to self-splice—using different mechanisms—without the aid of any proteins and are thus referred to as ribozymes (104). A self-splicing group I intron from *Tetrahymena thermophila* was one of the first ribozymes to be described, in the early 1980s (61). Ribozymes are considered legacies of a primordial RNA world, where RNA possessed both information-encoding and catalytic properties, before the advent of DNA and protein-based life forms (35).

Group I introns are small RNAs (~250 to 500 nucleotides) that have invaded protein-, rRNA-, and tRNA-encoding genes in a variety of organisms, including algae, fungi, lichens, some lower eukaryotes, and a few bacteria (47). While the first bacterial group I intron was not discovered until 1990 (63, 116), the recent availability of inexpensive and accurate whole-genome sequencing technologies has made it possible to identify these mobile elements in a number of bacterial species from diverse ecosystems. All bacterial group I introns analyzed to date have been shown to self-splice. An exception is an intron of *Simkania negevensis*, which reportedly remains unspliced in the mature 23S rRNA (33). Also, akin to the scenario in eukaryotes, *C. burnetii* and some *Synechococcus* strains contain multiple introns interrupting the same gene (45, 93).

All group I introns share a conserved secondary structure (Fig. 1A), which consists of paired elements (P) that assist in self-splicing by using a guanosine (or GMP or GTP) as a cofactor (110). P4-P5-P6 and P3-P7-P9 form two separately folding helices within the core. Helix P3-P7-P9 contains the binding site for the guanosine (G-binding site [GBS]) and is the minimal catalytic domain required for splicing (51). P1 and P10 are complementary to 5' and 3' exons, respectively, and are collectively termed the internal guide sequence (IGS) (113). Based on secondary structure, group I introns are classified further into 13 subgroups (78, 105).

In the first step of splicing (Fig. 1B), the 3'-OH group of an exogenous guanosine bound to a GBS carries out a nucleophilic attack on the 5' splice site, which is marked by a conserved G·U wobble pair within P1. After the first step, this guanosine is covalently bound to the free 5' end of the intron

* Corresponding author. Mailing address: Division of Biological Sciences, The University of Montana, Missoula, Montana 59812. Phone: (406) 243-5972. Fax: (406) 243-4184. E-mail: mike.minnick@mso.umt.edu.

† Present address: Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721.

[▽] Published ahead of print on 7 August 2009.

FIG. 1. (A) Predicted secondary structure of a group I intron (Cbu.L1951) (93). Paired, conserved helices common to group I introns are designated P1 to P10. The 5'- and 3'-terminal intron bases are encircled. The intron sequence is in uppercase; 5' and 3' exons are in lowercase and colored red and blue, respectively. P1 and P10 together form the IGS. The site of HE insertion in P8 is indicated in green. (B) Mechanism of group I intron splicing (110). 5' and 3' exons are in red and blue, respectively. Ω G, terminal intron guanine. G*, exogenous guanosine. (Step 1) Nucleophilic attack on the 5' splice site by the 3'-OH of G* in GBS. (Step 2) Nucleophilic attack on the 3' splice site by the free 3'-OH of the 5' exon. (Step 3) Free intron and spliced exons.

and leaves the GBS, allowing the conserved terminal guanine (Ω G) to occupy the GBS and mark the 3' splice site. In the second splicing step, the 3' OH group of the free 5' exon attacks the 3' splice site, in a reaction that is chemically equivalent to the reverse of step 1, resulting in ligation of 5' and 3' exons and release of the intron (103). Excised introns have been observed to circularize, but the significance of this property is not clearly understood (82). An exception to the splicing mechanism described above was observed in an intron (Cbu.L1917) located in the 23S rRNA gene of *C. burnetii* (93). This intron has a 3' terminal adenine in place of the otherwise conserved guanine. Consequently, Cbu.L1917 has a reduced rate of self-splicing in vitro (91).

Group I introns are mainly found inserted in tRNA and rRNA genes of bacteria. Increasingly, they are being found in a variety of protein-coding genes, including those for recombinase A and ribonucleotide reductase (77, 108). In bacteriophages, they are seen both in tRNA genes and in some protein-coding genes, like those for DNA polymerase, ribonucleotide reductase, and thymidylate synthase (45). A bias toward dis-

rupting structural RNA genes in bacteria could be due to the coupling of transcription and translation, which might prevent the ribozyme from attaining its optimum tertiary structure required to splice efficiently (84). However, introns that require concordant translation for efficient splicing are also known (96, 99), showing that introns adapt to their environment. Sexual reproduction in eukaryotes brings intron-containing and intronless alleles of the host gene together, providing an opportunity for introns to spread by homing (see below). Even though rampant HGT provides ample opportunity for the movement of introns and inteins, a lack of sexual reproduction is commonly invoked as an explanation for the apparent scarcity of group I introns in bacteria compared to mitochondria and chloroplasts of lower eukaryotes (29). Another possible reason for this phenomenon is the observed inhibition of bacterial growth caused by these elements. For example, group I introns of *Tetrahymena* and *Coxiella* expressed in *Escherichia coli* were found to associate with ribosomes, inhibit translation, and retard bacterial growth (83, 92). Due to this low-fitness trait, group I introns, like any other gene that de-

creases reproductive success of the host, would presumably be lost from the population by negative selection.

Various group I introns have evolved to associate with other parasitic elements. Twintrons (30), where two distinct group I introns are associated with each other, and IStrons (77, 107), where a group I intron and an insertion sequence (IS) are seen together, are two such cases. Rarely, spliceosomal introns are also found within group I introns (48). As seen below, group I introns also associate with endonucleases.

Inteins. Similar to ribozymes, another iconoclastic discovery was that of inteins (internal proteins). These elements are transcribed and translated together with the host protein but self-excise, leaving the flanking sequences (exteins) spliced together. The first intein was discovered in yeast vacuolar ATPases in 1990 (56). Since then, hundreds of inteins have been discovered in *Bacteria*, *Archaea*, and *Eukarya*. An intein database, InBase, has been established to provide information on all known inteins (87). In bacteria, inteins are found inserted in a variety of conserved proteins, including DNA polymerase, helicase, gyrase, recombinase A, and ribonucleotide reductase, whereas in bacteriophages, inteins have been found in DNA polymerase and ribonucleotide reductase (87). The ratio of intein size to host protein size varies widely, with some inteins being four times as large as the host protein and others only one-tenth the size of the host protein (87).

Inteins have a modular organization consisting of three functional domains: N- and C-terminal splicing domains and an optional endonuclease domain (Fig. 2A) (90). The N-terminal domain is comprised of four motifs, A, B, N2, and N4. The C-terminal domain contains two motifs, F and G. The central endonuclease domain consists of four motifs, C, D, E, and H. The N- and C-terminal motifs are involved in protein splicing and are conserved in most inteins, whereas motifs C, D, E, and H are absent in a number of inteins (referred to as mini-inteins). The first and last amino acids of the intein and the first amino acid of the C-terminal extein are involved in the splicing reaction (39). The first amino acid (motif A) in all inteins is Cys or Ser; the terminal amino acid is a conserved Asn, and the first amino acid that follows the intein is Cys, Ser, or Thr. Inteins splicing involves four successive nucleophilic displacements (Fig. 2B) (87). The first step is an N-O/S acyl shift, where the OH or SH side chain of the amino-terminal Ser or Cys attacks the carbonyl carbon of the preceding amino acid to generate an ester/thioester intermediate linking the N-terminal extein to the side chain of the first intein amino acid, thereby breaking the peptide bond between N-extein and the intein. The second step is a transesterification, where the OH or SH side chain of the first C-extein amino acid attacks the N-terminal ester/thioester bond formed in step 1. This results in the transfer of the N-extein to the side chain of the first C-extein amino acid, forming a branched intermediate. In the third step, the peptide bond between the intein and C-extein is broken by cyclization of the conserved C-terminal Asn to form a succinimide, resulting in intein excision. The N-extein is now attached via an ester bond to the side chain of the first C-extein amino acid. In the final step, the ester bond rapidly undergoes an acyl rearrangement to the thermodynamically more stable, normal peptide bond. In most inteins, the amino acid preceding the terminal Asn is an His, which is thought to assist in Asn cyclization. Some or all of the remaining residues are also important for

FIG. 2. (A) Inteins modular structure (87, 90). An intein with flanking exteins is shown. The N- and C-terminal splicing domains and the optional HE domain with their conserved motifs are shown. Conserved amino acids Cys or Ser at the 5' end of the intein, Asn at the 3' end of the intein, and Cys, Ser, or Thr at the first position on the 3' extein are also indicated. (B) Inteins splicing chemistry (39, 87). N and C exteins are in red and blue, respectively.

proper intein folding to generate the active site (21). Some noncanonical inteins with variations in structure and splicing chemistries have been identified (3, 42, 88).

In some cases, an intein and its host protein (e.g., DnaE) are split into two separate fragments (32, 115). The N and C termini of DnaE, containing the N and C termini of the intein, respectively, are encoded on two separate genes, *dnaE-n* and *dnaE-c*, located on different parts of the genome. Functional

DnaE protein is recreated from the two fragments by the *trans*-splicing activity of the split intein (76). Split inteins have also been found in other enzymes, like ribonucleotide reductase, DNA ligase, gp41, and IMP dehydrogenase (23).

INTRON AND INTEIN MOBILITY THROUGH HES

One of the most successful parasitic genetic elements is the HE. HEs are simple and elegant parasitic elements; a single gene encodes a single protein, and they are inherited in a dominant, non-Mendelian manner (12). HEs are small (<40-kDa) proteins that recognize and cleave long DNA target sequences (usually 14 to 40 bp) which typically occur only once per host genome, thus minimizing any potential negative impact on the host (80). Based on conserved sequence motifs, HEs can be classified into several families (36, 62) that have evolved in parallel to achieve the optimal balance of size, target sequence specificity, and attenuated fidelity to allow for maximum success (97, 120). The LAGLIDADG family has one or two conserved motifs, called the dodecapeptide motifs, with a consensus LAGLIDADG sequence (22), and most HEs found to date belong to this family. The $\beta\beta\alpha$ -Me family contains two subfamilies. The His-Cys box subfamily contains an ~30-amino-acid region with two His and three Cys residues (54), whereas the H-N-H subfamily contains an ~30-amino-acid region with conserved His and Asn residues (102). The GIY-YIG family contains conserved GIY and YIG tripeptides flanking an ~10-amino-acid segment (60). Some HEs utilize catalytic domains acquired from other proteins (or vice versa). An HE that utilizes the PD(D/E)XK motif commonly employed by restriction endonucleases (120), another which uses a domain similar to that of DNA resolvases (119), and a novel HE related to very-short-patch repair endonucleases (23) were recently described. Interestingly, an H-N-H endonuclease makes up the cytotoxic domain of colicin E9, a group A colicin that kills bacteria by nonspecific degradation of chromosomal DNA (81). In theory, any endonuclease that can cause a double-strand break and initiate recombination through flanking-sequence similarity can function as an HE. In fact, the restriction enzyme EcoRI was experimentally made to stimulate intron homing (27).

Insertion of an HE sequence into a gene can potentially impair its function. In order to limit the negative impact on a host and loss by negative selection, HEs tend to associate with other self-splicing elements, like group I introns and inteins that are nearly neutral to selection (6). Together, HEs and introns/inteins have formed a successful, mutually beneficial association wherein the HE provides mobility to the intron/intein and the intron/intein provides a "safe haven" for HE. The process by which the composite element moves from one site to another is called homing (Fig. 3A) (55). The homing mechanism requires that the protein be translated. When an intron/intein-containing allele comes together with an intron/intein-lacking allele, the HE protein binds to a homing site composed of the flanking exon/extein sequences and cleaves it. The host repairs this double-strand DNA break using homologous recombination between the alleles, which results in insertion of the parasitic element into the target sequence. The site is now "immune" to further HE cleavage because the inserted element disrupts the target sequence. Some HEs be-

longing to the H-N-H subfamily create single-strand nicks instead of double-strand breaks and also recognize intron-positive DNAs as substrates (41). The process by which recombination and homing occur after a single-strand nick is not clearly understood. HEs tolerate a degree of variation within their long recognition sequence, which enables them to coevolve with the host target sequence (97) and move to ectopic sites (19). Another reason for HEs' success is their adaptability to a new host, which conceivably helps to explain the divergent DNA binding regions observed between similar HEs from different hosts (18, 73).

HEs go through a dynamic cycle that includes invasion, fixation, inactivation, elimination, and eventual reinvasion (Fig. 3B) (37–39). Once an HE-containing parasitic element invades a new host, it spreads to all the individuals in that host's population and becomes fixed in that population. Once fixed, the HE becomes nonfunctional (since there are no available target sequences), starts to degenerate, and eventually becomes lost (13). In fact, a large number of group I introns and inteins have lost their respective HE genes. The intron/intein itself will maintain its sequence, because any change will affect its splicing ability, thus negatively impacting the host. Eventually, the whole element is lost from the population by a precise deletion event. The parasitic genetic element reappears in the population only through a new HGT event, a critical process for the long-term maintenance of a parasitic genetic element in a population. Some exceptions to this "homing cycle" model have been described (38). To prevent being purged from a genome, some HEs utilize an intriguing strategy: they have evolved a maturase function (98). Maturase activity promotes intron splicing by stabilizing RNA folding. To function as a maturase, HEs have evolved an RNA-binding site in addition to their DNA-binding site, showing their adaptability (72). Some HEs confer beneficial functions upon their hosts. VDE (also known as PI-SceI), an HE found inserted in a self-splicing intein in the *VMA-1* gene of *Saccharomyces cerevisiae*, is one of the main regulators of the host's high-affinity glutathione transporter (79). HO (F-SceII), a free-standing HE in *S. cerevisiae*, mediates mating-type switching (14). An intron-encoded HE was shown to provide a selective advantage to intron-containing *Sulfolobus acidocaldarius* cells over cells without the intron (1), and I-HmuI, an intron-encoded HE of *Bacillus subtilis* phage SP82, is required for exclusion of DNA from the related phage SPOI in the progeny of mixed infections (41).

Another mechanism of intron mobility involves reverse splicing. In this process, excised intron-RNA base pairs with host RNA sequences that are complementary to its IGS, followed by integration into the transcript by a reversal of the splicing process. The intron then becomes inserted into the corresponding gene through reverse transcription followed by recombination. Reverse splicing has been demonstrated in the lab (95) and inferred from intron transposition patterns (8, 49).

DISPARATE ORIGINS, CONVERGENT PARASITIC STRATEGIES

Both group I introns and inteins arose from preexisting molecules with autocatalytic abilities (Fig. 4). The progenitor of group I introns is thought to be a prebiotic catalytic RNA

FIG. 3. (A) Intron or group I intron homing through HE-mediated DNA double-strand-break repair and recombination. (B) Cycle model of HE gain, degeneration, and loss within host populations (37, 38, 39).

FIG. 4. Convergence of evolutionary paths of group I introns, inteins, and HEs.

(66). These primordial ribozymes were part of the “RNA world,” which predates current DNA- and protein-based biology. Some of the relics of the era when RNA functioned as a catalyst are found strewn across the contemporary biosphere: RNase P, ribosomes, spliceosomes, telomerase, and self-splicing introns (16). Although RNA is a good informational molecule and a powerful catalyst, it is not clear how these molecules replicated themselves in the preprotein world. Recently, Vicens and Cech showed that group I introns have the potential to polymerize RNA chains by forming 3',5' phosphodiester bonds, a milestone in the search for a prebiotic replicase (109). In addition, earlier in vitro evolution studies demonstrated that group I introns can catalyze RNA ligation (52, 118). Taken together, these observations suggest that modern group I introns arose when the self-splicing activity of a primordial replicase-like ribozyme was exploited for mobilization of the molecule as a parasitic genetic element.

Intein, the protein analog of a group I intron, is thought to have evolved from an ancient protein domain possessing the ability to self-cleave. A number of proteins involved in important biological processes are known to self-cleave (86). Among them, Hedgehog developmental proteins found in eumetazoans are evolutionarily related to inteins (43, 58). The self-cleaving C-terminal portion (Hog domain) of these proteins has a domain called Hint (Hedgehog and intein) that shares structural, sequence, and biochemical similarities with inteins (58). The Hog domain autocatalyzes its cleavage and attaches a cholesterol moiety to the N-terminal Hedge domain. The

cleaved Hedge domain with the attached cholesterol is secreted from the cell and serves as a signaling molecule (11). In addition to inteins and Hedgehog proteins, domains related to Hint have been identified and termed bacterial intein-like domains (BILs). Three types of BILs have been identified so far (BIL-A, -B, and -C) from diverse bacteria, including human and plant pathogens and predatory bacteria (2, 26). Each type of BIL is as different from the others as it is from Hedgehog-Hint and intein (11). BILs are hypothesized to generate host protein diversity and aid in host microevolution (2, 26). Since the initial N-S/O acyl shift common to all Hint domains is employed by numerous other self-cleaving proteins involved in diverse biological processes (117), it is presumed that the progenitor Hint domain arose by positive selection for some advantageous biological function. Duplication of the primordial Hint domain followed by a loop exchange would have resulted in a protein that autocleaves but leaves the host protein intact by splicing together the N- and C-terminal flanks (69), setting the stage for the evolution of a parasitic protein.

The primary requirement of a successful genetic parasite is to cause no harm to the host, which both group I introns and inteins accomplish by splicing out at the RNA and protein stages, respectively. This attribute aids in their maintenance only to a point, because the lack of benefit to the host puts them under constant threat of being purged from the compact bacterial genome, which is biased toward deletion. To counter loss by deletion, group I introns and inteins have evolved convergent strategies to improve mobility, maximize insertion

site availability, and minimize the probability of being lost from a genome.

Group I introns can independently move from their insertion site to a new site in the same organism (transposition) or to the same site in a different organism (HGT) through reverse splicing (95, 114). A 4- to 6-bp sequence complementarity between the intron's IGS and target RNA is all that is required to initiate reverse splicing, effectively providing any group I intron with a large pool of potential targets (114). However, successful integration of the intron into DNA occurs only if the target RNA along with the intron is reverse transcribed and subsequently undergoes recombination, surely a rare sequence of events that limits the spread of introns (29). Even when an intron reverse splices successfully and inserts into a new target DNA sequence, the forward splicing efficiency is much lower than at its natural site (89, 94), further restricting the spread of introns via reverse splicing. In the case of mini-inteins, it is not clear how they move from one insertion site to another. They do not employ recognizable transposition mechanisms, and their coding DNA or RNA is not known to separate from the host sequence. They might be able to transpose to a new location along with their flanking sequences by a rare nonhomologous recombination event. However, for this insertion to be successful, the intein must be in-frame and must be active at its new site, and the host protein should be functional even with the inserted intein flanks (89). Many group I introns and mini-inteins have solved these mobility limitations and broadened their potential target repertoire by linking up with endonucleases (25, 47).

The evolutionary steps that brought group I introns and inteins together with HEs to form composite mobile genetic elements have intrigued scientists since their discovery. Several studies have shown that HEs and their corresponding introns or inteins have separate phylogenetic histories (36, 46). Also, similar inteins and introns were found to contain different types of HEs, and closely related HEs are known to be encoded within distantly related introns (48). Moreover, "free-standing" endonucleases that are mobile, even without being associated with an intron or an intein, have been found abundantly in some bacteriophage genomes (7, 101). Taken together, these observations suggest that mobile inteins and group I introns evolved by repeated, independent invasions of mini-inteins and endonuclease-free introns by HEs (25). Introns/inteins and HEs can come together by recombination. The newly formed bipartite intron/intein will be maintained only if intron/intein splicing is maintained and if the HE can promote the spread of its host element by homing. To this end, HEs are always found inserted in peripheral loops that do not play a role in intron splicing or in intein-domains that are not essential for splicing. Evolutionary forces that drive this process recently came to light when David Shub and colleagues showed that the propensity of introns and HEs for targeting highly conserved sequences within conserved genes might have brought them together (9, 119) (Fig. 4).

The best strategy for parasitic elements to increase the prospect of finding an insertion site is to target DNA sequences that are most commonly encountered in a gene pool. Genes that play essential biological roles tend to be conserved across the biological spectrum and consequently serve as frequent insertion sites for parasitic genetic elements. Within these

genes, group I introns and inteins insert into sequences that remain static due to their functional importance (59, 97), further maximizing target availability. Only a rare, precise deletion event that removes the element and exactly recreates the host gene sequence will result in functional gene products, whereas an imprecise deletion can potentially mutate critical regions within the gene, resulting in deleterious consequences to the host. Hence, targeting highly conserved sequences within conserved genes offers the additional benefit of minimizing the likelihood of the molecular parasite being purged from the host genome (28).

It has been known for some time that free-standing HEs can move between bacteriophage genomes through a process called intronless homing (7). An example is the free-standing HE (SegF) in T4 phage that targets and cleaves a conserved sequence within the adjacent gene 56 of related T2 phage during mixed infections. Repair of the double-strand break results in the replacement of T2's gene 56 with that of T4's gene 56 along with the insertion of SegF. In this manner, even though the HE is inserted in a less-conserved intergenic region, its maintenance and mobility are maximized by targeting a neighboring conserved sequence (28). As discussed above, introns and inteins also target conserved genes for insertion. Since bacteriophages have a limited repertoire of conserved genes, it is inevitable that an intron/intein and an HE targeting the same conserved sequence come together in the same phage genome. When this happens, the intron/intein and the free-standing HE can move together from one host to another by a process termed collaborative homing (9, 119). A rare recombination event can insert the HE into the intron/intein without affecting its splicing, thereby giving rise to a composite parasitic element. This stable chimera can now efficiently spread through the population by homing to cognate sites and rarely even to ectopic sites. Alternatively, in some cases an HE might invade an intron due to the presence of a cleavage site within a peripheral loop of the intron (48, 71). In this event, the HE must adapt quickly to recognize and cleave intronless alleles of the target gene so as to facilitate homing, which in turn will ensure its maintenance within the intron. The intron/intein provides a "safe haven" for the HE, where it can evolve quickly to acclimate to the new target sequences. HEs are fast-evolving genes (18, 44) that are known to use a wide array of protein scaffolds (119, 120) of diverse origins. Some HEs are even chimeras, with N and C termini from different sources (7). Moreover, double-LAGLIDADG-motif HEs that have evolved from single-motif ancestors are more successful at invading divergent target sites, thereby promoting the spread of the composite genetic parasite to ectopic sites (44).

Bacteriophages are an ancient and genetically nonhomogeneous group that is thought to be the most abundant biological entity on earth (10, 15, 34). High rates of homologous and nonhomologous recombination, rapid evolution, and profuse genetic exchange provide phages with powerful means of innovation. Hence, they are considered the "start-up" entities where several new bacterial genes originate (24). In the same vein, phages most likely are the melting pot where inteins and group I introns associate with free-standing HEs to form composite parasitic elements. For homing to take place, both intron-positive and intron-negative alleles of the host gene must come together. It is not clear how this could happen frequently

enough in bacteria to facilitate the spread of group I introns and inteins (13). By contrast, bacteriophages have a global gene pool with intense horizontal exchange occurring among genetically related local groups of phages and low-grade exchange of sequences occurring over wide phylogenetic distances (34), making them ideal vehicles for the transfer of HEs, introns, and inteins between phages, among bacteria, and between phages and bacteria. Further, in lysogeny the prophage provides a silent locus from which HEs can invade homologous host genes in bacteria (108). Introns and inteins also tend to target conserved genes that have both phage and bacterial copies, making insertions less toxic and improving the odds of homologous recombination (59, 77). In addition to transduction, a specialized conjugation system (111) and natural competence (17) might explain the observed abundance of inteins and introns, respectively, in *Mycobacterium* (87) and *Bacillus* (108) species.

HARNESSING GENETIC PARASITES

Similar to inteins, group I introns have also been shown to mediate *trans*-splicing of exons contained on different RNAs (57). This property has potential use as a gene therapy tool. In fact, group I introns have been used to convert sickle β -globin transcripts into RNAs encoding γ -globin in an effort to treat sickle cell disease (64). *trans*-splicing ribozymes also have the potential to be used for targeted gene delivery and as therapeutic cytotoxins (5, 57).

The protein splicing ability of inteins has been exploited as a biotechnology tool. Inteins can be used as tags to purify fusion proteins in place of traditional histidine tags. After purification, the intein tag can be removed by utilizing the self-cleaving property of the intein (112). Other potential uses for inteins include the semisynthesis of cytotoxic proteins (31) and introducing nuclear magnetic resonance labels into part of a large protein (85). The *trans*-splicing ability of split inteins has been utilized for the synthesis of cyclic peptide libraries and in gene therapy (67, 106).

The ability to introduce specific double-strand DNA breaks makes HEs a very useful genetic tool. Some HEs are commercially available, e.g., I-Ceu I, I-Sce I, and PI-Psp I (New England Biolabs). HEs, along with other rare-cutting restriction enzymes, have been used to map bacterial genomes, especially to analyze chromosomal organization (68). HEs have been used to study double-strand-break repair mechanisms in phages, yeasts, plants, and mammalian cells and to study chromosomal repair systems in *Drosophila* (40, 53). Recently, artificial HEs were engineered with the aim of using them in human gene therapy (20).

In conclusion, self-splicing group I introns and inteins in bacteria originated from disparate self-cleaving sources but evolved convergently to target conserved gene sequences and in turn to associate with HEs to maximize their persistence and spread. Bacteriophages might be the vessels where composite inteins and introns originate and the vehicles for their spread. With the current expansion of genomic data, more mobile genetic elements are being reported, and a concerted effort is needed to analyze and understand all of them. Decoding the unique biological and chemical properties of these intriguing

elements will provide us with novel tools for industrial and medical applications.

ACKNOWLEDGMENTS

We thank the members of our lab for helpful discussions and technical assistance. We are grateful to the reviewers for their comments, which improved the manuscript immensely.

Our work was supported by the NIH Rocky Mountain Regional Center of Excellence for Biodefense and Emerging Infectious Disease grant U54 AI065357-040023 and by NIH grant R21 AI078125.

REFERENCES

1. Aagaard, C., J. Z. Dalggaard, and R. A. Garrett. 1995. Intercellular mobility and homing of an archaeal rDNA intron confers a selective advantage over intron- cells of *Sulfolobus acidocaldarius*. *Proc. Natl. Acad. Sci. USA* **92**:12285–12289.
2. Amitai, G., O. Belenkiy, B. Dassa, A. Shainskaya, and S. Pietrokovski. 2003. Distribution and function of new bacterial intein-like protein domains. *Mol. Microbiol.* **47**:61–73.
3. Amitai, G., B. Dassa, and S. Pietrokovski. 2004. Protein splicing of inteins with atypical glutamine and aspartate C-terminal residues. *J. Biol. Chem.* **279**:3121–3131.
4. Avise, J. C. 2001. Evolving genomic metaphors: a new look at the language of DNA. *Science* **294**:86–87.
5. Ayre, B. G., U. Köhler, H. M. Goodman, and J. Haseloff. 1999. Design of highly specific cytotoxins by using *trans*-splicing ribozymes. *Proc. Natl. Acad. Sci. USA* **96**:3507–3512.
6. Belfort, M. 2003. Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes Dev.* **17**:2860–2863.
7. Belle, A., M. Landthaler, and D. A. Shub. 2002. Intronless homing: site-specific endonuclease SegF of bacteriophage T4 mediates localized marker exclusion analogous to homing endonucleases of group I introns. *Genes Dev.* **16**:351–362.
8. Bhattacharya, D., V. Reebe, D. M. Simon, and F. Lutzoni. 2005. Phylogenetic analyses suggest reverse splicing spread of group I introns in fungal ribosomal DNA. *BMC Evol. Biol.* **5**:68.
9. Bonocora, R. P., and D. A. Shub. 2009. A likely pathway for formation of mobile group I introns. *Curr. Biol.* **19**:223–228.
10. Brüssow, H., and R. W. Hendrix. 2002. Phage genomics: small is beautiful. *Cell* **108**:13–16.
11. Bürglin, T. R. 2008. The Hedgehog protein family. *Genome Biol.* **9**:241.
12. Burt, A., and R. Trivers. 2005. *Genes in conflict: the biology of selfish genetic elements*. Harvard University Press, Cambridge, MA.
13. Burt, A., and V. Koufopanou. 2004. Homing endonuclease genes: the rise and fall and rise again of a selfish element. *Curr. Opin. Genet. Dev.* **14**:609–615.
14. Butler, G., C. Kenny, A. Fagan, C. Kurischko, C. Gaillardin, and K. H. Wolfe. 2004. Evolution of the MAT locus and its Ho endonuclease in yeast species. *Proc. Natl. Acad. Sci. USA* **101**:1632–1637.
15. Canchaya, C., G. Fournous, S. Chibani-Chennoufi, M. L. Dillmann, and H. Brüssow. 2003. Phage as agents of lateral gene transfer. *Curr. Opin. Microbiol.* **6**:417–424.
16. Cech, T. R. 2009. Crawling out of the RNA world. *Cell* **136**:599–602.
17. Chen, L., P. J. Christie, and D. Dubnau. 2005. The ins and outs of DNA transfer in bacteria. *Science* **310**:1456–1460.
18. Chevalier, B., M. Turmel, C. Lemieux, R. J. Monnat, Jr., and B. L. Stoddard. 2003. Flexible DNA target site recognition by divergent homing endonuclease isoschizomers I-CreI and I-MsoI. *J. Mol. Biol.* **329**:253–269.
19. Chevalier, B. S., and B. L. Stoddard. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* **29**:3757–3774.
20. Collins, C. H., Y. Yokobayashi, D. Umeno, and F. H. Arnold. 2003. Engineering proteins that bind, move, make and break DNA. *Curr. Opin. Biotechnol.* **14**:665.
21. Cooper, A. A., Y. J. Chen, M. A. Lindorfer, and T. H. Stevens. 1993. Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J.* **12**:2575–2583.
22. Dalggaard, J. Z., A. J. Klar, M. J. Moser, W. R. Holley, A. Chatterjee, and I. S. Mian. 1997. Statistical modeling and analysis of the LAGLIDADG family of site-specific endonucleases and identification of an intein that encodes a site-specific endonuclease of the HNH family. *Nucleic Acids Res.* **25**:4626–4638.
23. Dassa, B., N. London, B. L. Stoddard, O. Schueler-Furman, and S. Pietrokovski. 2009. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res.* **37**:2560–2573.
24. Daubin, V., and H. Ochman. 2004. Start-up entities in the origin of new genes. *Curr. Opin. Genet. Dev.* **14**:616–619.
25. Derbyshire, V., D. W. Wood, W. Wu, J. T. Dansereau, J. Z. Dalggaard, and

- M. Belfort. 1997. Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proc. Natl. Acad. Sci. USA* **94**:11466–11471.
26. Dori-Bachash, M., B. Dassa, O. Peleg, S. A. Pineiro, E. Jurkevitch, and S. Pietrokovski. 2009. Bacterial intein-like domains of predatory bacteria: a new domain type characterized in *Bdellovibrio bacteriovorus*. *Funct. Integr. Genomics* **9**:153–166.
 27. Eddy, S. R., and L. Gold. 1992. Artificial mobile DNA element constructed from the EcoRI endonuclease gene. *Proc. Natl. Acad. Sci. USA* **89**:1544–1547.
 28. Edgell, D. R. 2009. Selfish DNA: homing endonucleases find a home. *Curr. Biol.* **19**:R115–117.
 29. Edgell, D. R., M. Belfort, and D. A. Shub. 2000. Barriers to intron promiscuity in bacteria. *J. Bacteriol.* **182**:5281–5289.
 30. Einvik, C., M. Elde, and S. Johansen. 1998. Group I twintrons: genetic elements in myxomycete and schizopyrenid amoeboflagellate ribosomal DNAs. *J. Biotechnol.* **64**:63–74.
 31. Evans, T. C., Jr., J. Benner, and M. Q. Xu. 1998. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* **7**:2256–2264.
 32. Evans, T. C., Jr., D. Martin, R. Kolly, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X. Q. Liu, and M. Q. Xu. 2000. Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803. *J. Biol. Chem.* **275**:9091–9094.
 33. Everett, K. D., S. Kahane, R. M. Bush, and M. G. Friedman. 1999. An unspliced group I intron in 23S rRNA links *Chlamydiales*, chloroplasts, and mitochondria. *J. Bacteriol.* **181**:4734–4740.
 34. Filée, J., P. Forterre, and J. Laurent. 2003. The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res. Microbiol.* **154**:237–243.
 35. Gilbert, W. 1986. Origin of life: the RNA world. *Nature* **319**:618.
 36. Gimble, F. S. 2000. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol. Lett.* **185**:99–107.
 37. Goddard, M. R., and A. Burt. 1999. Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. USA* **96**:13880–13885.
 38. Gogarten, J. P., and E. Hilario. 2006. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol. Biol.* **6**:94.
 39. Gogarten, J. P., A. G. Senejani, O. Zhaxybayeva, L. Olendzenski, and E. Hilario. 2002. Inteins: structure, function, and evolution. *Annu. Rev. Microbiol.* **56**:263–287.
 40. Gong, W. J., and K. G. Golic. 2003. Ends-out, or replacement, gene targeting in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **100**:2556–2561.
 41. Goodrich-Blair, H., and D. A. Shub. 1996. Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell* **84**:211–221.
 42. Gorbalenya, A. E. 1998. Non-canonical inteins. *Nucleic Acids Res.* **26**:1741–1748.
 43. Hall, T. M., J. A. Porter, K. E. Young, E. V. Koonin, P. A. Beachy, and D. J. Leahy. 1997. Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* **91**:85–97.
 44. Haugen, P., and D. Bhattacharya. 2004. The spread of LAGLIDADG homing endonuclease genes in rDNA. *Nucleic Acids Res.* **32**:2049–2057.
 45. Haugen, P., D. Bhattacharya, J. D. Palmer, S. Turner, L. A. Lewis, and K. M. Pryer. 2007. Cyanobacterial ribosomal RNA genes with multiple, endonuclease-encoding group I introns. *BMC Evol. Biol.* **7**:159.
 46. Haugen, P., V. Reeß, F. Lutzoni, and D. Bhattacharya. 2004. The evolution of homing endonuclease genes and group I introns in nuclear rDNA. *Mol. Biol. Evol.* **21**:129–140.
 47. Haugen, P., D. M. Simon, and D. Bhattacharya. 2005. The natural history of group I introns. *Trends Genet.* **21**:111–119.
 48. Haugen, P., O. G. Wikmark, A. Vader, D. H. Coucheron, E. Sjøttem, and S. D. Johansen. 2005. The recent transfer of a homing endonuclease gene. *Nucleic Acids Res.* **33**:2734–2741.
 49. Hoshina, R., and N. Imamura. 2009. Phylogenetically close group I introns with different positions among *Paramecium bursaria* photobionts imply a primitive stage of intron diversification. *Mol. Biol. Evol.* **26**:1309–1319.
 50. Hurst, G. D., and J. H. Werren. 2001. The role of selfish genetic elements in eukaryotic evolution. *Nat. Rev. Genet.* **2**:597–606.
 51. Ikawa, Y., H. Shiraishi, and T. Inoue. 2000. Minimal catalytic domain of a group I self-splicing intron RNA. *Nat. Struct. Biol.* **7**:1032–1035.
 52. Jaeger, L., M. C. Wright, and G. F. Joyce. 1999. A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proc. Natl. Acad. Sci. USA* **96**:14712–14717.
 53. Jasin, M. 1996. Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.* **12**:224–228.
 54. Johansen, S., T. M. Embley, and N. P. Willassen. 1993. A family of nuclear homing endonucleases. *Nucleic Acids Res.* **21**:4405.
 55. Jurica, M. S., and B. L. Stoddard. 1999. Homing endonucleases: structure, function and evolution. *Cell. Mol. Life Sci.* **55**:1304–1326.
 56. Kane, P. M., C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebel, and T. H. Stevens. 1990. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**:651–657.
 57. Kohler, U., B. G. Ayre, H. M. Goodman, and J. Haseloff. 1999. Trans-splicing ribozymes for targeted gene delivery. *J. Mol. Biol.* **285**:1935–1950.
 58. Koonin, E. V. 1995. A protein splice-junction motif in hedgehog family proteins. *Trends Biochem. Sci.* **20**:141–142.
 59. Koufopanou, V., M. R. Goddard, and A. Burt. 2002. Adaptation for horizontal transfer in a homing endonuclease. *Mol. Biol. Evol.* **19**:239–246.
 60. Kowalski, J. C., M. Belfort, M. A. Stapleton, M. Holpert, J. T. Dansereau, S. Pietrokovski, S. M. Baxter, and V. Derbyshire. 1999. Configuration of the catalytic G1Y-Y1G domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res.* **27**:2115–2125.
 61. Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech. 1982. Self-splicing RNA: autocleavage and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**:147–157.
 62. Kühmann, U. C., G. R. Moore, R. James, C. Kleanthous, and A. M. Hemmings. 1999. Structural parsimony in endonuclease active sites: should the number of homing endonuclease families be redefined? *FEBS Lett.* **463**:1–2.
 63. Kuhse, M. G., R. Strickland, and J. D. Palmer. 1990. An ancient group I intron shared by eubacteria and chloroplasts. *Science* **250**:1570–1573.
 64. Lan, N., R. P. Howrey, S. W. Lee, C. A. Smith, and B. A. Sullenger. 1998. Ribozyme-mediated repair of sickle beta-globin mRNAs in erythrocyte precursors. *Science* **280**:1593–1596.
 65. Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
 66. Lehman, N. 2009. A ghost in the RNA machine. *Nat. Chem. Biol.* **5**:73–74.
 67. Li, J., W. Sun, B. Wang, X. Xiao, and X. Q. Liu. 2008. Protein trans-splicing as a means for viral vector-mediated in vivo gene therapy. *Hum. Gene Ther.* **19**:958–964.
 68. Liu, S. L., and K. E. Sanderson. 1996. Highly plastic chromosomal organization in *Salmonella typhi*. *Proc. Natl. Acad. Sci. USA* **93**:10303–10308.
 69. Liu, X. Q. 2000. Protein-splicing intein: genetic mobility, origin, and evolution. *Annu. Rev. Genet.* **34**:61–76.
 70. Logsdon, J. M., Jr. 1998. The recent origins of spliceosomal introns revisited. *Curr. Opin. Genet. Dev.* **8**:637–648.
 71. Loizos, N., E. R. Tillier, and M. Belfort. 1994. Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. *Proc. Natl. Acad. Sci. USA* **91**:11983–11987.
 72. Longo, A., C. W. Leonard, G. S. Bassi, D. Berndt, J. M. Krahn, T. M. Hall, and K. M. Weeks. 2005. Evolution from DNA to RNA recognition by the b13 LAGLIDADG maturase. *Nat. Struct. Mol. Biol.* **12**:779–787.
 73. Lucas, P., C. Otis, J. P. Mercier, M. Turmel, and C. Lemieux. 2001. Rapid evolution of the DNA-binding site in LAGLIDADG homing endonucleases. *Nucleic Acids Res.* **29**:960–969.
 74. Lykke-Andersen, J., C. Aagaard, M. Semiononov, and R. A. Garrett. 1997. Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem. Sci.* **22**:326–331.
 75. Marraffini, L. A., and E. J. Sontheimer. 2008. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**:1843–1845.
 76. Martin, D. D., M. Q. Xu, and T. C. Evans, Jr. 2001. Characterization of a naturally occurring trans-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry* **40**:1393–1402.
 77. Meng, Q., Y. Zhang, and X. Q. Liu. 2007. Rare group I intron with insertion sequence element in a bacterial ribonucleotide reductase gene. *J. Bacteriol.* **189**:2150–2154.
 78. Michel, F., and E. Westhof. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**:585–610.
 79. Miyake, T., H. Hiraishi, H. Sammoto, and B. Ono. 2003. Involvement of the VDE homing endonuclease and rapamycin in regulation of the *Saccharomyces cerevisiae* GSH11 gene encoding the high affinity glutathione transporter. *J. Biol. Chem.* **278**:39632–39636.
 80. Moran, J. V., S. Zimmerly, R. Eskes, J. C. Kennell, A. M. Lambowitz, R. A.

- Butow, and P. S. Perlman. 1995. Mobile group II introns of yeast mitochondrial DNA are novel site-specific retroelements. *Mol. Cell. Biol.* **15**: 2828–2838.
81. Mosbahi, K., C. Lemaître, A. H. Keeble, H. Mobasher, B. Morel, R. James, G. R. Moore, E. J. Lea, and C. Kleanthous. 2002. The cytotoxic domain of colicin E9 is a channel-forming endonuclease. *Nat. Struct. Biol.* **9**:476–484.
 82. Nielsen, H., T. Fiskaa, A. B. Birgisdottir, P. Haugen, C. Einvik, and S. Johansen. 2003. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA* **9**:1464–1475.
 83. Nikolcheva, T., and S. A. Woodson. 1997. Association of a group I intron with its splice junction in 50S ribosomes: implications for intron toxicity. *RNA* **3**:1016–1027.
 84. Ohman-Hedén, M., A. Ahgren-Stålhandske, S. Hahne, and B. M. Sjöberg. 1993. Translation across the 5'-splice site interferes with autocatalytic splicing. *Mol. Microbiol.* **7**:975–982.
 85. Otomo, T., K. Teruya, K. Uegaki, T. Yamazaki, and Y. Kyogoku. 1999. Improved segmental isotope labeling of proteins and application to a larger protein. *J. Biomol. NMR* **14**:105–114.
 86. Paulus, H. 2000. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* **69**:447–496.
 87. Perler, F. B. 2002. InBase: the Intein database. *Nucleic Acids Res.* **30**:383–384.
 88. Pietrovski, S. 1998. Identification of a virus intein and a possible variation in the protein-splicing reaction. *Curr. Biol.* **8**:R634–635.
 89. Pietrovski, S. 2001. Intein spread and extinction in evolution. *Trends Genet.* **17**:465–472.
 90. Pietrovski, S. 1998. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci.* **7**:64–71.
 91. Raghavan, R., L. D. Hicks, and M. F. Minnick. 2009. A unique group I intron in *Coxiella burnetii* is a natural splice mutant. *J. Bacteriol.* **191**:4044–4046.
 92. Raghavan, R., L. D. Hicks, and M. F. Minnick. 2008. Toxic introns and parasitic intein in *Coxiella burnetii*: legacies of a promiscuous past. *J. Bacteriol.* **190**:5934–5943.
 93. Raghavan, R., S. R. Miller, L. D. Hicks, and M. F. Minnick. 2007. The unusual 23S rRNA gene of *Coxiella burnetii*: two self-splicing group I introns flank a 34-base-pair exon, and one element lacks the canonical Ω G. *J. Bacteriol.* **189**:6572–6579.
 94. Roman, J., M. N. Rubin, and S. A. Woodson. 1999. Sequence specificity of in vivo reverse splicing of the *Tetrahymena* group I intron. *RNA* **5**:1–13.
 95. Roman, J., and S. A. Woodson. 1998. Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. *Proc. Natl. Acad. Sci. USA* **95**:2134–2139.
 96. Sandegren, L., and B. M. Sjöberg. 2007. Self-splicing of the bacteriophage T4 group I introns requires efficient translation of the pre-mRNA in vivo and correlates with the growth state of the infected bacterium. *J. Bacteriol.* **189**:980–990.
 97. Scalley-Kim, M., A. McConnell-Smith, and B. L. Stoddard. 2007. Coevolution of a homing endonuclease and its host target sequence. *J. Mol. Biol.* **372**:1305–1319.
 98. Schäfer, B., B. Wilde, D. R. Massardo, F. Manna, L. Del Giudice, and K. Wolf. 1994. A mitochondrial group-I intron in fission yeast encodes a maturase and is mobile in crosses. *Curr. Genet.* **25**:336–341.
 99. Semrad, K., and R. Schroeder. 1998. A ribosomal function is necessary for efficient splicing of the T4 phage thymidylate synthase intron in vivo. *Genes Dev.* **12**:1327–1337.
 100. Seshadri, R., I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser, and J. F. Heidelberg. 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl. Acad. Sci. USA* **100**:5455–5460.
 101. Sharma, M., R. L. Ellis, and D. M. Hinton. 1992. Identification of a family of bacteriophage T4 genes encoding proteins similar to those present in group I introns of fungi and phage. *Proc. Natl. Acad. Sci. USA* **89**:6658–6662.
 102. Shub, D. A., H. Goodrich-Blair, and S. R. Eddy. 1994. Amino acid sequence motif of group I intron endonucleases is conserved in open reading frames of group II introns. *Trends Biochem. Sci.* **19**:402–404.
 103. Stahley, M. R., and S. A. Strobel. 2006. RNA splicing: group I intron crystal structures reveal the basis of splice site selection and metal ion catalysis. *Curr. Opin. Struct. Biol.* **16**:319–326.
 104. Strobel, S. A., and J. C. Cochrane. 2007. RNA catalysis: ribozymes, ribosomes, and riboswitches. *Curr. Opin. Chem. Biol.* **11**:636–643.
 105. Suh, S. O., K. G. Jones, and M. Blackwell. 1999. A group I intron in the nuclear small subunit rRNA gene of *Cryptosporidium parvum*, an ascomycetous fungus: evidence for a new major class of group I introns. *J. Mol. Evol.* **48**:493–500.
 106. Tavassoli, A., and S. J. Benkovic. 2007. Split-intein mediated circular ligation used in the synthesis of cyclic peptide libraries in *E. coli*. *Nat. Protoc.* **2**:1126–1133.
 107. Tourasse, N. J., E. Helgason, O. A. Økstad, I. K. Hegna, and A. B. Kolstø. 2006. The *Bacillus cereus* group: novel aspects of population structure and genome dynamics. *J. Appl. Microbiol.* **101**:579–593.
 108. Tourasse, N. J., and A. B. Kolstø. 2008. Survey of group I and group II introns in 29 sequenced genomes of the *Bacillus cereus* group: insights into their spread and evolution. *Nucleic Acids Res.* **36**:4529–4548.
 109. Vicens, Q., and T. R. Cech. 2009. A natural ribozyme with 3',5' RNA ligase activity. *Nat. Chem. Biol.* **5**:97–99.
 110. Vicens, Q., and T. R. Cech. 2006. Atomic level architecture of group I introns revealed. *Trends Biochem. Sci.* **31**:41–51.
 111. Wang, J., L. M. Parsons, and K. M. Derbyshire. 2003. Unconventional conjugal DNA transfer in mycobacteria. *Nat. Genet.* **34**:80–84.
 112. Wood, D. W., W. Wu, G. Belfort, V. Derbyshire, and M. Belfort. 1999. A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotechnol.* **17**:889–892.
 113. Woodson, S. A. 2005. Structure and assembly of group I introns. *Curr. Opin. Struct. Biol.* **15**:324–330.
 114. Woodson, S. A., and T. R. Cech. 1989. Reverse self-splicing of the tetrahymena group I intron: implication for the directionality of splicing and for intron transposition. *Cell* **57**:335–345.
 115. Wu, H., M. Q. Xu, and X. Q. Liu. 1998. Protein *trans*-splicing and functional mini-inteins of a cyanobacterial *dnaB* intein. *Biochim. Biophys. Acta* **1387**: 422–432.
 116. Xu, M. Q., S. D. Kathe, H. Goodrich-Blair, S. A. Nierzwicki-Bauer, and D. A. Shub. 1990. Bacterial origin of a chloroplast intron: conserved self-splicing group I introns in cyanobacteria. *Science* **250**:1566–1570.
 117. Xu, Q., D. Buckley, C. Guan, and H. C. Guo. 1999. Structural insights into the mechanism of intramolecular proteolysis. *Cell* **98**:651–661.
 118. Yoshioka, W., Y. Ikawa, L. Jaeger, H. Shiraishi, and T. Inoue. 2004. Generation of a catalytic module on a self-folding RNA. *RNA* **10**:1900–1906.
 119. Zeng, Q., R. P. Bonocora, and D. A. Shub. 2009. A free-standing homing endonuclease targets an intron insertion site in the *psbA* gene of cyanophages. *Curr. Biol.* **19**:218–222.
 120. Zhao, L., R. P. Bonocora, D. A. Shub, and B. L. Stoddard. 2007. The restriction fold turns to the dark side: a bacterial homing endonuclease with a PD-(D/E)-XK motif. *EMBO J.* **26**:2432–2442.