

Pre-Proof Version of submitted manuscript

Final version published in
Bacterial Integrative Mobile Genetic Elements,
edited by Adam P. Roberts and Peter Mullany.
©2011 Landes Bioscience.

Contributors' information

Kristen S. Swithers, Graduate Student, Department of Molecular and Cell Biology,
University of Connecticut, 91 North Eagleville Road, Storrs, CT 06269-3125, USA
Phone: 1 860 486 1887, FAX: 1 860 486 4331, Email: kswithers@gmail.com

Johann Peter Gogarten, Board of Trustees Distinguished Professor, Department of
Molecular and Cell Biology, University of Connecticut, 91 North Eagleville Road, Storrs, CT
06269-3125, USA
Phone: 1 860 486 4061, FAX: 1 860 486 4331, Email: jpgogarten@gmail.com

Inteins and Introns

Kristen S. Swithers and J. Peter Gogarten*

* Corresponding author: J. Peter Gogarten, Department of Molecular and Cell Biology,
University of Connecticut, Storrs, CT 06269-3125, USA. Email: jpgogarten@gmail.com

ABSTRACT

Introns and Inteins compose a group of mobile genetic elements that are ubiquitous across the tree of life. This chapter describes these parasitic genetic elements, specifically group I introns, group II introns, and inteins. These genetic elements are spliced out of either RNA or proteins after transcription or translation. Some of these molecular parasites encode homing endonucleases that help in the invasion of intein/intron free alleles. In addition, free-standing homing endonucleases are considered parasites in their own right. The molecular parasites that use a homing endonuclease have a life cycle of their own called the homing cycle and have played a role in gene and genome evolution.

INTRODUCTION

The discovery of the intron is one of the most fundamental discoveries of the twentieth century. The concept of what is now termed the intron was first introduced in 1977 by several groups. A large mRNA early transcript in the eukaryotic nucleus was observed in adenovirus 2 and in HeLa cells that appeared to be post transcriptional modified to produce a significantly smaller late mRNA transcript.¹⁻⁴ The intragenic culprit of this peculiar phenomena was later termed to be the intron.⁵ These spliceosomal introns are processed out of the mRNA during maturation by the spliceosome, which can be composed of up to 200 unique proteins.⁶ Spliceosomal introns are present in varying numbers in all eukaryotic lineages, and they are completely absent from prokaryotic lineages. Some supposedly deep branching lineages, such as the diplomonad *Giardia*, contain many fewer spliceosomal introns, and fewer recognized spliceosomal proteins. However, the reduced complexity of the spliceosome and the paucity of introns not necessarily reflect the ancestral state but in part may reflect the outcome of genome streamlining.⁷ The complete loss of the spliceosome might be counterselected, because it is involved in transsplicing of essential proteins.⁸

Soon after the discovery of introns, theories for the purpose of these intragenic elements were proposed; the most noteworthy being the genes in pieces idea.^{5,9} This idea proposes that introns are present to accelerate protein evolution by allowing for shuffling of protein domains to generate new protein variants. The abundance of introns in eukaryotes implies that eukaryotes may have evolved more complex functions more rapidly than prokaryotes.⁹ This theory was subsequently developed into the introns early theory, which postulates that introns were present in the protein coding genes of the ancestor of the eukaryotes and prokaryotes, and that spliceosomal introns were retained in the eukaryotes and lost in the prokaryotes through genome streamlining.¹⁰⁻¹³ The opposing theory is the introns late theory. After the discovery of group II introns (discussed further in the next section), the intron late hypothesis was formulated which proposed that introns are actually selfish genetic elements that spread into the nuclear genome from the alphaproteobacterial ancestor of the mitochondria to the eukaryotic genome.¹⁴⁻¹⁷ More recent evidence has provided a compelling argument for the introns late hypothesis over the intron early hypothesis,¹⁸ although it appears that the most recent ancestor of all eukaryotes already possessed spliceosomes and spliceosomal introns.

GROUP II INTRONS

Group II introns are self-splicing mobile genetic elements. They are thought to be the predecessors of both the Eukaryotic spliceosomal intron and retrotransposons, and have had a profound role in shaping evolution across the tree of life.¹⁹⁻²¹ Group II introns are found in all domains of life; Bacteria, Archaea and in the mitochondria and chloroplasts of fungi, plants, protists and annelid worms.²² The RNA component of group II introns tends to be about 500 nt long and has six conserved secondary structural domains (DI-DVI). Group II introns are divided into three classes of RNA secondary structures (IIA, IIB and IIC) and different subclasses of structural variation, IIA1, IIA3 IIB1 and IIB2. Open reading frames are often found within the DIV domain encoding between one and four intron

encoded proteins (IEPs); these include proteins with domains for reverse transcriptase (RT), maturase/splicing (X), DNA-binding (D), and endonuclease (En) activity.

Group II introns present in prokaryotes are considered to be self-splicing and mobile. Group II introns present in chloroplasts and mitochondria tend to lack open reading frames; if an ORF is present, it tends to be pseudogene-like, with premature stop codons. In addition, the RNA domains tend to be degenerate which suggests that they are immobile and have lost their self-splicing ability. For example, of about 20 group II introns in the organelles of plants none are known to self-splice.^{23,24} However, to maintain functional genes, the introns must be spliced out. These degenerate self-splicing elements now rely on host encoded splicing factors to facilitate splicing.²⁵⁻²⁷

MOBILITY MECHANISMS FOR GROUP II INTRONS

Similar to spliceosomal introns, group II introns splice through two transesterification reactions that lead to exon ligation and intron excision (figure 1 A). However, for group II introns the reactions are catalyzed by the RNA, which folds into conserved secondary and tertiary structures, creating an active site for Mg²⁺ ions to bind to and catalyze the reactions.²⁸⁻³¹ This ribozyme activity results in the intron forming a lariat loop that is free to move to a new intron less target site. Group II introns are mobile by at least two different self-splicing mechanisms, retrohoming and retrotransposition.³² In these mechanisms the intron and intron encoded proteins come together and function as a ribonucleoprotein (RNP) to integrate into an intron less site. Retrohoming occurs for group II introns that have the En domain among their IEPs.³³⁻³⁵ Retrohoming is a site-specific process that is 100 times more efficient than retrotransposition.³⁵ In the site specific copy and paste mechanism of retrohoming the RNP recognizes specific nucleotide bases or DNA structural features, and the intron reverse splices into the insertion site as a linear intron between the two DNA exons. Then the En domain cleaves the opposite DNA strand a short distance downstream and uses the 3' OH end of the cleaved DNA as a primer to initiate reverse transcription and invoke the DNA repair and/or recombination systems of the host organism. The group II introns that lack an En domain or have a mutated En domain are still mobile but are less successful than those that do. These elements use a retrotransposition mechanism where they reverse splice into the single stranded DNA that is exposed during replication and transcription. They preferentially reverse splice to the lagging stranding where they can use the Okazaki fragments as primers to reverse transcribe into the host genome.

EVOLUTION OF GROUP II INTRONS

Group II introns may have evolved from a retroelement. This hypothesis states that the ancestor of extant group II introns was a bacterial ribozyme containing a RT domain. Phylogenetic trees based on the RNA component and ORF of introns suggest a coevolution of both elements.³⁶⁻³⁸ A more recent study revealed incongruence between the phylogeny of the CL subclasses of proteins and the group II introns they reside in, possibly suggesting these ORFs move between introns.³⁹ One possible mechanism for this is the occurrence of twintrons.⁴⁰ These are found in Archaea and Cyanobacteria, and are created when one

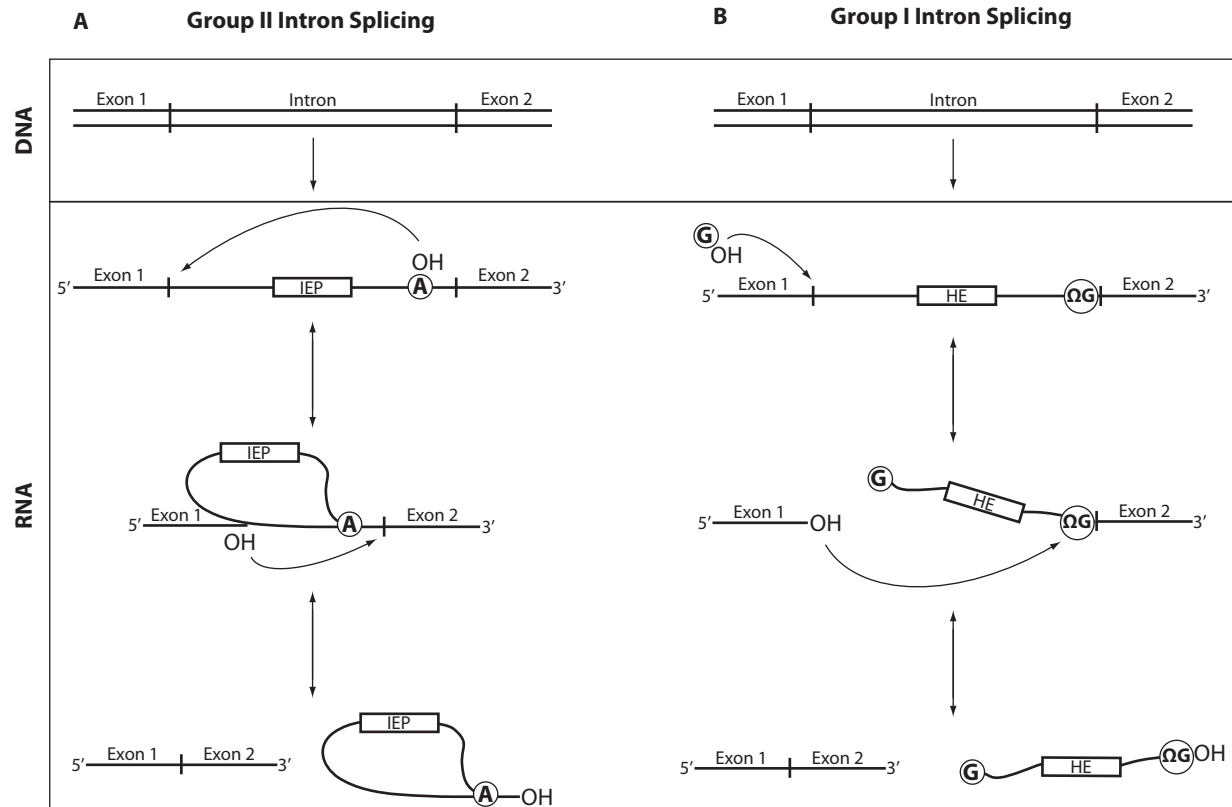


Figure 1. Schematic of the group II and group I intron splicing mechanisms. Panel A depicts the general group II intron splicing mechanism and the two transesterification reactions that occur (adapted from Toro *et al.*).³² The intron encoded protein (IEP) aids in mobility and/or splicing. The final products are the ligated exons and a lariat loop RNA intermediate, which can move to an intronless allele either by retrohoming or retrotransposition. Panel B shows the general group I intron splicing mechanism and the two transesterification reactions that occur. G-OH indicates the guanosine (or GMP, GDP or GTP) bound to the catalytic site of the intron. ΩG indicates the conserved terminal guanine or adenine⁶³ residue at the 3' end of the intron. The open reading frame encoding the homing endonuclease (HE), which aids in group I intron mobility, is depicted in the middle of the intron. Similar to the group II intron, one of the final products are the ligated exons and in contrast to the group II intron the other product is the linear RNA intermediate.

group II intron inserts into another group II intron.⁴¹ The inner intron must be spliced out before the outer intron. One can imagine a scenario where the inner intron loses splicing activity and decays, leaving one of its ORFs behind resulting in an intron with an ORF from a different group II intron class.³⁹

Group II introns are thought to be the ancestors of spliceosomal introns and non-LTR-retrotransposons.¹⁹⁻²¹ Their evolutionary relationship with the spliceosomal intron is

supported by the fact that group II introns and spliceosomal introns have a very similar splicing mechanism, and have comparable boundary sequences and structural similarities.⁴²⁻⁴⁴ It is hypothesized that group II introns evolved in bacteria and were transferred to the eukaryotic genome from the alphaproteobacterial endosymbiont ancestor of the mitochondria and the cyanobacterial endosymbiont ancestor of the plastid, then proliferated in the eukaryotic genome and evolved into the spliceosomal intron. Group II introns have played a role in shaping both eukaryote and prokaryote evolution. A recent study of group II intron proliferation in *Wolbachia*, an extant alphaproteobacteria, whose lifestyle is similar to the endosymbiont ancestor of the mitochondria, showed that mobile genetic elements are associated with genomic rearrangements and gene conversion events.⁴⁵

APPLICATIONS OF GROUP II INTRONS

Many gene disruption studies using derivatives of group II introns, targetrons, have been deployed in organisms such as *Lactococcus lactis*, *E. coli*, *Staphylococcus aureus*, *Francisella*, *Clostridia*.⁴⁶⁻⁴⁹ Targetrons are a site directed mutagenesis system that use the group II intron from the *ltrB* gene of *Lactococcus lactis* adapted to function in clostridial hosts. Availability of this system has greatly facilitated gene inactivation studies in the genus *Clostridium*. This targetron system may also facilitate gene function studies in other organisms for which a genetic system has not yet been developed.

GROUP I INTRONS

Group I introns are a self-splicing introns that splice out after transcription via an autocatalytic two-step transesterification reaction. The first group I intron was identified in *Tetrahymena thermophila* and it was one of the first ribozymes discovered in the early 1980s.⁵⁰ They are small RNAs that range in size from 250 to 500 nucleotides and they are composed of a catalytic RNA domain and often contain a homing endonuclease (HE) encoding ORF.⁵¹⁻⁵³ The HE recognizes a large (14-40) region in an intronless allele and catalyzes a double strand cut. See the section on homing endonucleases below for discussion of target site selection and the HE's role in the life cycle of the self splicing elements. Group I introns are present in the nuclear and organellar genomes of Eukaryotes, in some Eukaryotic viruses, Bacteria, and bacteriophage; to date they have not been identified in Archaea.⁵⁴ Group I introns are found within genes essential to the cell, specifically in 16S and 23S rRNA, tRNAs, ribonucleotide reductase, NADH dehydrogenase, recombinase A, DNA polymerase, thymidylate synthase, and genes involved in photosynthesis.^{53,55}

Interestingly these elements are found more frequently in structural RNAs in bacteria. This could be due the fact that transcription and translation are coupled in bacteria, which might not allow the newly synthesized mRNA sufficient time before translation to form the tertiary structure necessary for splicing.⁵⁶ Group I introns tend to be more abundant in the mitochondria and chloroplasts of eukaryotes than bacteria. The traditional explanation for this was the lack of sexual reproduction among prokaryotes. Sexual reproduction in eukaryotes may bring homologous intron-less and intron-containing alleles together more

frequently, providing more opportunities for rapid intron invasion in a population. However, in light of horizontal gene transfer (HGT), which now appears to be rampant in prokaryotes,⁵⁷ HGT should provide ample opportunities for the introduction of intron containing alleles, making the traditional explanation for the disparity in group I intron abundances unsatisfactory.^{57,58} The co-existence of multiple genomes in the same cell that is found after phage and virus infections of prokaryotes and in mitochondria was also proposed as a factor favoring intron movement.⁵⁸ Another possible explanation for the inequality of intron frequency in eukaryotes and bacteria is that some bacterial group I introns may inhibit growth. For example, when the group I introns of *Tetrahymena* and *Coxiella* are expressed in *E. coli*, the introns were shown to associate with ribosomes and inhibit translation, which ultimately slowed growth.⁵⁹⁻⁶¹

Although group I introns are not well conserved at the sequence level they all share a conserved secondary structure, and use guanosine as a cofactor in splicing.^{62,63} Based on the conserved structure group I introns are classified into 14 subgroups. The secondary structure is labeled by paired elements P1-P10. Ikawa et al.⁶⁴ provides a detailed description of these elements and of their three dimensional arrangement. P3-P7-P9 and P4-P5-P6 form helices that make up the core of the secondary structure. Helix P3-P7-P9 forms the guanosine binding site, which on binding activates the self-splicing reaction.⁶⁴ In contrast to group II introns, the excised intron is generally linear (figure 1B); however, it has been observed to circularize in some instances.⁶⁵

INTEINS

In the late 1980's, the first intein was discovered in the yeast vacuolar ATPases.^{66,67} The first hint at something unusual resulted from comparisons of vacuolar ATPase catalytic subunits from *Neurospora crassa*⁶⁸ and carrot⁶⁹ with a gene in yeast, initially thought to encode a calcium ion transporter.⁷⁰ Hirata et al.⁶⁷ showed that an additional sequence in the yeast homolog was retained in the processed mRNA, but absent from the functioning host protein. Inteins are present in all domains of life; as of April 2nd 2011 the intein database (InBase) lists 523 inteins residing in 65 distinct insertion sites within 36 different proteins.⁷¹ Inteins are found within proteins that are essential to cell function. These elements sit within a protein coding sequence and are translated with the gene it resides in. Through an autocatalytic mechanism the intervening sequence (intein) removes itself from the protein and the flanking protein sequences (exteins) are spliced together resulting in a functional host protein.⁷²⁻⁷⁵ Four conserved motifs are recognized in the general intein structure called either Blocks A, B, F and G or N1, N3, C2 and C1.^{71,76} Recently inteins have been divided into three classes based on different splicing mechanisms.^{77,78} Most inteins are members of class 1, and are considered to follow the standard intein splicing mechanism. Here we will discuss only class 1 inteins (figure 2); for more information on class 2 and 3 inteins see the article by Tori and Perler.⁷⁸ Inteins are also divided into two groups based on their size: large and mini inteins, which are differentiated by the presence or absence of a homing endonuclease domain.^{74,79}

All inteins share a low degree of sequence similarity in their splicing domain, which is formed by the amino and carboxy-terminal parts of the intein sequence. This similarity is

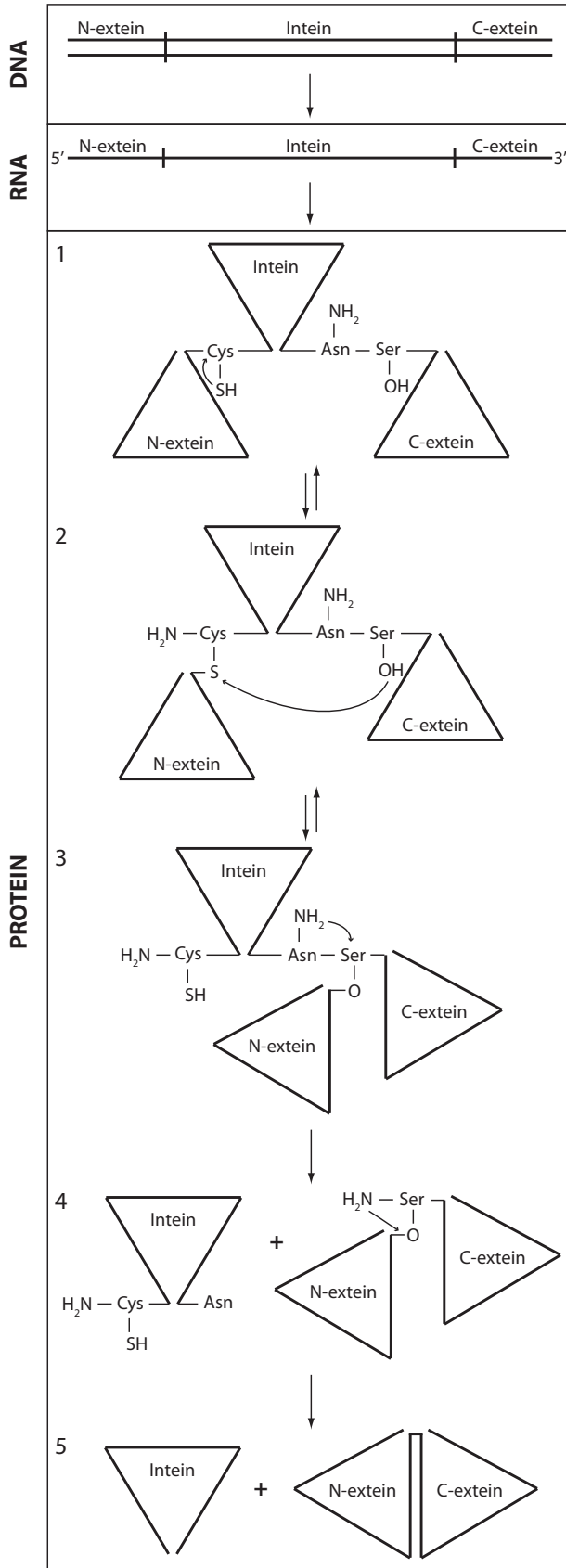


Figure 2. Schematic of the class I intein splicing mechanism adapted from Elleuche and Poggeler.⁷⁹ The four step reaction starts with an N-S acetyl rearrangement of the peptide bond on the N-terminus (1), proceeds to a transesterification between the nucleophile on the C-extein and the thioester on the N-terminus of the intein (2), then to an Asn-cyclization and peptide bond cleavage (3), and finally an N-O acyl shift rearrangement of the ester bond, releasing the intein from the extein and ligating the flanking exteins together to generate the functional host protein (4). The excised intein can move to an inteinless allele and cleave it with the help of a homing endonuclease domain.

sufficient to identify inteins in PSI-blast searches, and suggests a common origin of the inteins splicing domain and an evolutionary relationship with the autocatalytic domain of hedgehog proteins. The latter domain cleaves the hedgehog protein and adds cholesterol to the other domain of the parent proteins.^{75,80,81} Inteins found in the same positions within a protein are more similar and in phylogenetic analyses group together with strong support values; however, too little phylogenetic information is retained in the sequences to reconstruct the relationship between inteins located in different insertion sites, and between inteins and proteins which contain homologous domains.^{75,82,83} The general mechanism for intein splicing is a rapid succession of four nucleophilic attacks (figure 2). 1. An N-S acetyl rearrangement of the peptide bond on the N-terminus of the intein, 2. A transesterification between the nucleophile on the C-extein and the thioester on the N-terminus of the intein, 3. An Asn-cyclization and peptide bond cleavage and 4. An N-O acyl shift rearrangement of the ester bond, releasing the intein from the extein and ligating the flanking exteins together to generate a functional protein.⁷⁸

Usually an intein is encoded as a contiguous piece of DNA that is flanked by the DNA encoding the host protein. The amino and carboxyterminal exteins are transcribed and translated together with the intein sequences. However, in an illustration of a selectively neutral pathway towards higher complexity, the *dnaE* gene in *Synechocystis* sp. PCC6803 encoding DNA polymerase III catalytic subunit, the primary polymerase in DNA replication, is broken into two separate genes, encoded in different parts of the genome.⁸⁴ The amino-terminal extein, together with the amino terminal part of the intein, is transcribed independently from the carboxy-terminal part of the intein, which is transcribed and translated together with the carboxy-terminal extein. The two parts of the intein, produced from separate genes, associate after translation and then catalyze the splicing reaction that joins the two exteins into the functional DNA polymerase subunit. At no point along the evolutionary history, from the first invasion of the host protein by the intein to the gene breaking into two separate units, does it appear that the events were associated with a significant selective advantage or disadvantage for the cyanobacterium. A functional DNA polymerase is produced at all steps along the way; however, initially the intein was superfluous, it could be considered a molecular parasite, and the organism presumably would have survived well without the intein being present; however, in the end the intein has become an indispensable part of the machinery that synthesizes the functioning DNA polymerase. A simple deletion of the intein is no longer possible. While the intein likely has never provided a selective advantage to the host organism, it has become an essential part of a complex machinery that is under strong purifying selection, because without its splicing activity, a functioning DNA polymerase could no longer be produced. It is possible that future mechanisms might evolve that use the complex machinery to better regulate the synthesis of the DNA polymerase; however, the complex system initially evolved without providing an adaptive advantage. The split intein in DnaE has a wide distribution in cyanobacteria,⁸⁵ suggesting this rather complex gene structure is an evolutionarily stable arrangement. Arlin Stoltzfus describes other examples for neutral pathways towards complexity, including spliceosomal introns, which might have evolved along a similar trajectory.⁸⁶ A recent genome survey found several split inteins separated only by free standing homing endonuclease gene.⁸⁷ This arrangement might be an intermediate on the path to split inteins encoded in different parts of the genome, and also might be a precursor

for large inteins that integrate homing endonuclease and splicing domains into a single transcript.⁸⁷

APPLICATIONS OF INTEINS

A wide range of applications in biotechnology were developed based on inteins. These include kits for intein mediated protein purification that are commercially available and use different mechanisms to activate the splicing reaction leading to the release of the pure protein (see ⁷⁹ for a recent review). Split inteins have been used to synthesize toxic proteins from two non-toxic precursors, to synthesize cyclic proteins, and to introduce labels or synthetic peptides into proteins.^{79,88}

INSERTION SITES OF INTRONS AND INTEINS

Understanding where in a protein inteins, group I introns, group II introns and spliceosomal introns tend to accumulate sheds light on how these parasitic genetic elements persist over long periods of time. Comparative studies have surveyed the plethora of genomic data available to attempt to answer this. Studies on intein and group I intron insertion sites have shown they are found in conserved sites, while studies on group II and spliceosomal introns did not reveal a significant conserved site preference.^{55,75,89-93} This disparity between inteins, group I introns and group II introns can be explained by differences in mobility between the elements. Inteins and group I introns are generally mobile by homing endonucleases, these recognize large sequences that occur infrequently in the genome. Targeting conserved sites in conserved proteins guarantees that excision happens and is exact. Mutations in the insertion site that would allow a host protein to become resistant against the homing endonuclease, would likely result in loss of function of the host protein, and inaccurate splicing would result in small insertions or deletions in the most conserved region of the host protein. Most conserved proteins provide a vital function to the cell so exact excision is required to maintain a functional protein, thus strong purifying selection acts on the intein to maintain accurate and efficient splicing activity. Targeting a conserved motif that occurs in different protein families might also aid the elements move to a new target site. Another reason for targeting conserved proteins is that protein will be present in a more distantly related organism, which may allow transfer and persistence in a new divergent population. The more organisms a specific group I intron is found in, the more conserved the insertion site is.⁵⁵ Although most group II introns are mobile via an endonuclease, it has a much smaller recognition site which occurs more frequently in a genome. It appears that group II introns evolved a different strategy that relies on frequent propagation to outpace loss.⁵⁵ An exemplary comparison of target site conservation of inteins, and group I and group II introns is depicted in figure 3. Both inteins and group I introns are found inserted in the most conserved region of the ribonucleotide reductase, whereas the group II introns do not show such a preference. See ⁵⁵ for further examples and a statistical evaluation of target site conservation.

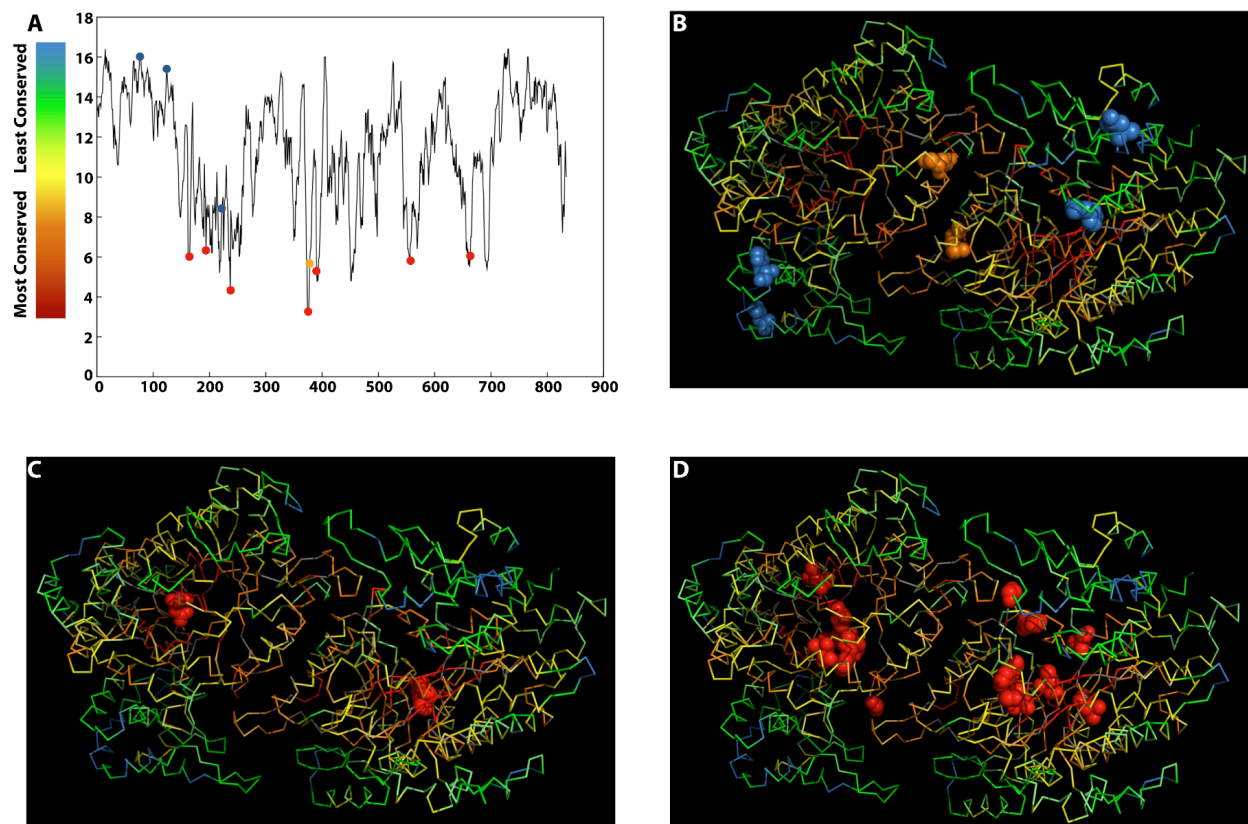


Figure 3. Positions of inteins, group I introns and group II introns along the protein sequence and in the structure of ribonucleotide reductases. Panel A shows the conservation profile of class I and class II ribonucleotide reductases. The x-axis in the amino acid positions along a multiple sequence alignment of ribonucleotide reductases and the y-axis is the conservation level in a five amino acid window centered around the indicated position. The lower the conservation score the more conserved a position is. The profile was calculated as described in⁵⁵ except the window size is five. The positions of inteins are shown as dark gray dots (or red in the colored image), the significance level for conserved site preference of inteins in the ribonucleotide reductase is $p < 0.0001$; group I introns are shown as medium gray dots (or orange in the colored image), the significance level for conserved site preference of group I introns is $p = 0.0256$, and group II introns are shown in light gray (or blue in the colored image), with a significance level for conserved site preference of $p = 0.7407$. I.e., the null hypothesis of insertion independent of site conservation is rejected for inteins and group I introns, but not for group II introns. The ribonucleotide reductase hosts seven inteins, one group I intron and three group II introns. Panel B – Panel D shows the crystal structure of a dimer of the human ribonucleotide reductase R1 subunit (RNR) (PDB ID: 2WGH).¹¹³ Each is colored according to sequence conservation. Panel B shows the crystal structure with the three group II introns insertion sites mapped to it as spheres. Panel C shows the RNR structure with the one group I intron insertion site mapped on to it as a sphere. Panel D shows the RNR crystal structure with seven intein insertion sites mapped to it as spheres. The group I intron and intein insertion sites map to conserved sites while the Group II intron do not.

HOMING ENDONUCLEASES

Homing endonucleases (HE) are a class of site-specific endonucleases that have a large recognition site (14-40 bp) and provide mobility to group I introns and inteins.⁹⁴⁻⁹⁶ They provide mobility to these genetic elements by targeting and cleaving intronless or inteinless alleles. The double stranded cleavage invokes the hosts DNA repair mechanisms, which uses the intron or intein containing allele as a template and copies the HE containing element into the newly cleaved target site.

How the homing endonucleases became associated with the splicing elements is an intriguing question. Among phages freestanding HEs are common along with HE-less group I introns. The free standing HEs are thought to provide a competitive advantage for the phage when it comes in contact with competing phage, as the HE will cleave the foreign phage genome.⁹⁷ The recent sequencing of a large number of phages provides a snap shot of how HEs could have possibly associated with introns and inteins. In cyanophages a free standing HE, F-CpH1, is encoded downstream of the *psbA* gene and within the *psbA* gene there is a HE free group I intron.⁹⁸ In related phages it was shown that the free standing HE can cleave the intronless intron insertion site. In T7 and T3 phages free standing HEs are able to cleave a gene encoding a DNA polymerase. Among T-phages there are some that have a free standing HE and the same HE sitting within a group I intron. Both of these elements were shown to cleave the intron insertion site.⁹⁹ Another study of split inteins and HEs in the global ocean survey (GOS) metagenomic data showed a possible progression of an HE becoming associated with inteins.⁸⁷ This progression from free standing HE to a nested HE provides a picture of how a HE could have become associated with its autocatalytic splicing element and suggests phage competition as a likely driver for the HE associating with its splicing element.

While homing endonucleases have large recognition sites, they do not require a 100% match to catalyze a double strand cut. HEs tolerate different nucleotides especially in positions that allow for synonymous substitutions of the host protein, i.e. mutations in a protein coding gene that do not change the encoded protein because of the redundancy of the genetic code.¹⁰⁰⁻¹⁰² This low precision of the HEs makes it more difficult for the host to evolve immunity to the HE through changes in the target site sequence. However, the low precision of the HEs also means that double strand cuts may also be created also in positions that do not correspond to the target site. This activity may result in genome rearrangements,¹⁰³ which is the process that has become the main function of the mating type switching HO endonuclease in yeast. This endonuclease evolved from a large intein whose descendants also include the intein in the yeast vacuolar ATPase catalytic subunit.^{75,104}

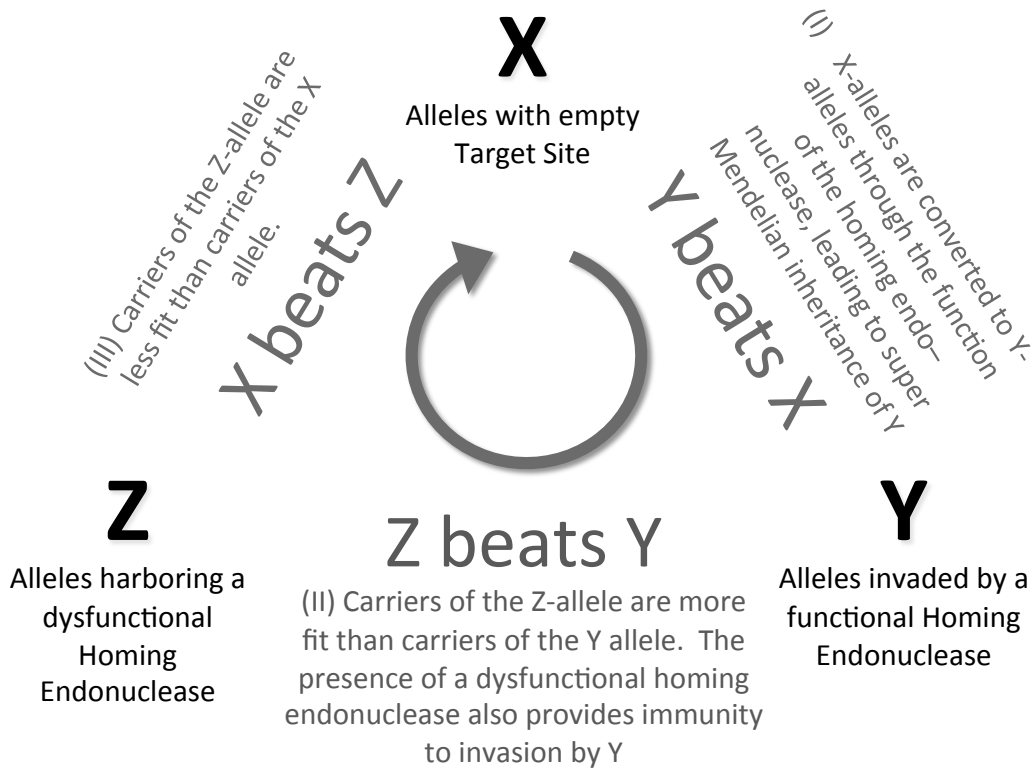


Figure 4: A molecular rock-paper-scissors game: The diagram depicts the pairwise relationships between carriers of the X allele, which does not contain a self-splicing element (intein or group I intron) nor a homing endonuclease (HE); the Y allele, containing a self-splicing element encoding a functioning HE; and the Z allele, which encodes a defective HE. The relative pairwise fitness differences are depicted by three inequalities: (I) Y beats X because the HE activity during sex or following gene transfer converts homing endonuclease free alleles into those that contain the molecular parasite; (II) Z beats Y because the defective HE produces less double strand breaks; the presence of the molecular parasite in the HE target site also provides immunity to the HE; (III) X beats Z, because its carriers have to expand less energy on replicating the molecular parasite and on synthesizing the defective HE. The three alleles also depict the main serial succession (central circular arrow) of the state of a particular homing endonuclease target site: The empty target site (X) is converted to Y through invasion by the homing endonuclease, the homing endonuclease in the Y alleles decays through mutation into the Z allele with dysfunctional homing endonuclease, and finally a precise deletion of the intein/intron restores the empty target site, i.e. Y converts to X. If each allele is successively fixed in the population, the central arrow also describes the homing cycle;^{75,105} however, the different alleles may also coexist in a population for long periods of time.^{96,108}

THE HOMING CYCLE

HEs are frequently found in inteins and group I introns, and provide a means for mobility for these parasitic genetic elements. The genetic elements using HEs go through cycles of invasion, decay and loss (compare figure 4). This cycle, termed the homing cycle, was first proposed by Goddard and Burt to explain the cyclic nature of group I intron gain and loss. This later was expanded to include inteins.^{75,96,105} The activity of the HE and its maintenance through purifying selection depend on the availability of unoccupied target sites.

The general model starts with the invasion of a population by a HE containing mobile element that is transferred from another gene, species, population or subpopulation. When an allele with HE and an allele that does not contain an intron/intein in the target site are present in the same cell as a consequence of sex or gene transfer, the HE has the opportunity to cleave the empty target site and trigger the conversion of the empty target site into an allele with inserted HE. The resulting super Mendelian inheritance leads to an increase in frequency of the HE containing allele throughout the invaded population. However, once all target sites are occupied the parasitic genetic element has no where else to go within the local population two possible outcomes may occur: 1) It acquires a beneficial function for the host and remains under purifying selection or 2) The homing endonuclease begins to lose function through random mutations in the absence of purifying selection then precise loss of the intron or intein occurs. If the latter occurs the cycle of invasion by an allele with a functioning HE can begin again.

The relationships between the alleles with empty target site, alleles that harbor a functioning HE, and alleles with a nonfunctioning HE form a non-transitive completion network (figure 4). For each of the three alleles one of the alleles outcompetes one of the other alleles, but loses out against the third allele. The progression of an individual target site is indicated by the central arrow. If each of the pairwise interactions leads to fixation of the winning allele, the traditional homing cycle results. However, in large populations the three types of alleles may coexist, and different subpopulations may be dominated by different alleles that propagate in the larger population similar to waves in a fluid medium, without reaching fixation in the population.⁹⁶ In a large, well mixed populations the frequencies of the three alleles may follow a trajectory where higher frequency waves for each of the alleles follow one another, oscillating in state space around an equilibrium, similar to a Lotka¹⁰⁶-Volterra¹⁰⁷ predator-prey model.¹⁰⁸ With the amount of metagenomic data that is being accumulated it seems possible that a quantitative and realistic model for the relations between the different alleles can be developed. One outcome of such a model would be a better understanding of gene flow within and between microbial populations.

ACQUISITION OF NEW FUNCTIONS

One way a HE and mobile genetic elements can avoid elimination and escape the homing cycle is to develop a function advantageous for the host cell. The catalysis of genome rearrangements has already been discussed above. Another example is the bacterial intein-like (BIL) domains, which are thought to be remnants of the HINT domain of inteins. They

are found in adhesin proteins and are thought to create protein diversity.^{109,110} Another example for an escape from the homing cycle is a *Naegleria* group I intron whose homing endonuclease has acquired a maturase function.¹¹¹

CONCLUSIONS/OUTLOOK

Many questions on the evolution and function of molecular parasites remain unanswered. For inteins the halophilic archaea have been proposed as a model system to further understand intein propagation and the effects inteins may have on their host.¹¹² Utilizing the vast amount of environmental metagenomic data has provided novel insights on how homing endonucleases might have become associated with the self splicing elements.⁸⁷ Metagenomic data, deep sequencing projects, and better experimental systems promise a quantitative understanding of the host parasite interactions that govern the life cycle of HEs. Because gene transfer is a crucial step in this life cycle, these studies will also provide information on gene flow within and between populations.

ACKNOWLEDGEMENT

This work was supported through the NASA Exobiology (NNX08AQ10G) and the National Science Foundation Assembling the Tree of Life program (DEB 0830024).

References

1. Mozer TJ, Thompson RB, Berget SM, Warner HR. Isolation and characterization of a bacteriophage T5 mutant deficient in deoxynucleoside 5'-monophosphatase activity. *J Virol.* Nov 1977;24(2):642-650.
2. Chow LT, Gelinis RE, Broker TR, Roberts RJ. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell.* Sep 1977;12(1):1-8.
3. Evans RM, Fraser N, Ziff E, Weber J, Wilson M, Darnell JE. The initiation sites for RNA transcription in Ad2 DNA. *Cell.* Nov 1977;12(3):733-739.
4. Goldberg S, Schwartz H, Darnell JE, Jr. Evidence from UV transcription mapping in HeLa cells that heterogeneous nuclear RNA is the messenger RNA precursor. *Proc Natl Acad Sci U S A.* Oct 1977;74(10):4520-4523.
5. Gilbert W. Why genes in pieces? *Nature.* Feb 9 1978;271(5645):501.
6. Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. *Nature.* Sep 12 2002;419(6903):182-185.
7. Chen XS, White WT, Collins LJ, Penny D. Computational identification of four spliceosomal snRNAs from the deep-branching eukaryote *Giardia intestinalis*. *PLoS One.* 2008;3(8):e3106.
8. Kamikawa R, Inagaki Y, Tokoro M, Roger AJ, Hashimoto T. Split introns in the genome of *Giardia intestinalis* are excised by spliceosome-mediated trans-splicing. *Curr Biol.* Feb 22 2011;21(4):311-315.

9. Doolittle WF. Genes in pieces: were they ever together? *Nature*. April 13 1978;272(5646):581-582.
10. Roger AJ, Doolittle WF. Molecular evolution. Why introns-in-pieces? *Nature*. Jul 22 1993;364(6435):289-290.
11. Doolittle WF, Stoltzfus A. Molecular evolution. Genes-in-pieces revisited. *Nature*. Feb 4 1993;361(6411):403.
12. Gilbert W, Glynias M. On the ancient nature of introns. *Gene*. Dec 15 1993;135(1-2):137-144.
13. Jeffares DC, Mourier T, Penny D. The biology of intron gain and loss. *Trends Genet*. Jan 2006;22(1):16-22.
14. Cavalier-Smith T. Intron phylogeny: a new hypothesis. *Trends Genet*. May 1991;7(5):145-148.
15. Stoltzfus A, Spencer DF, Zuker M, Logsdon JM, Jr., Doolittle WF. Testing the exon theory of genes: the evidence from protein structure. *Science*. Jul 8 1994;265(5169):202-207.
16. Logsdon JM, Jr. The recent origins of spliceosomal introns revisited. *Curr Opin Genet Dev*. Dec 1998;8(6):637-648.
17. Logsdon JM, Jr., Tyshenko MG, Dixon C, J DJ, Walker VK, Palmer JD. Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc Natl Acad Sci U S A*. Aug 29 1995;92(18):8507-8511.
18. Koonin EV. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct*. 2006;1:22.
19. Cech TR. The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell*. Jan 31 1986;44(2):207-210.
20. Sharp PA. On the origin of RNA splicing and introns. *Cell*. Sep 1985;42(2):397-400.
21. Zimmerly S, Guo H, Perlman PS, Lambowitz AM. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*. Aug 25 1995;82(4):545-554.
22. Lambowitz AM, Zimmerly S. Mobile group II introns. *Annu Rev Genet*. 2004;38:1-35.
23. Watkins KP, Rojas M, Friso G, van Wijk KJ, Meurer J, Barkan A. APO1 Promotes the Splicing of Chloroplast Group II Introns and Harbors a Plant-Specific Zinc-Dependent RNA Binding Domain. *Plant Cell*. Mar 18 2011.
24. Kroeger TS, Watkins KP, Friso G, van Wijk KJ, Barkan A. A plant-specific RNA-binding domain revealed through analysis of chloroplast group II intron splicing. *Proc Natl Acad Sci U S A*. Mar 17 2009;106(11):4537-4542.
25. Hausner G, Olson R, Simon D, et al. Origin and evolution of the chloroplast trnK (matK) intron: a model for evolution of group II intron RNA structures. *Mol Biol Evol*. Feb 2006;23(2):380-391.
26. Turmel M, Otis C, Lemieux C. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol Biol Evol*. Jun 2006;23(6):1324-1338.
27. Stern DB, Goldschmidt-Clermont M, Hanson MR. Chloroplast RNA metabolism. *Annu Rev Plant Biol*. Jun 2 2010;61:125-155.
28. Chanfreau G, Jacquier A. Catalytic site components common to both splicing steps of a group II intron. *Science*. Nov 25 1994;266(5189):1383-1387.
29. Sigel RK, Vaidya A, Pyle AM. Metal ion binding sites in a group II intron core. *Nat Struct Biol*. Dec 2000;7(12):1111-1116.

30. Gordon PM, Piccirilli JA. Metal ion coordination by the AGC triad in domain 5 contributes to group II intron catalysis. *Nat Struct Biol.* Oct 2001;8(10):893-898.
31. Gordon PM, Fong R, Piccirilli JA. A second divalent metal ion in the group II intron reaction center. *Chem Biol.* Jun 2007;14(6):607-612.
32. Toro N, Jimenez-Zurdo JI, Garcia-Rodriguez FM. Bacterial group II introns: not just splicing. *FEMS Microbiol Rev.* Apr 2007;31(3):342-358.
33. Huang T, Shaikh TR, Gupta K, et al. The group II intron ribonucleoprotein precursor is a large, loosely packed structure. *Nucleic Acids Res.* Dec 3 2010.
34. Cousineau B, Smith D, Lawrence-Cavanagh S, et al. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell.* Aug 21 1998;94(4):451-462.
35. Coros CJ, Piazza CL, Chalamcharla VR, Smith D, Belfort M. Global regulators orchestrate group II intron retromobility. *Mol Cell.* Apr 24 2009;34(2):250-256.
36. Fontaine JM, Goux D, Kloareg B, Loiseaux-de Goer S. The reverse-transcriptase-like proteins encoded by group II introns in the mitochondrial genome of the brown alga *Pylaiella littoralis* belong to two different lineages which apparently coevolved with the group II ribosome lineages. *J Mol Evol.* Jan 1997;44(1):33-42.
37. Toor N, Hausner G, Zimmerly S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA.* Aug 2001;7(8):1142-1152.
38. Simon DM, Clarke NA, McNeil BA, et al. Group II introns in eubacteria and archaea: ORF-less introns and new varieties. *RNA.* Sep 2008;14(9):1704-1713.
39. Simon DM, Kelchner SA, Zimmerly S. A broadscale phylogenetic analysis of group II intron RNAs and intron-encoded reverse transcriptases. *Mol Biol Evol.* Dec 2009;26(12):2795-2808.
40. Dai L, Zimmerly S. ORF-less and reverse-transcriptase-encoding group II introns in archaeobacteria, with a pattern of homing into related group II intron ORFs. *RNA.* Jan 2003;9(1):14-19.
41. Drager RG, Hallick RB. A complex twintron is excised as four individual introns. *Nucleic Acids Res.* May 25 1993;21(10):2389-2394.
42. Keating KS, Toor N, Perlman PS, Pyle AM. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA.* Jan 2010;16(1):1-9.
43. Madhani HD, Guthrie C. A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell.* Nov 27 1992;71(5):803-817.
44. Shukla GC, Padgett RA. A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol Cell.* May 2002;9(5):1145-1150.
45. Leclercq S, Giraud I, Cordaux R. Remarkable abundance and evolution of mobile group II introns in *Wolbachia* bacterial endosymbionts. *Mol Biol Evol.* Jan 2011;28(1):685-697.
46. Chen Y, McClane BA, Fisher DJ, Rood JI, Gupta P. Construction of an alpha toxin gene knockout mutant of *Clostridium perfringens* type A by use of a mobile group II intron. *Appl Environ Microbiol.* Nov 2005;71(11):7542-7547.
47. Rodriguez SA, Yu JJ, Davis G, Arulanandam BP, Klose KE. Targeted inactivation of *francisella tularensis* genes by group II introns. *Appl Environ Microbiol.* May 2008;74(9):2619-2626.

48. Yao J, Zhong J, Fang Y, Geisinger E, Novick RP, Lambowitz AM. Use of targetrons to disrupt essential and nonessential genes in *Staphylococcus aureus* reveals temperature sensitivity of Ll.LtrB group II intron splicing. *RNA*. Jul 2006;12(7):1271-1281.
49. Yao J, Zhong J, Lambowitz AM. Gene targeting using randomly inserted group II introns (targetrons) recovered from an *Escherichia coli* gene disruption library. *Nucleic Acids Res*. 2005;33(10):3351-3362.
50. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*. Nov 1982;31(1):147-157.
51. Michel F, Westhof E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol*. Dec 5 1990;216(3):585-610.
52. Suh SO, Jones KG, Blackwell M. 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*. May 1999;48(5):493-500.
53. Nielsen H, Johansen SD. Group I introns: Moving in new directions. *RNA Biol*. Sep-Oct 2009;6(4):375-383.
54. Haugen P, Simon DM, Bhattacharya D. The natural history of group I introns. *Trends Genet*. Feb 2005;21(2):111-119.
55. Swithers KS, Senejani AG, Fournier GP, Gogarten JP. Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol Biol*. 2009;9:303.
56. Ohman-Heden M, Ahgren-Stalhandske A, Hahne S, Sjoberg BM. Translation across the 5'-splice site interferes with autocatalytic splicing. *Mol Microbiol*. Mar 1993;7(6):975-982.
57. Swithers KS, Gogarten JP, Fournier GP. Trees in the web of life. *J Biol*. 2009;8(6):54.
58. Edgell DR, Belfort M, Shub DA. Barriers to intron promiscuity in bacteria. *J Bacteriol*. Oct 2000;182(19):5281-5289.
59. Nikolcheva T, Woodson SA. Association of a group I intron with its splice junction in 50S ribosomes: implications for intron toxicity. *RNA*. Sep 1997;3(9):1016-1027.
60. Raghavan R, Hicks LD, Minnick MF. Toxic introns and parasitic intein in *Coxiella burnetii*: legacies of a promiscuous past. *J Bacteriol*. Sep 2008;190(17):5934-5943.
61. Raghavan R, Minnick MF. Group I introns and inteins: disparate origins but convergent parasitic strategies. *J Bacteriol*. Oct 2009;191(20):6193-6202.
62. Vicens Q, Cech TR. Atomic level architecture of group I introns revealed. *Trends Biochem Sci*. Jan 2006;31(1):41-51.
63. Raghavan R, Hicks LD, Minnick MF. A unique group I intron in *Coxiella burnetii* is a natural splice mutant. *J Bacteriol*. Jun 2009;191(12):4044-4046.
64. Ikawa Y, Shiraishi H, Inoue T. Minimal catalytic domain of a group I self-splicing intron RNA. *Nat Struct Biol*. Nov 2000;7(11):1032-1035.
65. Nielsen H, Fiskaa T, Birgisdottir AB, Haugen P, Einvik C, Johansen S. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA*. Dec 2003;9(12):1464-1475.

66. Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science*. Nov 2 1990;250(4981):651-657.
67. Hirata R, Ohsumik Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem*. Apr 25 1990;265(12):6726-6733.
68. Bowman EJ, Tenney K, Bowman BJ. Isolation of genes encoding the *Neurospora* vacuolar ATPase. Analysis of *vma-1* encoding the 67-kDa subunit reveals homology to other ATPases. *J Biol Chem*. Oct 5 1988;263(28):13994-14001.
69. Zimniak L, Dittrich P, Gogarten JP, Kibak H, Taiz L. The cDNA sequence of the 69-kDa subunit of the carrot vacuolar H⁺-ATPase. Homology to the beta-chain of F₀F₁-ATPases. *J Biol Chem*. 1988;263(19):9102-9112.
70. Shih CK, Wagner R, Feinstein S, Kanik-Ennulat C, Neff N. A dominant trifluoperazine resistance gene from *Saccharomyces cerevisiae* has homology with F₀F₁ ATP synthase and confers calcium-sensitive growth. *Mol Cell Biol*. 1988;8(8):3094-3103.
71. Perler FB. InBase: the Intein Database. *Nucleic Acids Res*. Jan 1 2002;30(1):383-384.
72. Pietrokovski S. Intein spread and extinction in evolution. *Trends Genet*. Aug 2001;17(8):465-472.
73. Perler FB, Davis EO, Dean GE, et al. Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature. *Nucleic Acids Res*. Apr 11 1994;22(7):1125-1127.
74. Liu XQ. Protein-splicing intein: Genetic mobility, origin, and evolution. *Annu Rev Genet*. 2000;34:61-76.
75. Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E. Inteins: structure, function, and evolution. *Annu Rev Microbiol*. 2002;56:263-287.
76. Pietrokovski S. Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci*. Jan 1998;7(1):64-71.
77. Tori K, Dassa B, Johnson MA, et al. Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *The Journal of biological chemistry*. Jan 22 2010;285(4):2515-2526.
78. Tori K, Perler FB. Expanding the Definition of Class 3 Inteins and Their Proposed Phage Origin. *J Bacteriol*. Feb 11 2011.
79. Elleuche S, Poggeler S. Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl Microbiol Biotechnol*. Jun 2010;87(2):479-489.
80. Perler FB, Olsen GJ, Adam E. Compilation and analysis of intein sequences. *Nucleic Acids Res*. 1997;25(6):1087-1093.
81. Burglin TR. The Hedgehog protein family. *Genome Biol*. 2008;9(11):241.
82. Koufopanou V, Goddard MR, Burt A. Adaptation for horizontal transfer in a homing endonuclease. *Mol Biol Evol*. Mar 2002;19(3):239-246.
83. Senejani AG, Hilario E, Gogarten JP. The intein of the *Thermoplasma* A-ATPase A subunit: structure, evolution and expression in *E. coli*. *BMC Biochem*. 2001;2:13.
84. Wu H, Hu Z, Liu XQ. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A*. 1998;95(16):9226-9231.

85. Caspi J, Amitai G, Belenkiy O, Pietrokovski S. Distribution of split DnaE inteins in cyanobacteria. *Mol Microbiol*. Dec 2003;50(5):1569-1577.
86. Stoltzfus A. On the possibility of constructive neutral evolution. *J Mol Evol*. Aug 1999;49(2):169-181.
87. Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrokovski S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. *Nucleic Acids Res*. May 2009;37(8):2560-2573.
88. Aranko AS, Zuger S, Buchinger E, Iwai H. In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS One*. 2009;4(4):e5185.
89. Goodwin TJ, Butler MI, Poulter RT. Multiple, non-allelic, intein-coding sequences in eukaryotic RNA polymerase genes. *BMC Biol*. 2006;4:38.
90. Lazarevic V. Ribonucleotide reductase genes of Bacillus prophages: a refuge to introns and intein coding sequences. *Nucleic Acids Res*. Aug 1 2001;29(15):3212-3218.
91. Dai L, Zimmerly S. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res*. Mar 1 2002;30(5):1091-1102.
92. Toro N. Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. *Environ Microbiol*. Mar 2003;5(3):143-151.
93. Sandegren L, Sjoberg BM. Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. *J Biol Chem*. May 21 2004;279(21):22218-22227.
94. Chevalier BS, Stoddard BL. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res*. Sep 15 2001;29(18):3757-3774.
95. Dujon B. Group I introns as mobile genetic elements: facts and mechanistic speculations--a review. *Gene*. Oct 15 1989;82(1):91-114.
96. Gogarten JP, Hilario E. Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol*. 2006;6:94.
97. Goodrich-Blair H, Shub DA. Beyond homing: competition between intron endonucleases confers a selective advantage on flanking genetic markers. *Cell*. Jan 26 1996;84(2):211-221.
98. Zeng Q, Bonocora RP, Shub DA. A free-standing homing endonuclease targets an intron insertion site in the psbA gene of cyanophages. *Curr Biol*. Feb 10 2009;19(3):218-222.
99. Bonocora RP, Shub DA. A likely pathway for formation of mobile group I introns. *Curr Biol*. Feb 10 2009;19(3):223-228.
100. Gimble FS. Degeneration of a homing endonuclease and its target sequence in a wild yeast strain. *Nucleic Acids Res*. Oct 15 2001;29(20):4215-4223.
101. Scalley-Kim M, McConnell-Smith A, Stoddard BL. Coevolution of a homing endonuclease and its host target sequence. *J Mol Biol*. Oct 5 2007;372(5):1305-1319.
102. Kurokawa S, Bessho Y, Higashijima K, et al. Adaptation of intronic homing endonuclease for successful horizontal transmission. *FEBS J*. May 2005;272(10):2487-2496.

103. Nishioka M, Fujiwara S, Takagi M, Imanaka T. Characterization of two intein homing endonucleases encoded in the DNA polymerase gene of *Pyrococcus kodakaraensis* strain KOD1. *Nucleic Acids Res.* 1998;26(19):4409-4412.
104. Bakhrat A, Baranes K, Krichevsky O, et al. Nuclear import of ho endonuclease utilizes two nuclear localization signals and four importins of the ribosomal import system. *J Biol Chem.* May 5 2006;281(18):12218-12226.
105. Goddard MR, Burt A. Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A.* Nov 23 1999;96(24):13880-13885.
106. Lotka AJ. Analytical Note on Certain Rhythmic Relations in Organic Systems. *Proc Natl Acad Sci U S A.* 1920;6:410-415.
107. Volterra V. Variations and fluctuations of the number of individuals in animal species living together in *Animal Ecology*. In: Chapman RN, ed. *Animal Ecology*. New York: McGraw-Hill; 1931.
108. Yahara K, Fukuyo M, Sasaki A, Kobayashi I. Evolutionary maintenance of selfish homing endonuclease genes in the absence of horizontal transfer. *Proc Natl Acad Sci U S A.* Nov 3 2009;106(44):18861-18866.
109. Dori-Bachash M, Dassa B, Peleg O, Pineiro SA, Jurkevitch E, Pietrokovski S. Bacterial intein-like domains of predatory bacteria: a new domain type characterized in *Bdellovibrio bacteriovorus*. *Funct Integr Genomics.* May 2009;9(2):153-166.
110. Amitai G, Belenkiy O, Dassa B, Shainskaya A, Pietrokovski S. Distribution and function of new bacterial intein-like protein domains. *Mol Microbiol.* Jan 2003;47(1):61-73.
111. Wikmark OG, Einvik C, De Jonckheere JF, Johansen SD. Short-term sequence evolution and vertical inheritance of the *Naegleria* twin-ribozyme group I intron. *BMC Evol Biol.* 2006;6:39.
112. Barzel A, Naor A, Privman E, Kupiec M, Gophna U. Homing endonucleases residing within inteins: evolutionary puzzles awaiting genetic solutions. *Biochem Soc Trans.* Jan 19 2011;39(1):169-173.
113. Fairman JW, Wijerathna SR, Ahmad MF, et al. Structural basis for allosteric regulation of human ribonucleotide reductase by nucleotide-induced oligomerization. *Nat Struct Mol Biol.* Mar 2011;18(3):316-322.