

Inteins as indicators of bio-communication

Shannon M. Soucy ^a, J. Peter Gogarten ^{b,d}

^a Dartmouth College, 78 College St, Hanover, NH 03755

Shannon.Soucy@dartmouth.edu

^b University of Connecticut, 91 North Eagleville Rd, Storrs, CT 06250

Gogarten@uconn.edu

^d Communicating author

Abstract

Genetic information is stored in DNA molecules, horizontal gene transfer shares this information between organisms. The transfer of genetic information thus is an important method of bio-communication. Communities of organisms made up of either the same species, or different species can transfer genetic material between community members. In Bacteria and Archaea these transfers are unidirectional, except for mating described for Haloarchaea (Naor et al. 2012) and possibly also occurring in other archaea. Inteins are selfish genetic elements found in all domains of life. As is the case for most mobile selfish genetic elements, horizontal gene transfer is an important part of the intein's life cycle. Several unique properties of inteins make them particularly well suited as markers of horizontal gene transfer pathways. In this chapter, we explore these properties, and provide examples of their uses to reveal patterns of horizontal gene transfer.

DOI of published version:

https://doi.org/10.1007/978-3-319-65536-9_16

1. Introduction to inteins

Inteins are selfish genetic elements that are found in all three domains of life and also in viruses (Pietrokovski 2001; Gogarten et al. 2002; Perler 2002; Swithers et al. 2009; Soucy et al. 2014). Inteins often are found in multiple sites within a single gene. The invaded gene that encodes the host protein is called an extein (Fig. 1a). Inteins are named after the gene they invade, followed by a letter indicating the insertion site. For example, *Hvo polB-c* denotes the intein located in insertion site c of the gene in *Haloferax volcanii* encoding DNA polymerase B. Each intein insertion site is referred to as an intein allele, though this term is misleading, as phylogenetic analysis of all inteins in the Halobacteria (colloquially referred to as the Haloarchaea) shows strong support for an independent origin of each intein allele (Soucy et al. 2014); *i.e.*, inteins inserted into a same site of the gene are much more similar to one another than to inteins occupying different insertion sites in the same gene. The letters denoting the intein allele are given in order of the discovery of the intein insertion site, and not necessarily in the order they are found in the gene. For example, the ribonucleotide reductase gene has 5 insertion sites in the Haloarchaea (Soucy et al. 2014); the order they appear in the gene is: *rir-1l*, *rir-1k*, *rir-1b*, *rir-1g*, and *rir-1m*. Inteins inserted into the other sites (e.g., *rir-1a*) so far have not been described as occupied in Haloarchaea.

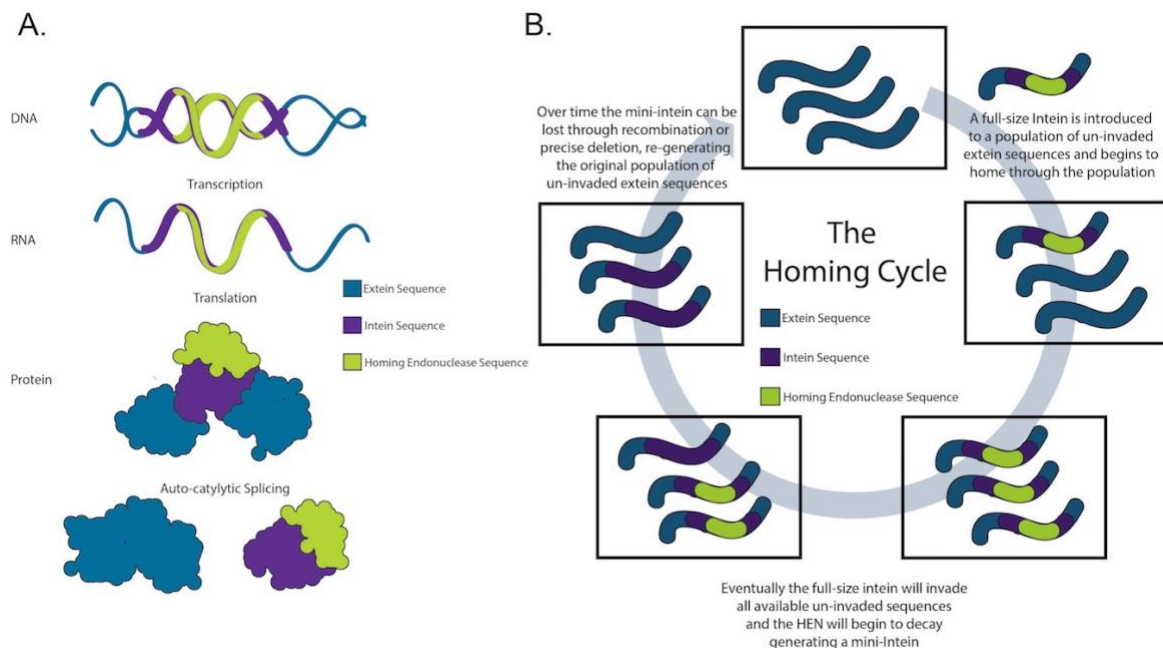


Figure 1: Overview of intein structure and function (Panel A) and the homing cycle (A)

1.1 Intein distribution and patterns in extein functions

Inteins are found in the most conserved parts of the most conserved proteins (Swithers et al. 2009), and most often in genes related to replication, recombination, and repair (COG category L) and to a lesser extent in other genes involved in nucleotide metabolism (COG category F) (Novikova et al. 2015). Many of the invaded genes

encode proteins that possess nucleotide binding sites (Gogarten et al. 2002). In Archaea, inteins were mainly found in the Euryarchaeota (68% of archaeal inteins) and Crenarchaeota (34% of archaeal inteins) (Novikova et al. 2015). At present debate continues regarding the question, if the observed intein distribution directly reflects the function of the genes, or only the between species conservation of the invaded genes. Different processes have been described to explain the observed distribution of inteins:

1. All sites in all genes are targeted by inteins; however, only genes and insertions that do not tolerate substitutions or small deletions without affecting important cellular functions (replication, recombination, repair, ATP synthesis and nucleotide metabolism) allow the inteins to persist over long periods of time (Gogarten et al. 2002; Swithers et al. 2009). In less important sites, small deletions may inactivate and remove the intein.
2. Inteins have evolved to target conserved sites in conserved genes because these sites are likely to have a similar sequence in an orthologous gene even in other, distantly related species (Swithers et al. 2009). This allows inteins to invade homologous genes present in other species, even across phyla and domain boundaries (Swithers et al. 2013).
3. The splicing activity of some inteins was found to be sensitive to redox and other conditions in the cell, and it was recently proposed that inteins could respond to the altered redox state during cellular stress, effectively pausing important cellular functions until conditions return to normal (Topilina et al. 2015; Novikova et al. 2015); however, the often limited within species distribution of inteins (Soucy et al. 2014; Naor et al. 2016) argues against inteins providing a strong selective advantage.

More work needs to be done to investigate the extent to which these three possible processes gave rise to the observed distribution of inteins.

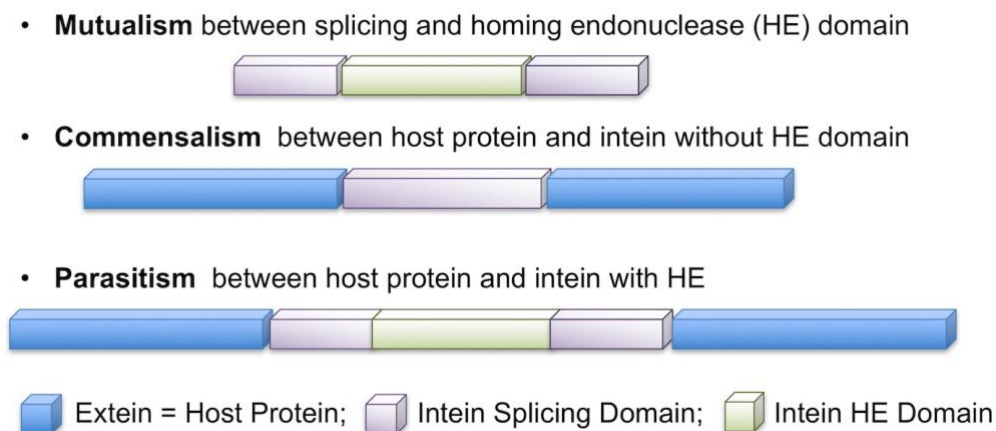


Figure 2: Types of symbiotic relationships between intein domains and the host protein. Most inteins contain two protein domains with distinct functions: the splicing domain, which removes the intein from the host protein after translation, and the homing endonuclease (HE) domain. In the linear sequence of the gene the HE domain is inserted inside the sequence encoding the splicing domain; however, in the folded protein the splicing domain forms a compact protein structure, distinct from the HE domain. The splicing domain prevents a high fitness cost due to interruption of the host protein, the HE domain provides mobility to the intein, because it allows invasion of uninvaded alleles. Inteins that have lost the HE domain (middle) have a commensal relationship to the host protein, whereas inteins with HE activity can be considered parasites of the host protein and the host organism (Gogarten and Hilario 2006).

1.2 Intein symbiotic state and the Homing Cycle

Most inteins are composed of two domains, a homing endonuclease domain, and a splicing domain which flanks the homing endonuclease domain on either side (Fig. 1a and Fig. 2). The splicing domain is the main component of the intein, this domain removes the intein sequence from the host protein sequence, also called extein, after translation. The homing endonuclease domain is considered an accessory domain, and is sometimes absent from intein sequences, these inteins are called mini-inteins (Derbyshire et al. 1997); inteins with homing endonuclease domains are called full-sized inteins (Soucy et al. 2014). The homing endonuclease domain initiates the conversion of uninvaded alleles into intein containing alleles.

The canonical model of intein invasion and loss, called the homing cycle (Goddard and Burt 1999), starts with an intein containing allele encountering an uninvaded allele. The homing endonuclease makes a double-strand break at the invasion site specific for the homing endonuclease. Using the invaded allele as a template the host DNA repair machinery copies the intein into the invasion site, thereby interrupting the invasion site and preventing future homing endonuclease activity (Fig. 1b). The homing cycle asserts that the invasion of a population goes to completion, and that subsequently homing endonuclease activity decays due to a lack of selection for function, because the opportunity to invade a new allele no longer exists. Recent findings revealed limitations of the homing cycle model. The different versions of the intein insertion site are in an intransitive fitness relationship (Barzel et al. 2011): The HE converts empty target sites, mini inteins are smaller than inteins with HE and therefore are expected to place less of a fitness burden on the host, and organisms with empty target sites may outcompete carriers of mini-inteins for the same reason. Theoretical studies showed that a wide range of fitness relationships between the three states of the intein insertion site allow for their coexistence over long periods of time even in well mixed populations (Yahara et al. 2009; Barzel et al. 2011). These theoretical considerations suggest that invasion might never go to completion and that some empty insertion sites might remain in the population. The persistence of empty target sites at low frequencies also solves the enigma of how the intein encoding part of a gene can be lost by a precise deletion – the last step of the homing cycle. In case of introns, a precise deletion may occur via recombination with a processed mRNA (Jeffares et al. 2006); however, in case of inteins the splicing reaction only occurs after translation, and how this information could flow back from the protein to DNA level remains enigmatic. The precise deletion through recombination with an intein free homolog, either persisting in the population at low copy numbers, or acquired through horizontal gene transfer provides a solution to this puzzle. The theoretical considerations also are in agreement with finding empty target sites and inteins with functioning homing endonuclease in the same population (Soucy et al. 2014; Naor et al. 2016) and with the persistence of functioning homing endonucleases in lineages over long periods of time (Gogarten and Hilario 2006; Butler et al. 2006).

In either of the above scenarios of intein invasion an intein without homing endonuclease, a mini-intein, reflects a long-term association of the intein with the host protein via vertical transmission from mother to daughter cells. Furthermore, since intein invasion relies on the presence of a functioning homing endonuclease, a full-size intein

invaded the extein more recently so that a selective pressure existed to maintain the homing endonuclease domain. Therefore, the presence of an intact homing endonuclease is a sign of horizontal gene transfer (HGT) in the recent (evolutionary time) history of the organism. The presence of a homing endonuclease reflects the symbiotic state of the intein (commensal or parasite, Fig. 2), and this symbiotic state provides a marker for the evolutionary history of that intein (Soucy et al. 2014). Curiously, the substitution rate of inteins are much higher than the adjacent extein sequence (Swithers et al. 2009). Except for the terminal residues required for splicing activity, the intein sequence accumulates substitutions at a high rate, leading to decay of the homing endonuclease sequence. This elevated substitution rate also creates informative characters, providing resolution to phylogenies, even between closely related species. The high substitution rate combined with the symbiotic state of inteins makes them valuable as markers of horizontal gene transfer, as this information provides information about which organisms were sharing genes (Swithers et al. 2013; Soucy et al. 2014).

1.3 Inteins as tools to trace gene flow

Genes acquired through horizontal gene transfer can have several selective outcomes, they can be beneficial to the recipient, they may be nearly neutral, or they may be detrimental. Undoubtedly the effect on the fitness of the host has an impact on the survival of the transferred gene in the recipient lineage. Therefore, an ideal tracer for naturally occurring gene transfer frequencies and patterns should not have an impact on the fitness of the recipient. The intein's self-splicing activity insures that the host protein remains functional, thereby avoiding a strong negative impact on the impact on the fitness of the host organism.

Transferred genes also need to integrate into the recipient genome, which can occur either through homologous recombination, or as an additive transfer. A good tracer should integrate into the recipient genome with a constant probability. Inteins force integration into the recipient genome through cutting the target site in the recipient genome. Another intein feature useful to reveal patterns of naturally occurring gene transfer is that inteins do not possess their own machinery to catalyze the transfer between cells. Therefore, their transfer provides a measure for the genetic exchange occurring in the host cells.

A potential problem is that inteins can only invade if an empty target site is present. A target site that is already occupied cannot be invaded again. However, observation shows that even within populations of the same species, empty target sites frequently persist (Soucy et al. 2014; Naor et al. 2016). Another concern is that full-size inteins can increase mating and recombination frequency of organisms that carry them (Giraldo-Perez and Goddard 2013; Naor et al. 2016); however, this does not detract for the inteins tracing transfer, even though the transfer rate maybe higher in the presence of the intein than in its absence.

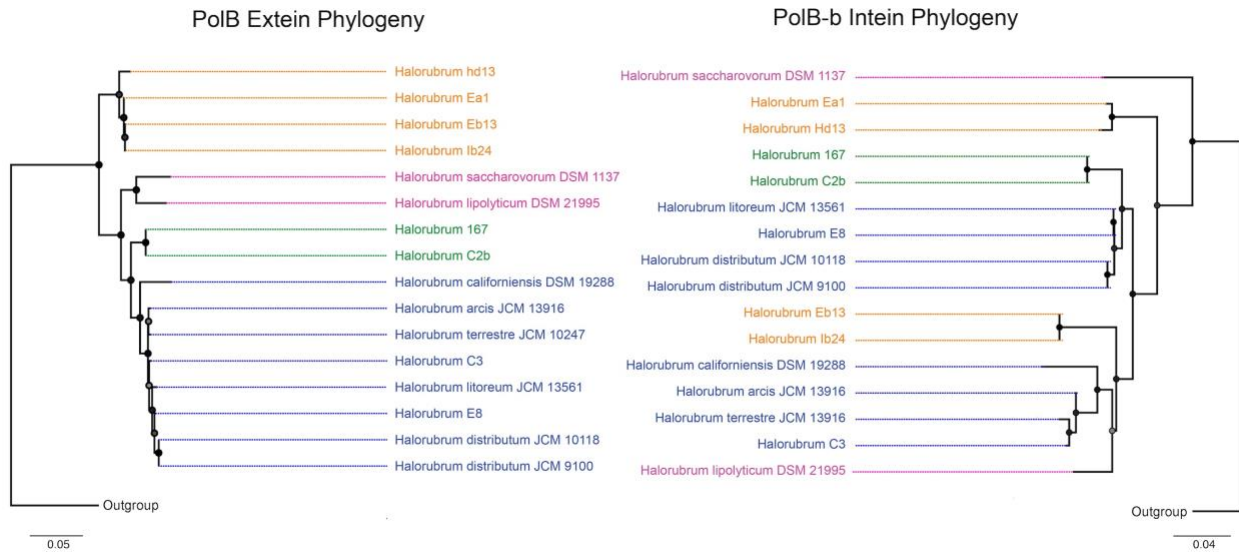


Figure 3: Comparison of the extein and intein evolutionary histories. The tree on the left gives the reconstructed phylogeny of the DNA polymerase B extein in the *Halorubrum* genus, the tree on the right gives the reconstructed phylogeny of the intein. Sequences were aligned using Muscle v.3.8.31 (Edgar 2004), and phylogenies were constructed using FastTree (Price et al. 2010). Strains and species names are given in the same color/grayscale value for both trees. The strain order in the extein phylogeny is strikingly different than in the intein phylogeny, indicating a plethora of horizontal gene transfer occurring within the *Halorubrum*.

2. Inteins as markers of bio-communication

2.1 Inteins as markers of bio-communication in closely related organisms

Part of the intein's life cycle (see section 1.2. and Fig. 1) involves transmission to new hosts via horizontal gene transfer. Therefore, we can use the distribution of inteins to identify HGT events between both closely related and more distantly related organisms (Swithers et al. 2013; Soucy et al. 2014; Fullmer et al. 2014; Naor et al. 2016). Especially between closely related organisms with high sequence similarity, gene transfer is difficult to detect using phylogenetic approaches. In these cases, the fast evolving intein sequences can serve as markers of bio-communication between organisms. Most intein alleles are shared by closely related species. In a 2014 survey of inteins in Haloarchaea, six out of 24 inteins were found exclusively in the Haloarchaea (i.e., they had not been found outside the Haloarchaea), and 16 out of 24 had the majority of sequences in the Haloarchaea (Soucy et al. 2014). Seven out of the ten inteins that had a majority in the Haloarchaea, but were shared with taxa outside the group were shared predominately with other Euryarchaeota (Soucy et al. 2014). This indicates that most inteins, at least in the Haloarchaea, are shared within phyla. In a similar analysis comparing intein and extein phylogenies for the *vma1-b* Swithers et al. showed that the intein in *Thermococcus litoralis* had been acquired from within the *Pyrococcus*, another genus within the Thermococcaceae (Swithers et al. 2013). Even more striking is the phylogenetic conflict between *polB* extein and *polB-b* intein in different *Halorubrum* strains, revealing transfer within the *Halorubrum* genus (Fig. 3), and the frequent transfer of the *pol II-a* between strains belonging to the *Halorubrum* genus and other Haloarchaea (Fig. 4). It is noteworthy that even though the *polB-b* and

pol II-a inteins were frequently transferred within *Halorubrum* and between *Halorubrum* and other Haloarchaea, these inteins are present in only 23 and 14 out of 37 *Halorubrum* genomes, respectively.

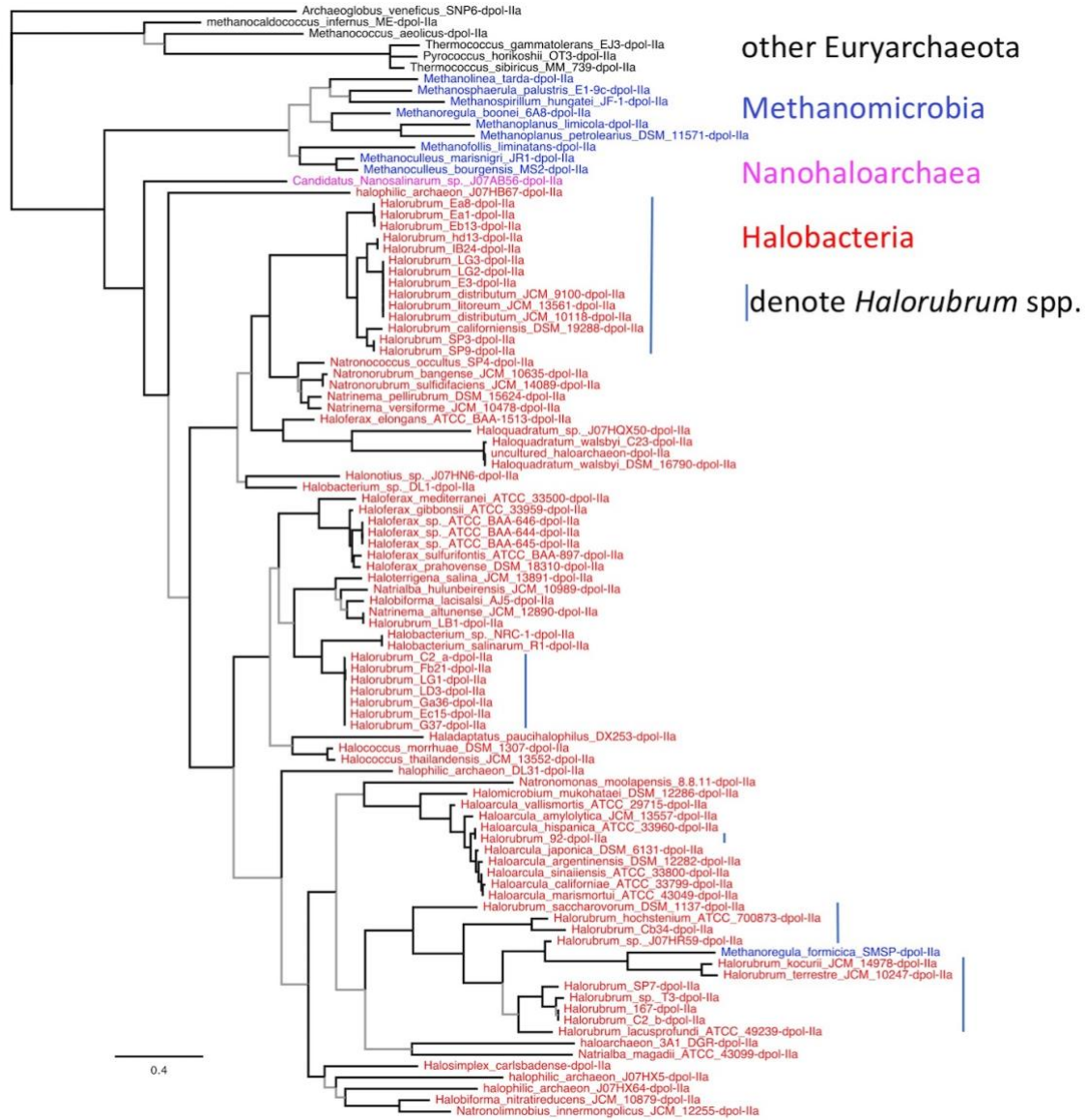


Figure 4: Maximum likelihood phylogeny of the intein in the large subunit of the DNA polymerase II, insertion site a (*pol II-a*). The phylogeny was calculated with PhyML (Guindon et al. 2010) using the LG, Gamma + I model and sequences as described in (Soucy et al. 2014). Branches with a support values (Anisimova and Gascuel 2006) below 0.85 are shown in gray. Sequences are as described in (Soucy et al. 2014). Font colors/grayscale denote different groups of archaea. Intein sequences from the genus *Halorubrum* group in several distinct groups within the Halobacteria and these groups, indicated by lines next to the strain names, are separated by well supported branches. One intein from a Methanomicrobia (*Methanoregula formicica* SMSP) groups within one of these *Halorubrum* clusters, whereas the other *pol II-a* inteins from Methanomicrobia group as a sistergroup to the halobacterial homologs, as is expected from the relationship between Haloarchaea and Methanomicrobia based on the phylogeny of the translation machinery (Woese 1987; Lasek-Nesselquist and Gogarten 2013).

2.2 Detecting the direction of bio-communication

Phylogenetic comparison between intein and extein phylogenies can provide clear indications that the intein has been transferred (Figs. 3 and 4); however, conflict between intein phylogeny and that of the host or host protein often is insufficient to determine the direction of transfer. For example, from the phylogenies depicted in Fig. 4 it is not always clear when a *Halorubrum* species was the donor and when the recipient of an intein transfer; however, the fact that the inteins from *Halorubrum* group in 5 well separated locations in this phylogeny reveals repeated transfer of this intein between Haloarchaeal genera. Sometimes the tree topology does provide a clue: the single intein from a methanomicrobia grouping within a group of inteins from *Halorubrum* strains, strongly suggest that this methanomicrobion, *Methanoregula formicica*, acquired the intein through genetic exchange with a Haloarchaeon, likely a member of the genus *Halorubrum* (Fig. 4). The comparison between *polB*-b intein and *polB* extein phylogenies (Fig. 3) suggests that the ancestor of *Halorubrum* Eb13 and Ib24 may have acquired the intein through transfer from the ancestor of *Hrr. californiensis*, *Hrr. arcis*, *Hrr. terrestre* and *Hrr. sp C3*. Similarly, the grouping of the *Thermococcus litoralis vma1-b* intein within a group of homologous inteins from several *Pyrococcus* species reveals *T. litoralis* as the recipient of the transfer (Swithers et al. 2013). A more sophisticated analysis involving the ratio of the pairwise phylogenetic distance between intein sequences and a reference sequence (the associated extein can sometimes serve as a reference) can help to infer additional transfers and to determine the direction of transfer (Swithers et al. 2013). The underlying rationale is that if two intein sequences are more similar to each other than expected from the similarity between the two corresponding extein sequences, then the intein was likely acquired through horizontal transfer and not only through vertical inheritance. However, this approach requires calibration of substitution rates for inteins that have been vertically transferred. Swithers et al. used a mini-intein to calibrate their dataset, but mini-inteins are not available for many intein datasets. More needs to be done to develop this method to determine and study gene transfer events between close relatives.

2.3 Inteins as markers of bio-communication in distantly related organisms

Perhaps more interesting than gene transfers between closely related organisms are transfer events that involve distantly related organisms. Many questions about the boundaries of gene transfer remain to be answered. For transfers between distantly related lineages the direction of transfer can often be inferred from the distribution of the intein. If there is a clear majority of close relatives of one of the partners, it is likely that organism is the donor. In the Haloarchaea 14 of 24 inteins are shared with at least one bacterial species. Surprisingly the shared inteins are not limited to other salt tolerant microbes but are found in bacteria from six different phyla (Soucy et al. 2014). Six of the 24 inteins found in the Haloarchaea are found in more bacterial species than archaeal species, indicating the intein may have been transferred from the bacteria and into the Haloarchaea. The other eight intein alleles that are shared between Bacteria and Haloarchaea have far more representatives in Haloarchaea than Bacteria, indicating that bio-communication is occurring in both directions. Thirteen out of 24 inteins in the Haloarchaea are shared with other Euryarchaeota (Soucy et al. 2014). The two inteins

with the highest number of other Euryarchaeota, *cdc21-a* and *pol-IIa*, also have the highest proportion of mini-inteins, implying these inteins have an ancient association with the euryarchaeal lineage (Soucy et al. 2014). The intein with the highest proportion of euryarchaeal sequences outside the Haloarchaea, *rhc-a*, also has several mini-inteins (Soucy et al. 2014). In contrast, 10 of the 14 inteins shared between Haloarchaea and Bacteria have no mini-inteins in the Haloarchaea, and few mini-inteins in the Bacteria (Soucy et al. 2014).

These findings support ancient and ongoing bio-communication through gene transfer with other Euryarchaeota and more recent bio-communication with bacterial species, supporting earlier evidence that haloarchaeal evolution included extensive gene transfers with the bacteria (Williams et al. 2012; Nelson-Sathi et al. 2012; Nelson-Sathi et al. 2014; Becker et al. 2014). No clear pattern has emerged regarding the types of genes that are invaded by inteins shared by Bacteria and Archaea, and those that are found only in Archaea. As with most inteins, most of the shared inteins fall into the replication, recombination, and repair COG category.

Inteins are valuable indicators of bio-communication through gene transfer within and between species. The phylogenetic information provided by inteins due to their high substitution rate has been used to identify gene transfer between organisms at all level of relationship: within populations, between related species belonging to the same order, between orders, phyla, and even between organisms from different domains. More needs to be done to integrate information of symbiotic state and substitution rates with the comparative analysis of topologies of gene and species phylogenies. To date, conclusions from using inteins as indicators of gene flow confirm results obtained from using whole genome data and individual gene phylogenies: Genes are transferred much more frequently between closely related organism than between divergent organisms (Andam and Gogarten 2011; Williams et al. 2012), but even transfers across domain boundaries have occurred, and some of these greatly impacted the physiology and ecology of the recipient organisms (Soucy et al. 2015).

References

- Andam CP, Gogarten JP (2011) Biased gene transfer in microbial evolution. *Nat Rev Microbiol* 9:543–55. doi: Research Support, U.S. Gov't, Non-P.H.S.
- Anisimova M, Gascuel O (2006) Approximate Likelihood-Ratio Test for Branches: A Fast, Accurate, and Powerful Alternative. *Syst Biol* 55:539–552. doi: 10.1080/10635150600755453
- Barzel A, Obolski U, Gogarten JP, Kupiec M, Hadany L (2011) Home and away- the evolutionary dynamics of homing endonucleases. *BMC Evol Biol* 11:324. doi: 10.1186/1471-2148-11-324
- Becker EA, Seitzer PM, Tritt A, Larsen D, Krusor M, Yao AI, Wu D, Madern D, Eisen JA, Darling AE, Facciotti MT (2014) Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet* 10:e1004784. doi: 10.1371/journal.pgen.1004784
- Butler MI, Gray J, Goodwin TJD, Poulter RTM (2006) The distribution and evolutionary history of the PRP8 intein. *BMC Evol Biol* 6:42. doi: 10.1186/1471-2148-6-42
- Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M (1997) Genetic

- definition of a protein-splicing domain: Functional mini-inteins support structure predictions and a model for intein evolution. *Genetics* 94:11466–11471.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–7. doi: 10.1093/nar/gkh340
- Fullmer MS, Soucy SM, Swithers KS, Makkay AM, Wheeler R, Ventosa A, Gogarten JP, Papke RT (2014) Population and genomic analysis of the genus *Halorubrum*. *Extrem Microbiol*. doi: 10.3389/fmicb.2014.00140
- Giraldo-Perez P, Goddard MR (2013) A parasitic selfish gene that affects host promiscuity. *Proc Biol Sci* 280:20131875. doi: 10.1098/rspb.2013.1875
- Goddard MR, Burt A (1999) Recurrent invasion and extinction of a selfish gene. *Proc Natl Acad Sci U S A* 96:13880–5.
- Gogarten JP, Hilario E (2006) Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol* 6:94. doi: 10.1186/1471-2148-6-94
- Gogarten JP, Senejani AG, Zhaxybayeva O, Olendzenski L, Hilario E (2002) Inteins: structure, function, and evolution. *Annu Rev Microbiol* 56:263–87. doi: 10.1146/annurev.micro.56.012302.160741
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. doi: 10.1093/sysbio/syq010
- Jeffares DC, Mourier T, Penny D (2006) The biology of intron gain and loss. *Trends Genet* 22:16–22. doi: 10.1016/j.tig.2005.10.006
- Lasek-Nesselquist E, Gogarten JP (2013) The effects of model choice and mitigating bias on the ribosomal tree of life. *Mol Phylogenet Evol* 69:17–38. doi: 10.1016/j.ympev.2013.05.006
- Naor A, Altman-Price N, Soucy SM, Green AG, Mitiagin Y, Turgeman-Grott I, Davidovich N, Gogarten JP, Gophna U (2016) Impact of a homing intein on recombination frequency and organismal fitness. *Proc Natl Acad Sci U S A* 113:E4654-61. doi: 10.1073/pnas.1606416113
- Naor A, Lapierre P, Mevarech M, Papke RT, Gophna U (2012) Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr Biol* 22:1444–8. doi: 10.1016/j.cub.2012.05.056
- Nelson-Sathi S, Dagan T, Landan G, Janssen A, Steel M, McInerney JO, Deppenmeier U, Martin WF (2012) Acquisition of 1,000 eubacterial genes physiologically transformed a methanogen at the origin of Haloarchaea. *Proc Natl Acad Sci U S A* 109:20537–42. doi: 10.1073/pnas.1209119109
- Nelson-Sathi S, Sousa FL, Roettger M, Lozada-Chávez N, Thiergart T, Janssen A, Bryant D, Landan G, Schönheit P, Siebers B, McInerney JO, Martin WF (2014) Origins of major archaeal clades correspond to gene acquisitions from bacteria. *Nature*. doi: 10.1038/nature13805
- Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI, Belfort M (2015) Intein Clustering Suggests Functional Importance in Different Domains of Life. *Mol Biol Evol*. doi: 10.1093/molbev/msv271
- Perler FB (2002) InBase: the Intein Database. *Nucleic Acids Res* 30:383–4.
- Petrokovski S (2001) Intein spread and extinction in evolution. *Trends Genet* 17:465–72.

- Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Soucy SM, Fullmer MS, Papke RT, Gogarten JP (2014) Inteins as indicators of gene flow in the halobacteria. *Front Microbiol* 5:299. doi: 10.3389/fmicb.2014.00299
- Soucy SM, Jinling H, Gogarten JP (2015) Horizontal gene transfer: building the web of life.
- Swithers KS, Senejani AG, Fournier GP, Gogarten JP (2009) Conservation of intron and intein insertion sites: implications for life histories of parasitic genetic elements. *BMC Evol Biol* 9:303. doi: 10.1186/1471-2148-9-303
- Swithers KS, Soucy SM, Lasek-Nesselquist E, Lapierre P, Gogarten JP (2013) Distribution and Evolution of the Mobile vma-1b Intein. *Mol Biol Evol*. doi: 10.1093/molbev/mst164
- Topilina NI, Novikova O, Stanger M, Banavali NK, Belfort M (2015) Post-translational environmental switch of RadA activity by extein–intein interactions in protein splicing. *Nucleic Acids Res* 43:6631–6648. doi: 10.1093/nar/gkv612
- Williams D, Gogarten JP, Papke RT (2012) Quantifying homologous replacement of loci between haloarchaeal species. *Genome Biol Evol* 4:1223–44. doi: 10.1093/gbe/evs098
- Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–71.
- Yahara K, Fukuyo M, Sasaki A, Kobayashi I (2009) Evolutionary maintenance of selfish homing endonuclease genes in the absence of horizontal transfer. *Proc Natl Acad Sci U S A* 106:18861–6. doi: 10.1073/pnas.0908404106