Chapter 2

Defining the Mobilome

Janet L. Siefert

Abstract

This chapter defines the agents that provide for the movement of genetic material which fuels the adaptive potential of life on our planet. The chapter has been structured to be broadly comprehensive, arbitrarily categorizing the mobilome into four classes: (1) **transposons**, (2) **plasmids**, (3) **bacteriophage**, and (4) **self-splicing molecular parasites**.

Our increasing understanding of the mobilome is as dynamic as the mobilome itself. With continuing discovery, it is clear that nature has not confined these genomic agents of change to neat categories, but rather the classification categories overlap and intertwine. Massive sequencing efforts and their published analyses are continuing to refine our understanding of the extent of the mobilome. This chapter provides a framework to describe our current understanding of the mobilome and a foundation on which appreciation of its impact on genome evolution can be understood.

Key words: Mobilome, mobile genetic elements (MGEs), transposable elements (TEs), transposons, plasmids, bacteriophage, Group II introns, jumping genes, insertion sequences (ISs).

1. Introduction

As important as the process of horizontal gene transfer is for biology, the mobilome is the agent of change that facilitates these events. Appreciating and understanding it is critical to understanding the constraints and freedoms that dictate how, why, and when gene transfer events are likely to take place. It is not overstating the fact to say that the mobilome is a driver of evolution. Our first indication for these agents of change can be traced to Barbara McClintock. In compelling and insightful experiments, she described the mobile elements that were affecting the expressions of genes in the chromosomes of maize *(1)*. She called them "controlling elements"; her discovery would earn her the Nobel

Maria B. Gogarten et al. (eds.), *Horizontal Gene Transfer: Genomes in Flux, vol. 532*

DOI 10.1007/978-1-60327-853-9_2 Springerprotocols.com

Prize and provide a glimpse into the process of transposition. We can now appreciate, and in many cases utilize, these diverse agents of change that we recognize as mobile genetic elements.

1.1. What Is the Mobilome?

Simply, the mobilome consists of all mobile genetic elements (MGE) in a cell. MGEs are any type of DNA that can move around within or between genomes. MGEs can be divided into categories based on their mechanism of movement and the character of the DNA sequences. For our purposes here, I will divide MGEs into four broad categories: (1) **transposons** (synonyms: transposable elements, TE, "jumping genes"), (2) **plasmids**, (3) **bacteriophage** elements, and (4) **self-splicing molecular parasites** (**Table 2.1**). The total of all these MGEs in any cell is referred to as its mobilome.

To define MGEs it is helpful to consider the mobilome in the context of the cellular organization of prokaryotes and eukaryotes (despite the strangely fervent, well-intentioned, but largely impractical ideas to dispense with this organization in biology) *(2)*. For eukaryotes the broad category of *transposable elements* (TE) are the major mobile elements. Eukaryotes are predominantly sexually reproducing entities. In multicellular eukaryotes, the division of cell types into somatic and gametic (germ line) cells greatly limits the number of vectors that can facilitate the movement of DNA. Complicating this endeavor is the reality that for any alien DNA to be passed to the next generation, the transposition must be taking place in the germ line of the eukaryote. Within the eukaryotes, animals are the ones most likely to suffer from this barrier, as the germ line and somatic cells are separated early in development. The likelihood of inter-cell transfer by MGEs for eukaryotes is low as the alien DNA would need to overcome these hurdles. This does happen, and there are certainly dramatic examples from the past involving unicellular eukaryotes (consider mitochondria, plastids, and the diversion of genes to the host nucleus). For the most part, we can confine our estimation of the eukaryotic mobilome to intragenomic dynamics, and frequently within somatic cells.

For prokaryotes, the repertoire of MGEs is defined as a broader set of agents because movement of genetic material is not primarily confined within a genome as it is for eukaryotes, but includes elements that provide movement *between* genomes. Indeed, as essentially non-sexual entities, prokaryotic diversity is fueled by HGT. What should be appreciated is that this is a relatively new understanding for prokaryotes. In the early 1960s, Margaret Dayhoff, combining biology and extraordinary computational skill and insight, created statistical methodologies for exploring the phylogeny of DNA and protein sequences. They were based on the assumption that HGT seldom occurred, and while in practice this view was instrumental in the rich progress

for the field of bioinformatics, the truth lies somewhere else. HGT does happen and in significant proportion. By 1994, Guttman and Dykhuizen demonstrated that recombination, rather than mutation, was a driving evolutionary force *(3)*. That is why you are reading this book.

The goal of this chapter is to review and describe our current understanding of the mobilome. I will include a definition and general categorization of all MGEs, and describe the current knowledge of their presence and effects on genome organization in both eukaryotes and prokaryotes.

Table 2.1
Major components of the mobilome

Mobilome
Transposons
Retrotransposons
LTR retrotransposases
non-viral, non-LTR transposases
LINEs
SINEs (e.g., Alu sequences)
Viral-like retrotransposases
DNA transposons
Insertion sequences
Plasmids
Conjugative
Non-conjugative
Insertion sequences in plasmids
Bacteriophages
Lytic/lysogenic/prophage
Gene transfer agents
Filamentous phages
Self-splicing molecular parasites
Group II introns
Group I introns
Inteins
Homing endonucleases

2. Transposons

Transposons, or "jumping genes," affect genomic dynamics within a single cell through the process of transposition and can alter gene expression by serving as promoters, enhancers, silencers, sites of epigenetic modification, or alternative splicing sites *(4)*. Sometimes transposons provide "molecular domestication" activities when their host adopts a transposon-encoded protein that can perform cellular functions *(4)*. Because copies of transposons can flood a genome, they serve as hotspots for recombination events, producing deletions, duplications, inversions, or translocations. Transposition can be disruptive, but it can have a positive role through stable integration and longlasting protein expression in cells. Through production of insertions and duplications, transposons greatly affect genome size as demonstrated by the extensive variation in nuclear genome size, or C-value, among eukaryotic species.

Based on their mechanisms of movement, transposons can be divided into three classes: **retrotransposons** (retroposons), **DNA transposons**, and **insertion sequences** (ISs).

2.1. Retrotransposons

Retrotransposons (also called retroposons) invade the genome by being transcribed to RNA and then back to DNA by reverse transcriptase; they are "copy and paste" reactions *(4)*. The reverse transcriptase is often, but not always, coded by the transposon itself and these MGEs can paste numerous copies into a genome, amplifying themselves in their host genome in tremendous numbers. They are ubiquitous in eukaryotic organisms, and are particularly abundant in plants. They make up between 49% and 78% of the maize genome *(5)* and 68% of the genome in wheat *(6)*. In mammals about half the genome is composed of retrotransposases and 42% of the human genome consists of remnants of retroposons *(7)*. This translates into millions of elements contained in our genomes.

There are two main categories of retrotransposons. The first group, LTR retrotransposons, encodes their own reverse transcriptase, contains long terminal repeats (LTRs), and is transcribed by RNA polymerase II. These are further classified into three subclasses: Ty1-copia-like, Ty3-gypsy-like, and Pao-BEL-like group, based on the degree of sequence similarity and the order of encoded gene products in them. Ty1-copia-like and Ty3-gypsy-like are found in high copy number in animals, fungi, protista, and plants with large genomes. Pao-BEL elements have been found only in animals so far *(8–10)*.

The second group is a non-viral, non-LTR superfamily, which is transcribed by RNA polymerase III and does not code for

reverse transcriptase. There are two subtypes of non-LTR retrotransposases, long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Originally considered "junk DNA", LINEs and SINEs play a significant role in gene evolution, structure, and determination of transcription levels. LINEs make up about 21% of the human genome and are used in forensics to generate genetic fingerprints. The non-coding SINEs are dependent upon their partner LINEs for reverse transcription *(11)*. Alu sequences are the most common SINE sequences in primates. SINEs are found in high copy numbers (up to 250,000) in the plant species and primates and make up about 13% of the human genomes *(7)*. These TEs have an impact on somatic cells. LINE 1 (L1) retrotransposases, along with Alu sequences, have been implicated in cancer through active mobilization, domestication of transposases, and by influencing genomic rearrangements *(4)*. The distribution of SINEs has been implicated in some genetic diseases and cancers. Alu sequences have been associated with breast cancer, Ewing's sarcoma, familial hypercholesterolemia, hemophilia, neurofibromatosis, and diabetes mellitus type II. SINEs and other retrotransposons are also used in forensic genomics through quantitative species-specific DNA detection, analysis of complex biomaterials, and the inference of geographic origin of human DNA samples *(12)*. Alu elements are the most abundant repetitive elements in the human genome; they emerged in the primate lineage 65 million years ago from a 5′ to 3′ fusion of the 7SL RNA gene and subsequently amplified throughout the human genome by retrotransposition to reach the present number of more than one million copies *(13)*.

A third category, viral-like retrotransposons, which are similar to retroviruses such as HIV, HIV-1 or HTLV-1, behave like retrotransposons and contain both reverse transcriptase and integrase. The integrase is the retrotransposon equivalent of the transposase of DNA-transposons. There is molecular evidence that endogenous retroviruses in the human genome may be involved in autoimmune diseases *(14)*.

2.2. DNA Transposons

This class of MGEs moves directly from one position to another within the genome using a transposase to "cut and paste" itself within the genome. The major difference of DNA Class II transposons from retrotransposons is that their transposition mechanism does not involve an RNA intermediate. Some DNA transposases can bind to any part of the DNA molecule and therefore the target site can be anywhere in the genome, while others bind to specific sequences. The transposase enzyme produces a staggered cut at the target site of DNA, producing sticky ends. The enzyme cuts out the transposon and ligates it into the target site, resulting in target site duplication. Insertion sites of DNA

transposons may be identified by short direct repeats followed by inverted repeats (which are important for the transposon excision by transposase). Not all DNA transposons transpose through a cut and paste mechanism. In some cases, a replicative transposition has been described in which the transposon replicates itself to a new target site.

To date there have been no published accounts of naturally occurring active DNA transposases in mammalian genomes. However, there are many copies of inactive fossil DNA transposons present *(7)*, and there are genes that were domesticated from these elements in their host genomes *(15)*. An example is the recombination-activating gene RAG1 and RAG2 proteins. These are important for generating somatic diversity in the immune system because they play an indispensable role in recombination during lymphocyte development. The *Rag* gene is hypothesized to have its origin from the transposase encoded by an ancient transposon superfamily named Transib *(16,17)*.

Analysis of the *Caenorhabditis elegans* genome indicates that approximately 12% of the *C. elegans* genome is derived from TEs. However, most of these sequences are fossil remnants that are no longer mobile but can be used by molecular archeologists to trace the interactions between parasitic "selfish DNA" *(18)* and a host genome. The investigation into transposition regulation in *C. elegans* has uncovered an unforeseen link between transposition and genome surveillance, genetic responses in the host genome involving chromatin modifications, and RNA interferences that have evolved to moderate or modify the effect of TEs *(19)*.

Tc1 and Tc3 are part of a superfamily of TEs, which is named after its two best-studied members: Tc1 and the related transposon *mariner*, which was identified in 1986 in *Drosophila mauritiana (20)*. Tc1/*mariner* elements are probably the most widespread DNA transposons in nature and can be found in fungi, plants, ciliates, and animals including ecdysozoans and vertebrates *(21)*. Molecular reconstruction of Sleeping Beauty, an ancient transposon in fish, represents a cornerstone in applying transposition-mediated gene delivery in vertebrate species, including humans *(22)*.

2.3. Insertion Sequences

An insertion sequence (IS, insertion sequence element, IS element) is a simple TE consisting of a short DNA sequence generally around 700–2500 bp in length. They do not carry any accessory genes and only code for proteins that are part of the transposition activity. This usually includes a transposase, which catalyses the enzymatic reaction, and a regulatory protein, which can stimulate or inhibit the activity. Inverted repeats typically flank the coding regions. ISs can be autonomous but may also be part

of composite transposons. In a composite transposon (also known as a "complex transposon"), two ISs flank one or more accessory genes, such as an antibiotic-resistance gene (e.g., Tn*10*, Tn*5*). ISs are proficient in moving neighboring genes.

Over 2500 different ISs have been identified to date. They are integral components of the mobile, dramatically affecting bacterial genomes as they reshuffle and shape them. Massive expansion of ISs is indicated in the emergence of some pathogenic bacterial species. They also play a role in assembling genes into complex plasmid structures *(23, 24)*. Today, a minimal and incomplete repertoire deposited in the ISfinder database (http://www-is.biotoul.fr) includes 2200 different ISs in over 295 eubacterial and archaeal species *(25)*. Using shared features, ISs are classified into about 20 families. An IS family is defined as a group of ISs with related transposases, strong conservation of the catalytic site, and conservation of organization. ISs are found in most but not all eubacterial and archaeal genomes *(25)*.

Analysis of a diverse set of 18 bacterial genomes revealed that the intra-genomic sequence diversity of a given IS is very low, suggesting that most ISs in an individual genome are evolutionarily young and might have been recently acquired. This observation can be explained by a period of IS expansion followed by a series of IS "extinctions" in bacterial lineages. Wagner *(26)* explains this scenario through the transitory selective advantage of IS elements to their host through lateral gene transfers and genomic rearrangements, which might prove to be detrimental to their host in the long term. However, the loss can be explained if the element is selectively neutral or nearly so. Although ISs are usually discussed in terms of prokaryotic genomes, certain eukaryotic DNA sequences belonging to the family of Tc1/*mariner* TEs may be considered to be ISs.

3. Plasmids

American molecular biologist Joshua Lederberg first coined the term plasmid in 1952 *(27)*. They are extra-chromosomal DNA molecules capable of autonomous replication separate from chromosomal DNA. They are typically circular and double-stranded and occur naturally in bacteria. As naturally occurring elements, they are usually not essential elements of the cell. Sometimes they are found in eukaryotic organisms (e.g., the *2-micrometre-ring* in *Saccharomyces cerevisiae*) *(28)*. Plasmid size varies from 1 to over 400 kbp. They may exist as a single copy in a cell or hundreds or thousands of copies of the same plasmid in a single cell.

Plasmids can be categorized in a variety of ways, including by function. These include five main functional classes: fertility

(F) plasmids, resistance (R) plasmids, Col-plasmids which contain genes that code for genes that kill other bacteria, degradative plasmids which are capable of the digestion of unusual substances, and virulence plasmids which can turn bacteria into pathogens. Interestingly, plasmids can belong to more than one functional group. Plasmids are also defined as high (over 100 molecules per cell) and low copy plasmids (1–25). Plasmids can be assigned into *compatibility groups*. It is possible for plasmids of different types to coexist in a single cell, but related plasmids are often incompatible, with only one of the relative surviving in the cell line. A good example of plasmid diversity is demonstrated in *Escherichia coli*, which hosts seven different plasmids. Plasmids are active participants in horizontal gene transfer (HGT) through the transfer of genetic material during conjugation *(29)*.

3.1. Conjugative Plasmids

Conjugation is bacterial sex, if you define sex as cell-to-cell contact with the exchange of material. Contrary to real sex in eukaryotes, there is no fusion of gametes, creation of zygotes, or equal exchange of genetic material (the result of sexual activity in eukaryote). In bacterial sex, the donor (a bacterium with the conjugative plasmid) using a pilus initiates contact and the bacterial recipient, while in contact, receives the plasmid in the exchange. This is accomplished through the action of a suite of genes, many of which are contained on the *tra* locus of the plasmid. Sometimes the plasmids become integrated into the host genome, and when these plasmids undergo conjugation, all or part of the host genome may be transferred.

There is an interesting case of a conjugative plasmid active across domains. Plasmids in *Agrobacterium* and *Rhizobium* contain elements that transfer to plant cells *(30)*. Once the genes are transferred to the plant cell, the plant cell protein machinery is hijacked to produce opines, which are used by the bacteria for energy and cell-building sources. This inter-kingdom transfer from bacterial plasmids produces galls or root nodules in the infected plants.

3.2. Non-conjugative Plasmids

Non-conjugative plasmids are incapable of initiating conjugation; they are transferred when they associate with conjugative plasmids. Researchers continue to discover their mobilization potential despite this dependence. For instance, non-conjugative plasmids have been implicated in events as varied as the ripening process of goat's milk cheese *(31)* to the isolation of non-conjugative plasmids from antibiotic-resistant *Staphylococcus* strains *(32)*.

3.3. IS Elements in Plasmids

Siguier and co-workers *(23)* did an initial survey for IS elements in plasmids and found the incidence of IS sequences is about zero for

plasmids smaller than ~20 kb. Once the IS sequence size is greater than 20 kb, there is an abrupt increase in their abundance, averaging between 5% and 15%, and reaching 20% to >40% in some cases. Siguier and co-workers hypothesize that plasmids capable of self-transfer between various species and genera are larger because they carry transfer functions and are capable of acquiring transposons, ISs, and accessory genes during passage between different host genomes. Siguier and co-workers cite evidence that certain TEs specifically target transmissible plasmids.

4. Bacteriophages

Bacteriophages (or more commonly, phages) are viruses that infect bacteria. They are typically composed of an outer protein shell surrounding the genetic material. Phages can range in size from 5 to 500 kb, can be either circular or linear in architecture, and be present as single-stranded RNA, double-stranded RNA, single-stranded DNA, or double-stranded DNA. They are estimated to be the most diverse and widely distributed entities in the biosphere, found anywhere where there are bacterial hosts. Despite their small size (between 20 and 200 nm), they are extremely effective predators and control the world's bacterial population *(33)*. Phages are major contributors to the process of HGT due to their environmental ubiquity, their large numbers, and functional affects on the hosts they infect *(34)*. In order to understand their role in HGT, we need to appreciate in basic terms, the "phage life cycle."

4.1. The Phage Life Cycle

The lytic cycle in phage depends upon chance encounters with the proper bacterial host receptors and, once the phage finds its match and attaches, with syringe-like action, injects its DNA into the host. Replication of phage genes begins almost immediately (at the expense of the host protein needs) and virion particle assembly begins. In about 15 min, separately constructed phage heads and tails will spontaneously assemble, packaged tightly with genetic material. In as little as 20 min after infection, phages can lyse their host, releasing 300 plus virus particles that can in turn infect other hosts.

With **lysogenic phage (also called temperate phage)** the infection will not initially cause lysis of the cell, but instead upon infection the phage genome integrates with the host genome (referred to as an endogenous phage or prophage). This arrangement continues until the host finds itself in deteriorating conditions. Once this happens, the dormant prophage will be excised from the host genome and the cycle resulting in lysis of the cell

begins. During this process, various pieces of host DNA can be introduced into the viral genome, and as the lysis continues, host DNA is packaged into the soon-to-be-dispersed phages. Interestingly, the lysogenic cycle allows the host cell to continue to survive and reproduce while the virus remains resident in the host's genome and is reproduced in all of the cell's offspring. Furthermore, in a process known as lysogenic conversion, prophages can provide benefits to its host while they are dormant by adding new functions to the bacterial genome. The harmless strain of *Vibrio cholerae* turns into the highly virulent one that causes cholera by this process *(35)*.

4.2. Filamentous Phages

Filamentous phages (F-phages) are examples of how evolution produces entities that do not fit into neatly defined categories. They co-opt their host's machinery typical of phage, but never integrate nor do they ever lyse their host cell. When in their infective state, F-phages are shaped like rods, contain a single-stranded DNA genome (the wild type with ~6400 bp), and will naturally infect Gram-negative bacteria. Arguably, the most famous filamentous phage is M13. Once M13 finds an *E. coli* host that contains an F plasmid (which provided the host with a pilus that it MUST have for infection), the phage can gain entry into its host. Once there, a life cycle begins that includes a double-stranded DNA intermediate as the replicative form, which is converted to a single-stranded DNA intermediate prior to encapsidation. This is certainly a key reason why M13 is a major component of the molecular biologists' toolbox. Recombinant technologies take advantage of the fact that it never lyses its host but continually pumps phage particles into the media, yet remain discreet entities within the host (they rarely ever become lysogenic). Currently, M13 is being investigated for uses as nanostructures and nanotechnology, and in 2006 MIT researchers using a modified M13 produced a protein that would complex with cobalt ion, producing a cobalt oxide material with energy storage capacity higher than current lithium-ion batteries *(36)*.

4.3. Gene Transfer Agents

A virus-like agent (colloquially called "gene transfer agents") has been described in *Rhodobacter capsulatus*, which at first glance might seem to be a defective prophage *(37)*. Currently, they are found only in prokaryotes and these gene transfer agents seem to function through genomic DNA transfer between cells with no negative effects associated with the transfer to the recipient. They are widespread in alpha proteobacteria species; and Lang and Beatty *(37)* suggest that because of their regulatory mechanisms (they are expressed through histidyl-aspartyl signaling and quorum sensing genes), they should be viewed as other genuine cellular structures analogous to flagella or pili.

5. Self-Splicing Molecular Parasites

5.1. Group II Introns

Group II Introns are a class of introns found in the rRNA, tRNA, and mRNA of organelles of fungi, plants, and protists and in the mRNA of bacteria. Because of their wide distribution, their possible role in the evolution of nuclear spliceosomal introns and nuclear non-LTR-retrotransposons *(38)*, and their ability to "retrohome" and also "retrotranspose" at low frequencies into ectopic sites that resemble the normal homing site, they are a fundamental component of the mobilome. They are ribozymes that are retroelements consisting of a highly structured RNA (typically exhibiting six stem-loops known as domains) and a multifunctional intron-encoded protein (IEP). The RNA carries out the splicing and reverse splicing (integration) activities, while the IEP facilitates these reactions by stabilizing the catalytic structure of the ribozyme *(38)*. The mobility of Group II introns occurs through a target DNA-primed reverse transcriptase. The excised intron RNA reverse splices directly into a DNA target site and is then reverse transcribed by the IEP. With minimal host damage, Group II introns "retrohome" at almost 100% into specific DNA target sites, usually at unoccupied site in an intronless allele. These elements comprise more than 45% of the human genome *(38)* and sometimes they can be found internal to IS sequences.

A brief description of group II intron phylogeny and evolution can provide some insight into their current and past roles in the mobilome. There are more than 200 ORF-containing group II introns sequenced and described, and almost all encode reverse transcriptase-related proteins. Based on phylogenetic analysis, the IEPs can be divided into eight major lineages denoted mitochondrial, chloroplast-like 1 and 2, and bacterial A–E *(31, 32)* (**Fig. 2.1**).

Each lineage of IEP is associated with a distinct RNA structural subclass, implying that the IEP was likely associated with the intron RNA prior to the divergence of different group II intron lineages. Because there is no obvious exchange of IEP, it appears that these lineages have experienced little or no HGT *(39)*. The "mitochondrial" and "chloroplast" lineages include a number of bacterial group II introns (e.g., the *Lactobacillus lactis* Ll.LtrB intron belongs to the "mitochondrial" lineage), and this can be explained as a reflection that the organellar group II introns have evolved from specific bacterial lineages *(40)*. While the IEPs have experienced very little HGT, the heterogeneous phylogenetic distribution of group II intron subclasses suggests that HGT of the introns themselves is relatively common, and cross-species transfer by conjugation has been demonstrated *(38)*.

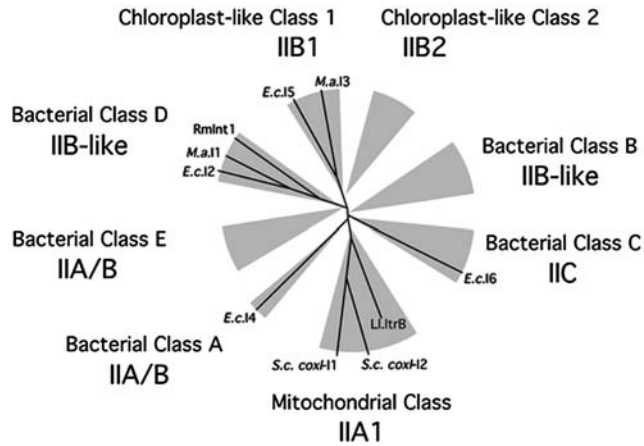

Fig. 2.1. Phylogeny of group II intron ORFs and correspondence with RNA structural classes. Phylogenetic relationships of group II intron ORFs are summarized based on neighbor-joining analyses. Group II intron ORFs are divided into eight clades, named mitochondrial, chloroplast-like 1 and 2, and bacterial A–E. Each ORF clade is associated with a distinct RNA structural class (IIA1, IIB1, IIB2, IIC, two other distinct IIB-like, and two distinct IIA/B hybrid classes) (from *(38)*, http://www.fp.ucalgary.ca/group2introns/).

Evidence for the conversion of Group II introns into spliceosomal introns is provided by degeneration of internal RNA structure, dependence on a common splicing apparatus, and the use of *trans*-acting RNAs. Because translation and transcription are separated in eukaryotes by the nuclear membrane, a Group II intron cannot bind the IEP immediately after the RNA is transcribed. This separation would favor substituting host protein in *trans*, which in turn favors the evolution of a common slicing apparatus. Mobility in early "spliceosomal" introns would be retained due to interaction with the RTs in *trans*. However, this would create a disadvantage as the number of introns grew, because the mobility would be increasingly detrimental to the host. If evolution then favored the replacement of the RT with other cellular splicing factors, then any Group II introns that had not inserted within genes would be under no selective pressure to retain splicing, enabling them to evolve into non-LTR-retrotransposons *(38)*.

5.2. Group I Introns, Inteins, and Homing Endonucleases

Three other classes of molecular parasites are worth mentioning. We will briefly describe Group I introns, inteins, and homing endonucleases. Group I introns are found in both eukaryotic and bacteriophage systems *(41)* and are considered part of the mobilome because they can transfer their intron to intronless alleles of the same gene. These introns catalyze their own splicing through a series of reactions initiated by a guanosine and ending in a linear segment of RNA. They can vary in size from 200 to 3000 nucleotides. Group I introns can have fairly sophisticated secondary and tertiary structure that is

conserved in structure but not necessarily in sequence homology. They are workhorses in the field of ribozyme in vitro evolution studies *(42)*.

Inteins are segments of a protein that are able to excise and rejoin through a peptide bond. They usually contain an endonuclease domain that plays a role in their propagation. Inteins have been found in all three super domains of life and a nice review of the role evolution has played in the spread and maintenance of these parasites is provided by Gogarten and Hilario *(43)*. In brief, the intein can maintain a functional endonuclease over the long term despite original models which would indicate that once inteins were fixed in a population, the homing endonuclease would be lost through random processes. Certainly, inteins provide a robust platform for the consideration of how parasitic genes move and become fixed in a population.

Homing endonucleases *(44)* deserve a few sentences of mention because they are such unique entities. Basically, these are rare-cutting enzymes that are encoded by the inteins and introns. There are four families, characterized by sequence motifs LAGLIDADG, GIY-YIG, H-N-H and His-Cys. They recognize stretches of DNA that are usually 12–40 bases long. They are considered highly invasive because they take refuge in the molecular parasites mentioned above and are therefore important elements of the mobilome community.

6. Conclusions/ Outlook

Our ability to sequence whole genomes in a matter of weeks provides us with opportunities to formulate and plan for strategies to quantify the mobilome in those genomes. It is an exciting prospect, especially given that microbial ecology and technology have allowed researchers to move the sequencing targets out of the laboratory and into the field. With the recognition that different bacterial lineages seemed to have different propensities for HGT *(45)*, researchers are eager to define the rules and constraints that could predict or define the transfer and the agents that accomplish them. Strides have been made. For instance, the classification and quantification of incompatibility groups of plasmids in certain lineages seems to be reaching a quantifiable threshold (e.g., 26 in the enterobacteriaceae, 14 in the pseudomonads, and 18 in the Gram-positive staphylococci) *(46)*. The burgeoning field is not without issues and they are significant ones. Current analysis schemes are hampered by the lack of large comparable data sets of MGE sequences to make gene predictions possible. The repetitive sequence, so common a feature of many MGEs, is the sequencer's nightmare. The complexity

and intermixing of MGEs makes identification and categorization in large-scale and automated projects a daunting task. Physical separation of autonomous MGE elements in bacterial species so that sequencing strategies can be performed is quite problematic. Despite these hurdles, the future is bright, because to understand the mobilome means that we will get at the very heart of evolution and the adaptive process of biology. The mobilome is at our fingertips and our appreciation of the mobilome will continue to expand as more sequence data accumulate and clever schemes are devised to analyze it. There is no doubt that mobilome discovery will continue to reveal the exquisite artistry that evolution has produced in the agents of change of biology.

References

1. McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* **36**, 344–355.
2. Pace, N. R. (2006) Time for a change. *Nature* **441**, 289.
3. Guttman, D. S., Dykhuizen, D. E. (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**, 1380–1383.
4. Collier, L. S., Largaespada, D. A. (2007) Transposable elements and the dynamic somatic genome. *Genome Biol* **8 Suppl 1**, S5.
5. SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima,Y., Bennetzen, J. L. (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet*, **20**, 43–45.
6. Li, W., Zhang, P., Fellers, J. P., Friebe, B., Gill, B. S. (2004) Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J*, **40**, 500–511.
7. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
8. King, R. D., Whelan, K. E., Jones, F. M., Reiser, P. G., Bryant, C. H., Muggleton, S. H., Kell, D. B., Oliver, S. G. (2004) Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **427**, 247–252.
9. Copeland, C. S., Mann, V. H., Morales, M. E., Kalinna, B. H., Brindley, P. J. (2005) The Sinbad retrotransposon from the genome of the human blood fluke, Schistosoma mansoni, and the distribution of related Pao-like elements. *BMC Evol Biol* **5**, 20.
10. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O. et al. (2007) A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**, 973–982.
11. Weiner, A. M. (2002) SINEs and LINEs: the art of biting the hand that feeds you. *Curr Opin Cell Biol* **14**, 343–350.
12. Ray, D. A., Walker, J. A., Batzer, M. A. (2007) Mobile element-based forensic genomics. *Mutat Res* **616**, 24–33.
13. Hasler, J., Strub, K. (2006) Alu elements as regulators of gene expression. *Nucleic Acids Res* **34**, 5491–5497.
14. Colmegna, I., Garry, R. F. (2006) Role of endogenous retroviruses in autoimmune diseases. *Infect Dis Clin North Am* **20**, 913–929.
15. Miller, W. J., Hagemann, S., Reiter, E., Pinsker, W. (1992) P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*. *Proc Natl Acad Sci U S A* **89**, 4018–4022.
16. van Gent, D. C., Mizuuchi, K., Gellert, M. (1996) Similarities between initiation of V(D)J recombination and retroviral integration. *Science* **271**, 1592–1594.
17. Kapitonov, V. V., Jurka, J. (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* **3**, e181.
18. Orgel, L. E., Crick, F. H. (1980) Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607.
19. Bessereau, J. L. (2006) Transposons in *C. elegans*. *WormBook* 1–13.
20. Jacobson, J. W., Medhora, M. M., Hartl, D. L. (1986) Molecular structure of a somatically

unstable transposable element in Drosophila. *Proc Natl Acad Sci U S A* **83**, 8684–8688.
21. Plasterk, R. H. (1991) The origin of footprints of the Tc1 transposon of *Caenorhabditis elegans*. *EMBO J* **10**, 1919–1925.
22. Ivics, Z., Izsvak, Z. (2006) Transposons for gene therapy! *Curr Gene Ther* **6**, 593–607.
23. Siguier, P., Filee, J., Chandler, M. (2006) Insertion sequences in prokaryotic genomes. *Curr Opin Microbiol* **9**, 526–531.
24. Chandler, M., Mahillon, J. (2002) Insertion sequences revisited. In *Mobile DNA vol II.* (Craig, N. L., Craigie, R., Gellernt, M., Lambowitz, A. M., eds.), ASM Press, Washington D.C., pp. 305–366.
25. Filee, J., Siguier, P., Chandler, M. (2007) Insertion sequence diversity in archaea. *Microbiol Mol Biol Rev* **71**, 121–157.
26. Wagner, A. (2006) Cooperation is fleeting in the world of transposable elements. *PLoS Comput Biol* **2**, e162.
27. Lederberg, J. (1952) Cell genetics and hereditary symbiosis. *Physiol Rev* **32**, 403–430.
28. Velmurugan, S., Mehta, S., Uzri, D., Jayaram, M. (2003) Stable propagation of 'selfish' genetic elements. *J Biosci* **28**, 623–636.
29. Thomas, C. M., Nielsen, K. M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat Rev Microbiol* **3**, 711–721.
30. Saito, K., Yamazaki, M., Murakoshi, I. (1992) Transgenic medicinal plants: *Agrobacterium*-mediated foreign gene transfer and production of secondary metabolites. *J Nat Prod* **55**, 149–162.
31. Martinez-Cuesta, M. C., Requena, T., Pelaez, C. (2001) Use of a bacteriocin-producing transconjugant as starter in acceleration of cheese ripening. *Int J Food Microbiol* **70**, 79–88.
32. Udo, E. E., Jacob, L. E. (1998) Conjugative transfer of high-level mupirocin resistance and the mobilization of non-conjugative plasmids in *Staphylococcus aureus*. *Microb Drug Resist* **4**, 185–193.
33. Kunin, V., He, S., Warnecke, F., Peterson, S. B., Garcia, M. H., Haynes, M., Ivanova, N., Blackall, L. L., Breitbart, M., Rohwer, F., McMahon, K.D., Hugenholtz, P. (2008) A bacterial metapopulation adapts locally to phage predation despite global dispersal. *Genome Res* **18**, 293–297.
34. Canchaya, C., Fournous, G., Brussow, H. (2004) The impact of prophages on bacterial chromosomes. *Mol Microbiol* **53**, 9–18.
35. Faruque, S. M., Nair, G. B. (2002) Molecular ecology of toxigenic Vibrio cholerae. *Microbiol Immunol* **46**, 59–66.
36. Nam, K. T., Kim, D. W., Yoo, P. J., Chiang, C. Y., Meethong, N., Hammond, P. T., Chiang, Y. M., Belcher, A. M. (2006) Virus-enabled synthesis and assembly of nanowires for lithium ion battery electrodes. *Science* **312**, 885–888.
37. Lang, A. S., Beatty, J. T. (2007) Importance of widespread gene transfer agent genes in [alpha]-proteobacteria. *Trends Microbiol* **15**, 54–62.
38. Lambowitz, A. M., Zimmerly, S. (2004) Mobile group II introns. *Annu Rev Genet* **38**, 1–35.
39. Toor, N., Hausner, G., Zimmerly, S. (2001) Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7**, 1142–1152.
40. Belhocine, K., Plante, I., Cousineau, B. (2004) Conjugation mediates transfer of the Ll.LtrB group II intron between different bacterial species. *Mol Microbiol* **51**, 1459–1469.
41. Lambowitz, A. M., Belfort, M. (1993) Introns as Mobile Genetic Elements. *Annu Rev Biochem* **62**, 587–622.
42. Joyce, G. F. (2007) Forty years of in vitro evolution. *Angew Chem Int Ed Engl* **46**, 6420–6436.
43. Gogarten, J. P., Hilario, E. (2006) Inteins, introns, and homing endonucleases: recent revelations about the life cycle of parasitic genetic elements. *BMC Evol Biol* **6**, 94.
44. Gogarten, J. P., Olendzenski, L., Hilario, E., Simon, C., Holsinger, K. E. (1996) Dating the cenancester of organisms. *Science* **274**, 1750–1751.
45. Ochman, H., Lawrence, J. G., Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.
46. Frost, L.S., Leplae, R., Summers, A. O., Toussaint, A. (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Micro* **3**, 722–732.