# Mobile genetic elements and horizontal gene transfer

**Ke Ji**

**Cancer Genomics and Developmental Biology, Utrecht University**

**Contents**

It has been more than half century ago that transposition was first discovered in corn by Barbara McClintock [1,2]. Ever since, geneticists have found various genetic events at which some genetic elements are capable of moving around the genome actively or passively. These genetic elements now referred to as mobile genetic elements (MGEs) [3] interest research communities for their bizarre behaviors, when comparing with conventional genetic events such as transcription, translation, replication and recombination etc., and for their impacts on evolution study, biotechnology development and health care.

In this thesis, various broadly defined MGEs are reviewed with the focuses on their genomic mobile mechanisms and their influences to biology and medicine.

# MGEs

It is not straightforward to make a structured classification of different mobile genetic elements due to either the ambiguous connections between their molecular mechanisms/possible origins, or the overlap of physiological functions caused by recombination. For the convenience of description in this thesis, MGEs are categorized into two main groups according to whether they move within a genome (intra-genomic mobility) or cross genomes (inter-genomic mobility).

MGEs with chromosomal integrative property are the main interests of this thesis since this can lead to the permanent change of genome composition. As a general definition of MGE, viruses can be considered as one special type of MGEs. While, because of their physiological complication (involving capsids and envelops) and parasitic property, virus related MGEs, like those involved in transduction, are not covered in this thesis. Plasmids, found in all three domains: Archaea, Bacteria and Eukaryota, are extra-chromosomal DNA molecules which carry the genes enabling them to self-replicate. Plasmids can be transferred between cells through transformation or conjugation, which is considered as one of the major forms of horizontal gene transfer (HGT). The common example of conjugative plasmids is F-plasmids ("F" represents "fertility factor"). F-plasmids are capable of integrating into chromosomal DNA and carrying chromosomal DNA during conjugation. Because of the chromosome integration property, conjugative/F-plasmids are the focus of this thesis.

**Intra-genomic MGEs**

The main form of intra-genomic mobility is transposition, which is defined as movement of DNA fragments (transposable elements/transposons) within the genome using non-sequence-homologous mechanisms [3]. Transposons are found ubiquitously in both prokaryotic and eukaryotic organisms. Transposon-derived sequences contribute to 1-5% fraction of prokaryote and lower eukaryote genomes. In higher eukaryotes, this proportion can even reach 40% or more [4]. One traditional way to categorize different transposons is based on the transposition intermediate. Under such classification, transposons can be assigned to Class I (retrotransposons) or Class II (DNA transposons) groups depending on whether a RNA intermediate is involved during transposition [5,6,7]. However this is a rough way to classify transposons without considering the detailed transposition mechanisms. A more informative categorizing way is based on the transposase encoded by the

transposon (see Figure 1), which characterizes the main molecular nature of each different transposition mechanism [4].
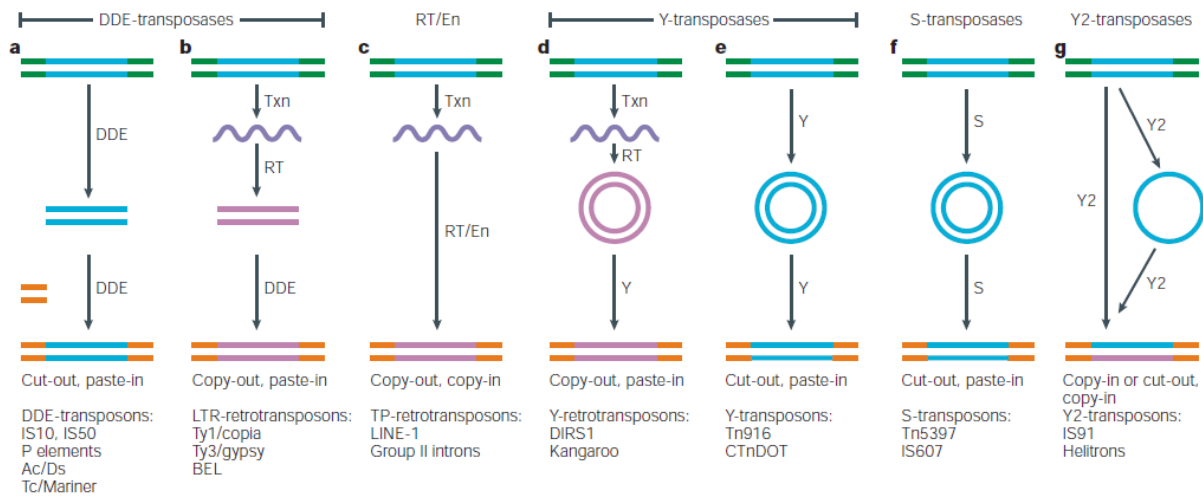


**Figure 1**. A simplified scheme of different types of transposons based on their transposases [4]. Five protein families dictate different transposition pathways: DDE-transposases, reverse transcriptase/endonucleases (RT/En), tyrosine (Y)-transposases, serine (S)-transposases and rolling-circle (RC)- or Y2-transposases. Transposons (blue) can be either 'cut-out' or 'copied-out' of the flanking donor DNA (green). Representatives of each type of transposon are listed below each pathway.

*DDE-transposases*. DDE-transposases feature a DDE (aspartic acid -- aspartic acid -- glutamic acid residues) motif at the catalytic domain of the tansposase. There are 2 main different transposition modes identified among the transposons using DDE-tranposases. In one mode, tranposons are excised from donor DNA and then inserted into target DNA. In the other mode, transposons use their encoded reverse transcriptases to generate a DNA copy from their full length transcripts, and then insert this copy into target DNA using DDE- tranposases.

*Y-transposases*. Y-transposases use nucleophilic tyrosine residue to attack DNA during excision and insertion of the transposons. Another feature of Y-transposases involved transpositions is the formation of a circular DNA intermediate before insertion into target sites. There are also 2 different transposition modes identified among the transposons using Y-tranposases. In one mode, excised transposon DNA is the intermediate. In the other mode, transposons are first transcribed into RNA, then reverse transcribed into DNA, circled and insert into target DNA.

*S-transposases*. S-transposases work in a slightly similar way as Y-transposases, in which nucleophilic serine, instead of tyrosine, residue is used to attack donor and target DNA. And the circular DNA intermediate is also involved in S-transposases mediated transpositions. No S-(retro)transposons have been recorded yet. Y- and S-transposases are close related to sites-specific (through homologous sites) Y- and S-recombinases. However Y- and S-transposases do not need to perform site-specific insertion even though their excision and integration mechanisms are likely to be identical to their site-specific relatives. An interesting thing is that Y- and S-transposons contribute most of so called conjugative tansposons [8,9], which encode enzymes enable transposons to be transferred not only within the same cell (intra-cellular mobility) but also from a donor to a recipient cell (inter-cellular mobility).

***Y2-transposases***. Y2-transposases feature a motif with 2 tyrosines separated by three other residues. During transposition, only one strand of the transposon DNA is excised from the donor site. This single strand DNA is inserted into one of the two strands at the target site. Then complementary strand is synthesized through passive DNA replication.

***TP- or RT/En-transposases***. TP-transposases (TP stands for target-primed) encode an enzyme which combines reverse transcriptase and endonuclease (RT/En) activities. During transposition, TP-transposases use TP-transposons' transcripts as templates to synthesize the DNA copies directly into the nicked target sites.

**Inter-genomic MGEs**

The process of inter-genomic movement of genetic materials is also referred to horizontal gene transfer (HGT) or lateral gene transfer (LGT), which contrasts to vertical gene transfer of inheriting genetic information from the ancestor or parent of a species [3]. HGTs can occur within the same species or between different species (even cross Domain). The variety of inter-genomic MGEs is much more diverse than intra-genomic MGEs. Some inter-genomic MGEs can be well characterized for the functional components which encode the transfer machineries. While some other inter-genomic MGEs can only be recognized as the remains of ancient transfer events.

Prokaryote — Prokaryote

HGT is a quite common and significant way for prokaryotic organisms to obtain new genetic traits besides endogenous mutation and recombination. Prokaryotic HGTs are discovered between bacteria [10], between archaea [11], and between bacteria and archaea [12]. Just as the whole archaea research lags behind bacterial studies, detailed knowledge about prokaryotic HGT mechanisms are most acquired from bacteria. The three main ways of prokaryotic HGT are transformation, conjugation and transduction (not covered in this thesis) [3,13,14]. Endosymbiotic gene transfer refers to the genetic material transfer between organelle (mitochondria and plastid) genome and nuclear genome of eukaryotic cells. The reason of placing this special content in the prokaryote-to-prokaryote HGT section is that the current popular hypothesis suggests: when endosymbiotic gene transfer started billions of years ago, these organelles (mitochondria and plastid) and the host cell were ancient (endo)symbiotic bacteria.

***Transformation***. Transformation is a process of the acquisition of naked DNA from extracellular environment into cytoplasm followed by genomic integration (not in all cases). Not all prokaryotes are capable of transformation at any time. Natural transformation happens when cells enter into a transient physiological state called competent. The specific triggers turning cells into competent are not clear for many species. The naked DNA from extracellular environment can be chromosomal DNA, plasmid DNA or viral DNA. They are derived from dead prokaryotic/eukaryotic cells or viral particles. The size of uptake DNA can range from few hundred nucleotides to more than 55,000 nt. Most time, only one strand of the uptake DNA is transported into cytoplasm, while the other strand is degraded outside cytoplasm [14]. In bacteria, transformation starts with the binding of uptake DNA onto DNA receptor proteins of competent cells. Then the transport machinery employs a cell-

envelope-spanning structure, which consists of type IV pilus and and type II secretion systems, and its coupled cytoplasmic membrane DNA translocation complex to import the single stranded uptake DNA into cytoplasm. The cytoplasmic membrane DNA translocation complex includes DNA receptor protein, channel protein and ATP-binding protein [15]. The imported single stranded DNA can be integrated into the host genome through homologous recombination or through rare illegitimate recombination. The homologous recombination relies on RecA protein and its analogs, while the illegitimate recombination does not. Other proteins involved in the recombination process include helicases, DNases, DNA binding proteins, DNA polymerase and ligase. If the imported single stranded DNA is derived from plasmid, it can go on plasmid reconstitution [14,15].

*Conjugation*. Conjugation is a process of the transfer of DNA from a donor to a recipient cell through cell-to-cell contact. The DNA transferred by conjugation can be conjugative plasmids, conjugative transposons (as mentioned above), or the recent years discovered varied MGEs which are capable of similar DNA transfer processes as conjugative plasmids and transposons [14]. All together, this broad group of MGEs is called integrative and conjugative elements (ICEs) (see Figure 2). The size of ICEs can range from 30 Kbp (kilo base pairs) to several hundred Kbp. ICEs have an extrachromosomal circular phase at certain time of their life cycle. ICEs can integrate into host chromosomes, and normally a single stranded form is generated during their conjugative transfer [16]. The knowledge about conjugation mechanism was initially obtained from studying F-plasmid conjugation processes. F-plasmids carry *tra* and *trb* loci, which together encode genes related to conjugation apparatus formation, cell surface contact and cell membrane fusion. F-plasmids contain an origin of transfer (*oriT*), which is nicked by the enzyme relaxase to initiate the single strand formation during conjugation. The linear single stranded DNA is reconstituted into double-stranded circular form after being transferred to donor cell. F-plasmids also contain their own origin of replication (*oriV*) enabling them autonomous replication. In some occasion, F-plasmids are found integrated into host genome by homologous recombination [17,18]. For non-conjugative-transposon ICEs, it is still unclear whether they can be maintained extrachromosomally through autonomous replication. It is known that they can integrate into host genome by homologous recombination (while conjugative transposons can insert into various genomic locations without the need of sequence similarity), and replicate together with the chromosome. As F-plasmids, non-conjugative-transposon ICEs also contain *oriT* for the initiation of the conjugative DNA processing. Genes similar but not identical to *tra* and *trb* loci genes are discovered in non-conjugative-transposon ICEs, which constitute a diverse range of conjugation mechanisms [16].
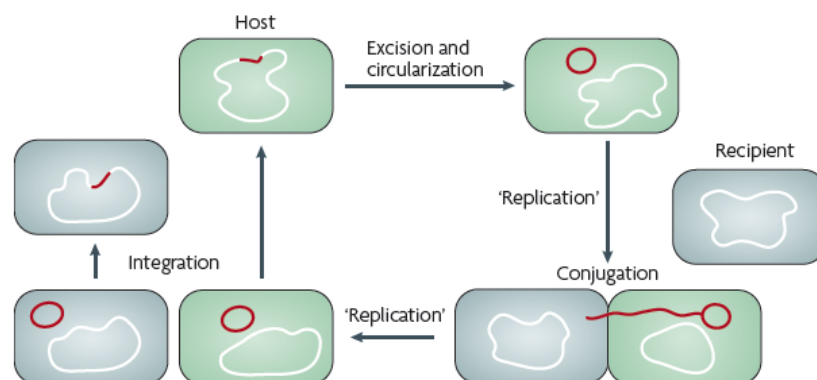


**Figure 2**. Schematic of a typical integrative and conjugative element life cycle [16]. An integrative and conjugative element (ICE) is integrated into one site in the hostchromosome and is bounded by specific

sequences on the right (attR) and left (attL). Excision yields a covalently closed circular molecule as a result of recombination between attL and attR to yield attP (in the ICE) and attB (in the host chromosome). An ICE-free cell can serve as a potential recipient. During conjugation, the donor and recipient are brought in close contact, and a single DNA strand is transferred to the new host through the action of rolling circle replication. Following transfer, DNA polymerase in the recipient synthesizes the complementary strand to regenerate the double-stranded, circular form. A recombination event between attP and attB results in integration into the host chromosome.

***Endosymbiotic gene transfer***. The widely accepted endosymbiotic theory suggests that, during eukaryote evolution, mitochondria, plastids, and possibly other organelles arose from the engulfment of free-living bacteria, proteobacteria as mitochondria precursors and cyanobacteria as chloroplast precursors for example, by the prokaryotic ancestors of eukaryotes (see Figure 3). One phenomenon accompanying organellar endosymbiosis is the reduction of organellar genome size. Today the known mitochondria and plastid genomes contain only several up to a few hundred of genes, which are far smaller than either nucleus genomes or modern prokaryote genomes. Some organellar genes vanished probably because the host nucleus orthologous genes have replaced the functions. While more organellar genome reductions are probably due to the endosymbiotic gene transfer of organellar genes into host genomes. It is evident that the majority of genes involved in the biological functions of mitochondria and plastids are now encoded at nuclear genomes. It is hard to evaluate the ancient transfer events since the organellar gene copies have disappeared. Evolutionarily recent transfers can still be identified by finding homologous gene copies located at both organellar and nuclear genomes [19]. Endosymbiotic gene transfer is highly oriented with the direction from organelle to nucleus. However, exceptions are discovered, in which copies of genes/genetic fragments were transferred from nucleus/plastid into mitochondria during recent transfer events, especially in plants [20]. A functional endosymbiotic gene transfer is a complicated process. To succeed, organellar genes/gene copies need to move out of organelles and integrate into the nuclear genome at proper locations without interrupting irreplaceable crucial nuclear genomic functions. The integrated genes also need to undergo certain mutations/modifications before being correctly regulated and expressed under the extraorganelle cytoplasmic system. Some integrated genes need to acquire the information for encoding transit peptides, so that their protein products can be transported to organelles to fulfill their biological functions. The detailed mechanisms of endosymbiotic gene transfer are yet unknown. There are two different theories, among the research community, about the general endosymbiotic gene transfer strategies termed "bulk DNA" and "cDNA intermediates". The former suggests that organellar genomic DNA, whether coding or non-coding, is the direct transfer intermediate. While the latter proposes that cDNA of organellar mRNA is used as the vehicle of endosymbiotic gene transfer [19].
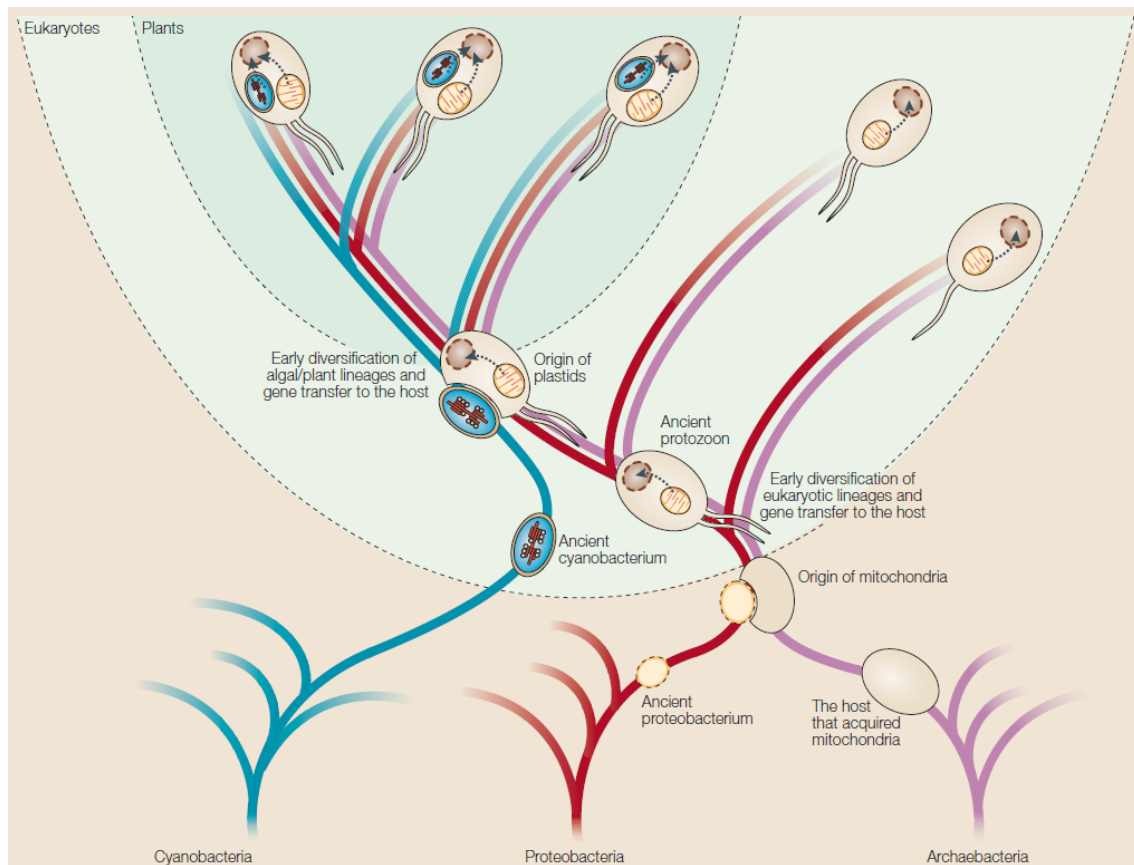
**Figure 3**. Endosymbiotic evolution and the tree of genomes [19]. Intracellular endosymbionts that originally descended from free-living prokaryotes have been important in the evolution of eukaryotes by giving rise to two cytoplasmic organelles.Mitochondria arose from α-proteobacteria and chloroplasts arose from cyanobacteria. Both organelles have made substantial contributions to the complement of genes that are found in eukaryotic nuclei today. The figure shows a schematic diagram of the evolution of eukaryotes, highlighting the incorporation of mitochondria and chloroplasts into the eukaryotic lineage through endosymbiosis and the subsequent co-evolution of the nuclear and organelle genomes. The host that acquired plastids probably possessed two flagella. The nature of the host cell that acquired the mitochondrion (lower right) is fiercely debated among cell evolutionists. The host is generally accepted by most to have an affinity to ARCHAEBACTERIA but beyond that, biologists cannot agree as to the nature of its intracellular organization (prokaryotic, eukaryotic or intermediate), its age, its biochemical lifestyle or how many and what kind of genes it possessed. The host is usually assumed to have been unicellular and to have lacked mitochondria.

Prokaryote — Eukaryote

In literature, most abundantly presented HGTs between prokaryotes and eukaryotes are the transfer of bacterium-origin genes into the genomes of unicellular eukaryotic organisms – protists [21,22]. DNA uptake processes like those in natural transformation and conjugation are rarely documented in eukaryotes. And no inter-domain transduction/transfection is yet known. Therefore one favorite way of bacterial DNA uptake by protists might be phagocytosis, which is supported by the fact: many protists eat bacteria as nutrient sources. However not all lineages of the protists, which are identified with HGT transferred bacterial origin genes, have the phagotrophic lifestyle. In these cases, the possible ways of bacterial DNA uptake by the protists could be through physical contact of

symbiotic/parasitic relationships or unknown inter-domain viral transduction/transfection [21]. It is hard to know the mechanisms with which bacterial genes survive cellular digestion/degradation and integrate into protist genomes. Prokaryote-to-eukaryote HGT is also found in the unicellular fungus *Saccharomyces cerevisiae* [23]. During the screen, two candidate HGT genes were identified in the *S. cerevisiae* genome: a dihydroorotate dehydrogenase gene was likely derived from a lactic acid bacterium and replaced the eukaryotic endogenous copy; an aryl- and alkyl-sulfatase was likely transferred from some alpha-proteobacteria. The authors suggested bacterium-to-fungus conjugation and natural transformation as possible mechanisms of the acquisition of bacterial DNA by the yeast. However there is no evidence supporting such theory. One interesting phenomenon observed during the screen is that foreign DNA is often found integrated near the yeast chromosome telomeres, which might offer some clues about the integration mechanism.

The examples of HGTs from prokaryotes to multicellular eukaryotes are sparse. The best studied case is *Agrobacterium*. This genus of bacteria is well known to cause tumors in plants by genetically transforming the genomes of targeted plants [24,25]. Not all strains of *Agrobacterium* are virulent. The virulence comes from the possession of tumour-inducing (Ti) or root-inducing (Ri) plasmid. Ti plasmid contains a segment called transfer DNA (T-DNA) which encodes the genes for transforming the infected plants. Ti plasmid also encodes virulence factor genes (*vir* genes) whose products facilitate the transfer and integration of T-DNA into plant genome. The process of T-DNA transfer is to some extent similar to conjugation. However only the single-stranded T-DNA not the whole Ti plasmid is transferred into plant cells. The transferred T-DNA is most likely integrated into plant chromosomes through illegitimate recombination or non-homologous end-joining (see Figure 4). T-DNA genes are then expressed in plant cells to produce enzymes synthesizing plant hormones auxin and cytokinin for plant tumor induction, and opines used by the *Agrobacterium* as carbon and nitrogen sources but not other microorganisms [26]. Astonishingly, *Agrobacterium* was shown to have the ability to genetically transform cultured human cells [27]. And even one human infection case was reported related to *Agrobacterium* [28]. Another evident example of HGTs from prokaryotes to multicellular eukaryotes is intracellular parasitic bacteria *Wolbachia*. *Wolbachia* genus bacteria are known to widely infect arthropod species including insects, spiders, mites and isopod species, plus many species of filarial nematodes. It was found that *Wolbachia* genome fragments were transferred into some host genomes ranging from nearly the entire *Wolbachia* genome (>1 megabase) to short (<500 base pairs) insertions [29]. The detailed mechanism of this intracellular parasite to host cell HGT is not yet known. It may or may not be similar to the organellar endosymbiotic gene transfer. The infection of *Wolbachia* causes various physiological changes in hosts like male killing, feminization, parthenogenesis, cytoplasmic incompatibility, viral resistance, and pathogenicity of filarial nematodes. So it will be interesting to check the consequence of transferred *Wolbachia* genes on host cells. Besides these two symbiosis-related examples of HGTs from prokaryotes to multicellular eukaryotes, there are reports of possible rhizobial (soil nitrogen fixing bacteria) genes transfer into plant-parasitic nematodes [30], and of bacterial cellulose synthase gene transfer into ascidians (urochordate) [31,32] which are the only animals known to produce cellulose in the whole metazoan kingdom.
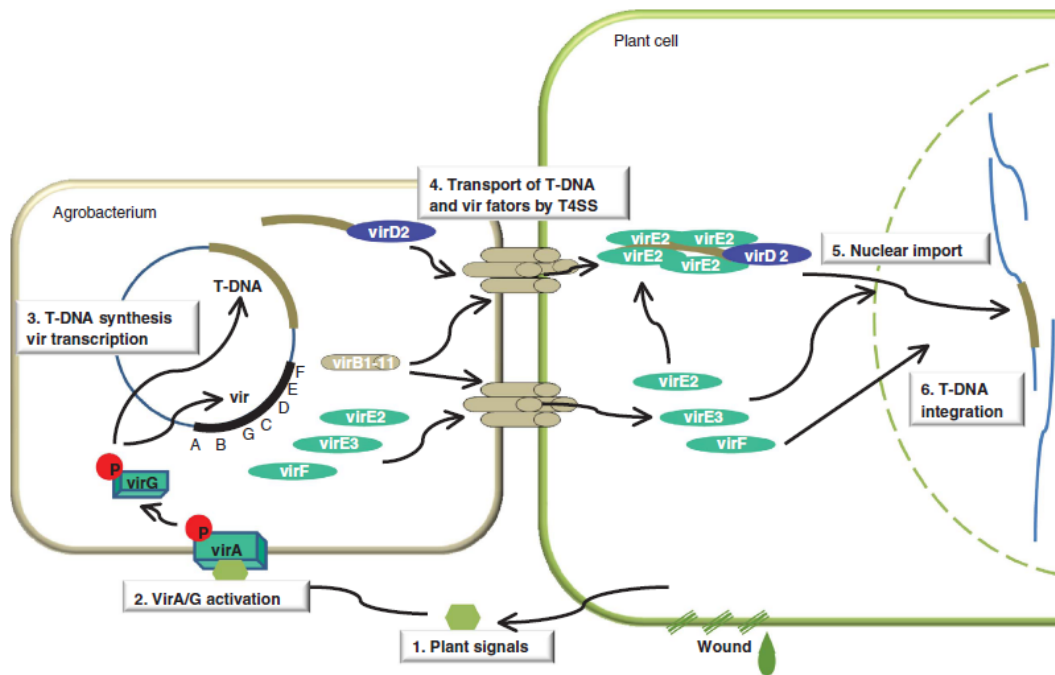
**Figure 4**. Overview of the Agrobacterium–plant interaction [26]. 1. Plant signals induce 2. VirA/G activation and thereby 3. T-DNA synthesis and vir gene expression in Agrobacterium. 4. Through a bacterial type IV secretion system (T4SS) T-DNA and Vir proteins are transferred into the plant cell to assemble a T-DNA/Vir protein complex. 5. The T-DNA complex is imported into the host cell nucleus in which 6. the T-DNA becomes integrated into the host chromosomes by illegitimate recombination.

HGTs between prokaryotes and eukaryotes favor the direction from prokaryotes to eukaryotes. Eukaryote-to-prokaryote HGTs are rare. On one hand, this phenomenon could be deceptive due to the observation bias caused by the uneven distribution of the amount of sequenced genomes between prokaryotes and eukaryotes. On the other hand, this phenomenon could also be biologically true: introns are barriers for transferred eukaryotic genes to be correctly expressed in prokaryotes; eukaryotic genomes are less diverse to offer prokaryotes fitness advantages; there is a far larger population of prokaryotes than the around eukaryotes in nature which statistically favors prokaryotes as the potential donors of the transferred genes [21,22]. Most frequently reported eukaryote-to-prokaryote HGTs include the presence of eukaryotic signal transduction system domains in prokaryotic genomes [33]; the transfer and replacement of aminoacyl-tRNA synthetases (aaRSs) from eukaryotes to prokaryotes [34]. But the most evident eukaryote-to-prokaryote HGTs are the discovery of eukaryote-like tubulins and actins (not homologous to bacterial cytoskeleton) in some bacterial strains [35]; the appearance of eukaryote-derived deoxyribose/fructose phosphate aldolase genes in some bacteria [36].

Eukaryote — Eukaryote

As mentioned before, prokaryotic HGT mechanisms such as transformation and conjugation are rare in eukaryotes, plus viral host-specificity is quite high, thus all these facts make the occurrence of eukaryote-to-eukaryote HGTs more mysterious. Maybe the most explainable example is from the chlorarachniophyte *Bigelowiella natans* (a protist from the Rhizaria group). The majority of the

nuclear genes encoding plastid-targeted proteins in *B. natans* have a chlorophyta green algal origin. But still a quite portion of the genes (≈21%) are more homologous to those from streptophyta green algae, red algae, and as well as bacteria [37]. Chlorarachniophytes are mixotrophic organisms capable of both photosynthesis and phagocytosis. This may explain that the acquisition of the other algae genes besides chlorophyta algal genes is probably due to the ingestion of these algae after the establishment of chlorophyta plastid during *B. natans* evolution history.

In plants, the first reported eukaryote-to-eukaryote nuclear HGT is related to transposons. Some Mu-like elements (MULEs) from genus *Setaria* are found highly similar to a small family of MULEs from rice genus *Oryza*. *Setaria* (Subfamily *Panicoideae*) and *Oryza* (Subfamily *Bambusoideae*) are from family *Poaceae* [38]. For the possible explanation of this HGT, the authors mentioned in the paper that: both rice and *Setaria* are obligate self-fertilizers, but both could outcross at a low frequency. However there is no evidence that these rather distant subfamilies can intercross. Another case of plant HGT was detected in the parasitic plant *Striga hermonthica*. A gene named *ShContig9483* (function unknown) from the eudicot *S. hermonthica* shows high similarity to genes in the monocots sorghum and rice but has no homologs in other eudicots [39]. *S. hermonthica* is known to infest monocots such as sorghum and rice by forming an invasive organ haustorium, which interconnects their vasculature with that of their hosts in order to extract nutrients, water, and even sometimes accidently host mRNAs. The authors thus suggested that: one possibility is that *ShContig9483* was originally transferred to *S. hermonthica* as mRNA or cDNA.

There are also interesting cases of HGTs in animals. *Hydra magnipapillata* (phylum *Cnidaria*) was found containing a gene homologous to the *flp* genes (function unclear) of the parabasalid *Trichomonas vaginalis* (a parasitic protist). This indicates a possible HGT between a *Trichomonas vaginalis*-like (genetically) protist and hydra even though the transfer direction cannot be determined [40]. A more devastating case is the acquisition of photosynthesis by the sea slug *Elysia chlorotica*. *E. chlorotica* is long known capable of acquiring plastids from their food source *Vaucheria litorea* (a yellow-green algae). These *V. litorea* plastids are sequestered in the sea slug's digestive epithelium cells, where they continue to photosynthesize for months (see Figure 5). This phenomenon leads to the suspicion that *E. chlorotica*, through HGT, obtains and replaces the original *V. litorea* nuclear-coded plastid-targeted proteins so that the plastid metabolism can be carried on without *V. litorea* cytoplasm. After genome screen, a *V. litorea* nuclear gene of oxygenic photosynthesis (*psbO*) was found integrated into *E. chlorotica* germline genome and expressed in the sea slug [41]. Another similar striking case of animal HGT is the discovery that aphids are able to synthesize carotenoids in addition to acquisition from their diet, which makes them the only known animals making their own carotenoids. The genome screen of pea aphid *Acyrthosiphon pisum* revealed multiple fungal-derived enzymes for carotenoid biosynthesis [42].
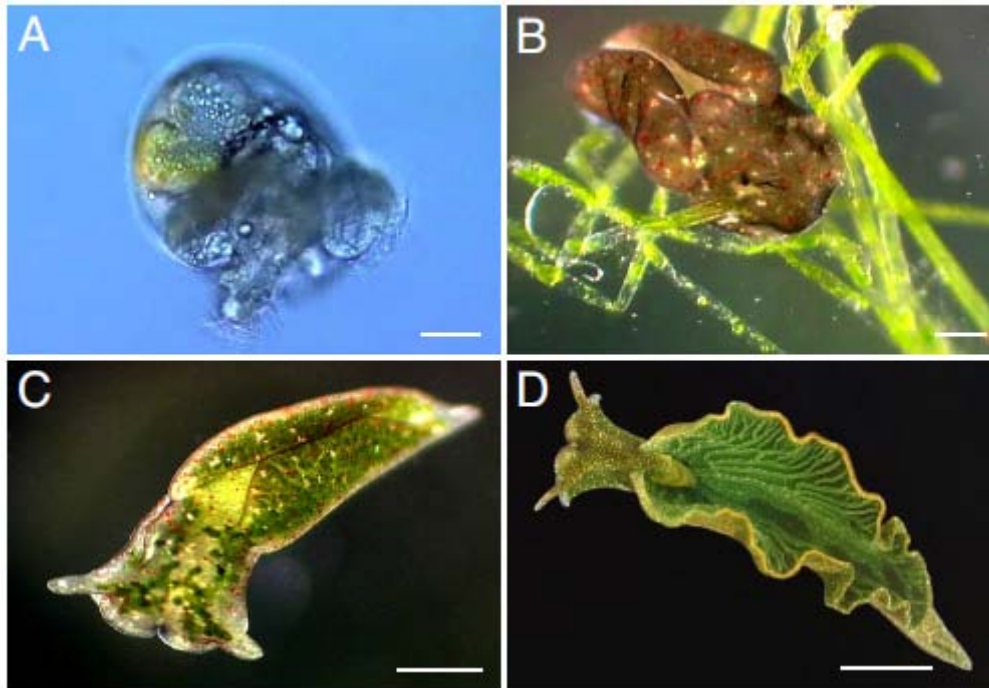
**Figure 5**. Laboratory culturing of E. chlorotica [41]. (A) Free-swimming E. chlorotica veliger larvae. (Scale bar, 100 µm.) Under laboratory conditions, the veliger larvae develop and emerge from plastid-free sea slug–fertilized eggs within approximately 7 days. The green coloring in the digestive gut is attributable to planktonic feeding and not to the acquisition of plastids at this stage. Metamorphosis of the larvae to juvenile sea slugs requires the presence of filaments of V. litorea. (B) Metamorphosed juvenile sea slug feeding for the first time on V. litorea. (Scale bar, 500 µm.) The grayish-brown juveniles lose their shell, and there is an obligate requirement for plastid acquisition for continued development. This is fulfilled by the voracious feeding of the juveniles on filaments of V. litorea. (C) Young adult sea slug 5 days after first feeding. (Scale bar, 500 µm.) By a mechanism not yet understood, the sea slugs selectively retain only the plastids in cells that line their highly branched digestive tract. (D) Adult sea slug. (Scale bar, 500 µm.) As the sea slugs further develop and grow in size, the expanding digestive diverticuli spread the plastids throughout the entire body of the mollusc, yielding a uniform green coloring. From these controlled rearing studies, we were able to conclude that the only source of plastids in our experimental sea slugs was V. litorea.

# MGE biology

### MGE identification

Unlike plasmid or viral DNAs, most integrative MGEs, as those mentioned in this thesis, do not abundantly exist in physically independent forms, and some integrated MGEs do not have isolate forms anymore when referring to those involved in ancient transfer events. Therefore they can only be sequenced and analyzed together with the host sequences (chromosomal DNAs), at most time, through bioinformatic approaches.

For intra-genomic MGEs, there are several methods developed to identify transposons. The repeat-discovery methods utilize the principle of detection of pairs of similar sequences (one of transposon

features) within the investigated genome, and then clustering these pairs into different families (transposon groups). One challenge of this approach is to filter out non-transposon repeats, which include chromosomal segmental duplications, tandem repeats and satellites. The homology-based methods detect transposons by homology search of the target genome with previous known transposon-coding sequences. The structure-based methods are designed to find the characteristic motifs of different transposons, or the nucleotide composition of transposon ORFs which often differs from that of genes in the same host genome. The comparative genomic methods locate transposon positions by pinpointing the insertion regions (caused by transposition) found during multiple alignments of orthologous genome sequences. Last, besides above de novo approaches, continuous expanding transposon reference set facilitates the fast screen of defined types of transposons cross genome data [43].

For inter-genomic MGEs, phylogenetic methods are nowadays the most promising ways to identify HGTs. The key of this type of approach is to find the phylogenetic incongruence between the (mostly) rRNA-based taxonomic phylogeny and the phylogeny of interested gene. This is the most robust method for ancient HGT detection. Bias will appear in this type of approach due to data incompleteness, when comparing to the relative abundance of rRNA databases, of interested genes, or due to incorrect construction of phylogenetic tree. Like for transposon detection, nucleotide composition patterns are also used to identify HGTs because of the existence of the pattern differences between donor and receipt sequences of HGTs. These patterns include base composition, G+C content, purine-to-pyrimidine ratio, oligonucleotide frequency and codon usage. However nucleotide composition intra-genome regional variation and pattern difference reduction of ancient HGTs will cause difficulties in detection. Homology search such as BLAST is the simplest and fastest way to detect HGTs. The drawback of this approach is being crude and inaccurate [13,21,22,33,44].

**MGE genomics**

Despite long history of the discovery and awareness of its importance in evolution, biodiversity, biohazard and etc., less attention has been paid to MGEs by research communities relatively. The infrastructure of MGE genomics is much imperfect than of conventional genomics. Very few independent genome projects have been dedicated to systematically screen MGEs cross genomes. Most time, MGEs are uncovered just as by-products of chromosomal genome sequencing projects. For MGE bioinformatics, most MGEs are poorly annotated. Besides the intrinsic difficulty of MGE detection and annotation, less effort has been put to develop specific bioinformatic tools. Without modifications, common gene annotation tools are not suitable for MGE annotation for reasons like lack of reference entries, hard to generate training sets and etc. Consequently, most MGE annotation is performed manually. As mentioned before, the ambiguous mosaic structure feature of MGEs make their classification and nomenclature chaos. There is no international standard or agreement concerning establishment of a systemic classifying and naming frame for MGEs. The similar state also exists for MGE gene ontology [3]. The CLAssification of Mobile genetic Elements (ACLAME) is probably currently the only specific database available for MGEs, which aims at building a comprehensive classification of the functional modules of MGEs at the protein, gene, and higher levels [45].

### MGEs in evolution

MGEs are extremely interesting for evolutionary biology. It has been a big transition among research communities, from considering the appearances of MGEs are trivial genetic incidences, to believing they are the main driving force behind life evolution. From molecular point of view, it is apparent that MGEs can produce much more new genetic information within a short period of time than point mutation does. Intra-genomic transposable elements cause genome rearrangement which creates potential opportunities for genome innovation; HGTs introduce new traits such as metabolic properties, virulence attribute, antibiotic resistance and etc. One hypothesis suggests: HGTs were extensive during the early stage of life evolution; such genetic material exchanges probably had started between pre-cellular macromolecular complexes even before the emergence of cellular life form, and had kept as common genetic events throughout primitive cell history; only when cell design complexity reached a critical point, so called "Darwinian Threshold", vertical genetic information transfer started to play a more important role, and this is also the time when "species" appeared [46]. In recent years, genome screens show that HGTs still commonly exist in prokaryotes, however much less in eukaryotes. This phenomenon logically reflects the idea of above mentioned hypothesis when considering the cell complexity difference between prokaryote and eukaryote.

# MGE health issues

### Mutagenesis

MGEs whether transposons, transformation integration elements, or ICEs are all potential mutagens. They are two sides of the same coin. On one hand, they can bring new traits for evolutionary competitive advantages; on the other hand, they can also damage the genome of host cell. When MGEs are inserted into a functional gene, most likely its function will be disrupted. When MGEs are excised out of a gene, incorrect repair of the resulting gap will ruin the gene structure. Some MGEs carry their own promoters, which can cause abnormal expression of adjacent host genes around the insertion site. Moreover multiple copies of the same MGE sequence in a genome, as for some transposons, can hinder the precision of chromosomal pairing during mitosis or meiosis in eukaryotes, resulting in genome abnormality such as chromosome duplication.

### Virulence

Genes coding virulent factors can be transferred from virulent strains to avirulent strains turning the latter pathogenic. Virulence transfer caused by HGT was first documented in a 1951 publication [47]. Victor J. Freeman described in this research paper the change of non-virulent strains of *Corynebacterium diphtheriae* into virulent strains though phage infection. The most recent reported example of possible virulence HGT is the 2011 outbreak of *Escherichia coli* (Strain TY2482) serotype O104:H4, which is at this moment causing serious health and food safety concern worldwide and especially in Europe. This particular enterohaemorrhagic *E. coli* (EHEC) strain can cause abdominal

pain, bloody diarrhea or haemolytic uraemic syndrome (HUS) which can lead to kidney failure (from WHO website). The Beijing Genomics Institute (BGI) sequenced and assembled the genome of this strain. The initial analyses on this draft genome showed the highly similarity, based on an identical Multi Locus Sequence Typing (MLST) on 7 important housekeeping genes, with two other *E. coli* strains: Strain 01-09591 (genome not available) isolated from Germany in 2001 and Strain 55989 (genome available) isolated from Central African Republic in 2002. The 2001 Germany strain also has identical profiles for 12 virulence/fitness genes as the 2011 outbreak strain. The 2002 Central African Republic strain, known to cause serious diarrhea, shares 93% genome sequence similarity with the 2011 outbreak strain, but does not contain the *Shiga* toxin gene, which encodes the toxin damaging blood cells (from BGI website). Based on current evidence, the 2001 Germany strain is most likely the ancestor of the 2011 outbreak strain even though the additional genomic information is lacking. Otherwise the 2011 outbreak strain could probably have derived from a the-2002-African-strain-like *E. coli*, which obtained virulent genes such as the *Shiga* toxin gene through HGT.


**Antibiotic resistance**

The abusive application of antibiotics since the 20th century has produced the disastrous consequence for medical practice and health care, which is the facilitation of evolution and spread of antibiotic resistant pathogenic strains. Antibiotic resistance of microorganisms is either established by point mutation or obtained through HGT. The first demonstrated antibiotic resistance caused by HGT is the discovery of multiple drug resistance transfer between *Shigellae* and *Escherichia coli* in the end of 1950s [48]. The most famous example of antibiotic resistance HGT is probably multidrug-resistant *Staphylococcus aureus* (MRSA), which refers to any strain of *Staphylococcus aureus* that has developed multiresistance to a broad range of antibiotics. MRSA is referred as a "superbug" and hospital troublesome because of the difficulty in treatment. The first penicillin-resistant MRSA was isolated in 1942; the first methiccillin-resistant MRSA strains was isolated in 1961; the first vancomycin-resistant MRSA was found in 2002; the first oxazolidinone/linezolid-resistance MRSA was reported in 2003. MRSA was also found resistant to other antibiotics such as dicloxacillin, nafcillin, oxacillin, cephalosporin, tetracycline and erythromycin [49]. Although features like short generation times and large population sizes boost the evolving rate of microorganisms, given the fact that MRSA has acquired plentiful new traits of antibiotic resistance within a relatively short period of time, it is highly expected that HGT has also made a contribution. The discovered evidences supporting this theory include: the transfer of methicillin-susceptible *Staphylococcus aureus* into methicillin-resistant *Staphylococcus aureus* by acquisition of a mobile genetic element called Staphylococcal Cassette Chromosome mec (SCCmec), which contains methicillin resistance gene *mecA*, from *Staphylocccus epidermidis* [50]; the development of vancomycin resistance by MRSA through obtaining a mobile genetic element Tn1546, which contains vancomycin resistance gene cluster *vanA*, from *Enterococcus faecalis* [51]. A recent new case of antibiotic resistance HGT is associated with New Delhi metallo-beta-lactamase-1 (NDM-1), which is a carbapenemase capable of inactivating carbapenem antibiotics. Carbapenems are a class of beta-lactam antibiotics with high resistance to most beta-lactamases, so that they were considered as one of the last resort antibiotics for many antibiotic-resistant bacterial infections. Now the situation will change since the appearance of NDM-1. The NDM-1 coding gene $bla_{NDM-1}$ is found mostly in plasmids, so it is easy to be transferred between bacteria like *Escherichia coli* and *Klebsiella pneumonia*. If the spread of NDM-1 cannot be

controlled, it will be worldwide catastrophe since there are currently very few antibiotics available to combat bacteria resistant to carbapenems [52].

# Conclusion

As described in this thesis, MGEs play important roles in earth life cycle. They make us reconsider the way life evolves and the definition of what a species is. Not mentioned are their application potentials in biotechnology. They also remind us the dangerous consequences when ignoring their existence in practice such as drug/antibiotic usage. In spite of these undebatable facts, the current research infrastructure and society attention on MGEs do not match their importance. Much more effort is needed to better understand their unique natures and their huge influences to the world.

**References**

1. Mcclintock B (1948) Mutable Loci in Maize. Carnegie Institute of Washington Year Book 47: 155-169.
2. Mcclintock B (1950) The Origin and Behavior of Mutable Loci in Maize. Proceedings of the National Academy of Sciences of the United States of America 36: 344-355.
3. Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol 3: 722-732.
4. Curcio MJ, Derbyshire KM (2003) The outs and ins of transposition: From mu to kangaroo. Nature Reviews Molecular Cell Biology 4: 865-877.
5. Grindley NDF, Leschziner AE (1995) DNA transposition: From a black box to a color monitor. Cell 83: 1063-1066.
6. Plasterk RHA (1993) Molecular Mechanisms of Transposition and Its Control. Cell 74: 781-786.
7. Haniford DB, Chaconas G (1992) Mechanistic aspects of DNA transposition. Curr Opin Genet Dev 2: 698-704.
8. Scott JR (1992) Sex and the Single Circle - Conjugative Transposition. Journal of Bacteriology 174: 6005-6010.
9. Scott JR, Churchward GG (1995) Conjugative Transposition. Annual Review of Microbiology 49: 367-397.
10. Garcia-Vallve S, Romeu A, Palau J (2000) Horizontal gene transfer in bacterial and archaeal complete genomes. Genome Research 10: 1719-1725.
11. Luo YN, Wasserfallen A (2001) Gene transfer systems and their applications in Archaea. Systematic and Applied Microbiology 24: 15-25.
12. Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, et al. (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima. Nature 399: 323-329.
13. Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405: 299-304.
14. Wackernagel W, Brigulla M (2010) Molecular aspects of gene transfer and foreign DNA acquisition in prokaryotes with regard to safety issues. Applied Microbiology and Biotechnology 86: 1027-1041.
15. Chen I, Dubnau D (2004) DNA uptake during bacterial transformation. Nature Reviews Microbiology 2: 241-249.

16. Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nature Reviews Microbiology 8: 552-563.
17. Willetts N, Skurray R (1980) The Conjugation System of F-Like Plasmids. Annual Review of Genetics 14: 41-76.
18. Frost LS, Ippenihler K, Skurray RA (1994) Analysis of the Sequence and Gene-Products of the Transfer Region of the F-Sex Factor. Microbiological Reviews 58: 162-210.
19. Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. Nature Reviews Genetics 5: 123-U116.
20. Blanchard JL, Lynch M (2000) Organellar genes - why do they end up in the nucleus? Trends in Genetics 16: 315-320.
21. Andersson JO (2005) Lateral gene transfer in eukaryotes. Cmls-Cellular and Molecular Life Sciences 62: 1182-1197.
22. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics 9: 605-618.
23. Hall C, Brachat S, Dietrich FS (2005) Contribution of horizontal gene transfer to the evolution of Saccharomyces cerevisiae. Eukaryotic Cell 4: 1102-1115.
24. Furner IJ, Huffman GA, Amasino RM, Garfinkel DJ, Gordon MP, et al. (1986) An Agrobacterium Transformation in the Evolution of the Genus Nicotiana. Nature 319: 422-427.
25. Intrieri MC, Buiatti M (2001) The horizontal transfer of Agrobacterium rhizogenes genes and the evolution of the genus Nicotiana. Molecular Phylogenetics and Evolution 20: 100-110.
26. Hirt H, Pitzschke A (2010) New insights into an old story: Agrobacterium-induced tumour formation in plants by plant transformation. Embo Journal 29: 1021-1032.
27. Kunik T, Tzfira T, Kapulnik Y, Gafni Y, Dingwall C, et al. (2001) Genetic transformation of HeLa cells by Agrobacterium. Proceedings of the National Academy of Sciences of the United States of America 98: 1871-1876.
28. Cain JR (1988) A Case of Septicemia Caused by Agrobacterium Radiobacter. Journal of Infection 16: 205-206.
29. Hotopp JCD, Clark ME, Oliveira DCSG, Foster JM, Fischer P, et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317: 1753-1756.
30. Scholl EH, Thorne JL, McCarter JP, Bird DM (2003) Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. Genome Biology 4.
31. Matthysse AG, Deschet K, Williams M, Marry M, White AR, et al. (2004) A functional cellulose synthase from ascidian epidermis. Proceedings of the National Academy of Sciences of the United States of America 101: 986-991.
32. Nakashima K, Yamada L, Satou Y, Azuma J, Satoh N (2004) The evolutionary origin of animal cellulose synthase. Development Genes and Evolution 214: 81-88.
33. Koonin EV, Makarova KS, Aravind L (2001) Horizontal gene transfer in prokaryotes: Quantification and classification. Annual Review of Microbiology 55: 709-742.
34. Wolf YI, Aravind L, Grishin NV, Koonin EV (1999) Evolution of aminoacyl-tRNA synthetases - Analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. Genome Research 9: 689-710.
35. Guljamow A, Jenke-Kodama H, Saumweber H, Ouillardet P, Frangeul L, et al. (2007) Horizontal gene transfer of two cytoskeletal elements from a eukaryote to a cyanobacterium. Current Biology 17: R757-R759.
36. Andersson JO, Sjogren AM, Davis LAM, Embley TM, Roger AJ (2003) Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Current Biology 13: 94-104.
37. Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga Bigelowiella natans. Proceedings of the National Academy of Sciences of the United States of America 100: 7678-7683.

38. Diao XM, Freeling M, Lisch D (2006) Horizontal transfer of a plant transposon. Plos Biology 4: 119-128.

39. Shirasu K, Yoshida S, Maruyama S, Nozaki H (2010) Horizontal Gene Transfer by the Parasitic Plant Striga hermonthica. Science 328: 1128-1128.

40. Steele RE, Hampson SE, Stover NA, Kibler DF, Bode HR (2004) Probable horizontal transfer of a gene between a protist and a cnidarian. Current Biology 14: R298-R299.

41. Rumpho ME, Worful JM, Lee J, Kannan K, Tyler MS, et al. (2008) Horizontal gene transfer of the algal nuclear gene psbO to the photosynthetic sea slug Elysia chlorotica. Proceedings of the National Academy of Sciences of the United States of America 105: 17867-17871.

42. Moran NA, Jarvik T (2010) Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids. Science 328: 624-627.

43. Bergman CM, Quesneville H (2007) Discovering and detecting transposable elements in genome sequences. Briefings in Bioinformatics 8: 382-392.

44. Ragan MA (2001) Detection of lateral gene transfer among microbial genomes. Current Opinion in Genetics & Development 11: 620-626.

45. Leplae R, Lima-Mendez G, Toussaint A (2010) ACLAME: A CLAssification of Mobile genetic Elements, update 2010. Nucleic Acids Research 38: D57-D61.

46. Woese CR (2002) On the evolution of cells. Proceedings of the National Academy of Sciences of the United States of America 99: 8742-8747.

47. Freeman VJ (1951) Studies on the virulence of bacteriophage-infected strains of Corynebacterium diphtheriae. Journal of Bacteriology 61: 675-688.

48. Watanabe T (1963) Infective Heredity of Multiple Drug Resistance in Bacteria. Bacteriological Reviews 27: 87-&.

49. Stobberingh EE, Deurenberg RH, Vink C, Kalenic S, Friedrich AW, et al. (2007) The molecular evolution of methicillin-resistant Staphylococcus aureus. Clinical Microbiology and Infection 13: 222-235.

50. Bloemendaal ALA, Brouwer EC, Fluit AC (2010) Methicillin Resistance Transfer from Staphylococcus epidermidis to Methicillin-Susceptible Staphylococcus aureus in a Patient during Antibiotic Therapy. PLoS One 5.

51. Weigel LM, Clewell DB, Gill SR, Clark NC, McDougal LK, et al. (2003) Genetic analysis of a high-level vancomycin-resistant isolate of Staphylococcus aureus. Science 302: 1569-1571.

52. Walsh TR, Kumarasamy KK, Toleman MA, Bagaria J, Butt F, et al. (2010) Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infectious Diseases 10: 597-602.