Horizontal Gene Transfer in *Thermus* spp.

Alba Blesa¹, Beate Averhoff² and José Berenguer¹*

¹Centro de Biología Molecular Severo Ochoa, Universidad Autónoma de Madrid- Consejo Superior de Investigaciones Científicas, Calle Nicolás Cabrera nº 1, 28049-Madrid, Spain
²Molecular Microbiology and Bioenergetics, Institute of Molecular Biosciences, Goethe University, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany

*Correspondence: jberenguer@cbm.csic.es

https://dx.doi.org/10.21775/cimb.029.023

Abstract

The small amount of genetic content in thermophiles generally limits their adaptability to environmental changes. In *Thermus* spp., very active horizontal gene transfer (HGT) mechanisms allow the rapid spread of strain-specific adaptive gene modules among the entire population. Constitutive expression of a rather particular and highly efficient DNA transport apparatus (DTA) is at the center of this HGT-mediated enhanced adaptability. The function of the DTA is dependent on the integrity and longevity of the extracellular DNA (eDNA) being transformed, which can be improved by the production of extracellular vesicles (EV) through lysis of a fraction of the population. The DTA must also contend with the recipient cell’s defensive barriers, namely restriction enzymes, a panoply of CRISPR-Cas systems, and the argonaute-like protein TtAgo, which may be bypassed by transjugation, a new class of bidirectional transformation-dependent conjugation. Efficient transjugation depends on the presence of the ICETh1, an integrative and conjugative element which promotes simultaneous, generalized DNA transfer from several points in the genome. Transjugation shows preference for genes located within a megaplasmid replicon, where the main strain-specific adaptive modules are located. Contribution of transformation, vesicle-mediated eDNAs, and transjugation to HGT in this genus is discussed.

Introduction

Horizontal gene transfer (HGT) and symbiosis constitute a basis to explain the explosive evolution and dramatic adaptability to new environments shown by both prokaryotes and eukaryotes. Among extreme thermophilic bacteria, the remnants of HGT events have been discovered through bioinformatics analysis of the large number of genome sequences currently available. The most accepted analyses support that 20-25% of the genes present in organisms belonging to ancient clades such as *Thermotoga* or *Aquifex* were acquired from *Archaea* (Aravind *et al.*, 1998; Deckert *et al.*, 1998; Nelson *et al.*, 1999), likely thermophilic ones, which may have played relevant roles in the thermal adaptation of these genera (Aravind *et al.*, 1998). This is the case for several isolates of the genus *Thermus*, whose genomes harbor several genes of archaeal origin in addition to others shared by their phylogenetic neighbors. Members of the genus *Thermus* are widespread among very different thermal environments, with optimal growth temperatures ranging from moderately high (50 °C) to more extreme (80°C) thermal environments, and in locations as diverse as thermal effluents on earth and in the sea to compost (Da Costa *et al.*, 2006).

At the time of writing (March 2018), there are 33 public genome projects involving *Thermus* spp. (https://gold.jgi-psf.org/index), 23 of which are kept as permanent drafts. The average genome size is about 2.3 Mbp, which in most strains is divided between a chromosome of around 2 Mbp (containing the core genome conserved among all strains in the species) and one or more megaplasmids ranking from 60 to 440 kbp that harbors strain-specific genes and thus, showing lower synteny among isolates (Bruggemann and Chen, 2006). Small plasmids <20 kbp are also found in different strains (Tripathi *et al.*, 2017). Most *T. thermophilus* strains contain a megaplasmid of...
around 200 kbp named pTT27 in the strains HB27 and HB8 (Ohtani et al., 2012). Although the pTT27-related megaplasmids of *T. thermophilus* and other *Thermus* spp concentrate "dispensable" genes, they also encode complete or partial housekeeping pathways, including the synthesis of relevant compounds such as carotenes, coenzyme B12, dATP, dGTP and dUTP, and siroheme (Henne et al., 2004) (Tripathi et al., 2017). Those *Thermus* isolates lacking a pTT27-like megaplasmid encode these housekeeping genes within different regions of the chromosome (Brumm et al., 2015; Fujino et al., 2017; Gounder et al., 2011). Remarkably, 4 or 5 chromosomal copies per cell have been reported, implying a polyploid nature at least for *T. thermophilus* HB8 (Ohtani et al., 2010), with copy numbers of the pTT27 megaplasmid similar to that of the chromosome.

One of the factors that likely contributes to strain diversity and differences in chromosomal and megaplasmid synteny is the presence of several insertion sequences (IS) distributed throughout the genome, concentrated to a larger extent in the megaplasmid (Henne et al., 2004). These IS have been ascribed to different families according to their differences in transposition mechanism, and though in some cases there is evidence of recent transposition activity (Gregory and Dahlberg, 2008; Swarts, 2014), in others only a pseudogene for the transposase remains (Henne et al., 2004). Regardless of activity, the presence of several copies of these elements and their conservation among different species and isolates contributes to the diversity found between otherwise close species, and likely facilitates the integration of DNA sequences flanked by conserved IS through recombination.

Among the genes present in the core genome and the pangenome of *Thermus* spp., there is a collection of homologs that are frequently found in thermophilic archaea and bacteria. Several of these genes are also located in the pTT27 megaplasmid, including defense-related adaptive CRISPR-Cas (Staals et al., 2014), Argonaute proteins (Swarts, 2014), and reverse gyrase (Brochier-Armanet and Forterre, 2007). In addition to genes of archaeal origin, genes encoding specific environmental adaptations can be found in the pTT27-like megaplasmid, including genes for the use of specific sugars or for the use of nitrogen oxides in anaerobic respiration (Alvarez et al., 2011). Some of these genes are associated with IS, and several of these strain-specific capabilities have been proven to easily transfer between strains, for example allowing a formerly aerobic strain to grow anaerobically by denitrification (Alvarez et al., 2011).

Thus, the genes mainly found in the megaplasmid of different *Thermus* spp. seem to constitute an evolutionary strategy, in which the sharing of a large pangenome allows rapid population adaptation without requiring individual maintenance of a large genome that would be risky to replicate under high temperature conditions. For this to be an effective adaptive strategy, the genus has developed very active DNA acquisition mechanisms that promote a dynamic gene exchange flux.

Regardless, such DNA acquisition mechanisms must be compatible both with the high environmental temperatures that partially melts extracellular DNA (eDNA), and with the formidable barrier posed by the cell envelope. *Thermus* spp. have a complex envelope that combines features of both Gram-negative and Gram-positive microorganisms (Cava et al., 2009), and electron microscopy studies have revealed the presence of an outer membrane (OM) surrounding the cell wall. However, the peptidoglycan layer (as thin as those found in Gram-negatives) includes dipeptide bridges between the D-Alanine at the fourth position in one muropeptide, and the L-ornithine at the third position in another chain, as in Gram-positives (Quintela et al., 1995). Moreover, the cell wall includes secondary cell wall polysaccharides (SCWP) covalently bound to the peptidoglycan, as found in some Gram-positives, and even rare modifications such as the presence of phenyl-acetate in the peptidoglycan of some isolates (Quintela et al., 1999). Actually, the OM is anchored to pyruvylated sugars of the SCWP through the N-terminal domain of a regular protein layer (SlpA) (Cava et al., 2004) that constitutes a class of membrane scaffold reminiscent of the OmpA porin in *Escherichia coli*. Thus, for eDNA to efficiently cross this multi-layer barrier, a similarly complex DNA transport apparatus (DTA) must be in place.

In reality, very efficient horizontal gene transfer (HGT) mechanisms exist in *Thermus* spp., studied extensively in the model *T. thermophilus* strain HB27, to which most of the specific aspects of HGT described in this chapter will be addressed. The main HGT mechanisms in *T. thermophilus* involve a DTA that can act directly on free eDNA (Averhoff, 2009) as described in other bacteria, or uptake eDNA from extracellular vesicles (EV) (Blesa and Berenguer, 2015), or directly transfer from donor
cells in intimate contact in a process called transjugation (Blesa et al., 2017).

**Natural competence**

Natural competence in *T. thermophilus* has been reported as one of the most efficient systems described to date in terms of rate (up to 40 kbp/s per cell) and promiscuity, as internalization of DNA from members of all three domains of life occurs with similar efficiency (Schwarzenlander and Averhoff, 2006). It is also noteworthy that the system functions during all growth phases (Hidaka et al., 1994). However, transformation frequency varies greatly from strain to strain (Koyama et al., 1986) despite the high degree of conservation of natural competence genes among strains, possibly due to barriers against HGT such as restriction-modification systems.

At least 16 genes have been implicated in the DTA of *T. thermophilus* HB27 through a combination of bioinformatics and mutational analysis (Averhoff, 2009), many of which play a dual role in natural transformation and biogenesis of type IV pili (T4P) (Averhoff, 2009; Friedrich et al., 2001; Friedrich et al., 2002; Friedrich et al., 2003), and they have been assigned to three different groups.

The first group includes the highly conserved competence proteins ComEA, ComEC and DprR proteins. ComEA was found to bind double-stranded DNA (dsDNA), thereby contributing to the transport of eDNA through the OM (Salzer et al., 2016b). ComEA is exclusively present in the inner membrane (IM). This finding was quite surprising, since all ComEA orthologues in Gram-negative bacteria are soluble. ComEC is a polytopic IM protein whose orthologues are also widespread in many bacteria. The IM localization together with the finding that comEC deletion mutants still take DNA into a DNase-resistant state suggests that ComEC mediates DNA transport from the periplasm to the cytoplasm (Schwarzenlander et al., 2009). ComEC also modulates the transcriptional regulation of DNA translocator and T4P components, thereby mediating a response to extracellular stimuli (Salzer et al., 2014c). The function of DprA has not been addressed in detail, but the finding that DNA binding and uptake in dprA knockout mutants are unaffected, together with functions conveying incoming single-stranded DNA (ssDNA) to RecA in orthologues (Mortier-Barriere et al., 2007; Yadav et al., 2014) is in line with the idea that DprA stabilizes incoming DNA and is important for strand exchange during recombination.

The second group of competence proteins consists of proteins which are very similar to components of the T4P biogenesis system, including four pilin-like proteins (PilA1, PilA2, PilA3 and PilA4), a leader peptidase (PilD), an AAA-ATPase (PilF), an IM protein (PilC), a PilM-homologue and a secretin-like protein (PilQ). Individual deletion of PilA1, PilA2 and PilA3 abrogated competence without affecting piliation, suggesting distinct roles for the individual pilins in DNA uptake but not in pilus assembly. In contrast, PilA4 plays an essential role in DNA uptake and is also the major subunit of T4P (Friedrich et al., 2003; Schwarzenlander et al., 2009). PilF is a unique zinc-binding AAA-ATPase essential for the transport of DNA through the OM and polymerization of T4P (Rose et al., 2011; Salzer et al., 2013; Salzer et al., 2014b; Schwarzenlander et al., 2009). It has a tripartite structure: a unique N-terminus containing three general secretory domains; a central region; and a C-terminal ATPase and zinc-binding domain. PilF assembles into hexameric complexes that have a disk and ring-like structure separated by a short stem-like structure (Rose et al., 2011, Collins et al., 2013). The C-terminal region encodes a tetracysteine motif which mediates zinc binding (Rose et al., 2011; Salzer et al., 2014a). Individual cysteine residues are also important for complex stability, probably beneficial for the functionality of PilF under high temperature environmental conditions (Salzer et al., 2014a). Analogous to PilA4 and PilF, the competence proteins PilC, PilM, PilN and PilQ were also found to be absolutely essential for T4P assembly (Friedrich et al., 2002). Structural analyses suggest that the IM protein PilC forms dimers which might interact with cytoplasmic and periplasmic proteins (Karuppiah et al., 2010). However, further work is needed to identify interaction partners and elucidate the function of PilC. PilM interacts with the IM protein PilN, forming dimeric PilMN complexes that initiate interaction with two PI0 monomers, generating a transmembrane platform for the assembly of pilins (Karuppiah et al., 2013; Karuppiah and Derrick, 2011). The secretin PilQ has a unique modular organization and forms homopolymeric complexes, which are essential for the passage of DNA through the OM (Burkhardt et al., 2011; Schwarzenlander et al., 2009). The PilQ complex has a unique structure consisting of a "cone", a "cup" and six rings with a large central channel. The non-conserved N-terminus of PilQ exhibits a modular architecture comprised of domains with alternating α-helices and β-sheets that assemble into the six ring system underneath the cup-like structure and are important
for pilus extrusion and function but are dispensable for natural transformation (Burkhardt et al., 2012; Salzer et al., 2016a). The structural features together with the unprecedented length (34 nm) of the secretin complex suggests that the PilQ complex spans the entire cell periphery, thereby mediating DNA transport across the OM and periplasmic space in a single step, which might be beneficial for DNA uptake in high temperature environments.

The similarities of competence proteins to components of T4P biogenesis led to the hypothesis that T4P plays a role in natural transformation. However, our findings that non-piliated Thermus mutants still take up free DNA, some with even higher frequencies than wild-type cells, led to the conclusion that piliation and natural competence, despite sharing some common genes, are separate systems (Burkhardt et al., 2012; Salzer et al., 2014b).

In addition to these conserved competence proteins, DTA depends on several unique proteins, such as ComZ, PilN, PilO and PilW (Friedrich et al., 2002; Friedrich et al., 2003). The role of ComZ remains to be determined, though it seems to be implicated in DNA uptake through the OM (Schwarzenlander et al., 2009). PilN and PilO are both part of an IM platform for initiation of the biogenesis of the DTA and T4P (Karuppiah et al., 2013; Rumszauer et al., 2006). The non-conserved competence protein PilW is also implicated in both DTA and T4P, and is essential for assembly and/or stability of the OM-spanning PilQ complexes (Rumszauer et al., 2006). Based on the numerous structural and functional data, a model of the DTA in T. thermophilus is presented in Figure 1.

---

**Figure 1. Model of the DNA translocator.** The secretin forms a channel spanning the outer membrane (OM) and the periplasm. This channel mediates extrusion of a pilus-like structure comprised of PilA1-4. PilW is important for the assembly of the secretin complex, and the AAA-ATPase (PilF) powers the polymerization of the pilus-like structure. The dimeric PilC complex links the energy released from PilF-mediated ATP hydrolysis to the assembly of a pilus-like structure. PilM, PilN and PilO form the inner membrane (IM) assembly platform, while ComEA binds incoming DNA and delivers it to the IM channel formed by ComEC.
The next question is: What is the fate of the incoming DNA? So far it is not clear whether one or both strands of the eDNA are taken up by *T. thermophilus*. In most natural competence systems, only one strand of the eDNA enters the cell, thus this pattern has also been proposed for *Thermus* spp. In any case, eDNA is incorporated into the genome by homologous recombination, with >90% similarity and sizes above 500 nt required for successful integration.

Recently, electron cryo-tomography and reconstruction has provided the first three-dimensional *in situ* structure of the secretin (Gold et al., 2015). The secretin channel was found to be highly dynamic, switching between closed and open states. Two gates safely close the channel in the absence of pili but open when pili are extruded. Moreover, the opening of the channel to make way for pilus extrusion is accompanied by major conformational changes of the N-terminal domains of the central secretin. Interestingly, only ~20% of the secretin channels extruded pili, and it is tempting to speculate that the idle complexes may be active in DNA uptake. Taken together, these *in situ* studies open avenues for further protein structure-function analyses, shedding more light onto the mechanisms of DNA translocation.

**Vesicle-protected HGT**

Efficient HGT through natural competence is limited by the size of the eDNA, as effective integration into the chromosome requires a rather long DNA fragment. This is especially relevant in high-temperature environments, where the stability of the DNA is further compromised (Soler et al., 2008). One of the mechanisms that helps alleviate problems with eDNA stability in high temperatures is protection by attachment to or enclosure within extracellular vesicles (EVs) or phage-like capsids called gene transfer agents (Lang et al., 2012). Vesicle-protected eDNA can withstand time and nuclease activity, allowing displacement of eDNA over long distances. EVs stored for more than a year at 4 °C maintained their ability to be taken up by *T. thermophilus*.

EVs are membrane-enclosed structures that are produced by members all three domains of life under different names (membrane vesicles, exosomes, shedding microvesicles, etc.) which are frequently generated by bacteria during the growth phase or when integrated within a biofilm (Turnbull

---

**Figure 2. Scheme of HGT processes detected in *T. thermophilus*.** *T. thermophilus* cells in close contact are represented, with the chromosome and pTT27 megaplasmid shown in black and blue lines, respectively. Production of polar type IV pili (broken black lines) involved in twitching motility is also shown. 1) Both cells can take in unprotected DNA (red) from the environment through the natural competence apparatus, likely via the PilQ protein (white cylinder) localized at the cell poles and shared by the type IV pili. Internalized DNA is shown as single-stranded, but this has not yet been demonstrated. 2) A transjugation model is depicted in which the ICEth1 (green), represented both as an inserted and circular element, generates the enzymes involved in the recognition of origins of transfer in the genome at different frequencies, in a distributive seemingly random manner. The fragments generated from the chromosome (black), the megaplasmid (blue), and the ICETh1 (green) are transported through the cell envelope (“push” step) in a TdtA-dependent manner (gray cylinders), and the secreted DNA is captured by the competence apparatus and internalized (“pull” step), in a DNase-protected process. 3) Both cells can capture DNAse-protected DNA from membrane vesicles generated from lysis of distant donor cells. The natural competence apparatus is required for such internalization, and no clear-cut conclusions can be made regarding the requirement of vesicle fusion to liberate trapped DNA. Mechanisms 1 and 3 are subject to the TiAgO surveillance defense system based on DNA-DNA interference, whereas mechanism 2 is not.
et al., 2016). In Gram-negative bacteria, the production of OM vesicles (OMVs) is quite common, driven by different mechanisms in response to diverse trigger signals. In addition to the expected OM and periplasmic components, these OMVs frequently contain cytoplasmic enzymes and/or DNA, playing a great diversity of roles, such as in cell communication, immune system evasion, and stress response, among others (Schwechheimer and Kuehn, 2015). Indeed, it is evident that OMV biogenesis diverges among bacteria and is influenced by the environmental context (Deatherage and Cookson, 2012; Orench-Rivera and Kuehn, 2016).

In contrast to the apparently sophisticated mechanisms involved in OMV generation in some bacteria, the generation of EVs in *Thermus* spp. seems to be a consequence of cell lysis, with fortuitous capture of DNA fragments from the genome within membrane fragments (Blesa and Berenguer, 2015). This lytic origin is also seen in other bacteria where it is linked to stressful conditions, suggesting it is either a consequence of a specific program to induce lysis or the result of irreversible cell damage (Orench-Rivera and Kuehn, 2016).

Very little is known about how EV-associated eDNA can enter the recipient cell, though a fusion mechanism has been proposed for OMVs that would allow eDNA access to the periplasm, with other internalization mechanisms required for entry to the cytoplasm. Regarding the EVs of lytic origin, access is likely related to the presence of a functional natural competence apparatus in the recipient (Fulsundar et al., 2014).

Different *Thermus* isolates produce DNAse-protected eDNA associated with EVs in a growth medium and phase-dependent manner, in such a way that fast-growing cells generate larger amounts of eDNA than cells in stationary phase (Blesa and Berenguer, 2015). This eDNA was acquired by the recipient cells at similar frequencies independently of its integration site in the chromosome or in the megaplasmid. Actually, the analysis of the eDNA fraction associated with EVs and protected from the DNAse revealed a size of around 20 kbp, similar to that obtained for genomic preparations of the strain following conventional methods, and further analysis showed that the whole genome was randomly represented in the EV-protected eDNA fraction (Blesa and Berenguer, 2015).

Proteomic assays of the EV fraction also revealed the presence of proteins from all the cell compartments, with the majority of proteins from the IM (20%), OM (12%), and periplasmic (27%) fractions, but with a significant presence of cytoplasmic proteins (14%). The presence of random genomic fragments and proteins clearly supports a lytic origin for *Thermus* EVs, similar to other bacteria.

Interestingly, the production of EVs was highest during the adaptive (latency) and exponential phases of *T. thermophilus* cultures, lower when the cells entered stationary phase, and almost undetectable when the cells were grown in minimal media, where the growth rate was much lower than in rich media. These data suggest that the production of EVs in laboratory conditions is a consequence of unbalanced growth in a non-natural medium with nutrient excess, but also reveals that cell lysis is a normal event, and that cell populations in natural environments may generate EVs as vehicles for HGT. In this sense, the possibility of a programmed lytic process in parts of the population, such as those described in *Pseudomonas aeruginosa* biofilms (Turnbull et al., 2016), cannot be excluded.

The mechanism underlying the entry of EV-associated eDNA into *T. thermophilus* is not clear, but it depends on the natural competence of the recipient cell, as mutants lacking different competence genes are unable to acquire such eDNA. Therefore, recipient cells are somehow capable of directly contacting the EV-protected eDNA: either the EV fuses to the OM, or the eDNA is actually adsorbed to surface of the EV and is thus accessible to the competence machinery, but hidden enough to avoid degradation by DNAse. Keeping in mind the great complexity of the multilayered *T. thermophilus* cell envelope and its intermediate structure between that of Gram-negative and Gram-positive bacteria (Cava et al., 2009), external display of eDNA on the EV surface seems to be a more likely option.

Whatever the mechanism of eDNA acquisition is, it is relevant to place it within the general context of DNAse-resistant HGT in *Thermus* spp., where much higher transfer frequencies are detected for other mechanisms such as transjugation, as described below.
Transjugation

Genes encoding proteins homologous to components of the classical conjugation machinery (including type 4 secretion systems, T4SS, relaxases and relaxosome elements involved in oriT recognition and transfer) have been identified in plasmids from *Thermus thermophilus* JL-18 (accession CP003254), *T. thermophilus* SGO, 5JP17-16 (accession CP002778) and *T. aquaticus* Y51MC23 (Brumm *et al*., 2015), though a conjugation phenotype has not yet been proven in these strains.

However, a conjugation-like cell-to-cell transfer process has been described in *T. thermophilus* strains that lack obvious homologs to these conjugation genes. For example, conjugation between a nitrate respiring strain of *T. thermophilus* (NAR1) and the aerobic strain HB27 was described that allowed for isolation of a nitrate-respiring HB27 derivative (Ramirez-Arcos *et al*., 1998). Sequencing of the transferred DNA identified an apparently mobile element encoding the nitrate respiration capability (NCE) (Cava *et al*., 2009). Additional conjugation experiments with other denitrifying strains of *T. thermophilus* that had no apparent standard conjugative apparatus (based on genome sequences with 98% coverage, laboratory results) also led to the isolation of denitrifying derivatives of HB27 (Alvarez *et al*., 2011). All these data supported the existence of a new mechanism for conjugation independent of T4SS in these *T. thermophilus* strains.

This unconventional conjugation-like mechanism was subsequently identified in the aerobic *T. thermophilus* HB27 (César *et al*., 2012), and a deeper analysis revealed transfer efficiencies similar to transformation, although the mechanism differed significantly from classical conjugation (Blesa *et al*., 2014). Firstly, it didn’t involve inhibition or incompatibility barriers for DNA donation, in such a way that the same cell could function both as donor and recipient, making parenthood analysis of the progeny a difficult task. No homologues to T4SS components were identified in the genome, and genes seemed to transfer without maintenance of their order in the chromosome, in a process more akin to generalized transduction than to conjugation. Finally, it was also noted that the genes encoded by the pTT27 megaplasmid were transferred at a rate one order of magnitude more efficiently than chromosomal genes, despite similar copy numbers per cell.

Still more surprising was the requirement of a fully active natural competence system in the recipient cells, in such a way that mutants deficient in virtually any described component of the DTA could serve only as DNA donor, a property that led to proposal of the term transjugation (transformation-dependent conjugation) (Blesa *et al*., 2017). Further parenthood analysis experiments, in which combinations of non-competent mutants were used, confirmed that the DNA-donation machinery was independent of the competence apparatus, both present in the same bacteria, implying a two-step (“push and pull”) model for transjugation.

The search for putative components of the DNA-pushing system led to the identification of transjugation donor translocase A (TdtA), a protein belonging to the FtsK-HerA family of DNA translocase-helicases. Its mutation impaired the ability of the cell to act as a donor in transjugation, without significant effect on natural competence or any other phenotypic trait (Blesa *et al*., 2017). Purified recombinant TdtA was subsequently shown to form hexameric rings in the presence of ATP, with a central pore wide enough to accommodate dsDNA, resembling the TraB protein encoded by conjugative plasmids of *Streptomyces* spp. In fact, TraB seems to be the only protein required for conjugative transfer of plasmids between *Streptomyces* spp., with additional proteins encoded by conjugative plasmids acting at the level of intramycelium spreading (Thoma and Muth, 2012) (Thoma and Muth, 2015).

Interestingly, the TdtA protein is only present in a few of the sequenced strains of *T. thermophilus*, encoded by a mobile genetic element (ICETh1) reminiscent of the integrative and conjugative elements (ICE) found in many conjugative bacteria (Blesa *et al*., 2017). This 14 kb element encodes a typical phage-like recombinase of the XerC family and a four-gene operon encoding: a type IIG restriction endonuclease identical in sequence to Th111I1 (Zhu *et al*., 2014); a putative nuclease of the NurA family; TdtA; and a putative DNA methylase. Additional genes in the ICETh1 include a transposase for which two more copies exist in the genome, and a putative hydrolase of unknown function. The whole ICETh1 is integrated within the chromosome flanked by 46-bp direct repeats belonging to the 3’-end of a tRNA gene, and its exogenous origin is supported by a much lower G +C content (58%) than that of the chromosome (68%).
The ICEth1 can excise from the chromosome but apparently does not replicate, as the circular form of the element is much more difficult to detect than the corresponding scar left in the chromosome after its excision (Blesa et al., 2017). Actually, the ICEth1 is apparently lost at high frequencies (up to 8% of cells) during exponential growth, supporting its non-replicative nature, and raising questions regarding its maintenance within the population.

In contrast to classical ICEs that have evolved to promote their self-transfer, ICETH1 stimulates the generalized and parallel transfer of all the genes in the genome, especially of those genes localized in the megaplasmid, at similar or higher frequencies than the ICETH1 itself. This apparently altruistic nature of the ICETH1 might constitute a sophisticated strategy for spreading, using traits encoded by the genome for co-selection, though a putative advantage of those cells harboring the ICETH1 cannot be discarded at this point.

One of the less understood aspects of transjugation is the way in which the origins of transfer (orIT) are selected. As the TdT protein itself does not contain DNA-binding motifs similar to those found in the TraB protein (Thoma and Muth, 2012), a very attractive hypothesis involves action of the restriction endonuclease TTh11IIi, which is itself encoded by the ICETH1, at multiple targets in the genome. Genes located near regions with a high density of TTh11IIi recognition sequences seem to be transferred at higher frequencies than genes located in regions with fewer recognition sites, and mutants defective in this restriction enzyme have transfer frequencies three orders of magnitude lower than the wild-type. Further efforts to analyze the frequency of transfer of plasmids with or without restriction sites should shed more light on the process.

The role of the NurA nuclease also remains to be elucidated, though its relevance seems clear as nurA null mutations abolish transjugation (Blesa et al., 2017). Archaeal homologs of NurA have been shown to act in coordination with the HerA helicase in repair of dsDNA-breaks involving homologous recombination. NurA-HerA complexes have been proposed in which the helicase pushes DNA through a NurA dimer that digests the DNA in a trimming-like process (Rzechorzek et al., 2014). In Thermus, the formation of a putative NurA-TdT complex would likely be involved in DNA donation, although its exact role must be studied further.

The role of innate and adaptive immunity as barriers against HGT
Several classes of barriers have been described that protect cells against infection and replication of viruses (Seed, 2015). In addition to the DNA restriction barriers present in most bacteria, and likely as a consequence of the great promiscuity and constitutive nature of its natural competence system, T. thermophilus wields an arsenal of adaptive and innate defensive systems. Adaptive defenses rely on a collection of 10-12 CRISPR sequences, most of them located in the megaplasmid, and the corresponding Cas protein systems belonging to three different families (I, IIa, IIb) that function as crRNA-DNA and crRNA-RNA interference systems against invading DNA (Staals et al., 2014). The role of these systems against intragenus or intraspecies HGT is not likely to be significant as these genes cannot be recognized by the CRISPR spacers of the recipient strains, for the same reasons that the donor strain does not target its own genome.

In addition to this adaptive system, Thermus spp. contain a homolog to the eukaryotic argonaute protein (TiAg0), whose role in defense against both plasmidic (Swarts, 2014) and genomic DNA (Blesa et al., 2014) acquired by natural competence has been well established. The protein somehow recognizes the entering eDNA and uses an ssDNA guide to screen for complementarity and degradation after DNA-DNA interference, in what seems to be an additional innate defense mechanism. In contrast to its activity against eDNA acquired by natural competence, TiAg0 does not limit access of DNA acquired by transjugation, despite the involvement of natural competence in both processes (Blesa et al., 2014). The possibility that its discrimination is based on a lower G+C content of the target DNA, as proposed after in vitro assays (Swarts et al., 2017; Swarts, 2014), seems not to be the case in vivo, as interference was also detected when isogenic DNA was transferred by transformation, but not by transjugation (Blesa et al., 2014). Therefore, the possibility exists for an activation mechanism of TiAg0 that is somehow triggered by eDNA.

Transduction
Most Thermus genomes sequenced to date do not harbor integrated prophages, supporting that temperate phages are not common in the biology of Thermus spp. Nevertheless, there are some exceptions to this general observation. For example, the genome of T. aquaticus Y51MC23 contains two
complete and two residual prophages (Brumm et al., 2015), whereas Thermus strains RL, 2.9, and CCB_US3_UF1 each contain a prophage (Dwivedi et al., 2012; Navas et al., 2015; Teh et al., 2012). However, the ability of these prophages to excise and replicate has not been studied, and their involvement in specialized transduction is yet to be assayed.

In contrast, a great variety and number of lytic phages have been described infecting different Thermus spp. (Yu et al., 2006), although sequences are only available for a few of them (phiYS40, TMA, phiOH2, P23-77, P23-45, p74-26, In93). This fact, together with the abundance of CRISPR sequences and Cas systems found in the genomes of most Thermus isolates (Staals et al., 2014), strongly supports frequent infection by environmental viruses, suggesting that generalized transduction may be a contributing factor to HGT gene flow in these bacteria.

**Concluding remarks**

The genus Thermus is ubiquitous and widespread in diverse thermal environments revealing an apparent metabolic flexibility which contrasts its small genome size. In what seems to be a strategy to cope with limitations on genetic content, most strains contain a genetically plastic megaplasmid that allocates many specific adaptation modules and very efficient HGT mechanisms that allow the rapid acquisition and spread of such traits among the population in response to a sudden change in the environment.

The natural competence machinery is central to all of the transfer mechanisms demonstrated so far: transformation, EVs and transjugation. Putative involvement of viral transduction or classical donor-recipient conjugation systems has yet to be proven, but the abundance of Thermus-specific phages and the existence of gene clusters reminiscent of the T4SS and relaxases in a few of the sequenced strains indicate that they also may contribute to HGT in the species.

Natural competence itself is demonstrated to be a very efficient method of HGT for some Thermus strains, and it has been proposed as the major mechanism involved in strain adaptation. Acquisition of traits such as new S-layers and the capability to grow anaerobically with nitrate are laboratory examples of how natural competence affords such adaptability. The DTA is a highly complex and very dynamic system composed of several unique proteins, highlighting the features which might have been triggered by the extreme environment and/or the complex cell envelope. It exhibits extraordinarily broad substrate specificity and a very high efficiency, which may be of major importance for thermoadaptation of Thermus and interdomain DNA transfer in hot environments. However, in its natural environment, the stability of naked eDNA is always compromised by the presence of DNAses and high temperatures, suggesting that eDNA transport in a protected manner is more likely. In this sense, EVs produced in laboratory conditions during exponential phase may mimic the natural process in which fragmented cell envelopes generate a protective coat that adsorbs and engulfs large genome fragments, allowing for their environmental persistence until capture by a recipient cell's natural competence apparatus.

Transjugation is a newly identified mechanism that promotes HGT between isogenic and non-isogenic strains of Thermus thermophilus bearing the ICETh1. The presence of this element results in the contact-mediated transfer of any marker in the genome from donor to recipient, and especially favors those genes located within the megaplasmid, which may be more useful for environmental adaptation. The mechanism is similar to the distributive transfer found in Mycobacterium, as it involves parallel transfer of genes from multiple sites in the genome in an apparently simultaneous manner, though it is still unknown if transjugation produces mosaic progenies, as in mycobacteria. The mechanism depends on the activity of a DNA translocase protein that can accommodate dsDNA, but also a nuclease and a restriction enzyme that are likely involved in selection of the origins of transfer. The most exciting property of transjugation, however, is that although it depends on the natural competence apparatus in the recipient cell, the transferred DNA avoids the interference mechanisms mediated by TtAgo, which degrades at least 90% of eDNA progenies, as in mycobacteria. The mechanism depends on the activity of a DNA translocase protein that can accommodate dsDNA, but also a nuclease and a restriction enzyme that are likely involved in selection of the origins of transfer. The most exciting property of transjugation, however, is that although it depends on the natural competence apparatus in the recipient cell, the transferred DNA avoids the interference mechanisms mediated by TtAgo, which degrades at least 90% of eDNA molecules that enter the cell by transformation. Considering the 10-fold preference for the transfer of megaplasmid-associated genes over chromosomal ones, all the evidence clearly points to this mechanism as a major tool for environmental adaptation in Thermus spp.

**Open questions**

Despite being the best known model among extreme thermophiles, several questions remain to be answered regarding the HGT mechanisms reviewed here, such as: How is the simultaneous
expression of distantly located competence genes regulated? How is the signal transduced from outside to inside? Which form of energy powers the uptake of DNA? How does the competence apparatus recognize and transport eDNA? Where is the apparatus located within the cell? What is the link between the competence apparatus and pili? How can the bacteria distinguish between eDNA acquired by transformation and DNA acquired by transjugation while using the same competence proteins?

The competence-dependence of the transjugation process itself continues to be a puzzle itself, with such fundamental questions as: How does the system recognize regions in the chromosome or megaplasmid as the origin of transfer? Which proteins in addition to those encoded by the ICEth1 are necessary for donor cells to secrete DNA through the complex cell wall? How is such DNA protected from DNases before the competence proteins of the recipient cells are able to hide it within the periplasm? Additional questions regarding the relevance of eDNA transfer through vesicles or the putative relevance of viruses as vehicles of HGT also await future research.

Acknowledgements

This work was supported by grants BIO2016-77031-R from the Spanish Ministry of Economy and Competitiveness and grants nº 324439 and nº 685474 from the FP7-PEOPLE-2012-IAPP and the 2020 research and innovation program of the European Union, respectively, to JB., and by a grant from the German Research Foundation (Av9/6-2) to BA. An institutional grant from Fundación Ramón Areces to the CBMSO is also acknowledged.

References


incoming ssDNA to RecA. Cell 130, 824-836 DOI: 10.1016/j.cell.2007.07.038.


