Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Towards a taxonomy of conjugative plasmids

Raul Fernandez-Lopez, Santiago Redondo, M Pilar Garcillan-Barcia and Fernando de la Cruz

Conjugative plasmids are the keystone of horizontal gene transfer. Metagenomic research and clinical understanding of plasmid transmission beg for a taxonomical approach to conjugative plasmid classification. Up to now, a meaningful classification was difficult to achieve for lack of appropriate analytical tools. The advent of the genomic era revolutionized the landscape, offering a plethora of plasmid sequences as well as bioinformatic analytical tools. Given the need and the opportunity, in view of the available evidence, a taxonomy of conjugative plasmids is proposed in the hope that it will leverage plasmid studies.

Address
Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria, Santander, Spain

Corresponding author: de la Cruz, Fernando (delacruz@unican.es)

Current Opinion in Microbiology 2017, **38**:106–113

This review comes from a themed issue on **MGE-HGT in prokaryotes**

Edited by **Andrew Lang**, **J Thomas Beatty** and **Phoebe Rice**

http://dx.doi.org/10.1016/j.mib.2017.05.005

Horizontal gene transfer and the origin of bacterial species

While central to evolutionary thinking, the concept of species has been problematic, especially in the prokaryotic domain [1–3]. Doolittle and Zhaxybayeva identify three main questions regarding the origin of bacterial species: whether bacterial species exist, whether a unitary definition is possible, and what are the evolutionary forces behind bacterial speciation [2]. Horizontal gene transfer (HGT) stands out because it conditions the answers to these three questions [2,4]. If rampant, HGT turns phylogenetic trees into reticulate networks, that is, different regions of the genome would show unrelated evolutionary lineages, and the species concept would become irrelevant [2,5]. If intense, but restricted to a fraction of the genome, HGT would resemble sexual reproduction, enabling the maintenance of genomic diversity within species [6,7••,8,9]. If HGT is highly variable among species, then genomic diversity will also exhibit high inter-species variability, thus complicating the task of achieving a common operational definition. Finally, HGT is intrinsically linked to recombination. When operating within the boundaries of a given species, HGT promotes homologous recombination, thus acting as a cohesive force that maintains intra-species allelic diversity [2,3,6]. When HGT events cross the species barrier, it becomes a major source of evolutionary innovation, fostering new ecological adaptations and speciation [9,10•,11,12••,13].

The role of mobile genetic elements

Natural transformation, phage transduction, and bacterial conjugation are three main routes of HGT in bacteria. Phage transduction and bacterial conjugation are encoded by mobile genetic elements (MGEs) [14]. Conjugative plasmids, integrative conjugative elements (ICEs) and bacteriophages propagate both by vertical expansion (piggybacking their host reproduction), and by infectious transfer (horizontally invading new hosts). This latter ability makes MGEs essential vehicles for bacterial recombination, since they often incorporate in their genomes genes and operons from past hosts [15]. HGT pathways are not mutually exclusive, as phages can induce cell lysis and liberate to the environment intact plasmid molecules that can be taken up by naturally competent species [16•].

The impact of MGEs on the speciation dynamics is twofold. On the one hand, the host range of MGEs infecting a species becomes the critical parameter determining the propensity of that species to recombine outside its own gene pool [17,18]. Broad host range MGEs thus favor genetic exchanges among distinct species, fostering evolutionary innovation [10•,17,19]. On the other hand, the rate of HGT within species (likely to be higher than transfer between species), imposes a higher limit on the rate of homologous recombination [6,7••]. A given species might exhibit highly efficient biochemical mechanisms for homologous recombination by RecA or similar mechanisms, yet if there were no MGEs to serve as bridges between different genotypes, evolution would proceed in a clonal way.

Understanding the biology of MGEs thus becomes an essential endeavor to unravel the role of HGT in the origin of bacterial species. Although the relative importance of different MGEs is likely to be variable, depending on the bacterial species considered, conjugative plasmids are generally recognized as fundamental vehicles for

HGT in many clinically and environmentally relevant species [14,15,17]. Conjugative plasmids have been instrumental in the epidemic propagation of antibiotic resistances [20], virulence determinants [19,21], resistance to xenobiotics [22] and evolutionary innovations such as biodegradative pathways [23], or nitrogen fixation genes [24]. Yet, despite their obvious relevance in the evolution of bacterial species, we know surprisingly little about their population genetics, host range, and transfer rates in the wild. Moreover, it is not a given that these parameters can be assigned to certain groups of bacterial plasmids as a whole. Are there plasmid clusters to which we can assign certain genotypic and phenotypic characteristics? Or are these characteristics so idiosyncratic that every plasmid molecule is fundamentally different from all others? In other words, does anything similar to a 'plasmid species' exist? While the existence and nature of viral species is generally accepted [25], and it has been even possible to observe speciation in real time [26••], no such consensus for conjugative plasmids exists yet.

Classifying plasmids

Taxonomy involves the organization of a significant classification, but not all classifications are taxonomical. Because plasmids are key in the spread of antibiotic resistances, several classification schemes have been developed for epidemiological tracking [27–29]. While both taxonomical and epidemiological classification might overlap to some extent (*e.g.*, *E. coli* phylogroups [30,31]), both endeavors have fundamentally different goals. Epidemiological classification (even when it involves some sort of phylogenetic analysis such as MLST) is oriented to track outbreaks, detect clonal expansions, and distinguish between different strains. On the other hand, systematic classification involves phylogenetic and functional information. Species must reveal certain features about the organization and biology of their members [2], even if the amount of information provided varies depending on the genetic/phenetic homogeneity of the species in question [1]. Moreover, species must be engraved in a hierarchy of clades, informative of the phylogenetic relationships between them. That is, we should be able to group species into higher taxonomic units.

That plasmids can be grouped into taxonomic units, is a realization that predates the genomic era [32,33]. Early studies recognized that plasmids could be organized into different phenotypic groups [34], which was interpreted as sign of a phylogenetic relationship between them [32,33]. Members of each group were unable to co-reside within the same bacterial host, thus these clusters were denominated incompatibility (Inc) groups [35]. DNA sequencing revealed that members of a given Inc group exhibit certain sequence identity among them. However, the extent of sequence identity among members varies between different Inc groups. Within certain groups, the similarity extends to most of the plasmid backbone. This is the case for plasmids belonging to IncP, IncN or IncW groups [36–38]. In other groups, like IncF, similarity among members is limited to the conjugation region [20,34,39]. Thus, although members of the same Inc group exhibit a certain phylogenetic relationship, this relationship is highly variable between different Inc groups.

Defining a plasmid taxonomic hierarchy (classes, families and species)

To set up an operational taxonomy, we need to construct phylogenies using some conserved genetic marker. Bacterial taxonomy employs 16S RNA and concatenated sets of conserved proteins, but no universally conserved gene exists in plasmids. The closest to a universally conserved plasmid gene would be the replication initiation protein (RIP). However, RIP-based phylogenies are difficult to make and interpret because (i) plasmids quite often contain more than one RIP, and (ii) there are many alternative mechanisms to initiate replication, so plasmid RIP genes, when identifiable, belong to a large number of protein families [40,41]. A feasible alternative would be to use the conjugative relaxase, the protein required to initiate plasmid mobilization through conjugation. It is known that all plasmids transmissible by conjugation contain a known relaxase [42]. This approach has the drawback of leaving out non-mobilizable plasmids, but provides a robust, universal evolutionary marker for mobilizable and conjugative plasmids [43]. According to their relaxases, plasmids can be divided in eight MOB (for mobility) classes [42,44]. Phylogenetic analysis of plasmids within these classes resolves well-supported clades (Figure 1). Evolutionary ancient clades, which will be called families, display an old association between a given mobilization system and a mating pair formation system (MPF) [42,44,45]. For example, MOB_F plasmids can be subdivided into several families according to their MPF system (Figure 1). MOB_{F12} plasmids always contain a MOB_{F12} type relaxase, plus an MPF_F; MOB_{F11} plasmids contain a MOB_{F11} relaxase and a MPF_T; and MOB_{F13} plasmids encode a MOB_{F13} relaxase associated to a MPF_C [39,45]. Thus, at the taxonomic family level, plasmids exhibit conservation of a 20–30 kb DNA segment containing a defined TRA system (MOB + MPF) with substantial DNA homology.

Phylogenetic trees made with components of the TRA system generally exhibit congruent topologies from this taxon level downwards [38,39,45,46]. This allows the generation of unique phylogenies, in which the overall DNA homology of the entire plasmid increases as we advance to more recent clades (Figure 1). For example, while the MOB_{F11} family includes plasmids that contain a MOB_{F11} relaxase and a MPF_T, its subclades encompass groups of plasmids that share the same relaxase, conjugative pilus, replication, and stability regions [38].