Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Flexible genomic islands as drivers of genome evolution

Francisco Rodriguez-Valera, Ana-Belen Martin-Cuadrado and Mario López-Pérez

Natural prokaryotic populations are composed of multiple clonal lineages that are different in their core genomes in a range that varies typically between 95 and 100% nucleotide identity. Each clonal lineage also carries a complement of not shared flexible genes that can be very large. The compounded flexible genome provides polyclonal populations with enormous gene diversity that can be used to efficiently exploit resources. This has fundamental repercussions for interpreting individual bacterial genomes. They are better understood as parts rather than the whole. Multiple genomes are required to understand how the population interacts with its biotic and abiotic environment.

Address
Evolutionary Genomics Group, Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, San Juan de Alicante, 03550 Alicante, Spain

Corresponding author: Rodriguez-Valera, Francisco (frvalera@umh.es)

Current Opinion in Microbiology 2016, **31**:154–160

This review comes from a themed issue on **Environmental microbiology**

Edited by **Monica Vasquez** and **Steven Hallam**

http://dx.doi.org/10.1016/j.mib.2016.03.014

1369-5274/© 2016 Elsevier Ltd. All rights reserved.

Introduction

In the wild, bacterial and archaeal populations reproduce clonally by simple fragmentation mechanisms (typically bipartition). However, non-reproductive sex is widespread and less confined than in eukaryotes [1]. Not limited by similarity requirements like meiosis prokaryotic cells can incorporate exotic DNA from distant microbes (including eukaryotes). The importance of prokaryotic sex, also known as horizontal gene transfer (HGT), in prokaryotic evolution has been proven so extensively [2], and in such a broad range of microbes, that the possibility of representing their evolution in a tree like form possess serious challenges [3]. However, seen from the other end of the time range, at short evolutionary time spans, the mechanisms and the dynamics of gene acquisition and loss appear more amenable to systematic description. Specifically, genes located in clusters or genomic islands can be more easily tracked as they vary among closely related genomes. Such genomic islands are essential in providing diversity at the population level but also in maintaining this diversity. In this review we will analyze recent evidence about the stability and dynamics of prokaryotic genomes in the short to medium term. Comparative genomics, metagenomics and single cell genomics have been instrumental in generating this evolving new perspective.

Prokaryotic species definition, a view from metagenomics

The definition of species, the basic units of classification, is a major conundrum of Biology at large. In animal and plants sexual reproduction is a great help since it provides populations with a gene flow that is responsible of coherence within the species and discrimination from other neighboring taxa. This is the biological species concept in which species are considered as evolutionary units [4]. This concept is already problematic for metazoan or plants that reproduce clonally or interbreed with variable frequency, but is largely useless when transferred to clonally reproducing prokaryotes in which exchange of genes has been proven to occur, at variable frequencies, but over vast phylogenetic distances. In prokaryotes even the existence of discontinuities in the diversity space that would justify the existence of discrete units of classification, that is, species, is controversial [5•]. In classical definitions, one clone, the type strain, is considered to represent the species, but the existence of many other clones that are closely related is acknowledged. During the pre-genomic era the 'golden rule' of 70% DNA–DNA hybridization was considered a formal threshold for two strains to be included in a single species. This value of 70% could be already considered indicative of a singularity of prokaryotes since the values for animal or plant species are much higher [6].

The advent of genomics brought the pan-genome paradigm that refers to all the genes present in a single prokaryotic species. The pan-genome has two components with different evolutionary trajectories and specialization [7]. The core genome comprising genes that are found in all or most strains of the species and the flexible genome (sometimes called adaptive or accessory) made up of genes found only in some strains but with no orthologs in others. This two-tiered nature is radically different from the eukaryotic species genomes in which nearly all genes have orthologs, not only within the species but even in other distant orders or classes [8].

A remarkable feature of the prokaryotic flexible genome is its size. For any given strain only about 50% of the genome is core. Given that each strain contributes hundreds of novel (different) flexible genes, the size of the compounded flexible pool is enormous. In *Escherichia coli* with barely 100 genomes analyzed over 45 000 gene families have been described [9], this is already twice the gene diversity of humans. Even more remarkable is recent evidences that indicate that these large genetic pools coexist within microbial populations (groups of related individuals sharing the same habitat) [10,11].

The most convincing reports of the existence of well-defined species of prokaryotes come from metagenomics. A metagenome has genomic fragments from all the cells present in a microbial habitat (sample) and can be compared to each other or to a reference genome. Genome recruitment is the bioinformatic comparison of a genome (typically from a pure culture) with a metagenome from a sample in which this pure culture (or single cell genome) is expected to be present (e.g. the same sample used for isolation of the culture or the cell). Genome recruitment has been used since the beginning of metagenomics to assess microbial abundance patterns, but it can also be used to analyze the population genomic structure. Individual metagenomic reads form a cloud that is typically located from 95 to 100% identity beyond which the discontinuity or boundary of the species is clearly proven by the empty gap showing no metagenomic fragments [12]. This pattern was taken as a reflection of natural discontinuities in the sequence space, that is, natural species. The threshold of 95% identity coincides with the values established to define species by analysis of genomes of pure culture isolates [13]. Figure 1 displays two recruitment plots generated by mapping metagenomic sequence information from two different environments onto separate reference genomes to illustrate the commonality of recruitment patterns for two very different aquatic habitats and microbes. *Haloquadratum walsbyi* is an archaeon that inhabits NaCl saturated brines while *Alteromonas mediterranea* is a marine bacterium. In spite of the enormous difference between the two reference genomes/habitats, the plots are remarkably similar. This pattern is recurrent in most aquatic (relatively homogeneous) habitats [14], and might be extrapolated to other habitats as well [15].

The local pangenome

These recruitment patterns are actually reflections of the local pangenome of the population. The cloud of fragments recruited over 95% reveals the diversity of core genes of the cells in the population (and the error rate of the sequencing method used), that is, the gene pools shared by most strains within a species that is well conserved in synteny but varies at the sequence level within this range of 95–100%. It is important to emphasize that 5% divergence over a 3 Mb long genome is equivalent to 150 000 single nucleotide polymorphisms, that is, the core genomes of the population is already significantly variable at the sequence level. The other significant fact that is revealed in Figure 1 is that the reference genome does not recruit evenly and there are gaps or under-recruiting regions. The name metagenomic island was proposed [16••] to describe these regions that under-recruited from a metagenome in which a large part of the genome recruited over 95% nucleotide identity. Their presence reflects the flexible component of the reference genome. Although flexible genes are often interspersed throughout the core genome, most are concentrated in flexible genomic islands (fGIs). Given that each of the clonal lineages contains different genes at these locations these areas of the reference genome appear as under-recruiting and are largely the explanation for metagenomic islands. There are two different kinds of fGIs with different kinds of biological roles and modes of variation: replacement and additive (Figure 2).

Replacement fGIs are clusters of genes that although present in all clonal lineages at equivalent (syntenic) positions show no sequence homology to each other. However, the biological function for which they code is equivalent. They are often important for basic functions of the cell (if not essential for survival in nature) and all clonal lineages carry a version. The most paradigmatic example would be the gene cluster coding for the O-chain polysaccharide of the lipopolysaccharide of Gram negative bacteria. Additive fGIs on the other hand are a product of illegitimate recombination and it is conceivable that they are in a continuous state of change. They are located in a number of hotspots along the genome, normally associated with a tRNA gene. The tRNA acts as a target for illegitimate recombination in which a variable number of gene cassettes coding for a broad diversity of biological functions are inserted. Integrons are typical examples of additive fGIs with a specific mechanism of variation associated to the integration of gene cassettes [17], but there are many other genome evolution hotspots with different integration mechanisms.

Replacement genomic islands coding for polysaccharides

Most replacement fGIs contain gene clusters involved in the synthesis of polysaccharides. In bacteria polysaccharides synthetized at the external surfaces of the cell have a number of common elements [18]. They require genes involved in the synthesis of sugars that sometimes are enzymatically modified by acylation, methylation or otherwise. Other genes are involved in the specific bonding to UTP to form the UDP-sugar intermediate that carries the energy required for glycosidic linkage. Finally, some specific genes produce specific transporters such as the flippase Wzx that carry the oligosaccharides to their extracellular location [19]. Replacement genomic islands often code for combinations of genes that lead to the

Download English Version:

https://daneshyari.com/en/article/6131543

Download Persian Version:

https://daneshyari.com/article/6131543

Daneshyari.com