
Article:
Young, J Peter W (2016) Bacteria Are Smartphones and Mobile Genes Are Apps. Trends in microbiology. pp. 931-932. ISSN: 0966-842X

https://doi.org/10.1016/j.tim.2016.09.002

Forum

Bacteria Are Smartphones and Mobile Genes Are Apps

J Peter W Young

Department of Biology, University of York, York YO10 5DD, UK

Correspondence: peter.young@york.ac.uk (J.P.W. Young).

Keywords

accessory gene, core genome, smartphone

Abstract

Bacterial core and accessory genome components are analogous to the operating system and applications of smartphones. The core genome provides stable taxonomy and species lists, but phenotypes reflect the mobile pool of accessory genes. This suggests changes to the ways we define bacterial species and describe bacterial communities.

We need bacterial names so that we know what we are talking about, but do the names tell us what we really need to know about the bacteria? Current 'polyphasic' taxonomy recognises species as clusters of strains that share a high level of DNA similarity, but also requires some degree of phenotypic distinctness [1]. It is often difficult to find phenotypic features that reliably distinguish related species and, now that we have multiple genome sequences from many species, it is easy to see why. There are core genes that are shared by all members of a species, and often by related species and genera [2], but there are also many accessory genes that are present in some strains but absent from others [3]. The same accessory genes are found in related species, but they have phylogenies that are different from those of the core genes. While the core genes perform vital routine functions, the diversity of bacterial phenotypes and adaptations is largely provided by the mobile pool of accessory genes that are readily gained and lost, creating the bewildering levels of intraspecific diversity revealed by population genomics surveys. The term 'pan-genome' was proposed for the sum of the core genome plus all accessory genes that could be found in a species [4], but the number of possible genes seems indefinitely large for most species, meaning that the pan-genome is 'open', i. e. undefined.

Modern technology provides a metaphor that may help us to understand bacteria. At the level of population genomics, there are striking parallels between bacteria and personal computing devices such as smartphones. Phones are shipped from the factories in a limited range of models (species) with standard operating systems (core genes), but users download apps (accessory gene modules) from the internet (community gene pool) so there are soon millions of distinct combinations of software (genotypes). Rather than accessing a central app store, bacteria adapt through peer-to-peer networking (horizontal gene transfer), and the process is generally passive as far as the recipient is concerned. It is more akin to email spam – new genes are constantly arriving in the inbox, despite advanced spam filters and antiviral

software (CRISPR and restriction). Bacteria are constantly hitting the 'delete' key, but every now and then something really useful turns up and is kept.

Listing the bacterial species and genera in a community, using a gene such as 16S rRNA, is like assessing the market share of different phone models. This is important information for sales managers, but provides little insight into community function. Kenyan farmers, Japanese schoolgirls and German postdocs might buy the same phone models, but the apps used in their communities will be very different. In the same way, a recent metagenomics survey of gut microbiomes found that accessory gene pools varied between Fijian villages even though the species composition was not differentiated [5]. There is now widespread recognition that a species list is not enough and functional genes are what matters, so metagenomic studies like this are increasingly common.

Given what we now know about genomic variation, we can see a way forward to a taxonomic framework that reflects the reality of bacterial evolutionary processes. Within each taxon (genus, family, etc.) there is a set of core genes that is shared by virtually all strains [6], and most of these genes have a consistent phylogeny, at least above the species level. This is fortunate, as the huge effort to determine bacterial phylogeny using 16S rRNA genes would have been a failure if this gene had a phylogeny that was independent of the rest of the genome. There are, of course, many examples of individual core genes that have aberrant phylogenies in parts of the tree because they have been transferred between lineages, but the consensus is generally consistent. We can, therefore, ignore accessory genes and use core gene phylogeny to establish a stable bacterial taxonomy – a classification of the 'hardware' and 'operating systems' [3].

On the other hand, phenotypes are less useful for taxonomy because the majority of phenotypes that interest us are conferred by mobile accessory genes that are not shared by all members of a species [7]. The abilities to resist antibiotics, cause diseases, form symbioses, catabolise unusual substrates, etc., are obvious examples. Their determinants are usually carried on phages, plasmids, transposons, genomic islands, integrons, and so on. We care about these phenotypes, as it is important to know whether a bacterium will cause anthrax, fix nitrogen or degrade trinitrotoluene, but these are not fixed characteristics of a particular species. Researchers studying rhizobia realised long ago that the genes that confer the nitrogen-fixing symbiosis with legumes were accessory genes that moved within and between species, and adopted the term biovar, and later the more specific symbiovar (sv.), to describe strains that shared a particular set of symbiosis determinants [8]. Thus, nodulation of peas may be effected by *Rhizobium leguminosarum* sv. *viciae*, or by *Rhizobium pisi* sv. *viciae* or *Rhizobium laguerreae* sv. *viciae*, all with similar, perhaps identical, plasmid-encoded symbiosis genes [9]. On the other hand, *R. leguminosarum* sv. *trifolii* and *R. pisi* sv. *trifolii* denote other strains of these same species that nodulate clover instead because they have different symbiosis genes. Note that this naming convention reflects the real organisation of the bacterial genome: the species name is based on the core genes, while the symbiovar describes a module of accessory genes that is of particular interest. Of course, every strain has an array of other accessory modules for resistance, catabolism, etc., and if these were the focus of interest then biovars could be defined for these, too: each strain can belong to multiple biovars. This approach can readily be extended to other phenotypes in other bacteria, such as the very different pathogenicity phenotypes of the *anthracis* and *thuringiensis* biovars of *Bacillus cereus* [10], as we have argued elsewhere [3]. In a recent population genomics study, a previously unsuspected module for adaptation to serpentine soil was identified in *Mesorhizobium* by a genome-wide association approach [11]. These are the 'apps' that adapt bacteria to their ways of life.

Just as a computer application that proves universally useful may be assimilated into later versions of the operating system, so there is no fixed partition between accessory and core genes. Key differences between bacterial families, genera and species reflect different sets of genes that were once optional but have been 'domesticated' and become part of the core [3,6,12].

A metaphor cannot explain everything, but it creates a perspective that may stimulate new questions. Smartphone apps are neat packages that are designed to interface cleanly with the operating system and to operate largely independently of each other. To what extent is the seething accessory gene pool similarly modular? Can we recognise clusters of genes that function together as discrete apps? What features of accessory modules enhance their ability to plug and play in different genomic backgrounds, and what limits the host range in which they can perform? The choice of hardware and operating system may seem less exciting, but can affect the user experience. What kinds of specialisation do bacteria build into their core genomes, and how and why do these functions differ from those that are conferred by accessory genes? What are the processes by which genes become assimilated into the core? By the time we have all the answers, smartphones will be museum pieces and we will need a new metaphor.

References

1 Vandamme, P. and Peeters, C. (2014) Time to revisit polyphasic taxonomy. *Antonie van Leeuwenhoek* 106, 57–65

2 Young, J.P.W. *et al.* (2006) The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology* 7, R34

3 Kumar, N. *et al.* (2015) Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biology* 5, 140133–140133

4 Tettelin, H. *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome." *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955

5 Brito, I.L. *et al.* (2016) Mobile genes in the human microbiome are structured from global to individual scales. *Nature* 535, 435–439

6 Harrison, P.W. *et al.* (2010) Introducing the bacterial 'chromid': not a chromosome, not a plasmid. 18, 141–148

7 Ormeño-Orrillo, E. and Martínez-Romero, E. (2013) Phenotypic tests in *Rhizobium* species description: an opinion and (a sympatric speciation) hypothesis. *Systematic and Applied Microbiology* 36, 145–147

8 Rogel, M.A. *et al.* (2011) Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Systematic and Applied Microbiology* 34, 96–104

9 Remigi, P. *et al.* (2016) Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. *Trends in Microbiology* 24, 63–75

10 Zwick, M.E. *et al.* (2012) Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of *Bacillus anthracis*. *Genome Research* 22, 1512–1524

11 Porter, S.S. *et al.* (2016) Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic *Mesorhizobium*. *ISME J* DOI: 10.1038/ismej.2016.88

12 Lassalle, F. *et al.* (2015) Ecological speciation in bacteria: reverse ecology approaches reveal the adaptive part of bacterial cladogenesis. *Research in Microbiology* 166, 729–741