

Pan-genomes and cross-tree of life questions.

J. Joets

Institut Diversité, Écologie et Évolution du Vivant, UMR Génétique Quantitative et Evolution - Le Moulon - Université Paris-Saclay, INRAE, CNRS, AgroParisTech, 12 route 128, 91190 Gif-sur-Yvette, France

johann.joets@inrae.fr

Académie d'Agriculture de France, session, April 13th 2022.

Keywords: pan-genome, genome, dispensable genome, core genome, adaptation, species.

Until the beginning of the century, the gene content of the genome was considered as an invariant of a species. Cases of variation in the number of copies of a gene between individuals, or even the absence of a gene in some individuals, were known, but they were thought to be exceptions limited to certain families of genes such as the genes for resistance to pathogens in plants.

With the rapid progress of sequencing over the last twenty years, it has become possible to assemble and compare the sequences of complete genomes of several individuals of the same species. It quickly became obvious that the number of genes that could be absent or present from one individual to another one was much higher than expected. This facultative fraction of the genome, also called accessory or dispensable, is generally small in animals (5% in pigs, 15% in humans and up to 25% in mussels) and larger in cultivated plants (between 25% in sunflower or rape and up to 50% in soya). In microorganisms, the accessory genome represents from 20% (*Chlamydia*) to more than 90% of the genome (*Rickettsia*). The genome sequence of an individual is therefore no longer sufficient to represent all the genetic information of a species and the term pan-genome, grouping together the core genome, shared by all individuals and the accessory genome, was coined in 2005 by Tettelin¹ following his work in bacteria.

This presentation will introduce the concepts underlying pan-genomics. It will present the main mechanisms (horizontal gene transfers, genome duplications, deletions, dynamics of transposable elements) and evolutionary forces (selection, drift) that, depending on environmental constraints, shape pan-genomes. We will examine the particular case of the reduction in size of the accessory genome under the effect of domestication and human selection of cultivated species. Unlike the core genome, the majority of genes in the accessory genome are of unknown function. But when this function is known, it is generally related to the adaptation of organisms to environmental variations. An emblematic example is the *SubIA* gene that confers submergence tolerance to rice only when it is present². A better knowledge of the accessory genome will therefore be very useful for tackling the major environmental challenges facing agriculture in coming decades.

Many applications have already emerged from this knowledge. Among these, in plant improvement, new quantitative genetic methods are being developed to better consider this part of genetic variability that has been poorly exploited until now. In medicine and animal health, this knowledge should help identify new antibiotic targets. Finally, these advances will make it possible to revise the notion of species and to distinguish groups that were previously believed to form a single species for lack of sufficient criteria to distinguish them.

1 Tettelin H., Masignani. V, *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A.* **102**(39):13950-5 (2005). <https://doi.org/10.1073/pnas.0506758102>.

2 Xu, K., Xu, X., Fukao, T. *et al.* *SubIA* is an ethylene-response-factor-like gene that confers submergence tolerance to rice. *Nature* **442**, 705–708 (2006).

<https://doi.org/10.1038/nature04920>