Commentaries

Vol. 3, 2 (December 2019)
ISSN: 2532-5876

DOI: 10.13133/2532-5876_6.3

The Universal Phenotype

Alessio Martino,[a] *Alessandro Giuliani,*[b] *Antonello Rizzi* [a]

[a] *Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, Italy*

[b] *Department of Environment and Health, Istituto Superiore di Sanità, Rome, Italy*

Corresponding author: Alessandro Giuliani alessandro.giuliani@iss.it

Citation: Martino, A, Giuliani, A, Todde, V, Rizzi, A, 2019, "The Universal Phenotype", *Organisms. Journal of Biological Sciences*, vol. 3, no. 2, pp. 8-10. DOI: 10.13133/2532-5876_6.3

Commentary on: Martino, A, Giuliani, A, Todde, V, Bizzarri, M, Rizzi, A, 2019, "Metabolic Networks Classification Knowledge Discovery by Information Granulation" *Computers in Biology and Chemistry*, pp. 107187. DOI: 10.1016/j.compbiolchem.2019.107187

1. A Species-Specific Fingerprint

There is a basic (and often overlooked) difference of status between genotype and phenotype. It is now widely accepted the many-to-many relation between the two (Noble, 2011) so that the same genotype can support different phenotypes and the other way around. The complexity of phenotype/genotype relation is at the basis of many speculations and the shift from instructive to permissive character of the genotype is deeply changing our view of both physiology and evolution (Po et al. 2019, Braun E. 2015).

The difference that none (at least to our knowledge) took into consideration, is that while genotype (in its basic meaning of DNA genome sequence) is a universal feature of all the living organisms this universality does not hold for phenotypes. In other words, we cannot make a phenetic all-encompassing classification based on characters as leaves shape (animals have no leaves), brain size (only present in animals) or sensitivity to antibiotics (only pertaining to bacteria and fungi).

On the other hand, the strict state and tissue dependence of apparently low-level (and thus universal) phenotypes like gene expression, proteomic or metabolomics profiles does not allow for among species unbiased comparisons.

In order to have a universal phenotype shared at all the layers of biological organization that in turn remains sufficiently stable to be considered as a "species-specific" fingerprint, we must look at a property shared by all living organisms (with the only exception of viruses whose living organism status is in any case questionable): metabolism.

Clearly, we intend for 'metabolism' the entire set of enzyme-catalysed chemical reactions that 'can in principle' took place in an organism, while those actually taking place are strongly state dependent and thus highly unstable. Metabolism thus corresponds to the entire metabolic network having as nodes the small organic molecules present in the organism with edges between all molecule <A,B> pairs that can be transformed one into another by a single chemical reaction.

The possibility of 'going to phenotype from genotype with a single jump' offered by metabolic networks analysis, complementing phylogenetic and ecological cues, was already explored (Braun E. 2015, Borenstein et al. 2008, Lewis et al. 2012). Along similar ways the possibility to individuate the lethal mutations (Palumbo et al. 2005) by the sole analysis of metabolic network, is another fertile research avenue.

Notwithstanding this interest, all the scholars explored specific biological problems without testing the possibility of considering metabolic network wiring as a 'phenotypic barcode' of biological species exactly in the same terms ribosomal RNA 16S is a 'genotypic barcode' (Sarangi et al. 2019).

Metabolic network wiring is as stable as genotype given it stems from the enzymatic proteins encoded in

the genome of single organisms (and consequently on the kind of chemical reaction those enzymes catalyse). Notwithstanding that, the metabolic network representation does not simply equate the genotype for three main reasons:

1. The presence/absence of the enzyme A is independent of eventual changes in its sequence (many to one genotype-phenotype mapping)
2. The same chemical reaction can be catalysed by different enzymes so allowing for both multiple edges between two metabolites (simplification of phenotypes) and to the same wiring by means of different enzyme species (many to one genotype-phenotype mapping).
3. The same enzyme can be inserted in different pathways in different organisms (one to many genotype-phenotype mapping).

This paper demonstrates the mutual distances between metabolic networks wiring are able to both discriminate different species and to reconstruct the known phylogenetic relations at all levels of biological classification (Martino et al. 2019b).

This was possible by means of a very refined computational approach based on Granular Computing able to conjugate discrimination efficiency and the possibility to get biologically meaningful hints.

2. The Computational Approach

The breakthrough of the Granular Computing paradigm as a component of the vast toolbox of machine learning techniques, allowed the development of advanced pattern recognition systems able to deal with non-conventional data, such as networks (Martino et al. 2018). According to the latter, the vast majority of the information contained in structured domains (e.g. networks, sequences) can be preserved by extracting a set of meaningful "information granules" (e.g., portions of the networks) and then by describing each original network according to the number of occurrences of each information granule within the network itself. As per the paper commented, the puzzling point is: can different organisms can be discriminated according to statistically relevant chemical reactions drawn from their respective metabolic networks?

This *modus operandi* allows to solve a 'global problem' (i.e., discriminating amongst organisms having different cellular architecture, organisms belonging to different species or different kingdoms, and so on) by relying on 'atomic entities' such as individual chemical reactions in a metabolic pathway (i.e., individual edges in a metabolic network). This facet is particularly crucial if the "global problem" is hard to be analysed in its entirety in order to gather further insights, while "atomic entities" are not.

Furthermore, whether this "global problem" can be cast as an optimization problem, one gets the full benefit of the biological interpretability of the learning system, paving the way to so called Explainable Artificial Intelligent systems. In fact, one can drive the data-driven learning machine towards the selection of the smallest subset of edges which, at the same time, hold the vast majority of the information, hence endowing the highest discriminative power.

This summarizes the computational aspect in (Martino et al. 2019b), in which the authors faced four different problems located at different definition scales (discrimination between different cellular architectures – i.e., prokaryotes vs. eukaryotes, discrimination amongst different kingdoms, discrimination amongst animals, and discrimination amongst bacteria). Other than obtaining remarkable discrimination capabilities, which accounts for the reliability of the proposed metric, all four problems returned the most meaningful set of information granules (chemical reactions) which gave rise to biologically meaningful hints, fostering the use of metabolic networks as universal phenotype.

On a larger scale, this work fosters the cooperation between biologists and pattern recognition engineers, unleashing the potential of data-driven techniques towards interpretable models (Martino et al. 2018, Martino et al. 2019a).

3. Conclusion

Besides the generation of theoretically relevant hints (e.g. which specific chemical reactions happen only in eukaryotes) the practical application of the results reported in (Martino et al. 2019b) are particularly evident in ecological settings.

Each ecological space is defined by the role played by different actors (e.g., predators, preys, primary producers), the existence of a healthy environment depending upon a balanced mixture of different ecological niches occupation. In the case of microbial communities, especially in the case of internal ecologies of mucosa microbiota (Gilbert et al. 2018), the comparison between healthy and 'disease' microbial profiles is computed in terms of genotype barcode that, by definition, does not convey any biological information other than species identification.

Shifting to 'phenotype barcoding' could be much more informative because allows us to discriminate between function preserving (the same metabolic functions are carried out by different species) and function altering (a given metabolic activity is no more present) changes. This could yield a major achievement in terms of both pathology (human microbiota) and environmental sciences (microbial ecology of soil and water).

The Granular Computing approach used for solving such a very hard computational problem (needing millions of atomic comparisons as applied to metabolic networks, each having hundreds of nodes and consequently thousands of edges) allowed a dimensionality collapse and the subsequent enucleation of "discriminant edges". This is, at least in our opinion, an example of a sensible approach to Big Data that saves both the prediction efficiency and the biological interpretation, paving the way to a productive collaboration between different disciplines. Machine learning (and in particular Granular Computing Inductive Modelling) is not only a useful information processing toolset for "in silico" experiments, but represents a true paradigm revolution in science, towards an efficient and effective way to identify meaningful regularities in Big Data, for knowledge discovery and nature understanding.

References

Noble, D. 2011. "A theory of biological relativity: no privileged level of causation". *Interface focus*, 2(1), 55-64.

Po, Agnese, et al. 2019 " Phenotypic transitions enacted by simulated microgravity do not alter coherence in gene transcription profile." *npj Microgravity* 5.1 (2019): 1-13.

Braun, E. 2015 "The unforeseen challenge: from genotype-to-phenotype in cell populations". *Reports on Progress in Physics,* 78(3), 036602.

Borenstein, E., Kupiec, M., Feldman, M. W., & Ruppin, E. 2008. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of Sciences,* 105(38), 14482-14487.

Lewis, N. E., Nagarajan, H., & Palsson, B. O. 2012. "Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods". *Nature Reviews Microbiology*, 10(4), 291.

Palumbo, M. C., Colosimo, A., Giuliani, A., & Farina, L. 2005. "Functional essentiality from topology features in metabolic networks: a case study in yeast". *FEBS letters*, 579(21), 4642-4646.

Sarangi, A. N., Goel, A., & Aggarwal, R. 2019. "Methods for studying gut microbiota: A primer for physicians". *Journal of clinical and experimental hepatology*, 9(1), 62-73.

Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. 2018. "Current understanding of the human microbiome". *Nature medicine*, 24(4), 392.

Martino, A., Giuliani, A. & Rizzi, A. 2018. "Granular computing techniques for bioinformatics pattern recognition problems in non-metric spaces". In *Computational Intelligence for Pattern Recognition*, 53-81.

Martino, A., Giuliani, A. & Rizzi, A. 2019a. "(Hyper)Graph Embedding and Classification via Simplicial Complexes". *Algorithms*, 12(11), 223.

Martino, A., Giuliani, A., Todde, V., Bizzarri, M. & Rizzi, A. 2019b. "Metabolic Networks Classification and Knowledge Discovery by Information Granulation". *Computational Biology and Chemistry*, 107187.