The orthoBackbone: an evolutionarily-conserved backbone of genetic networks

Rion Brattig Correia[1,2,*], Paulo Navarro-Costa[1,3], and Luis M. Rocha[1,4]

[1] Instituto Gulbenkian de Ciência, Oeiras, Portugal
[2] CAPES Foundation, Ministry of Education of Brazil, Brasília, DF, Brazil
[3] Institute of Environmental Health, Lisbon Medical School, Lisbon, Portugal
[4] Department of Systems Science and Industrial Engineering, Binghamton University, Binghamton, NY, USA
*rionbr@gmail.com

Extracting knowledge from large gene expression datasets is a complex task, particularly if the analysis requires comparison between different species. We present the *orthoBackbone* (orthoBb), an evolutionarily-conserved subgraph of genetic networks across multiple species. We cast genetic networks as weighted graphs where nodes are genes of interest and edges correspond to strength of observed functional interactions. The latter can be derived from large-scale integrated databases of protein interactions, such as StringDB [1], that compile high-throughput experimental data, literature data mining, and genomic context predictions. We built a network for each species of interest—i.e. human, mouse and insect—where nodes and edges denote the interaction knowledge of the biological system being studied. Comparison between the networks of different species allows the identification of key biological processes that have been conserved throughout evolution. For this, matching nodes in the networks from different species have to be identified via orthologous relationships, i.e., genes that have kept the same functional role across evolution and speciation. This is cast as a multilayer network where each layer represents a particular species, and genes that are orthologous are connected across layers—a gene in one layer may connect to multiple orthologous genes in another. However, these networks are often large and dense and thus methods to reduce their complexity, integrate cross-species knowledge, and steer biological interpretation are much desired by domain experts. For these reasons, we developed the orthoBb, building upon our previous single-layer metric-backbone of complex networks [2, 3].

For each species network (i.e., each network layer) we compute its *metric backbone*. The metric-backbone is the subgraph that is sufficient to compute all shortest paths, thus removing edges that break the triangle inequality. This means that only metric edges are kept and thus edges that are redundant in the original species network, with regards to shortest-paths, are removed. This drastically lowers network density as 80-90% of all edges are deemed redundant and thus removed. In practice, we use a modified version of the Dijkstra algorithm to compute the metric-backbone. A mathematical description and a python implementation of the metric backbone can be seen in Simas *et al* [3]. Our biological interpretation is that the backbone represents the most important gene relationships in our species of interest. It follows that a fraction of these important gene interactions must have been evolutionarily-preserved, as they encode for functional relationships that are crucial for the survival of different species. Therefore,

the orthoBb is the subgraph of the metric-backbone where each edge has an analogous edge that connects orthologous genes in all other metric-backbone layers. Importantly, an assumption is made for cases where there is a $1 \rightarrow n$ gene orthology relation, as at least one backbone edge must be analogous in another species' backbone.

Preliminary results to be presented at the conference show that the orthoBb helps to identify fundamental biological programs underlying male fertility across species (Publication in preparation). As an example, in Fig 1 we plot the genetic circuit responsible for the control of the Wee1 cell cycle regulator across different vertebrate and invertebrate species. Despite a distance of more than 700 million years of evolution, in all three species the *ubiquitin* molecule regulates *Wee1* via the SCF complex. The orthoBb was fundamental for the identification of this circuit, as the *ubiquitin* molecule is connected to large hundreds of different target molecules. Both the complexity reduction and focus on fundamental evolutionarily-conserved processes ensured by the orthoBb uncovered the specific disruption of this circuit in a male infertility background.

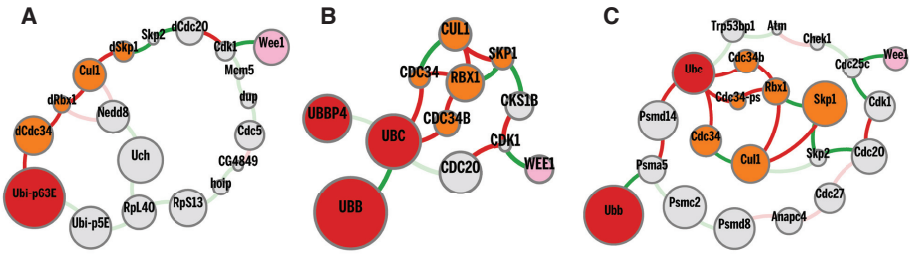

Fig. 1. Genes involved in the *Ubiquitin* $\rightarrow$ *Wee1* cell cycle regulation, shown as the metric and orthoBackbone genetic network in three different species. **A.** *Drosophila melanogaster* (the common fruit fly). **B.** *Mus musculus* (mouse). **C.** *Homo sapiens* (human). *Ubiquitin* genes (including orthologues) shown in red and *Wee1* genes in pink. Units of the mediating SCF complex are represented in orange. Only shortest paths among colored genes shown for simplification. Edges in the metric-backbone shown in green and edges in the orthoBackbone shown in red. Functional pathway shown with darker color.

References

1. von Mering, C., *et al.* (2005) *STRING: known and predicted protein-protein associations, integrated and transferred across organisms*. Nucleic Acids Res., 2015, 1; 33(Database issue):D433-7.
2. Simas, T. & Rocha, L. M. (2015) *Distance closures on complex networks*. Network Science, **3**(2), 227–268.
3. Simas, T., Correia, R.B. & Rocha L.M. (2021) *The distance backbone of complex networks*. Journal of Complex Networks. In Press. arXiv:2103.04668