Final Publishable Summary Report

EVOGREN proposes the study of the evolution of gene regulatory networks (GRN), extensions of the basic coalescent model and the development of a method to detect sites of GRN genes that evolve under positive selection pressure. The advent of modern DNA sequencing technology is the driving force to address questions that we could not answer in previous decades and eventually extend modern population genetics. Such an extension will be possible in the next few years because: (i) the developments in DNA sequencing technologies (NGS) are steadily contributing to an accelerating accumulation of *accurate* molecular sequence data; (ii) the field of bioinformatics provides many advanced kernels that can be used for the analysis of large scale genomic data; (iii) population genetics researchers realize that mathematical models involved in the analysis of genomic data are often too simplistic to provide additional insights into the evolutionary processes that give rise to the genetic structure of the present-day populations; (iv) it becomes evident that studying DNA sequences only, may be not adequate to understand the nature of forces that operate on the evolution of populations. More than four decades ago, it was first proposed that regulatory changes could lead to species-specific adaptations as well as phenotypic variability (King and Wilson, 1975; Richards 2008; Hernando-Herraez et al. 2015). Regulation of gene expression can be achieved by gene regulatory networks. Twenty years ago, Andreas Wagner proposed a simple model of gene regulatory network evolution which leads to substantial robustness to the detrimental effect of mutations (canalization). Wagner's seminal publication (Wagner 1996) inspired since then several other studies (Crombach and Hogenweg, 2008; Tsuda and Kawaka, 2010) that either extend the basic model or they study specific aspects and implications of Wagner's ideas.

EVOGREN contributes to the afore-mentioned extension of the population genetics theory by (i) providing scalable computational algorithms able to analyze and detect recent positive selection using sequences from thousands of individuals and millions of segregating sites (by a very fast implementation of the OmegaPlus algorithm); (ii) proposing extensions of the current coalescent theory (CoMuS - Multiple Species Coalescent; coalescent model of gene expression); (iii) the implementation of a method that uses the so-called 2-dimensional Site Frequency Spectrum (2D-SFS) to detect selection; (iv) the delineation of evolutionary forces shaping the human salivary adaptation, and (v) the evolution of human structural variants.

Algorithm for the simulation of gene regulatory networks

I implemented a GRN evolution algorithm to study (i) the evolution of the regulatory matrix under neutrality or selection pressure. The algorithm heavily uses optimizations (bitwise operations, integer arithmetics) that allows a fast execution time even for large populations (~10,000 individuals) for a large amount of generations (~20,000 generations). The algorithm is available from github (https://github.com/idaios/grn). Results from simulations suggest that patterns of polymorphisms when selection operates on gene regulatory networks resemble partially neutrality and/or soft selective sweeps and it is very difficult to detect them with current algorithms (SweeD, OmegaPlus) that detect recent strong positive selection.

Multi-Species Coalescent

I implemented a multi-species coalescent model to study the interaction between speciation and population genetics processes. With the new algorithm, called CoMuS (Papadantonakis S, Poirazi P, and Pavlidis P, MER accepted), we demonstrated the usage of CoMuS to infer parameter values such as speciation time or gene flow between Neandertal and *Homo sapiens*.

Fast implementation of the OmegaPlus algorithm to detect positive selection using the linkage-disequilibrium patterns.

I implemented a fast, scalable version of the OmegaPlus algorithm that results in better load balance and therefore faster analyses of very large datasets of DNA sequences. OmegaPlus exploits the linkage-disequilibrium (LD) to detect locations affected by positive selection. The new version of OmegaPlus (entitled as OmegaPlus-G) optimizes the load-balance and is suitable for very large datasets.

Study of the evolutionary forces on structural variants

We identified 427 polymorphic human deletions that are shared with archaic hominin genomes, approximately 87% of which originated before the Human–Neandertal divergence (ancient). Our analyses indicate that the genomic landscapes of both ancient and introgressed deletion variants were primarily shaped by purifying selection, eliminating large and exonic variants (Lin, Pavlidis, Karakoc, et al. 2015; Eaaswarkhanth, Pavlidis and Gokcumen, 2015).

Evolution of *muc7*

The salivary *MUC7* gene provides an exceptional opportunity for studying such impact, since it harbors copy number variable subexonic repeat sequences that encode for microbe-interacting protein domains. Here, we showed that *MUC7* has rapidly evolved under episodic positive selection in primates after it originated in the placental mammal ancestor. Analysis of pairwise distances among the majority of human haplotypes suggests admixture from an archaic African hominin (Xuo, Pavlidis, Flanagan et al. 2016, The 85th Annual Meeting of the American Association of Physical Anthropologists (2016)).

Conclusions and socio-economical impact

EVOGREN's main results provide novel ways to study the impact of evolutionary forces on GRNs. We developed fast algorithms to detect selection and study evolution of both single and multiple species simultaneously. Furthermore, with the study of structural variations in humans and neandertals we delineated the forces that affect the evolution of such variants.

Additionally, EVOGREN suggests novel questions and challenges. For example, we currently implement an extension of the simulation package that will work with a greater number of genes. This new project is assigned as one undergraduate and one MSc thesis. An additional undergraduate thesis is implemented that studies the evolution of bacterial genes that are in operons and thus it springs from EVOGREN. Therefore, EVOGREN contributes to strengthening computational evolutionary biology at FORTH and University of Crete. Furthermore, EVOGREN critically contributed to the election of the applicant as a researcher C at the Institute of Computer Science (ICS).

Website of the project: https://sites.google.com/site/evogren/home

Contact Details: pavlidisp@gmail.com, poirazi@imbb.forth.gr

References

- Anton Crombach and Paulien Hogeweg. Evolution of Evolvability in Gene Regulatory Networks, PLoS Comput Biol. 2008 Jul; 4(7): e1000112.
- King MC, Wilson AC. Evolution at two levels in humans and chimpanzees. Science. 1975;188: 107–116. pmid:1090005 doi: 10.1126/science.1090005
- Richards EJ. Population Epigenetics. Curr Opin Genet Dev. 2008 Apr;18(2):221-6. doi: 10.1016/j.gde.2008.01.014. Epub 2008 Mar 11.
- Tsuda ME, Kawata M. Evolution of Gene Regulatory Networks by Fluctuating Selection and Intrinsic Constraints. Babu MM, ed. *PLoS Computational Biology*. 2010;6(8):e1000873. doi:10.1371/journal.pcbi.1000873.