How To Innovate:
A Novel Method to Explore Evolutionary Novelty

Saurin Parikh[1], Anne-Ruxandra Carvunis[1,2]

1. Department of Computational & Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA
2. Center for Evolutionary Biology & Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA

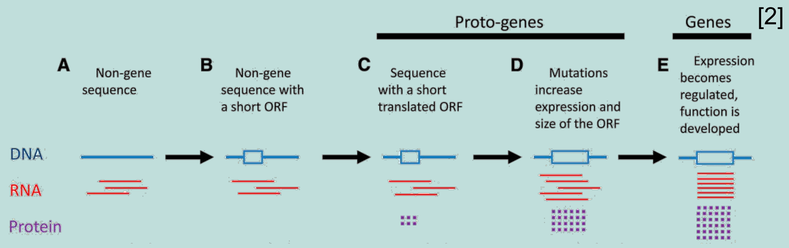
Pitt Medicine

The Carvunis Lab
Change and Innovation in Biological Systems

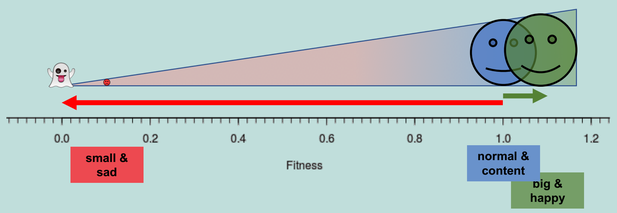
CENTER FOR EVOLUTIONARY BIOLOGY AND MEDICINE
CEBaM PITTSBURGH

Introduction

- The proto-gene model for *de novo* gene birth introduces a hypothesis of how novelty arises within the DNA [1]
- Proto-gene expression is predicted to provide the cell with adaptive potential making them better suited to their environment



- Detecting this adaptive potential in high-throughput screens (HTS) requires a robust and accurate bioinformatics pipeline, because:
  1. It is thought to be rare
  2. Produces a very small effect size, which makes it harder to detect



- Software packages like gitter [3], balony [4], HT colony grid analyzer [5] and the Matlab Colony Analyzer toolkit [6] have been developed to analyze HTS but none of them focus on increased fitness

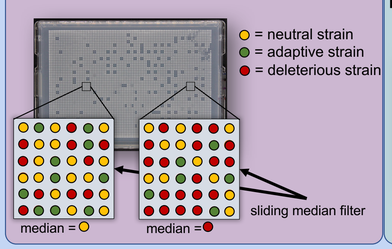
Goals

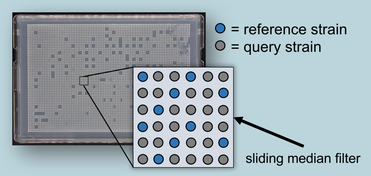
- To modify existing ways of examining fitness in HTS to detect adaptiveness
- To inspect whether proto-gene overexpression leads to an adaptive fitness
- To examine if proto-genes have a differential effect on fitness as compared to other genes

Method

1. Our current screening method relies on growing *Saccharomyces cerevisiae* on a solid agar surface in a grid pattern. These grids are then analyzed in parallel to quantify for phenotypes such as colony size.
2. We used over 4700 *S. cerevisiae* strains of the same genetic background where each contained an expression vector plasmid with a unique open reading frame (ORF), that could either be a gene or a proto-gene.
3. Using colony size as a proxy for fitness we aimed to identify those ORFs whose overexpression led to larger colony sizes as compared to the reference strain. This is what we defined as adaptive fitness.
4. We modified both *in silico* and *in vitro* approaches to make HTS more sensitive towards adaptive fitness. All analyses were done using the Matlab Colony Analyzer Toolkit.

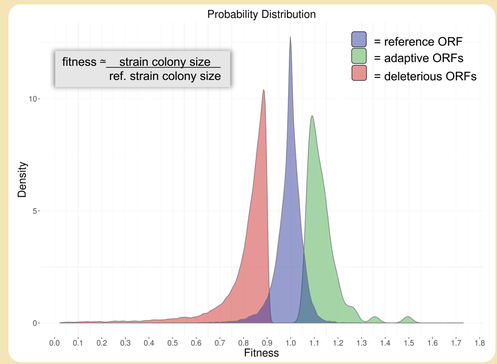
Old Method	Our Method
Keeping Track of All the Strains	
The toolkit has no built-in way to differentiate results of one strain from the other	Database management with the help of MySQL to maintain detailed records and layout of the screen
Dominant Phenotype ≠ Reference Phenotype	
The default spatial normalization method is strain agnostic and applies a median filter on a sliding window to determine background intensity for each point on the grid. This can result in a bias from the dominant phenotype.	Reference colony based normalization is implemented using the prior knowledge of experimental layout such that background intensity is determined using reference strain only. Making within and between plate comparisons more accurate.





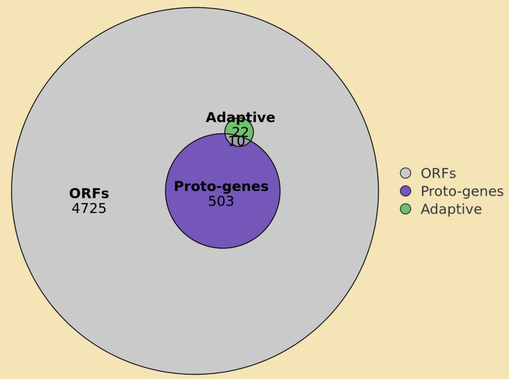
Results

Fitness distribution reveals that adaptive ORFs are detected



We see ORFs whose fitness distributions are significantly different than the reference

Adaptiveness of proto-genes and genes



Of the 32 adaptive ORFs that were detected, 22 were genes and 10 were proto-genes!

Our results show a 4-fold enrichment of adaptive ORFs in the proto-gene category shedding light on their potential role in adaptation and in exposing evolutionary novelty.

Conclusion

- Our method was successful in identifying adaptive ORFs
- These are promising results that indicate a differential effect of proto-genes on fitness of the organism
- A more complete exploration of these proto-genes will help us identify the inherent adaptive potential and understand how it plays an important role in adaptation and mechanisms of *de novo* gene birth.

References

1. Carvunis A-R, Rolland T, Wapinski I, et al. Proto-genes and *de novo* gene birth. *Nature*. 2012;487(7407):370-374. doi:10.1038/nature11184.
2. Frietze S, Leatherman J. Examining the Process of *de Novo* Gene Birth: An Educational Primer on "Integration of New Genes into Cellular Networks, and Their Structural Maturation." *Genetics*. 2014;196(3):593-599. doi:10.1534/genetics.113.160895.
3. Wagih O, Parts L. gitter: A Robust and Accurate Method for Quantification of Colony Sizes From Plate Images. *G3: Genes|Genomes|Genetics*. 2014;4(3):547-552. doi:10.1534/g3.113.009431.
4. Young BP, Loewen CJ. *Balony:* a software package for analysis of data generated by synthetic genetic array experiments. *BMC Bioinformatics*. 2013;14:354. doi:10.1186/1471-2105-14-354.
5. Collins SR, Schuldiner M, Krogan NJ, Weissman JS. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biology*. 2006;7(7):R63. doi:10.1186/gb-2006-7-7-r63.
6. Bean GJ, Jaeger PA, Bahr S, Ideker T. Development of Ultra-High-Density Screening Tools for Microbial "Omics." Hoheisel JD, ed. *PLoS ONE*. 2014;9(1):e85177. doi:10.1371/journal.pone.0085177.

Acknowledgements

We thank Nelson Coelho and John Iannotta for their contribution to the experimentation part of this study.
This work was supported by the grant R00-GM108865 from the National Institutes of Health awarded to Anne-Ruxandra Carvunis.

= saur1n