

# Transcription factor binding sites detection in *Paramecium*

Matthias Grenié \*, Jean-François Goût † and Michael Lynch †

\*ENS de Lyon, Département de Biologie, and †Indiana University, Biology Department, Lynch Lab

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed vehicula metus sapien. Suspendisse pulvinar, felis ut hendrerit aliquet, dui nisi bibendum erat, fermentum mattis enim nibh id arcu. Vestibulum ultrices eros sed odio tincidunt bibendum. Pellentesque fermentum ante vel nulla commodo fermentum. Vestibulum in augue sit amet libero viverra accumsan eu at magna. Sed at ligula quis nibh pharetra facilisis non eu libero. Suspendisse non quam sit amet massa luctus interdum sit amet in purus. Integer id orci elit, vitae sollicitudin lectus.

Keyword1 | Keyword2 | Keyword3

Abbreviations: SAM, self-assembled monolayer; OTS, octadecyltrichlorosilane

## Introduction

Structure of the introduction

- Whole Genome Duplications background, major evolutionary force
- the *Paramecium* project, why *Paramecium* is interesting, the aurelia complex
- Here, focusing on the computational part, developing pipeline, showed that etc.

Since Ohno first hypothesized the influence of Whole-Genome Duplications (WGD), scientists kept showing that a broad number of species experienced WGD at least once : yeast, Drosophila, Vertebrates, Salmonids, *Paramecium*. WGD are evolutionary event when the genome of a given individual is duplicated, meaning that the whole genome is in two copies, duplicated pairs of genes are paralogs. WGD may be involved in many evolutive radiations as it creates a context of loosen selection. According to the Duplication-Degeneration-Complementation model,

To understand the consequences of WGDs we have been studying the *Paramecium aurelia* complex. As one of the only

free-living eukaryotes studied, other than yeasts, *Paramecium* is a very attractive model. The diversity of the *Paramecium* ciliates is well studied.

We focused on four species of *Paramecium* : *P. biaurelia*, *P. sexaurelia*, *P. tetraurelia* and *P. caudatum* as an outgroup (see phylogenetic tree). The three *aurelia* species underwent two rounds of WGDs, WGD<sub>X</sub> (... years ago) and WGD<sub>Y</sub> (... years ago).

**Biological questions :** Do gene expression is linked, in *P.*, with specific motifs? How are TFBS affected by WGD? Are they conserved among species, is this linked to expression level? Conserved among each species? Is there a bias of TFBS usage in certain species?

## Results

## Discussion

Conservation among species. Major results is that. Divergent resolution of WGD → divergence in motifs?

Motifs detection should take phylogeny into account for comparative analysis. Not the same value.

## Materials and Methods

developped a whole pipeline (show simple pipeline graph)

- Families upstream sequences extraction
- CDSs extraction and alignment
- CDSs phylogenetic tree
- BigFoot identification (explanation of phylogenetic score and alignment score)
- MEME research
- Comparison MEME and BigFoot
- Identification of given motif in species genome
- Correlation between motifs and expression levels

**ACKNOWLEDGMENTS.** Thanks to M. Lynch and J.-F. Goût. Funding by NSF.

1. M. Belkin and P. Niyogi, Using manifold structure for partially labelled classification, *Advances in NIPS*, 15 (2003).
2. P. Bérard, G. Besson, and S. Gallot, Embedding Riemannian manifolds by their heat kernel, *Geom. and Fun. Anal.*, 4 (1994), pp. 374–398.
3. R.R. Coifman and S. Lafon, Diffusion maps, *Appl. Comp. Harm. Anal.*, 21 (2006), pp. 5–30.
4. R.R. Coifman, S. Lafon, A. Lee, M. Maggioni, B. Nadler, F. Warner, and S. Zucker, Geometric diffusions as a tool for harmonic analysis and structure definition of data. Part I : Diffusion maps, *Proc. of Nat. Acad. Sci.*, (2005), pp. 7426–7431.
5. P. Das, M. Moll, H. Stamati, L. Kavvaki, and C. Clementi, Low-dimensional, free-energy landscapes of protein-folding reactions by nonlinear dimensionality reduction, *P.N.A.S.*, 103 (2006), pp. 9885–9890.
6. D. Donoho and C. Grimes, Hessian eigenmaps : new locally linear embedding techniques for high-dimensional data, *Proceedings of the National Academy of Sciences*, 100 (2003), pp. 5591–5596.
7. D. L. Donoho and C. Grimes, When does isomap recover natural parameterization of families of articulated images?, *Tech. Report Tech. Rep. 2002-27*, Department of Statistics, Stanford University, August 2002.
8. M. Grüter and K.-O. Widman, The Green function for uniformly elliptic equations, *Man. Math.*, 37 (1982), pp. 303–342.
9. R. Hempel, L. Seco, and B. Simon, The essential spectrum of neumann laplacians on some bounded singular domains, 1991.
10. Kadison, R. V. and Singer, I. M. (1959) Extensions of pure states, *Amer. J. Math.* 81, 383-400.
11. Anderson, J. (1981) A conjecture concerning the pure states of  $B(H)$  and a related theorem. in *Topics in Modern Operator Theory*, Birkhäuser, pp. 27-43.
12. Anderson, J. (1979) Extreme points in sets of positive linear maps on  $B(H)$ . *J. Funct. Anal.* 31, 195-217.
13. Anderson, J. (1979) Pathology in the Calkin algebra. *J. Operator Theory* 2, 159-167.
14. Johnson, B. E. and Parrott, S. K. (1972) Operators commuting with a von Neumann algebra modulo the set of compact operators. *J. Funct. Anal.* 11, 39-61.
15. Akemann, C. and Weaver, N. (2004) Consistency of a counterexample to Naimark's problem. *Proc. Nat. Acad. Sci. USA* 101, 7522-7525.
16. J. Tenenbaum, V. de Silva, and J. Langford, A global geometric framework for nonlinear dimensionality reduction, *Science*, 290 (2000), pp. 2319–2323.
17. Z. Zhang and H. Zha, Principal manifolds and nonlinear dimension reduction via local tangent space alignment, *Tech. Report CSE-02-019*, Department of computer science and engineering, Pennsylvania State University, 2002.

Reserved for Publication Footnotes

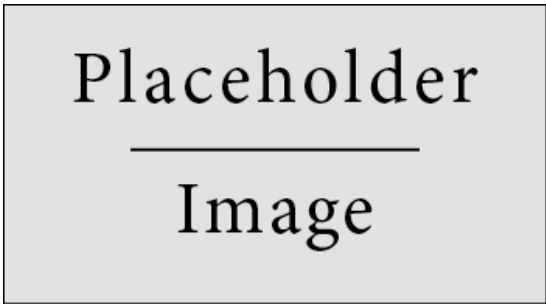


FIGURE 1. Figure caption

TABLE 1. Table caption

Treatments	Response 1	Response 2
Treatment 1	0.0003262	0.562
Treatment 2	0.0015681	0.910
Treatment 3	0.0009271	0.296