

Curriculum Vitæ

Macha Nikolski¹

born on the 22 November 1972, St Petersburg, Russia

Laboratoire Bordelais de Recherche en Informatique (LaBRI)

Université Bordeaux, 351, cours de la Libération, 33405 Talence, France

Tel. +33 (0) 54000 35 06

e-mail macha.nikolski@labri.fr

<http://www.labri.fr/perso/macha/>

Education and Employment

since 2010: **Head** of Bordeaux Bioinformatics Center; **Head** of Bioinformatics axis of the SIRIC BRIO

2010-2013: **Head** of Bioinformatics group within the MABioVis team, LaBRI, Bordeaux

2009 - 2010: **Research Scientist** at the CNRS Poncelet Lab, Moscow, Russia; **group leader** at the GosNII Genetika Institute

october 2009: **Habilitation à Diriger les recherches**, defended at the Bordeaux 1 University in front of M. Jens Stoye (Univ. Bielefeld, Germany), Betül Kirdar (Univ. Bogazici, Turkey), Michel Aigle (Univ. Lyon, France), Serge Dulucq (Univ. Bordeaux, France), Guy Melançon (Univ. Bordeaux, France), Pascal Weil (CNRS, Bordeaux, France). Title : *From Genomic to Functional Models*.

2002 - 2009: **Research Scientist** with the CNRS at the LaBRI, Bordeaux, France

2001 - 2002: **Research Scientist** at Synopsys, Boston, USA

1999 - 2001: **Assistant Professor** (ATER) at the University of Bordeaux, France

2000: **PhD** in Computer Science defended at the Bordeaux 1 University. Title : *Binary Decision Diagrams and Applications to Reliability Analysis*

1996: Engineering school diploma (**Masters**) at the Ecole Nationale d'Electronique, Informatique et Radioélectrique de Bordeaux (ENSEIRB)

1994: **Bachelors** (majoring in Computer Science) at the University of California at Berkeley

¹Note that prior to obtaining French citizenship my name was spelled Maria Nikolskaia.

Research Activities

I am a computational biologist working mainly in genomics. Computational genomics requires the development of algorithms, statistical methodology and mathematical foundations, and a large part of my research is therefore on methods. Validating these methods, that is answering biological questions by using computational approaches is central in my work. Therefore, method development goes always hand in hand with application, which in practical terms requires close collaboration with colleagues in the life sciences.

My career began in comparative genomics, and initially I was interested in genome annotation, determination of protein families using consensus clustering techniques and genome rearrangements. In this context I contributed to the yeast genomes' sequencing consortia. In parallel I have worked on building rigorous models of the dynamical behaviour of biological systems (eg. networks), on how is it possible to combine different models and on establishing properties about models' behaviors.

Currently, I'm particularly interested in applications of high-throughput sequencing data. My group develops methods for Next Generation Sequencing data analysis and de novo sequence assembly and its downstream analysis (genome assembly, variant detection, diversity analysis, biomarker identification, functional analysis). Methodological solutions are often a mix of algorithmics and machine learning.

I have supervised 8 PhD students. Five of them have successfully defended (A. Sarkar in 2012, N. Vyahhi in 2010, H. Soueidan in 2009, G. Jean in 2008 and F. Iragne in 2007) and are all currently employed either as a permanent researcher (3), post-doctoral researcher (1) or as a research engineer (1). Three of them still pursue their PhDs: T. Hume, L-A. Schmitt and J. Rudewicz. I have supervised more than 10 Master's students and 5 postdocs. I have been a dissertation reviewer and opponent for 5 PhDs. Since starting with the CNRS in 2002, I have 26 papers in refereed international journals (h-index 13 and 2001 total citations according to Google Scholar) and am actively involved in the development of various software tools.

Involvement in scientific community

- I am strongly involved in three initiatives for developing **international standards**: Hupo PSI (Protein Standards Initiative), MIASE (Minimum Information About a Simulation Experiment) and COSMOS (COordination Of Standards In MetabOlomicS);
- Member of the Bordeaux SIRIC (SItes de Recherche Intégrée sur le Cancer) **Bioinformatics Steering Committee**
- Member of the LaBRI **Scientific Council**
- Co-organizer of the **Recomb-CG 2013** conference in Lyon;
- Co-organizer of a Special Session **BioNetVis** at the **ISMB 2013** conference in Berlin;
- Organizer of the **Workshop on Next Generation Sequencing Technologies**, Zvenigorod, Russia, 2010;

- Member of the International Society for Computational Biology (ISCB) since 2005;
- Member of the **Scientific Program Committee** of Recomb-CG 2013, WABI 2012, MCCB 2009, WABI 2009, AIBio 2008, JOBIM 2007;
- Reviewer for **Scientific Journals** TSI, TCS, TCBB, Plos. Comp. Bio, BMC Bioinformatics;
- Reviewer for **Grant Calls** by ANR (PEPII, AMPBio, Masses de Données, PEPII, PDOC) and EU (Offensive Sciences RMT);
- Member of **Hiring Committees** (Université d'Evry 2005-2007, INRIA Saclay 2008-2009, Université de Bordeaux 2009 and 2011);

Grants

I am currently a PI the the following grants:

- FP7 Coordinated Action Cosmos *COordination Of Standards In MetabOlogicS*, 2012-2015 (PI for Bordeaux);
- ANR Chaire d'Excellence ABRIWG *Apricot Whole Genome Sequence Genotyping*, 2011-2014 (in charge of the bioinformatics WP);
- Infrastructure d'Avenir *Institut Francais de Bioinformatique*, 2012-2022 (member through ReNaBi-SO).
- Investissement d'Avenir Cloud computing, Big data, *BioDataCloud*, 2013-2015
- Peps IdEx / CNRS *PepTeam: A novel biomarker and targeting agents identification protocol*, 2013-2014

I have been in charge of, or participated in the coordination of the following grants:

- PI for grant from the Russian Agency for Science Education and Innovation *Return of Competence*, 2009-2010;
- PI for Bordeaux in the *Yeast Systems Biology Network* (CA under the FP6), 2005-2008;
- In charge of ACI "Jeunes Chercheurs" *Saccaromyces siliceus*, 2003-2005.

I have participated in the following grants:

- ANR *Evolmyco*, 2008-2012 (step in for Antoine de Daruvar);
- GDR *Génolevures* 2, 3 and 4 (CNRS GDR 2354), 2002-2010;
- ANR Blanc *Genarise* (ANR-05-BLAN-0331), 2005-2008;
- ACI IMPBio *Génolevures En Ligne* (IMPB114), 2003-2005.

Selected publications

- [1] D. Alonso-Aleman, A. Barre, S. Beretta, P. Bonizzoni, M. Nikolski, and G. Valiente. Further steps in tango: improved taxonomic assignment in metagenomics. *Bioinformatics*, 2013.
- [2] H. Soueidan, F. Maurier, A. Groppi, P. Sirand-Pugnet, F. Tardy, C. Citti, V. Dupuy, and M. Nikolski. Finishing bacterial genome assemblies with mix. *BMC Bioinformatics*, 14(Suppl 15):S16, 2013.
- [3] A. Sarkar, M. Nikolski, and P. Durrens. Introducing individual family based variability in protein family expansion. *International Journal of Bioinformatics Research and Applications (IJBRA)*, 9(2):121–133, 2013.
- [4] H. Ferry-Dumazet, L. Gil, C. Deborde, A. Moing, S. Bernillon, D. Rolin, M. Nikolski, A. de Daruvar, and D. Jacob. Mery-b: a web knowledgebase for the storage, visualization, analysis and annotation of plant 1h-nmr metabolomic profiles. *BMC Plant Biology*, 11:104:10.1186/1471–2229–11–104, 2011.
- [5] A. Sarkar, H. Soueidan, and M. Nikolski. Identification of conserved gene clusters in multiple genomes based on synteny and homology. *BMC Bioinformatics*, 12(Suppl 9):S18, 2011.
- [6] D. Waltemath, R. Adams, D. Beard, F. Bergmann, U. Bhalla, R. Britten, V. Chelliah, M. Cooling, J. Cooper, E. Crampin, A. Garny, S. Hoops, M. Hucka, P. Hunter, E. Klipp, C. Laibe, A. Miller, I. Moraru, D. Nickerson, P. Nielsen, M. Nikolski, S. Sahle, H. Sauro, H. Schmidt, J. Snoep, D. Tolle, O. Wolkenhauer, and N. LeNovère. Minimum information about a simulation experiment (miase). *PLoS Computational Biology*, 7:e1001122. doi:10.1371/journal.pcbi.1001122, 2011.
- [7] The Genolevures Consortium. Comparative genomics of protoploid genomes of saccharomycetaceae defines the orthologous gene set and basic yeast proteome repertoire. *Genome Research*, 19:1696–709, 2009.
- [8] D. Sherman, T. Martin, M. Nikolski, C. Cayla, J.L. Souciet, and P. Durrens. Genolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes. *Nucleic Acids Research*, 37(database issue):D550–D554, 2009.
- [9] G. Jean, D. Sherman, and M. Nikolski. Mining the semantics of genome super-blocks to infer ancestral architectures. *Journal of Computational Biology*, 16:1267–84, 2009.
- [10] H. Soueidan, G. Sutre, and M. Nikolski. Qualitative transition systems for the abstraction and comparison of transient behavior in parametrized dynamic models. In *Proceedings of CMSB’09, Lecture Notes in Bioinformatics 5688 (ISBN 978-3-642-03844-0)*, pages 313–327, 2009.

- [11] N. Vyahhi, A. Goeffon, M. Nikolski, and D. Sherman. Swarming along the evolutionary branches sheds light on genome rearrangement scenarios. In *accepted for publication in GECCO*, 2009.
- [12] P. Durrens, M. Nikolski, and D. Sherman. Fusion and fission of genes define a metric between fungal genomes. *Plos Computational Biology*, 4(10), 2008.
- [13] M. Cvijovic, H. Soueidan, D. Sherman, E. Klipp, and M. Nikolski. Exploratory simulation of cell ageing using hierarchical models. In *Genome Informatics Series, ISSN: 0919-9454*, volume 21, pages 114–125, 2008.
- [14] A. Goeffon, M. Nikolski, and D. Sherman. An efficient probabilistic population-based descent for the median genome problem. In *Proceedings of GECCO 2008 (ISBN:978-1-60558-130-9)*, pages 315–322, 2008.
- [15] G. Jean, D. Sherman, and M. Nikolski. Reconstruction and visualization of genome rearrangements within the kuyveromyces. Proceedings of the ESF-EMBO Symposium on Comparative Genomics of Eukariotic Microorganisms, Poster, 2008.
- [16] D.J. Sherman H. Soueidan and M. Nikolski. Biorica: A multi model description and simulation system. In *Proceedings of the 2nd Foundations of Systems Biology in Engineering (FOSBE)*, pages 279–287, Stuttgart, 2007. Fraunhofer IRB Verlag. ISBN 978-3-8167-7436-5.
- [17] G. Jean and M. Nikolski. Genome rearrangements: a correct algorithm for optimal capping. *Information Processing Letters*, 104(1):14–20, 2007.
- [18] F. Iragne, M. Nikolski, and D. Sherman. Extrapolation of metabolic pathways as an aid to modelling completely sequenced non-Saccharomyces yeasts. *FEMS Yeast Res.*, 8:132–139, 2007.
- [19] M. Nikolski and D. Sherman. Family relationships: should consensus reign? - consensus clustering for protein families. *Bioinformatics*, 23(2):e71–e76, 2007.
- [20] D. Sherman, P. Durrens, E. Beyne, M. Nikolski, and J.-L. Souciet. Genolevures complete genomes provide data and tools for comparative genomics of hemiascomycetous yeasts. *Nucleic Acids Research*, 34(Database Issue):D432–D435, 2006.
- [21] A. Barre, V. Jouffe, C. Lartigue, M. Nikolski, A. Blanchard, and P. Sirand-Pugnet. Annotation transfert based on orthology relationships: reannotation of mycoplasma genomes from the pneumoniae group. JOBIM’06 poster, 2006.
- [22] F. Iragne, M. Nikolski, B. Mathieu, D. Auber, and D. Sherman. ProViz: protein interaction visualization and exploration. *Bioinformatics*, 21(2):272–274, 2005.

- [23] B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. de Montigny, C. Marck, C. Neuvéglise, E. Talla, N. Goffard, L. Frangeul, M. Aigle, V. Anthouard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J.-M. Beckerich, E. Beyne, C. Bleykasten, A. Boiramé, J. Boyer, L. Cattolico, F. Confanioleri, A. de Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye, C. Hennequin, N. Jauniaux, P. Joyet, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J.-M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potier, G.-Franck Richard, M.-L. Straub, A. Suleau, D. Swennene, F. Tekaia, M. Wésolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C. Bouchier, B. Caudron, C. Scarpelli, C. Gaillardin, J. Weissenbach, P. Wincker, and J.-L. Souciet. Genome evolution in yeasts. *Nature*, 430(6995):35–44, July 2004.
- [24] D. Sherman, P. Durrens, E. Beyne, M. Nikolski, and J.-L. Souciet. Génolevures: comparative genomics and molecular evolution of hemiascomycetous yeasts. *Nucleic Acids Research*, 32 Database issue:D315–D318, 2004.
- [25] H. Hermjakob, L. Montecchi-Palazzi, G. Bader, J. Wojcik, L. Salwinski, A. Ceol, S. Moore, S. Orchard, U. Sarkans, C. von Mering, B. Roechert, S. Poux, E. Jung, H. Mersch, P. Kersey, M. Lappe, Y. Li, R. Zeng, D. Rana, M. Nikolski, H. Husi, C. Brun, K. Shanker, SG. Grant, C. Sander, P. Bork, W. Zhu, A. Pandey, A. Brazma, B. Jacq, M. Vidal, D. Sherman, P. Legrain, G. Cesareni, I. Xenarios, D. Eisenberg, B. Steipe, C. Hogue, and R. Apweiler. The hupo psi’s molecular interaction format—a community standard for the representation of protein interaction data. *Nature Biotech*, 22(2):177–83, Feb 2004.
- [26] M. Nikolskaia and L. Nikolskaia. Size of OBDD representation of 2-level redundancies functions. *Theoretical Computer Science*, 255(1-2):615–625, 2001.
- [27] P.F. Williams, M. Nikolskaia, and A. Rauzy. Bypassing BDD construction for reliability analysis. *Information Processing Letters*, 75(1–2):85–89, 2000.