



ABCDE



Scientific Report

First name / Family name

Joana Gonçalves

Nationality

Portuguese

Name of the *Host Organisation*

Centrum Wiskunde & Informatica

First Name / family name
of the *Scientific Coordinator*

Gunnar Klau

Period of the fellowship

01/09/2013 to 31/08/2014



I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

My fellowship was carried out in the Algorithmic Computational Biology group within the Life Sciences research theme at CWI, the Netherlands Research Institute for Mathematics and Computer Science, under the supervision of Prof. Gunnar Klau.

Throughout the fellowship, I focused mostly on developing novel algorithms (and software tools) for the discovery of local patterns in time series data.

I developed an efficient algorithm for biclustering of time series data with the ability to group similar patterns occurring with potential temporal delays (time lags), termed LateBiclustering. The following steps were involved in this work: design and implementation of the algorithm; deriving of theoretical proofs of correctness and analysis of complexity; collection of suitable data; design and conduction of experiments; analysis of results; writing up of the paper and responses to reviewers' comments. This research resulted in a first-author journal publication in IEEE/ACM Transactions on Computational Biology and Bioinformatics.

In parallel, I contributed to the integration of the TFRank method I previously proposed to prioritise transcription factors for groups of target genes of interest into YEASTRACT, a database and information system for the analysis of gene and genomic transcriptional regulation in *Saccharomyces cerevisiae* (baker's yeast). This work resulted in a co-authored journal publication in the Nucleic Acids Research's 2014 Database Issue.

Another project I invested in was a major upgrade to a free software tool I previously developed for the biclustering analysis of gene expression time series data, BiGGES-TS. This included the following upgrades: integration of the LateBiclustering algorithm enabling the efficient discovery of time-lagged biclusters; development and integration of specific visualisations for time-lagged biclusters; major changes to the calculation and presentation of statistical enrichment of biological functions for all bicluster types. This resulted in the preparation of a manuscript for submission to the Nucleic Acids Research's 2015 Web Server Issue.

Furthermore, I wrote a research proposal entitled "Revealing the Building Blocks of Complex Systems". This proposal was submitted for consideration for a Veni grant, included within the Innovational Research Incentives Scheme of NWO, the Netherlands Organisation for Scientific Research.

Additionally, I worked further on an algorithm for temporal three-dimensional clustering of multiple time series data matrices. The manuscript describing this algorithm is currently in preparation and will be submitted to a relevant journal in the field of Computational Biology in the near future.



Finally, I presented my work at the following events / venues:

04/10/2013 - Life Sciences group seminar, CWI, Amsterdam, Netherlands.

31/10/2013 - ERCIM ABCDE Annual Seminar III, Athens, Greece.

14/11/2013 – Computational Cancer Biology group seminar, Netherlands Cancer Institute (NKI), Amsterdam, Netherlands.

05/12/2013 - Database Architectures think tank, CWI, Amsterdam, Netherlands.

15/04/2014 - Pattern Recognition and Bioinformatics group, Delft University of Technology, Delft, Netherlands.

17/06/2014 - Delft Bioinformatics Lab interview, TUDelft, Delft, Netherlands.

21/08/2014 - Swedish Institute of Computer Science, Stockholm (Kista), Sweden.

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Teixeira MC, Monteiro PT, Guerreiro JF, **Gonçalves JP**, Mira NP, dos Santos SC, Cabrito TR, Palma M, Costa C, Francisco AP, Madeira SC, Oliveira AL, Freitas AT and Sá-Correia I (2014) The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*, *Nucleic Acids Research*, 42(D1):D161-D166. ([doi: 10.1093/nar/gkt1015](https://doi.org/10.1093/nar/gkt1015))

Goncalves JP and Madeira SC (2014) LateBiclustering: Efficient Heuristic Algorithm for Time-Lagged Bicluster Identification, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 11(5):801-813. ([doi: 10.1109/TCBB.2014.2312007](https://doi.org/10.1109/TCBB.2014.2312007))

Gonçalves JP and Madeira SC (?) An efficient three-dimensional temporal clustering algorithm, in preparation.

Gonçalves JP and Madeira SC (?) BiGGesTS 2.0: upgraded temporal and time-lagged biclustering analysis, in preparation.

III – ATTENDED SEMINARS, WORKHOPS, CONFERENCES

- ABCDE Annual Seminar III, Athens, Greece, 31/10-01/11/2013.
- “From Molecular Networks to Phenotypes” (SB@NL 2013), Egmond aan Zee, Netherlands, 25-26/11/2013. (<http://www.ncsb.nl/sbnl2013>)
- 20th ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD 2014), New York City, NY, USA, 24-27/08/2014. (<http://www.kdd.org/kdd2014/>)



IV – RESEARCH EXCHANGE PROGRAMME (REP)

Research Exchange Programme 1

University of Geneva / Swiss Institute of Bioinformatics (SIB)

7-11 July 2014

My first Research Exchange Programme visit was to the University of Geneva and the Swiss Institute of Bioinformatics (SIB) in Geneva, Switzerland, on 7-11 July 2014. My host was Dr. Frédérique Lisacek, the senior researcher responsible for Proteome Informatics at SIB. During the visit I met with Dr. Lisacek, who introduced me to the different research themes and groups at SIB and referred me to potential partners for collaboration. Additionally I met with Prof. Patrick Ruch, leader of the Text Mining group at SIB, with whom I had fruitful discussions about the work I developed during my fellowship, but also current advances in text mining with applications to biology. Finally, I had a meeting with Dr. Daniel Keller, from the École Polytechnique Federale de Lausanne (EPFL), Lausanne, who kindly introduced me to their work in the context of the Blue Brain Project. On all occasions, I had the opportunity to pitch my work.

Research Exchange Programme 2

Swedish Institute of Computer Science (SICS)

18-22 August 2014

My second Research Exchange Programme visit was to the Swedish Institute of Computer Science (SICS), in Stockholm (Kista), Sweden, on 18-22 August 2014. My host was Dr. Daniel Gillblad, leader of the Decisions, Networks and Analytics Laboratory (DNA lab) at SICS. During the visit, Dr. Gillblad introduced me to the different research themes and projects within the group. I also had the opportunity to present extensively on the research developed during my ERCIM fellowship, which motivated a more focused meeting with Dr. Ian Marsh on the overlap between my work and the research being done at the DNA lab. Additionally, I had fruitful discussions with other ERCIM fellows hosted by the DNA lab and other groups at SICS, some of which I had met on the occasion of the ABCDE Annual Seminar III in Athens.