



ABCDE



Scientific Report

First name / Family name

Murray PATTERSON

Nationality

Canadian

Name of the *Host Organisation*

Centrum Wiskunde & Informatica
(CWI)

First Name / family name
of the *Scientific Coordinator*

Gunnar KLAU

Period of the fellowship

01/02/2013 to 31/01/2014

I – SCIENTIFIC ACTIVITY DURING YOUR FELLOWSHIP

The major project for this period of the fellowship, as stated in the research training programme for this period, was to come up with a method for assembling haplotypes from long-read sequencing data, that is, reads containing many single nucleotide polymorphisms (SNPs). The inherent difficulty in haplotype assembly comes from errors in the reads, and so one criterion (that we use) for judging the quality of an inferred haplotype is the one that involves correcting the minimum number of errors in the reads – the minimum error correction (MEC). We identified this project because sequencing technologies are moving towards the direction of longer reads, and as far as we know, no currently existing method can manage such data in a practical setting.

Here we developed a novel dynamic programming approach to haplotype assembly. It is the first approach that yields an exact solution for the MEC criterion in runtime linear in the number of SNPs per read, making it suitable for reads from future generation sequencing technologies. In a simulation study, our method has been shown to outperform existing reads-based assembly methods, but also performs well against other types of methods, such as statistical phasing methods. This paper was accepted for publication at RECOMB 2014 (see II – Publications).

A small side project that I worked on here at CWI was on the mapping of paralogs in protein-protein interaction (PPI) networks. Most methods tackle this problem using the protein (sequence) similarity matrix, or a phylogenetic tree on the proteins (usually constructed from this matrix). What we did is to break down such a mapping into its elementary units, based on evidence (from this matrix) for pairs of proteins that may be co-evolving – what we coined as 'units of co-evolution'. This paper was accepted for publication at RECOMB Comparative Genomics 2013 (see II – Publications).

II – PUBLICATION(S) DURING YOUR FELLOWSHIP

Cedric Chauve*, Murray Patterson* and Ashok Rajaraman*. Hypergraph covering problems motivated by genome assembly questions. In Proceedings of the *International Workshop on Combinatorial Algorithms (IWOCA, Rouen, France, 2013)*, LNCS **8288**:428–432 (2013).

Murray Patterson, Gergely Szöllösi, Vincent Daubin, Eric Tannier. Lateral gene transfer, rearrangement, reconciliation. *BMC Bioinformatics* 2013, **14**(Suppl 15):S4 (15 October 2013). see <<http://www.biomedcentral.com/1471-2105/14/S15/S4>>

Mohammed El-Kebir*, Tobias Marschall*, Inken Wohlers*, Murray Patterson, Jaap Heringa, Alexander Schönhuth and Gunnar W. Klau. Mapping proteins in the presence of paralogs using units of coevolution. *BMC Bioinformatics* 2013, **14**(Suppl 15):S18 (15 October 2013). see <<http://www.biomedcentral.com/1471-2105/14/S15/S18>>

Murray Patterson*, Tobias Marschall*, Nadia Pisanti, Leo van Iersel, Leen Stougie, Gunnar W. Klau+ and Alexander Schoenhuth+. WhatsHap: Haplotype Assembly for Future-Generation Sequencing Reads. To appear at RECOMB 2014 – see <<http://www.compbio.cmu.edu/recomb/acceptedpapers.html>>

III – ATTENDED SEMINARS, WORKSHOPS, CONFERENCES

Netherlands Bioinformatics Centre (NBIC) Conference, Lunteren, Netherlands, 2013 <<http://www.nbic.nl/about-nbic/nbic-conferences/nbic-2013/>>

Cologne-Twente Workshop (CTW), Enschede, Netherlands, 2013 <<http://wwwhome.math.utwente.nl/~ctw/>>

International Workshop on Combinatorial Algorithms (IWOCA), Rouen, France, 2013 <<http://iwoca2013.colloques.univ-rouen.fr/>>

Intelligent Systems for Molecular Biology (ISMB), Berlin, Germany, 2013 <<http://www.iscb.org/ismbecb2013>>

RECOMB Comparative Genomics, Lyon, France <<http://rcg2013.sciencesconf.org/>>

Utrecht Graphs Workshop, Utrecht, Netherlands, 2013 <<http://www.staff.science.uu.nl/~Kang0002/UGW2013.html>>

Systems Biology Netherlands (SBNL), Egmond aan Zee, Netherlands, 2013 <<http://www.ncsb.nl/sbnl2013>>

IV – RESEARCH EXCHANGE PROGRAMME (REP)

N/A