

COST Action BM1006 (SeqAhead): MC Business Meeting and Scientific Meeting

Andreas Gisel¹, Teresa Attwood², Jacques van Helden³, José R. Valverde⁴, Ana Conesa⁵, Ralf Herwig⁶, Eija Korpelainen⁷, Veli Mäkinen⁸, Steve Pettifer⁹, Alberto Policriti¹⁰, Thomas Svensson¹¹, Gert Vriend¹², Erik Bongcam-Rudloff¹³

¹CNR, Institute for Biomedical Technologies, Bari, Italy

²Faculty of Life Sciences and School of Computer Science, University of Manchester, Manchester, United Kingdom

³Lab. Technological Advances for Genomics and Clinics, Université d'Aix-Marseille, Marseille, France

⁴Centro Nacional de Biotecnología, CSIC, Madrid, Spain

⁵Centro de Investigaciones Príncipe Felipe, Valencia, Spain

⁶Max Planck Institute for Molecular Genetics, Berlin, Germany

⁷IT Center for Science, CSC, Helsinki, Finland

⁸Department of Computer Science, University of Helsinki, Helsinki, Finland

⁹School of Computer Science, The University of Manchester, United Kingdom

¹⁰University of Udine, Italy

¹¹Karolinska Institutet, Stockholm, Sweden

¹²Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands

¹³Swedish University of Agricultural Sciences, Uppsala, Sweden

SeqAhead is a COST Action created by a group of researchers involved in the development of Next Generation Sequencing (NGS) data-analy-

sis software and pipelines. The primary objective of SeqAhead is to develop a coordinated action plan to help the scientific community to handle the flood of NGS data in an efficient and coherent manner, using state-of-the-art bioinformatics. Establishment of a strong European network of NGS, data-analysis and informatics centres will facilitate and stimulate the exchange of data, protocols, software, experiences and ideas (^{1,2}).

Following its kick-off meeting on 13 March 2011, SeqAhead organised its first Management Committee (MC) meeting, combining the event with a scientific meeting and Working Group (WG) discussions, from 7-9 November 2011, in Brussels, Belgium.

The first part of the 3-day event was the official MC meeting. The local organiser, Jacques van Helden, and the Action Chair, Erik Bongcam-Rudloff, opened the event. Erik Bongcam-Rudloff gave an overview of the Action, pointing out that it had grown from the 18 member countries represented at the kick-off meeting to 22 signatories, highlighting both the relevance of, and the need for, such an Action within the biological community.

The budget for the first year of the Action was also presented. This included 3 meetings, one training school, 2 Short-Term Scientific Missions (STSMs), and a variety of other ad hoc dissemi-



All participants of the COST scientific meeting

1 www.sequahead.eu

2 <http://journal.embnet.org/index.php/embnetjournal/article/view/218>

nation activities. Preliminary ideas were outlined for the training school, planned for the end of May 2012, in Uppsala, in conjunction with a COST Action workshop; Erik confirmed that he had already booked the training facilities, with access to computer resources at the UPPMAX computer centre, to allow hands-on classes with 'real' NGS data.

The second part of the MC meeting was dedicated to the Working Groups. Each WG Chair briefly presented the Group's principal tasks and how these might be achieved.

WG1, Technology Watch, Chaired by Ralf Herwig and Thomas Svensson: Ralf Herwig summarised the role of this group in providing timely alerts on new technology developments in topics such as sequence technology, analysis tools, applications and projects. The major activity of this WG will be scanning, reading and summarising scientific and technological articles. He proposed to organise a 'journal club', which will meet on a monthly basis to exchange the latest news on NGS technology developments.

WG2, Action Plan for NGS Bioinformatics, Chaired by Andreas Gisel and Ana Conesa:



COST scientific session: Robert Lyle presenting the Norwegian Sequencing Center

Andreas Gisel described the role of this group first, in reviewing the challenges and gaps in analysis pipelines in parallel with WG1, and then in formulating actions that would be tackled in collaboration with WG3 and WG4. The creation of sub-committees on specific topics was discussed in order to galvanise as many Action and non-Action members as possible to work on these topics.



Laurent Falquet presenting the SIB NGS infrastructure



Day 3: Joint work group 2 and work group 3 meeting

WG3, Software, Chaired by Eija Korpelainen and Steve Pettifer: Eija Korpelainen described the role of WG3 in gathering information on current data-analysis tools, including those under development, aiming to collaborate with WG4 to provide solutions in cases where these tools need to be customised to handle vast amounts of NGS data. A list of NGS tools developed by Action members has already been seeded on the WG4 page of the Action website.

WG4, Generic Informatics Topics, Chaired by Veli Makinen and Alberto Policriti: Veli Makinen described how this WG will focus on computer technology problems, such as data storage, interoperability, Grid and Cloud computing, and semantic applications.

WG5, Dissemination, Education and Training, Chaired by Gert Vriend and Jacques van Helden: Jacques van Helden explained that this WG will use several different media, including the portal and printed matter, to distribute information about NGS; it will also implement courses and teaching materials. An important role for this WG will be to propose standards for publishing NGS tools.

Before lunch, the COST Action BM0902, "Network of experts in the diagnosis of myeloproliferative disorders (MPD)" was presented by Sylvie Hermouet and Robert Kralowicz; their aim was to establish a close collaboration with SeqAhead,

as they will be involved in extensive data-analysis scenarios using NGS technology in future.

During the afternoon, there was an open session on common aims and planned activities. In particular, a joint training school with TD0801 (Statseq, "Statistical challenges on the €1000 Euro genome sequences in plants") and FA0806 (Plantivax, "Plant virus control employing RNA-based vaccines: a novel non-transgenic strategy") was discussed, as were the form and location of the next MC meeting, together with a summary of the remaining activities and proposals for year 2 of the Action.

The second day of this COST event was organised as a scientific session, in which selected Action members and non-members (according to their submitted abstracts) were invited to present their work on NGS data-analysis platforms, tools and applications. There were presentations on 4 different platforms, given by Robert Lyle from The Norwegian Sequencing Centre (NSC), Oslo, Norway; Kjell Petersen, representing NGS research and services at the Computational Biology Unit (CBU), Bergen, Norway; Laurent Falquet from the Swiss Institute for Bioinformatics, Lausanne, Switzerland, representing the Vital-IT HPC and Swiss-Prot groups; and Ning Li, from Beijing, China, presenting the Beijing Genomics Institute (BGI) sequencing and bioinformatics strategy.

A range of applications, providing broad coverage across the NGS data-analysis debate, were also presented: Frank Picard outlined bioinformatics developments for NGS data analysis at PRABI; Ana Conesa reviewed NOseq, an RNA-seq differential-expression method robust for sequencing-depth biases; Eric Rivals presented a combinatorial and integrated method to analyse RNA-seq reads; Jacques van Helden introduced RSAT peak-motifs, a pipeline for discovering motifs in massive ChIP-seq peak sequences; Luca Pireddu spoke about the Seal suite of distributed software for high-throughput sequencing; Keijo Heljanko presented scalable Cloud computing solutions for NGS data; Andreas Gisel discussed smallRNA data analysis; Eija Korpelainen presented Chipster 2.0, a user-friendly NGS data-analysis suite with built-in genome browser and workflow functionality; and Petr Baldrian reviewed the current possibilities and limitations in data analysis of environmental metagenomes and metatranscriptomes.

There was also a 'Miscellaneous' session, in which Jean Imbert presented HTS Science and gave a technology-watch tour; Matthias Steinbrecher spoke about innovation and trends with In-Memory technology; and José Ramón Valverde talked of NGS data-analysis from the user perspective. All presentations are available on the [SeqAhead](#)³ site.

The final day of the event involved a series of parallel meetings in which each WG met to discuss its activities for the forthcoming Action year.

WG1 agreed to meet frequently and exchange information from publications they planned to analyse, and to provide frequent updates on the WG1 page of the Action website.

WG2, WG3 and WG4 agreed both to a number of important actions that SeqAhead should initiate, and to establish a primary repository of NGS data-analysis tools. The latter will be made available via the WG3 page of the Action website, and on the [SeqAnswers](#)⁴ software hub. The agreed actions were to formulate, develop and publish on the website:

- a list of existing tools and platforms for NGS data analysis
- parallelisation (distributed computing) approaches for NGS data analysis
- protocols (descriptions on how to analyse NGS data), focusing initially around:
 - oncogenomics (e.g., how to align ChIP-seq data to a normal reference)
 - metagenomics

Protocols for other topics will be proposed and formulated in future, including, for example:

- variant annotation, especially for non-coding variants, association with phenotype
- genome annotation quality
- ncRNA analysis and annotation

A small group, mainly members of WG4, broke out from the WG2, WG3 and WG5 discussions to hold a first action meeting focusing on parallelisation approaches in NGS data analysis. The group agreed to organise a second action meeting on 'Hadoop technology', in February or March 2012, to develop solutions for the parallelisation of NGS data analysis.

WG5 discussed the practicalities and processes both for accepting applicants on Training Schools, and for awarding STSMs. The group discussed sets of criteria to facilitate these processes, and agreed to post further information, template application forms, etc., on the WG5 page of the Action website. An important role for this WG will be to propose standards for documenting NGS tools in order to make them usable by external users (user manual, demos, annotated study cases, utilization protocols).

The next SeqAhead, COST Action BM1006, events will be its inaugural training school and workshop, in Uppsala, Sweden, at the end of May and beginning of June 2012. Follow the Action on www.seqahead.eu and become active as an external expert in NGS data analysis.

³ <http://seqahead.cs.tu-dortmund.de/meetings:slides-2011-11>

⁴ <http://seqanswers.com/wiki/SEQanswers>