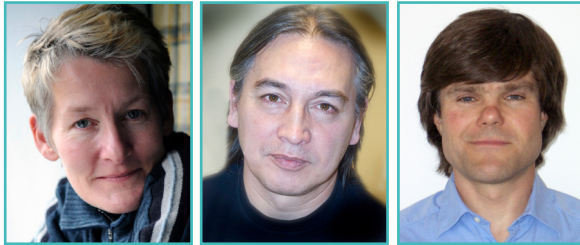# SEQAHEAD - COST Action BM1006: Next Generation Sequencing Data Analysis Network

Teresa Attwood[1], Erik Bongcam-Rudloff[2], Andreas Gisel[3]

[1]University of Manchester, United Kingdom
[2]Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences and Department of Immunology, Genetics and Pathology, Uppsala University, Sweden
[3]CNR, Institute for Biomedical Technologies, Bari, Italy

COST (European Cooperation in Science and Technology) is one of the longest-running European instruments supporting cooperation, collaboration and orchestration among scientists and researchers across Europe working in the same field. Some of the organisers of the two EMBRACE workshops on 'Next Generation Sequencing' (NGS) saw this type of Action as exactly the right kind of mechanism to try to tame the data tsunami being generated by the furiously fast developing NGS technologies. Their aim was to tackle the bioinformatics challenges inherent in managing and analysing these data, and to support researchers who use NGS technologies but do not have direct access to the necessary underpinning bioinformatics resources. The history of the NGS initiative is short, but explosive. It is imperative for the life science community to be prepared for the enormous growth in NGS data, the challenges this presents, and the opportunities it affords. Recognising these issues, and the need for global cooperation, gave birth to the idea for this COST Action proposal; it developed into the concerted action of today.

The initial proposers of the Action, Erik Bongcam-Rudloff and Andreas Gisel, Eija Korpelainen and Peter Rice were members of Work Package 4 (WP4) Test Cases in the FP6 Network of Excellence (NoE) EMBRACE. Many of the test cases collected by WP4 involved problems relating to the use of NGS technologies,
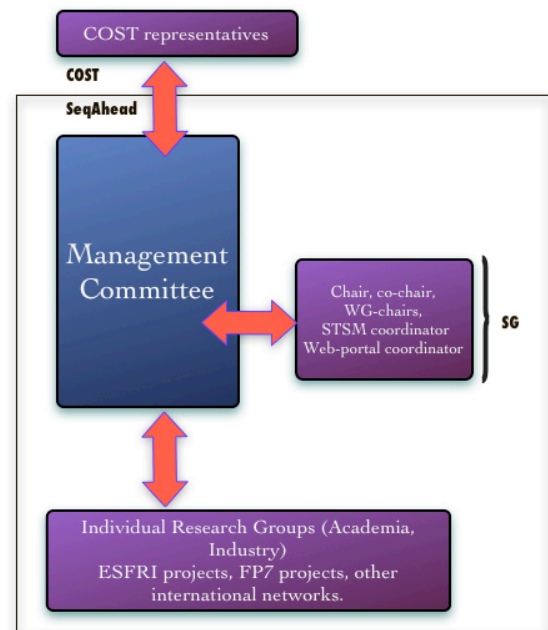

Figure 1. Graphic of the Action management structure.

demonstrating the growing need for NGS data analysis procedures of various types. As a first step to try to get to the bottom of the problems identified in WP4, the WP4 group was motivated to form a Task Force, including Erik Bongcam-Rudloff, Eija Korpelainen, Inge Jonassen, Nils-Einar Eriksson, Etienne deVilliers, Andreas Gisel, Laurent Falquet, JR Valverde and Gert Vriend. With the support of the Bioinformatics Italian Society (BITS), the Task Force organised an EMBRACE NGS data management and analysis workshop and hack-a-thon for bioinformaticians, which took place on November 18-20 (2009), in Rome at CASPUR, the Inter-University Consortium for the Application of Super-Computing for Universities and Research, hosted by Node manager Tiziana Castrignano. A comprehensive report on this workshop has been published on the EMBnet.news journal [1].

The acuteness of NGS data-analysis issues raised in the workshop led the Task Force to write a COST pre-proposal, *SEQAHEAD: Next Generation Sequencing Data Analysis Network*. The text proposed a strong network to monitor, on one side, the rapidly growing sequencing technologies and, on the other, the growing needs in data management and analysis. The idea was that a highly orchestrated action should guarantee efficient, agile procedures to cope with the NGS data flood. The pre-proposal

Figure 2. Group picture of the participants of the COST Action kick-off meeting in Brussels.

was written by six partners, in consultation with several of the workshop participants, who had indicated their interest in joining the Action if it were funded. In May 2010, the pre-proposal passed with almost the maximum score, and the way was therefore open to prepare a full proposal. In parallel with this development, and encouraged by the success of the first workshop, a second EMBRACE NGS workshop was organised on June 16 -17 (2010), in Ruvo di Puglia, Italy, affiliated with the EMBnet annual general meeting (for the programme, see [2]), with the full support of the Italian EMBnet Node manager Domenica D'Elia.

The latter meeting had multiple purposes. It was both a workshop to exchange ideas and information, and to work together, via a hack-a-thon, on real life NGS problems, and it was a platform to discuss the structure of, and ideas for, the full COST proposal.New partners among the participants were invited to sign up to the Action, and they drew our attention to colleagues not present at the meeting but who would likely also be interested to join (for the report, see [3]). In the end, forty-eight scientists with different knowledge within the NGS domain, from important institutions across sixteen European countries, created the full proposal. Again, this passed with almost the maximum score, and Erik Bongcam-Rudloff, the coordinator of the Action, successfully defended it in front of the Domain Committee. Early in December 2010, the proposed COST Action was officially approved [4].

With eighteen countries signed up to, or intending to sign, the Memorandum of Understanding (MoU), the COST Action BM1006 *Next Generation Sequencing Data Analysis Network* was initiated

with a kick-off meeting on March 13 (2011), in Brussels. The meeting was organised by the COST office, and chaired by Magdalena Radwanska and Anja Van Der Snickt, who introduced the participants to the 'world' of COST Actions, mainly running us through the labyrinthine rules and regulations. They explained that the designated rapporteur for the Action will be Dr. Tanja Gmeiner Stopar (University Medical Centre Ljubljana, Slovenia), who will be responsible for monitoring our activities as the project moves forward. The primary goals of the kick-off meeting were to establish the budget for the first year, and to elect key personnel from within the consortium to manage the Action over the next four years. Erik Bongcam-Rudloff (SE) was consequently elected as the Action Chair and Teresa Attwood (UK) as Vice-Chair. A Chair and Vice-Chair were also allocated to each Working Group, as follows: WG1, *Technology watch for new developments*: Ralf Herwig (DE) and Thomas Svensson (SE); WG2, *Development of action plan for NGS bioinformatics to cope with challenges for the EU research area*: Andreas Gisel (IT) and Ana Conesa (ES); WG3, *Design, implementation and incorporation of software solutions*: Eija Korpelainen (FI) and Steve Pettifer (UK); WG4, *Generic informatics topics*: Veli Makinen (FI) and Alberto Policriti (IT); WG5, *Development of strategic dissemination and education program for NGS bioinformatics*: Gert Vriend (NL) and Jacques van Helden (BE). A number of other people were also elected to key roles within the Action: hence, Eric Rivals (FR) and JR Valverde (ES) will be responsible for coordinating the Short Term Scientific Missions (STSMs), which are mainly to support exchange activities of young researchers between European research institutes; and Sven Rahman (DE) will be responsible for the Action's Web portal [5], which will be an important medium for disseminating the Action's many activities in 'real-time'.

The budget for the first year was divided into four parts, mainly covering the first event, during October 2011, for all Management Committee members to meet, and to organise and start the Action's main activities. The rest of the budget will cover education, student exchange and publication and dissemination. Ultimately, the overall budget will be flexible, depending on the number of additional partners who sign up to the MoU as the Action progresses.

For now, we are extremely grateful to everyone who helped to get SEQAHEAD funded, and to those who have taken on important management roles within the Action; together, we eagerly look forward to vigorous and fruitful collaborations during the coming years!

References

1. http://journal.embnet.org/index.php/embnet-news/article/view/60/207
2. www.nextgenerationsequencing.org
3. http://journal.embnet.org/index.php/embnet-journal/article/view/176/382
4. http://w3.cost.eu/index.php?id=212&action_number=BM1006
5. http://www.seqahead.eu