The Biogems Community: Challenges in distributed software development in bioinformatics

George Githinji[1,10], Ben Woodcroft[2,10], Joachim Baran[3,10], Francesco Strozzi[4,10], Raoul Bonnal[5,10], Naohisa Goto[6,10], Toshiaki Katayama[7,10], Hiroyuki Mishima[8,10], and Pjotr Prins[9,10]

[1]*KEMRI-Wellcome Trust Research Programme, Kenya,* [2]*The Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, The University of Queensland, Australia,* [3]*National Evolutionary Synthesis Center, Durham, United States of America,* [4]*Bioinformatics Core Facility, Parco Tecnologico Padano, Italy,* [5]*Integrative Biology Program, Istituto Nazionale Genetica Molecolare, Milan 20122, Italy,* [6]*Department of Genome Informatics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka, 565-0871, Japan,* [7]*Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo 108-0071, Japan,* [8]*Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, 852-852, Japan,* [9]*University Medical Centre, Utrecht, Netherlands,* [10]*The Biogems Community*

Abstract

The BioRuby project is a comprehensive bioinformatics library for the Ruby programming language. In its more than fifteen years of existence it has evolved from a monolithic to a distributed code-base.

With over 40 contributors and 134 projects and a plan to enlist other Bio* projects, it is important to assess potential challenges of the distributed contributions model such as testing, documentation, coding standards and communication.

In this talk we present metrics on test coverage, documentation and communication within the biogems projects. We also highlight successful distributed software development models and patterns that encourage new contributions, collaboration and build trust among developers.