

BioJava Project Update

Holland, R.C.G.¹, Down, T.², Pocock, M.³, Prlic, A.⁴, Huen D.⁵, James K.⁴, Foisy, S.⁶, Dräger, A.⁷, Yates, A.⁸, Heuer M.⁹, and Schreiber, M.J.¹⁰

¹ Eagle Genomics Ltd., Cambridge, UK ² Wellcome Trust/Cancer Research UK Gurdon Institute, Cambridge CB2 1QN, UK ³ University Newcastle Upon Tyne, Newcastle Upon Tyne, NE1 7RU, UK ⁴ Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK ⁵ Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK ⁶ Laboratory in Genetics and Genomic Medicine of Inflammation, Montreal Heart Institute, Montreal, Québec, Canada. ⁷ Eberhard Karls University Tübingen, Center for Bioinformatics (ZBIT), Germany ⁸ European Bioinformatics Institute (EMBL-EBI), Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK ⁹ Harbinger Partners, Inc., St. Paul, Minnesota, USA ¹⁰ Novartis Institute for Tropical Diseases, 10 Biopolis Road, Chromos #05-01, Singapore 138670

BioJava was conceived in 1999 by Thomas Down and Matthew Pocock as an API to simplify bioinformatics software development using Java (Pocock et al., 2000). It has since then evolved to become a fully-featured framework with modules for performing many common bioinformatics tasks.

As a free and open-source project, BioJava is developed by volunteers coordinated by the Open Bioinformatics Foundation (O|B|F, <http://open-bio.org/>) and is one of several Bio* toolkits (Mangalam, 2002). Over the past eight years, the BioJava has brought together nearly fifty different code contributors, hundreds of mailing list subscribers, and several wiki contributors. All code and related documentation is distributed under version 2.1 of the GNU Lesser General Public License (LGPL) license (Free Software Foundation, Inc., 1999). All wiki documentation is made available online under version 1.2 of the GNU Free Documentation License (Free Software Foundation, Inc., 2000).

BioJava has been used in a number of real-world applications, including Bioclipse (Spjuth et al., 2007), BioWeka (Gewehr et al., 2007), Cytoscape (Shannon et al., 2003), and Taverna (Oinn et al., 2004), and has been referenced in over fifty published studies. A list of these can be found on the BioJava website.

The latest BioJava release (version 1.6, released on 13 Apr 2008) offers more functionality and stability over the previous official releases. The phylogenomics package was improved and expanded by our 2007 Google Summer of Code (GSOC'07) student Boh-Yun Lee. It now contains fully-functional Nexus and Phylip parsers, and tools for calculating UPGMA and Neighbour Joining, Jukes-Kantor and Kimura Two Parameter, and MP. The PDB file parser was improved by Jules Jacobsen for better dealing with PDB header records. Andreas Dräger provided several patches for improving the genetic algorithm packages. The version 1.6 release also contains numerous bug fixes and documentation improvements.

The BioJava website is <http://biojava.org/>. The version 1.6 release can be downloaded from <http://biojava.org/wiki/BioJava:Download>.

References

- Free Software Foundation, Inc. (1999) GNU Lesser General Public License, version 2.1, <http://www.gnu.org/licenses/old-licenses/lgpl-2.1.html>, accessed 10 May 2008.
- Free Software Foundation, Inc. (2000) GNU Free Documentation License, version 1.2, <http://www.gnu.org/licenses/fdl-1.2.html>, accessed 10 May 2008.
- Gewehr JE, Szugat M, Zimmer R. (2007) BioWeka—extending the Weka framework for bioinformatics *Bioinformatics* 2007 23(5):651-653.
- Mangalam H. (2002) The Bio* toolkits – a brief overview. *Brief Bioinform.*, 3, 396-302.
- Oinn T, Addis M, Ferris J, Marvin D, Greenwood M, Carver T, Pocock MR, Wipat A, Li P. (2004) Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics*, 20, 3045–3054.
- Pocock M, Down T, Hubbard T. (2000) BioJava: Open Source Components for Bioinformatics. *ACM SIGBIO Newsletter* 20(2), 10-12.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 2003 Nov; 13(11):2498-504.
- Spjuth O, Helmus T, Willighagen EL, Kuhn S, Eklund M, Wagener J, Murray-Rust P, Steinbeck C, Wikberg JE. (2007) Bioclipse: an open source workbench for chemo- and bioinformatics. *BMC Bioinformatics*. 2007 Feb 22;8:59.