

# BioJava

<http://www.biojava.org/>

BOSC 2007, Vienna

Richard Holland  
<[holland@ebi.ac.uk](mailto:holland@ebi.ac.uk)>

# What is it?

- A Java library (LGPL) for use in bioinformatics.
- Object model.
  - Sequences, Alignments, Phylogenetic trees, etc.
- Object manipulation.
  - Alignment slices, Tree navigation/reordering, etc.
- Object I/O.
  - Read/write common formats (Fasta, Nexus, Blast, etc.)
  - Read/write to BioSQL databases.

# What is it not?

- A workbench.
- A GUI toolkit.
  - Although it does have some Swing components for displaying BioJava objects in Swing applications.
- A novel algorithm.
  - What algorithms it does contain are reimplementations of existing ones.
  - It enables users to simplify common bioinformatics programming tasks.

# BioJava 1.5 vs. BioJavaX

- 8 years since first release in 1999.
- 1.5 released in June 2007 after a 3 year wait.
- 1.5 will be last release to support Java 1.4.
  - 1.6 will use features from Java 6 (not 5!).
- 1.5 incorporates the BioJavaX extensions.
  - (described at BOSC 2005, Detroit, as part of DengueInfo project)
- BioJava2 is an entirely unrelated project.

# What is new in 1.5?

- The BioJavaX extensions.
  - BioJavaX is like the javax packages.
  - Extends and improves, but doesn't replace or modify.
  - Improved sequence, feature, and location objects.
  - Genetic Algorithms.
  - More detailed file parsers.
  - Much improved BioSQL interaction.
- Many bugfixes to existing code.

# What's happening right now?

- December '06 – NESCent Phyloinformatics Hackathon.
  - Developed Nexus parser and strategy for phylogenetics object models and toolkits.
  - Led to ongoing NESCent/Google Summer of Code project with our own dedicated student, Boh-Yun Lee.
- User testing and unit test development.
- Working on our future plans.
  - February '07 – strategy meeting held at EBI.

# What's next?

- Further phylogenetics development.
- Variation data. (dbSNP)
- Microarrays.
- Possible shift away from sequence-centric to feature-centric object model.
- Simplification of APIs.
  - Provide helper methods to mask complex tasks.
  - Tighten up overly flexible code.

# What's next?

- Groovy. (scripting interface).
- XPath and XQuery for parsing XML data.
- J2EE integration.
- Maven (vs. Ant) and Subversion (vs. CVS).
- EnsJ (the Java API for Ensembl).
  - lives in our repository since Ensembl retired it.
  - nobody working on it at present.
- Publication of a BioJava paper.

# How can you help us?

- Tell us what you are using it for!
- Email us if you get stuck, or use bugzilla.
  - <http://bugzilla.open-bio.org/>
  - [<biojava-1@lists.open-bio.org>](mailto:biojava-1@lists.open-bio.org)
- We welcome all code contributions and ideas.
- Offers of documentation and JUnit tests will be much appreciated.
- Tell us what features you think we should add.

# Thanks

- Matthew Pocock, Thomas Down, Mark Schreiber, Michael Heuer, Keith James and David Huen for getting it where it is today.
- Hilmar Lapp for organising the NESCent Hackathon and Google Summer of Code projects.
- OBF for providing us with server infrastructure.
- Ensembl for funding my trip to BOSC this year.
- BioMart for allowing me to spend a bit of time supporting BioJava.

# Thanks

- All contributors past and present:
  - Alan Li, Andreas Drager, Andreas Prlic, Andy Yates, Aroul Ramadass, Benjamin Schusteri-Backler, Boh-Yun Lee, Bradford Powell, Brian Gilman, Brian King, Bubba Puryear, Cambridge Antibody Technology Group plc, Colin Hardman, David Allen, David H. Klatt, David Huen, David Scott, David Waring, Eric Haugen, Ewan Birney, Francois Pepin, Frans Verhoef, George Waldon, Gerald Loeffler, Gero Greiner, Greg Cox, Hanning Ni, Horvath Tamas, Jason Stajich, Jim Balhoff, Jolyon Holdstock, Jonathan Warren, Kalle Naslund, Keith James, Kim Rutherford, Lachlan Coin, Laurent Jourden, Leen Ammeraal, Lei Lai, Len Trigg, Lorna Morris, Lukas Kall, Mark Schreiber, Mark Southern, Markus Brosch, Martin Senger, Martin Szugat, Matthew Pocock, Mathieu Wierpert, Mayo Foundation, Michael Heuer, Michael Jones, National Institute of Standards and Technology, Nimesh Singh, Paul Seed, Peter Lackner, Rhett Sutphin, Richard Holland, Ronald Tschalar, Ron Kuhn, Russell Smithies, Samiul Hasan, Scott McCrory, Shuvankar Mukherjee, Simon Brocklehurst, Simon Foote, Stuart Johnston, Tamas Horvath, Tanya Vavouri, Thad Welch, The MathWorks Inc., Thomas Down, Tim Dilks, Tobias Thierer, Todd Riley, Unidata Program Center.