LibrAlign — A Java library with powerful GUI components for multiple sequence alignment and attached raw and meta data

Why developing LibrAlign?

Several applications currently developed in our group and by our cooperators deal with multiple sequence alignments (MSA) or associated raw and meta data (e.g., trace files, positions of repeats or inversions, ...), and allow the user to view and edit them in a graphical user interface (GUI). Instead of implementing independent solutions for these different tasks, we decided to create a common library that contains powerful and reusable GUI components and is easily extensible by others.

Features and design

- AlignementArea is the main GUI component of LibrAlign
  - It contains the aligned sequences and allows to edit these (q)
  - as well as a set of DataAreas displaying attached data (g).
- Sequence data can come from any data source by an implementation of the Interface SequenceDataProvider (g).
- LibrAlign contains several default implementations that provide compatibility to e.g. Biojava or allow a compressed storage of very large data sets.
- The GUI components allow the user to easily edit the alignment (g).
- The parent class DataTree can also be extended by application developers to create new custom data areas (see q) and example in fig. 5.

Software based on LibrAlign

Currently there are a number of software projects available or under development which make use of the LibrAlign API in their current or future versions.

- Taxonomic editor of the EDIT platform: The EDIT platform is currently extended to support sequence data for the Campanula portal of EDIT using LibrAlign (see fig. 2). EDITor available at http://cybertaxonomy.eu/taxeditor/.
- AlignmentComparator: Visualizes differences between alternative alignments of the same data set and allows annotating the comparison (see fig. 3), available at http://bioinfweb.info/AlignmentComparator/.
- PhyDE: Phylogenetic data editor: An alignment editor that provides features especially useful for phylogenetic purposes (see fig. 4, available at http://phyde.de/). New versions will also be based on LibrAlign.
- HIR-Finder: Locates microstructural mutations like tandem repeats possibly associated with hairpins and visualizes these in a special data area (fig. 5).

Compatibility and IO

- All components have a Swing and a SWT version, so they can be integrated into nearly all thinkable kinds of Java GUIs, including more complex projects based on the Eclipse Rich Client Platform http://www.eclipse.org/home/categories/rcp.php or Bioclipse (Spjuth et al., 2009).
- The toolkit independence is made possible by the JTC library of biowin.commons (http://commons.bioinfweb.info/java/).
- LibrAlign is fully interoperable with the Biojava 3 API (Pitlick et al., 2012) and also contains bridges to Biojava 1.9 legacy release.
- NeXML (Voß et al., 2012) is the main format of LibrAlign, since its flexible meta data annotation allows to store data associated with the data areas. (The definition of according ontologies for the different data areas is currently ongoing.)
- Several additional alignment formats are supported for im- and export.

Acknowledgements

The funding of parts of the development of LibrAlign with grant MI 1879/1-3 to KPM by the DFG (German research foundation) is highly appreciated. BCS wants to thank the DECHHEMA (Society for Chemical Engineering and Biotechnology) for financing the presentation of this poster at the German Conference on Bioinformatics 2014.

Furthermore the authors are very thankful to the developers of the other open source projects Libalign is build on (Biojava, Apache commons, NeXML), especially to Rutger Vos for his advice on the NeXML integration. A special thank you is also due to Joachim Röschenbleck for providing demonstration data and to Ruth Mendenhall for very helpful comments on this poster.

