

Bioinformatics simplified with seqpad

De Beule, D.¹, Decouttere, F.¹, Trooskens, G.², Devisscher, M.¹ & Van Criekinge, W.²

¹ Genohm
Technologiepark 3 bus 9
B9052 Zwijnaarde, Belgium

www.genohm.com

{david|frederik|martijn}@genohm.com

² Laboratory for Bioinformatics and
Computational Genomics
Dept. Molecular Biotechnology
Fac. Bioscience Engineering, Ghent University
Coupure Links 653
B9000 Gent, Belgium
{geert.trooskens|wim.vancriekinge}@ugent.be

Url: www.seqpad.org

Source code: trac.seqpad.org | svn.seqpad.org

License: LGPL

Abstract

We present seqpad, an open source sequence visualisation and annotation suite. Seqpad has been available for several years as a commercial package from Genohm.com. Today, in line with our business shift from a product based to a service oriented model, we release our flagship product to the open source community.

Seqpad is a Swing application built in Java 1.6 and interfacing with BioJava 1.6 and BioSQL. It runs on Windows, Mac and Linux. Seqpad is available from www.seqpad.org, or can be checked out from our svn repository.

While several good visualisation and annotation packages exist in the open source community, we believe that Seqpad represents important contributions in usability, functionality, visualisation and automation.

First off, Seqpad has been developed with ease of use as one of the main design concepts, and targets as a user the biologist rather than the bioinformatician, while still leveraging the versatility and power of the Biojava platform. Data in Seqpad is organized in projects. Projects can contain a wide variety of typical bioinformatics files, such as sequences, multiple alignments, phylogenetic trees and protein structure models, but also spreadsheets and word processor documents. We feel this is a very intuitive way for scientists to organize data. The project view has very intuitive controls, e.g. copying a fasta sequence from a browser window and dropping it on the project view will automatically result in a file being created in the project view holding the sequence. The fasta format is automatically recognized, so the sequence can immediately be visualised in a visualisation dock.

The visualisation of sequences, alignments, trees and dotplots is based on the Piccolo 2D Graphics framework (<http://www.piccolo2d.org>). Piccolo is a framework that introduces the concept of Zoomable User Interfaces. A ZUI is a new kind of interface that enables showing a huge canvas of information on a traditional computer display by letting the user smoothly zoom in, to get more detailed information, and zoom out for an overview.

Seqpad features automated switching between protein/dna views, and also has a custom built OpenGL based 3D renderer for protein structure models. All visualisations are tightly coupled: for example, selecting a dna sequence in the zoomable sequence interface also highlights the relevant part in protein 3D structure, a very useful feature when interpreting biological relevance of conserved regions for example. The 3D viewer also facilitates quick docking experiments.

With the 'time line' feature it is possible to slide a virtual time line over a phylogenetic tree and see the matches in the coupled multiple alignment change over time.

Seqpad also features an automatic scheduler: Blast queries launched from seqpad for example automatically create a poll job on an internal stack, so the user is not required to poll for results, and results automatically end up in the project view.

The presentation will include a use case demonstrating features of the product. Our main developers will be present, so people wanting to get involved can get started immediately !