Universität Bielefeld

The Bielefeld University Bioinformatics Server

Alexander Sczyrba, Jan Krüger, Henning Mersch and Robert Giegerich

Abstract

The Bielefeld University Bioinformatics Server (BiBiServ) supports Internet-based collaborative research and education in bioinformatics. Currently, 29 software tools and various educational media are available. These include tools from different areas such as Genome Comparison, Alignments, Primer Design, RNA Structures, and Evolutionary Relationships. The BiBiServ resources are accessible at

http://bibiserv.techfak.uni-bielefeld.de/

BiBiServ Tools

Here, we focus on the seven most prominent or recent tools on BiBiServ.

REPuter
Efficient detection of various types of repeats

REPuter [1] uses efficient algorithms and data structures to locate exact repeats in linear space and time. These exact repeats are used as seeds from which significant degenerate repeats are constructed allowing for mismatches, insertions and deletions. REPuter provides an evaluation of significance and interactive visualization (see figure 1). REPuter is available online on BiBiServ and as standalone application for most Unix based operating systems.
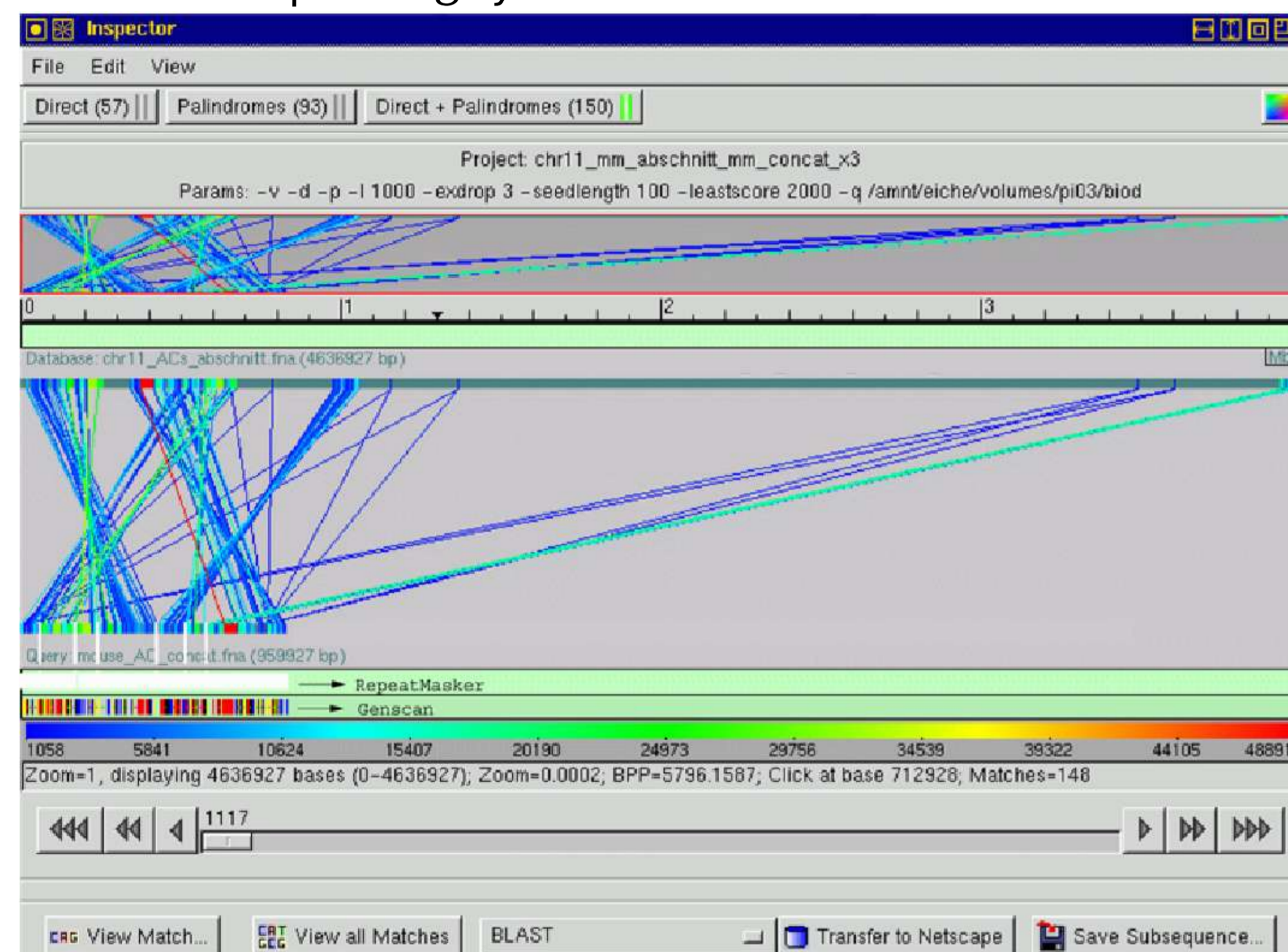
Figure1: REPuter: Comparison of mouse chromosome 11 assembled sequence (NCBI - Sep 02) (upper sequence) with small unordered draft contigs (bottom sequence). Matches between the sequences are indicated by connecting lines. The color code of the matches corresponds to the length of the match, red being the longest match of 48891 nt, short matches (about 1000 nt) are colored blue.

E2G
#### Efficiently mapping large EST sets to genomic sequence

E2G [3] is a web based-tool which efficiently maps large EST and cDNA data sets to genomic DNA. The server hosts huge EST databases of a few GB in size in indexed data structures. This allows users to rapidly detect new genes, verify the exon-intron structure of predicted genes and determine splice variants in the uploaded genomic region of interest. The Web interface provides a graphical view of the results (see figure 2). Alignments can be visualized and matching sequences downloaded for further processing. E2G is also available as WebService.

RNAforester
RNA Structure Alignment

While there are several programs available that compare two RNA secondary structures and compute a distance or a similarity score, none of them produces a structure alignment. This is the purpose of RNA Forester. RNA structures are represented as trees or more generally, sequences of trees, called forests. The alignment model underlying RNA Forester is the mathematically faithful generalization of sequence alignments to trees.
1. global and local similarity alignments of RNA structures
2. finds the most similar substructure within two arbitrary structures
3. can discover conserved structural motifs in the absence of sequence conservation
The computed alignments are drawn in a color code, highlighting common features, deviations, and compensatory base changes (see figure 4).
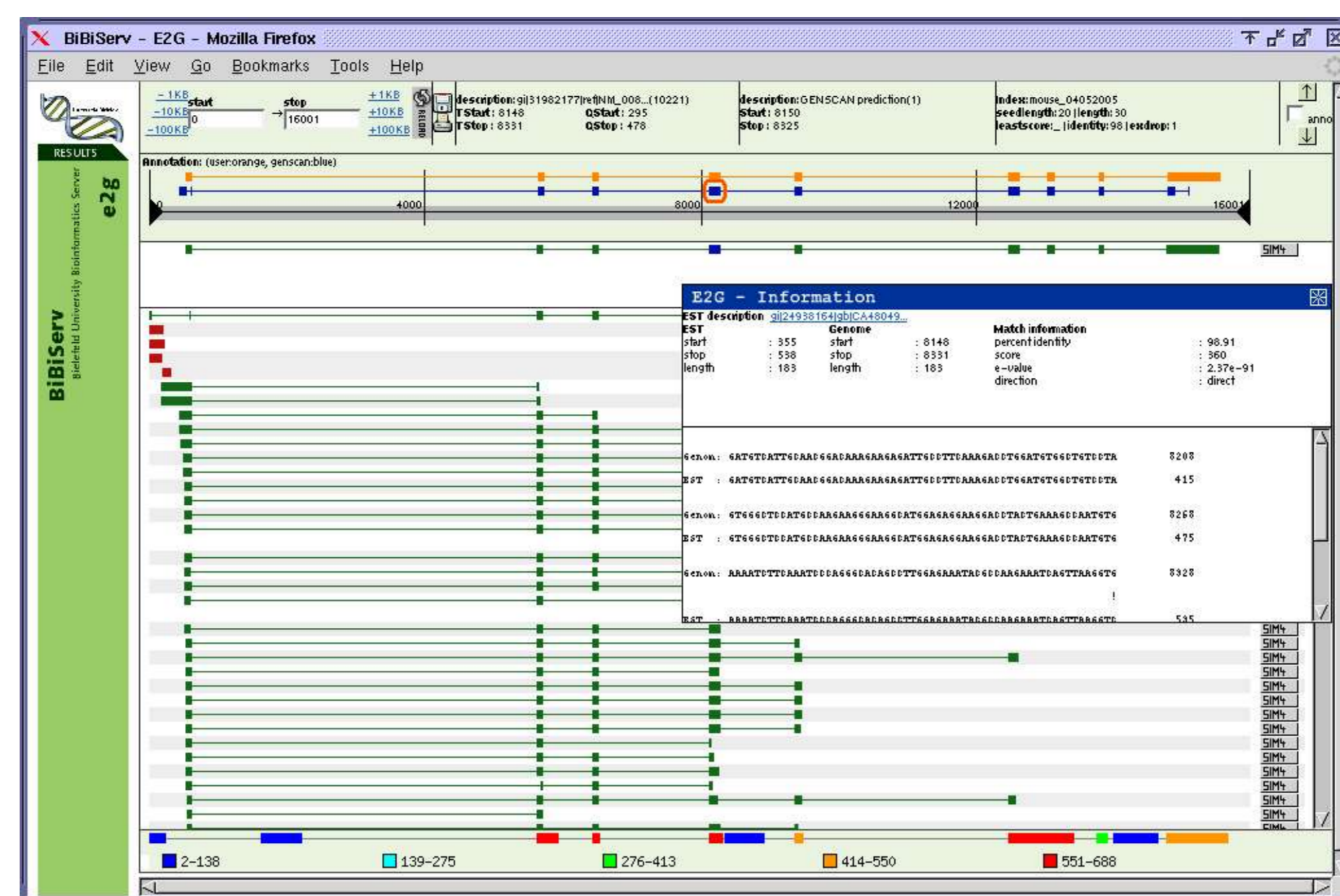
Figure 2: Graphical overview produced by e2g by mapping 16.5Kbp genomic sequence from M. musculus (Genbank GI: 28515921) to 4.1 Mill. mouse ESTs.

PknotsRG
RNA folding, including pseudoknot

PknotsRG [5] is a tool for folding RNA secondary structures, including the class of simple recursive pseudoknots. The program runs in O(n^4) time and O(n^2) space, therefore its application here on the BiBiserv is limited to sequences of length up to 600 bases. The energy parameters for structures containing no pseudoknots are the same as in the actual mfold 3.1. The energy for pseudoknots is computed with a model similar to that used by Rivas&Eddy in pknotsRE.

RNAhybrid
Finding the minimum free energy hybridization

RNAhybrid [6] is a tool for finding minimum free energy (mfe) hybridization of a long (target) and a short (query) RNA. The hybridization is performed in a kind of domain mode, ie. the short sequence is hybridized to the best fitting parts of the long one. The tool is primarily meant as a means for microRNA target prediction. In addition to mfes, the program calculates p-values based on extreme value distributions of length normalized energies. RNAhybrid is available online as classical webinterface, as WebService and as download for various platforms.
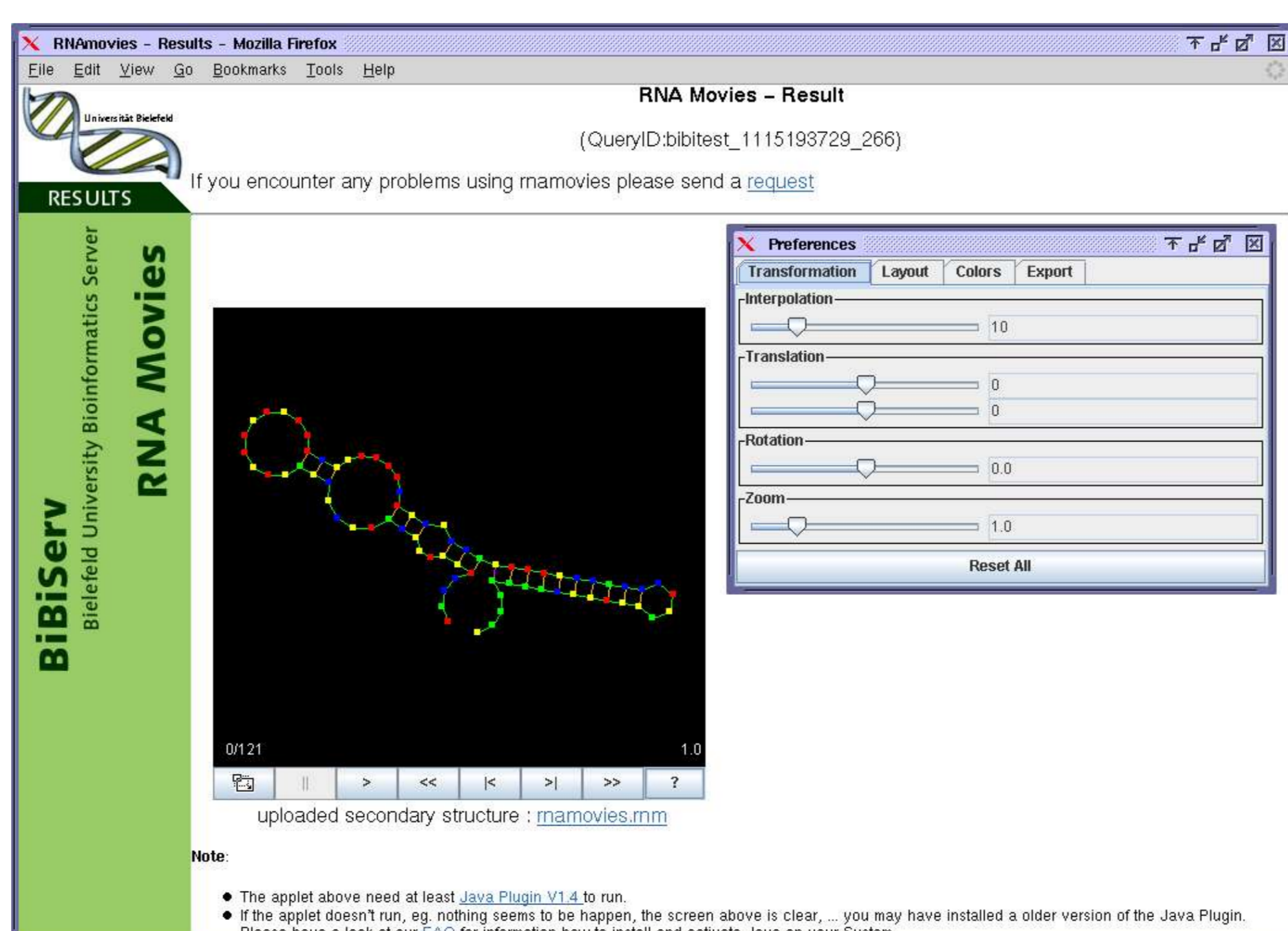
RNAmovies
Animated visualization of RNA secondary structures

RNAmovies [7] is a system for the visualization of RNA secondary structure spaces. Its input is a script consisting of primary and secondary structure information. From this script, the system generates animated graphical structure representations. In this way, it creates the impression of an RNA molecule exploring its own 2D structure space. In this new version, RNAmovies has been extended to support the DCSE input format, display pseudoknots and "entangled helices" and provide resolution independed publication quality output via PostScript. Furthermore a Java "Online" version has been developed to interactively display RNAmovies on the web.

RNAshapes
Folding of "relevant" RNA secondary structures

RNAshapes [8] is a tool for folding of "relevant" RNA secondary structures. In the analysis of RNA and especially when regarding the secondary structure it is often required to take also suboptimal structures into account. As the number of suboptimal structures is very large and even the number of near-optimal structures grows rapidly with sequence length, these kind of analyses are quite expensive. One property of

Figure 3: RNAmovies (Java version) running as applet in a browser, shown one structure of LeptomonasCollosoma 5'End

suboptimal structures is that they often only differ in a few base pairs, while maintaining the same overall shape. These are less interesting because in most cases one is interested in those structures that are rather dissimilar, meaning those having a different shape. RNAshapes is an approach to the direct folding of only those structures having a different shape.
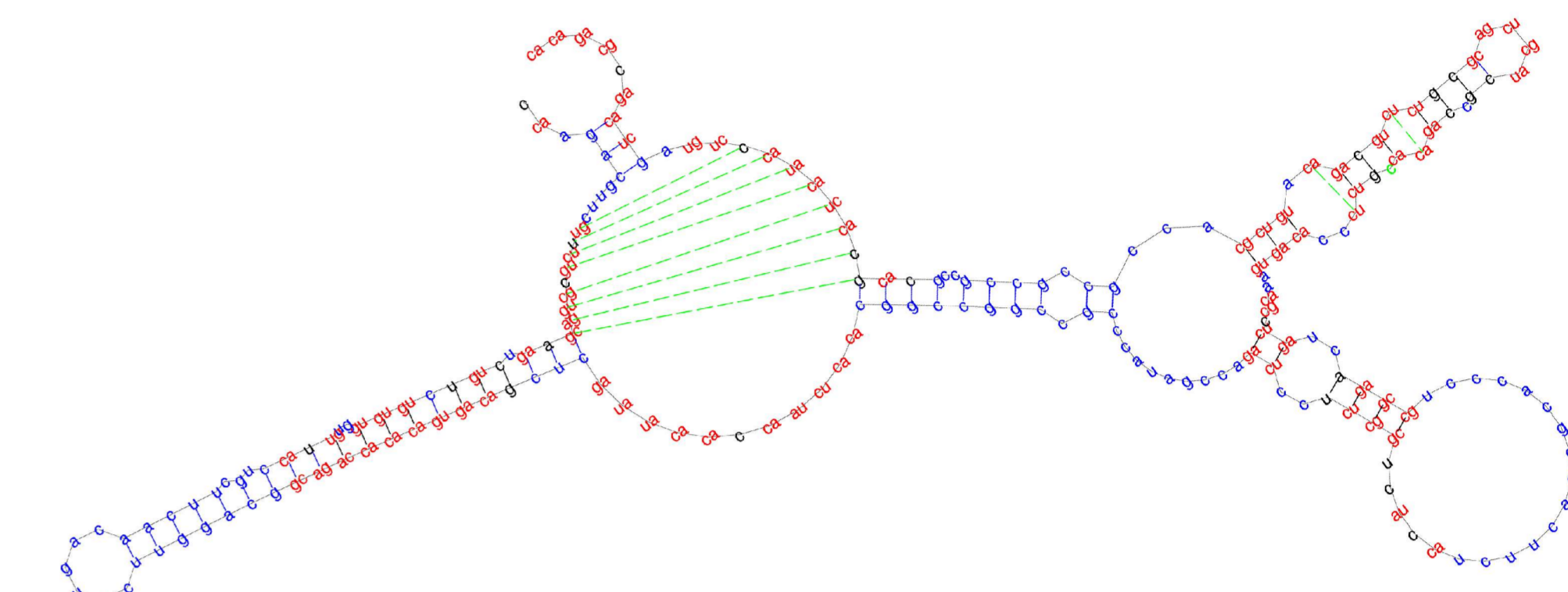
Figure 4: RNAforester: Structure alignment of the human ferritin 5'UTR and the Drosophila melanogaster succinate dehydrogenase 5'UTR. Bases printed in black show structure elements that occur in both structures with the same sequence. Sequence variations are displayed by using red letters. Bases or base pairs that can only be found in ferritin are printed in blue, while bases that only occur in succinate dehydrogenase are printed in green.

WebServices

A new technology for invoking programs on foreign computers are WebServices. Providing access using WebServices makes it possible to use the programs from local computers and compute results remotely without technical knowledge of the program.
A WebService is described in a so called WSDL document, it includes all required information, like the location of the WebService and the interface. BiBiServ is part of the HOBIT project [9], which aims at providing and interconnecting bioinformatics tools as WebServices to offer easy and automated usage of these tools. WebService technology is also used internally; the web server offers user interfaces which can be accessed via a webbrowser. Internally the jobs are submitted to the WebServices server which processes the requests on a computing grid. This way, we can provide a high amount of distributed compute resources, which can be used through both interactive web and WebService interfaces. Currently BiBiServ offers four tools on basis of WebServices: E2G[2], REPuter[1], SplitsTree[10] and RNAhybrid[6]. We are currently working on integrating more tools, which will be finished shortly. An overview is also available [11].

Educational Media

The BiBiServ Media & Distance Education section supports teaching in bioinformatics with internet-based multimedia courses. Currently there are five online courses and tutorials available on BiBiServ. The recent ones are :
• The ADP pages: Interactive pages that allow to study and experiment with classical dynamic programming algorithms
• Sequence Analysis with Distributed Resources: A Web-based practical course on sequence analysis using resources from all over the world (see figure 5).
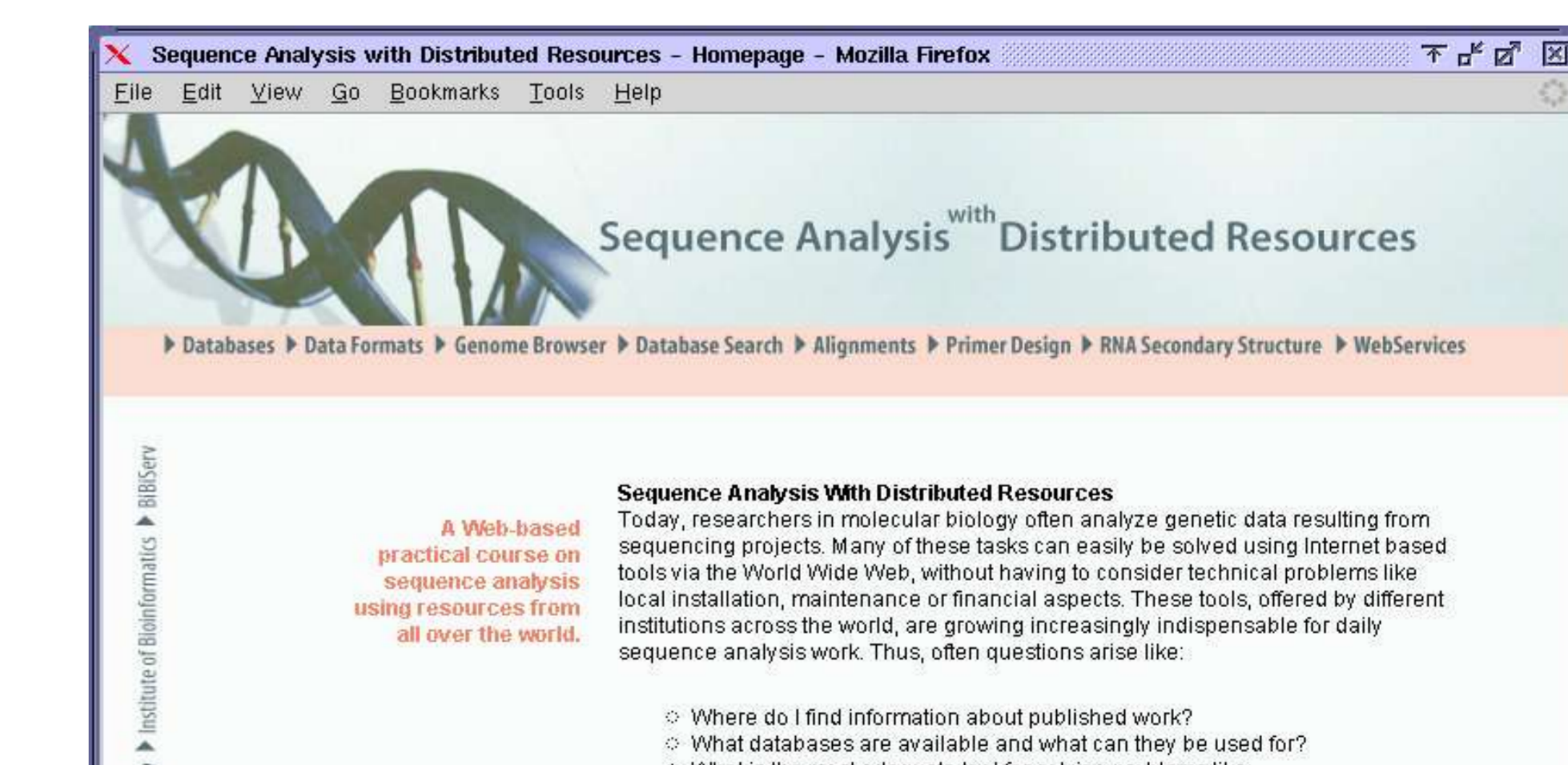
Figure 5: SADR (Sequence Analysis with Distributed Resources) online course

References

[1] S. Kurtz, J. V. Choudhuri, E. Ohlebusch, C. Schleiermacher, J. Stoye, R. Giegerich: REPuter - The Manifold Applications of Repeat Analysis on a Genomic Scale. Nucleic Acids Research, 29 (22):4633-4642, 2001.
[2] B. Morgenstern: DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 15, 211 – 218, 1999.
[3] J. Krüger, A. Sczyrba, S. Kurtz, R. Giegerich: e2g - An Interactive Web-Based Server for Efficiently Mapping large EST and cDNA sets to Genomic Sequences. Nucleic Acids Research, 2004, Vol.32, Web Server issue W301-304
[4] R. Giegerich, D. Haase and M. Rehmsmeier: Prediction and visualization of structural switches in RNA, Pac. Symp. Biocomput., Pages: 126-137, 1999
[5] J.Reeder, R.Giegerich: Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics, BMC Bioinformatics, 5:104, 2004.
[6] M. Rehmsmeier, P. Steffen, M. Höchsmann, R. Giegerich : Fast and effective prediction of microRNA/target duplexes. RNA, 10:1507-1517, 2004.
[7] R. Giegerich, D. Evers: RNA Movies - visualizing RNA secondary structure spaces. Bioinformatics,Volume 15, Issue 1, January 1999, pp. 32-37
[8] R.Giegerich, B.Voss, M.Rehmsmeier: Abstract Shapes of RNA, Nucleic Acids Research, 32 (16), Pages:4843-4851, 2004
[9] HOBIT (Helmholtz Open Bioinformatic Technology): http://hobit.sourceforge.net
[10] D.H. Huson: SplitsTree- Analyzing and visualizing evolutionary data,Bioinformatics, 14(1): 68-73 (1998)
[11] http://bibiserv.techfak.uni-bielefeld.de/wsdl