Bielefeld University Bioinformatics Server

Webservice Pipelines in Bioinformatics

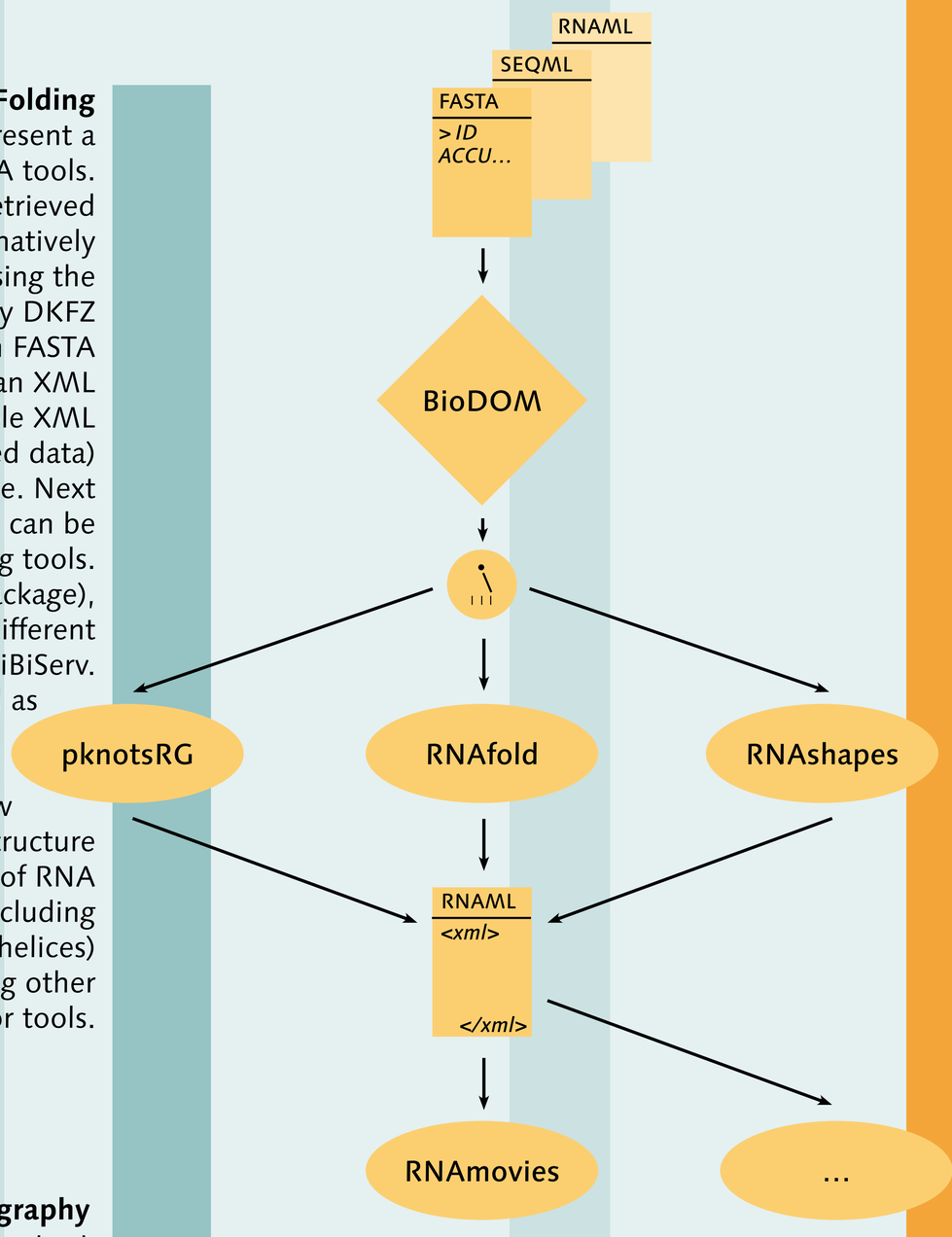
Jan Krüger and Henning Mersch

Introduction

We present an automatism for easily connecting bioinformatical projects based on webservice technology, which are mainly offered by the Bielefeld University Bioinformatics Server (BiBiServ - http://bibiserv.techfak.uni-bielefeld.de) and other collaborators of the Helmholtz Open Bioinformatics Technology project (HOBIT - http://hobit.sourceforge.net).

RNA Folding

As first example, we present a pipeline of different RNA tools. Some sequences will be retrieved from local files or alternatively from the EMBL database using the SoapDB webservice offered by DKFZ Heidelberg. The sequences (in FASTA format) will be converted to an XML format (RNAML is a suitable XML format for most RNA related data) using the BioDOM webservice. Next the converted sequences can be folded using different folding tools. RNAfold (Vienna RNA Package), pknotsRG or RNAshapes are different folding tools offered by BiBiServ. The result - represented as RNAML - can be visualized using RNAmovies. This webservice can be used to view a single RNA secondary structure or sequential animation of RNA secondary structures (including pseudoknots and entangled helices) or further processed using other webservices or tools.



Webservice Choreography

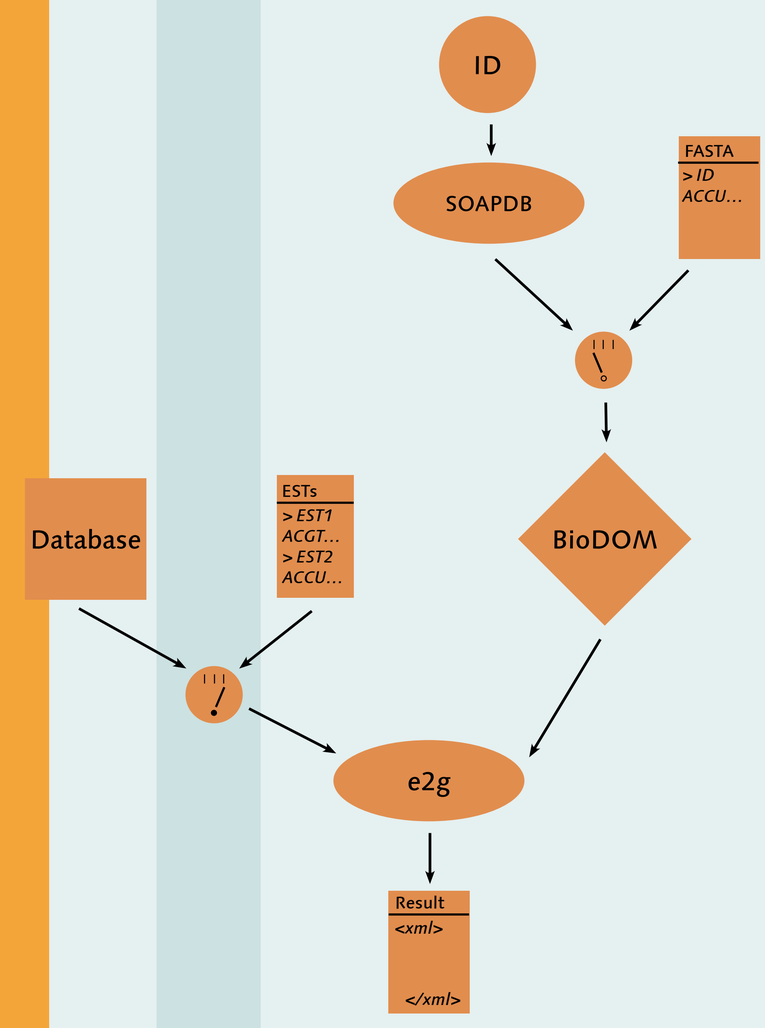
Using standardized input and output formats makes it easy to combine webservices to a pipeline, thus one webservices choreography (composed of different tasks) could easily be used multiple times without any user interaction. Using standardized input and output formats makes it easy to combine webservices to a pipeline, thus one webservices choreography (composed of different tasks) could easily be used multiple times without any user interaction.

EST Matching

As second example, we present a pipeline of SoapDB offered by DKFZ Heidelberg and e2g offered by BiBiServ, both within the HOBIT network. A sequence will be retrieved in FASTA format from the EMBL database via SoapDB. This sequence will be pushed to the e2g webservice, which maps it to a specified genomic sequence. The result will be presented as alignments corresponding to the EBIApplicationResult Schema, which is also used by the BLAST webservices offered by EBI. This can be visualized or further processed.

BioDOM

Connecting webservices is mostly a problem of incompatible data formats which require more or less complex data conversion. To solve this problem using webservices in bioinformatics, we support XML formats beside proprietary formats for easy data exchange between common bioinformatics webservices. BiBiServ offers an API called „BioDOM“, which is also available as webservice. BioDOM provides a wide range of converters from and to more or less standardized grammars (XML Schemas), like SequenceML, AlignmentML, RNAML and EBIApplication. BioDOM is an open project hosted at SourceForge (http://biodom.sourceforge.net).



Webservices at BiBiServ

	Tool	Description	Input Format	Output Format	URL*
Genome Comp.	REPuter	REPuter computes all maximal duplications and reverse, complemented and reverse complemented repeats in a DNA input sequence.	FASTA	proprietary (ASCII), EBI Application Format (XML)	reputer/
Alignments	e2g	A web based tool for efficiently aligning genomic sequence to EST and cDNA data.	FASTA, (FASTA)	proprietary (ASCII), EBI Application Format (XML)	e2g/
	DIALIGN	DIALIGN is a novel alignment program based on segment-to-segment comparison. It is especially suited to detect local similarities among distantly related sequences.	SequenceML (XML)	AlignmentML (XML)	dialign/
RNA Studio	pknotsRG	PknotsRG is a tool for folding RNA secondary structures, including the class of simple recursive pseudoknots.	RNAML (XML)	RNAML (XML)	pknotsrg/
	RNAshapes	RNAshapes is a tool for RNA secondary structure prediction, making use of the abstract shape representation of RNA secondary structure.	RNAML (XML)	RNAML (XML)	rnashapes/
	RNAmovies	RNAmovies is a system for the visualization of RNA secondary structure landscapes.	DotBracket, RNAML (XML)	visualization of RNA structures	rnamovies/
	RNAfold	RNA folding tool of the Vienna RNA Package.	RNAML (XML)	RNAML (XML)	rnafold/
Evolutionary Relationship	ROSE	ROSE implements a new probabilistic model of RNA-, DNA-, or protein-sequence evolution.	array of parameter/ value, PhyloML (XML)	AlignmentML, SequenceML, PhyloML (XML)	rose/
	SplitsTree	SplitsTree uses the split decomposition method to analyze and visualize distance data, especially extracted from biological sequences.	NEXUS	PostScript	splits/

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

BiBiServ
Faculty of Technology
Bielefeld University
33594 Bielefeld

http://bibiserv.techfak.uni-bielefeld.de/
bibi-help@techfak.uni-bielefeld.de

References

REPuter
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 2001. Volume 29 (22), pages 4633-4642

e2g
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. Volume 32, Web Server issue, pages W301-304

DIALIGN
B. Morgenstern: DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinformatics 1999. Volume 15, pages 211 - 218

pknotsRG
J. Reeder, R. Giegerich: Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics, BMC Bioinformatics 2004. Volume 5, page 104

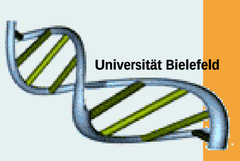
RNAshapes
R. Giegerich, B. Voss, M. Rehmsmeier: Abstract Shapes of RNA, Nucleic Acids Research 2004. Volume 32(16), pages 4843-4851

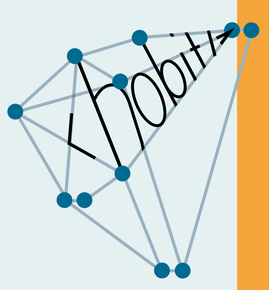
RNAmovies
R. Giegerich, D. Evers: RNA Movies: visualizing RNA secondary structure spaces. Bioinformatics 1999. Volume 15(1), pages 32-37

RNAfold
I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, P. Schuster: Fast Folding and Comparison of RNA Secondary Structures. Monatshefte für Chemie 1994. Volume 125, pages 167-188

ROSE
J. Stoye, D. Evers, F. Meyer: Bioinformatics 1998, Volume 14(2), pages 157-163

SplitsTree
D.H. Huson: SplitsTree: Analyzing and visualizing evolutionary data. Bioinformatics 1998. Volume 14(1), pages 68-73





Universität Bielefeld

design Susanne Konermann (BiBiServ Media)