HOBIT at the BiBiServ

Jan Krüger   Henning Mersch

Bielefeld Bioinformatics Service
Institute of Bioinformatics
CeBiTec Bielefeld University

jkrueger@TechFak.Uni-Bielefeld.DE
hmersch@TechFak.Uni-Bielefeld.DE

Cologne, March 2005

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

BiBiServ Environment
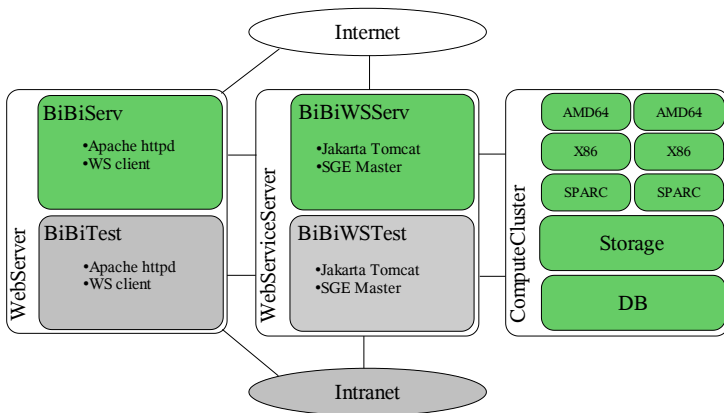
Solaris 10: Zones

- new feature of Solaris 10 release
- virtualize OS services
- a global zone (GZ) serves up to 8192 non-global zones (NGZ)
- communictation between NGZ only via network
- GZ can restrict resources of NGZ
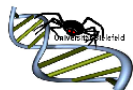 (CPU-time, memory usage, network,IO . . .)

Tomcat / Java

- allows to define very fine granulated access policies to Java sandbox behaviour
- available since Java 1.2
- configuration via policy file
  - hard to read for human eyes :-(
  - aweful GUI : Policy Configuration Tool (part of J2SDK)
- not very well documented

Architecture
http://bibiserv.techfak.uni-bielefeld.de/bibiws/
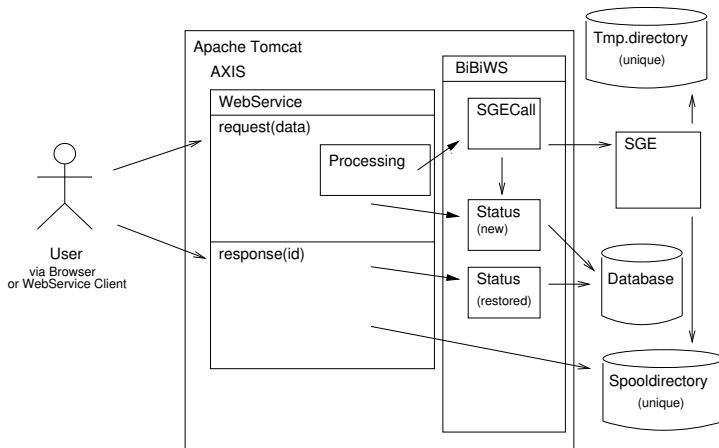
BiBiWS - Update
http://bibiserv.techfak.uni-bielefeld.de/bibiws/

- returns XML-DOM-Document

Overview	General Work	WebServices	Requests to HOBIT
	○○○○●○	○○○○○	

BiBiWS

BiBiWS - Update
http://bibiserv.techfak.uni-bielefeld.de/bibiws/
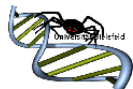
- returns XML-DOM-Document
- Tmp. dir (auto flush)

BiBiWS - Update
http://bibiserv.techfak.uni-bielefeld.de/bibiws/

- returns XML-DOM-Document
- Tmp. dir (auto flush)
- Spooldir: delete (3 days) / clean (30 days)

BiBiWS - Update
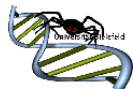http://bibiserv.techfak.uni-bielefeld.de/bibiws/

- returns XML-DOM-Document
- Tmp. dir (auto flush)
- Spooldir: delete (3 days) / clean (30 days)
- WebService - documentation site

BiBiWS - Update
http://bibiserv.techfak.uni-bielefeld.de/bibiws/

- returns XML-DOM-Document
- Tmp. dir (auto flush)
- Spooldir: delete (3 days) / clean (30 days)
- WebService - documentation site
- Perl SOAP::Lite Patch

BiBiWS - Update
http://bibiserv.techfak.uni-bielefeld.de/bibiws/
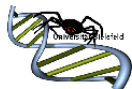
- returns XML-DOM-Document
- Tmp. dir (auto flush)
- Spooldir: delete (3 days) / clean (30 days)
- WebService - documentation site
- Perl SOAP::Lite Patch
- DRMAA (not yet)

WSDL-*Repository*
http://bibiserv.techfak.uni-bielefeld.de/wsdl/

BiBiServ
Bielefeld University Bioinformatics Server

Tools · Education · Administration · News · Links

Tools

Genome Comparison
  AGenDa
  REPuter
  more
Alignments
  E2G
  PoSSuMsearch
  more
Primer Design
  GeneFisher
RNA Studio
  RNAshapes
  RNAhybrid
  more
Evolutionary Relationship
  ROSE
  SplitsTree
  more
Others
  PREdictor
  GenDB
  more

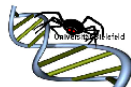WebService Overview

Tools, which are currently offered as a webservice

Name	WSDL	Description
bibiws	link	Demonstration WebService for BiBiWS Lib
e2g	link	e2g is a web-based tool which aligns genomic sequence to cDNA and EST data sets. The server hosts huge EST databases of a few gigabases in size in indexed data structures. This allows users to rapidly (i) detect new genes, (ii) verify the exon–intron structure of predicted genes, and (iii) determine splice variants in the uploaded genomic region of interest. Alignments can be visualized and matching sequences downloaded for further processing.
REPuter	link	REPuter computes all maximal duplications and reverse, complemented and reverse complemented repeats in a DNA input sequence.
SplitsTree	link	SplitsTree uses the split decomposition method to analyse and visualize distance data, e.g. data derived from biosequences. See also the original side of splitstree at the University of Tuebingen.

Tools, which are currently offline (due to maintaince)

Name	Description
BLAST	A sample webservice using NCBI blast for aligning two sequences that was developed during a HOBIT workshop in Bielefeld, 6/7th October 2004.
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.

BiBiWS

WSDL-*Repository*
http://bibiserv.techfak.uni-bielefeld.de/wsdl/

WebService Overview

Tools, which are currently offered as a webservice

Name	WSDL	Description
bibiws	link	Demonstration WebService for BiBiWS Lib
e2g	link	e2g is a web–based tool which aligns genomic sequence to cDNA and EST data sets. The server hosts huge EST databases of a few gigabases in size in indexed data structures. This allows users to rapidly (i) detect new genes, (ii) verify the exon–intron structure of predicted genes, and (iii) determine splice variants in the uploaded genomic region of interest. The Web interface provides a graphical overview of the results. Alignments can be visualized and matching sequences downloaded for further processing.
REPuter	link	REPuter computes all maximal duplications and reverse, complemented and reverse complemented repeats in a DNA input sequence.
SplitsTree	link	SplitsTree uses the split decomposition method to analyse and

WSDL-*Repository*
http://bibiserv.techfak.uni-bielefeld.de/wsdl/

| SplitsTree | link | SplitsTree uses the split decomposition method to analyse and visualize distance data, e.g. data derived from biosequences. See also the original side of splitstree at the University of Tuebingen. |

Tools, which are currently offline (due to maintaince)

Name	Description
BLAST	A sample webservice using NCBI blast for aligning two sequences that was developed during a HOBIT workshop in Bielefeld, 6/7th October 2004.
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.

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

- design of oligo sequences for microarrays

S. Rahmann
Fast large scale oligonucleotide selection using the longest common factor approach
Journal of Bioinformatics and Computational Biology, 1(2):343-361, 2003

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

- design of oligo sequences for microarrays
- input: string containing target and background sequences (FASTA), parameter array (key → value)

S. Rahmann
Fast large scale oligonucleotide selection using the longest common factor approach
Journal of Bioinformatics and Computational Biology, 1(2):343-361, 2003

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

- design of oligo sequences for microarrays
- input: string containing target and background sequences (FASTA), parameter array (key $\rightarrow$ value)

- output: Oligo/Target matrix (new XSD)

S. Rahmann
Fast large scale oligonucleotide selection using the longest common factor approach
Journal of Bioinformatics and Computational Biology, 1(2):343-361, 2003

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

- design of oligo sequences for microarrays
- input: string containing target and background sequences (FASTA), parameter array (key → value)
- uses *OligoSelection* (FU-Berlin) for non-unique design
- output: Oligo/Target matrix (new XSD)

S. Rahmann
Fast large scale oligonucleotide selection using the longest common factor approach
Journal of Bioinformatics and Computational Biology, 1(2):343-361, 2003

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

- design of oligo sequences for microarrays
- input: string containing target and background sequences (FASTA), parameter array (key → value)
- uses *OligoSelection* (FU-Berlin) for non-unique design
- output: Oligo/Target matrix (new XSD)
- under development

S. Rahmann
Fast large scale oligonucleotide selection using the longest common factor approach
Journal of Bioinformatics and Computational Biology, 1(2):343-361, 2003

Alignments

e2g
http://bibiserv.techfak.uni-bielefeld.de/e2g/

- e2g WebService efficiently maps large EST/cDNA data sets to genomic using VMatch
- input: parameter array (key → value),
  string containing a single sequence (FASTA)
- output: EBI - Application Alignment format
- XML-based input format, Java based GUI client
  (under development)

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

Universität Bielefeld

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

- REPuter WebService allows fast computation of degenerate repeats in complete genomes, based on VMatch.
- input: string containing one or more sequences in FASTA, parameter array (key $\rightarrow$ value)
- output: EBI - Application Alignment format
- XML-based input format, Java based graphically client (under development)

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 Res. 2001

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

- SplitsTree WebService uses the split decomposition method to analyze and visualize distance data
- input: string containing distance data in NEXUS format
- output: string containing visualized data in PostScript format
- under development: XML-based input and output format (SVG instead of PostScript)

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

.

Other WebServices
currently in development

- **dialign** constructs pairwise and multiple sequence alignments
  by comparing whole segments of the sequences

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

A.W.M. Dress, G.Füllen, S.W. Perrey
A Divide and Conquer Approach to Multiple Alignment
AAAI Press, Menlo Park, CA, USA, 107-113, 1995

J. Stoye, D. Evers, F. Meyer
Generating Benchmarks for Multiple Sequence Alignments and Phylogeentic Reconstructions
Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology, AAAI Press

Other WebServices
currently in development

- **dialign** constructs pairwise and multiple sequence alignments by comparing whole segments of the sequences
- **dca** produces fast, high quality simultaneous multiple sequence alignments of amino acid, RNA, or DNA sequences

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

A.W.M. Dress, G.Füllen, S.W. Perrey
A Divide and Conquer Approach to Multiple Alignment
AAAI Press, Menlo Park, CA, USA, 107-113, 1995

J. Stoye, D. Evers, F. Meyer
Generating Benchmarks for Multiple Sequence Alignments and Phylogenetic Reconstructions
Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology, AAAI Press

Other WebServices
currently in development

- **dialign** constructs pairwise and multiple sequence alignments by comparing whole segments of the sequences
- **dca** produces fast, high quality simultaneous multiple sequence alignments of amino acid, RNA, or DNA sequences
- **Rose** implements a probabilistic model of the evolution of RNA-, DNA-, or protein-like sequences.

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

A.W.M. Dress, G.Füllen, S.W. Perrey
A Divide and Conquer Approach to Multiple Alignment
AAAI Press, Menlo Park, CA, USA, 107-113, 1995

J. Stoye, D. Evers, F. Meyer
Generating Benchmarks for Multiple Sequence Alignments and Phylogeentic Reconstructions
Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology, AAAI Press

Requests to HOBIT

- MSA Grammar (*BSML* ?)

Requests to HOBIT

- MSA Grammar (*BSML* ?)
- *Hobit Statuscode* discussion

Overview

General Work
oooooo

WebServices
ooooo

Requests to HOBIT

Requests to HOBIT

- MSA Grammar (*BSML* ?)
- *Hobit Statuscode* discussion
- WebService Repository

Requests to HOBIT

- MSA Grammar (*BSML* ?)
- *Hobit Statuscode* discussion
- WebService Repository
- Reliability of WebServices