Netherlands
Bioinformatics
Centre

National Collaborative Platform for Genomics and Proteomics Data Analysis

Hailiang (Leon) Mei

(https://wiki.nbic.nl/index.php/Next_Generation_Sequencing)

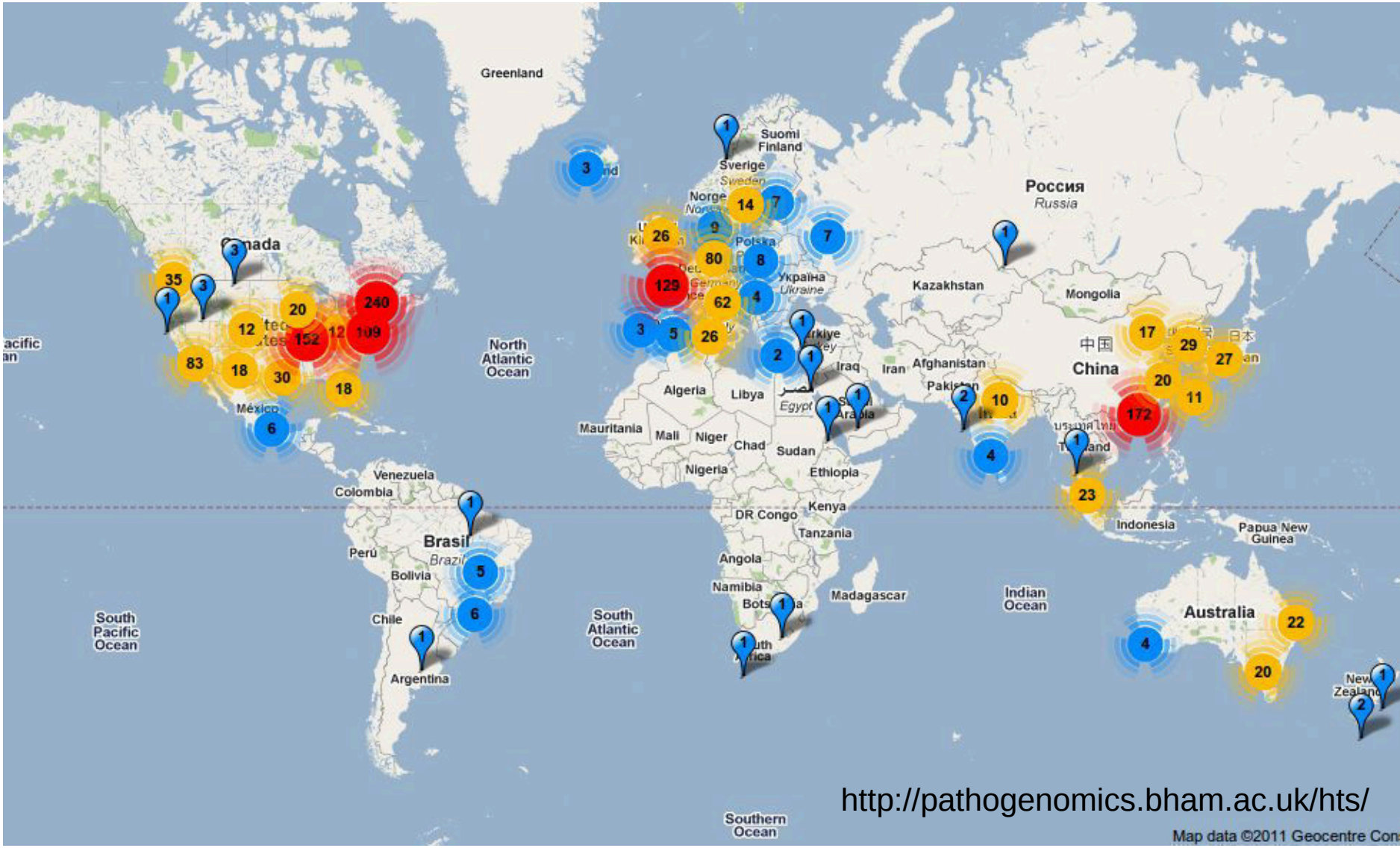
Outline

- NBIC BioAssist program
- galaxy.nbic.nl

Netherlands Bioinformatics Centre

Next generation sequencing capacity

- Your $1000 Genome



However, bioinformatics tools!@#$

nbic

[[BioAssist]]

page | discussion | edit | history | move | watch

NGS Tools

Return to the main page of Next Generation Sequencing

Please feel free to add your tools and your experiences! You can make a new page about a tool if you want to describe it in some detail.

navigation
- BioAssist Main Page
- Current events
- Recent changes
- Random page
- Help
- BioRange private

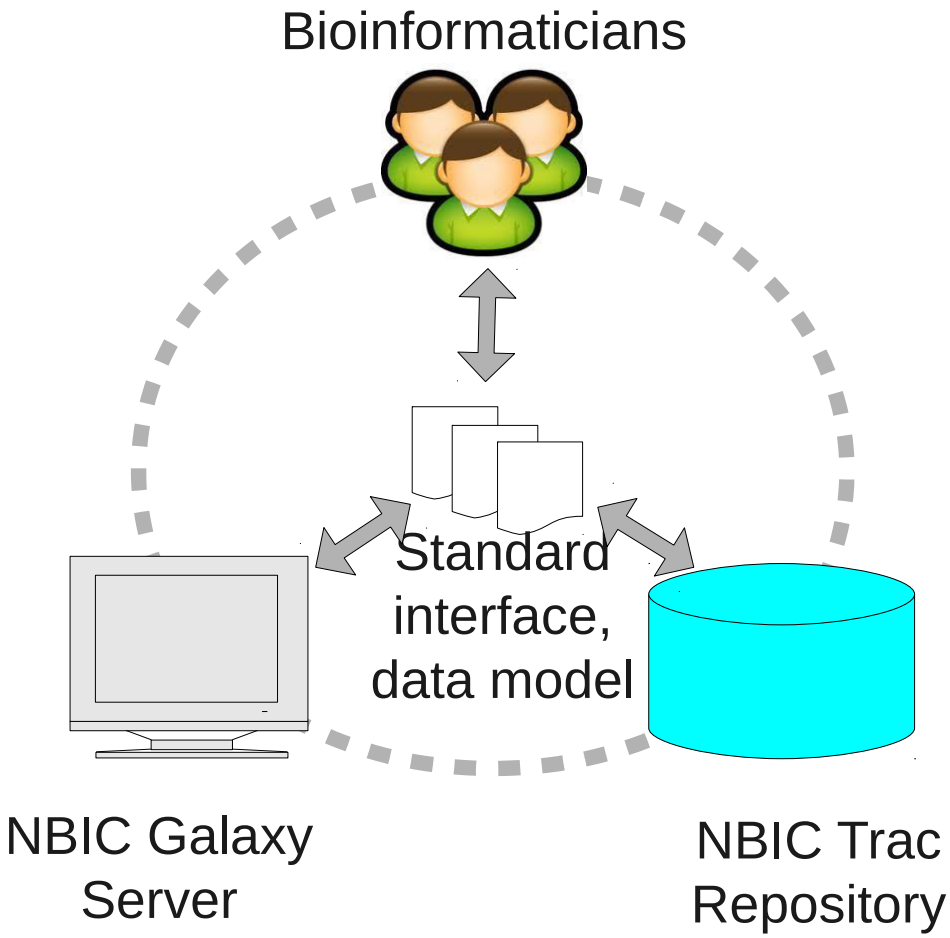
search
Go | Search

toolbox
- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

Category	Package	Description	Performance experience
Assembly (de novo)	ALLPATHS-LG	Introduced on January 2011 by broad institute. "It works on both small and large (mammalian size) genomes. To use it, you should first generate ~100 base Illumina reads from two libraries: one from ~180 bp fragments, and one from ~3000 bp fragments, both at about 45x coverage. Sequence from longer fragments will enable longer-range continuity."	None
Viewer	EagleView genome viewer	EagleView is an information-rich genome assembler viewer with data integration capability. EagleView can display a dozen different types of information including base qualities, machine specific trace signals, and genome feature annotations.	None
Alignment	MUMmerGPU	MUMmerGPU is a low cost, ultra-fast sequence alignment program designed to handle the increasing volume of data produced by new, high-throughput sequencing technologies. MUMmerGPU demonstrates that even memory-intensive applications can run significantly faster on the relatively low-cost GPU than on the CPU.	None
Methylation	Batman	Bayesian tool for methylation analysis (Batman)—for analyzing methylated DNA immunoprecipitation (MeDIP) profiles	None
Base-calling	Alta-Cyclic	Alta-Cyclic is a novel Illumina Genome-Analyzer (Solexa) base caller. Alta Cyclic Features: Longer Reads, More Accurate Reads (compared to Solexa's default base caller), Reduces systematic bias towrsd a certain nucleotide in later cycles. On a GAII platform, Alta Cyclic was able to provide a large amount of useful reads after 78 cycles.	None
Enrichment/peak calling	FindPeaks 3.1	Findpeaks was developed to perform analysis of ChIP-Seq experiments. It uses a naive algorithm for identifying regions of high coverage, which represent Chromatin Immunoprecipitation enrichment of sequence fragments, indicating the location of a bound protein of interest.	None
Assembly (de novo)	ALLPATHS	De novo assembly of whole-genome shotgun microreads.	None
Assembly (de novo)	SHARCGS	SHARCGS is a suitable tool for fully exploiting novel sequencing technologies by assembling sequence contigs de novo with high confidence and by outperforming existing assembly algorithms in terms of speed and accuracy. Authors are Dohm JC, Lottaz C, Borodina T and Himmelbauer H. from the Max-Planck-Institute for Molecular Genetics.	None
Assembly (de novo)	Velvet	Velvet is a de novo genomic assembler specially designed for short read sequencing technologies, such as Solexa or 454. Need about 20-25X coverage and paired reads. Developed by Daniel Zerbino and Ewan Birney at the European Bioinformatics Institute (EMBL-EBI).	None
Assembly (de novo)	EDENA	De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Made by Hernandez D et al.	None
Assembly	SSAKE	The Short Sequence Assembly by K-mer search and 3' read Extension (SSAKE) is a genomics application for aggressively assembling millions of short nucleotide sequences by progressively searching for perfect 3'-most k-mers using a DNA prefix tree. SSAKE is designed to help leverage the information from short sequences reads by stringently clustering them into contigs that can be used to characterize novel sequencing targets. Authors are René Warren, Granger Sutton, Steven Jones and Robert Holt from the Canada's Michael Smith Genome Sciences Centre. Perl/Linux.	None
Alignment	qpalma	QPalma is an alignment tool targeted to align spliced reads produced by Next Generation sequencing platforms such as Illumina Solexa or 454. QPalma aligns short reads to the genomic sequences in an optimal way according to its underlying algorithm and trained parameters. It creates an alignment using dynamic programming (written in C++), and returns the alignment in a psl like format. The algorithms computes optimal local alignments, so if no alignment has been found it is because no alignment got a sufficiently high alignment score.	None
Alignment	SOAP	SOAP (Short Oligonucleotide Alignment Program) is a program for efficient gapped and ungapped alignment of short oligonucleotides onto reference sequences. The program is designed to handle the huge amounts of short reads generated by parallel sequencing using the new generation Illumina-Solexa sequencing technology. SOAP is compatible with numerous applications, including single-read or pair-end resequencing, small RNA discovery and mRNA tag sequence mapping. SOAP is a command-driven program, which supports multi-threaded parallel computing, and has a batch module for multiple query sets. Author is Ruiqiang Li at the Beijing Genomics Institute. C++ for Unix.	None
Assembly (de novo)	SOAPdenovo	SOAPdenovo, a short read de novo assembly tool, is a package for assembling short oligonucleotide into contigs and scaffolds.	None
SNP/Indel Discovery	SOAPsnp	SOAPsnp is an accurate consensus sequence builder based on soap1 and SOAPaligner/soap2's alignment output. It calculates a quality score for each consensus base, which can be used for any latter process to call SNPs.	None
Workflows	Galaxy	Galaxy allows you to do analyses you cannot do anywhere else without the need to install or download anything. You can analyze multiple alignemnts, compare genomic annotations, profile metagenomic samples and much much more...	None
Communities	SeqAnswers	Next generation sequencing community.	None
Integrated solutions	CLCbio Genomics Workbench	de novo and reference assembly of Sanger, 454, Solexa, Helicos, and SOLiD data. Commercial next-gen-seq software that extends the CLCbio Main Workbench software. Includes SNP detection, browser and other features. Runs on Windows, Mac OS X and Linux.	None
Integrated solutions	NextGENe	de novo and reference assembly of Illumina and SOLiD data. Uses a novel Condensation Assembly Tool approach where reads are joined via "anchors" into mini-contigs before assembly. Requires Win or MacOS.	None
Integrated solutions	SeqMan Genome Analyser	Software for Next Generation sequence assembly of Illumina, 454 Life Sciences and Sanger data integrating with Lasergene Sequence Analysis software for additional analysis and visualization capabilities. Can use a hybrid templated/de novo approach. Early release commercial software. Compatible with Windows® XP X64 and Mac OS X 10.4.	None
Alignment	ELAND	Efficient Large-Scale Alignment of Nucleotide Databases. Whole genome alignments to a reference genome. Written by Illumina author Anthony J. Cox for the Solexa 1G machine.	None
Assembly	EULER	Short read assembly. By Mark J. Chaisson and Pavel A. Pevzner from UCSD (published in Genome Research).	None
Alignment	Exonerate	Various forms of alignment (including Smith-Waterman-Gotoh) of DNA/protein against a reference. Authors are Guy St C Slater and Ewan Birney from EMBL. C for POSIX.	None
Alignment & Mapping	GMAP	GMAP (Genomic Mapping and Alignment Program) for mRNA and EST Sequences. Developed by Thomas Wu and Colin Watanabe at Genentec. C/Perl for Unix.	None
Alignment & Assembly	MOSAIK	Reference guided aligner/assembler. Written by Michael Strömberg at Boston College.	None
Alignment & Mapping	MAQ	Mapping and Assembly with Qualities (renamed from MAPASS2). Particularly designed for Illumina-Solexa 1G Genetic Analyzer, and has preliminary functions to handle ABI SOLiD data. Written by Heng Li from the Sanger Centre.	None
Alignment	MUMmer	MUMmer is a modular system for the rapid whole genome alignment of finished or draft sequence. Released as a package providing an efficient suffix tree library, seed-and-extend alignment, SNP detection, repeat detection, and visualization tools. Version 3.0 was developed by Stefan Kurtz, Adam Phillippy, Arthur L Delcher, Michael Smoot, Martin Shumway, Corina Antonescu and Steven L Salzberg - most of whom are at The Institute for Genomic Research in Maryland, USA. POSIX OS required.	None
Alignment	Novocraft	Tools for reference alignment of paired-end and single-end Illumina reads. Uses a Needleman-Wunsch algorithm. Available free for evaluation, educational use and for use on open not-for-profit projects. Requires Linux or Mac OS X.	None
Assembly	RMAP	Assembles 20 - 64 bp Solexa reads to a FASTA reference genome. By Andrew D. Smith and Zhenyu Xuan at CSHL. (published in BMC Bioinformatics). POSIX OS required.	None
Alignment	SeqMap	Works like ELand, can do 3 or more bp mismatches and also INDELs. Written by Hui Jiang from the Wong lab at Stanford. Builds available for most OS's.	None
Assembly	SHRIMP	Assembles to a reference sequence. Developed with Applied Biosystem's colourspace genomic representation in mind. Authors are Michael Brudno and Stephen Rumble at the University of Toronto. Works with data in letterspace (Roche, Illumina), colourspace (AB) and Helicos space.	probcalc gave a 'segmentation fault' on a x86_64-pc-linux-gnu with Roche data

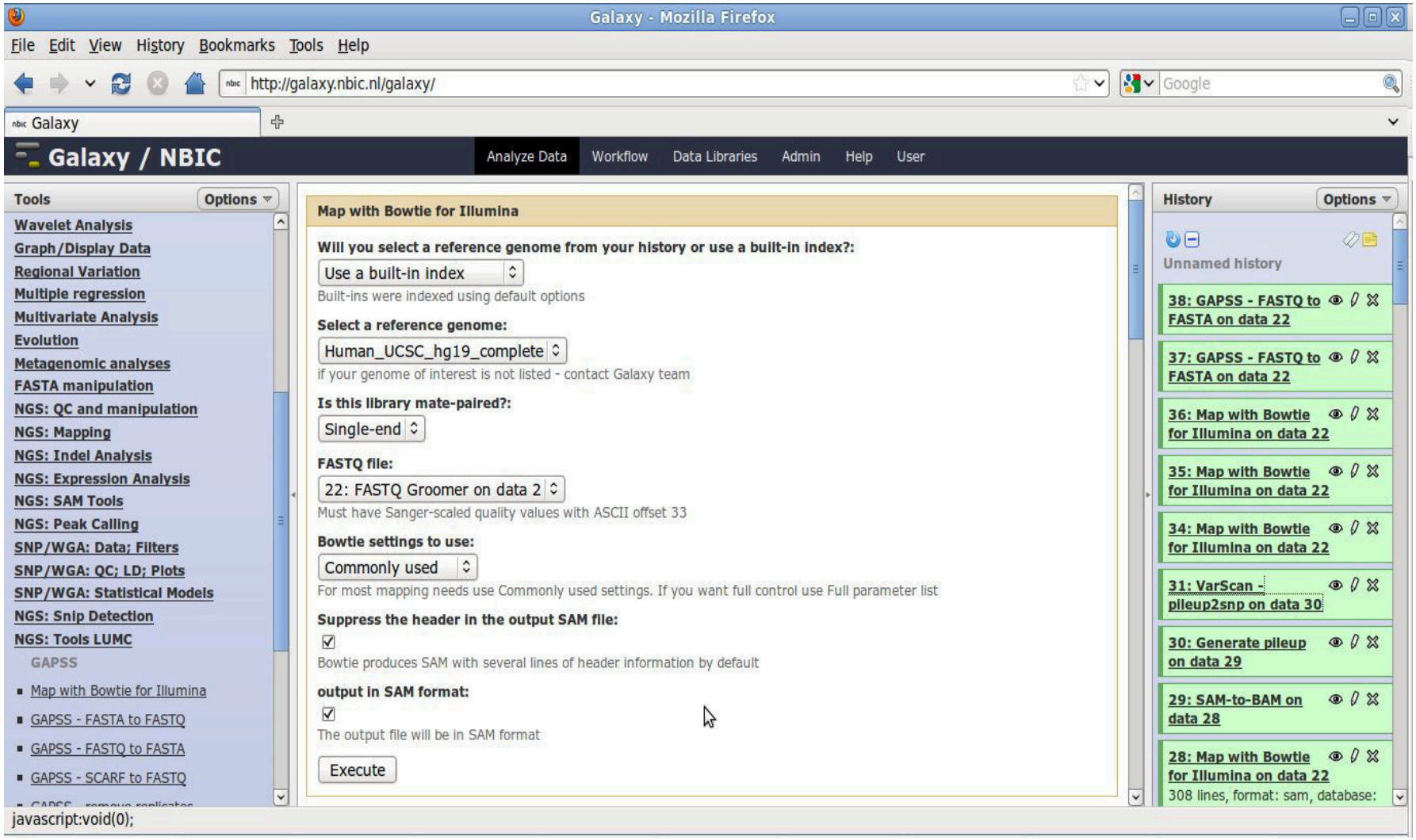
nbic Netherlands Bioinformatics Centre

NBIC BioAssist



nbic Netherlands Bioinformatics Centre

NBIC Galaxy server



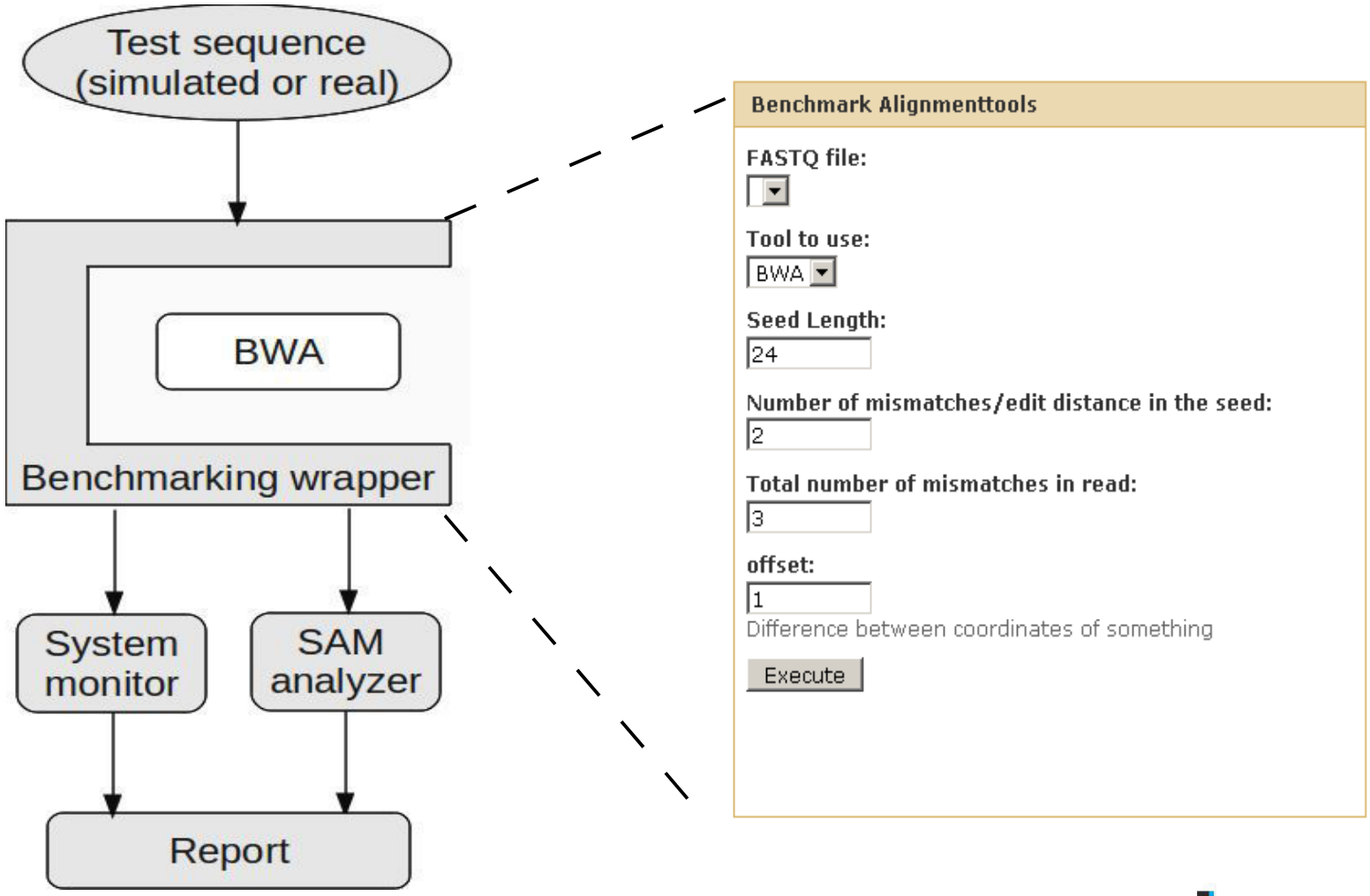
nbic Netherlands Bioinformatics Centre

NBIC Galaxy pipelines

- Alignment & variant calling pipelines
  - GAPSSv1 (Illumina, Bowtie, Varscan, Ensembl @ LUMC)
  - GAPSSv2 (Illumina, Stampy, SeattleAPIs @ LUMC)
  - Genome of the Netherlands (Illumina, BWA, Picard, GATK @ UMCG)
  - SAP42 (Solid, BWA, SAMtools @ Hubrecht)
- Alignment software benchmarking (Hubrecht)
- Proteomics msCompare (RUG)
- Chip-seq (EMC)
- NGS QC for Illumina, Solid, 454
- De novo assembly software benchmarking (WUR)
- SV software benchmarking (LUMC, Hubrecht)

nbic Netherlands Bioinformatics Centre

Alignment tool benchmarking



Netherlands Bioinformatics Centre

Galaxy VM

- Easy installation of local Galaxy
- Run jobs locally so no security&privacy worries
- Shipped with NBIC recommended pipelines
- A step towards Cloud based computation
- Support reproducibility
- Problems
  - >20G :(
  - How to lift over users, history datasets, etc

nbic Netherlands Bioinformatics Centre

Galaxy RPM repository

- An international collaboration
  - Joachim Jacob, Luc Ducazu, (Vlaams Instituut voor Biotechnologie, BITS)
  - David van Enckevort (NBIC)
  - Adam Huffman (Manchester Univ.)
- A stable repository of easily installable packages for NGS tools
- Open

nbic Netherlands Bioinformatics Centre

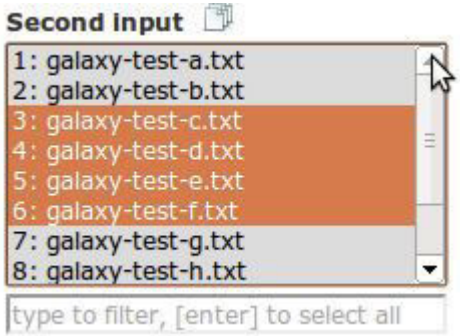
NBIC Galaxy hackathon (April 13,14, 2011)

- Alex Bossers (WUR), Dannon Baker (Emory), Frans Paul Ruzius (Hubrecht), Freddy de Bree (CVI, WUR), Freek de Bruijn (NBIC), Henk van den Toorn (UU), Ishtiaq Ahmad (RUG), Joachim Jacob (VIB - BITS), Martijn Vermaat (LUMC), Nate Coraor (Penn State), Rob Hooft (NBIC), Wil Koetsier (UMCG)

nbic Netherlands Bioinformatics Centre

Hackathon results

- Tool tags & tool_conf.xml autogeneration
- Looping through multiple files



nbic Netherlands Bioinformatics Centre

Acknowledgement

- Hubrecht/UMCU

Pieter Neerincx

Frans Paul Ruzius

Bas van Breukelen

Edwin Cuppen

Victor Guryev

- Erasmus MC

Rutger Brouwer

Wilfred van Ijcken

- AMC

Barbera van Schaik

Carsten Byrman

Perry Moerland

Antoine van Kampen

Silvia Olabarriaga

- VIB

Joachim Jacob

Luc Ducazu

Michiel Bataillie

Stéphane Plaisance

- UMCG

Freerk van Dijk

Morris Swertz

Berend Hoekman

Peter Horvatovich

- LUMC

Jeroen Laros

Matthew Hestand

Kai Ye

Kostas Karasavvas

Johan T. den Dunnen

- Wageningen UR

Jan van Haarst

Alex Bossers

Roeland van Ham

- NBIC

David van Enckevort

Marc van Driel

Freek de Bruijn

Victor de Jager

Christine Chichester

Rob Hooft

Barend Mons

nbic Netherlands Bioinformatics Centre

Acknowledgement - continued

- The Galaxy team
- Developers of open source tools
- You

Netherlands Bioinformatics Centre


nbic
Netherlands
Bioinformatics
Centre

NGS task force members

- Amsterdam Medical Center, Clinical Epidemiology, Biostatistics and Bioinformatics group
- Erasmus Medical Center, Biomics group
- Erasmus Medical Center, Bioinformatics group
- Erasmus Medical Center, Complex Genetics group
- Hubrecht Institute, Genome Biology group
- Leiden University Medical Center, Center for Human and Clinical Genetics
- Leiden University Medical Center, Molecular Epidemiology group
- Nijmegen Centre for Molecular Life Sciences, Bacterial Genomics group
- Nijmegen Centre for Molecular Life Sciences, Department of Human Genetics
- SARA, High Perfomance Computing and Visualization
- Technical University Delft, Bioinformatics group
- University Medical Center Groningen, Genomics Coordination Centre
- Wageningen University, Bioinformatics group
- Wageningen University, Central Veterinary Institute
- KeyGene N.V.

nbic Netherlands Bioinformatics Centre

Bioinformatics groups in NL

Server performance

Galaxy VM – bring computation to data!