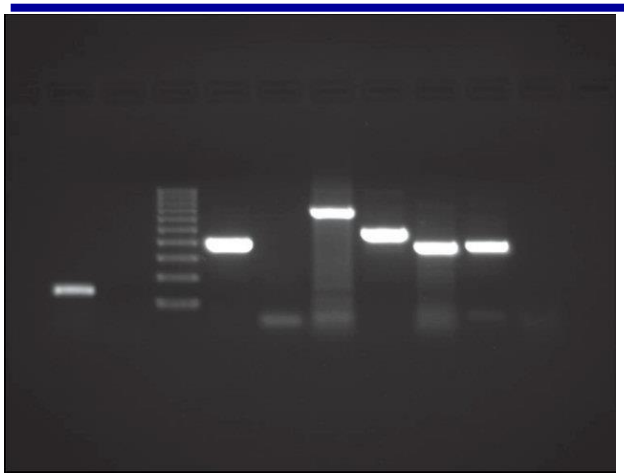
Systems and servers for NGS data analysis

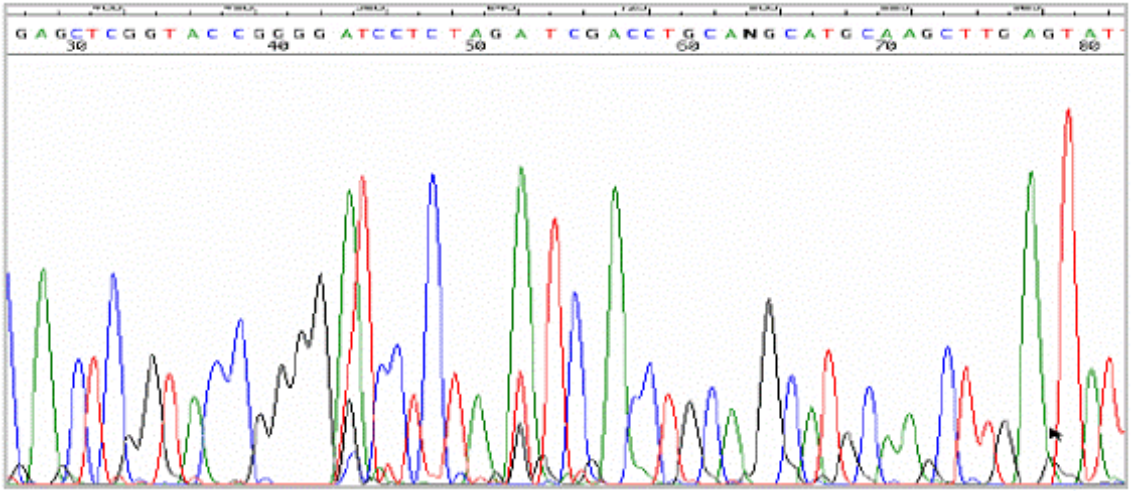
Rosangela Tozzoli

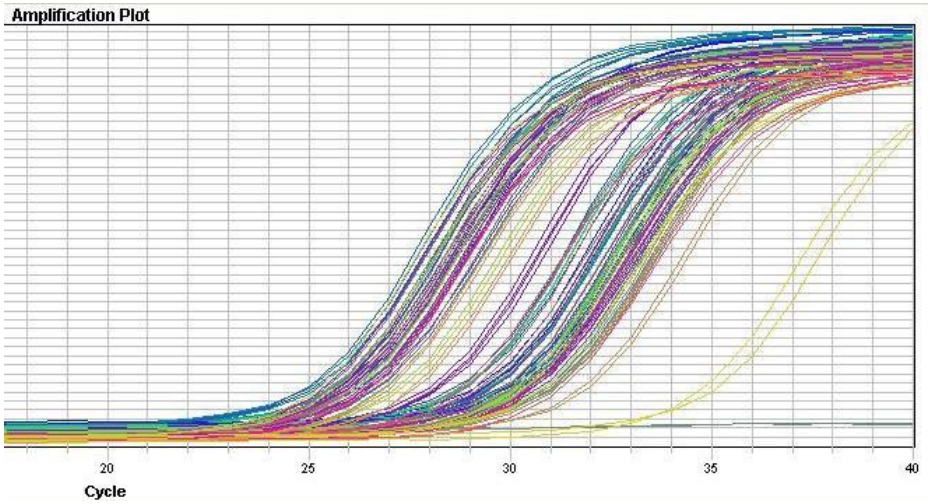
WGS course,
October 2020

Istituto Superiore di Sanità, Dep. of Food Safety, Nutrition and Veterinary Public Health
European Union and National Reference Laboratory for *E. coli*, Rome, Italy

Data Analysis: A new syntax







Data Analysis: A new syntax

Mobio_16Run_400_hiq_Pool17062015.fastq

```
@C9IBY:00426:00452
ATCAATTAAAATTTTATCTAGCGCATTACATGCACTGATTTTATCCATTTTTGCATTCACCACCACATCGAGCAATTTTTCCCAG
TCCGCATCGCTGGCAATATAGGCGAAGTTATTCCCTCTACCACTAACCAGAACGGCGCACCGAGCATGCTCTTTAACGAAAGCAA
TCAATTGCTCGCCTCCACGTGGAACAATCAAATCGATCGGTTCATCGGGGTTTTCAGGAAGGCTCGCGTAGCCTCACGATCCAAC
GTGATGAGTTTAATCCAATCTGTCGTCAATCCATTCTGCCGTAACGCGTC
+
BBCDAC6;;;/;CC8CCCCCC?>??C@CCEEDDD;;;;<<</<<<7<<<<<.;BAA@CCCADDABCDCCCCCCC@@@@@.;;
1;;?CACC???CCCCACCA?;;:=9=>CACE@CC@CC>CCCCC;>?CCACACAA=@@>B;;:@288888:@=888?,82::?
D<BBB=CB@5828=CBBC?CC?;;>:>7<9B>=@@@A:@BBAA??;??
B@@8888*888*8<A=A2848888=:::@@=@@@CB?;;8;?BBBB@@@ABBD3:2:0171777000:008700*//*///
828<<4:;@?87
@C9IBY:03696:02678
CTTGGTGGTAATGGTGGTGTGGTTGGGCCGGAAGCCATCGCCGGGGTTGGCGCTATTCTTGAAATGCTGCTCGGTGCAGCAGGTG
CTGATGGCCAGGTTTTTGACTCGCTGGGCGATACCATCTCTGGTCTGATGGGAGACACAGGTACCTCGGCGTCTGAAGCGGCCGA
TGCCATCACCGCCAGCCTGGGAGGAGCCCTGGATGGTAGCCTGGCGGGCCTGGGTATCGATGCTCCAGCCGAAGCCGTGATCGGA
GCCCTGTTGAGTGGTAAGGCGGGGGCGGTGATGCCGTCTTGAC
+
?@ACACCA????CACCA?;;:5:5:C/:5:<B=@B<;>CCCACDD:DADADDCCCFAB@?:::/;BBB>@?CD@CCCC>;;@<??
ACCCCD@C@CA4888:08@CCC@;;CC>CC@@@CACCCCCDDACBCCCCCD=CCCCC@@@AC@@;??>28888:?
@ACCE@C@CDDDC>?>C@B?C?:CCCACDE?C?7<888,:@4>?C?B@;>;>B7;?C=CAFCD<CC??>CC?@?
CADDCACD9::?CC???@A=???B;BBC?A@@@A=AB?E:?>BBB3:;7;?BBBB>B???>B??
@C9IBY:01239:00533
GGTGATGGTGGTGGTTCGGGTGTGGGTGTGTGTCGGGCGTGTGGGGGCGGGTTGCGGTTGGCTGGCGCGCATGGGCGCCCGGTCA
CGGTGAACGTCACCATCCCGCTGTCCACCCTGCTCGGCCTGGACGACCACCCGCG
+
D@DDCCC@CCADDAC>;;;1;;;;D?D>?>=:::@@;B>??DDGEE5CCC>CAC;;8;>B>????A;;;?>:=>:?CCF?
D@CCD??>?CCAF??=BA5::::/:88::B4:::/:::AAC4:5;B><<000*00<*000
@C9IBY:02674:01311
TATTACCACAGGTAAGTGAATTAAATGATATTAATATTTTTCAACGGTTAGCAGAAATTTATTCGACAGCTGGGGAATTTGAAAA
AGCAATTCACTATTATGAACGTGTTTTAGAAGACAGACTTGATATTGACTGCTTATTTGGCTACGGGTTAACAGCATATCAAGCA
```

Data Analysis: Software stand-alone

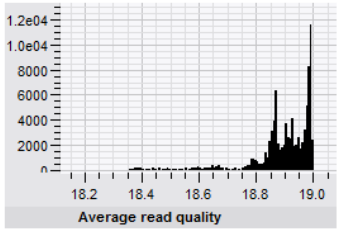
BioNumerics

Manage next generation sequencing data

The new sequence read sets experiment type offers an integrated environment for importing, preprocessing and analyzing sets of reads from high throughput sequencers or public repositories.

An integrated NGS data analysis platform

- Fast import of sequence read sets from various next generation sequencing platforms, such as Roche 454, Illumina Solexa, IonTorrent, etc.
- Storage of large amounts of short sequences (including paired-end reads) and quality scores.
- Comprehensive data preprocessing and quality control settings for demultiplexing, splitting paired-end reads, primer removal, structural and quality trimming, chimera detection and cleaning up sequence read sets.
- Global statistics calculation of sequence reads: creation of read length histograms, revision of base distribution, and quality score distribution. Generation of reports in rich text, table and chart formats.
- Sequence read sets are *database objects*, meaning that they can be annotated using custom information fields and that *user privileges* determine who is allowed to access and/or modify the data.
- Create comparisons for Kmer based clustering of sequence read sets, using all available similarity coefficients and hierarchical clustering methods.



User-Friendly Interface, Processing, RAM needed

geneious

Transform biological data into knowledge and actionable insights

Biological sequence data made easy

- Painfree analysis with an intuitive, user-friendly interface
- Simply import and convert a vast range of data types
- Cross-platform: Mac, Windows, Linux
- Licensing options to suit your users and your budget cycle
- Comprehensive support for technical issues, set up and training
- Customize with your own algorithms, plugins or workflows
- Increase process efficiency and improve data organisation
- High interoperability with good API to link LIMS and other tools

Data Analysis: Software stand-alone

CLC Genomics Workbench

QIAGEN

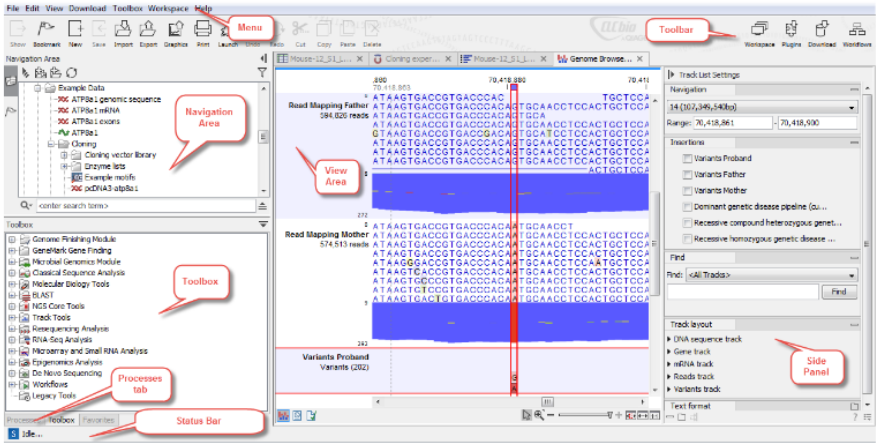
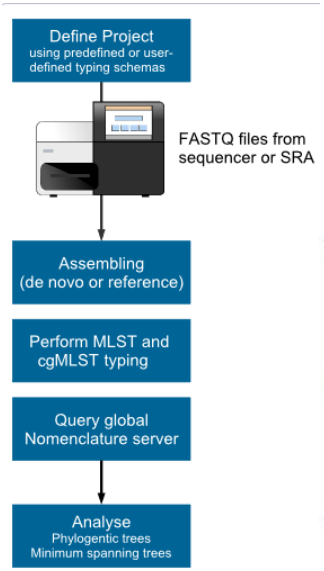


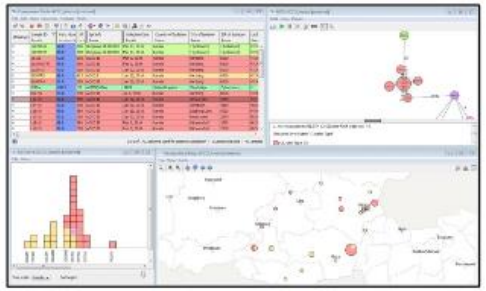
Figure 2.1: *The user interface.*

ridom BIOINFORMATICS

Ridom SeqSphere+



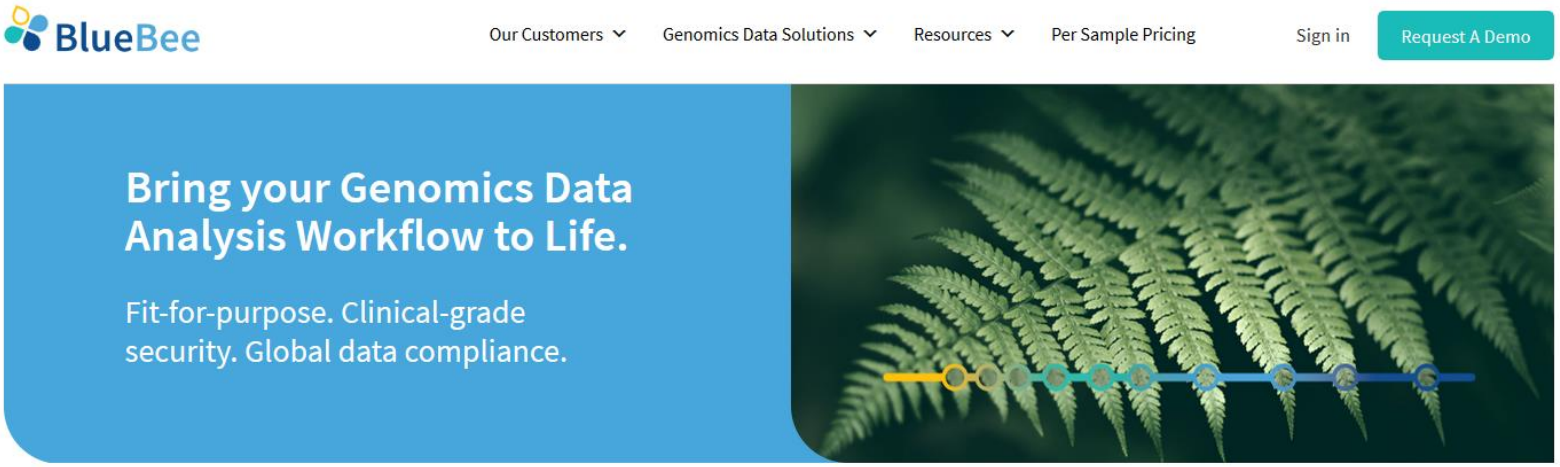
Pipeline for automated sequence analysis



Bacterial Genome characterization

Genome- wide allele and SNP calling

Data Analysis: Software stand-alone

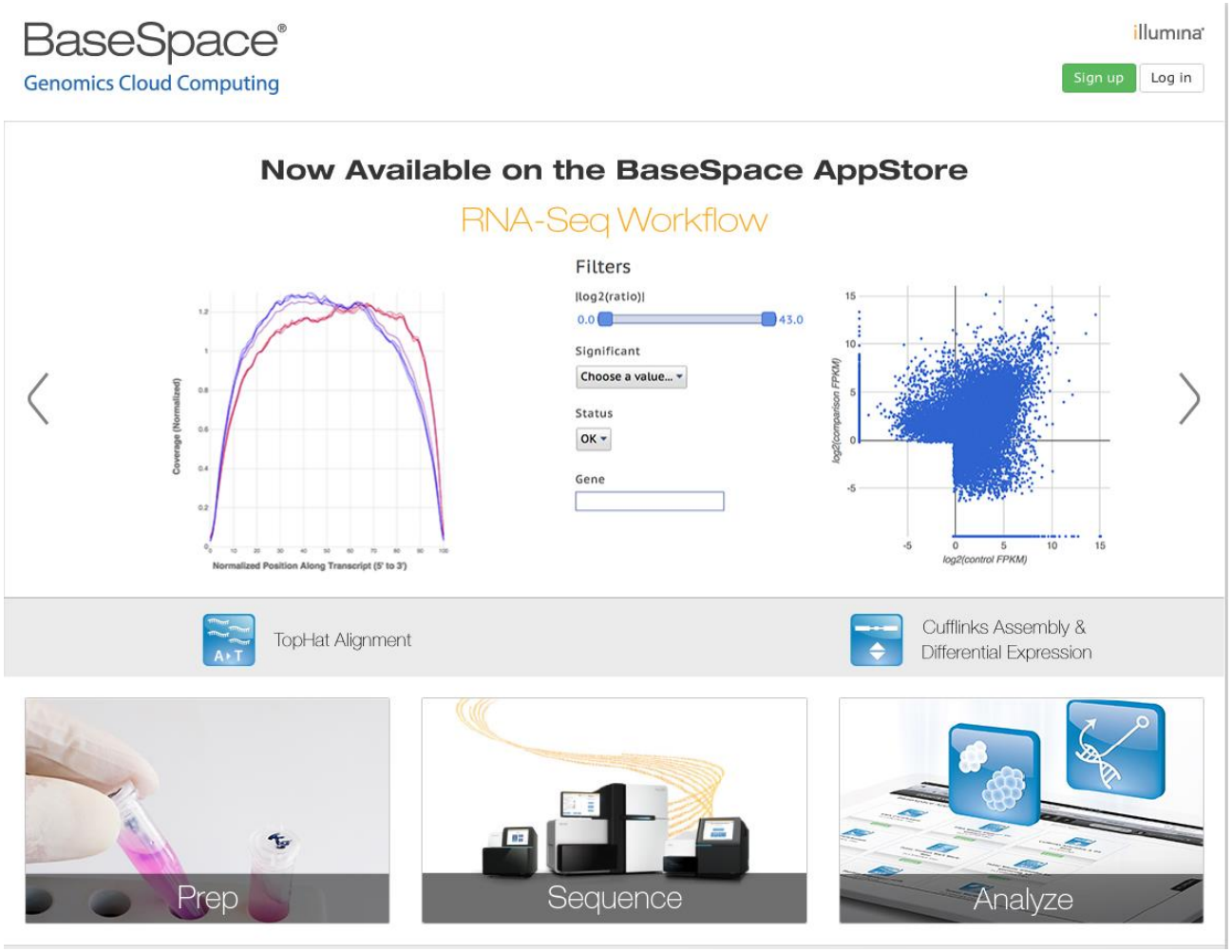


Data Analysis: Software stand-alone

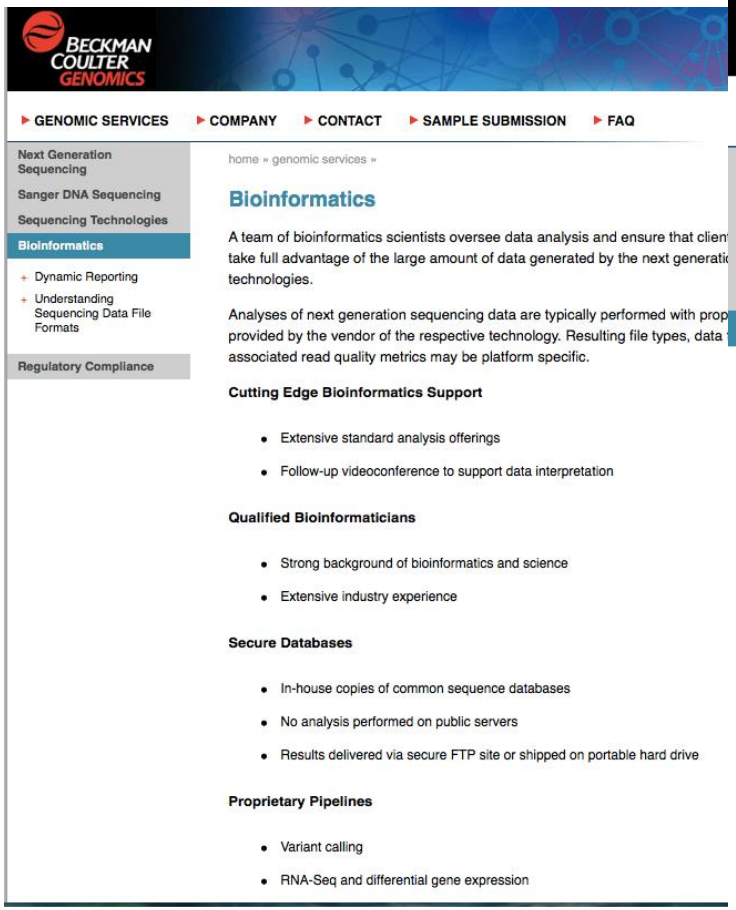
- *de novo* assembly
- Search for interesting genes
- Alignment of sequences, production of VCF files

BUILT IN THE ION TORRENT TECHNOLOGY PACKAGE

Data Analysis: Cloud-based Software



Data Analysis: Outsourcing



BECKMAN COULTER GENOMICS

GENOMIC SERVICES | COMPANY | CONTACT | SAMPLE SUBMISSION | FAQ

- Next Generation Sequencing
- Sanger DNA Sequencing
- Sequencing Technologies
- Bioinformatics
  - Dynamic Reporting
  - Understanding Sequencing Data File Formats
- Regulatory Compliance

home » genomic services »

Bioinformatics

A team of bioinformatics scientists oversee data analysis and ensure that clien take full advantage of the large amount of data generated by the next generati technologies.

Analyses of next generation sequencing data are typically performed with prop provided by the vendor of the respective technology. Resulting file types, data associated read quality metrics may be platform specific.

Cutting Edge Bioinformatics Support

- Extensive standard analysis offerings
- Follow-up videoconference to support data interpretation

Qualified Bioinformaticians

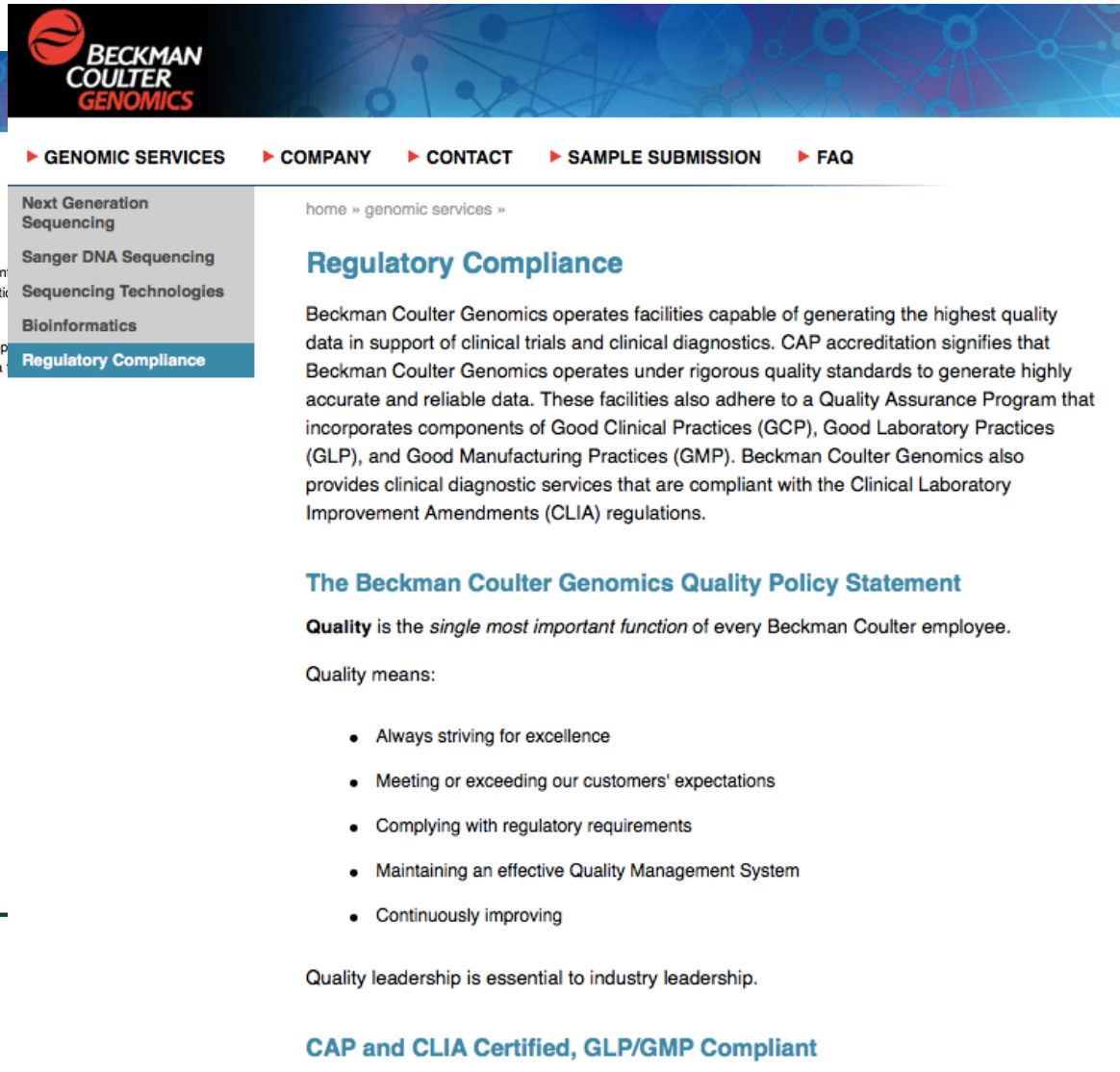
- Strong background of bioinformatics and science
- Extensive industry experience

Secure Databases

- In-house copies of common sequence databases
- No analysis performed on public servers
- Results delivered via secure FTP site or shipped on portable hard drive

Proprietary Pipelines

- Variant calling
- RNA-Seq and differential gene expression



BECKMAN COULTER GENOMICS

GENOMIC SERVICES | COMPANY | CONTACT | SAMPLE SUBMISSION | FAQ

- Next Generation Sequencing
- Sanger DNA Sequencing
- Sequencing Technologies
- Bioinformatics
- Regulatory Compliance

home » genomic services »

Regulatory Compliance

Beckman Coulter Genomics operates facilities capable of generating the highest quality data in support of clinical trials and clinical diagnostics. CAP accreditation signifies that Beckman Coulter Genomics operates under rigorous quality standards to generate highly accurate and reliable data. These facilities also adhere to a Quality Assurance Program that incorporates components of Good Clinical Practices (GCP), Good Laboratory Practices (GLP), and Good Manufacturing Practices (GMP). Beckman Coulter Genomics also provides clinical diagnostic services that are compliant with the Clinical Laboratory Improvement Amendments (CLIA) regulations.

The Beckman Coulter Genomics Quality Policy Statement

Quality is the *single most important function* of every Beckman Coulter employee.

Quality means:

- Always striving for excellence
- Meeting or exceeding our customers' expectations
- Complying with regulatory requirements
- Maintaining an effective Quality Management System
- Continuously improving

Quality leadership is essential to industry leadership.

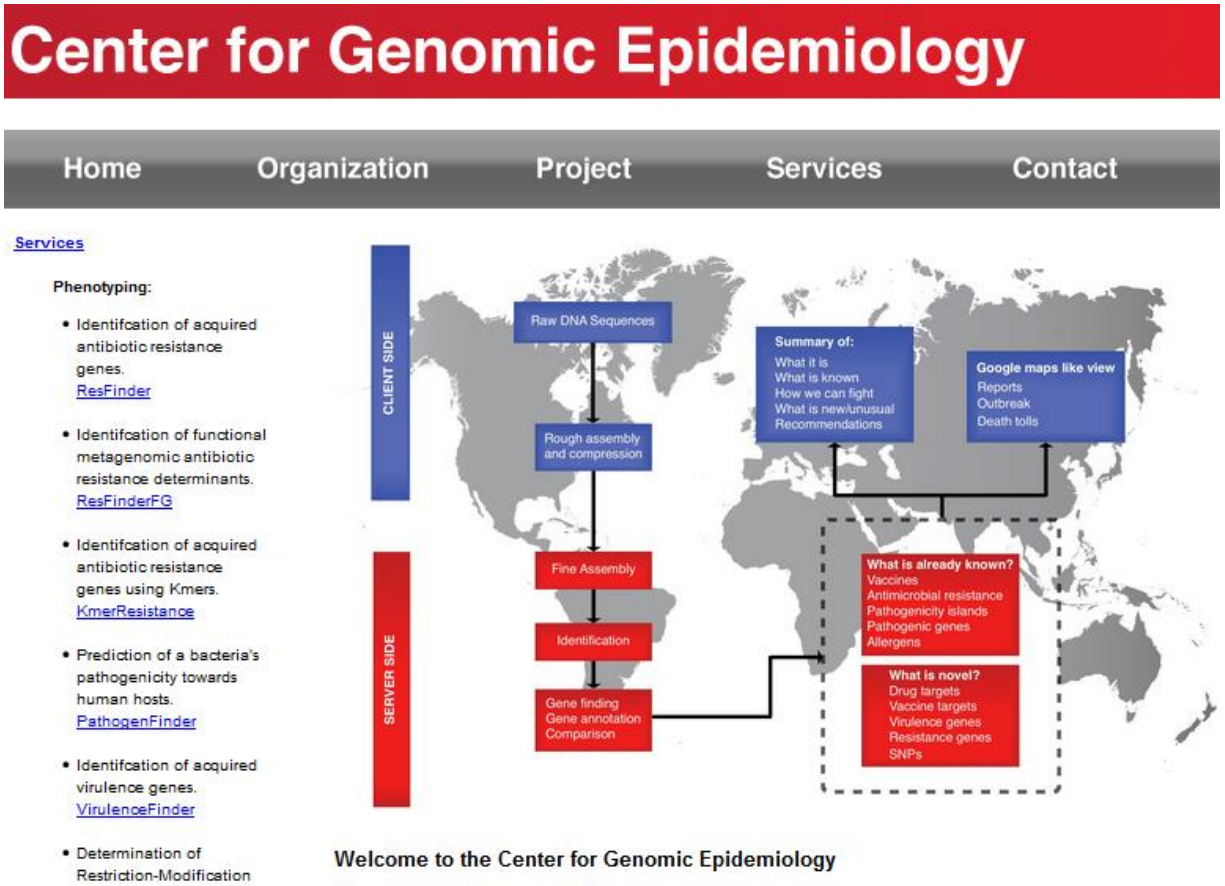
CAP and CLIA Certified, GLP/GMP Compliant

Data Analysis: Outsourcing

Data Analysis: Public servers

- Species identification
- *de novo* assembly tools
- VirulenceFinder
- SerotypeFinder
- ResFinder
- MLST
- SNPs tree and newly deleveloped NGS-driven philogenetic tools

FREE, USER-FRIENDLY WEB INTERFACE



Closed Public server

Data Analysis: Public servers

Unipro UGENE

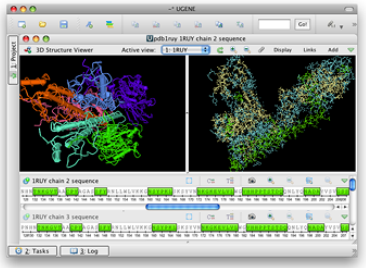
Home | Documentation | Learn | Community | Downloads | Support | Contacts | Donate

Unipro UGENE v. 36

September, 2020

UGENE is free open-source cross-platform bioinformatics software
It works perfectly on Windows, macOS and Linux and requires only a few clicks to install

Download PDF | Download UGENE



News

September 21st, 2020
UGENE 36 has been released

June 17th, 2019
UGENE's 11th anniversary

Podcast

Ep56: What's new in UGENE 33, 34 and 35, part 1.

Ep54: What's new in UGENE 1.31 & 1.32 – Part 1. Metagenomic workflows

Cite Us

Okonechnikov K, Golosova O, Fursov M, the UGENE team. **Unipro UGENE: a unified bioinformatics toolkit**. *Bioinformatics* 2012 28: 1166-1167. doi:10.1093/bioinformatics/bts091

Golosova O, Henderson R, Vaskin Y, Gabrielian A, Grekhov G, Nagarajan V, Oler AJ, Quiñones M, Hurt D, Fursov M, Huyen Y. **Unipro UGENE NGS pipelines and components for variant calling, RNA-seq and ChIP-seq data analyses.** *PeerJ 2014* 2:e644. doi:10.7717/peerj.644

Support and Services

Feel free to contact us if you need technical support.

Free Resources

- Browse the documentation
- Watch instructional videos
- Discuss UGENE on the public forum
- Report issues to the bug-tracking system
- Write us— any feedback is appreciated

Commercial Support

We provide the following fee-based services:

- Fix a defect or add a new feature into UGENE at the

Board

>Feature Requests: Donate a little bit

>Feature Requests: Re: DNA-Star file conversion

Issues

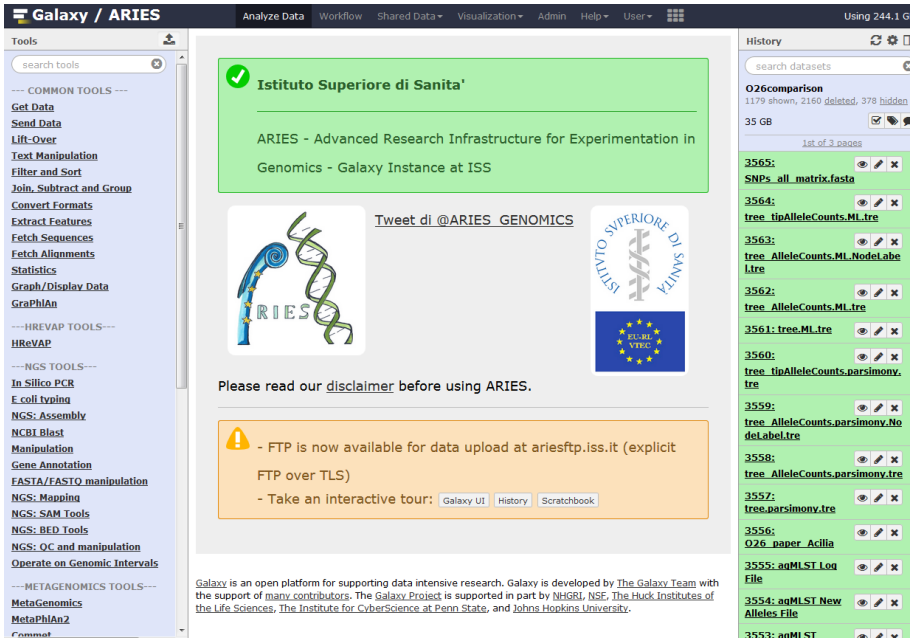
Tested: UGENE crashes with the attached workflow output [UGENE-6757]

Open: Some buttons on the "Search in alignment" tab don't work until you change focus [UGENE-6880]

Data Analysis: Public servers

Galaxy

- *de novo* assembly tools
- BLAST search of genes of interest
- Alignment of sequences, typing tools, production of dendrograms

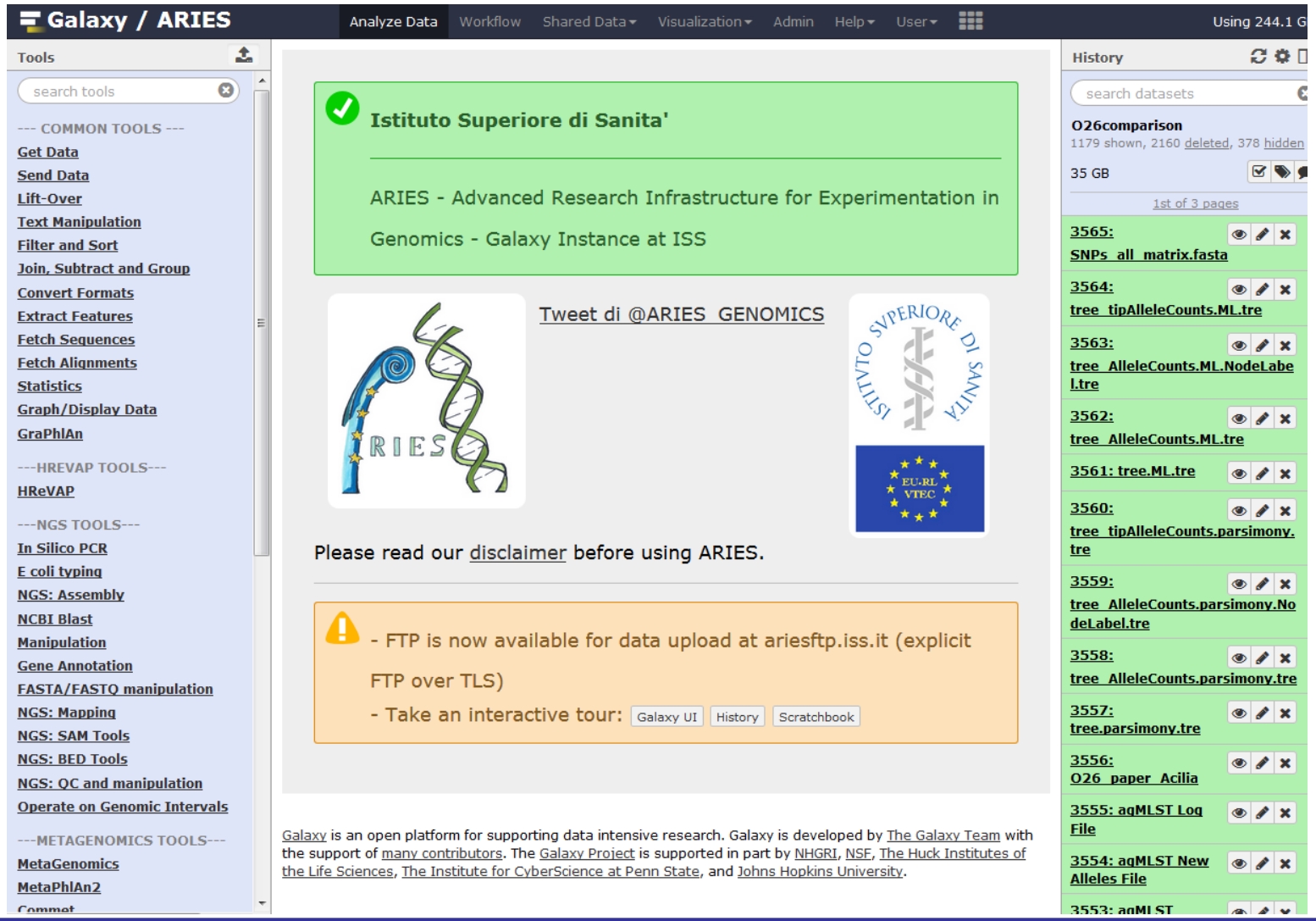


ARIES - ISS

OPEN SOURCE, USER-FRIENDLY WEB INTERFACE, OPEN FOR INTRODUCTION OF CUSTUMIZED TOOLS, ELECTION PLATFORM FOR DEVELOPING AND SHARING OF NEW TOOLS

Open Public server

ARIES: A Galaxy-based workspace for intensive data analyses



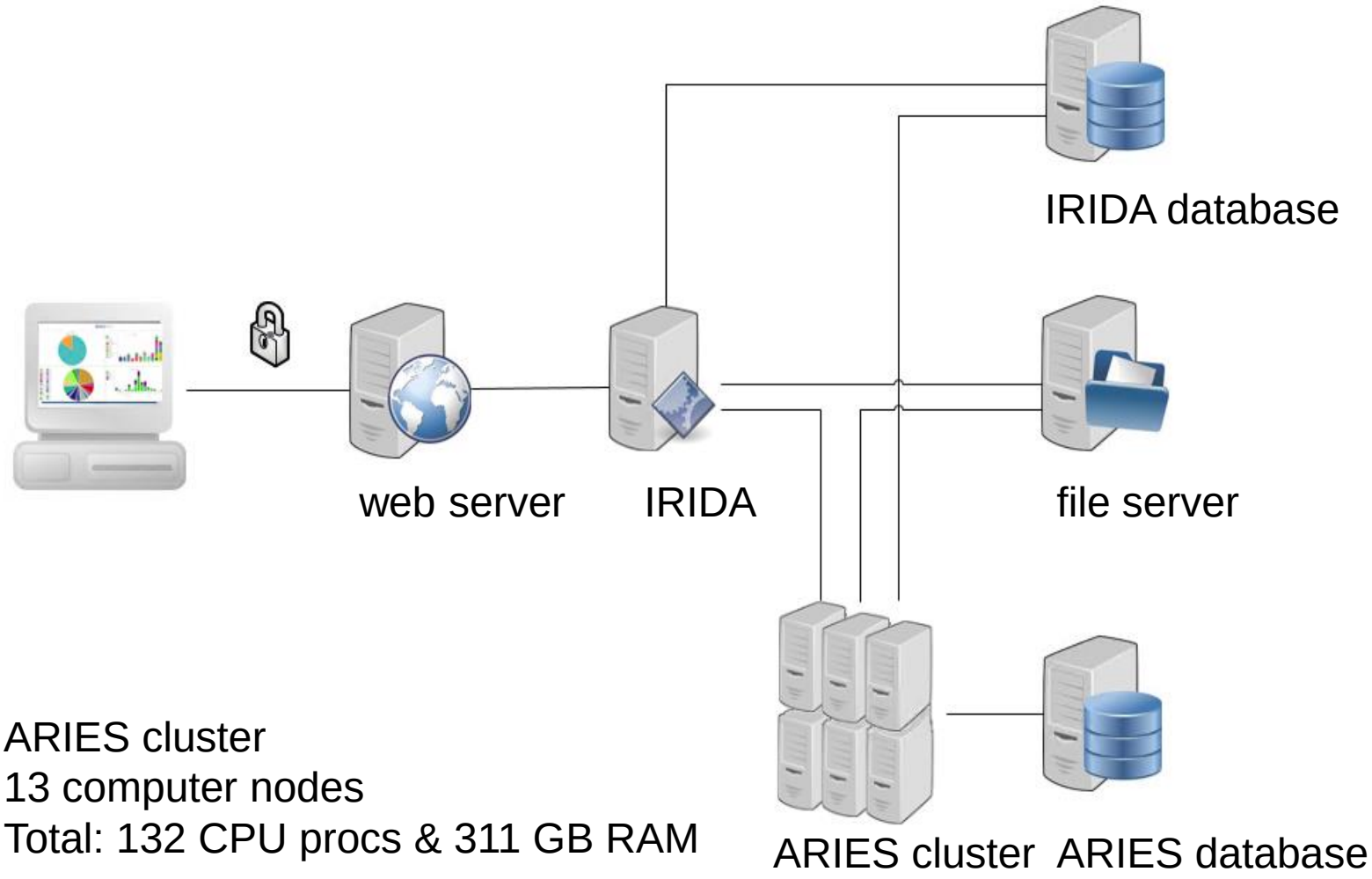
ARIES geographic spread

183 total Users

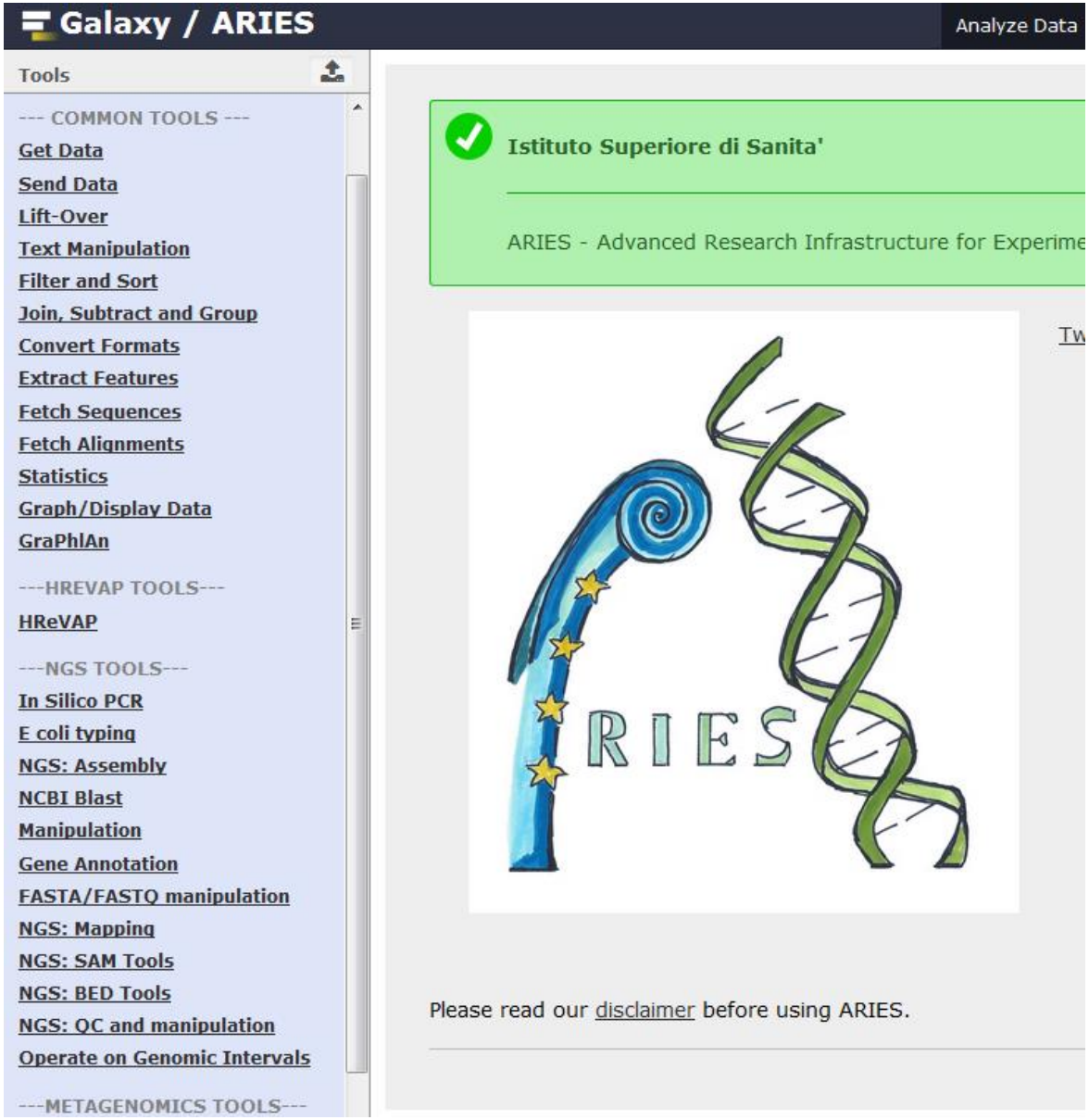
156 European Users (25 NRLs)

27 Users from outside EU

ARIES Under the hood



ARIES cluster
13 computer nodes
Total: 132 CPU procs & 311 GB RAM



QC

Assembly de novo

Mapping

Microbial genome annotation

NCBI Databases

Databases shared with CGE/SSIs

Custom Databases

***E. coli* typing:**

Virulotyping
Serotyping
Clermont phylogrouping
HReVAP
MLST
ksnp3 for ref-free wgSNPs
cgMLST