Advanced Research Infrastructure for Experimentation in Genomics (ARIES): A concept of a bioinformatics framework for the analysis of genomic data from zoonotic agents

Valeria Michelacci[1], Arnold Knijn[1], Massimiliano Orsini[2], Stefano Morabito[1]

[1]*Istituto Superiore di Sanità, Rome , Italy*

[2]*Istituto Zooprofilattico Sperimentale dell'Abruzzo e Molise, Teramo, Italy*

Contact: valeria.michelacci@iss.it; aries@iss.it

Introduction

The European Reference Laboratory for *E. coli*, and the IT services of the Istituto Superiore di Sanità in Rome, have developed a web platform for the analysis of genomic data in the field of public health and food safety, with the aim of deploying a comprehensive bioinformatics approach to the study of food-borne zoonoses and infectious diseases at the human and animal interface. The web portal is termed ARIES (Advanced Research Infrastructure for Experimentation in genomicS) and is powered by the Galaxy framework.

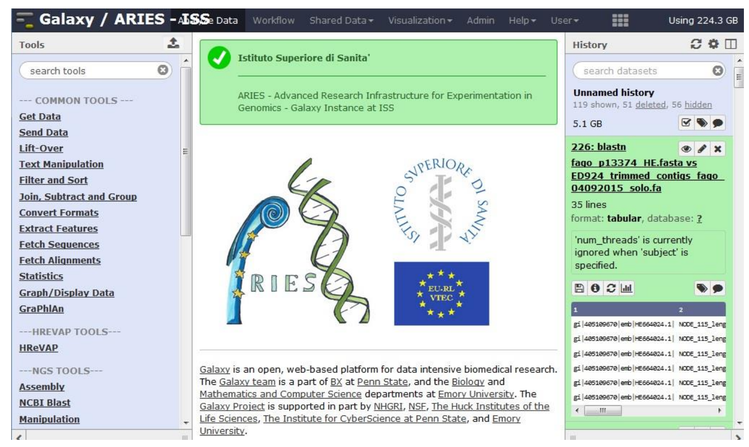
Aims

- Supporting scientists in genomic investigation of pathogens, by providing at the same time a unique pre-configured genomic analyses environment characterised by reproducibility, transparency and easy data sharing as well as an open platform for implementation of new bioinformatics resources.
- Development of an Information System for the collection of genomic and epidemiological data to enable the Next Generation Sequencing (NGS)-based surveillance of infectious epidemics, foodborne outbreaks and diseases at the animal-human interface.
- Development of analytical pipelines enabling harmonized, real time multi-genome comparisons, to improve the detection of clusters of cases of infections and allowing the global bio-tracing of pathogens.
- Development of metagenomics models for the culture-independent detection and typing of pathogens and the study of their interactions with the microbiota in human and animal samples and in the vehicles of infections.

Concept

NGS Data production

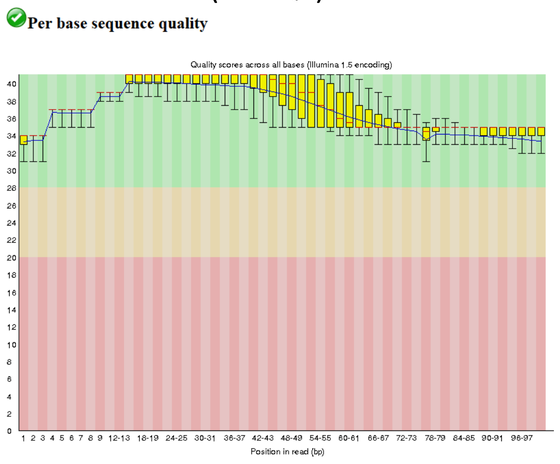
NGS Data Analysis



1. Choose the tool
2. Choose the data and set the parameters
3. Get the result

ARIES contains all the tools for basic and specialized genomics:

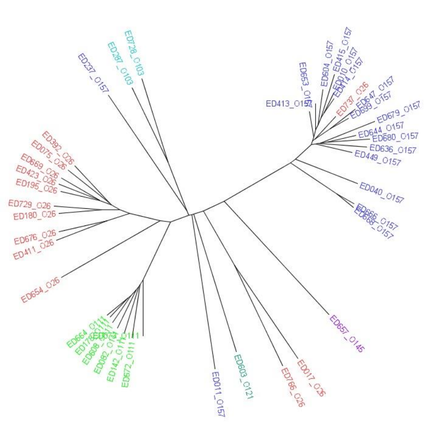
Quality check of the raw data
(FastQC)



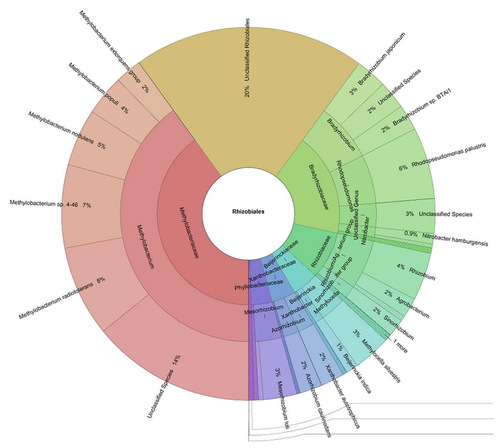
De novo assembly
(SPADES, Edena, MIRA, Velvet)

```
>NODE_215_length_1206_cov_29.0156_ID_429
CGCACCGGAAAGGTGGCGGGGGATTCACAACGATAATATCTGTCTGAAATTATTTAGGAT
GCAGGTATCTTCTGATAATCATACATCCCGAATTTTGTTCTTAATAGTTGAGTAGTGGTA
TTTGGTTTCCCTGAGAGAATATAAAGGTGAATCATGCTGCATTGAGTTTGATTCAATTTG
TACTGAGACAAGTATCTTCATGATTATAATCAAAAGTTGGGTGGATTGGTCTGGCGCTCA
TCATTTTTTCGAATAAACTCACTAATTATTGCCCCCCCCCAAAGGAGATGATTTCATGTAT
AATTGTGGTGAATCATATTCAGAATACGACATATCATTATCTGACGATGAAAACTATGAC
GATGACTATTATCGTGAACGCCCGTGTCCGGGAGAAGTTTTCGGCTATGATAATACTGAG
GATAATGATGAGGCGGCATCTGATTATGCATTAAAAATAAATGAAAACTCTGATTAATAC
GGGGGTAGCTGATAATAAGACAATCAATGACAGATATTAATGAATAAACATAATCAGGCG
CTCAATACTTGGTTTAATCTGGGCGCCTGATTAAAATCTGAGGAAACATATAATAATTTT
CATCAAAATAGTATTTATGTACTAACTAAGCTGGATTTGAACCTCTATTTCTGGGTATTT
GTTATCACTTAATCCATTACTGCTCCTCCTTCAGCATATGTCGTATCCGTTGAAGGATTT
ACGCATATCGGTAACGTCGGCAACGAGCCAGATGCGGGTACCTGACGGGAGTGAGATCAT
CGTCCCCTCCCGGTCAGTTCACGGATCAACAGCGTGAGCAGCTCTGGCGATGGATTTTCC
AGCGTCATGTTACCGTGACGGAACTCCACCTTGCAGGAGCCGGCACTGACACTGGTCTGA
GTGGAAGTGGATAAAGACGGCGCAATGGCCGCCACCGGTTCTTTCTGCTCATCCGGCGTT
ATTTCTACAGGTAATAATTCAACGCCAGTGTCAGAAGAGGTCGTTACCTGAAGACGCCGT
GATATACGCCCTTCGTTCTGCCAGAGCCTGAGCCATTTGAACAGGAGGTTATCATTGATA
TCGTGTTCCCGGGCAATACGGGCAACAGAGGCTCCTGGTTGTGATGCCAGTTCCACCATA
CGAAGTTTGAATTCATTCGAATAGTTTTTACGAGGTTCTTTTCGCCAGTCCTGTAATTCC
ATACTT
>NODE_212_length_1234_cov_52.9584_ID_423
TGTACAGTCTGGTCGCCGCCGGCATAAACCACAGTGGTATTTGCCGTTCCTTCAGTTCTC
ACAATCTGGCGACCGTTTTTTATTGATGGTTGTGCGTACGGCATCCCCGCGAACAAACTGC
CCGGTATCACCGTTTTCTGCATCCGGTCCCCCTTCCGCCCCGGTATTAACAACGGTATCC
```

Whole genome SNPs comparisons
(Ksnp3)



Metagenomics tools
(Krona, Commet)



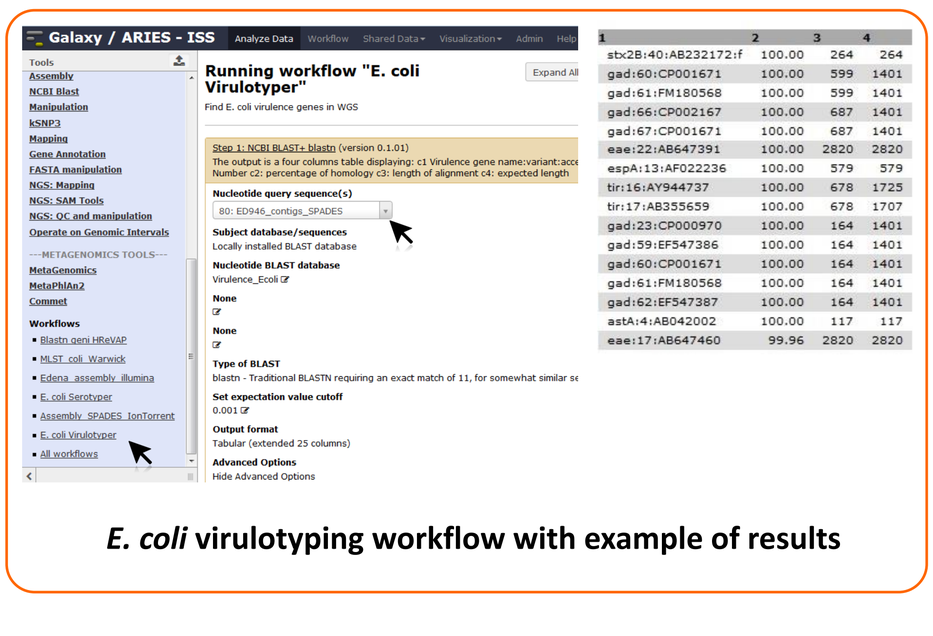
Other tools:

- **Mapping** (BWA, Bowtie2)
- **Gene annotation** (Prokka, Glimmer)
- **Alignment** (BLAST+ suite)

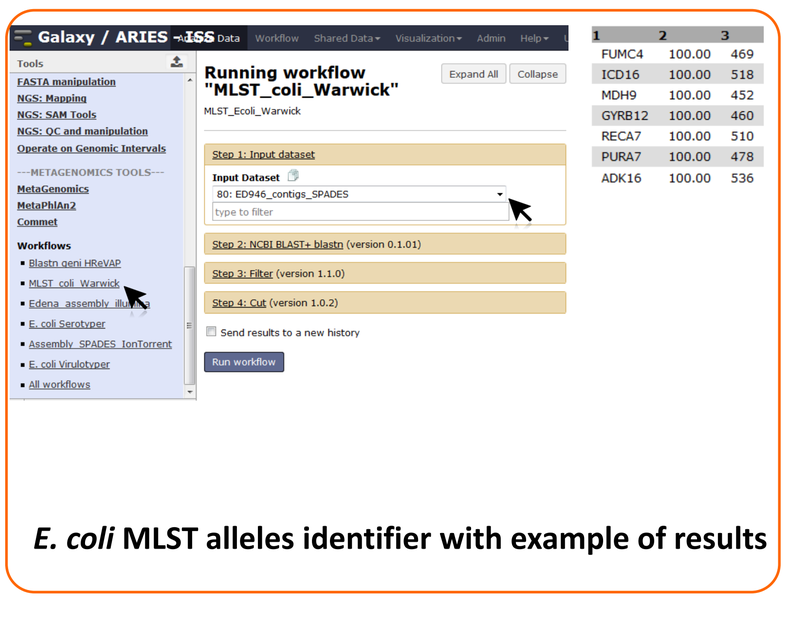
and counting…

At present ARIES features workflows specifically developed for the characterization of pathogenic *E. coli*, such as:

- identification of the presence of virulence genes
- *in silico* serotyping
- MLST typing



***E. coli* virulotyping workflow with example of results**



***E. coli* MLST alleles identifier with example of results**

Possibility to easily develop and share new analytic pipelines

e.g. The pipeline for the **High Resolution Virulence Allelic Profiling of Shiga-toxin producing *E. coli*** was developed and is currently running on **ARIES** (see abstract from the corresponding presentation at the High throughput PCR pre-symposium workshop, October 7th)

CONCLUSIONS

- ARIES is currently optimized for the analysis of *E. coli* genomes, but the tools implemented can be adapted to the analysis of other pathogens upon request, by integrating reference databases
- The use of a common and easy-to-use platform will help the formation of a network of laboratories for the exchange of developed pipelines and results, driving towards the harmonization of NGS data analysis.
- The development of an integrated system for the production, collection and analysis of genomic data on agents of foodborne zoonoses and infectious diseases will be the ground for an enhanced real time monitoring and surveillance.

The platform is planned to open for external access in early 2016. Contact us at aries@iss.it to join us!

REFERENCES

- Goecks, J, Nekrutenko, A, Taylor, J and The Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. 2010 Aug 25;11(8):R86.
- Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. "Galaxy: a web-based genome analysis tool for experimentalists". Current Protocols in Molecular Biology. 2010 Jan; Chapter 19:Unit 19.10.1-21.
- Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. "Galaxy: a platform for interactive large-scale genome analysis." Genome Research. 2005 Oct; 15(10):1451-5.