



D2.4 List of bioinformatic analyses pipelines

WP2 Implementation of whole genome sequencing protocols

Author(s): Daniel Polzer, Daza Prieto Beatriz



Co-funded by
the European Union

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them.

Copyright message

© HERA 2 Consortium, 2024

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorized provided the source is acknowledged.

D2.4 List of bioinformatic analyses pipelines

Document Information

Grant Agreement Number	101102440	Acronym	HERA 2	
Full Title	Consolidation of WGS/RT-PCR and infrastructure processes in surveillance and outbreak investigation activities			
Call	EU4H-2022-DGA-MS-IBA-1			
Topic	EU4H-2022-DGA-MS-IBA-01-02			
Type of Action	EU4H-PJG			
Start Date	01.10.2022	Duration (in months)		48
HaDEA Project Officer	Diana Burghardt			
Task	T2.4 Harmonization of bioinformatic analysis pipelines			
Work Package	2			
Date of Delivery	Contractual	M42/March 2026	Actual	M42/March 2026
Nature	R- Document report	Dissemination Level		PU - Public
Lead Beneficiary	AGES			
Lead Author	Daniel Polzer	Organization		AGES
Other authors	Daza Prieto Beatriz			
Reviewer(s)	Jennifer Rao, Judit Henzko, Grigoris Spanakos			

Document History

Version	Issue Date	Stage	Changes	Contributor
0.1	2.2.26	First Draft		Daniel Polzer

D2.4 List of bioinformatic analyses pipelines

0.2	16.2		Refined and send for feedback	
1	30.3	Final Version	All Feedback incorporated	All Partners

TABLE OF CONTENTS

EXECUTIVE SUMMARY	5
BACKGROUND INFORMATION.....	6
LIST OF PIPELINES	7
CONCLUSIONS	9

EXECUTIVE SUMMARY

Bioinformatic pipelines are automated workflows that process biological data through a structured and defined computational step. These steps may include quality control, data cleaning, alignment or assembly, variant detection, statistical analysis, and visualization.

Pipelines integrate multiple software tools and reference datasets into a coordinated end-to-end process that transforms raw experimental data into interpretable scientific results. The organization of work steps into pipelines enhances reproducibility as well as scalability.

As part of this project each partner developed new pipelines and enhanced existing ones specifically tailored to their individual use cases.

BACKGROUND INFORMATION

Before the start of this project, each partner organization relied on its own bioinformatic pipelines to process sequencing data. These pipelines had been developed independently, using different parameter settings, reference databases, and quality-control criteria. Additionally, commercially available pipelines are used as well. As a result, the same biological dataset processed by different partners could yield inconsistent outputs.

These discrepancies often made it difficult to reach a consensus interpretation across partners. The lack of harmonization not only introduced uncertainty into collaborative analyses but also hindered the comparability, reproducibility, and scalability of results. This fragmentation represented a clear methodological gap. The challenge was to standardize or improve the existing pipelines. The project aimed to bridge this gap by developing new pipelines and enhancing existing ones, ensuring that all partners could rely on more consistent workflows.

Below we give an overview of the bioinformatic pipelines currently used by the consortium members. Some of these pipelines are publicly available, others are commercial, and some are in-house methods. Further we differentiate between “generalist” pipelines which work either pathogen agnostic or for a wide range of pathogens and “pathogen specific” pipelines.

LIST OF PIPELINES

Pipeline Name	Input	Public/i n-house	Purpose	Country	Type (Specific/ Generalistic)
SeqSphere	fastq/fastq	public	cgMLST of various bacterial species	Austria Croatia	Generalistic
MTBseq/TB-Profiler	fastq	public	Subtyping of Mycobacteria, resistance gene detection	Austria Croatia	specific
Corona-Pipeline	fastq	In-house	Subtyping of corona virus	Austria Croatia	specific
AMR-DB Pipeline	fastq/fastq	public	Typing; annotation; plasmid detection	Austria	generalistic
Norovirus pipeline	fastq	In-house	Subtyping of norovirus	Austria	specific
Food metabarcoding Pipeline	fastq	public	Detecting food compositions	Austria	generalistic
Salmonella outbreak detection pipeline	cgMLST distance matrix	In-house	Detection of salmonella outbreak clusters	Austria Croatia	specific
SNPy Pipeline bacterial	fastq/fastq	public	On-demand SNP analysis	Austria Croatia	generalistic
Bacterial subtyping workflow	fasta	In-house	Various bacterial subtyping tools (e.g VirulenceFinder)	Austria Croatia	generalistic
Kraken	Fastq/fastq	public	Metagenomic classification	Austria Croatia	generalistic
Influenza pipeline	fastq	In-house	Influenza subtyping	Austria Croatia	specific
RSV	fastq	In-house	RSV subtyping	Croatia	specific

D2.4 List of bioinformatic analyses pipelines

Measles	fastq	In-house	Mesles subtyping	Croatia	specific
Legionella sp/Legsta	fasta	public	Legionella subtyping	Croatia	specific
Klebsiella sp/ Kleborate	fasta	public	Klebsiella pneumoniae subtyping	Croatia	specific
Neisseria gonorrhoeae/ ngmaster	fasta	public	Neisseria gonorrhoeae subtyping	Croatia	specific
Enterovirus	Fasta	Public	EV subtyping	Croatia	specific
Routine analysis workflow for bacteria	fastq	In-house	Short + long reads QC, assembly, typing, AMR+virulence+plasmid detection, annotation	Greece	generalistic
SNP analysis workflow	fastq/fasta	In-house	SNP-based analysis of sample collections	Greece	generalistic
cgMLST analysis workflow	fasta	In-house	cgMLST-based analysis of sample collections	Greece	generalistic
Torrent Server running the Torrent Suite Software (TSS)	fasta	public	SARS-CoV-2 subtyping	Greece	specific
DRAGEN Automated Pipeline	fastq/fasta	public	RSV subtyping	Greece	specific
CoVpipe2	fastq	public	SARS-CoV-2 lineage/clade assignment & consensus generation, variant calling	Greece	specific
DRAGEN Microbial Enrichment Plus App	fastq	public	Enrichment/secondary analysis for microbial sequencing (consensus generation)	Greece	generalistic

CONCLUSIONS

Before the start of this project, consortium partners differed considerably in their levels of expertise and technical infrastructure for next-generation sequencing (NGS) and the associated bioinformatic workflows. Through the project's coordinated capacity-building and knowledge-sharing activities, these disparities have been significantly reduced, leading to a more balanced and strengthened analytical capability across the consortium.

Full harmonization of pipelines across all partners was not feasible, primarily due to differences in local IT infrastructures and specific analytical requirements. As a result, distinct pipelines remain in use at individual institutions. Nevertheless, the structured exchange of expertise and best practices within the consortium has enabled all partners to benefit from one another's experience, significantly advancing pipeline development and methodological alignment across the network.