D3.6 Nextflow Workflow, Database scheme and API definition hosted on public code repository

WP3: Enhancement & Consolidation of WGS- & PCR-based methods for public health action

Author(s): Patrick Hyden, Tobias Mösenbacher

Co-funded by
the European Union

Disclaimer

Copyright message

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)**	42
HaDEA Project Officer	Diana Burghardt		
Task	T3.6		
Work Package	3		
Date of Delivery	**Contractual** M18/March 2024	**Actual**	M18/March 2024
Nature	Report	**Dissemination Level**	Public
Lead Beneficiary	AGES		
Lead Author	Patrick Hyden	**Organization**	AGES
Other authors	Tobias Mösenbacher		
Reviewer(s)	Karin Rainer, Julia Steger		

Document History

Version	Issue Date	Stage	Changes	Contributor
0.1	04.05.2023	First template generated in Steering Workshop		AGES, CIPH, EODY, NCPHP
0.2	22.03.2023	First draft shared with consortium partners		AGES
1.0	28.03.2024	Quality checked version after revision		AGES, CIPH, EODY, NCPHP
1.1	28.11.2024	Revised after discussion with PO	Formatting	Julia Steger

Table of Contents

1. EXECUTIVE SUMMARY

This document summarizes the Nextflow Workflow, Database scheme and API definition hosted on public code repository. Its main objective is to provide the software code for bioinformatic resources created in the task 3.5 “WGS-based AMR monitoring via in-silico AMR-gene detection”.

To fulfil this task, a bioinformatic workflow was created along with an appropriate database structure to store all relevant results. This software code is made freely available as two open-source projects which are further developed during the remaining duration of this project

2. BACKGROUND INFORMATION

One aspect of the HERA 2 project focuses on the monitoring of antimicrobial resistance genes. To get a better understanding of the situation in each country, already generated sequencing data of bacterial pathogens may be used with a bioinformatic approach to collect genetic information about resistances.

There are various well established antibiotics resistance gene databases and prediction tools in the bioinformatic community. However, these tools are frequently redundant in their prediction capabilities and there’s a lack of consistency in the report data structure.

Our goal is to facilitate the usage of multiple software tools and collect all predictions in a similarly structured table collection. In addition to resistance genes, information about the potential mobility of the surrounding nucleotides should be generated and stored in the same format.

3. OBJECTIVES AND CONTENTS

The main objective of this task is to provide a basic technology which enables each country to create their own antibiotics resistance gene database. For this purpose, a multi-component solution was developed.

The bioinformatic workflow consists of several well-established tools in the community:

- ResFinder – A tool developed by SSI/DTU for in-silico prediction of antimicrobial resistance genes and point-mutations
- AMRFinderPlus – A tool maintained by NCBI for in-silico prediction of antimicrobial resistance genes and point-mutations
- Bakta – A tool to provide bacterial gene annotation
- ISEScan – A tool to annotate mobile genetic elements

- Mob_suite – A tool to predict mobility of nucleotide sequences
- PlasmidFinder – A tool to identify and type plasmid sequences
- SpeciesFinder – A tool to identify bacterial species

To facilitate the organization of sequences and tabular data, a database scheme is proposed and implemented in a python package. This module serves as python API and provides helper-scripts to easily set up the database (currently mysql, mariadb or sqlite are supported) and import bioinformatic results.
The python package automatically detects previously unseen variants of known resistance genes and allows for replacing the temporarily provided name with correct nomenclature in a future update of the underlying detection database (resfinder_db or amrfinder_db). This facilitates the update procedure and aims at avoiding re-calculation for minor updates.

Both modules are provided as open source software (MIT License) in public repositories on https://github.com/ages-bioinformatics

4. RESOURCES

Python Package a*gesamrdb*

https://github.com/ages-bioinformatics/agesamrdb

Nextflow-Workflow: *wf-antibiotic-resistance*
https://github.com/ages-bioinformatics/wf-antibiotic-resistance