Deliverable 2.5

Workpackage 2

Responsible Partner: 36-INSA

Contributing partners: 2-AGES, 7-SZU, 14-UT, 23-UoS, 25-NUIG, 33-NVI

GENERAL INFORMATION

European Joint Programme full title	**Promoting One Health in Europe through joint actions on foodborne zoonoses, antimicrobial resistance and emerging microbiological hazards**
European Joint Programme acronym	**One Health EJP**
Funding	**This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 773830.**
Grant Agreement	**Grant agreement n° 773830**
Start Date	**01/01/2018**
Duration	**60 Months**

DOCUMENT MANAGEMENT

Title OHEJP deliverable	**D-JRP15-FED-AMR-WP2.5**
WP and task	WP2, JRP15-R2-WP2-T3.1
Leader	36-INSA
Other contributors	2-AGES, 7-SZU, 14-UT, 23-UoS, 25-NUIG, 33-NVI
Due month of the deliverable	M36
Actual submission month	M46
Type *R: Document, report* *DEC: Websites, patent filings, videos, etc.; OTHER*	R **Save date:**: 25-Nov-21
Dissemination level *PU: Public (default)* *CO: confidential, only for members of the consortium (including the Commission Services).*	PU *See updated Grant Agreement*
Dissemination *Author's suggestion to inform the following possible interested parties.*	OHEJP WP 1 ☐ OHEJP WP 2 ☐ OHEJP WP 3 ☒ OHEJP WP 4 ☐ OHEJP WP 5 ☐ OHEJP WP 6 ☐ OHEJP WP 7 ☐ Project Management Team ☒ Communication Team ☒ Scientific Steering Board ☐ National Stakeholders/Program Owners Committee ☐ EFSA ☐ ECDC ☐ EEA ☐ EMA ☐ FAO ☐ WHO ☐ OIE ☐ Other international stakeholder(s): …… Social Media: ..... **Other recipient(s):** .....

TABLE OF CONTENTS

D-JRP15-FED-AMR-WP2.5

Shotgun Sequencing: ARG Diversity in Tested Environmental Compartments (T2.3.1)

I. Introduction

1. Management of the Sub-Task: JRP15-R2-WP2-T3.1

This deliverable corresponds with Sub-Task: JRP15-R2-WP2-T3.1, entitled **Shotgun sequencing and bioinformatic analyses of AMR genes and MGEs**, which only began in Y4, with the following Leaders and participants:
Sub-Task Leader: Alexandre De Menezes
Deputy Sub-Task Leader: Werner Ruppitsch
Sub-Task Participants: 2-AGES, 7-SZU, 14-UT, 23-UoS, 25-NUIG, 33-NVI, 36-INSA

2. Aim

In this sub-task we performed the analysis of DNA sequence data from the shotgun sequencing of soil, water, manure and faeces samples from the FED-AMR consortium. The sequencing was carried out using Illumina sequencing technologies and it was applicable to samples sequenced with gene enrichment. Both NGS and gene enrichment were carried out by an external Company using ARESdb (Ferreira et al., 2020).

3. Initial plan of the task

To perform the shot gun sequencing to obtain an overview over the variety of ARGs present in the diverse environmental compartments will be started after the DNA of all samples was purified. Certain reference samples as decided by the Scientific Board will be analysed according to the cheapest offer obtained. Bioinformatics and statistical analysis will be performed at the Irish Centre for High-End Computing (ICHEC) at 25-NUIG using metaSPAdes, SNPFinder, MEGAN and similar appropriate software. Resistomes will be determined using the Resfinder reference database. Normalized fragments resistance genes per kilobase reference per million bacterial fragments (RPKM) will be calculated for quantitative evaluations. This subtask is based upon results from and methodology applied by the EFFORT project. The remaining partners supply samples. The results will be forwarded to WP5 and WP6.

4. Modifications in the task

As indicated in **D-JRP15-FED-AMR-WP2.2 (***Preliminary data collection on ARG prevalence and ARG background load in the compartments analysed so far (T2.4)***)**, a deviation from the original plan was preferred by the consortium, which involved testing a higher number of samples by shotgun sequencing plus gene enrichment and 16S amplicon sequencing. Applying the initial plan described in 3. would involve the in-house development of the probes. With the increased number of samples we decided to sequence, it was more cost-effective to outsource this analysis, which is also sufficient for quantitative comparisons across samples.

II. Description of deliverable:

After the DNA extraction of each sample obtained by each of the partners, the DNA samples were sent to the company Ares Genetics (Vienna, Austria) in orderto obtain harmonized results.

1. Advanced AMR Marker Profiling by Next Generation Sequencing

This comprised the following steps:

Next-Generation Sequencing[1]	NGS target enrichment
Quality Control	QC of NGS raw data
	QC of *de novo* assemblies
Resistance Profiling[2]	Comprehensive resistance gene analysis
	NGS raw data

[1] Target enriched sequencing of AMR markers from client provided DNA.
[2] Based on proprietary ARESdb

1.1. Advanced AMR Marker Profiling

In Ares Genetics (Fig. 1), after assessment of DNA quality, DNA samples were randomly fragmented using enzymatic shearing. Sequencing adapters were ligated to both ends of the DNA fragments and amplified by PCR, prior to pre capture library QC. Following, libraries were enriched with a capture-based NGS panel designed based on Ares Genetics' proprietary reference database ARESdb. Ready-to-sequence libraries were paired-end sequenced on an Illumina platform. All samples were demultiplexed and Illumina adaptor residuals trimmed.

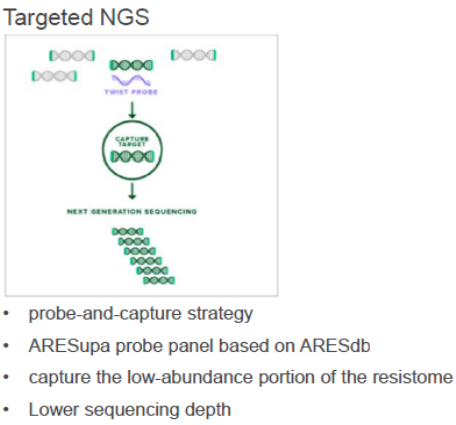

Fig. 1. Targeted sequencing.

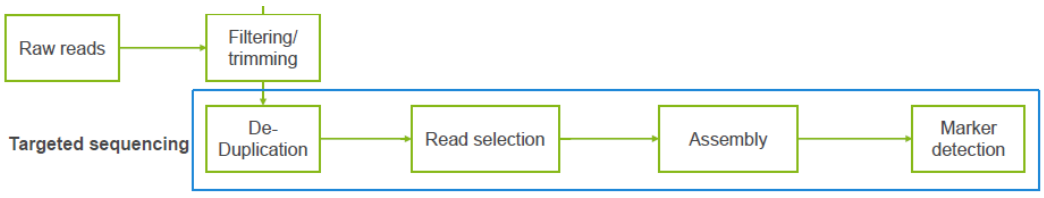

Fig. 2. Bioinformatics workflow.

A no template control (NTC) in which the template DNA is replaced by molecular grade water was included to assess cross-contamination throughout library preparation. Library QC after library preparation showed negative results in the fragment analysis as well as no concentration by fluorometric measurement.

1.2. Advanced AMR Marker Profiling: Raw Reads Quality Control

Raw reads were checked using FastQC, quality filtered/trimmed with Trimmomatic and de-duplicated with Fastuniq.

2. Advanced AMR Marker Detection

AMR markers were detected via sequence alignment of captured and assembled reads (targets) against ARESdb (Ferreira et al., 2020). All detected AMR markers are listed per sample in a supplementary file*.xlsx*. 'Identity' describes the percentage of how many nucleotides in the target sequence are identical with the reference sequence. 'Contig Length' describes the contig's length in base pairs. If the 'Sequence type' is 'Protein', 'Alignment Length', and 'Marker Length' describe the respective lengths in amino acids. 'Coverage' describes the percentage of marker reference sequence which is covered by the target sequence. 'RPKM' describes the reads per kilobase million (RPKM), a contig-length weighted read depth measure. Only markers with a coverage ≥50% and identity ≥90% were included in the final reports.

3. Data Availability

3.1. NGS Raw Data, Assemblies and Supplementary Data

Sequencing raw data and all supplementary files were accessed via a SFTP server, which could be accessed by FileZilla and Cyberduck.

3.2. Sample and Data Storage

Digital data generated for the analysis of the samples were stored by each partners until a common storage of all raw data be defined by the Coordination, WPL, WPDL, as well as WPSubTaskL, WPDSubTaskL and the other WP2 partners (indeed the Sub-Task Participants). Data will be uploaded to public sequence databases before publication in scientific journals.

Two main data types were obtained: excel files containing the list of the detected ARGs in each sample and the paired-end raw reads in compressed FASTQ format used for advanced AMR profiling for each sample (=gene enrichment).

4. Data analysis

The data obtained (through excel files) using the methodology described above is already very consistent to the aim of the task. However, to be able to answer the several questions of the project we need to compare data from all countries, compartments, collection date (month/season), exDNA vs. total DNA, and several other variables (e.g. comparison with isolate data; comparison with 16S rRNA gene sequencing data, comparison with WP4 data: elements, ATBs, herbicides; mobile genetic elements: transposons, IS, plasmids, phages, and others), namely using descriptive statistics,**which will take some months and so we decided to complete this deliverable now, as the data already exists**.

Some columns within the excel files were common to all WP2 partners and are designated as "C" in the example below. These included the marker name, the marker family or class, the marker drug class, the marker type, the RPKM, the marker length, identity and coverage. Variable columns were those sample-specific information like external sample ID, FED-AMR reference ID or DNA type and are designated as "V".Additional variable columns to be included by each partner upon the receipt of their results included "compartment, "collection date",

5. Extended sequence-based analysis

If the consortium still decides to perform further analysis on the raw sequence data, this will occur by using typing tools, e.g. Phylogeny (cgMLST/SNP analysis, software…), ARGs (Resfinder, CARD, PlasmidFinder and other CGE Server tools). The rationale for this re-analyis could be justified by the fact that the gene-enrichment sequence data files may contain many non-target (non-AMR) sequences which may have valuable information about microbial metabolism in the different compartments analysed. Although the non-AMR genes are not a focus of the study, their exploration may lead to further scientific outputs from the FED-AMR project. To obtain this information, it will be necesssary to perform additional bioinformatic analysis to characterise non-AMR genes and metabolic pathways. To do this, the sequence data will be uploaded to the MGnify database (https://www.ebi.ac.uk/metagenomics/), which is maintained by the European Molecular Biological Laboratory (EMBL-EBI), which provides powerful sequence annotation tools. The MGnify analyses will be complemented by statistical analysis of the annotated datasets. **However, as this is an extra analysis it will only performed if we have time.**

6. Linkage to other WPs

The results obtained in this task will be forwarded to WP5 and WP6, as soon as WP2 has all concluded all statistic analysis and comparisons, **planned in the project**.

7. Some brief conclusion

The enrichment performed using the ARESupa panel was essential, as without it metagenomics should have far fewer AMR genes.

This was the best option found by the consortium. Performing shotgun metagenomics would be more expensive and, on the other hand, with this methodology there is also a previous analysis of the results carried out by the company Ares Genetics, which has proved to be an asset considering the huge amount of data that the project produced.

8. Reference

Ferreira I, Beisken S, Lueftinger L, Weinmaier T, Klein M, Bacher J, Patel R, von Haeseler A, Posch AE. Species identification and antibiotic resistance prediction by analysis of whole-genome sequence data by use of ARESdb: an analysis of isolates from the Unyvero lower respiratory tract infection trial. J Clin Microbiol, 2020. 58:e00273- 20. https://doi.org/10.1128/JCM.00273-20

9. Annexes

Annex 1: Ares Genetics FED-AMR Sample Shipping

Annex 2: Example of data