Project report:

Project E20: FED-AMR-Estonia – AMR (antimicrobial resistance) markers over ecosystem boundaries along the food/feed chain.

Team: Krõõt Arbo

Project repository: https://github.com/kroota/AMR_Est

Business understanding

Background

Antimicrobial resistance is one of the most serious public health threats facing humanity. Misuse and overuse of antimicrobials in humans and animals are the main drivers in the development of drug-resistant pathogens. The role of the environment as an important source of resistance has been increasingly recognized, but our understanding of its contribution is still limited. About 80% of sold antibiotics are used in agriculture. Majority of antibiotics are not completely metabolized in the bodies of human and animals, high percentage, sometimes up to 90%, of administrated drugs are coming out with urine and feces. Antibiotic residues are found in wastewater, animal manure, sludge, and biosolids that are frequently used to fertilize agricultural lands. From there antibiotics move to soil and water, where they end up in plants (crops, harvest) what we or animals are eating. Because of that it is important to research the food/feed production chain.

Business goals

- Monitoring the antimicrobial resistance (AMR) genes over different ecosystem boundaries
- Determination of the resistome in the tested environmental compartments
- Extracellular DNA (exDNA) vs totalDNA AMR markers. Are there any unique markers in exDNA or totalDNA samples?
- Is the soil/crops resistome influenced by AMR markers found in the farm?

Inventory of resources

- Dataset 1 (605 kb): Private dataset of target enrichment sequencing data of AMR markers in 103 samples from different compartments along the food/feed chain.
- Krõõt Arbo – data miner
- ASUS UX305CA 8GB 64-bit

Requirements, assumptions, and constraints

- Submission deadline: Monday, 12th of December, at noon (12:00)

Risks and contingencies

Blackout - As in last month I have had two power failures at home, it might be a risk. For that I should keep my phone and computer batteries full so I could use my phone internet while working with the project.

Antimicrobial resistance (AMR) - AMR occurs when bacteria, viruses, fungi and parasites change over time and no longer respond to medicines making infections harder to treat and increasing the risk of disease spread, severe illness and death.

Resistome – All the antibiotic resistance genes in communities of both pathogenic and non-pathogenic bacteria

Rpkm - reads per kilobase million - The measure RPKM was devised as a within-sample normalization method; as such, it is suitable to compare gene expression levels within a single sample, rescaled to correct for both library size and gene length

Costs and benefits

Benefits: Surveillance of AMR genes in agricultural ecosystem

Data-mining goals

Data-mining success criteria

Data understanding

Gathering data

- Dataset 1 (605 kb): Private dataset of target enrichment sequencing data of AMR markers in 103 samples from different compartments along the food/feed chain.

Outline data requirements

- Data is gathered in one year period (covering acrop growing period)

Verify data availability

- Dataset is verified

Define selection criteria

- Dataset is XLSX Worksheet 605kb

Describing data

- 8165 objects/cases and 11 attributes. Attributes are „Customer ID", „NGS ID", „Identity[%]", „Alignment lenght", „Contig lenght", „Marker lenght", „Coverage[%]", „Sequence type", „Marker", „Mechanism", „rpkm".
- Objects are AMR markers in different samples
- 7 different sample compartments – pig feces, manure, wastewater, feed, farmers feces, crop, river(pond), soil with 5 different fertilization practices and two different crops (barley, potato)

Exploring data

- We have some differences in technical parallels (3 parallels), where in one parallel there are some outliers, which must be investigate further

Verifying data quality

- Data quality is verified

Planning your project

- Data preparation (15h)
    - Cleaning data
    - Constructing data
    - Integrating data
    - Formatting data
- Modeling (20h)
- Evaluation (20h)
- Making a poster (5h)