

SIEGA: Integrated System for Genomic Epidemiology of Andalusia

Main author: J. ALBERTO CHAVES SANCHEZ (REGIONAL MINISTRY OF HEALTH AND FAMILY AFFAIRS (CSAFA). REGIONAL GOVERNMENT OF ANDALUSIA.)

Co-authors: Carlota M. Lozano Izquierdo, Carlos S. Casimiro-Soriguer, Ana M. Carriazo, Joaquón Dopazo Blazquez

INTRODUCTION

In June 2014, one of EFSA's recommendations was that '[EFSA and ECDC] should instil a sense of urgency in all partners regarding implementation of WGS [(Whole Genome Sequencing)] for food and public health safety across the EU'.

More than 300 000 cases of food-borne diseases (FBD) are declared each year in the EU. Some of them are clearly linked to the primary food production stage but others are very difficult to identify, such as *Listeria*.

Andalusia is a region in the south of Europe, and is an eminent food-producing region as well as a tourist destination. Therefore, food and environmental safety is of special importance and goes beyond the territory of the region.

Andalusia wants to take a qualitative step in the area of health protection, specifically in the field of food safety, thus facilitating decision-making in the prevention and control of FBD, which in turn, will result in greater protection of public health, through the implementation of the 'SIEGA' initiative, which was devised in 2020. It was created with a 'One Health' approach within the Food Safety Unit of the Directorate-General of Public Health and Pharmacies within the Regional Ministry of Health and Families of Andalusia.

METHODOLOGY

Samples from the Andalusian network of Public, Animal and Plant Health Laboratories, together with samples from designated hospital clinical analysis laboratories are being sequenced.

Sequences are obtained using commercial standards in several sequencing units.

The Clinical Bioinformatics unit of the Andalusian health system is in charge of treating the sequences and creating a genomic database.

An automatic workflow has been developed:

1. Quality control of the readings (fastp, qualimap).

2. Search for contamination in the samples (refseq-masher).
3. Assembly of the genome (Unicycler).
4. Location of genes (PROKKA).
5. Functional gene annotation (PROKKA).
6. Location of possible plasmids (Plasmidfinder).
7. Location of virulence and resistance genes (MEGARES, Resfinder, CARD, VFDB, ARG-ANNOT).
8. Assignment of serotype (LisSero, SeqSero).
9. MLST and cgMLST (python script).
10. Alignment of genomes and obtainment of core genome (parsnp).
11. Phylogeny from the core genome (parsnp, nextstrain).

As a result, a report is generated for each sample and the phylogeny and locations can be viewed on a nextstrain server.

RESULTS

During the first stage, and as a result of the Listeria outbreak in Andalusia in summer 2019, a genomic database of Listeria and Salmonella strains was created. Moreover, E.Coli and Campylobacter have recently been incorporated into the food safety area and Legionella into the environmental health area.

A website was created: <https://www.clinbioinfospa.es/projects/siega/>

16 public health laboratories have analysed and extracted DNA and 8 laboratories have carried out sequencing. More than 600 samples have been collected in the SIEGA database during the first year.

Samples	Clinical origin	Food origin	Animal origin	Total
Salmonella	116	64	70	250
Listeria	371	77	448	
E.Coli	1	6	7	
Campylobacter	8	2	10	
Total	496	149	70	715

Other countries are collaborating with Andalusia:

- For Listeria: Portugal, Austria and Morocco.
- For Salmonella: Finland, Italy, Denmark and Sweden.

This genomic analysis accurately identifies each microorganism by means of the cgMLST and SNP, which makes it possible to precisely locate the samples that may arise from the same outbreak.

DISCUSSION

The first conclusion is that, from the 2019 Listeria outbreak, tools were put in place to improve the management of food and environmental outbreaks and SIEGA was created.

SIEGA has a structure that is growing not only in terms of participating clinical laboratories but also sequencing laboratories. Public entities/agencies have also been incorporated in a way that fosters mutual collaboration, with the exchange of sequences of responsible pathogens involved in outbreaks taking place.

In one year, SIEGA has registered numerous sequences of pathogenic microorganisms that have been decisive in identifying outbreaks inside and outside Andalusia. To date, SIEGA has provided pioneering precision in the exact identification of microorganisms, which, until now, would have been impossible in Andalusia due to its intrinsic characteristics.

Compared with the volume of samples that hospitals have, few samples are sent for sequencing. This fact highlights the need to create a standardised protocol together with Andalusian hospitals.

Although the genomic database of microorganisms of Andalusia could advance research in food safety and environmental health, more resources are required.