# MDST (Multi Dimensional Sequence Typing) An Universal Sequence Typing System

**Val F Lanza[1,2,3,4], Fernando Baquero[1,2,3], Teresa M. Coque[1,2,3]**

*[1]Department of Microbiology, Ramón y Cajal University Hospital, Ramón y Cajal Health Research Institute (IRYCIS), Madrid, Spain, [2]Joint Unit of Antibiotic Resistance and Bacterial Virulence associated with the Spanish National Research Council (CSIC), [3]Network Research Center for Epidemiology and Public Health (CIBER-ESP), Madrid, Spain,[4]National Center of Biotechnology, CSIC, Madrid, Spain*

**Background:** One of the major tasks in microbial epidemiology is the need to type the isolates to allow tracking the outbreaks in different locations. Up to now the most extensively sequence-based tool is Multi Locus Sequence Typing (MLST). MLST have some intrinsic weakness: first, It is a not a correlative scheme (the designation of a particular ST say nothing about its phylogenetic relation). Second, MLST is not always enough precise to define the real clones, and might overestimate the clone concept. Third, two identical MLSTs in strains of the same species can include a set of isolates with different phylogenetic backgrounds. On the other hand, new sequence types are constantly appearing with the expansion of NGS technology. Approaches as cgMLST or wgMLST increase the typing accuracy but they do not provide the possibility of any possible designation to closely phylogenetically related sequence types. **MDST is a new approach that creates a new continuous scheme classification detect phylogenetically relations between clones and is applicable to all bacterial organisms.**

## Hypothesys

*"Given a set of genomes that define a coordinate system we can create a euclidean space where each strain is defined by its spatial coordinates."*

## Aim

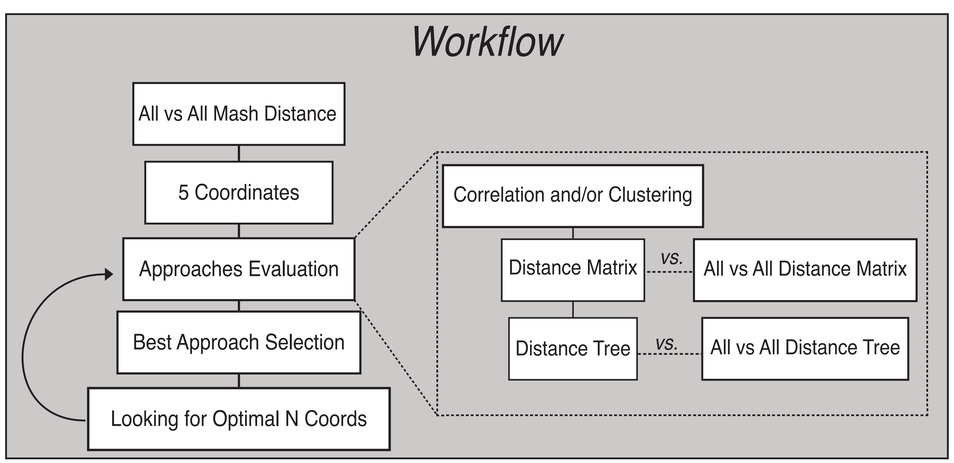
*"Create a system to find the best heuristic set of genomes that define the euclidean space."*

## Material & Methods

Five different approaches were used to find the coordinate genome set and among 2 and 20 coordinates:

- **Correlation** (The set of less correlated genomes)
- Clustering using **Clara**
- Clustering using **Affinity** Propagation
- **Clara** and less **cor**related cluster
- **Affinity** and less **cor**related cluster

The system was tested in 1.890 *Klebsiella pneumoniae*, 3.891 *Escherichia coli* and 501 *Enterococcus faecium* genomes.



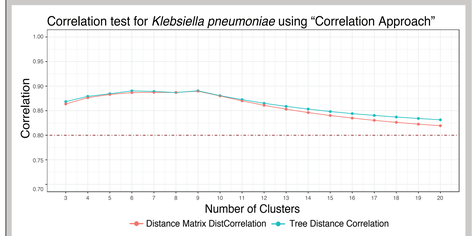

## Results

### *Klebsiella pneumoniae*

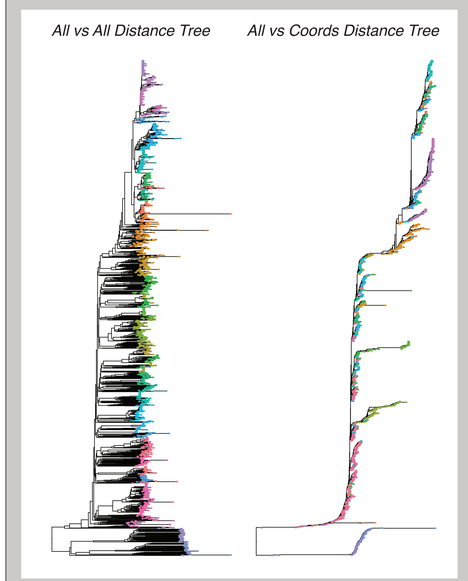
Best result: Correlation Approach and 6 coordinates

**Example**

*Klebsiella pneumoniae* subsp. pneumoniae HS11286
(35,34,33,34,20,34)




*All vs All Distance Tree* *All vs Coords Distance Tree*

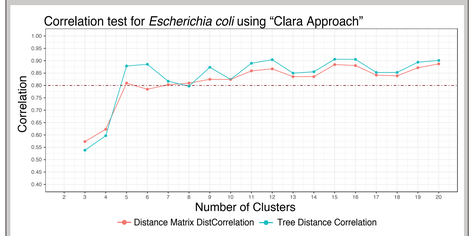


### *Escherichia coli*

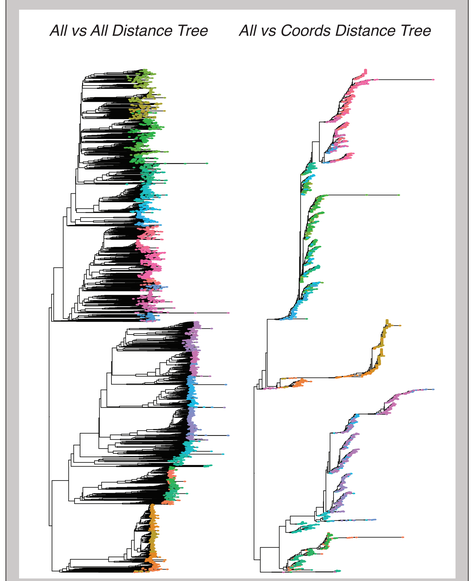
Best result: Clara Approach and 5 coordinates

**Example**

*Escherichia coli* str. K-12 substr. MG1655
(55,25,57,44,64)




*All vs All Distance Tree* *All vs Coords Distance Tree*

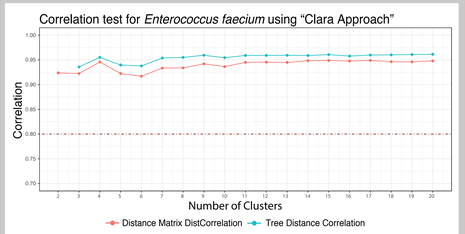


### *Enterococcus faecium*

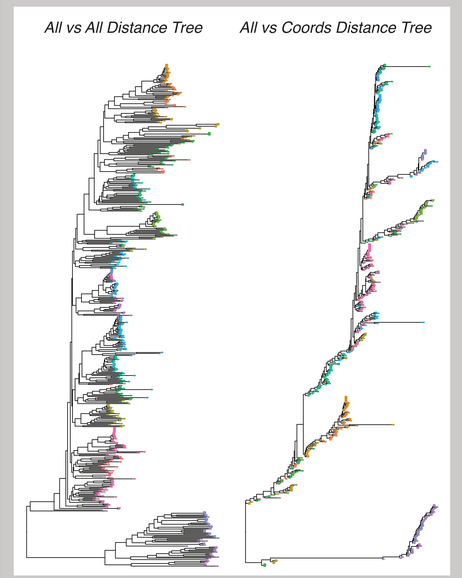
Best result: Clara Approach and 4 coordinates

**Example**

*Enterococcus faecium* DO
(28,40,27,25)




*All vs All Distance Tree* *All vs Coords Distance Tree*



## Conclusions

- MDST is a new typing system that take advantage of the euclidean properties to classified and organized the genomes.
- MDST is valid to different species offering similar resolution to all of them.
- MDST use the whole genome information to type the genomes.
- The euclidean properties establish a framework in where the strains could be related by a simple distance as well as an hypergeometric area would define an outbreak.

**For more information please visit:**
**https://github.com/valflanza/PosterECCMID2017**

Funded by PI15/01307