MLST approaches:

7 hosekeeping genes

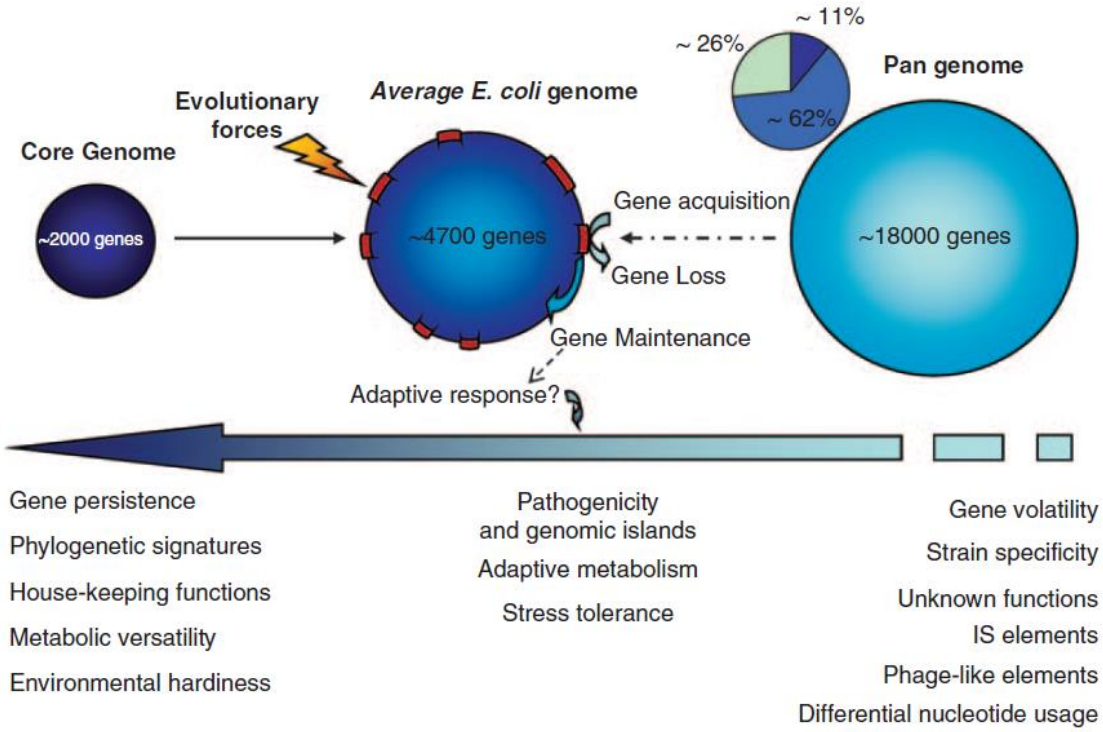
whole-genome

core-genome

Valeria Michelacci

Bioinformatics training,
June 2018

The *E. coli* pangenome



Van Elsas J.D. et al., 2011

Pangenome **Whole genome** **Core genome** **Accessory genome**

Housekeeping

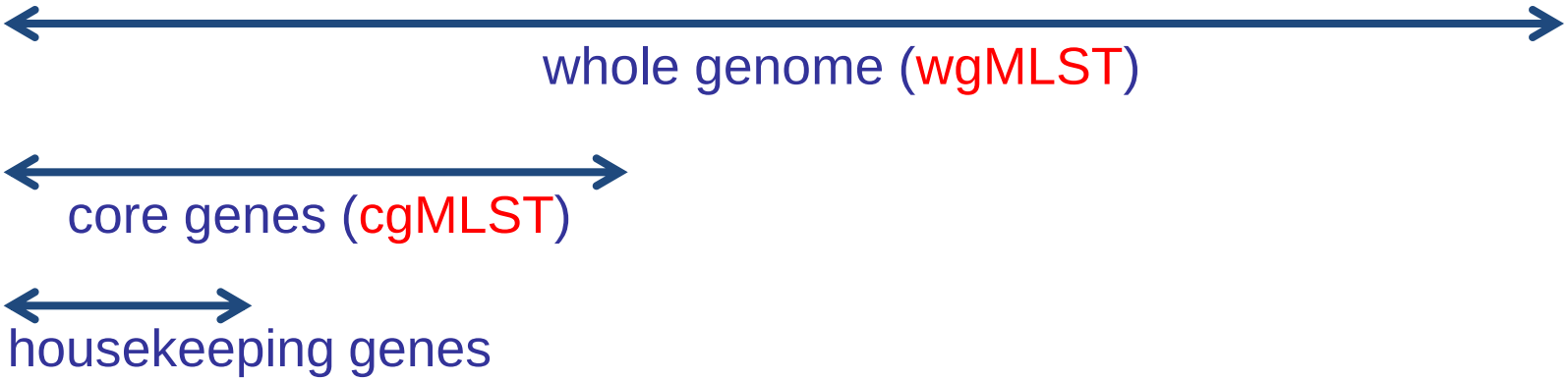
Applying MLST to *E. coli*

Conventional MLST

7 housekeeping genes

Low sensitivity	Good for phylogenetic analysis
High robustness	Not good enough for outbreak investigation

MLST from WGS data



Available Databases

Salmonella

Strains:98638

Assembled
- Legacy:**7229**
- From NGS:**86597**
- In Progress:**0**

Schemes
- Achtman 7 Gene:**93469**
- cgMLST V2:**86597**
- wgMLST:**86597**
- CRISPOL:**38216**
- CRISPR:**51315**
- rMLST:**86597**

Database Home

Escherichia/Shigella

Strains:65783

Assembled
- Legacy:**8836**
- From NGS:**49455**
- In Progress:**3**

Schemes
- Achtman 7 Gene:**58267**
- wgMLST:**49451**
- cgMLST V1:**49446**
- rMLST:**49451**

Database Home

Yersinia

Strains:2978

Assembled
- Legacy:**1165**
- From NGS:**1627**
- In Progress:**0**

Schemes
- Achtman 7 Gene:**2418**
- McNally 7 Gene:**1994**
- wgMLST:**1627**
- cgMLST V1:**1627**
- rMLST:**1626**

Database Home

Moraxella

Strains:555

Assembled
- Legacy:**420**
- From NGS:**104**
- In Progress:**6**

Schemes
- Achtman 7 Gene:**505**
- rMLST:**84**

Database Home

WARWICK
Warwick Medical School

Need Help? Not sure where to start? Click here to read the manual (Wiki)
Any Questions or comments? please post to our Issue tracker (BitBucket)
Development team:
- Martin Sergeant(m.j.sergeant@warwick.ac.uk)
- Zhemin Zhou(zhemin.zhou@warwick.ac.uk)
- Nabil-Fareed Alikhan(n-f.alikhan@warwick.ac.uk)
- Mark Achtman

BBSRC
bioscience for the future

chewBBACA: assembly based allele-calling of wgMLST

Developed by INNUENDO (EFSA-funded project)

Based on wgMLST scheme developed by Enterobase

https://github.com/B-UMMI/chewBBACA

MICROBIAL GENOMICS

Methods paper template

chewBBACA: A complete suite for gene-by-gene schema creation and strain identification

Mickael Silva[1], Miguel Machado[1], Diogo N. Silva[1], Mirko Rossi[2], Jacob Moran-Gilad[3,4], Sergio Santos[1], Mario Ramirez[1] and João André Carriço[1]*

1 Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal 2 Department of Food Hygiene and Environmental Health, Faculty of Veterinary Medicine, University of Helsinki, Finland 3 Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer Sheva, Israel 4 Public Health Services, Ministry of Health, Jerusalem, Israel

chewBBACA: assembly based allele-calling of wgMLST

Developed by INNUENDO (EFSA-funded project)

It works on pre-assembled contigs (.fasta)

Loci and alleles are defined by Prodigal

BLAST Score Ratio>0.6 to identify the allele

7601 curated loci
About 3400 loci called per strain

chewBBACA results - Statistics

Genome	EXC	INF	LNF	PLOT	NIPH	ALM	ASM
NC_017162.fna	892	2319	1909	0	104	5	37
NC_011586.fna	1563	1697	1809	0	116	6	75

The column headers stand for:

- ***EXC*** - alleles which have exact matches (100% DNA identity) with previously identified alleles
- ***INF*** - inferred new alleles using Prodigal CDS predictions
- ***LNF*** - loci not found. No alleles were found for the number of loci in the schema shown. This means that, for those loci, there were no BLAST hits or they were not within the BSR threshold for allele assignment.
- ***PLOT*** - possible loci on the tip of the query genome contigs (see image below). A locus is classified as *PLOT* when the CDS of the query genome has a BLAST hit with a known larger allele that covers the CDS sequence entirely and the unaligned regions of the larger allele exceeds one of the query genome contigs ends. This could be an artifact caused by genome fragmentation resulting in a shorter CDS prediction by Prodigal. To avoid locus misclassification, loci in such situations are classified as *PLOT*.

chewBBACA results on ARIES

Statistics

Genome	EXC	INF	LNF	PLOT	NIPH	ALM	ASM
Genome	EXC	INF	LNF	PLOT	NIPH	ALM	ASM
ED1032_contigs	3543	16	4007	0	27	3	5
ED1088_contigs	3348	77	4120	3	16	2	35
ED1089_contigs.fasta	3105	116	4263	6	11	16	84
ED1104_contigs.fasta	3493	4	4055	1	13	5	30
ED1105_contigs.fasta	3433	12	4098	1	14	4	39

Contigs info

FILE	b0073.fasta	b0074.fasta	b0075.fasta
ED1032_contigs	scaffold_0&199417-198324&-	scaffold_0&200988-199415&-	LNF
ED1088_contigs	NODE_1_length_228150_cov_40.8159_ID_1&198543-197450&-	NODE_1_length_228150_cov_40.8159_ID_1&200114-198541&-	LNF
ED1089_contigs.fasta	NODE_1_length_227956_cov_19.4419_ID_1&197229-196136&-	NODE_1_length_227956_cov_19.4419_ID_1&198800-197227&-	LNF
ED1104_contigs.fasta	NODE_4_length_186376_cov_34.6136_ID_7&29626-30717&+	NODE_4_length_186376_cov_34.6136_ID_7&28055-29626&+	LNF
ED1105_contigs.fasta	NODE_4_length_186376_cov_40.795_ID_7&29626-30717&+	NODE_4_length_186376_cov_40.795_ID_7&28055-29626&+	LNF

Alleles

FILE	b0073.fasta	b0074.fasta	b0075.fasta	b0076.fasta	b0077.fasta	b0078.fasta
ED1032_contigs	10	11	LNF	460	13	2
ED1088_contigs	10	11	LNF	3	13	2
ED1089_contigs.fasta	10	11	LNF	3	13	2
ED1104_contigs.fasta	10	11	LNF	3	13	2
ED1105_contigs.fasta	10	11	LNF	3	13	2

File to use for cluster analysis

Logging info

Repeated loci

Minimum spanning tree on Phyloviz online

PHYLOViZ Online

Home
About
News
API
Public Data sets
Upload Data sets

PHYLOViZ Online

Web-based tool for visualization, phylogenetic inference, analysis and sharing

PHYLOViZ Online is an online version of the software PHYLOViZ, a software that allows the analysis of sequence-based typing methods that generate allelic profiles and their associated epidemiological data. Our motivation was to give an user-friendly solution for data analysis and sharing without installing any specific software.

The application is freely available to all users and there is no login requirement. All users can upload and perform data analysis. There is the additional possibility of storing data on the application for future access upon registration.

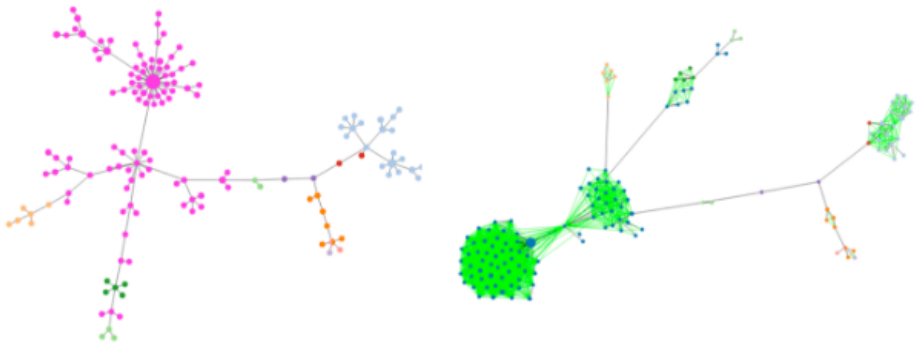
CHECK OUT THE ONLINE VIDEO TUTORIAL IN YOUTUBE OR THE WALKTHROUGH OF THE AVAILABLE FEATURES

Sample data sets available!

More information on the data formats can be found here.

Try all PHYLOViZ Online different functionalities using:

- Login-free upload.
- The common user: ***demo*** and password: ***demo***.
- Using public datasets.



Minimum spanning tree on Phyloviz online – e.g.