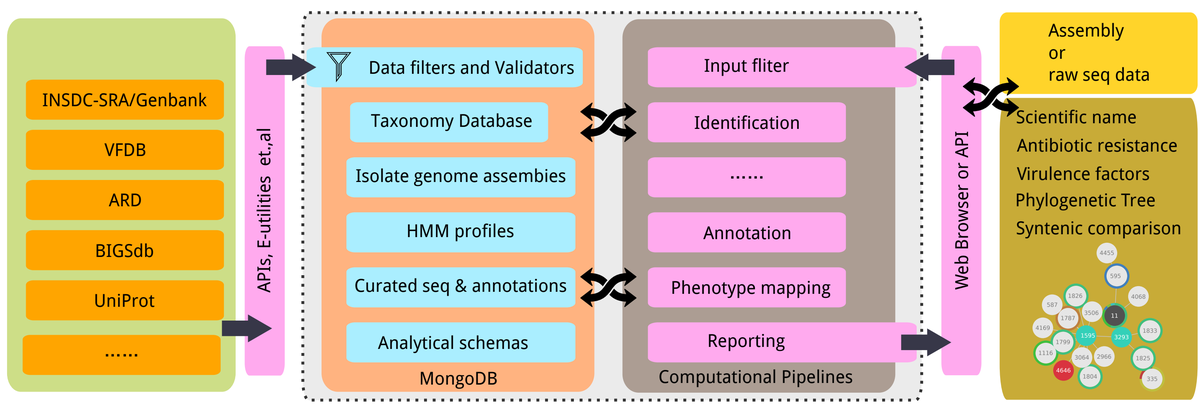
MicrobeTrakr: A Web Service for Whole-genome Sequence-based Microbial Identification

Shiming Chen, Ya Wang, Xiaolin Hu, Junning Wang
E-mail: info@zetabio.com - Phone: +86-21-61841318
Zeta Biosciences, 285 Jianguo W Rd. Xuhui District, Shanghai, P. R. China

Introduction:

MicrobeTrakr is a web service for WGS based microbial species identification and molecular epidemiology. MicrobeTrakr can perform smart data analysis, including species identification, sub-typing, clinical and epidemiological related genes identification. Additional knowledge related to the query will be retrieved to help data interpretation. MicrobeTrakr takes assembled bacterial genome sequences as input(web browser) and can be accessed via API to upload raw data of HTS sequencer. MicrobeTrakr uses Needleman-Wunsch global alignment to identify house keeping genes in MLST schemes, ARGs and VFGs automatically. Further more MicrobeTrakr using well established algorithm to group STs and display relationship between the closed-related isolates of particular bacterial species. MicrobeTrakr generates the eBURST diagram and phylogeny tree to present evolution relationship of input isolate. Phenotypes of the query also predicted by mapping detected genotype to curated data.



Architecture:

MicrobeTrakr is a knowledge engine like service, it can generates sense-making results for user at all education levels. MicrobeTrakr use the standard data exchange format and controlled-vocabularies defined by the INSDC, and can be invoked programmatically.

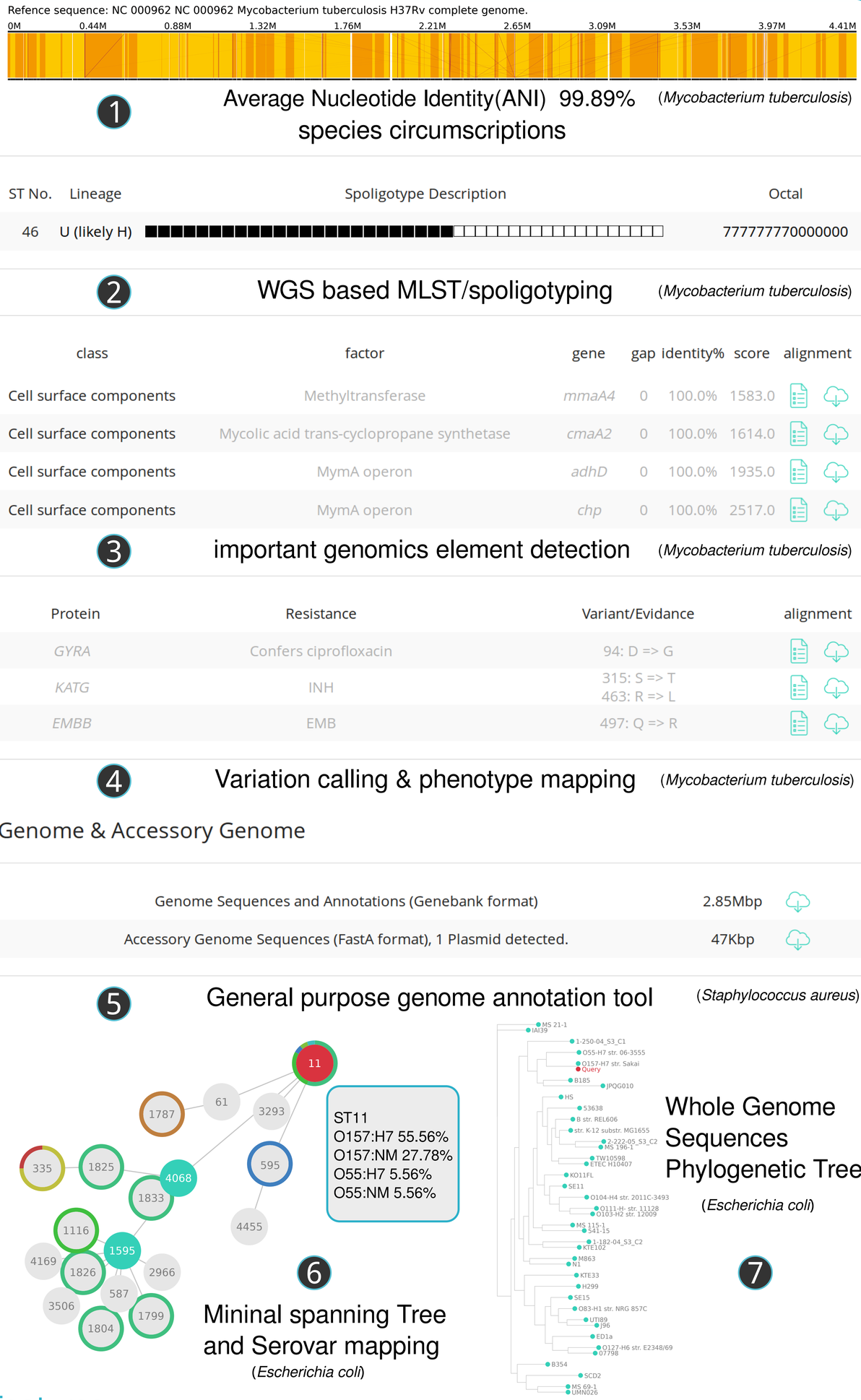
Strain	Phenotype	PCR-Genotype	WGS-Genotype	WGS-ST	PCR-ST	#WGS-Plasmids
AUS 0002	VSE	*nil*	*vanB*	202	202	12
AUS 0006	VSE	*nil*	*ND*	17	17	11
AUS 0007	VSE	*nil*	*ND*	17	17	9
AUS 0010	VSE	*nil*	*ND*	32	32	9
AUS 0012	VRE	*vanB*	*vanB*	17	17	13
AUS 0015	VSE	*nil*	*ND*	17	17	7
AUS 0016	VRE	*vanB*	*vanB*	18	18	7
AUS 0018	VSE	*nil*	*ND*	252	252	14
AUS 0019	VSE	*nil*	*ND*	17	17	9
AUS 0021	VRE	*vanB*	*vanB*	17	17	9
AUS 0029	VSE	*nil*	*ND*	203	203	14
AUS 0030	VRE	*vanB*	*vanB*	252	252	7
AUS 0031	VRE	*vanB*	*vanB*	252	252	7
AUS 0038	VSE	*nil*	*ND*	252	252	8
AUS 0039	VSE	*nil*	*ND*	204	204	14
AUS 0040	VSE	*nil*	*ND*	203	203	9
AUS 0041	VSE	*nil*	*ND*	203	203	9
AUS 0045	VRE	*vanB*	*vanB*	203	203	11
AUS 0048	VRE	*vanB*	*vanB*	203	203	9
AUS 0050	VRE	*vanB*	*vanB*	252	252	13
AUS 0053	VSE	*nil*	*vanB*	203	203	13
AUS 0055	VRE	*vanB*	*vanB*	203	203	9
AUS 0057	VSE	*nil*	*ND*	252	252	9
AUS 0063	VSE	*nil*	*ND*	132	132	6
AUS 0064	VSE	*nil*	*ND*	252	252	4
AUS 0065	VSE	*nil*	*ND*	341	341	6
AUS 0069	VRE	*vanB*	*vanB*	414	414	7
AUS 0070	VRE	*vanB*	*vanB*	203	203	10
AUS 0071	VRE	*vanB*	*vanB*	252	252	10
AUS 0078	VSE	*nil*	*ND*	203	203	7
AUS 0080	VSE	*nil*	*ND*	17	17	6
AUS 0081	VRE	*vanB*	*vanB*	483	483	12
AUS 0084	VRE	*vanB*	*vanB*	17	17	13
AUS 0087	VRE	*vanB*	*vanB*	203	203	9
AUS 0090	VRE	*vanB*	*vanB*	203	203	7
AUS 0091	VRE	*vanB*	*vanB*	203	203	8
AUS 0092	VSE	*nil*	*ND*	55	55	3
AUS 0094	VSE	*nil*	*ND*	17	17	7
AUS 0098	VSE	*nil*	*ND*	17	17	7
AUS 0100	VSE	*nil*	*ND*	17	17	9
AUS 0101	VSE	*nil*	*ND*	22	22	8
RPH1	VRE	*vanB*	*vanB*	341	341	4
RPH2	VRE	*vanB*	*vanB*	203	203	8
RPH3	VRE	*vanB*	*vanB*	203	203	10

Results of WGS and PCR based genotyping

Evaluation:

Illumina reads of 44 *Staphylococcus aureus* isolates were downloaded from the NCBI-SRA database(PRJNA205886)[1] and processed by MicrobeTrakr pipeline. WGS-based MLST types of all 44 isolates confirmed the results from the PCR and conventional sequencing method[1]. Truncated *vanB* operon were detected in 2 isolates(AUS0002 and AUS0053) using WGS method which PCR assays is *vanB* negative. Our preliminary evalution suggest that WGS can give precies MLST results as conventional sequencing, and provide higher resolution for genotyping.

[1] Howden, Benjamin P., et al. "*Genomic insights to control the emergence of vancomycin-resistant enterococci.*" MBio 4.4 (2013): e00412-13.



Features:

MicrobeTrakr can take bacterial genome assembly in FASTA format sequence(s) to identify bacterial species, subtyping and further analysis.

1. Use whole genome wide average nucleotide identity (ANI) to define species.
2. Smart subtyping using MLST or spoligotyping schema.
3. Context based virulence and antibiotic resistance related genes detection.
4. Accessory genome recognization and precise variation phenotypes mapping.
5. Whole genome-based phylogenetic tree reconstruction.

Online Demo

Conclusions:

We believe that the WGS would change the diagnostic paradigm of microbiology and epidemiology in near future. However, standards for WGS and analysis is still under development. Easy use and automatically data analytic tools are also claimed. MicrobeTrakr is a comprehensive and ideal automatically data analytic tools for whole-genome sequencing based bacterial identification and subtyping. It can be use for epidemiological tracking and drug resistance prediction. MicrobeTrakr is freely available at http://**microbetrakr.org** and can be accessed programmatically.

Zeta Biosciences
Making sense of biological data