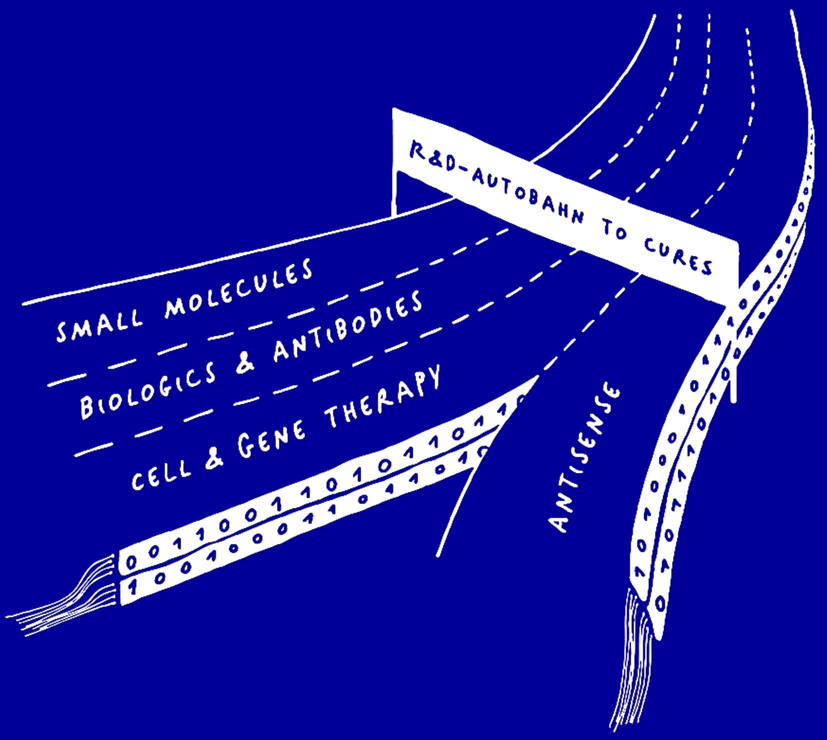
#RESEARCHNEVERSTOPS

Bioinformatics supporting Infectious Disease

We create knowledge for efficient discovery and development of better medicines



Bioinformatics implements solutions enabling data-driven A.INF pipeline

Our mission: decipher compound MoA, assess and describe MoR

Provide improved and optimised **microbiology tools** for compound assessment:

- Strain genome complete characterisation:
  - Genome sequence (hybrid assembly)
  - Genome annotation (focus on virulence, resistance and stress genes)
  - *In silico* taxonomic identification (species, MLST)
- Use extensive comparative genomics to propose bespoke strain panels
- Validation of new genomic constructs

Data assets:

- Data collection of genomes from clinical and WT bacterial strains coupled with phenotypic experimental data

Analyses & Models

		WGS	TnSeq	RNAseq
Identify and characterise Mode of Action (**MoA**)	• Identify **Target** by investigating compound impacted pathways • Describe compound **metabolism** in bacteria • Genome-wide gene **essentiality** and **fitness** analysis under cpd challenge	✓	✓	✓
Characterise Mechanism of Resistance (**MoR**) Evaluate **risk of resistance**	• Identify **genomic loci** responsible for spontaneous & induced resistance • Possible resistance **"pathways"** description by poolseq (bacterial population sequencing) analysis • Genome-wide gene **essentiality** and **fitness** analysis under cpd challenge • **GWAS** • *In vitro/in vivo* **fitness** analysis	✓[1]	✓[1]	✓
Translational studies	• Evaluate relevance of MoA & MoR • Compare and contrast preclinical models • Identify translational medium for a screening • Identify essential genes for *in vivo* colonization • *In vitro* vs *in vivo* models pathway regulation comparison		✓	✓
	• (Multi-)Omics **data integration** & functional interpretation using genome-scale metabolic networks	**Metabolic Network Analysis**		

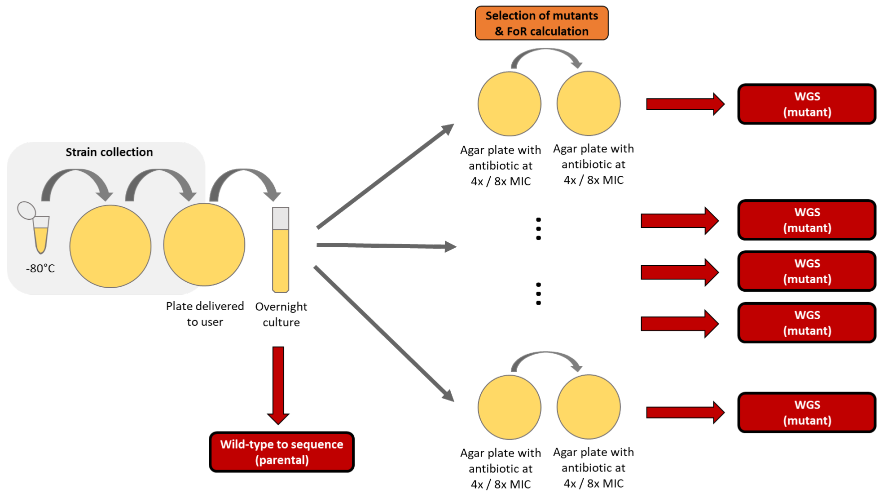
---

[1] Pre-existing, partial solutions for de novo genome assembly and TnSeq analysis

Resistant isolate genome analysis to identify mutated loci

Describe modified genomic loci in resistant isolates

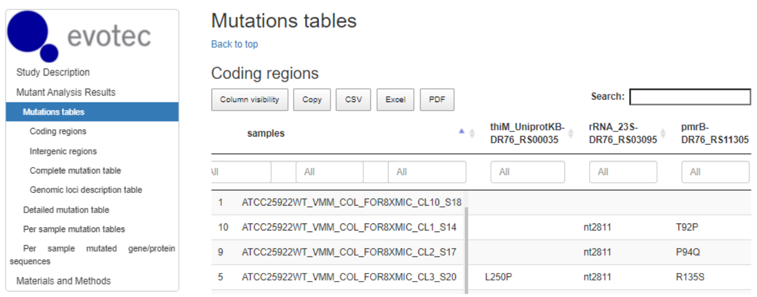
Use-case: Determining the mutations present in the surviving isolates of FoR[1)]



Other applications:

- Strain characterisation from patients with mono- or multi-species/strain infections
- Strain fitness analysis
  (genomic identification of variants from fitness experiment)

Mutation table

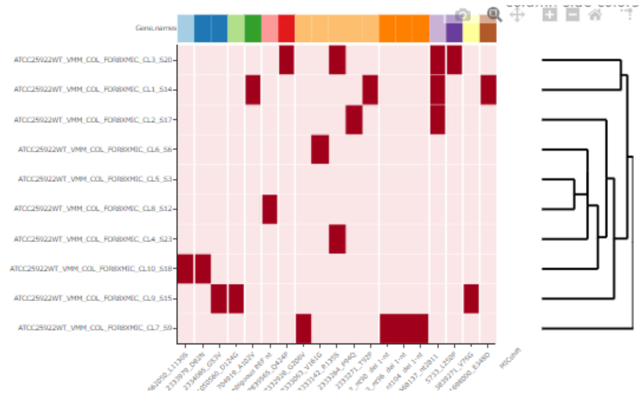


Mutations tables

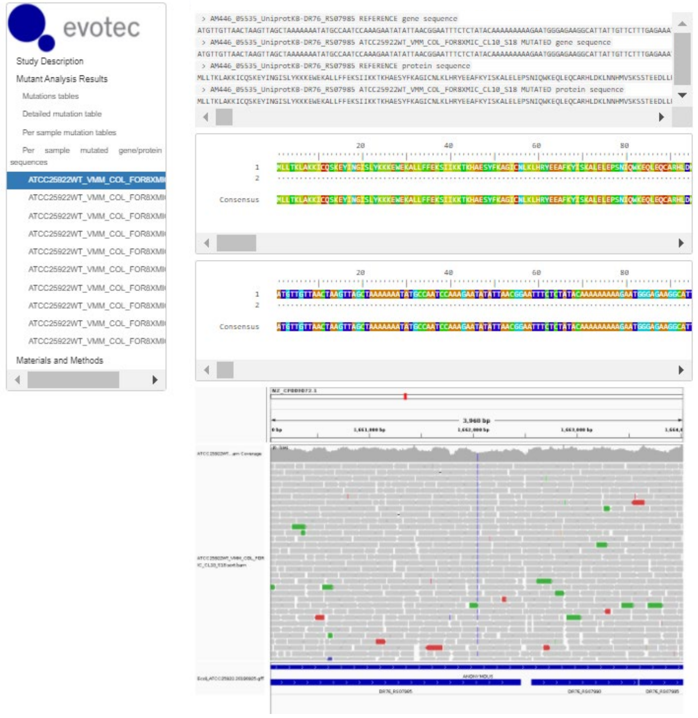
Coding regions

	samples	thiM_UniprotKB-DR76_RS00035	rRNA_23S-DR76_RS03095	pmrB-DR76_RS11305
1	ATCC25922WT_VMM_COL_FOR8XMIC_CL10_S18			
10	ATCC25922WT_VMM_COL_FOR8XMIC_CL1_S14		nt2811	T92P
9	ATCC25922WT_VMM_COL_FOR8XMIC_CL2_S17		nt2811	P94Q
5	ATCC25922WT_VMM_COL_FOR8XMIC_CL3_S20	L250P	nt2811	R135S

Sample & mutation clustering



Mutated gene/protein sequences/read support



[1)] FoR: frequency of resistance

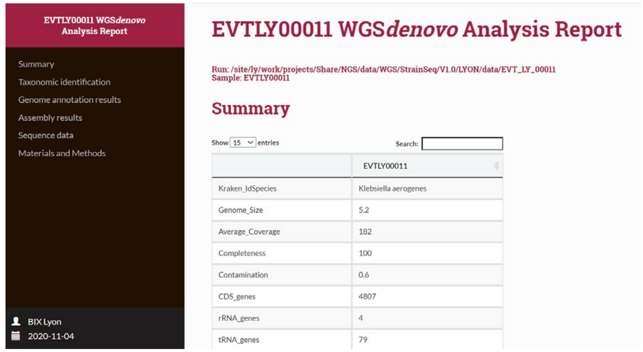
Bacterial isolate genome analysis

Deliver complete genome sequence and gene annotation

Comparative genomics

- Pangenome analysis
- Phylogenetic analysis
- Gene/mutation prevalence in a strain collection
- Gene replacement in reference strain validation
- Tn insertion mutant validation

Assembly QC



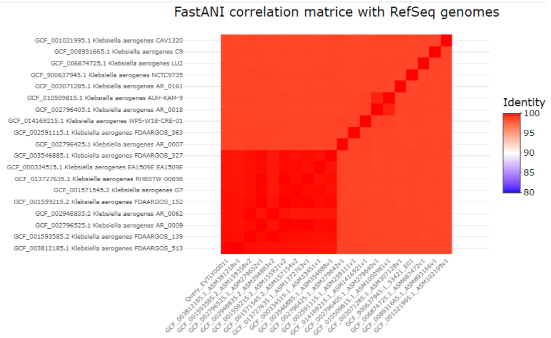
EVTLY00011 WGS*denovo* Analysis Report

Run: /site/ly/work/projects/Share/NGS/data/WGS/StrainSeq/V1.0/LYON/data/EVT_LY_00011
Sample: EVTLY00011

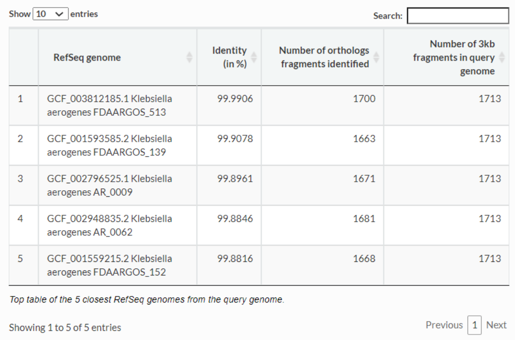
Summary

	EVTLY00011
Kraken_IdSpecies	Klebsiella aerogenes
Genome_Size	5.2
Average_Coverage	182
Completeness	100
Contamination	0.6
CDS_genes	4807
rRNA_genes	4
tRNA_genes	79

Strain diversity analysis (FastANI)



Genome annotation

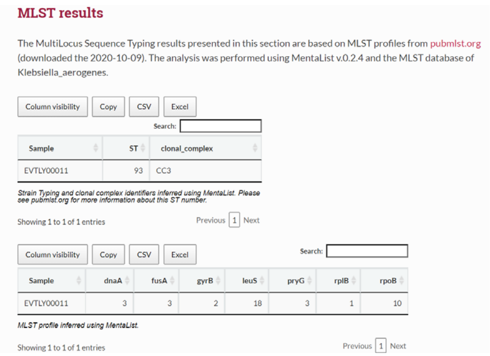


	RefSeq genome	Identity (in %)	Number of orthologs fragments identified	Number of 3kb fragments in query genome
1	GCF_003812185.1 Klebsiella aerogenes FDAARGOS_513	99.9906	1700	1713
2	GCF_001593585.2 Klebsiella aerogenes FDAARGOS_139	99.9078	1663	1713
3	GCF_002796525.1 Klebsiella aerogenes AR_0009	99.8961	1671	1713
4	GCF_002948835.2 Klebsiella aerogenes AR_0062	99.8846	1681	1713
5	GCF_001559215.2 Klebsiella aerogenes FDAARGOS_152	99.8816	1668	1713

Top table of the 5 closest RefSeq genomes from the query genome.

Showing 1 to 5 of 5 entries

Strain typing



MLST results

The MultiLocus Sequence Typing results presented in this section are based on MLST profiles from pubmlst.org (downloaded the 2020-10-09). The analysis was performed using MentaList v.0.2.4 and the MLST database of Klebsiella_aerogenes.

Sample	ST	clonal_complex
EVTLY00011	93	CC3

Strain Typing and clonal complex identifiers inferred using MentaList. Please see pubmlst.org for more information about this ST number.

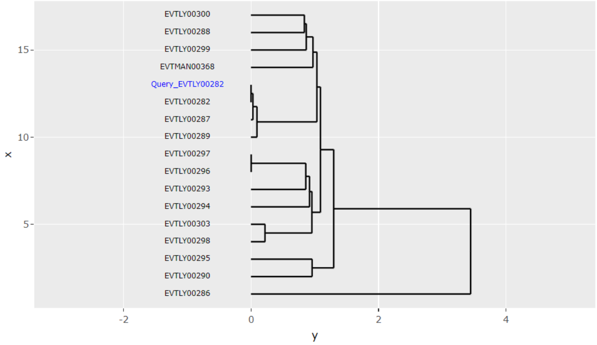
Showing 1 to 1 of 1 entries

Sample	dnaA	fusA	gyrB	leuS	pryG	rplB	rpoB
EVTLY00011	3	3	2	18	3	1	10

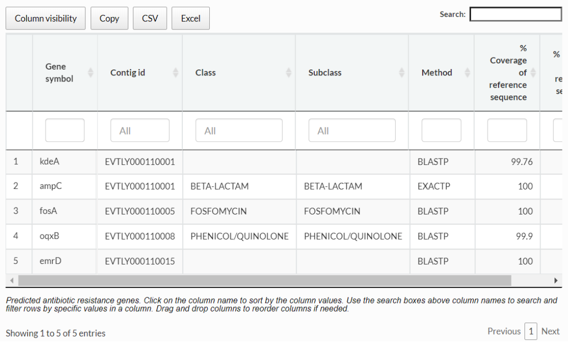
MLST profile inferred using MentaList.

Showing 1 to 1 of 1 entries

Strain diversity analysis (FastANI)



Antibiotic resistance, stress and virulence genes



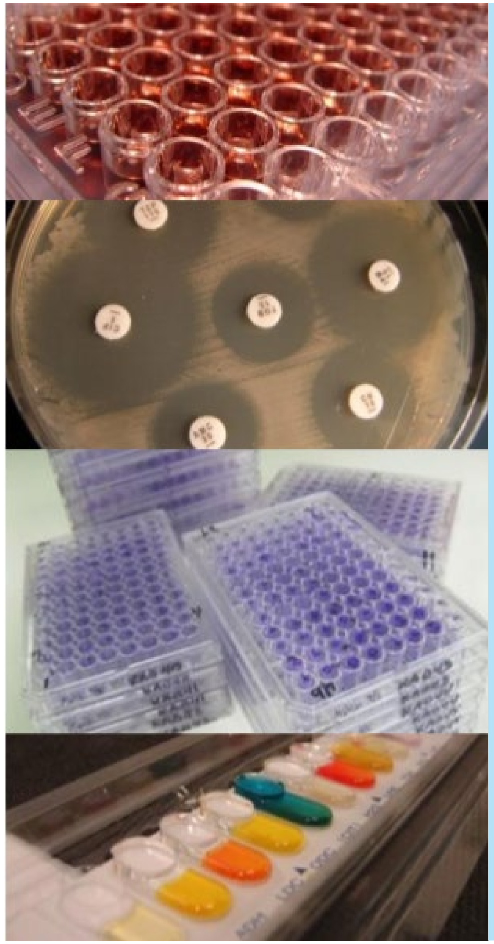
	Gene symbol	Contig id	Class	Subclass	Method	% Coverage of reference sequence
1	kdeA	EVTLY000110001			BLASTP	99.76
2	ampC	EVTLY000110001	BETA-LACTAM	BETA-LACTAM	EXACTP	100
3	fosA	EVTLY000110005	FOSFOMYCIN	FOSFOMYCIN	BLASTP	100
4	oqxB	EVTLY000110008	PHENICOL/QUINOLONE	PHENICOL/QUINOLONE	BLASTP	99.9
5	emrD	EVTLY000110015			BLASTP	100

Predicted antibiotic resistance genes. Click on the column name to sort by the column values. Use the search boxes above column names to search and filter rows by specific values in a column. Drag and drop columns to reorder columns if needed.

Showing 1 to 5 of 5 entries

Genome analysis put to scale: EvostrAIn™

A large and rapidly evolving collection of highly characterised strains and clinical isolates



EvostrAIn™
~ 8,000 strains from the clinic and reference collections

EvostrAIn™– genomics digital counterpart enabling

- Improved collection quality
- Informed enrichment
- Accelerated proposal2contract
- Design of bespoke screening panels

Strain data

- Strain metadata

Compounds

Phenotype data

- MIC
- FoR
- Vitek2

Genomic information

- Genome sequence
- Genome annotation
- Orthologous gene group annotation
- Antibiotic resistance, stress and virulence genes ID
- Taxonomic identification
- Strain Typing
- etc.

EvostrAIn™

METACYC

ENA European Nucleotide Archive

RefSeq

JGI IMG/M INTEGRATED MICROBIAL GENOMES & MICROBIOMES

Public microbial genomes

Bacterial genomics Toolbox

- ✓ Pangenome analysis
- ✓ Phylogenetic tree analysis
- ✓ Blast sequence/gene search
- ✓ GWAS analysis
- ✓ Optimized strain panel building

↔

EvostrAIn™-gDB
EvostrAIn™
genomics database

↔

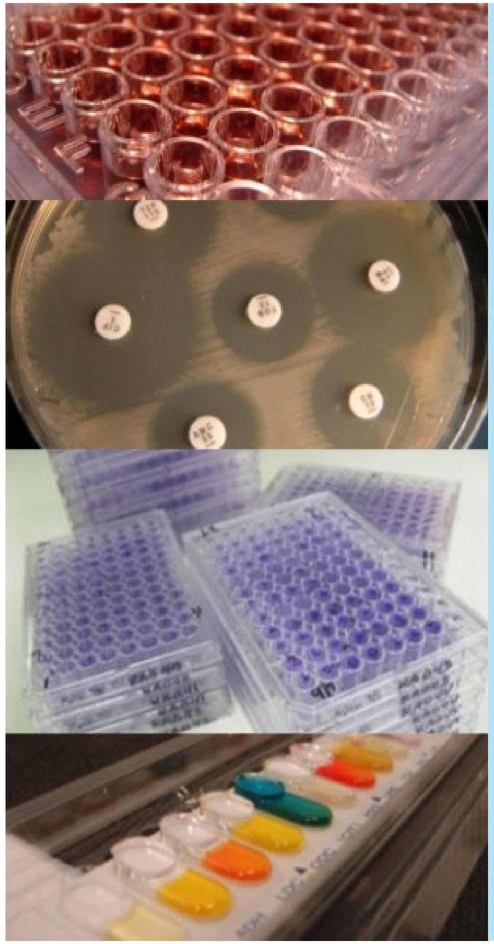
Web-based search engine

- ✓ Query any genome annotation field
- ✓ Obtain genomes sequences
- ✓ Obtain gene and protein sequences
- ✓ Query resistance results
- ✓ Query genes for short-range variants (SNPs, short indels)

High quality data set

Genome data access

EvostrAIn^{TM} genomics: allowing microbiologists access to genomics and phenotype data



EvostrAIn^{TM}
~ 8,000 strains from the clinic and reference collections

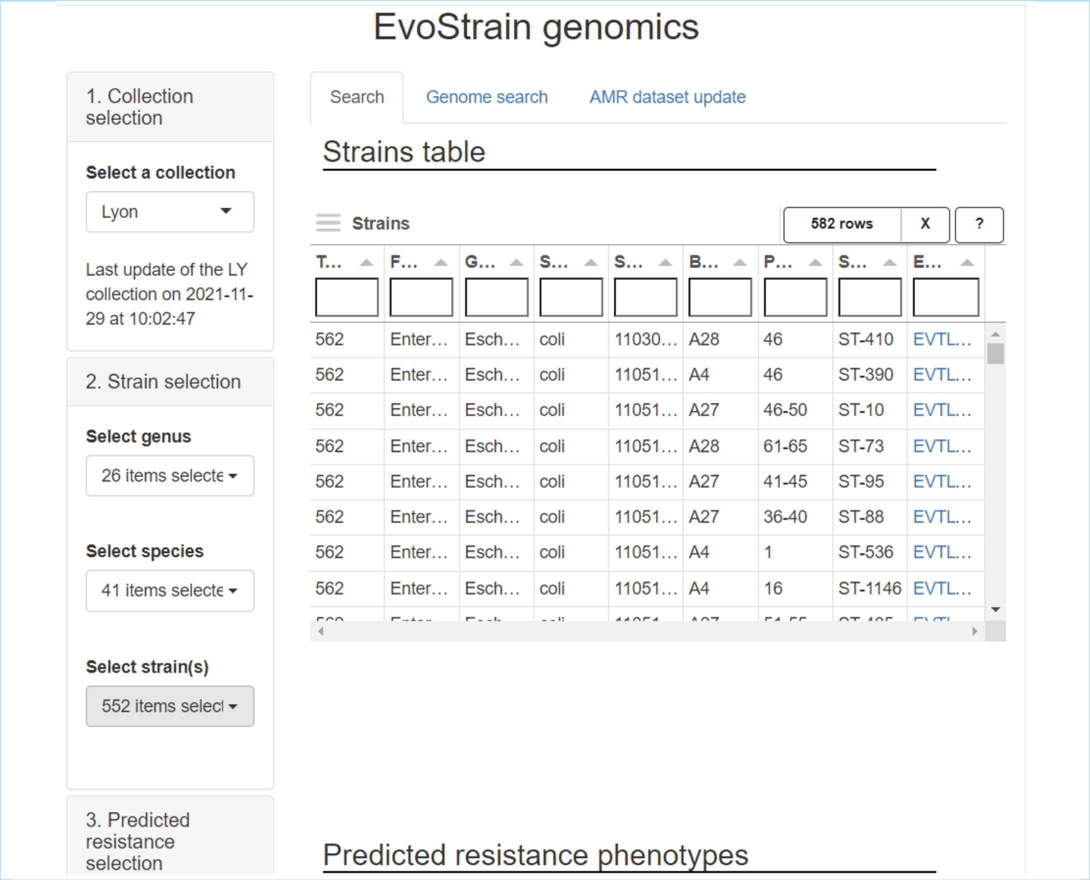
EvostrAIn^{TM}– genomics digital counterpart enabling

- Improved collection quality
- Informed enrichment
- Accelerated proposal2contract
- Design of bespoke screening panels

Full or (interactively) filtered data:

- Strain and genome information
- Genome sequence
- Gene sequence
- Protein sequence
- Antibiogram data
- Gene resistance table/heatmap
- …

Completed with scripted solutions to fully answer to the platform needs



EvoStrain genomics

1. Collection selection

Select a collection: Lyon

Last update of the LY collection on 2021-11-29 at 10:02:47

2. Strain selection

Select genus: 26 items selecte

Select species: 41 items selecte

Select strain(s): 552 items select

3. Predicted resistance selection

Strains table

Strains — 582 rows

T…	F…	G…	S…	S…	B…	P…	S…	E…
562	Enter…	Esch…	coli	11030…	A28	46	ST-410	EVTL…
562	Enter…	Esch…	coli	11051…	A4	46	ST-390	EVTL…
562	Enter…	Esch…	coli	11051…	A27	46-50	ST-10	EVTL…
562	Enter…	Esch…	coli	11051…	A28	61-65	ST-73	EVTL…
562	Enter…	Esch…	coli	11051…	A27	41-45	ST-95	EVTL…
562	Enter…	Esch…	coli	11051…	A27	36-40	ST-88	EVTL…
562	Enter…	Esch…	coli	11051…	A4	1	ST-536	EVTL…
562	Enter…	Esch…	coli	11051…	A4	16	ST-1146	EVTL…

Predicted resistance phenotypes

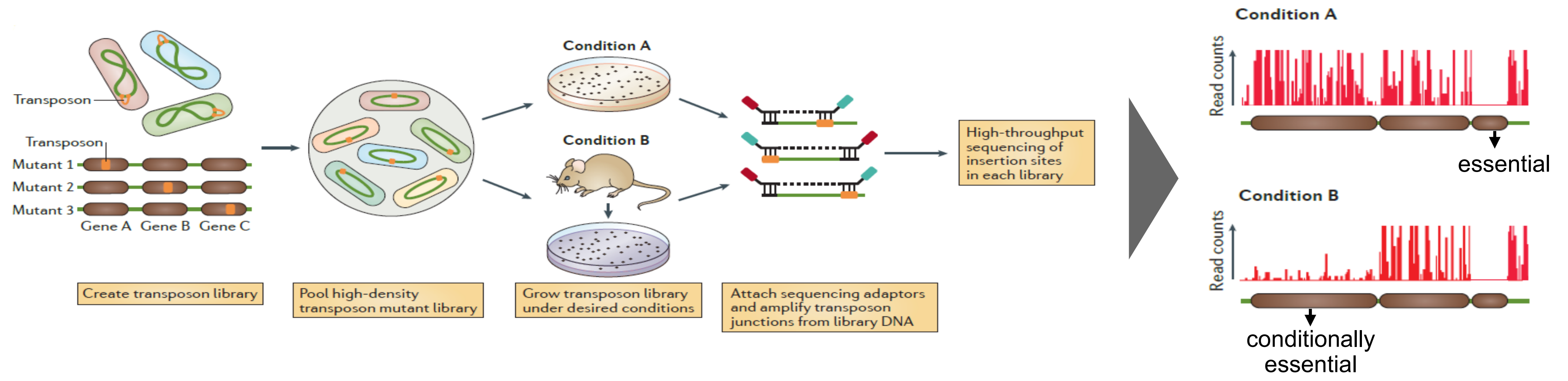
TnSeq platform to characterise compound MoA / MoR

Genotype-phenotype studies for gene essentiality and fitness analyses

High density Tn mutant profiling enables the Discovery and Development of Novel Antimicrobials by revealing molecular targets and mechanisms of resistance of hit compounds

Tn5 based saturated libraries available for 2 species

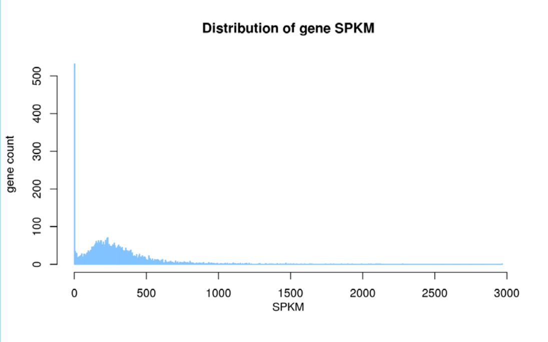
- *E. coli* (ATCC 35218), *K. pneumoniae* (ATCC 13883)
  - construction of a *P. aeruginosa* (PA103) Tn*5* library is ongoing
- Libraries validated and tested for compounds profiling and culture conditions comparison (VMM)

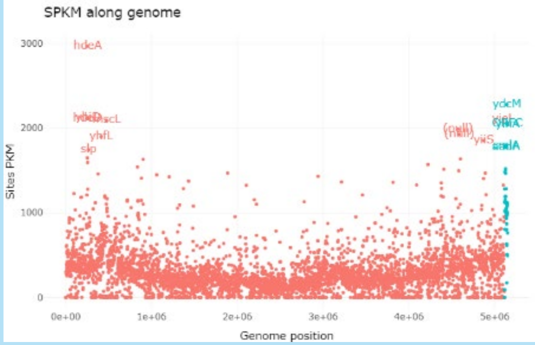


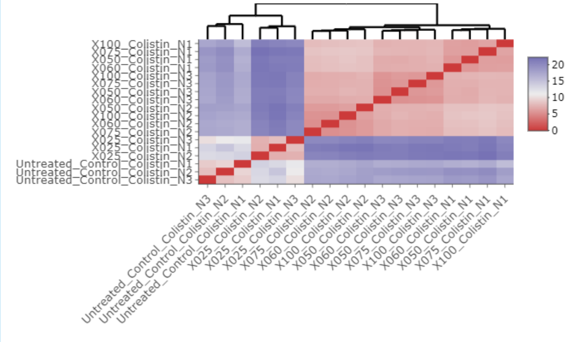
MC Chao et al. The design and analysis of transposon insertion sequencing experiments. Nature Rev Microbiol. 2016

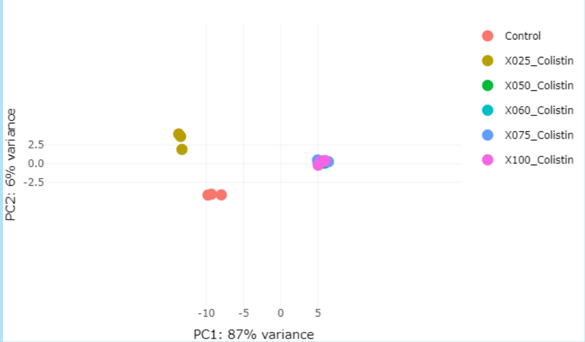
Complete gene analysis-TnSeq pipeline is available

Quality Control

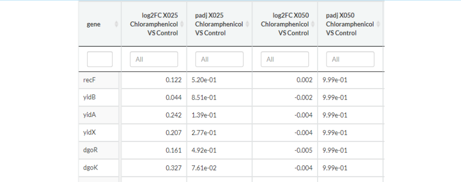




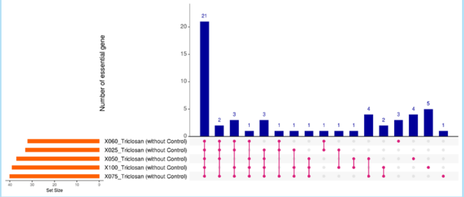




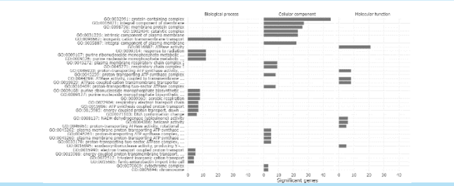
Essential Gene Analysis

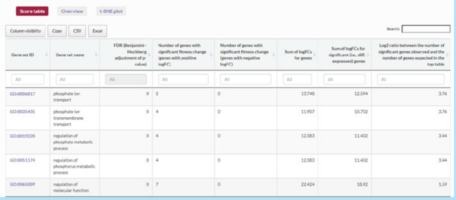


gene	log2FC X025 Chloramphenicol VS Control	padj X025 Chloramphenicol VS Control	log2FC X050 Chloramphenicol VS Control	padj X050 Chloramphenicol VS Control
	All	All	All	All
recF	0.122	5.20e-01	0.002	9.99e-01
yidB	0.044	8.51e-01	-0.002	9.99e-01
yidA	0.242	1.39e-01	-0.004	9.99e-01
yidX	0.207	2.77e-01	-0.004	9.99e-01
dgoR	0.161	4.92e-01	-0.005	9.99e-01
dgoK	0.327	7.61e-02	-0.004	9.99e-01

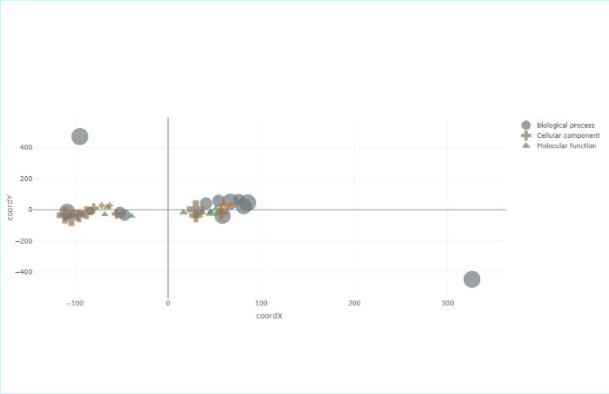


Functional Analysis

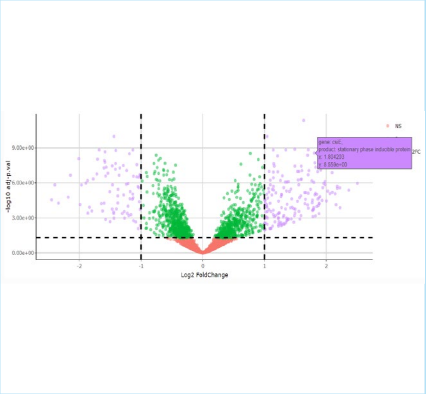


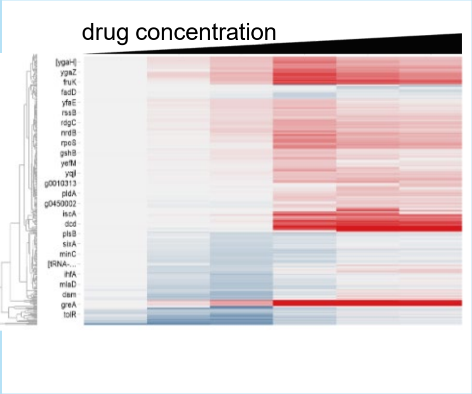


Gene set ID	Gene set name	FDR (Benjamini-Hochberg adjustment of p-value)	Number of genes with significant fitness change (genes with positive logFC)	Number of genes with significant fitness change (genes with negative logFC)	Sum of logFCs for genes	Sum of logFCs for significant (i.e., diff. expressed) genes	Log2 ratio between the number of significant genes observed and the number of genes expected in the top table
All	All	All	All	All	All	All	All
GO:0006817	phosphate ion transport	0	5	0	13.748	12.594	3.76
GO:0035435	phosphate ion transmembrane transport	0	4	0	11.907	10.702	3.76
GO:0019220	regulation of phosphate metabolic process	0	4	0	12.383	11.402	3.44
GO:0051174	regulation of phosphorus metabolic process	0	4	0	12.383	11.402	3.44
GO:0065009	regulation of molecular function	0	7	0	22.424	18.92	1.59



Gene Fitness Analysis



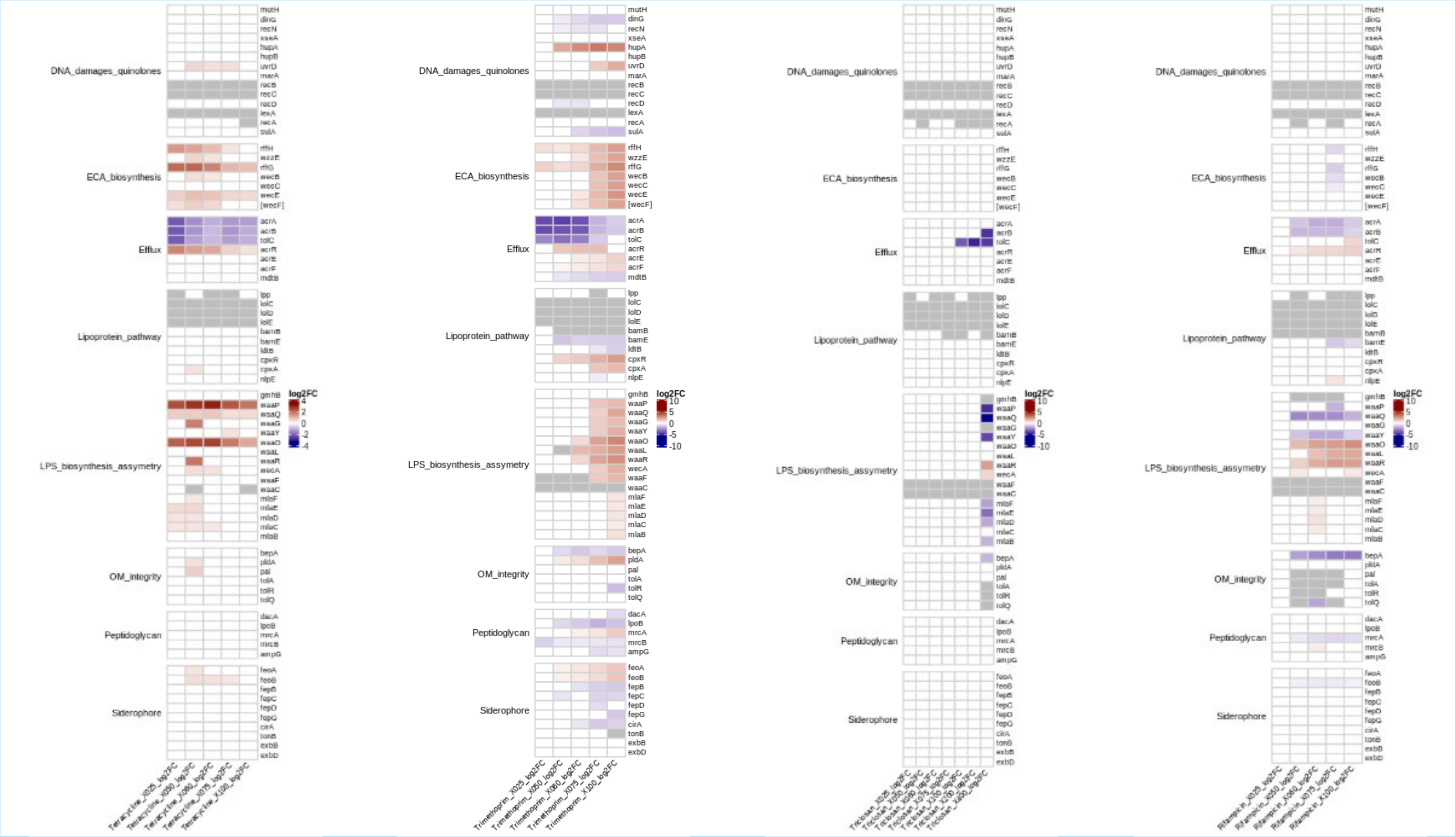


TnSeq platform to characterise compound MoA / MoR

Building a data collection of reference compounds profile for MoA classification (ML)

- **Construction of reference compound data collection on in house Tn-libraries**
  - >30 reference compounds were profiled on *E. coli* and *K. pneumoniae*
  - Stress and toxic compounds are currently tested
  - Meta-analysis and MoA family signature identification
  - Comparison of novel unknown compound with data collection

Manual curation of MoA related genes



Compound microbiological profiling by Phenotypic Microarrays

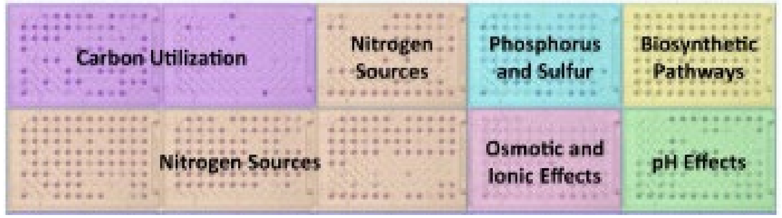
Assessment of growth media impact, compounds synergy and antagonism

- **Robustness of a compound / translation to vivo**
  Medium dependent modulation of compound activity and resistance
- **MoA pathway insight**
  From metabolism/growth medium dependency

960 media in one assay

Assess global metabolism, osmotic conditions and pH sensitivity

- 10 x 96 wells microplates = 960 media
- Compound dose effect
- Growth on 20h



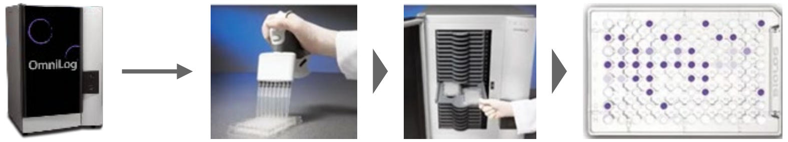
Microbiological profiling - Home made plates in development

- to resume long and fastidious classic assays (salts, cations, BSA etc…)
- to test complex media (serum, urine etc…)

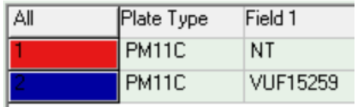
96 media in one assay

- Collect the data

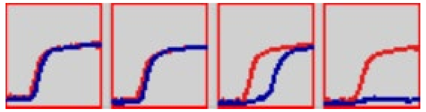
ΔT = 15 min on 24h



- Metabolic profiles comparison



All	Plate Type	Field 1
🟥 1	PM11C	NT
🟦	PM11C	VUF15259

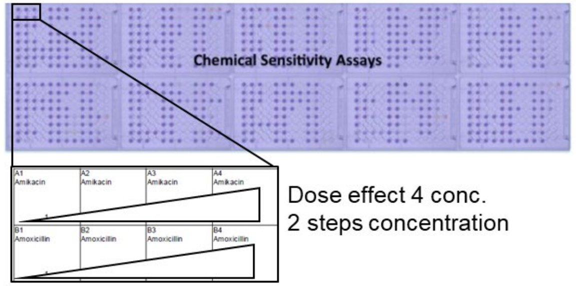


- **Synergy, potentiation, antagonism & Cross-resistance**
- Plates with marketed antibiotics and chemicals in dose response to assess a compound or a strain

240 antibiotics and chemicals in one assay

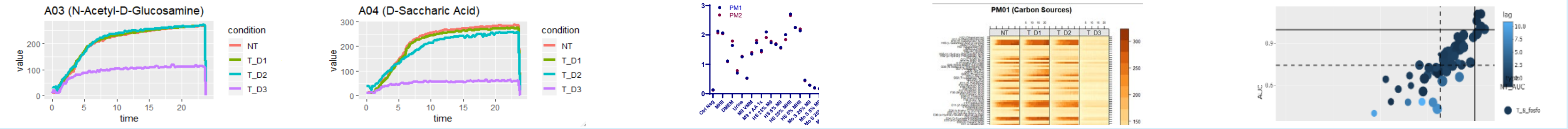
10 MicroPlates to assess chemical interactions

PM11 to 20: Rich medium + 240 ATB and chemicals – Dose effect 4 conc.



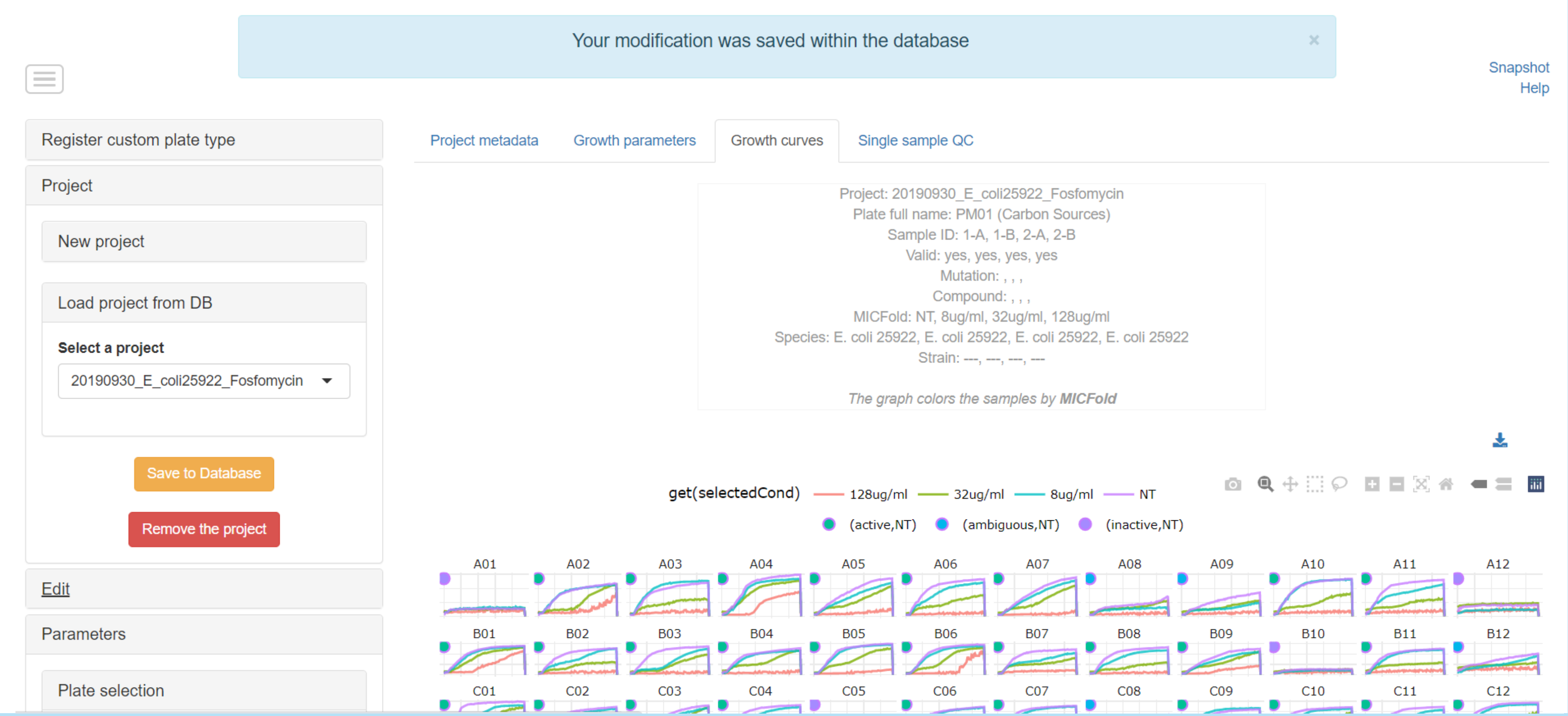
Dose effect 4 conc.
2 steps concentration

QC and data analysis pipeline development by ~~BIX~~ ~~(~~Bioinformatics team ~~LYO)~~



Shiny application for Biolog-QC and data management

- Robust data management
- Graphical user interface for data quality analysis and validation before project import within project data – database
- Rich visualisation toolbox



Your modification was saved within the database

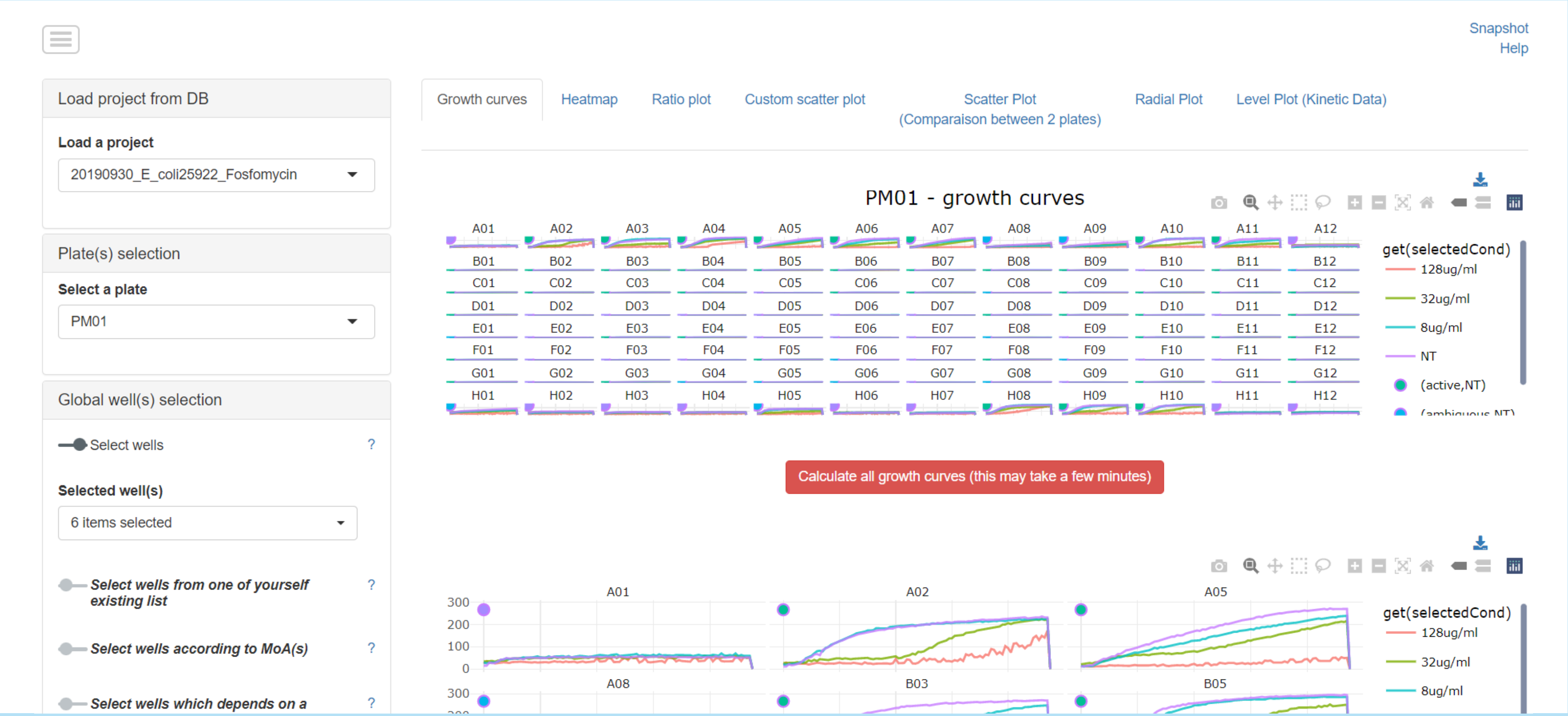
Select a project
20190930_E_coli25922_Fosfomycin

Project: 20190930_E_coli25922_Fosfomycin
Plate full name: PM01 (Carbon Sources)
Sample ID: 1-A, 1-B, 2-A, 2-B
Valid: yes, yes, yes, yes
Mutation: , , ,
Compound: , , ,
MICFold: NT, 8ug/ml, 32ug/ml, 128ug/ml
Species: E. coli 25922, E. coli 25922, E. coli 25922, E. coli 25922
Strain: ---, ---, ---, ---

The graph colors the samples by ***MICFold***

Shiny application for Biolog data analysis

- Robust data management
- Graphical user interface for data quality analysis and validation before project import within project data – database
- Rich visualisation toolbox



Load project from DB

Load a project: 20190930_E_coli25922_Fosfomycin

Plate(s) selection

Select a plate: PM01

Global well(s) selection

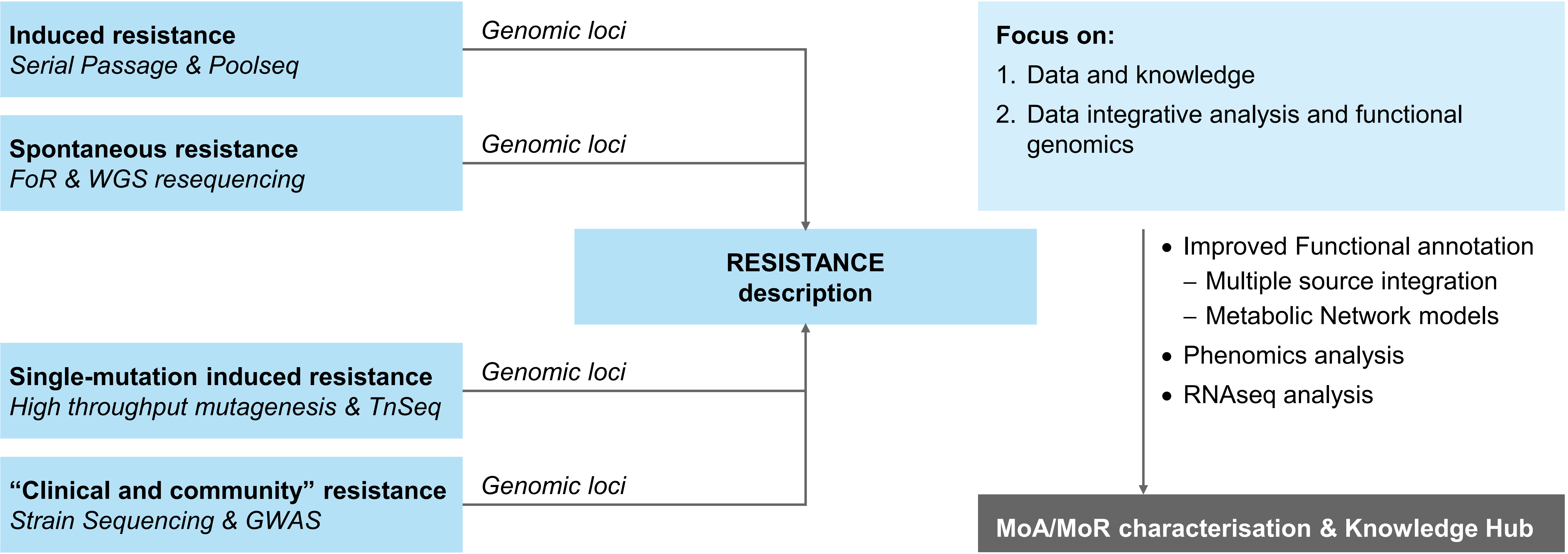
Selected well(s): 6 items selected

Select wells from one of yourself existing list

Select wells according to MoA(s)

Select wells which depends on a

Incremental development: from description to integration & interpretation



Induced resistance
Serial Passage & Poolseq

Genomic loci

Spontaneous resistance
FoR & WGS resequencing

Genomic loci

Single-mutation induced resistance
High throughput mutagenesis & TnSeq

Genomic loci

"Clinical and community" resistance
Strain Sequencing & GWAS

Genomic loci

RESISTANCE description

Focus on:

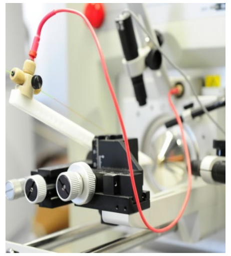
1. Data and knowledge
2. Data integrative analysis and functional genomics

- Improved Functional annotation
  - Multiple source integration
  - Metabolic Network models
- Phenomics analysis
- RNAseq analysis

MoA/MoR characterisation & Knowledge Hub

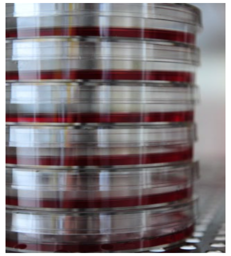
Quantitative MS supporting drug discovery

Chemical proteomics, global proteomics and metabolomics



Global Proteomics Platforms

- High-end quantitative mass spectrometry to monitor protein expression, phosphorylation, glycosylation, acetylation, ubiquitination or arginine methylation
- Targeted mass spectrometry assay development and deployment



Chemical proteomics

- Evotec Cellular Target Profiling^{TM} technology to both identify and quantify interactions with cellular compound targets
- Drug photoaffinity labelling and activity-based protein profiling for covalent target capture

Metabolomics

- *In vitro* and *in vivo* quantification of metabolites in complex sample using SPE-LC-MS/MS

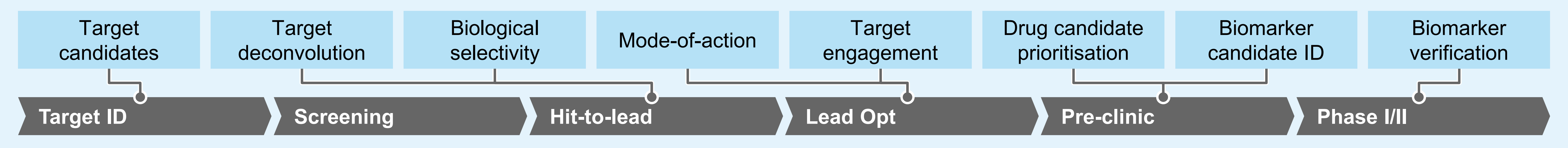
1 ***In vivo* mode-of-action analysis in cells, tissues or patients**

2 **Discovery and verification of biomarker candidates**

3 **Cellular compound selectivity analysis in a native context**

4 **Target de-convolution of hit com-pounds from phenotypic screens**

5 **Targeted analysis in cells, tissues, body fluids or awake animals**



evotec

#RESEARCHNEVERSTOPS

Your contact:

Business Development
114 Innovation Drive, Milton Park, Abingdon
Oxfordshire OX14 4RZ, UK

T: +44.(0)1235.86 15 61
F: +44.(0)1235.86 31 39
info@evotec.com

2022/04 | V1