



Genome Assembly and Differential Expression of *Enterococcus faecium* E745

Paper I – RNA-seq and Tn-seq reveal fitness
determinants of vancomycin-resistant *Enterococcus*
faecium during growth in human serum (Zhang *et al.* 2017)

Overview

- 01 Introduction**
Background and aim of study
- 02 Preprocessing**
Quality control, trimming
- 03 Assembly**
Different assemblers
- 04 Annotation**
Finding genes

- 05 Synteny**
Aligning different species
- 06 Transcriptomics**
Mapping, counting, comparing RNA
- 07 Results**
- 08 Conclusions**

Enterococcus faecium

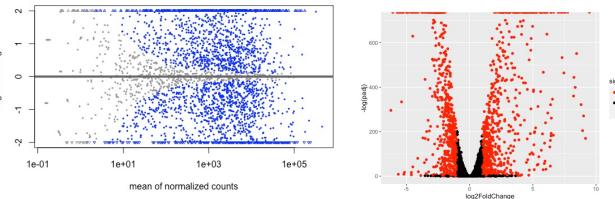
- Multi-resistant
- Hospital disease (clade A)
- Urinary tract and bloodstream
- Strain: E745 from rectal swab (Clade A-1, NL)



Photo: <https://www.indiamart.com/proddetail/enterococcus-faecium-probiotics-8411604530.html>

Aim

Identify genes responsible for growth in human serum



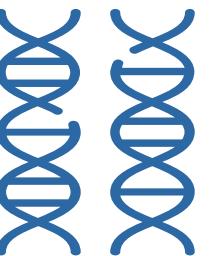
Data

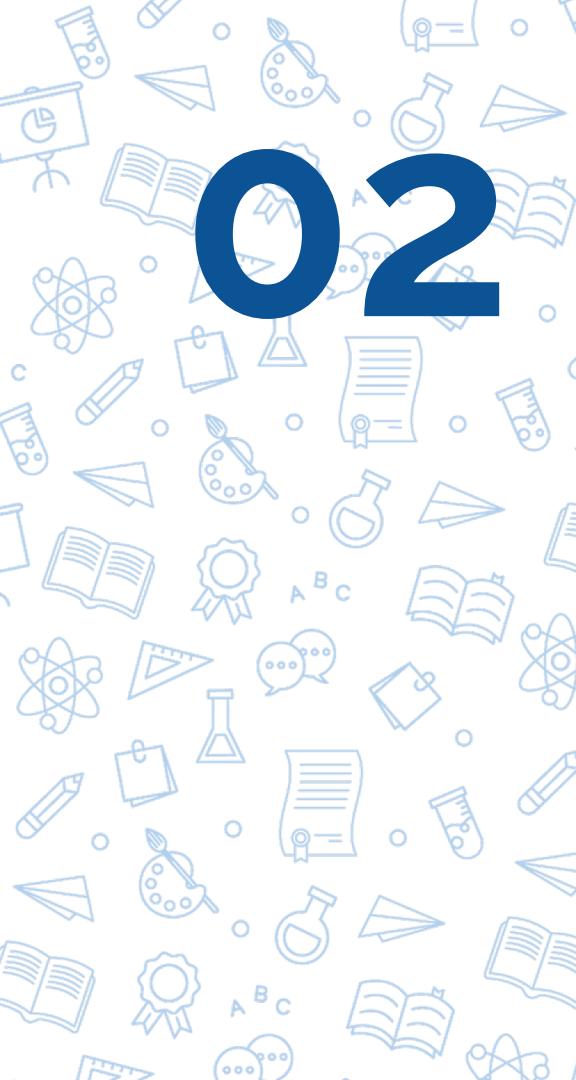
DNAseq from:

- PacBio RSII
- illumina HiSeq 2500
- NanoPore MinION 2D

RNAseq (and Tn) from
medium and serum:

- illumina HiSeq 2500





02

Preprocessing

Quality Control with
FastQC

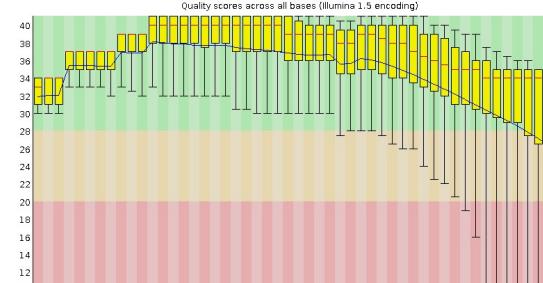
Trimming with
Trimmomatic

FastQC example

Illumina DNA read 1

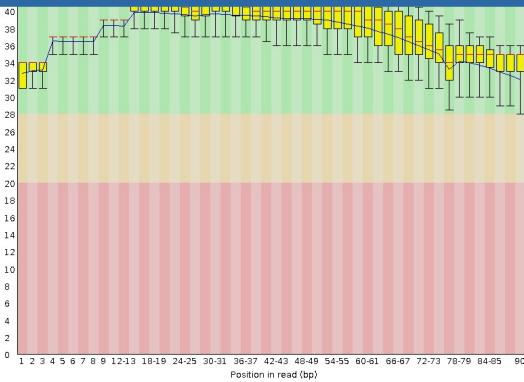


Illumina DNA read 2

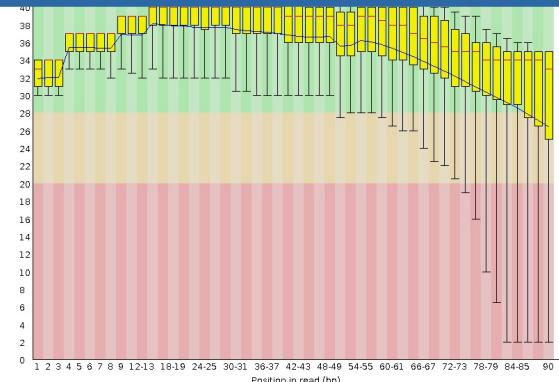


Before trimming

No significant difference – files already trimmed



After trimming





03

Assembly + evaluation

Assembly

Canu

PacBio

Spades

illumina
illumina + NanoPore
illumina + PacBio
illumina + NanoPore + PacBio

Pilon

Canu_PacBio

+

BWA aligned
illumina + NanoPore

Evaluation

BCF

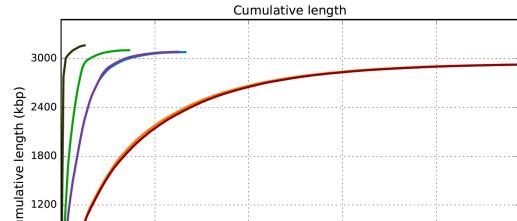
MUMmerPlot

QUAST

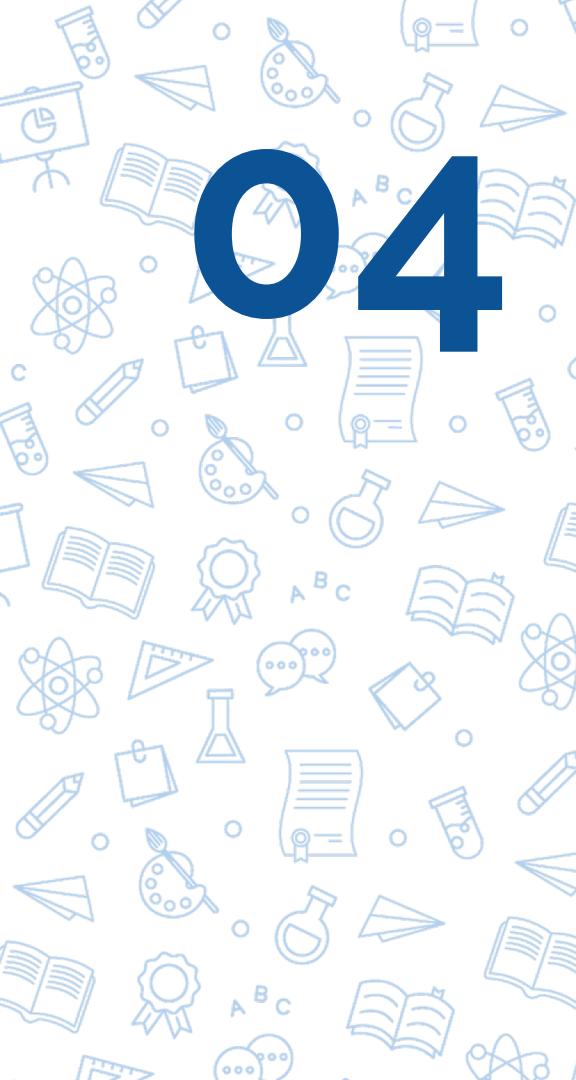
Article:

1 circular chromosome: 2.764Mb
6 plasmids: 9.3 - 223.7kbp
Total length: 3.168Mb
3095 predicted coding sequences

QUAST report



Column1	<input checked="" type="checkbox"/> illumina.contigs	<input type="checkbox"/> illumina_Nanopore.contigs	<input type="checkbox"/> PacBio.contigs	<input checked="" type="checkbox"/> Pilon_illumina_Nanopore_PacBio	<input type="checkbox"/> Pilon_illumina_PacBio	<input type="checkbox"/>
#contigs(>=0bp)	488	175	10			10
#contigs(>=1000bp)	163	46	10			10
#contigs(>=50000bp)	17	16	2			2
Total_length(>=0bp)	2975893	3101302	3162825		3162825	3162984
Total_length(>=1000bp)	2905560	3075336	3162825		3162825	3162984
Total_length(>=50000bp)	1402640	2724412	3006416		3006416	3006523
#contigs	195	53	10		10	10
Largest_contig	114283	458911	2775639		2775639	2775725
Total_length	2926792	3080395	3162825		3162825	3162984
GC(%)	37.65	37.64	37.77		37.77	37.77
N50	49767	162629	2775639		2775639	2775725
N75	19664	102479	2775639		2775639	2775725
L50	19	6	1		1	1
L75	43	11	1		1	1
Mismatches: #N's per 100k ^t	0.00	0.00	0.00		0.00	0.00



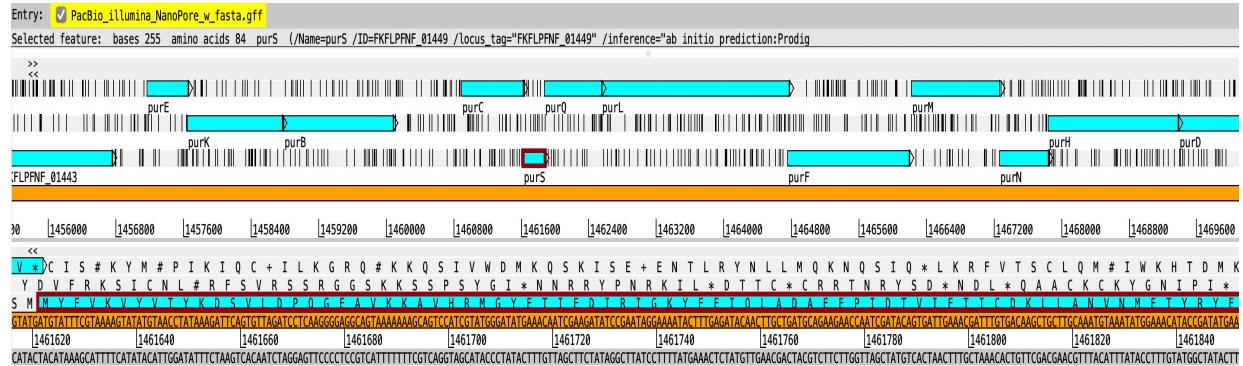
04

Annotation Prokka

Findings:

- rRNA
- tRNA
- tmRNA
- CDS (3126)

Prokka annotation



(Artemis)

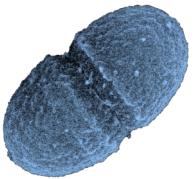


05

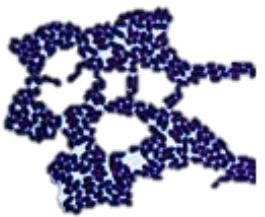
Synteny

Align, compare

Synteny – *E. faecium*



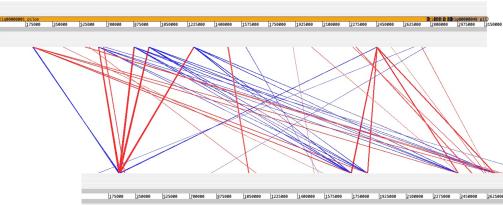
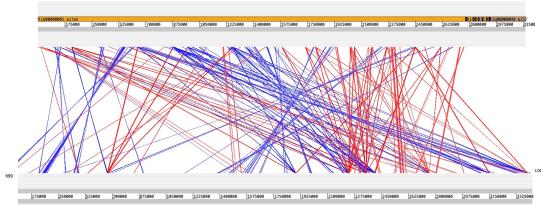
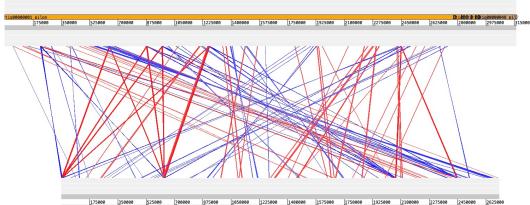
E. faecalis



E. casseliflavus



L. monocytogenes

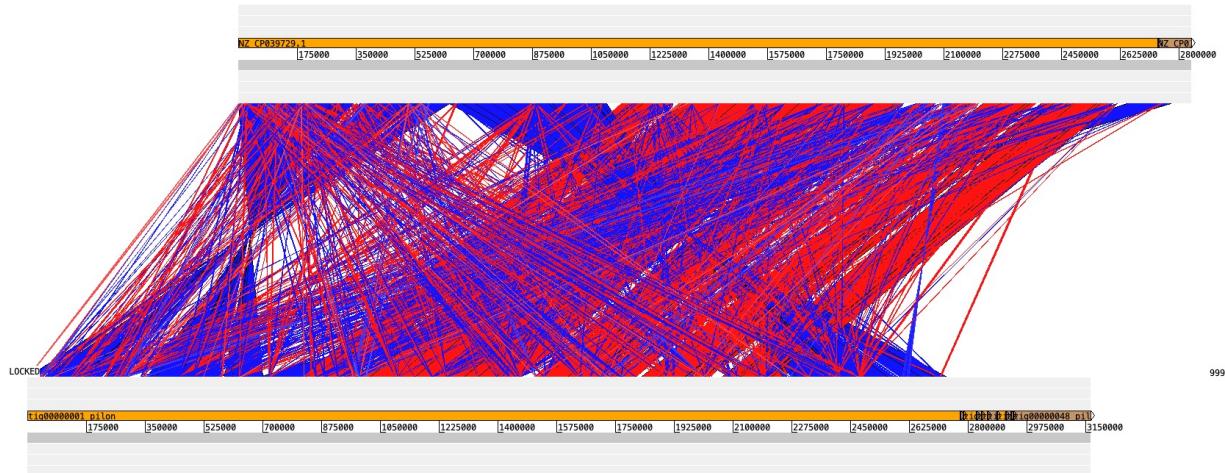


blastn alignment of whole genome. Visualized with Artemis ACT

(Photos: https://www.google.com/url?sa=i&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FEnterococcus_faecalis
<https://microbe-canvas.com/Bacteria.php?p=1239>
<https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.istockphoto.com%2Fphotos%2Flisteria-monocytogenes>)

Synteny

E. faecium E745 vs Reference E. faecium



blastn alignment of whole genome. Visualized with Artemis ACT

06

Transcriptomics

Mapping, counting reads, compare
E. faecium in Medium vs Serum

RNA seq

(Medium vs Serum)

BWA (index, mem)
Samtools (sort, index)

Mapping RNA reads
onto genome
assembly

HTSeq count

Count number of
transcripts mapped
to each gene

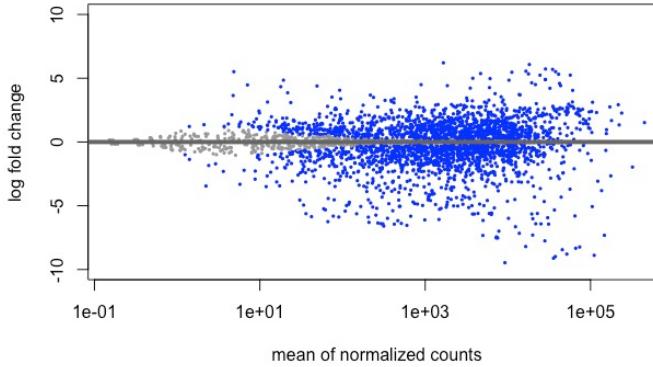
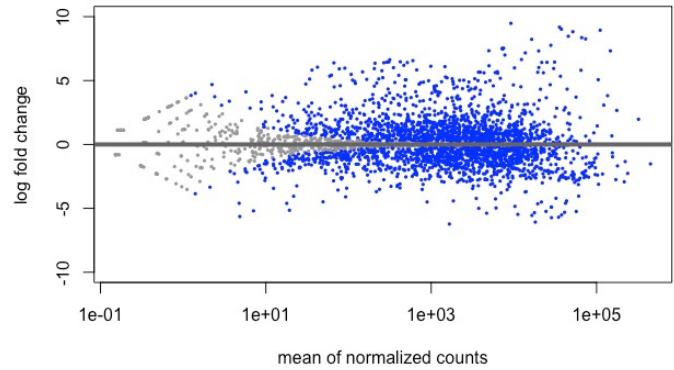
DeSeq2
RStudio

Differential
expression
Plots

07

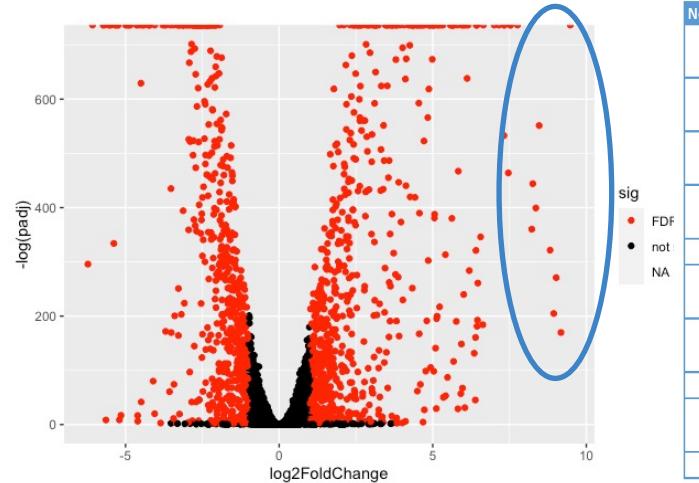
Results

DeSeq2



3216 transcription units. 1216 units (40%) were upregulated and 1212 units (40%) were down-regulated.
(RStudio plots. Log fold change vs LFC shrinkage)

DeSeq2



No.	Gene	Gene name	Product	log2FoldChange	pvalue
1	FKFLPFNF_01449	<i>purS</i>	Phosphoribosylformylglycinamide synthase subunit	9.47689	0.00000e+00
2	FKFLPFNF_01450	<i>purQ</i>	Phosphoribosylformylglycinamide synthase subunit	9.17443	4.36682e-75
3	FKFLPFNF_01448	<i>purC</i>	Phosphoribosylaminoimidazole-succinocarboxamide synthase	9.01904	3.63111e-119
4	FKFLPFNF_01451	<i>purL</i>	Phosphoribosylformylglycinamide synthase subunit	8.93882	2.35838e-90
5	FKFLPFNF_01452	<i>purF</i>	Amidophosphoribosyltransferase	8.82134	3.03795e-141
6	FKFLPFNF_01453	<i>purM</i>	Phosphoribosylformylglycinamide cyclo-ligase	8.46419	3.17345e-241
7	FKFLPFNF_01455	<i>purH</i>	Bifunctional purine biosynthesis protein	8.35599	4.26558e-175
8	FKFLPFNF_01456	<i>purD</i>	Phosphoribosylamine-glycine ligase	8.25687	1.39884e-194
9	FKFLPFNF_01454	<i>purN</i>	Phosphoribosylglycinamide formyltransferase	8.22430	4.60185e-158
10	FKFLPFNF_01510	-	Hypothetical protein	7.75820	0.00000e+00

(RStudio plot. After false rate discovery adjustment)

Example of Transposon effect

- Transposons in these genes: *pyrK*, *pyrF*, *purH*, *purD*, *manY*
- Showing number of mutant survivors after 24h in serum and medium
- Significant negative effect on viability in serum

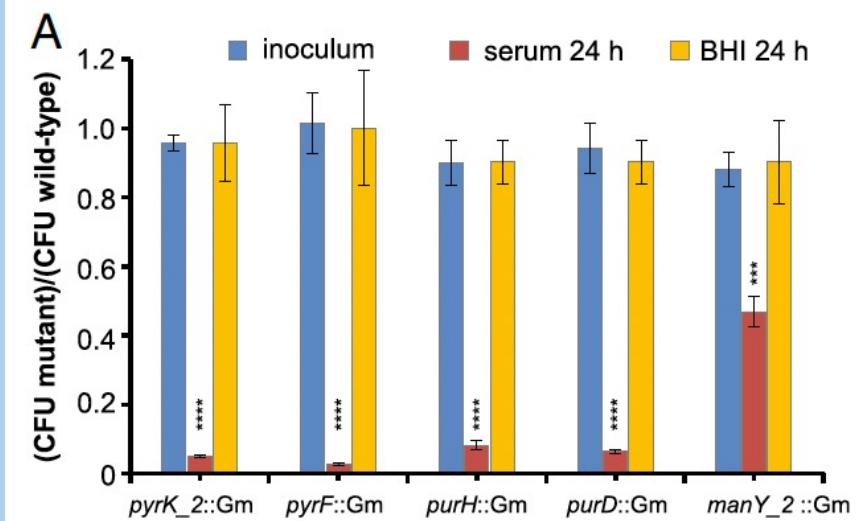
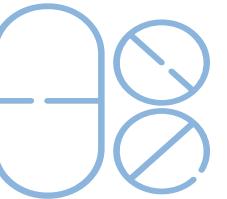


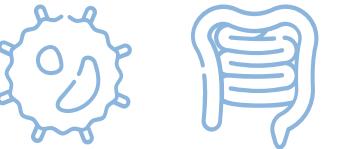
Diagram from Zhang et al. 2017

Conclusion



Good targets:

Genes involved in the nucleotide synthesis pathway:
purS, purQ, purC, purL, purF, purM, purH, purD, purN
(and carbohydrate metabolism manY, manZ, Tn-seq)



THANK YOU

References

- Zhang X, de Maat V, Guzmán Prieto AM, Prajsnar TK, Bayjanov JR, de Been M, Rogers MRC, Bonten MJM, Mesnage S, Willems RJL, van Schaik W. 2017. RNA-seq and Tn-seq reveal fitness determinants of vancomycin-resistant Enterococcus faecium during growth in human serum. BMC Genomics 18: 893.
- <https://canu.readthedocs.io/en/latest/tutorial.html#canu-the-command>
- <https://cab.spbu.ru/files/release3.14.0/manual.html>
- <https://support.illumina.com/bulletins/2016/04/fastq-files-explained.html>
- <http://bioconductor.org/packages/devel/bioc/vignettes/DESeq2/inst/doc/DESeq2.html#htseq-count-input>