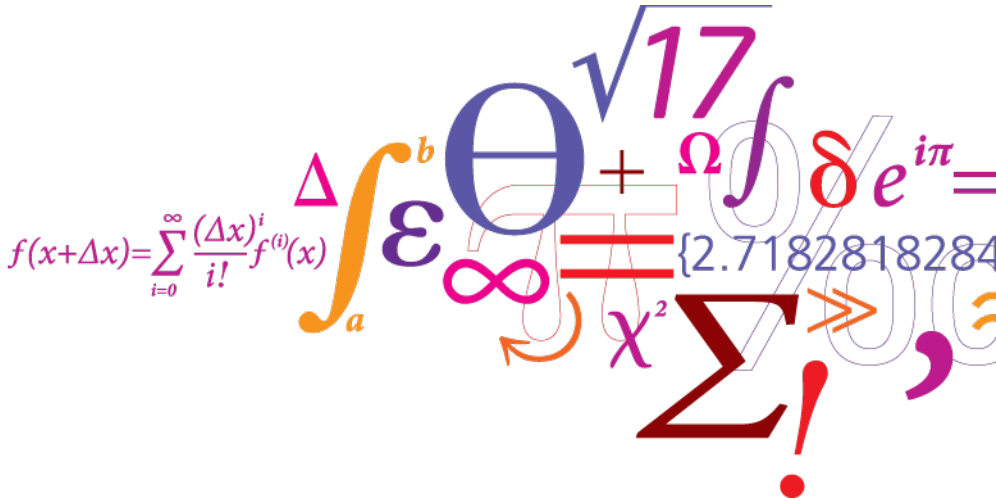
EURL-AR Training Course 2018

Exercise 1 – Revision and Discussion

28 September 2018



DTU Food
National Food Institute

Exercise 1 – Part 1

Objectives: Evaluate data quality and decide which data can be used for downstream analysis

Isolate ID	Expected species	Output from Quality Control pipline					
		Bases (MB)	Reads	N50	Number of contigs	Longest	Total BPs
Isolate A	Escherichia coli	994	8019656	288501	46	545016	4976587
Isolate B	Escherichia coli	405	2746488	254172	101	399093	5302962
Isolate C	Escherichia coli	385	2595090	173253	196	386355	10511210
Isolate D	Escherichia coli	1532	11072180	354248	57	629640	5060120
Isolate E	Escherichia coli	816	5875512	188239	66	556756	4779513
Isolate F	Escherichia coli	602	4292106	66436	1897	492856	7310556
Isolate G	Klebsiella pneumoniae	40	289816	2920	2346	29790	4994958
Isolate H	Yersinia enterocolitica	852	6069858	121701	73	245199	4636672
Isolate I	Staphylococcus aureus	134	921174	116189	51	232965	2809977
Isolate J	Staphylococcus aureus	151	1053706	135887	117	342404	5038817
Isolate K	Staphylococcus aureus	560	3935968	125616	43	504236	2746951
Isolate L	Staphylococcus aureus	446	3124582	539934	19	992722	2709295
Isolate M	Staphylococcus aureus	14	101884	1102	1501	20858	1545320
Isolate N	Streptococcus group C (Hemolytic)	185	1327182	73298	139	287028	4190334
Isolate O	Streptococcus group C (Hemolytic)	311	2292512	73298	58	182696	2093457
Isolate P	Enterococcus sp.	184	1331140	100659	220	449018	6889533

Isolate ID	Expected species	Output from Quality Control pipline					
		Bases (MB)	Reads	N50	Number of contigs	Longest	Total BPs
Isolate A	Escherichia coli	994	8019656	288501	46	545016	4976587
Isolate B	Escherichia coli	405	2746488	254172	101	399093	5302962
Isolate C	Escherichia coli	385	2595090	173253	196	386355	10511210
Isolate D	Escherichia coli	1532	11072180	354248	57	629640	5060120
Isolate E	Escherichia coli	816	5875512	188239	66	556756	4779513
Isolate F	Escherichia coli	602	4292106	66436	1897	492856	7310556
Isolate G	Klebsiella pneumoniae	40	289816	2920	2346	29790	4994958
Isolate H	Yersinia enterocolitica	852	6069858	121701	73	245199	4636672
Isolate I	Staphylococcus aureus	134	921174	116189	51	232965	2809977
Isolate J	Staphylococcus aureus	151	1053706	135887	117	342404	5038817
Isolate K	Staphylococcus aureus	560	3935968	125616	43	504236	2746951
Isolate L	Staphylococcus aureus	446	3124582	539934	19	992722	2709295
Isolate M	Staphylococcus aureus	14	101884	1102	1501	20858	1545320
Isolate N	Streptococcus group C (Hemolytic)	185	1327182	73298	139	287028	4190334
Isolate O	Streptococcus group C (Hemolytic)	311	2292512	73298	58	182696	2093457
Isolate P	Enterococcus sp.	184	1331140	100659	220	449018	6889533

High number of contigs with a low N50

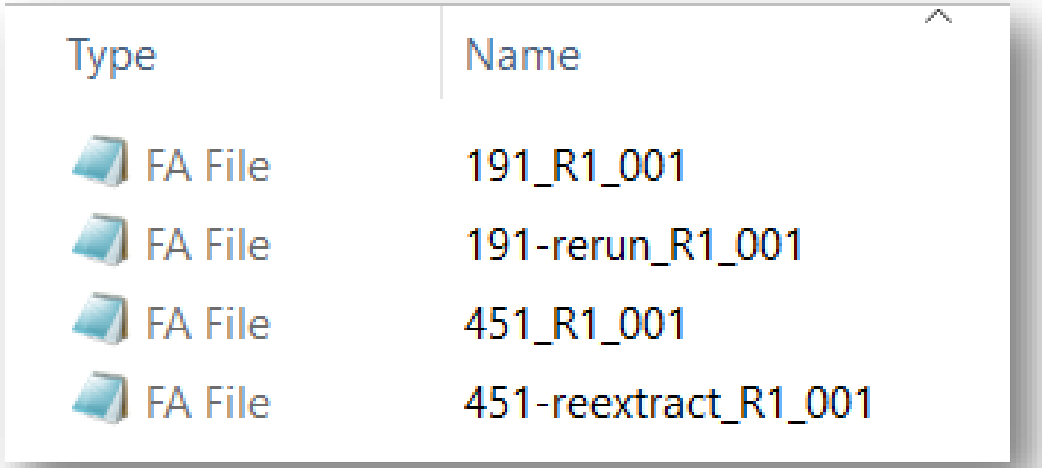
Unexpected genome size

Impact of data quality in results

Isolate ID	Expected species	Output from Quality Control pipline						Downstream analysis		
		Bases (MB)	Reads	N50	Nr. of contigs	Longest	Total BPs	Species	MLST	Resistance
Isolate C	Escherichia coli	385	2595090	173253	196	386355	10511210	Escherichia coli	ecoli[Unknown ST]	blaSHV-33,fosA-like,oqxA-like,oqxB-like
Isolate C - reextract	Escherichia coli	149	1082756	171241	90	452935	5117962	Escherichia coli	ecoli[ST-73]	
Isolate J	Staphylococcus aureus	151	1053706	135887	117	342404	5038817	Escherichia coli	ecoli[ST-155]	aadA1-like,blaOXA-1,floR-like,sul1,sul2
Isolate J - reextract	Staphylococcus aureus	145	1061786	105253	143	306865	5040601	Escherichia coli	ecoli[ST-155]	aadA1-like,blaOXA-1,floR-like,sul1,sul2
Isolate M	Staphylococcus aureus	14	101884	1102	1501	20858	1545320	Staphylococcus aureus	saureus[Unknown ST]	blaZ
Isolate M - rerun	Staphylococcus aureus	313	2332424	103842	92	232550	4634362	Yersinia enterocolitica	yersinia[ST-12]	vat(F)-like
Isolate M - rerun nr. 2	Staphylococcus aureus	475	2140740	671954	14	1132628	2695805	Staphylococcus aureus	saureus[ST-7]	blaZ,
Isolate H	Yersinia enterocolitica	852	6069858	121701	73	245199	4636672	Yersinia enterocolitica	yersinia[ST-12]	vat(F)-like

Exercise 1 - Part 2

Objectives: Understand that data quality will influence the downstream analysis of genome attributes.



Center for Genomic Epidemiology

MLST 2.0 (Multi-Locus Sequence Typing)

ResFinder 3.0

	Identification		Output from Quality Control pipline					
	ID	Expected species	Bases (MB)	Reads	N50	Number of contigs	Longest	Total BPs
Isolate G	191	Klebsiella pneumoniae	40	289816	2920	2346	29790	4994958
	191_rerun	Klebsiella pneumoniae	335	2311772	375077	46	832865	5490232
Isolate P	451	Enterococcus sp.	184	1331140	100659	220	449018	6889533
	451_reextract	Enterococcus sp.	540	2281928	94809	197	692220	3288819

Impact of data quality in results

Isolate 191

ResFinder Results – original sequence

ResFinder Results – rerun

MLST Results – original sequence

MLST Results - rerun

Isolate 451

ResFinder Results – original sequence

ResFinder Results – reextract

MLST Results – original sequence

MLST Results - reextract

ResFinder-3.0 Server - Results

Input Files: *191_R1_001.fa*

Show Acquired antimicrobial resistance results

Acquired antimicrobial resistance gene - Results

Aminoglycoside

No resistance genes found.

Beta-lactam

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
blaSHV-40	99.88	861/861	NODE_1116_length_1612_cov_0.739394	157..1017	Beta-lactam resistance	AF535128

Colistin

No resistance genes found.

Fluoroquinolone

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
oqxB	98.95	3153/2286	NODE_754_length_2286_cov_0.534044	1..2286	Quinolone resistance	EU370913
oqxA	99.23	1176/1176	NODE_165_length_5002_cov_0.762462	3009..4184	Quinolone resistance	EU370913

Fosfomycin

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
fosA	98.54	420/412	NODE_419_length_3254_cov_0.833067	2535..2946	Fosfomycin resistance	ACWO01000079

Fusidic Acid

No resistance genes found.

Glycopeptide

No resistance genes found.

MLS - Macrolide, Lincosamide and Streptogramin B

No resistance genes found.

Nitroimidazole

No resistance genes found.

Oxazolidinone

No resistance genes found.

Phenicol

No resistance genes found.

Rifampicin

No resistance genes found.

Sulphonamide

No resistance genes found.

Tetracycline

No resistance genes found.

Trimethoprim

No resistance genes found.

ResFinder-3.0 Server - Results

Input Files: *191-rerun_R1_001.fa*

Show Acquired antimicrobial resistance results

Acquired antimicrobial resistance gene - Results

Aminoglycoside

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
aph(6)-Id	100.00	837/837	NODE_31_length_4638_cov_1.636444	136..972		M28829
aph(3")-Ib	100.00	804/804	NODE_31_length_4638_cov_1.636444	972..1775		AF321551

Beta-lactam

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
blaSHV-40	99.88	861/861	NODE_2_length_669309_cov_4.564226	471090..471950	Beta-lactam resistance	AF535128
blaTEM-1B	100.00	861/861	NODE_35_length_2170_cov_1.549682	662..1522	Beta-lactam resistance	AY458016

Colistin

No resistance genes found.

Fluoroquinolone

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
oqxB	98.95	3153/3153	NODE_4_length_437966_cov_5.772371	421397..424549	Quinolone resistance	EU370913
oqxA	99.23	1176/1176	NODE_4_length_437966_cov_5.772371	424573..425748	Quinolone resistance	EU370913

Fosfomycin

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
fosA	98.54	420/412	NODE_5_length_375077_cov_5.547881	14046..14457	Fosfomycin resistance	ACWO01000079

Fusidic Acid

No resistance genes found.

Glycopeptide

No resistance genes found.

MLS - Macrolide, Lincosamide and Streptogramin B

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
mph(A)	100.00	906/906	NODE_29_length_6934_cov_2.188776	195..1100	Macrolide resistance	D16251

Nitroimidazole

No resistance genes found.

Oxazolidinone

No resistance genes found.

Phenicol

No resistance genes found.

Rifampicin

No resistance genes found.

Sulphonamide

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
sul2	100.00	816/816	NODE_31_length_4638_cov_1.636444	1836..2651	Sulphonamide resistance	HQ840942

Tetracycline

No resistance genes found.

Trimethoprim

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
dfrA14	100.00	474/474	NODE_29_length_6934_cov_2.188776	4806..5279	Trimethoprim resistance	KF921535

MLST-2.0 Server - Results

mlst Profile: *kpneumoniae*

Organism: *Klebsiella pneumoniae*

Sequence Type: *Unknown*

Nearest ST: *391*

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
gapA	100	100	450	450	0	gapA_10
infB	100	100	318	318	0	infB_5
mdh	100	86.3732	412	477	0	mdh_33?
pgi	100	50.2315	217	432	0	pgi_4?
phoE	100	100	420	420	0	phoE_6
rpoB	100	100	501	501	0	rpoB_4
tonB	100	100	414	414	0	tonB_4

Notes: ? alleles with less than 100% coverage found

? pgi: *Uncertain hit, ST can not be trusted.*

? mdh: *Uncertain hit, ST can not be trusted.*

MLST-2.0 Server - Results

mlst Profile: *kpneumoniae*

Organism: *Klebsiella pneumoniae*

Sequence Type: *391*

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
gapA	100	100	450	450	0	gapA_10
infB	100	100	318	318	0	infB_5
mdh	100	100	477	477	0	mdh_1
pgi	100	100	432	432	0	pgi_4
phoE	100	100	420	420	0	phoE_6
rpoB	100	100	501	501	0	rpoB_4
tonB	100	100	414	414	0	tonB_4

extended output

Input Files: *191-rerun_R1_001.fa*

ResFinder-3.0 Server - Results

Input Files: *451_R1_001.fa*

Show Acquired antimicrobial resistance results

Acquired antimicrobial resistance gene - Results

Aminoglycoside

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
aph(3')-III	100.00	795/795	NODE_56_length_37853_cov_2.928537	2435..3229	Aminoglycoside resistance	M26832
ant(6)-Ia	100.00	909/909	NODE_56_length_37853_cov_2.928537	3861..4769	Aminoglycoside resistance	AF330699

Beta-lactam

No resistance genes found.

Colistin

No resistance genes found.

Fluoroquinolone

No resistance genes found.

Fosfomycin

No resistance genes found.

Fusidic Acid

No resistance genes found.

Glycopeptide

No resistance genes found.

MLS - Macrolide, Lincosamide and Streptogramin B

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
lsa(A)	99.33	1497/1497	NODE_9_length_150319_cov_2.509794	92392..93888	Macrolide resistance	AY225127

Nitroimidazole

No resistance genes found.

Oxazolidinone

No resistance genes found.

Phenicol

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
cat	98.17	654/655	NODE_30_length_73258_cov_2.224966	69367..70020	Phenicol resistance	M11587

Rifampicin

No resistance genes found.

Sulphonamide

No resistance genes found.

Tetracycline

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
tet(M)	99.22	1920/1920	NODE_11_length_143639_cov_2.844626	107692..109611	Tetracycline resistance	AM990992
tet(J)	99.25	1197/1197	NODE_18_length_111773_cov_2.597299	48339..49535	Tetracycline resistance	ACLE01000065

Trimethoprim

No resistance genes found.

ResFinder-3.0 Server - Results

Input Files: *451-reextract_R1_001.fa*

Show Acquired antimicrobial resistance results

Acquired antimicrobial resistance gene - Results

Aminoglycoside

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
ant(6)-Ia	100.00	909/909	NODE_38_length_14191_cov_61.640572	8380..9288	Aminoglycoside resistance	AF330699
aph(3')-III	100.00	795/795	NODE_38_length_14191_cov_61.640572	9920..10714	Aminoglycoside resistance	M26832

Beta-lactam

No resistance genes found.

Colistin

No resistance genes found.

Fluoroquinolone

No resistance genes found.

Fosfomycin

No resistance genes found.

Fusidic Acid

No resistance genes found.

Glycopeptide

No resistance genes found.

MLS - Macrolide, Lincosamide and Streptogramin B

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
lsa(A)	99.33	1497/1497	NODE_2_length_248220_cov_60.996401	56432..57928	Macrolide resistance	AY225127

Nitroimidazole

No resistance genes found.

Oxazolidinone

No resistance genes found.

Phenicol

No resistance genes found.

Rifampicin

No resistance genes found.

Sulphonamide

No resistance genes found.

Tetracycline

Resistance gene	Identity	Query/HSP	Contig	Position in contig	Phenotype	Accession no.
tet(M)	99.22	1920/1920	NODE_7_length_98172_cov_73.813229	61007..62926	Tetracycline resistance	AM990992

Trimethoprim

No resistance genes found.

MLST-2.0 Server - Results

mlst Profile: *efaecalis*

Organism: *Enterococcus faecalis*

Sequence Type: *21*

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
aroE	100	100	459	459	0	aroE_1
gdh	100	100	530	530	0	gdh_1
gki	100	100	438	438	0	gki_1
gyd	100	100	395	395	0	gyd_7
pstS	100	100	583	583	0	pstS_9
xpt	100	100	456	456	0	xpt_1
yqiL	100	100	436	436	0	yqiL_1

extended output

Input Files: *451_R1_001.fa*

MLST-2.0 Server - Results

mlst Profile: *efaecalis*

Organism: *Enterococcus faecalis*

Sequence Type: *21*

Locus	Identity	Coverage	Alignment Length	Allele Length	Gaps	Allele
aroE	100	100	459	459	0	aroE_1
gdh	100	100	530	530	0	gdh_1
gki	100	100	438	438	0	gki_1
gyd	100	100	395	395	0	gyd_7
pstS	100	100	583	583	0	pstS_9
xpt	100	100	456	456	0	xpt_1
yqiL	100	100	436	436	0	yqiL_1

extended output

Input Files: *451-reextract_R1_001.fa*

Why do we see **less** AMR genes for the reextraction of 451?

Why is there **no change** in MLST results?

Conclusion

What do we conclude?

- **We need to be critical when analysing QC reports and look at several parameters**
- **We also need to consider the sequencing process as a whole (e.g. to detect cases of sample mix-up)**
- **We have to keep in mind that it is possible to obtain unexpected results (e.g. in cases where the preliminary species identification was wrong)**
- **Sequence quality will absolutely affect the downstream analysis and it is better to be safe than sorry.**

Thank you!

anrire@food.dtu.dk