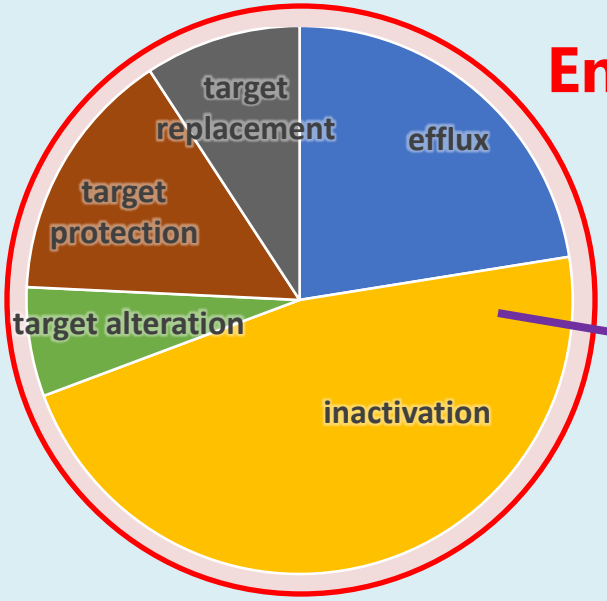
Normalization of metagenomic ARG profiles

Entire bacterial community (population unit: RPK-16S.)



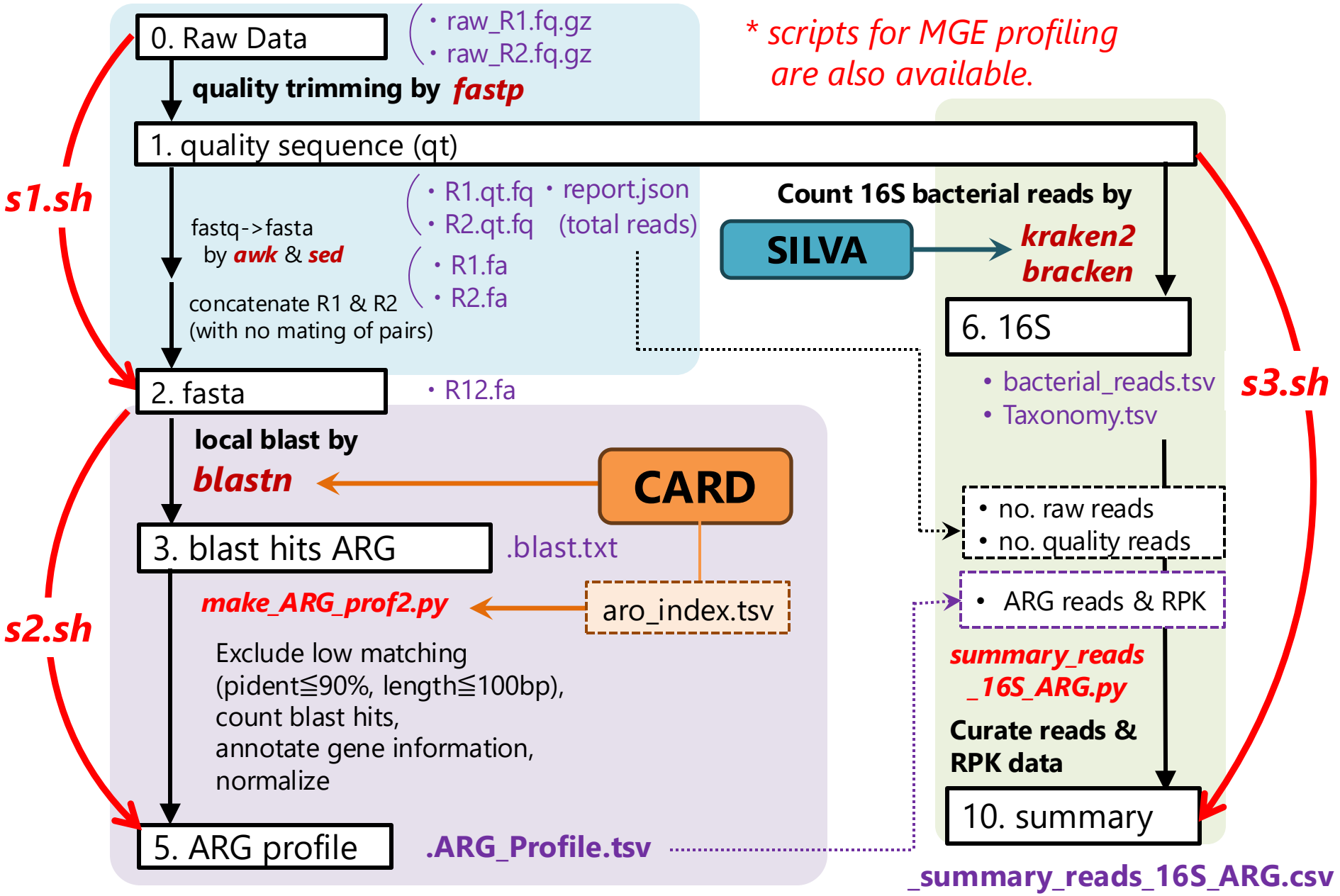
Entire ARGs
(unit: RPK)

ARG composition
(unit: proportion of RPK)

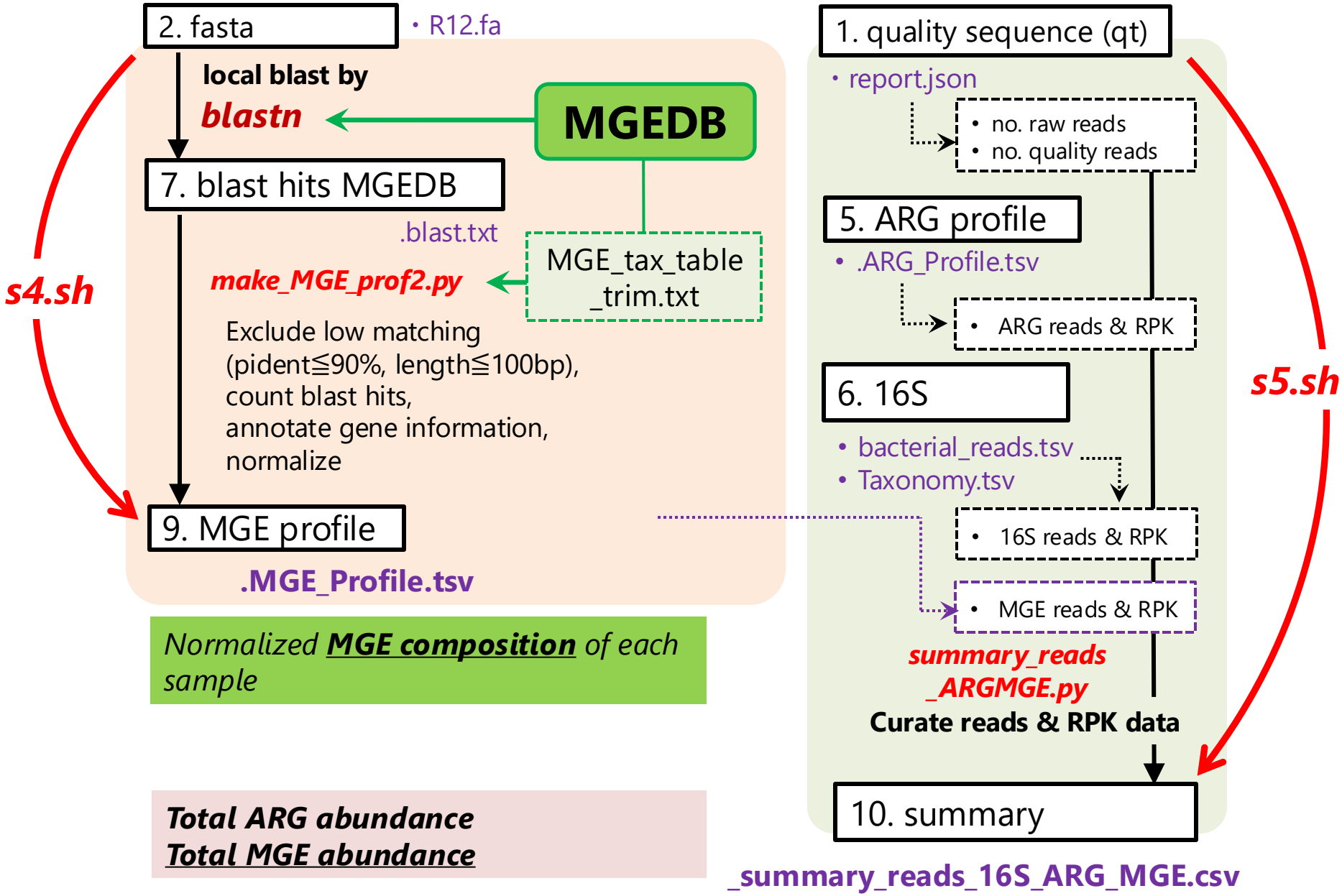
$$\text{ARG composition} = \frac{\text{Copies (RPK) of each ARG group}}{\text{Entire ARG copies (RPK)}}$$

$$\text{Total ARG abundance} = \frac{\text{Entire ARG copies (RPK)}}{\text{Entire bacterial population (RPK of 16S)}}$$

Analytical workflow of metagenomic AMR profiling



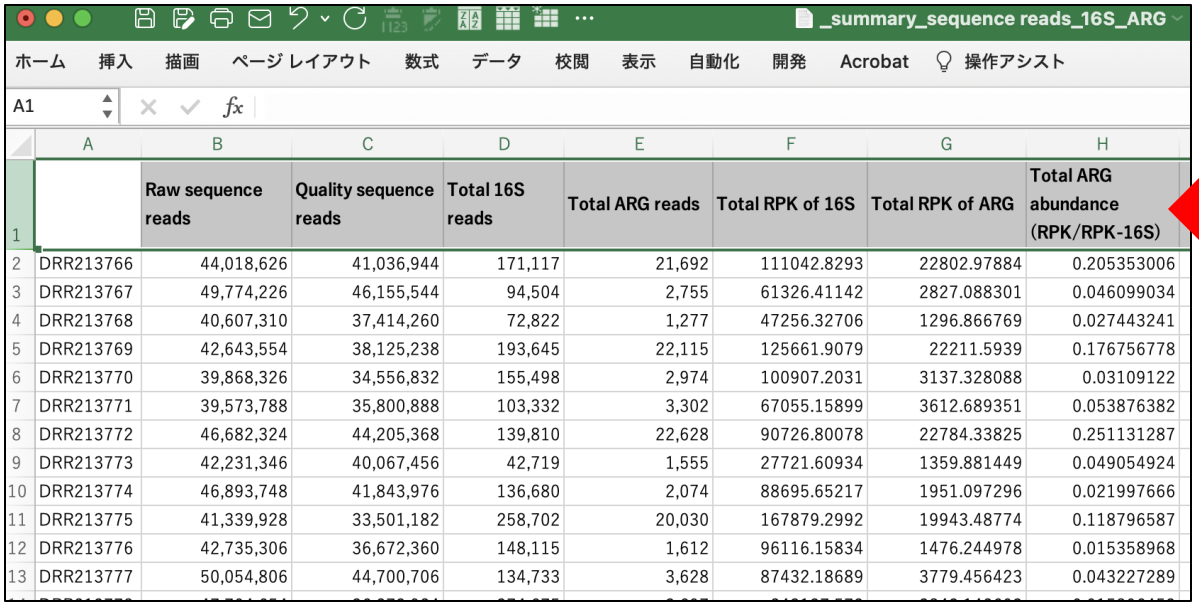
Analytical workflow of metagenomic MGE profiling



*Normalized **MGE composition** of each sample*

Total ARG abundance
Total MGE abundance

Output data of ARG & MGE profiles



_summary_sequence reads_16S_ARG

ホーム 挿入 描画 ページ レイアウト 数式 データ 校閲 表示 自動化 開発 Acrobat 操作アシスト

A1

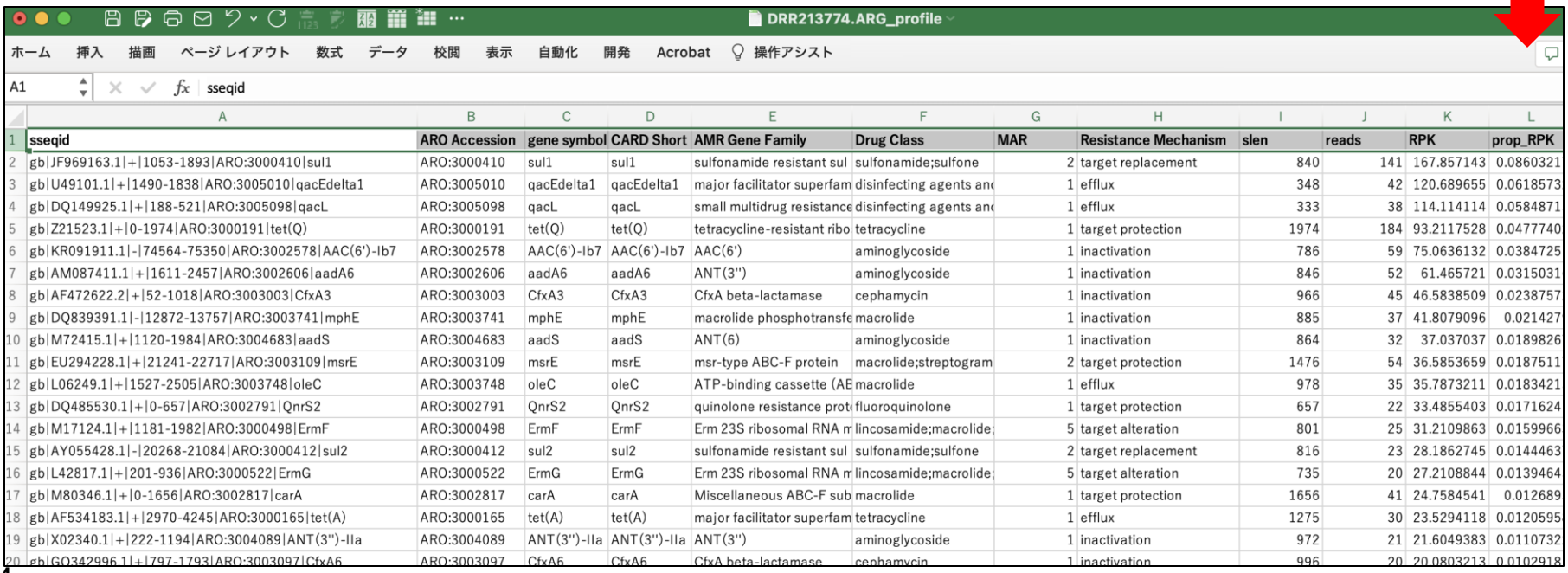
	A	B	C	D	E	F	G	H
1		Raw sequence reads	Quality sequence reads	Total 16S reads	Total ARG reads	Total RPK of 16S	Total RPK of ARG	Total ARG abundance (RPK/RPK-16S)
2	DRR213766	44,018,626	41,036,944	171,117	21,692	111042.8293	22802.97884	0.205353006
3	DRR213767	49,774,226	46,155,544	94,504	2,755	61326.41142	2827.088301	0.046099034
4	DRR213768	40,607,310	37,414,260	72,822	1,277	47256.32706	1296.866769	0.027443241
5	DRR213769	42,643,554	38,125,238	193,645	22,115	125661.9079	22211.5939	0.176756778
6	DRR213770	39,868,326	34,556,832	155,498	2,974	100907.2031	3137.328088	0.03109122
7	DRR213771	39,573,788	35,800,888	103,332	3,302	67055.15899	3612.689351	0.053876382
8	DRR213772	46,682,324	44,205,368	139,810	22,628	90726.80078	22784.33825	0.251131287
9	DRR213773	42,231,346	40,067,456	42,719	1,555	27721.60934	1359.881449	0.049054924
10	DRR213774	46,893,748	41,843,976	136,680	2,074	88695.65217	1951.097296	0.021997666
11	DRR213775	41,339,928	33,501,182	258,702	20,030	167879.2992	19943.48774	0.118796587
12	DRR213776	42,735,306	36,672,360	148,115	1,612	96116.15834	1476.244978	0.015358968
13	DRR213777	50,054,806	44,700,706	134,733	3,628	87432.18689	3779.456423	0.043227289

_summary_reads.csv

Total ARG abundance in all samples (RPK/RPK-**16S**)

ARG composition of each sample (RPK/RPK-**total ARG**)

.ARG_Profile.tsv



DRR213774.ARG_profile

ホーム 挿入 描画 ページ レイアウト 数式 データ 校閲 表示 自動化 開発 Acrobat 操作アシスト

A1 sseqid

	A	B	C	D	E	F	G	H	I	J	K	L
1	sseqid	ARO Accession	gene symbol	CARD Short	AMR Gene Family	Drug Class	MAR	Resistance Mechanism	slen	reads	RPK	prop_RPK
2	gb|JF969163.1|+|1053-1893|ARO:3000410|sul1	ARO:3000410	sul1	sul1	sulfonamide resistant sul	sulfonamide;sulfone	2	target replacement	840	141	167.857143	0.0860321
3	gb|U49101.1|+|1490-1838|ARO:3005010|qacEdelta1	ARO:3005010	qacEdelta1	qacEdelta1	major facilitator superfam	disinfecting agents an	1	efflux	348	42	120.689655	0.0618573
4	gb|DQ149925.1|+|188-521|ARO:3005098|qacL	ARO:3005098	qacL	qacL	small multidrug resistance	disinfecting agents an	1	efflux	333	38	114.114114	0.0584871
5	gb|Z21523.1|+|0-1974|ARO:3000191|tet(Q)	ARO:3000191	tet(Q)	tet(Q)	tetracycline-resistant ribo	tetracycline	1	target protection	1974	184	93.2117528	0.0477740
6	gb|KR091911.1|-|74564-75350|ARO:3002578|AAC(6')-Ib7	ARO:3002578	AAC(6')-Ib7	AAC(6')-Ib7	AAC(6')	aminoglycoside	1	inactivation	786	59	75.0636132	0.0384725
7	gb|AM087411.1|+|1611-2457|ARO:3002606|aadA6	ARO:3002606	aadA6	aadA6	ANT(3")	aminoglycoside	1	inactivation	846	52	61.465721	0.0315031
8	gb|AF472622.2|+|52-1018|ARO:3003003|CfxA3	ARO:3003003	CfxA3	CfxA3	CfxA beta-lactamase	cephamycin	1	inactivation	966	45	46.5838509	0.0238757
9	gb|DQ839391.1|-|12872-13757|ARO:3003741|mphE	ARO:3003741	mphE	mphE	macrolide phosphotransfe	macrolide	1	inactivation	885	37	41.8079096	0.021427
10	gb|M72415.1|+|1120-1984|ARO:3004683|aadS	ARO:3004683	aadS	aadS	ANT(6)	aminoglycoside	1	inactivation	864	32	37.037037	0.0189826
11	gb|EU294228.1|+|21241-22717|ARO:3003109|msrE	ARO:3003109	msrE	msrE	msr-type ABC-F protein	macrolide;streptogram	2	target protection	1476	54	36.5853659	0.0187511
12	gb|L06249.1|+|1527-2505|ARO:3003748|oleC	ARO:3003748	oleC	oleC	ATP-binding cassette (AE	macrolide	1	efflux	978	35	35.7873211	0.0183421
13	gb|DQ485530.1|+|0-657|ARO:3002791|QnrS2	ARO:3002791	QnrS2	QnrS2	quinolone resistance proti	fluoroquinolone	1	target protection	657	22	33.4855403	0.0171624
14	gb|M17124.1|+|1181-1982|ARO:3000498|ErmF	ARO:3000498	ErmF	ErmF	Erm 23S ribosomal RNA m	lincosamide;macrolide;	5	target alteration	801	25	31.2109863	0.0159966
15	gb|AY055428.1|-|20268-21084|ARO:3000412|sul2	ARO:3000412	sul2	sul2	sulfonamide resistant sul	sulfonamide;sulfone	2	target replacement	816	23	28.1862745	0.0144463
16	gb|L42817.1|+|201-936|ARO:3000522|ErmG	ARO:3000522	ErmG	ErmG	Erm 23S ribosomal RNA m	lincosamide;macrolide;	5	target alteration	735	20	27.2108844	0.0139464
17	gb|M80346.1|+|0-1656|ARO:3002817|carA	ARO:3002817	carA	carA	Miscellaneous ABC-F sub	macrolide	1	target protection	1656	41	24.7584541	0.012689
18	gb|AF534183.1|+|2970-4245|ARO:3000165|tet(A)	ARO:3000165	tet(A)	tet(A)	major facilitator superfam	tetracycline	1	efflux	1275	30	23.5294118	0.0120595
19	gb|X02340.1|+|222-1194|ARO:3004089|ANT(3")-IIa	ARO:3004089	ANT(3")-IIa	ANT(3")-IIa	ANT(3")	aminoglycoside	1	inactivation	972	21	21.6049383	0.0110732
20	gb|GQ342996.1|+|797-1793|ARO:3003097|CfxA6	ARO:3003097	CfxA6	CfxA6	CfxA beta-lactamase	cephamycin	1	inactivation	996	20	20.0803213	0.0102918

Scripts for data curation and analysis

金沢大学 KANAZAWA UNIVERSITY

crtARG.py ***crtMGE.py*** ***crt16S.py***

curate profile data to create comparison tables for:
compositions and further multivariate analysis

crtDrugClass.py

creates sample-comparison tables for:
abundance of each drug class

ARG.*.propRPK.csv

ARG.gene_symbol.prop_RP

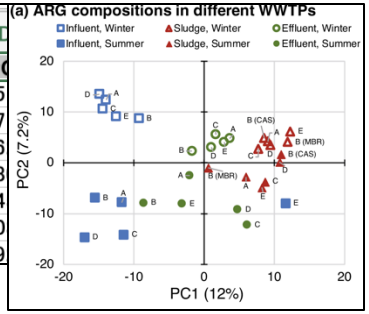
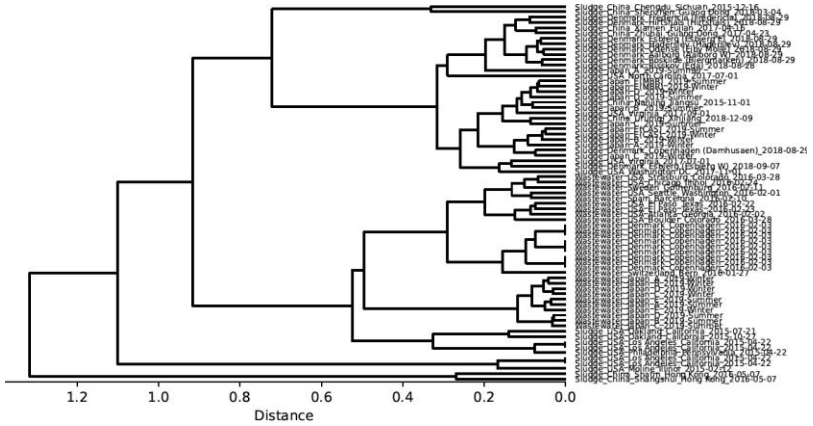
		sul1	OXA	aadA6	qacEdelta1	msrE	mphE
2	Wastewater_Japan_B_2019-Winter	0.07491711	0.06452201	0.01612124	0.06993944	0.05603556	0.0548
3	Sludge_Japan_B_2019-Winter	0.16338533	0.04669117	0.03135823	0.12908777	0.01509784	0.0115
4	Wastewater_Japan_C_2019-Winter	0.06474525	0.05623737	0.01000479	0.0582175	0.0585951	0.0512
5	Sludge_Japan_C_2019-Winter	0.16354529	0.10712291	0.02373615	0.14838016	0.06694458	0.0403
6	Wastewater_Japan_D_2019-Winter	0.06280662	0.05651053	0.01120634	0.05801747	0.06497485	0.0667
7	Sludge_Japan_D_2019-Winter	0.10505117	0.04270929	0.03389949	0.06550605	0.02690335	0.0282
8	Wastewater_Japan_E_2019-Winter	0.0690045	0.05318488	0.00924598	0.06901685	0.08801971	0.0773
9	Sludge_Japan_E(MBR)_2019-Winter	0.13064034	0.02931354	0.02642318	0.10121984	0.02615954	0.0344
10	Sludge_Japan_E(CAS)_2019-Winter	0.17450229	0.04039248	0.03158797	0.13153385	0.04230545	0.027

ARG.drug_class.per16S.xlsx

ARG.drug_class.per_16S

		aminocoumarin	aminoglycoside	acterial free fatty	bicyclomycin-like	carbapenem
2	Wastewater_Japan_B_2019-Winter	0.002099958	0.040502764	7.797E-06	0	0.023450
3	Sludge_Japan_B_2019-Winter	0.001222841	0.012677544	0	0	0.0044448
4	Wastewater_Japan_C_2019-Winter	0.00177174	0.030084906	0	6.58218E-06	0.0179363
5	Sludge_Japan_C_2019-Winter	0.000432452	0.006105121	0	0	0.004332
6	Wastewater_Japan_D_2019-Winter	0.002662921	0.045998838	1.44368E-05	3.64682E-05	0.0233103
7	Sludge_Japan_D_2019-Winter	0.001897183	0.010284338	0	8.9511E-05	0.0053995
8	Wastewater_Japan_E_2019-Winter	0.001139812	0.020289589	0	0	0.0105710
9	Sludge_Japan_E(MBR)_2019-Winter	0.000550718	0.003193026	0	0	0.0014940
10	Sludge_Japan_E(CAS)_2019-Winter	0.000726708	0.010938528	0	0	0.0037288
11	Wastewater_Japan_A_2019-Winter	0.002670011	0.059503734	9.663E-06	1.84628E-05	0.0386482

pca_hc_scale.py

performs **principal component analysis**
and **hierarchic cluster analysis** to
compare the resistome among samples.

		PC1	PC2	PC
2	Wastewater_Japan_B_2019-Winter	7.086841	18.77344	-3.5
3	Sludge_Japan_B_2019-Winter	-8.04637	1.861994	2.37
4	Wastewater_Japan_C_2019-Winter	7.999758	15.79485	-2.6
5	Sludge_Japan_C_2019-Winter	-6.42381	1.859793	0.08
6	Wastewater_Japan_D_2019-Winter	9.525514	17.83392	-2.4
7	Sludge_Japan_D_2019-Winter	-10.8172	3.585288	6.90
8	Wastewater_Japan_E_2019-Winter	5.064808	11.46496	-1.9