

Supplementary Materials for Aggarwal *et al.*

Supplementary Tables

Name	Sequence
V1FW_S D501	AATGATACGGCGACCACCGAGATCTACACAAGCAGCAacactctttccctacacgacgct cttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V1FW_S D502	AATGATACGGCGACCACCGAGATCTACACACGCGTGAacactctttccctacacgacgct cttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V1FW_S D503	AATGATACGGCGACCACCGAGATCTACACCGATCTACacactctttccctacacgacgctc ttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V1FW_S D504	AATGATACGGCGACCACCGAGATCTACACTGCGTCACacactctttccctacacgacgct cttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V1FW_S D505	AATGATACGGCGACCACCGAGATCTACACGTCTAGTGacactctttccctacacgacgctc ttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V1FW_S D506	AATGATACGGCGACCACCGAGATCTACACCTAGTATGacactctttccctacacgacgctc ttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V1FW_S D507	AATGATACGGCGACCACCGAGATCTACACGATAGCGTAcactctttccctacacgacgct cttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V1FW_S D508	AATGATACGGCGACCACCGAGATCTACACTCTACTacactctttccctacacgacgctc ttccgatctNNNNAGMGTTYGATYMTGGCTCAG
V2RV_S D701	CAAGCAGAAGACGGCATACGAGATACCTAGTAgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D702	CAAGCAGAAGACGGCATACGAGATACGTACGTgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D703	CAAGCAGAAGACGGCATACGAGATATATCGCGgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D704	CAAGCAGAAGACGGCATACGAGATCACGATAGgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D705	CAAGCAGAAGACGGCATACGAGATCGTATCGCGgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D706	CAAGCAGAAGACGGCATACGAGATCTGCGACTgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D707	CAAGCAGAAGACGGCATACGAGATGCTGTAACgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D708	CAAGCAGAAGACGGCATACGAGATGGACGTTAgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D709	CAAGCAGAAGACGGCATACGAGATGGTCGTAGgtgactggagttcagacgtgtgctcttccg atctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D710	CAAGCAGAAGACGGCATACGAGATTAAGTCTCgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D711	CAAGCAGAAGACGGCATACGAGATTACACAGTgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT
V2RV_S D712	CAAGCAGAAGACGGCATACGAGATTTGACGCAgtgactggagttcagacgtgtgctcttccga tctNNNNGCTGCCTCCCGTAGGAGT

Supplementary Table S1: Full length Illumina tagged primers used in study

Original taxa	BLAST assigned taxa	OTU
<i>Solanum melongena</i> (eggplant)	<i>Moraxella osloensis</i>	7278
<i>Ambystoma mexicanum</i> (axolotl)	<i>Acinetobacter lwoffii</i>	16873
gamma proteobacterium enrichment culture clone JF9-3	<i>Pseudomonas stutzeri</i>	7869
<i>Ceuthophilus</i> sp.	<i>Pseudomonas</i> sp.	17786
bacterium enrichment culture clone Guo10	<i>Acinetobacter</i> sp.	16545
uncultured gamma proteobacterium	<i>Acinetobacter johnsonii</i>	16234
uncultured gamma proteobacterium	<i>Acinetobacter</i> sp.	6386
<i>Coregonus lavaretus</i> (common whitefish)	<i>Cutibacterium acnes</i>	1449
uncultured gamma proteobacterium	<i>Pseudomonas</i> sp.	15694
uncultured alpha proteobacterium	<i>Agrobacterium</i> sp.	2447
<i>Athetis lepigone</i>	<i>Pseudomonas putida</i>	15973
uncultured gamma proteobacterium	<i>Pantoea agglomerans</i>	19430
bacterium enrichment culture clone Guo10	<i>Acinetobacter johnsonii</i>	16549
<i>Triticum aestivum</i> (bread wheat)	<i>Pantoea agglomerans</i>	1713
Cyanobacteria/Melainabacteria group bacterium S15B-MN24 CBMW_12	bacterium	2448
<i>Triticum aestivum</i> (bread wheat)	<i>Pseudomonas</i> sp.	14880
uncultured gamma proteobacterium	<i>Acinetobacter</i> sp.	22396
uncultured organism	<i>Streptococcus</i> sp.	18567
<i>Bacteroidetes</i> endosymbiont of <i>Metaseiulus occidentalis</i>	<i>Sphingobacterium multivorum</i>	18113
uncultured Xanthomonadales bacterium	<i>Stenotrophomonas</i> sp.	24764
uncultured <i>Kocuria</i> sp.	<i>Kocuria</i> sp.	11727
uncultured <i>Magnetococcus</i> sp.	<i>Magnetococcus</i> sp.	17306
uncultured organism	<i>Morganella morganii</i>	13974
uncultured <i>Stenotrophomonas</i> sp.	<i>Stenotrophomonas</i> sp.	14445
<i>Solanum melongena</i> (eggplant)	<i>Moraxella osloensis</i>	7664
uncultured <i>Frigoribacterium</i> sp.	<i>Frigoribacterium</i> sp.	8729
uncultured gamma proteobacterium	<i>Acinetobacter</i> sp.	22692
<i>Serratia plymuthica</i> RVH1	<i>Serratia plymuthica</i>	12905
uncultured organism	<i>Streptococcus</i> sp.	9583
uncultured gamma proteobacterium	<i>Pseudomonas</i> sp.	14885
uncultured organism	<i>Streptococcus</i> sp.	18518
uncultured organism	<i>Streptococcus</i> sp.	18519
uncultured proteobacterium	<i>Moraxella osloensis</i>	7312

<i>uncultured organism</i>	<i>Stenotrophomonas sp.</i>	24912
<i>Kocuria palustris</i> PEL	<i>Kocuria palustris</i>	4869
<i>Antarctic seawater bacterium Bsw10170</i>	<i>Psychrobacter sp.</i>	8875
<i>bacterium enrichment culture clone Guo10</i>	<i>Acinetobacter sp.</i>	6395
<i>uncultured organism</i>	<i>Streptococcus sp.</i>	18569
<i>uncultured organism</i>	<i>Streptococcus sp.</i>	9603
<i>uncultured organism</i>	<i>Morganella morganii</i>	13970
<i>Lepisosteus oculatus</i> (spotted gar)	<i>Cutibacterium sp.</i>	26069
<i>halophilic and alkaliphilic bacterium HN24</i>	<i>Alteribacillus sp.</i>	4290
<i>uncultured Bacteroidetes bacterium</i>	<i>Sphingobacterium multivorum</i>	8736
<i>uncultured organism</i>	<i>bacterium</i>	26105
<i>Trichuris trichiura</i> (human whipworm)	<i>bacterium</i>	9363
<i>Lepisosteus oculatus</i> (spotted gar)	<i>bacterium</i>	11552
<i>gamma proteobacterium symbiont of Graphosoma rubrolineatum</i>	<i>Pantoea sp.</i>	19435
<i>uncultured organism</i>	<i>Streptococcus sp.</i>	10002
<i>uncultured gamma proteobacterium</i>	<i>Pseudomonas sp.</i>	7868
<i>uncultured organism</i>	<i>Segatella copri</i>	11683
<i>uncultured beta proteobacterium</i>	<i>Massilia sp.</i>	15297
<i>Sargassum henslowianum</i>	<i>Pseudomonas sp.</i>	15771
<i>Type-C symbiont of Plautia stali</i>	<i>Pantoea dispersa</i>	11877
<i>Frankliniella occidentalis</i> (western flower thrips)	<i>Erwinia sp.</i>	20685
<i>uncultured organism</i>	<i>Haemophilus parainfluenzae</i>	13484
<i>uncultured organism</i>	<i>Lactobacillus gasseri</i>	4438
<i>uncultured organism</i>	<i>Staphylococcus sp.</i>	5125
<i>uncultured organism</i>	<i>Staphylococcus aureus</i>	14222
<i>unidentified</i>	<i>Streptococcus sp.</i>	6344
<i>Ceuthophilus sp.</i>	<i>Pseudomonas oleovorans</i>	23177
<i>uncultured organism</i>	<i>bacterium</i>	18471
<i>uncultured organism</i>	<i>Escherichia coli</i>	26623
<i>bacterium 1A5</i>	<i>Chryseobacterium oncorhynchi</i>	1979
<i>uncultured organism</i>	<i>Stenotrophomonas maltophilia</i>	17898
<i>uncultured organism</i>	<i>Streptococcus sp.</i>	9588
<i>uncultured microorganism</i>	<i>Helcococcus sp.</i>	8325
<i>uncultured alpha proteobacterium</i>	<i>Agrobacterium sp.</i>	26581
<i>Triticum urartu</i>	<i>bacterium</i>	4985
<i>Trichuris trichiura</i> (human whipworm)	<i>bacterium</i>	9366

<i>unidentified marine bacterioplankton</i>	<i>Salmonella enterica</i>	21359
<i>human gut metagenome</i>	<i>bacterium</i>	3703
<i>uncultured organism</i>	<i>Streptococcus sp.</i>	4119
<i>uncultured gamma proteobacterium</i>	<i>Acinetobacter radioresistens</i>	22712
<i>uncultured organism</i>	<i>Staphylococcus sp.</i>	5119
<i>uncultured organism</i>	<i>bacterium</i>	17984
<i>uncultured organism</i>	<i>bacterium</i>	17911
<i>Coregonus lavaretus (common whitefish)</i>	<i>Cutibacterium acnes</i>	20637
<i>uncultured organism</i>	<i>Neisseria sp.</i>	26003
<i>uncultured organism</i>	<i>Lactobacillus crispatus</i>	3520
<i>Triticum aestivum (bread wheat)</i>	<i>bacterium</i>	11971
<i>Teleogryllus commodus</i>	<i>Pseudomonas sp.</i>	24751
<i>uncultured organism</i>	<i>Staphylococcus aureus</i>	15046
<i>uncultured organism</i>	<i>Streptococcus sp.</i>	18501
<i>Trichuris trichiura (human whipworm)</i>	<i>bacterium</i>	9368
<i>unidentified marine bacterioplankton</i>	<i>bacterium</i>	12169
<i>Solanum torvum</i>	<i>Paracoccus sp.</i>	12792
<i>halophilic and alkaliphilic bacterium HN24</i>	<i>Dolosigranulum pigrum</i>	14546
<i>Coregonus lavaretus (common whitefish)</i>	<i>Cutibacterium acnes</i>	1457
<i>Ambystoma mexicanum (axolotl)</i>	<i>Acinetobacter lwoffii</i>	26268
<i>Solanum melongena (eggplant)</i>	<i>Moraxella osloensis</i>	26387
<i>Solanum melongena (eggplant)</i>	<i>Moraxella osloensis</i>	7286
<i>uncultured organism</i>	<i>Stenotrophomonas sp.</i>	25790
<i>uncultured alpha proteobacterium</i>	<i>Agrobacterium sp.</i>	26410
<i>Teleogryllus commodus</i>	<i>Enterobacteriaceae bacterium</i>	4745
<i>halophilic and alkaliphilic bacterium HN24</i>	<i>Alteribacillus sp.</i>	14551
<i>uncultured organism</i>	<i>Acinetobacter sp.</i>	15599
<i>uncultured organism</i>	<i>Stenotrophomonas sp.</i>	25791
<i>uncultured eubacterium WD264</i>	<i>bacterium</i>	6823
<i>halophilic and alkaliphilic bacterium HN24</i>	<i>Alteribacillus sp.</i>	21604
<i>uncultured organism</i>	<i>bacterium</i>	9586
<i>uncultured organism</i>	<i>bacterium</i>	17934
<i>uncultured organism</i>	<i>Morganella morganii</i>	21474
<i>bacterium enrichment culture clone Guo10</i>	<i>Acinetobacter sp.</i>	26388
<i>Ceuthophilus sp.</i>	<i>Pseudomonas sp.</i>	26487
<i>uncultured gamma proteobacterium</i>	<i>Pseudomonas sp.</i>	15975
<i>uncultured organism</i>	<i>Granulicatella adiacens</i>	6828
<i>Solanum melongena (eggplant)</i>	<i>Moraxella osloensis</i>	26075

<i>Eogystia hippophaecolus</i>	<i>Acinetobacter guillouiae</i>	25143
<i>Ambystoma mexicanum</i> (axolotl)	<i>Acinetobacter lwoffii</i>	26305
uncultured organism	<i>Morganella morganii</i>	13077
<i>Ceuthophilus</i> sp.	<i>Pseudomonas</i> sp.	26560
<i>Ambystoma mexicanum</i> (axolotl)	<i>Acinetobacter lwoffii</i>	26024
<i>Coregonus lavaretus</i> (common whitefish)	<i>Cutibacterium acnes</i>	26118
uncultured cyanobacterium	bacterium	26032
uncultured gamma proteobacterium	<i>Acinetobacter johnsonii</i>	22462
<i>Ambystoma mexicanum</i> (axolotl)	<i>Acinetobacter lwoffii</i>	16338
bacterium enrichment culture clone Guo10	<i>Acinetobacter</i> sp.	6394
uncultured organism	<i>Staphylococcus</i> sp.	10024
unidentified marine bacterioplankton	<i>Lactobacillales</i> group	3609
BTEX-degrading bacterium enrichment culture	<i>Stenotrophomonas maltophilia</i>	66
uncultured organism	bacterium	17943
halophilic and alkaliphilic bacterium HN24	<i>Alteribacillus</i> sp.	21605
<i>Trichuris trichiura</i> (human whipworm)	bacterium	18332
uncultured compost bacterium	<i>Corynebacterium</i> sp.	934
uncultured gamma proteobacterium	<i>Acinetobacter johnsonii</i>	22448
<i>Ceuthophilus</i> sp.	<i>Pseudomonas</i> sp.	26431
<i>Ambystoma mexicanum</i> (axolotl)	<i>Acinetobacter lwoffii</i>	26026
uncultured organism	bacterium	10382
uncultured organism	bacterium	17988
<i>Solanum melongena</i> (eggplant)	<i>Moraxella osloensis</i>	26025
<i>Ambystoma mexicanum</i> (axolotl)	<i>Acinetobacter lwoffii</i>	26584
uncultured gamma proteobacterium	<i>Acinetobacter johnsonii</i>	26005
gamma proteobacterium enrichment culture clone JF9-3	<i>Pseudomonas stutzeri</i>	26035
uncultured organism	<i>Staphylococcus</i> sp.	14223
uncultured gamma proteobacterium	<i>Acinetobacter</i> sp.	26554
<i>Solanum melongena</i> (eggplant)	<i>Moraxella osloensis</i>	25996
uncultured gamma proteobacterium	<i>Acinetobacter</i> sp.	16057
human gut metagenome	<i>Faecalibacterium prausnitzii</i>	8225
<i>Triticum aestivum</i> (bread wheat)	<i>Pantoea agglomerans</i>	26244
<i>Ambystoma mexicanum</i> (axolotl)	<i>Acinetobacter lwoffii</i>	26100
<i>Mycobacterium tuberculosis</i>	<i>Rothia mucilaginosa</i>	4908
<i>Mycobacterium tuberculosis</i>	<i>Rothia mucilaginosa</i>	4903

Supplementary Table S2: Mis-assigned taxa by ARB. A few operational taxonomic units were evaluated with BLAST where the species was discordant with the taxa genus and/or not a prokaryote, as assigned by ARB, or where an

uncultured bacterium was seen across the mock microbial community suggesting it was either an expected mock microbial community member or a significant contaminant.

Species	Low DNA Qubit R1	Low DNA Qubit R2	Extraction Location Associated	PCR Location Associated	Known Contaminant Correlation R1	Known Contaminant Correlation R2	Decontam (neg control)	Biological Plausibility	Contaminant
<i>S. bongori</i>	No	No	NA	NA	NA	NA	Possible	No	Yes
<i>S. aureus</i>	No	No	No	No	Yes	No	Possible	Yes	No
<i>D. pigrum</i>	No	No	No	No	No	No	Possible	Yes	No
<i>Corynebacterium sp.</i>	No	No	No	No	No	No	Possible	Yes	No
<i>C. pseudodiphtheriticum</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>C. jeikeium</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>C. accolens</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>C. koseri</i>	No	No	No	No	Yes	No	Possible	Yes	No
<i>K. michiganensis</i>	No	No	No	No	No	No	Possible	Yes	No
<i>S. epidermidis</i>	No	Yes	No	No	No	No	Unlikely	Yes	No
<i>K. pneumoniae</i>	No	No	No	No	Yes	No	Unlikely	Yes	No
<i>B. cereus</i>	No	No	No	No	Yes	No	Possible	No	Yes
<i>K. oxytoca</i>	No	No	No	No	Yes*	No	Possible	Yes	No
<i>Acinetobacter sp.</i>	No	No	No	No	Yes	Yes	Possible	No	Yes
<i>M. catarrhalis</i>	No	No	No	No	No	No	Possible	Yes	No
<i>K. aerogenes</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>A. ursingii</i>	No	No	No	No	Yes	Yes*	Possible	No	Yes
<i>Acinetobacter sp.</i>	No	No	No	No	Yes	Yes	Possible	No	Yes
<i>E. coli</i>	No	No	No	No	Yes	Yes	Possible	Yes	No
<i>C. freundii</i>	No	No	No	Yes	No	Yes	Possible	Yes	No
<i>Pseudomonas sp.</i>	No	No	No	Yes	Yes	Yes*	Possible	No	Yes
<i>M. osloensis</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>Psychrobacter sp.</i>	No	No	No	No	Yes*	No	Possible	No	Yes
<i>A. lwoffii</i>	No	No	No	No	Yes*	Yes	Possible	No	Yes
<i>M. nonliquefaciens</i>	No	No	No	No	Yes	No	Possible	Yes	No
<i>Rhodococcus sp.</i>	No	No	No	No	Yes	No	Unlikely	No	Yes

<i>Enhydrobacter sp.</i>	No	No	Yes	No	No	Yes	Unlikely	No	Yes
<i>Alkanindiges sp.</i>	No	No	No	No	No	No	Unlikely	No	Yes
<i>P. stutzeri</i>	No	No	No	No	Yes	Yes	Possible	No	Yes
<i>S. marcescens</i>	No	No	No	No	No	No	Possible	Yes	No
<i>M. lincolnii</i>	No	No	No	No	Yes	Yes	Unlikely	Yes	No
<i>Chryseobacterium sp.</i>	No	No	No	Yes	No	Yes	Possible	No	Yes
<i>Ralstonia sp.</i>	No	Yes	No	Yes	Yes	Yes*	Unlikely	No	Yes
<i>Pseudomonas sp.</i>	No	No	No	Yes	Yes	Yes*	Possible	No	Yes
<i>Enterobacter sp.</i>	No	No	No	Yes	Yes*	Yes	Possible	No	Yes
<i>A. johnsonii</i>	No	No	No	No	Yes*	Yes	Unlikely	No	Yes
<i>Stenotrophomonas sp.</i>	No	No	No	No	Yes*	Yes*	Possible	No	Yes
<i>P. mirabilis</i>	No	No	No	No	Yes	No	Unlikely	No	Yes
<i>Magnetococcus sp.</i>	No	No	No	No	Yes	No	Unlikely	No	Yes
<i>E. faecalis</i>	No	No	No	No	Yes	Yes	Unlikely	No	Yes
<i>P. coleopterorum</i>	No	No	No	No	Yes	No	Possible	No	Yes
<i>A. parvus</i>	No	No	No	No	No	Yes	Unlikely	No	Yes
<i>A. johnsonii</i>	No	No	No	No	Yes*	Yes	Unlikely	No	Yes
<i>M. lacunata</i>	No	No	No	No	No	Yes	Unlikely	Yes	No
<i>P. proteolytica</i>	No	No	No	No	Yes	Yes	Unlikely	No	Yes
<i>P. gergoviae</i>	No	No	No	No	Yes	Yes	Possible	No	Yes
<i>R. mucilaginoso</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>S. haemolyticus</i>	No	No	No	No	Yes	No	Unlikely	Yes	No
<i>S. maltophilia</i>	No	No	Yes	No	Yes	Yes*	Possible	No	Yes
<i>E. cloacae</i>	No	No	No	No	Yes	Yes	Unlikely	No	Yes
<i>S. sciuri</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>P. agglomerans</i>	No	No	No	No	Yes	Yes	Unlikely	No	Yes
<i>P. aeruginosa</i>	No	No	Yes	Yes	Yes*	Yes	Possible	No	Yes
<i>Streptococcus sp.</i>	No	No	No	No	Yes	Yes	Unlikely	Yes	No
<i>P. putida</i>	No	No	No	No	Yes*	Yes	Possible	No	Yes
<i>Klebsiella sp.</i>	No	No	No	No	Yes*	Yes	Unlikely	Yes	No

<i>M. morganii</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>C. acnes</i>	No	Yes	No	No	Yes	Yes	Unlikely	Yes	No
<i>P. oryzihabitans</i>	No	No	NA	NA	No	Yes	Unlikely	No	Yes
<i>M. morganii</i>	No	No	No	No	No	Yes	Unlikely	Yes	No
<i>P. frederiksbergensis</i>	No	No	No	No	Yes	Yes	Unlikely	No	Yes
<i>R. terrigena</i>	No	No	No	No	No	Yes	Possible	No	Yes
<i>S. dysgalactiae</i>	No	No	No	No	No	No	Unlikely	Yes	No
<i>Stenotrophomonas</i> <i>sp.</i>	No	No	No	No	Yes*	Yes*	Possible	No	Yes
<i>Moraxella sp.</i>	No	No	No	No	No	No	Unlikely	Yes	No

Supplementary Table S3: Determination of possible contaminants from low biomass anterior nares' samples.

Species	Comparison	Log-fold change	Standard error	adjusted p-value
<i>Corynebacterium accolens</i>	Persistent ~ Non-carrier	-0.95	0.30	0.03
<i>Corynebacterium jeikeium</i>	Persistent ~ Non-carrier	-1.40	0.29	<0.001
<i>Corynebacterium sp.</i>	Persistent ~ Non-carrier	-1.66	0.28	<0.001
<i>Dolosigranulum pigrum</i>	Persistent ~ Non-carrier	-1.63	0.29	<0.001
<i>Moraxella catarrhalis</i>	Persistent ~ Non-carrier	-0.64	0.21	0.04
<i>Staphylococcus aureus</i>	Persistent ~ Non-carrier	5.50	0.24	<0.001
<i>Staphylococcus aureus</i>	Intermittent ~ Non-carrier	1.45	0.29	<0.001
<i>Staphylococcus epidermidis</i>	Persistent ~ Non-carrier	-0.75	0.23	0.02

Supplementary Table S4: Differential abundance of species by nasal colonisation status using ANCOM-BC2 displaying significant associations. Log fold (natural log) changes as compared to *Staphylococcus aureus* non-carriers. Log-fold change between the compared groups shown with associated standard error and adjusted p-values. P-values adjusted with holm method.

Intercept	Odds Ratio	Lower C.I	Upper C.I.
CST II	0.59	0.37	0.95
CST III	0.86	0.56	1.32
CST IV	0.64	0.40	1.02
CST V	0.22	0.11	0.43
CST VI	0.50	0.30	0.84
CST VII	2.00	1.41	2.84
Sex	Odds Ratio	Lower C.I	Upper C.I.
CST II	0.63	0.38	1.04
CST III	0.70	0.45	1.09
CST IV	0.69	0.42	1.14
CST V	1.73	0.88	3.42
CST VI	0.53*	0.30	0.93
CST VII	0.67*	0.47	0.96
Current Smoker	Odds Ratio	Lower C.I	Upper C.I.
CST II	0.59	0.16	2.17
CST III	0.86	0.31	2.42
CST IV	0.60	0.16	2.24
CST V	1.00	0.21	4.76
CST VI	0.79	0.21	2.94
CST VII	0.73	0.31	1.70
Pets	Odds Ratio	Lower C.I	Upper C.I.
CST II	1.13	0.68	1.86
CST III	0.87	0.55	1.36
CST IV	0.94	0.56	1.56
CST V	0.68	0.35	1.35
CST VI	1.08	0.62	1.87
CST VII	0.93	0.65	1.33
Healthcare contact	Odds Ratio	Lower C.I	Upper C.I.
CST II	0.87	0.39	1.94
CST III	1.07	0.54	2.10
CST IV	0.71	0.30	1.67
CST V	0.86	0.28	2.67

CST VI	0.77	0.31	1.90
CST VII	0.90	0.52	1.58
Chronic Skin Condition	Odds Ratio	Lower C.I	Upper C.I.
CST II	1.00	0.45	2.20
CST III	1.09	0.55	2.15
CST IV	0.79	0.34	1.86
CST V	1.30	0.49	3.44
CST VI	0.89	0.36	2.19
CST VII	1.21	0.70	2.08
Asthma	Odds Ratio	Lower C.I	Upper C.I.
CST II	1.05	0.41	2.69
CST III	0.88	0.38	2.07
CST IV	1.28	0.52	3.15
CST V	0.00	0.00	2.32e+201
CST VI	0.76	0.24	2.38
CST VII	0.62	0.30	1.27
Allergy	Odds Ratio	Lower C.I	Upper C.I.
CST II	0.73	0.42	1.24
CST III	0.77	0.48	1.23
CST IV	0.69	0.40	1.19
CST V	0.73	0.36	1.46
CST VI	0.79	0.44	1.43
CST VII	0.90	0.62	1.30

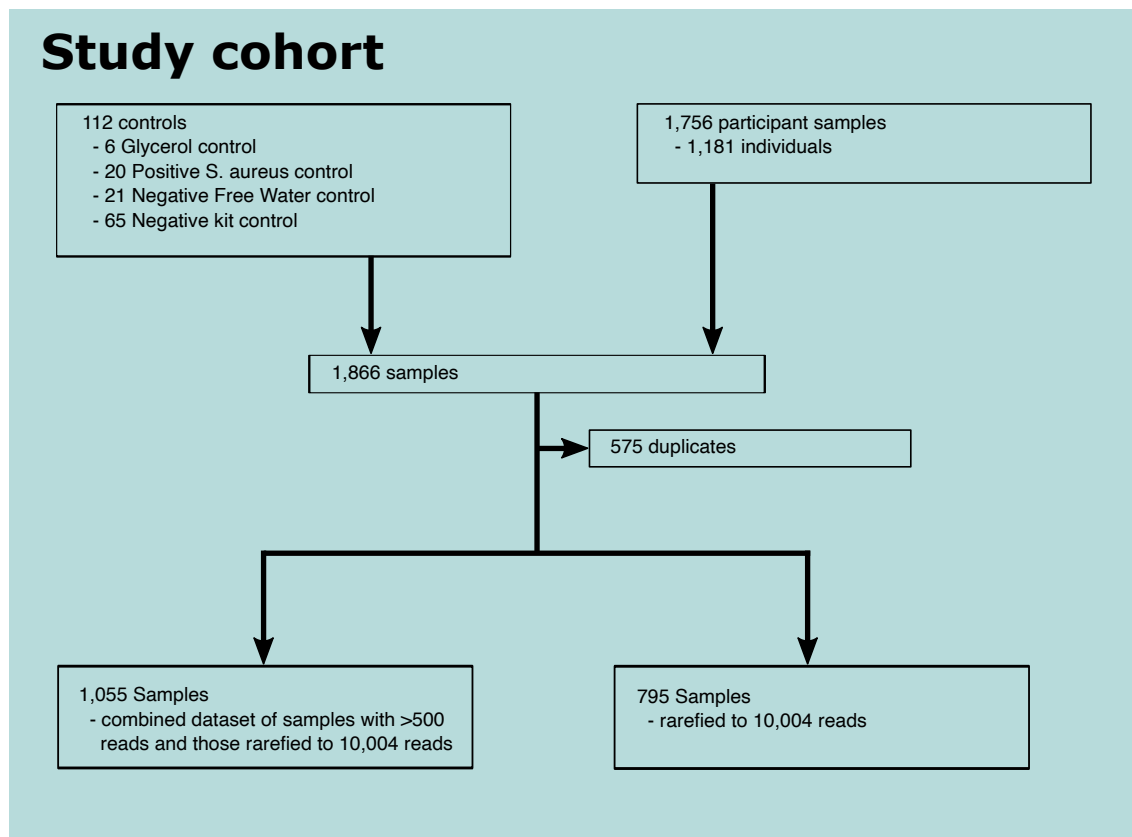
Supplementary Table S5: Multinomial logistic regression model of major lifestyle and clinical risk factors for *S. aureus* Community State Types. Odds ratios are presented with 95% Confidence Intervals (C.I.). We found men had a reduced relative risk for association with CST VI (OR = 0.53, 95% CI = 0.30-0.93, p=0.03) and CST VII (OR=0.67, 95% CI= 0.47-0.96, p=0.03) compared with CST1. * p<0.05, ** p<0.01, *** p<0.001.

Taxon	Group	Eigenvector value (2 d.p.)
<i>Serratia marcescens</i>	0	5.09e-17
<i>Morganella morganii</i> subsp.	0	1.96e-17
<i>Klebsiella aerogenes</i>	0	0
<i>Dolosigranulum pigrum</i>	1	1
<i>Moraxella catarrhalis</i>	1	0.71
<i>Corynebacterium jeikeium</i>	1	0.70
<i>Corynebacterium pseudodiphtheriticum</i>	1	0.66
<i>Corynebacterium accolens</i>	1	0.54
<i>Moraxella nonliquefaciens</i>	1	0.34
<i>Moraxella lincolnia</i>	1	0.32
<i>Klebsiella oxytoca</i>	1	0.23
<i>Staphylococcus aureus</i>	2	0.79
<i>Corynebacterium</i> sp.	2	0.69
<i>Cutibacterium acnes</i>	2	0.64
<i>Staphylococcus epidermidis</i>	2	0.62
<i>Rothia mucilaginosa</i>	2	0.42
<i>Moraxella osloensis</i>	2	0.28
<i>Moraxella</i> sp.	2	0.06
<i>Citrobacter freundii</i>	3	0.07
<i>Klebsiella michiganensis</i>	3	0.01
<i>Escherichia coli</i>	3	0.01
<i>Moraxella lacunata</i>	3	0.01
<i>Klebsiella pneumoniae</i>	4	9.88e-17
<i>Citrobacter koseri</i>	4	5.38e-17

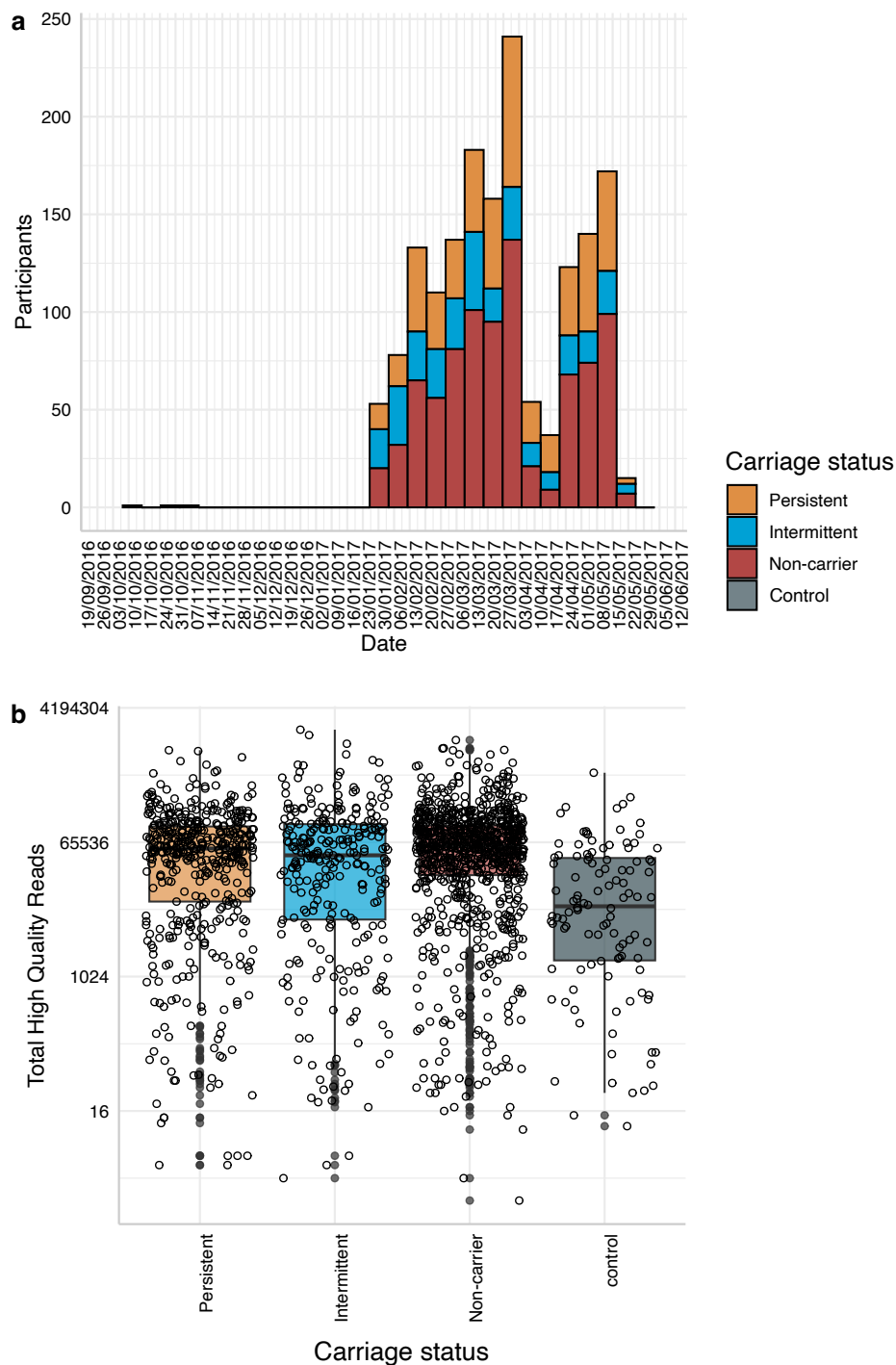
Supplementary Table S6: Eigenvector centrality values from network analysis.

Species-level networks were inferred with NetCoMi (v1.2) using SparCC correlations (zeroes replaced by a pseudocount; centred log-ratio transform; 1000 bootstraps), then analysed for centrality and community structure via fast-greedy clustering

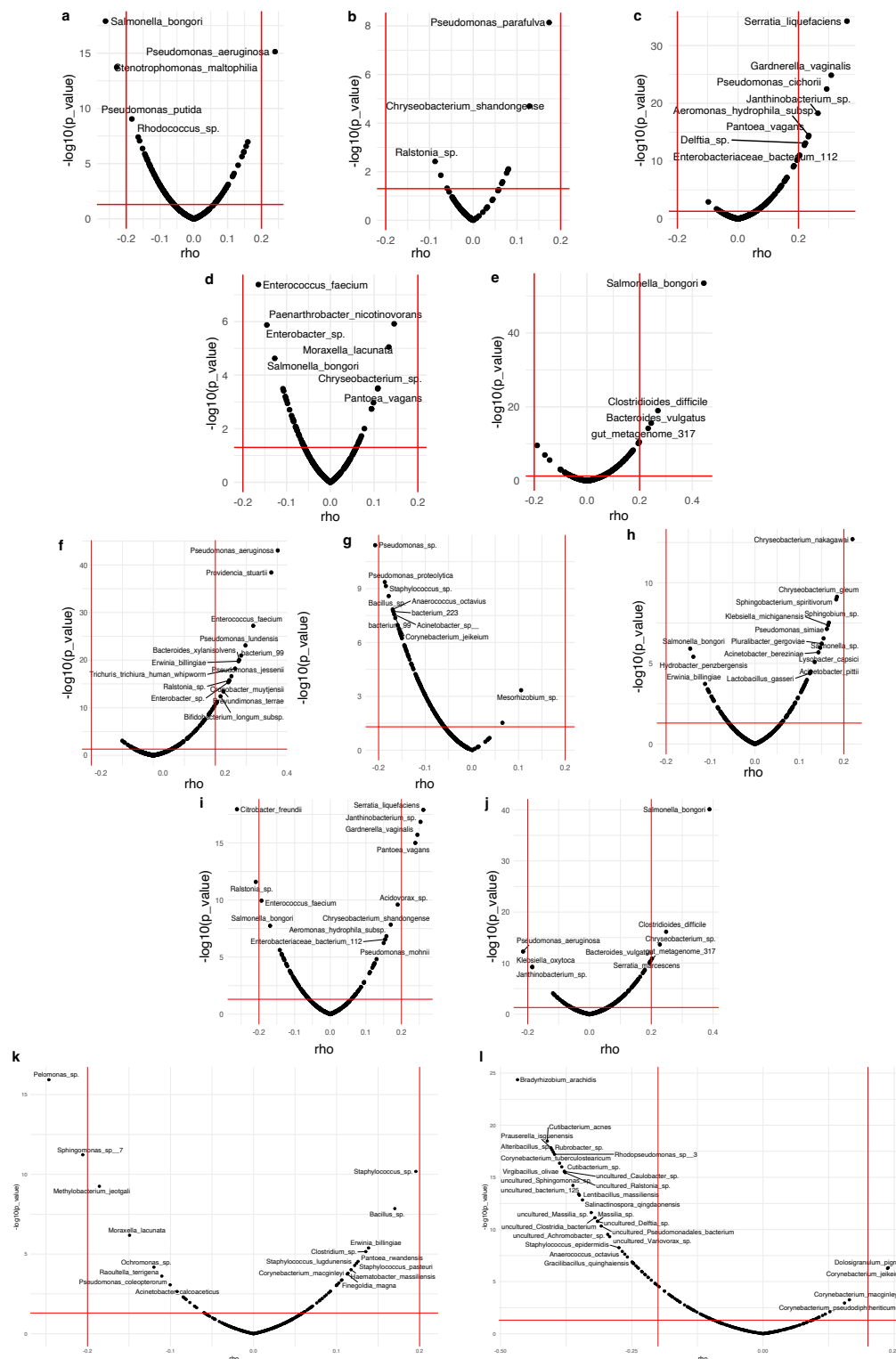
Supplementary Figures



Supplementary Fig. S1: Study cohort and available 16S rRNA sequences.

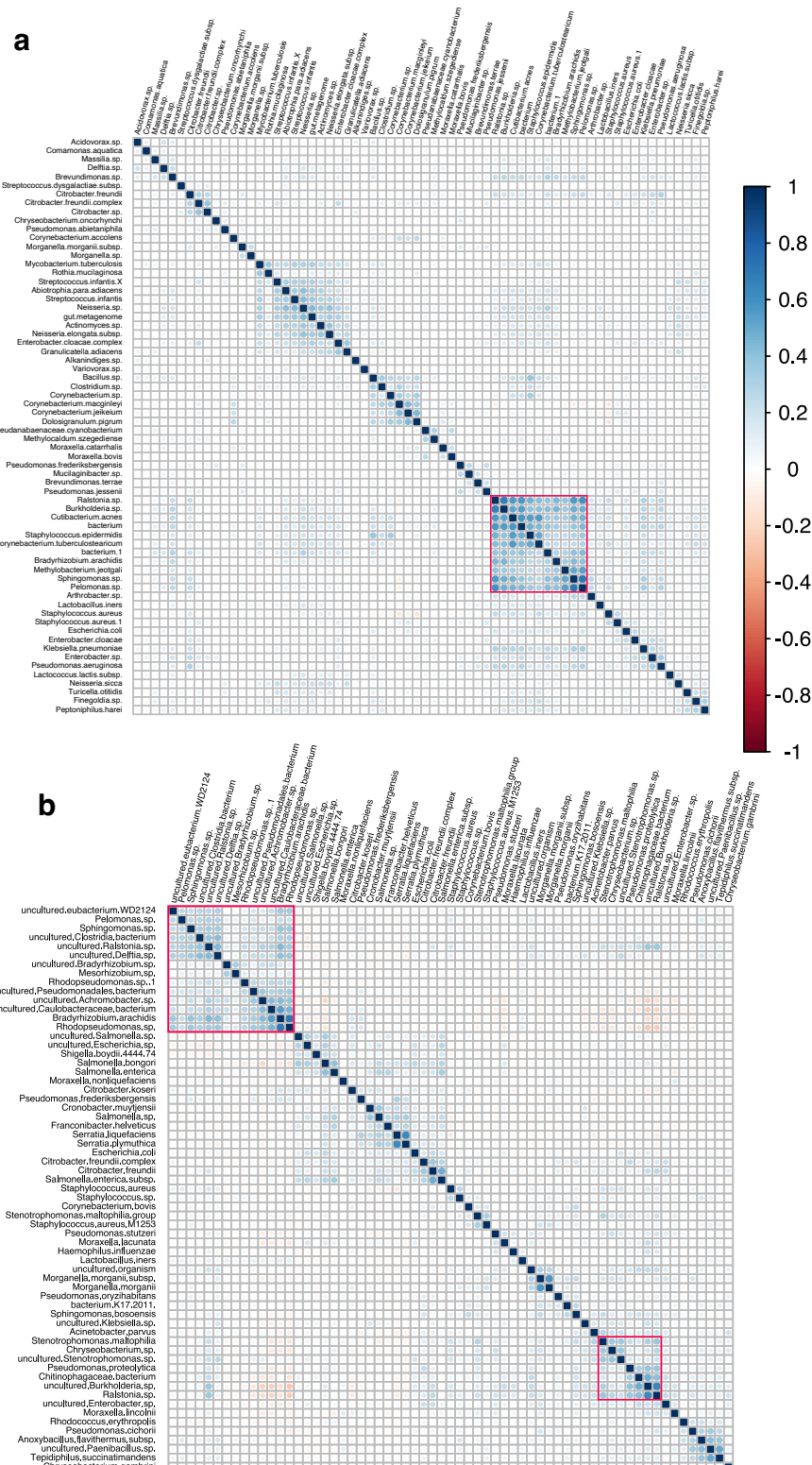


Supplementary Fig. S2: Nasal samples processed for 16S rRNA gene sequencing from the CARRIAGE study with high-quality reads after mothur processing. (a) all samples processed over the study period (b) box plot to show variation in the total read count (i.e. 16S rDNA gene copies) by carriage or control status. For (a) and (b), n=1866 samples.

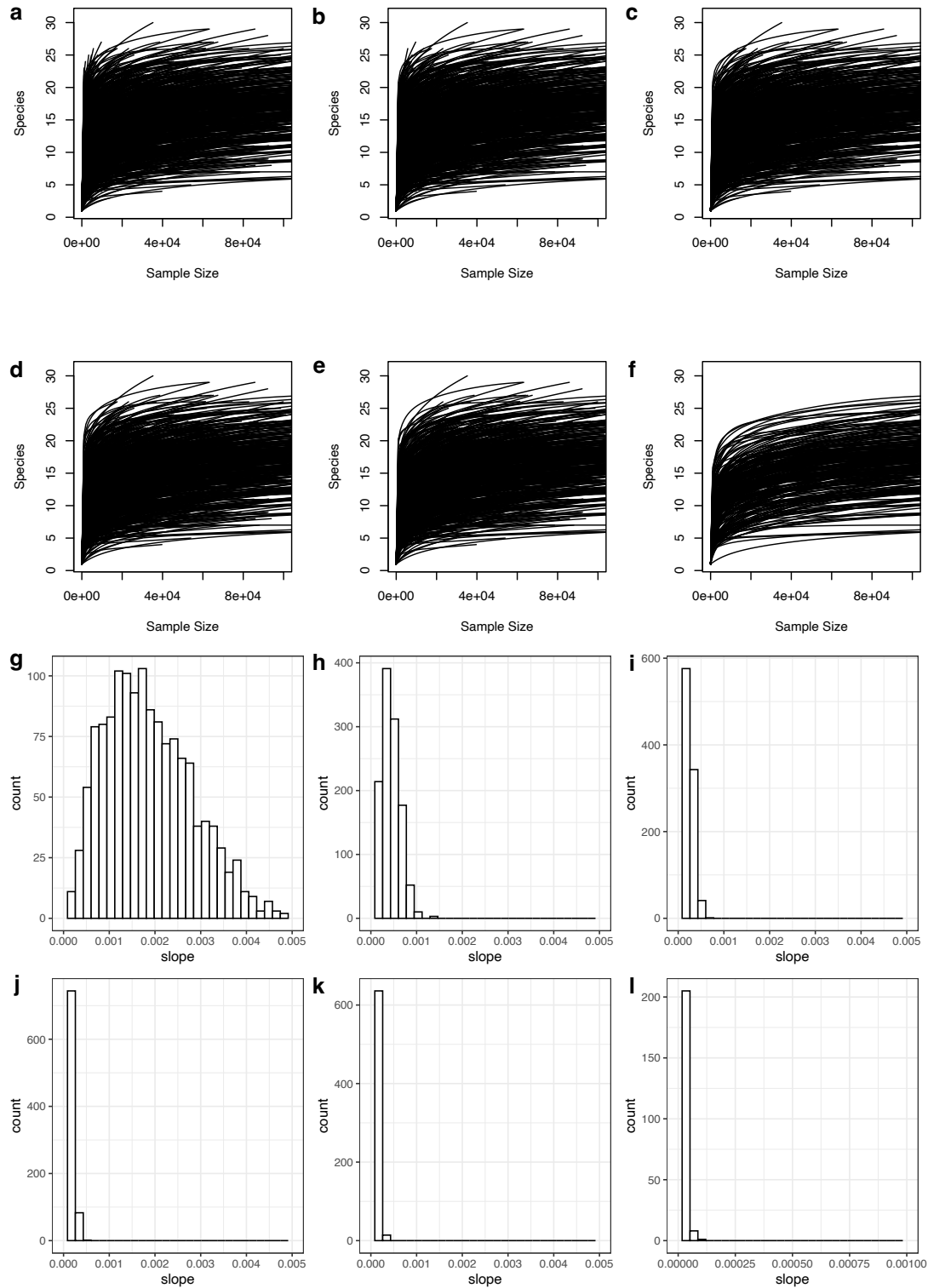


Supplementary Fig. S3: Identification of contaminants through batch effects and association with sample DNA concentration. Y-axis represents log 10 of p-value of spearman's correlation coefficient calculation, and horizontal 'red-line' intercepts y-axis at p=0.05. Vertical red-line intercepts spearman rho values at -0.2 and 0.2. **(a-e)** correlation of taxa relative abundance at a species level with each extraction location where samples from the first batch were processed. **(f-j)** correlation of taxa relative abundance at a species level with each PCR amplification location where samples

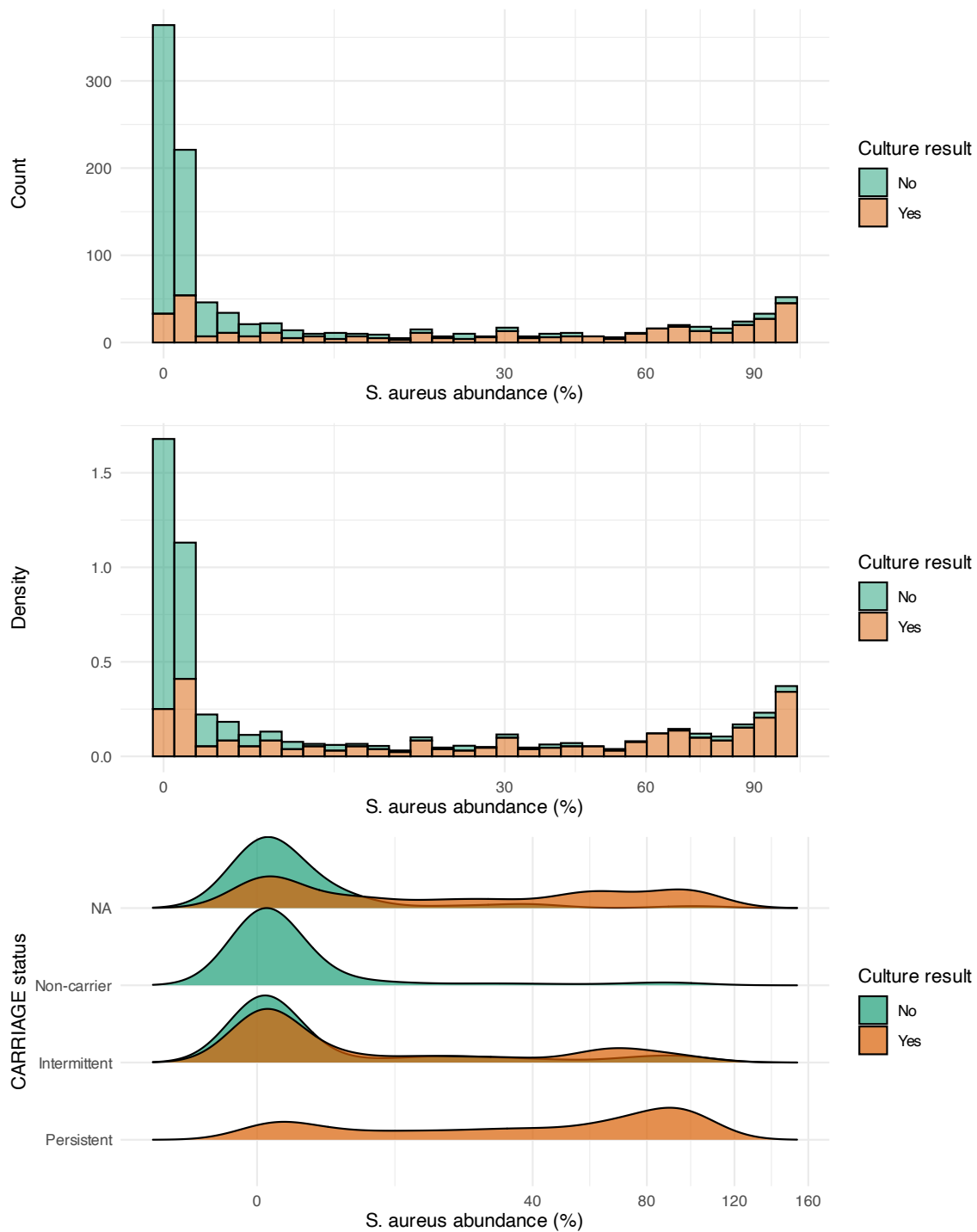
Supplementary Fig. S4: Heatmap demonstrating correlation of contaminants with one another. All species were correlated with one another and ordered with



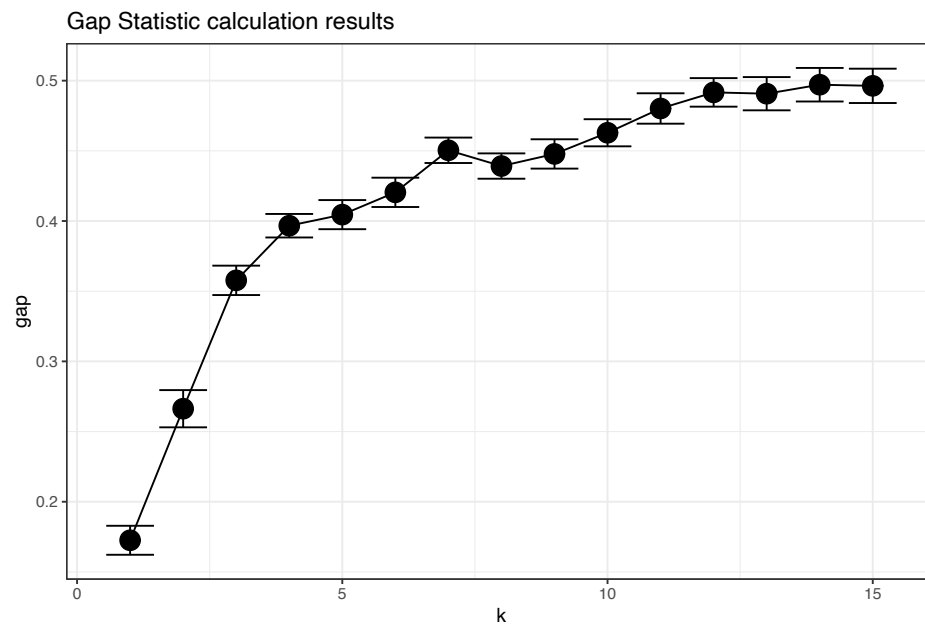
hierarchical clustering of the spearman's correlation coefficient rho values. Only significant correlation values are displayed ($p < 0.05$). The correlation heatmap was then subset to allow for visualisation of examples that displayed clusters involving known contaminants that are highlighted with the red boxes. **(a)** subset plot from species-to-species correlation plot of the first batch of samples processed (n=1099 samples). **(b)** subset plot from species-to-species correlation plot of the second batch of samples processed (n=767 samples).



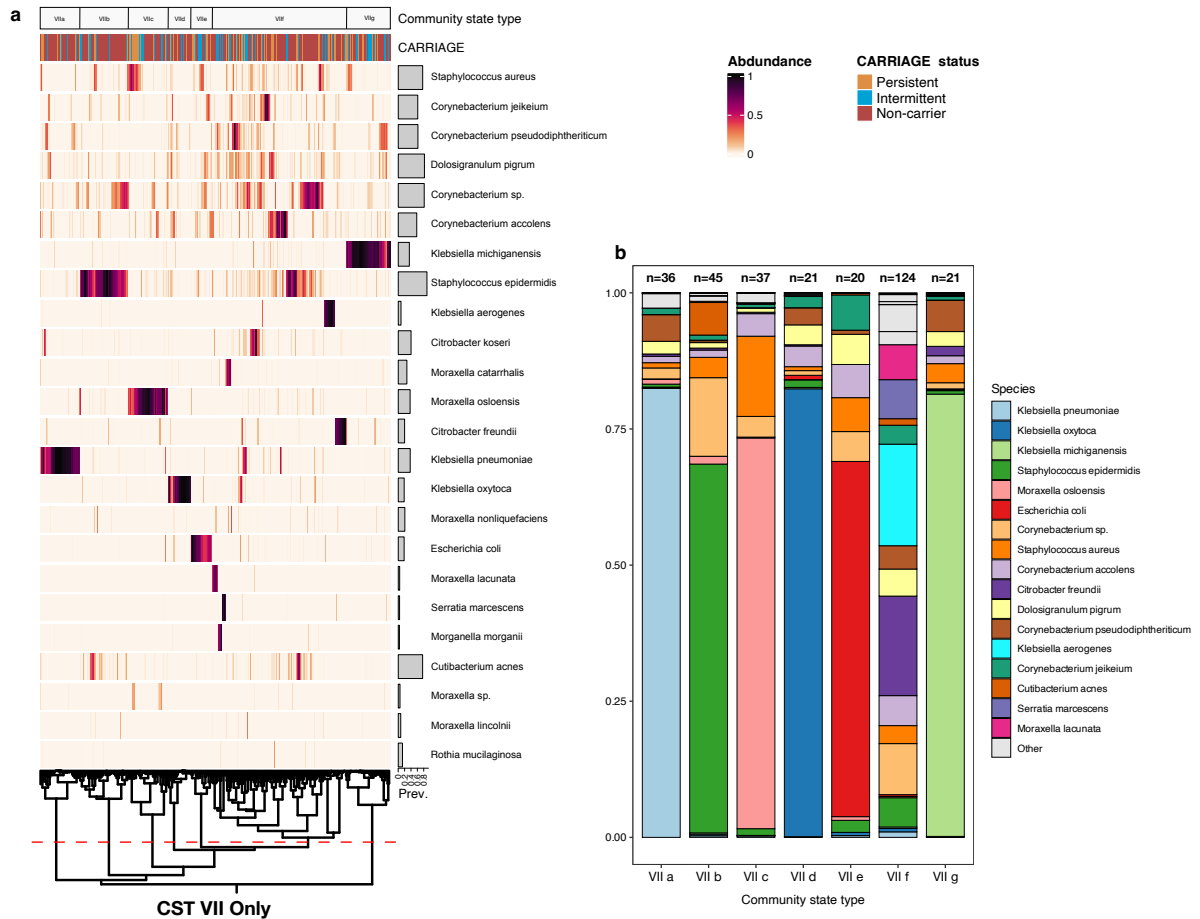
Supplementary Fig. S5: Setting a rarefaction threshold. (a-f) Rarefaction curves generated for samples ($n=1,756$ samples) after the removal of samples with less than a) 1,000 b) 5,000 c) 10,000 d) 15,000 e) 20,000 and f) 100,000 high-quality reads. (g-l) Histogram representing the slope of the rarefaction curves generated in figures a to f at the respective high-quality read count threshold below which samples were removed g) 1,000 h) 5,000 i) 10,000 j) 15,000 k) 20,000 and l) 100,000 high-quality reads.



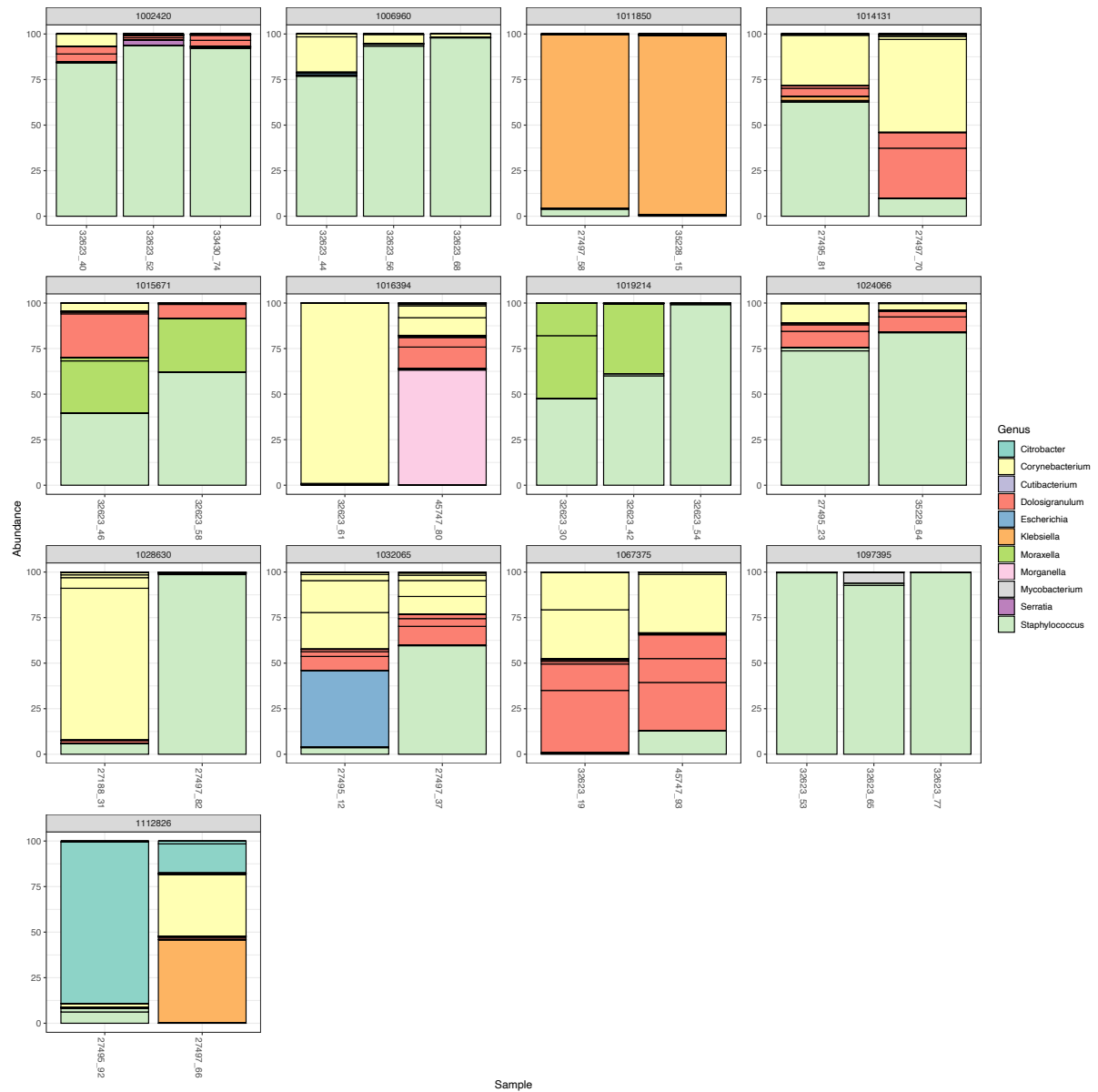
Supplementary Fig. S6: Relationship of *Staphylococcus aureus* culture and *S. aureus* abundance (a) Histogram distribution (counts) of *S. aureus* abundance (%) by culture result. Negative culture, n=672 samples, Positive culture, n=382 samples. (b) Density distribution of *S. aureus* abundance (%) by culture result. Negative culture, n=672 samples, Positive culture, n=382 samples. (c) Ridgeline densities of *S. aureus* abundance (%) across non-, intermittent-, and persistent- colonisation states and culture outcomes: persistent, n=275 samples; intermittent, n=169 samples; non-carrier, n=532 samples.



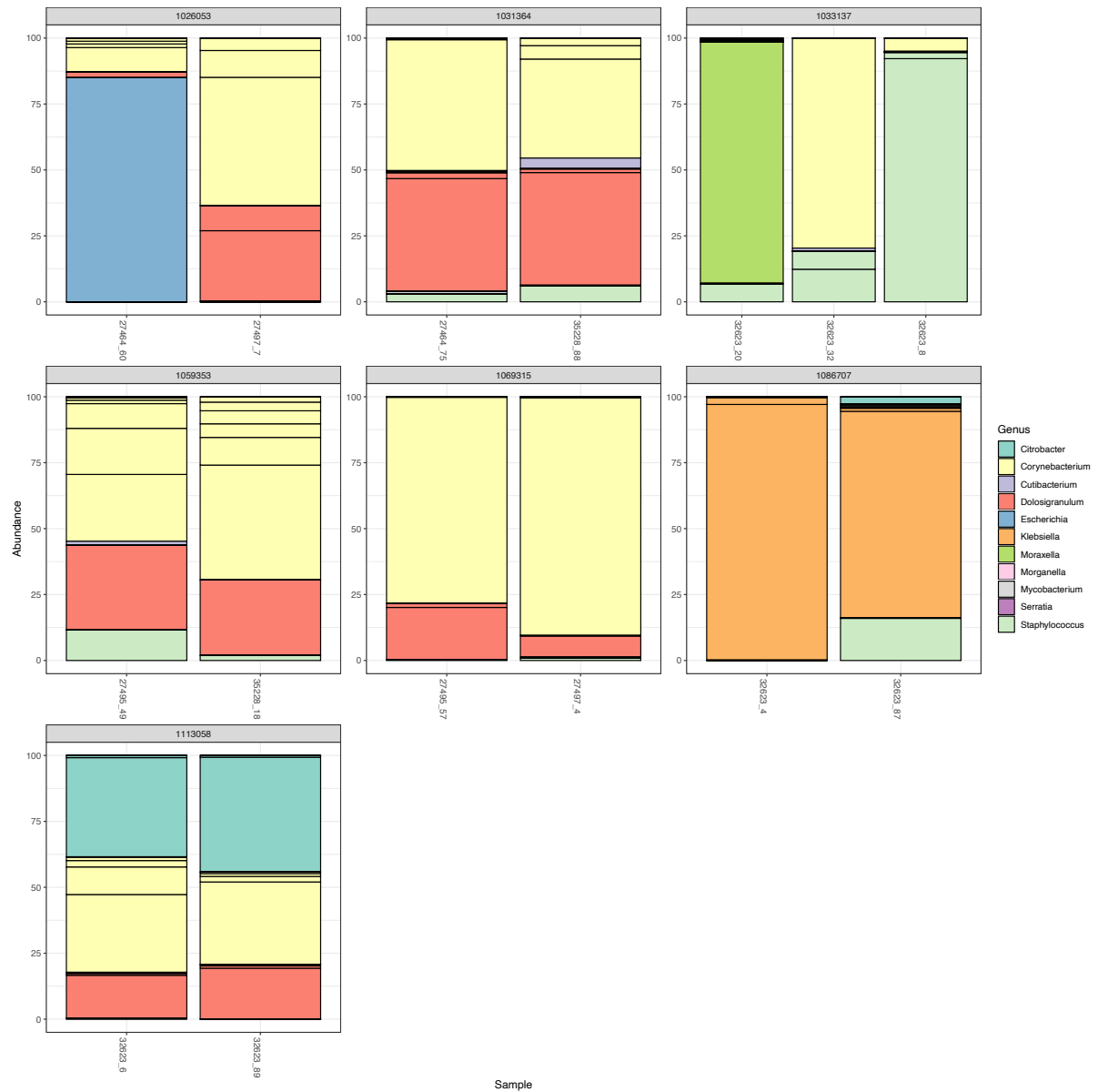
Supplementary Fig. S7: Gap statistic calculation results with number of clusters (k) against goodness of clustering measure (gap).



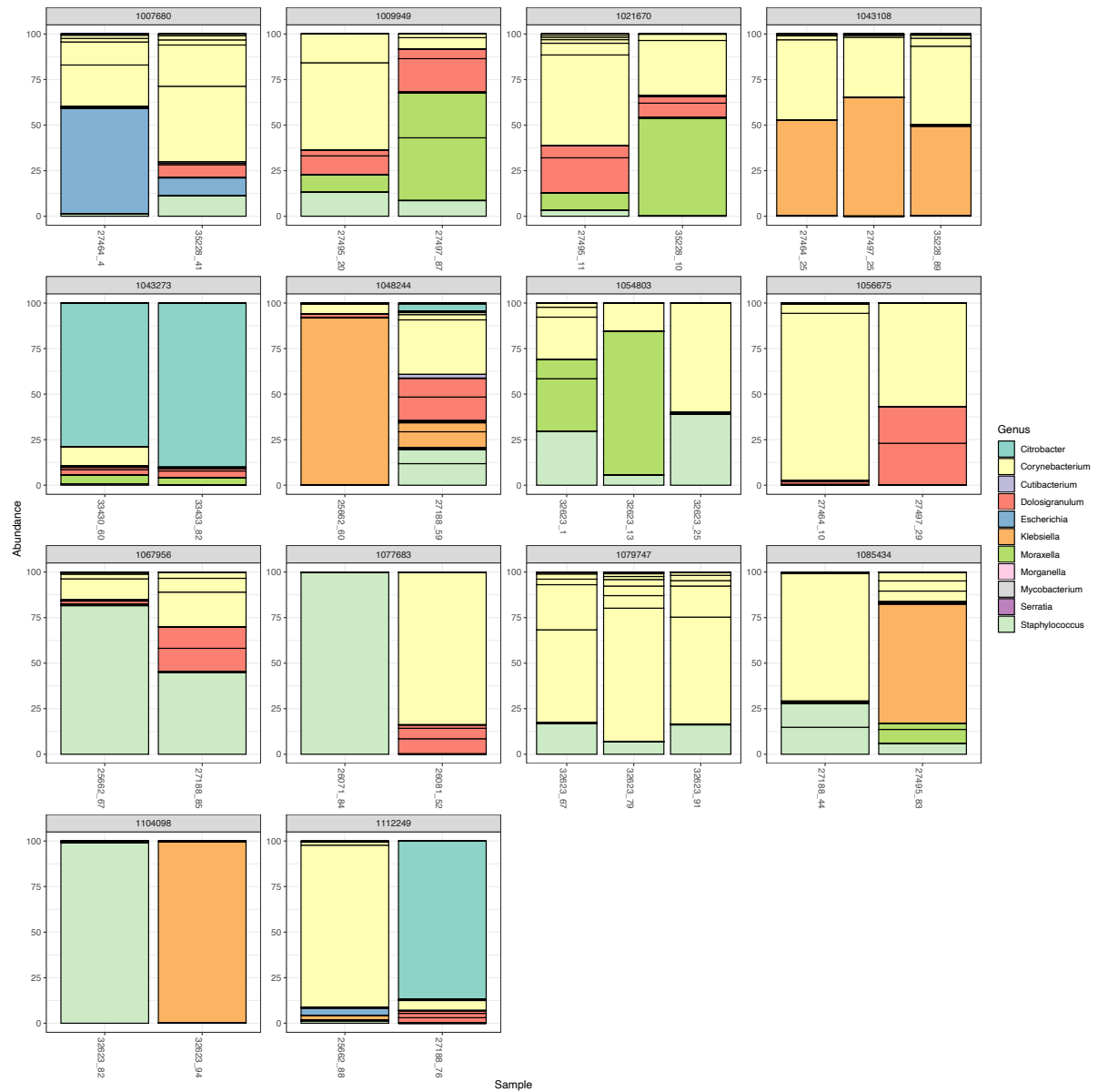
Supplementary Fig. S8: Sub-clusters and associated mean species abundance observed in CST-VII (a) sub-set plot of CST-VII alone from Figure 2A. Samples are ordered by hierarchical clustering using Bray-Curtis distances based on the compositional, relative abundance data, represented by the dendrogram. Prevalence of each species across the samples is represented by the horizontal bar plots. Community state types (CSTs) and *S. aureus* colonisation status of samples are represented above the heatmap. Sub-clusters are visually defined by partitioning the samples based on the dendrogram (dashed red-line) **(b)** Mean microbial composition of samples for each CST-VII sub-cluster at a species level. Sub-clusters are dominated by a single species, revealing additional structure within the nasal microbiome across the sampled population. For (a, b), n=324 samples from unique individuals.



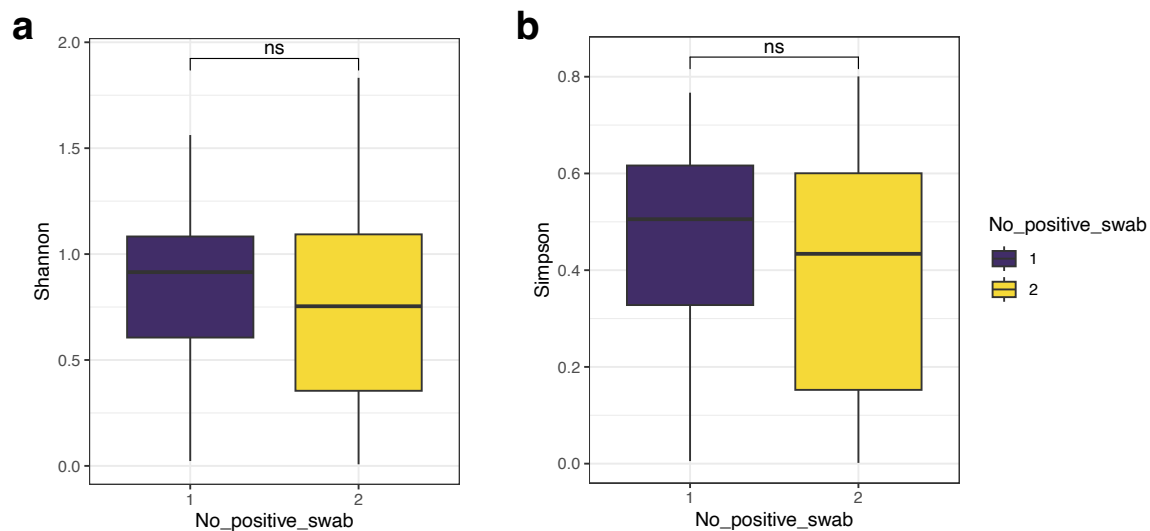
Supplementary Fig. S10: Relative abundance of serially sampled *Staphylococcus aureus* persistent carriers at a genus level. Relative abundance is calculated after rarefaction of high-quality reads. N=13 individual participants.



Supplementary Fig. S11: Relative abundance of serially sampled *Staphylococcus aureus* intermittent carriers at a genus level. Relative abundance is calculated after rarefaction of high-quality reads. N=7 individual participants.



Supplementary Fig. S12: Relative abundance of serially sampled *Staphylococcus aureus* non-carriers at a genus level. Relative abundance is calculated after rarefaction of high-quality reads. N=14 individual participants.



Supplementary Fig. S13: Alpha diversity from nasal samples by frequency of nasal swabs positive for *Staphylococcus aureus* amongst intermittent carriers. (a-b) Box plot comparing alpha diversity (Shannon and Simpson) from nasal samples by the number of swabs positive (one or two) for *Staphylococcus aureus* from intermittent carriers; n=169 samples from unique individuals. **(a)** Shows no significant difference in alpha diversity (Shannon) between intermittent carriers with one swab positive for *S. aureus* vs those with two swab (p=0.21). **(b)** Shows no significant difference in alpha diversity (Simpson) between intermittent carriers with one swab positive for *S. aureus* vs those with two swab (p=0.19). 'ns', 'non-significant'.

Supplementary Results

Random Forest model performance of combined dataset

We assessed the performance of the random forest model to predict carriage status with the combined dataset including all samples above 500 reads and those rarefied above 10,000.

We determined the best number of candidates to be sampled at each tree (mtry) to be 4. The estimated test classification accuracy of the trained model was 70.6% (1-estimated out of box error). We determined the accuracy of the model with the test data. The overall accuracy of the model was 69.9% (C.I.= 62.3-76.2%, p=0.001) significantly exceeding the no information rate.