**Supplemental Material**


**Evaluating bias for Illumina-based bacterial 16S rRNA gene profiles**

Katherine Kennedy[1], Michael W. Hall[1], Michael D. J. Lynch[1], Gabriel Moreno-Hagelsieb[2], and Josh D. Neufeld[1]#


[1]University of Waterloo, Department of Biology, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada

[2]Wilfrid Laurier University, Department of Biology, Waterloo, Ontario, Canada N2L 3C5, Canada

Running Head: Evaluating bias for 16S rRNA gene profiles

#Address correspondence to Josh D. Neufeld, jneufeld@uwaterloo.ca

Table S1. All samples sequenced with number of sequences contributed to analysis. Samples in light grey were excluded due to a low number of sequences and samples in dark grey were negative controls.

| Lane | Barcode | Sample | Pooled | Template Concentration | Barcode | Sequences Contributed |
|------|---------|--------|--------|------------------------|---------|-----------------------|
| 1 | V4_1R | S1 | Yes | high | ATCACG | 29907 |
| 1 | V4_2R | S1 | Yes | high | CGATGT | 40939 |
| 1 | V4_3R | S1 | Yes | high | TTAGGC | 40012 |
| 1 | V4_4R | S1 | Yes | high | TGACCA | 24218 |
| 1 | V4_5R | S1 | Yes | high | ACAGTG | 43532 |
| 1 | V4_6R | S1 | Yes | low | GCCAAT | 16085 |
| 1 | V4_7R | S1 | Yes | low | CAGATC | 222 |
| 1 | V4_8R | S1 | Yes | low | ACTTGA | 15911 |
| 1 | V4_9R | S1 | Yes | low | GATCAG | 36273 |
| 1 | V4_10R | S1 | Yes | low | TAGCTT | 15054 |
| 1 | V4_11R | 6TD | Yes | high | GGCTAC | 41985 |
| 1 | V4_12R | 6TD | Yes | high | CTTGTA | 77059 |
| 1 | V4_13R | 6TD | Yes | high | AGTACG | 70548 |
| 1 | V4_14R | 6TD | Yes | high | TCAGTC | 50318 |
| 1 | V4_15R | 6TD | Yes | high | TTGAGC | 97766 |
| 1 | V4_16R | 6TD | Yes | low | AAGCGA | 15909 |
| 1 | V4_17R | 6TD | Yes | low | TCCTCA | 15161 |
| 1 | V4_18R | 6TD | Yes | low | GGTTGT | 12052 |
| 1 | V4_19R | 6TD | Yes | low | TGAGGT | 23508 |
| 1 | V4_20R | 6TD | Yes | low | TACCGT | 24562 |
| 1 | V4_21R | 10AS | Yes | high | CCAACT | 71393 |
| 1 | V4_22R | 10AS | Yes | high | AGAGAG | 42154 |
| 1 | V4_23R | 10AS | Yes | high | CACTTG | 69607 |
| 1 | V4_24R | 10AS | Yes | high | TCAAGG | 63817 |
| 1 | V4_25R | 10AS | Yes | high | AGTGGT | 54384 |
| 1 | V4_26R | 10AS | Yes | low | GACACT | 13211 |
| 1 | V4_27R | 10AS | Yes | low | CCTTCT | 27644 |
| 1 | V4_28R | 10AS | Yes | low | GGATAA | 20895 |
| 1 | V4_29R | 10AS | Yes | low | CCTTAA | 14797 |
| 1 | V4_30R | 10AS | Yes | low | CAAGAA | 12505 |
| 1 | V4_31R | S3 | Yes | high | GTTGAA | 58081 |
| 1 | V4_32R | S3 | Yes | high | TCACAA | 58995 |
| 1 | V4_33R | S3 | Yes | high | AGTCAA | 54104 |
| 1 | V4_34R | S3 | Yes | high | CGAATA | 70280 |
| 1 | V4_35R | S3 | Yes | high | GCTATA | 60227 |
| 1 | V4_36R | S3 | Yes | low | GAGTTA | 17204 |
| 1 | V4_37R | S3 | Yes | low | TTGGTA | 17907 |
| 1 | V4_38R | S3 | Yes | low | AACGTA | 29850 |
| 1 | V4_39R | S3 | Yes | low | GTACTA | 16578 |

| 1 | V4_40R | S3 | Yes | low | CATCTA | 257 |
|---|--------|-----|-----|------|--------|-----|
| 1 | V4_41R | S1 | No | high | TGTAGA | 56188 |
| 1 | V4_42R | S1 | No | high | ATCAGA | 48967 |
| 1 | V4_43R | S1 | No | high | ACATGA | 44309 |
| 1 | V4_44R | S1 | No | high | TAGACA | 42964 |
| 1 | V4_45R | S1 | No | high | GAGAAT | 3699 |
| 1 | V4_46R | S1 | No | low | CTCAAT | 2279 |
| 1 | V4_47R | S1 | No | low | AGGTAT | 9222 |
| 1 | V4_48R | S1 | No | low | TTGCAT | 1835 |
| 1 | V4_49R | S1 | No | low | TGGATT | 36714 |
| 1 | V4_50R | S1 | No | low | ACCATT | 5904 |
| 1 | V4_51R | 6TD | No | high | CTAGTT | 39474 |
| 1 | V4_52R | 6TD | No | high | AGTGTT | 277 |
| 1 | V4_53R | 6TD | No | high | TCTCTT | 46185 |
| 1 | V4_54R | 6TD | No | high | GTAAGT | 137519 |
| 1 | V4_55R | 6TD | No | high | CAATGT | 34483 |
| 1 | V4_56R | 6TD | No | low | ATTCGT | 26978 |
| 1 | V4_57R | 6TD | No | low | ATGACT | 14975 |
| 1 | V4_58R | 6TD | No | low | ACTTCT | 43590 |
| 1 | V4_59R | 6TD | No | low | CATAAG | 4315 |
| 1 | V4_60R | 6TD | No | low | TTCTAG | 20219 |
| 1 | V4_61R | 10AS | No | high | AAGATG | 73682 |
| 1 | V4_62R | 10AS | No | high | TATGTG | 52649 |
| 1 | V4_63R | 10AS | No | high | AATTGG | 27085 |
| 1 | V4_64R | 10AS | No | high | TAATCG | 507 |
| 1 | V4_65R | 10AS | No | high | ACTAAC | 80471 |
| 1 | V4_66R | 10AS | No | low | TGTTAC | 4272 |
| 1 | V4_67R | 10AS | No | low | ATACAC | 32074 |
| 1 | V4_68R | 10AS | No | low | CTTATC | 45677 |
| 1 | V4_69R | 10AS | No | low | AGATTC | 3006 |
| 1 | V4_70R | 10AS | No | low | ACGGAA | 30973 |
| 1 | V4_71R | S3 | No | high | TGCGAA | 147 |
| 1 | V4_72R | S3 | No | high | GACCAA | 49796 |
| 1 | V4_73R | S3 | No | high | CTGTCA | 27838 |
| 1 | V4_74R | S3 | No | high | GCAGAT | 27438 |
| 1 | V4_75R | S3 | No | high | TCGTGT | 42015 |
| 1 | V4_76R | S3 | No | low | GAACCT | 26648 |
| 1 | V4_77R | S3 | No | low | GTCATG | 85676 |
| 1 | V4_78R | S3 | No | low | GATAGC | 4595 |
| 1 | V4_79R | S3 | No | low | AAGTCC | 31185 |
| 1 | V4_80R | S3 | No | low | ATTGCC | 6566 |
| 1 | V4_81R | SX | yes | high | CCGAGA | 53661 |
| 1 | V4_82R | SX | yes | high | CGCTGA | 143162 |
| 1 | V4_83R | SX | yes | high | GGCACA | 52548 |
| 1 | V4_84R | SX | yes | low | CGTGCA | 27642 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | V4_85R | SX | yes | low | GGCCTT | 27417 |
| 1 | V4_86R | SX | yes | low | CCTGGT | 20185 |
| 1 | V4_87R | SX | yes | high | CAGGCT | 41885 |
| 1 | V4_88R | SX | yes | high | GTCGCT | 53682 |
| 1 | V4_89R | SX | yes | high | GCGTAG | 35433 |
| 1 | V4_90R | SX | yes | low | CTGGAG | 37627 |
| 1 | V4_91R | SX | yes | low | CTACGG | 35151 |
| 1 | V4_92R | SX | yes | low | ACACCG | 29416 |
| 1 | V4_93R | negcont | na | na | GTTCCG | 241 |
| 1 | V4_94R | negcont | na | na | CAGCAC | 642 |
| 1 | V4_95R | negcont | na | na | CCGTTC | 464 |
| 1 | V4_96R | negcont | na | na | GCATCC | 371 |
| 2 | V4_1R | S1 | Yes | high | ATCACG | 23312 |
| 2 | V4_2R | S1 | Yes | high | CGATGT | 30580 |
| 2 | V4_3R | S1 | Yes | high | TTAGGC | 0 |
| 2 | V4_4R | S1 | Yes | high | TGACCA | 28383 |
| 2 | V4_5R | S1 | Yes | high | ACAGTG | 31086 |
| 2 | V4_6R | S1 | Yes | low | GCCAAT | 19715 |
| 2 | V4_7R | S1 | Yes | low | CAGATC | 24126 |
| 2 | V4_8R | S1 | Yes | low | ACTTGA | 19623 |
| 2 | V4_9R | S1 | Yes | low | GATCAG | 20025 |
| 2 | V4_10R | S1 | Yes | low | TAGCTT | 16685 |
| 2 | V4_11R | 6TD | Yes | high | GGCTAC | 31789 |
| 2 | V4_12R | 6TD | Yes | high | CTTGTA | 37765 |
| 2 | V4_13R | 6TD | Yes | high | AGTACG | 41893 |
| 2 | V4_14R | 6TD | Yes | high | TCAGTC | 34343 |
| 2 | V4_15R | 6TD | Yes | high | TTGAGC | 46193 |
| 2 | V4_16R | 6TD | Yes | low | AAGCGA | 13917 |
| 2 | V4_17R | 6TD | Yes | low | TCCTCA | 18927 |
| 2 | V4_18R | 6TD | Yes | low | GGTTGT | 8183 |
| 2 | V4_19R | 6TD | Yes | low | TGAGGT | 16947 |
| 2 | V4_20R | 6TD | Yes | low | TACCGT | 9733 |
| 2 | V4_21R | 10AS | Yes | high | CCAACT | 44346 |
| 2 | V4_22R | 10AS | Yes | high | AGAGAG | 36900 |
| 2 | V4_23R | 10AS | Yes | high | CACTTG | 44436 |
| 2 | V4_24R | 10AS | Yes | high | TCAAGG | 43535 |
| 2 | V4_25R | 10AS | Yes | high | AGTGGT | 50034 |
| 2 | V4_26R | 10AS | Yes | low | GACACT | 14967 |
| 2 | V4_27R | 10AS | Yes | low | CCTTCT | 15885 |
| 2 | V4_28R | 10AS | Yes | low | GGATAA | 22784 |
| 2 | V4_29R | 10AS | Yes | low | CCTTAA | 14842 |
| 2 | V4_30R | 10AS | Yes | low | CAAGAA | 10201 |
| 2 | V4_31R | S3 | Yes | high | GTTGAA | 42576 |
| 2 | V4_32R | S3 | Yes | high | TCACAA | 38499 |
| 2 | V4_33R | S3 | Yes | high | AGTCAA | 43807 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | V4_34R | S3 | Yes | high | CGAATA | 51703 |
| 2 | V4_35R | S3 | Yes | high | GCTATA | 57034 |
| 2 | V4_36R | S3 | Yes | low | GAGTTA | 16813 |
| 2 | V4_37R | S3 | Yes | low | TTGGTA | 20506 |
| 2 | V4_38R | S3 | Yes | low | AACGTA | 20498 |
| 2 | V4_39R | S3 | Yes | low | GTACTA | 12793 |
| 2 | V4_40R | S3 | Yes | low | CATCTA | 458 |
| 2 | V4_41R | S1 | No | high | TGTAGA | 33518 |
| 2 | V4_42R | S1 | No | high | ATCAGA | 30673 |
| 2 | V4_43R | S1 | No | high | ACATGA | 16622 |
| 2 | V4_44R | S1 | No | high | TAGACA | 26813 |
| 2 | V4_45R | S1 | No | high | GAGAAT | 23804 |
| 2 | V4_46R | S1 | No | low | CTCAAT | 4158 |
| 2 | V4_47R | S1 | No | low | AGGTAT | 8406 |
| 2 | V4_48R | S1 | No | low | TTGCAT | 21612 |
| 2 | V4_49R | S1 | No | low | TGGATT | 29469 |
| 2 | V4_50R | S1 | No | low | ACCATT | 25817 |
| 2 | V4_51R | 6TD | No | high | CTAGTT | 29569 |
| 2 | V4_52R | 6TD | No | high | AGTGTT | 22184 |
| 2 | V4_53R | 6TD | No | high | TCTCTT | 23581 |
| 2 | V4_54R | 6TD | No | high | GTAAGT | 57618 |
| 2 | V4_55R | 6TD | No | high | CAATGT | 24669 |
| 2 | V4_56R | 6TD | No | low | ATTCGT | 14641 |
| 2 | V4_57R | 6TD | No | low | ATGACT | 18141 |
| 2 | V4_58R | 6TD | No | low | ACTTCT | 30553 |
| 2 | V4_59R | 6TD | No | low | CATAAG | 16383 |
| 2 | V4_60R | 6TD | No | low | TTCTAG | 15738 |
| 2 | V4_61R | 10AS | No | high | AAGATG | 35874 |
| 2 | V4_62R | 10AS | No | high | TATGTG | 32479 |
| 2 | V4_63R | 10AS | No | high | AATTGG | 22976 |
| 2 | V4_64R | 10AS | No | high | TAATCG | 11814 |
| 2 | V4_65R | 10AS | No | high | ACTAAC | 46183 |
| 2 | V4_66R | 10AS | No | low | TGTTAC | 15111 |
| 2 | V4_67R | 10AS | No | low | ATACAC | 16895 |
| 2 | V4_68R | 10AS | No | low | CTTATC | 19676 |
| 2 | V4_69R | 10AS | No | low | AGATTC | 13698 |
| 2 | V4_70R | 10AS | No | low | ACGGAA | 15970 |
| 2 | V4_71R | S3 | No | high | TGCGAA | 23264 |
| 2 | V4_72R | S3 | No | high | GACCAA | 33802 |
| 2 | V4_73R | S3 | No | high | CTGTCA | 27377 |
| 2 | V4_74R | S3 | No | high | GCAGAT | 30528 |
| 2 | V4_75R | S3 | No | high | TCGTGT | 20118 |
| 2 | V4_76R | S3 | No | low | GAACCT | 28758 |
| 2 | V4_77R | S3 | No | low | GTCATG | 55074 |
| 2 | V4_78R | S3 | No | low | GATAGC | 14274 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | V4_79R | S3 | No | low | AAGTCC | 16168 |
| 2 | V4_80R | S3 | No | low | ATTGCC | 15946 |
| 2 | V4_81R | SX | yes | high | CCGAGA | 44568 |
| 2 | V4_82R | SX | yes | high | CGCTGA | 41293 |
| 2 | V4_83R | SX | yes | high | GGCACA | 35443 |
| 2 | V4_84R | SX | yes | low | CGTGCA | 16991 |
| 2 | V4_85R | SX | yes | low | GGCCTT | 16868 |
| 2 | V4_86R | SX | yes | low | CCTGGT | 12231 |
| 2 | V4_87R | SX | yes | high | CAGGCT | 32029 |
| 2 | V4_88R | SX | yes | high | GTCGCT | 35515 |
| 2 | V4_89R | SX | yes | high | GCGTAG | 25929 |
| 2 | V4_90R | SX | yes | low | CTGGAG | 24184 |
| 2 | V4_91R | SX | yes | low | CTACGG | 21652 |
| 2 | V4_92R | SX | yes | low | ACACCG | 19598 |
| 2 | V4_93R | -ve ctrl | n/a | n/a | GTTCCG | 248 |
| 2 | V4_94R | -ve ctrl | n/a | n/a | CAGCAC | 866 |
| 2 | V4_95R | -ve ctrl | n/a | n/a | CCGTTC | 429 |
| 2 | V4_96R | -ve ctrl | n/a | n/a | GCATCC | 610 |

Table S2. Analysis of variance between replicates of each group. The *F*-ratio (*F*) represents the group variance divided by the unexplained within-group variance; *p* represents the probability that the observed ratio would be expected based on chance alone.

| Sample | Template concentration | Pooling | Lane |
|--------|------------------------|---------|------|
| 6TD | $F = 139.6$ <br> $p = 6.0^{-14}$ | $F = 1.77$ <br> $p = 0.19$ | $F = 0.197$ <br> $p = 0.66$ |
| 10AS | $F = 188.2$ <br> $p = 7.2^{-16}$ | $F = 3.31$ <br> $p = 0.077$ | $F = 0.0214$ <br> $p = 0.88$ |
| S1 | $F = 11.2$ <br> $p = 0.0021$ | $F = 1.22$ <br> $p = 0.28$ | $F = 3.16$ <br> $p = 0.085$ |
| S3 | $F = 3.74$ <br> $p = 0.061$ | $F = 0.944$ <br> $p = 0.34$ | $F = 0.275$ <br> $p = 0.60$ |
| SX | $F = 17.8$ <br> $p = 0.00035$ | $F = 0.169$ <br> $p = 0.69$ | $F = 1.82$ <br> $p = 0.19$ |