**Supplemental Information**

**Appendix A**

**Table A.1** Keywords used to identify environmental information

|  |
| --- |
| Keyword |
| Case-sensitive |
| 'Source: ' |
| 'Source type strain:' |
| Case-insensitive |
| 'freshwater bacterium' |
| 'isolates originate from' |
| 'habitat' |
| 'first isolated' |
| 'isolated by' |
| 'were isolated' |
| 'was isolated' |
| 'obtained from' |
| 'isolates from' |
| 'recovered from' |
| 'primary reservoir' |
| 'identified from' |
| 'isolated from' |
| 'occurs in' |
| 'found in' |
| 'found on' |
| 'endosymbiont of' |
| 'associated with' |

**Appendix B**

**Table B.1** Database statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Level | Species | Genus | Family | Order | Class | Phylum |
| Count | 11633 | 2339 | 408 | 175 | 77 | 34 |
| HabitatInformation | 72.0% | 85.6% | 92.6% | 96.6% | 97.4% | 100% |
| Pre-definedEnvironment | 45.4% | 81.3% | 91.4% | 95.4% | 96.1% | 97.1% |

**Table B.2** Percentages of taxa passing each pre-defined filter

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Category | Habitat | Species | Genus | Family | Order | Class | Phylum |
| animal | human | 8.4 | 17.6 | 14.1 | 27.4 | 32.4 | 53.2 | 35.4 | 56.6 | 36.4 | 55.8 | 47.1 | 61.8 |
| other mammal | 6.4 | 12.0 | 31.9 | 37.1 | 35.1 | 38.2 |
| bird | 1.7 | 4.7 | 18.1 | 21.1 | 23.4 | 29.4 |
| fish | 0.8 | 3.6 | 13.2 | 16.6 | 19.5 | 23.5 |
| reptile | 0.28 | 1.0 | 4.7 | 9.7 | 15.6 | 17.6 |
| amphibian | 0.06 | 0.47 | 2.7 | 5.7 | 7.8 | 14.7 |
| insect/tick/spider | 1.7 | 4.7 | 17.6 | 23.4 | 26.0 | 32.3 |
| marine invertebrate | 1.3 | 5.8 | 19.9 | 28 | 28.6 | 32.3 |
| other invertebrate | 0.48 | 2.8 | 11.0 | 18.9 | 23.4 | 29.4 |
| other animal | 2.5 | 9.1 | 28.4 | 32.6 | 35.1 | 35.3 |
| body sites | skin | 1.4 | 14.1 | 4.5 | 20.5 | 15.0 | 41.7 | 18.3 | 47.4 | 22.1 | 46.8 | 20.6 | 52.9 |
| eye | 0.57 | 1.6 | 6.6 | 11.4 | 11.7 | 17.6 |
| oral | 2.4 | 4.6 | 14.7 | 21.7 | 24.7 | 29.4 |
| ear/nose/throat | 3.5 | 6.0 | 18.6 | 22.3 | 22.1 | 23.5 |
| female reproductive | 1.2 | 3.2 | 11.5 | 17.7 | 22.1 | 23.5 |
| male reproductive  | 0.51 | 1.2 | 5.4 | 11.4 | 16.9 | 23.5 |
| general reproduction | 0.78 | 3.0 | 11.8 | 17.7 | 18.2 | 23.5 |
| lymph/circulatory  | 2.7 | 7.9 | 25.0 | 25.7 | 24.7 | 23.5 |
| nervous system | 0.61 | 2.9 | 10.8 | 17.7 | 19.5 | 17.6 |
| bone/muscle | 0.73 | 2.0 | 9.6 | 16 | 18.2 | 20.6 |
| liver/urinary/pancreas | 1.5 | 4.8 | 15.7 | 19.4 | 22.1 | 23.5 |
| gut/digestive system | 6.5 | 13.4 | 32.1 | 40.6 | 45.5 | 50 |
| food | fruits/vegetables | 1.8 | 7.0 | 6.3 | 16.2 | 20.6 | 38.5 | 25.7 | 40.6 | 29.9 | 41.6 | 35.2 | 47.1 |
| dairy | 1.4 | 3.8 | 14.2 | 17.7 | 23.4 | 20.6 |
| grains | 1.5 | 5.3 | 16.2 | 20 | 22.1 | 26.5 |
| meat/eggs | 2.0 | 5.9 | 21.1 | 23.4 | 23.2 | 29.4 |
| alcohol | 0.68 | 2.0 | 8.1 | 11.4 | 13.0 | 14.7 |
| other food | 0.76 | 3.0 | 9.3 | 11.4 | 13.0 | 14.7 |
| plants | trees | 2.0 | 8.4 | 7.2 | 19.3 | 25.0 | 39.0 | 32 | 43.4 | 29.9 | 39.0 | 26.5 | 41.2 |
| flowers | 0.52 | 1.9 | 8.3 | 12.6 | 14.3 | 14.7 |
| lichen/algae/fungi | 0.83 | 4.1 | 13.0 | 18.3 | 19.5 | 23.5 |
| other plants | 6.0 | 14.3 | 32.4 | 37.1 | 35.1 | 41.2 |
| environ-ment | natural environments | 2.1 | 37.4 | 7.3 | 63.3 | 21.6 | 78.9 | 30.9 | 86.3 | 31.2 | 85.7 | 35.3 | 91.2 |
| farmed environments | 2.9 | 10.3 | 29.7 | 36 | 40.3 | 44.1 |
| fresh water | 7.5 | 25.2 | 48.8 | 54.3 | 58.4 | 67.6 |
| salt water | 11.5 | 31.2 | 53.7 | 62.3 | 63.6 | 85.3 |
| general water | 7.0 | 25.4 | 48.5 | 55.4 | 59.7 | 76.5 |
| soil | 17.5 | 28.9 | 49.5 | 57.7 | 64.9 | 64.7 |
| sand | 1.1 | 5.1 | 17.6 | 24.6 | 24.7 | 26.5 |
| rock/stone | 0.52 | 2.5 | 10.3 | 16.6 | 16.9 | 23.5 |
| other | waste | 3.5 | 12.1 | 12.5 | 34.8 | 32.6 | 66.4 | 38.9 | 76.0 | 37.7 | 77.9 | 50 | 85.3 |
| extreme | 6.5 | 21.0 | 47.3 | 61.7 | 67.5 | 73.5 |
| symbiont | 0.28 | 2.1 | 9.1 | 13.7 | 18.2 | 26.5 |
| built | 1.3 | 5.7 | 20.8 | 22.3 | 18.2 | 20.6 |
| oil/coal | 1.1 | 5.6 | 22.1 | 29.7 | 31.2 | 29.4 |
| microbial mat | 0.73 | 3.2 | 10.5 | 17.1 | 24.7 | 38.2 |

**Appendix C**

**Table C.1** Analysis of taxa failing the human OR other mammal OR any body-site filter for our forehead microbiome samples

|  |  |  |
| --- | --- | --- |
| Genus | Max. Rank of Failure | Potential Explanation |
| *Geodermatophilus* | Order | \*present in 2/475 samples\*one forehead swab on person 13; the initial control swab for person 13\*a common contaminant in blanks (Salter, Cox, Turek, Calus, Cookson, Moffatt et al. 2014)***\*likely contaminant*** |
| *Rhodothermus* | Order | \*present in 3/475 samples\*one forehead swab on person 40; the final control swab for person 40; the  initial control swab for person 41**\**likely contaminant***  |
| *Thermus* | Order | \*present in 3/475 samples\*one forehead swab on person 45; the final control swabs for persons 40, 47**\**probably contaminant*** |
| *Steroidobacter* | Order | \*present in 1/475 samples\*one forehead swab on person 43; not in controls**\**uncertain origin*** |
| *Methylotenera* | Order | \*present in 2/475 samples\*two forehead swabs on person 31; not in controls**\**uncertain origin*** |
| *Modestobacter* | Order | \*present in 5/475 samples\*one forehead swab on persons 6, 10, 36, two on person 43; not in controls\*also present at ~0.1% in other skin studies (Oh, Byrd, Deming, Conlan, Barnabas, Blakesley et al. 2014)**\**likely on skin*** |
| *Enhydrobacter* | Family | \*present in 159/475 samples\*present on 31/50 people and in 14/50 control sets\*we have cultured this directly from skin***\*likely on skin*** |
| *Adhaeribacter* | Family | \*present in 1/475 samples\*one forehead swab on person 13\*not found in any controls***\*uncertain origin*** |
| *Hymenobacter* | Family | \*present in 8/475 samples\*one forehead swab on persons 13, 17, 24, 37, 43, 45, two on person 32; not  in controls\*recently reported as a potentially novel human-associate (Romano-Bertrand, Frapier, Calvet, Colson, Albat, Parer et al. 2015)***\*likely on skin*** |
| *Methylobacterium* | Family | \*present in 50/475 samples\*present on 14/50 people and in 16/50 control sets \*simultaneously present in controls and samples for 4/50 people\*a common contaminant in blanks (Salter, Cox, Turek, Calus, Cookson, Moffatt et al. 2014)***\*likely contaminant*** |
| *Rheinheimera* | Family | \*present in 20/475 samples\*present on 4/50 people and in 11/50 control sets\*simultaneously present in controls and samples for 3/50 people***\*likely contaminant*** |
| *Demequina* | Family | \*present in 1/475 samples\*present on the forehead of person 23***\*uncertain origin*** |
| *Skermanella* | Family | \*present in 4/475 samples\*present on 2/50 people\*present in 2/50 control sets\*simultaneously present in controls and samples for 0/50 people***\*uncertain origin*** |
| *Cellvibrio* | Family | \*present in 6/475 samples\*present on 2/50 people\*present in 4/50 control sets\*simultaneously present in controls and samples for 0/50 people**\**uncertain origin*** |