**Supplemental Information for:**

**A total crapshoot?   
Evaluating bioinformatic decisions in animal diet metabarcoding analyses**

Devon R. O’Rourke (corresponding author: [devon@outermostlab.com](mailto:devon@outermostlab.com))

Nicholas A. Bokulich

Matthew D. MacManes

Jeffrey T. Foster

**Table of Contents:**

|  |  |
| --- | --- |
| **Table S1 – bat guano site collection information and sample size** | Page 2 |
| **Table S2 – alpha diversity estimates for mock community samples** | Page 3 |
| **Table S3 – Kruskal-Wallis test for differences in bat guano alpha diversity estimates among denoising software and filtering regimes** | Page 3 |
| **Table S4 – Dunn’s test for pairwise differences in bat guano alpha diversity estimates among denoising software and filtering regimes** | Page 4 |
| **Table S5 – PERMANOVA (Adonis) test for differences in bat guano community composition among denoising software and filtering regimes** | Page 5 |
| **Table S6 – Per-sample mock community ASV counts among denoising software and filtering regime** | Page 7 |
| **Figure S1 – bat guanos site collection map** | Page 8 |

**Table S1**. Guano samples were passively collected from 15 sites in New Hampshire (14) and Maine (1) in 2016. The following table includes the site abbreviation used in **Figure S1**, the town name where samples were collected, and the number of samples sequenced from each site.

**Site Town                                  Samples**

ALS    Alstead, NH 120

BRN    Brown Lane, Hollis, NH 178

CHI    Chichester, NH 92

CNB    Canterbury, NH 162

EPS    Epsom, NH 85

FAR    Fairfield, ME 52

FOX    Fox State Forest, Hillsborough, NH   177

GIL    Gilsum, NH 44

HOL    Holderness, NH 165

HOP    Hopkinton, NH 182

MAP    Maple Hill, Hollis, NH 209

MAS    Massabseic Lake, Auburn, NH         85

MTV    Mont Vernon, NH                     91

PEN    Penacook, NH 47

ROL    Rollinsford, NH 13

**Table S2**. Alpha diversity estimates among four mock community samples. Number of sequences or sequence equivalents are provided for each combination of denoising program (Denoiser) and filtering regime (basic, standard, or extra) for each Hill Number (q = 0 is equivalent to observed richness; q = 1 is equivalent to Shannon’s Entropy; q = 2 is equivalent to Simpsons’s 1-D diversity).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Denoiser** | **Sample** | **q = 0** | | | **q = 1** | | | **q = 2** | | |
| **basic** | **standard** | **extra** | **basic** | **standard** | **extra** | **basic** | **standard** | **extra** |
| DADA2 | libA | 25.00 | 24.00 | 24.00 | 23.82 | 23.55 | 23.54 | 23.31 | 23.17 | 23.16 |
| libB | 24.00 | 25.00 | 24.00 | 23.62 | 23.65 | 23.62 | 23.30 | 23.31 | 23.30 |
| libC | 24.00 | 24.00 | 24.00 | 23.48 | 23.48 | 23.49 | 23.01 | 23.02 | 23.03 |
| libD | 25.00 | 24.00 | 24.00 | 23.55 | 23.52 | 23.52 | 23.11 | 23.11 | 23.11 |
| Deblur | libA | 31.00 | 29.00 | 24.00 | 23.74 | 23.68 | 23.53 | 23.19 | 23.17 | 23.13 |
| libB | 27.00 | 25.00 | 24.00 | 23.64 | 23.58 | 23.56 | 23.16 | 23.14 | 23.15 |
| libC | 30.00 | 24.00 | 24.00 | 23.78 | 23.60 | 23.60 | 23.30 | 23.24 | 23.23 |
| libD | 42.00 | 25.00 | 24.00 | 24.26 | 23.53 | 23.50 | 23.29 | 23.06 | 23.05 |
| VSEARCH | libA | 39.00 | 23.00 | 23.00 | 22.49 | 22.04 | 22.02 | 21.01 | 20.95 | 20.91 |
| libB | 30.00 | 25.00 | 24.00 | 22.31 | 22.18 | 22.16 | 21.08 | 21.05 | 21.10 |
| libC | 34.00 | 29.00 | 24.00 | 22.35 | 22.25 | 22.13 | 20.89 | 20.94 | 21.00 |
| libD | 30.00 | 25.00 | 24.00 | 22.09 | 21.95 | 21.94 | 20.73 | 20.67 | 20.69 |

**Table S3**. Kruskal-Wallis statistic and Benjamini-Hochberg adjusted significance values for bat guano data.

|  |  |  |  |
| --- | --- | --- | --- |
| **Filtering parameter** | **Hill Number** | **KW statistic** | **Adjusted p-value** |
| basic | 0 | 2.024 | 0.363471 |
| standard | 0 | 20.472 | 5.38E-05 |
| extra | 0 | 124.8 | 2.38E-27 |
| basic | 1 | 67.851 | 5.54E-15 |
| standard | 1 | 55.37 | 9.48E-13 |
| extra | 1 | 64.151 | 1.76E-14 |
| basic | 2 | 48.699 | 7.98E-11 |
| standard | 2 | 39.868 | 2.2E-09 |
| extra | 2 | 41.344 | 1.58E-09 |

**Table S4**. Dunn’s test for pairwise differences among denoising groups alpha diversity estimates for bat guano data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Filtering parameter** | **Hill Number** | **Comparison** | **Z** | **Adjusted p-value** |
| basic | 0 | dada2 - deblur | -1.395 | 0.489 |
| basic | 0 | dada2 - vsearch | -0.925 | 0.533 |
| basic | 0 | deblur - vsearch | 0.494 | 0.621 |
| standard | 0 | dada2 - deblur | -0.800 | 0.424 |
| standard | 0 | dada2 - vsearch | -4.275 | 0.000 |
| standard | 0 | deblur - vsearch | -3.364 | 0.001 |
| extra | 0 | dada2 - deblur | 4.605 | 0.000 |
| extra | 0 | dada2 - vsearch | 11.129 | 0.000 |
| extra | 0 | deblur - vsearch | 6.242 | 0.000 |
| basic | 1 | dada2 - deblur | 6.185 | 0.000 |
| basic | 1 | dada2 - vsearch | 7.767 | 0.000 |
| basic | 1 | deblur - vsearch | 1.386 | 0.166 |
| standard | 1 | dada2 - deblur | 5.979 | 0.000 |
| standard | 1 | dada2 - vsearch | 6.790 | 0.000 |
| standard | 1 | deblur - vsearch | 0.616 | 0.538 |
| extra | 1 | dada2 - deblur | 5.598 | 0.000 |
| extra | 1 | dada2 - vsearch | 7.736 | 0.000 |
| extra | 1 | deblur - vsearch | 1.933 | 0.053 |
| basic | 2 | dada2 - deblur | 5.146 | 0.000 |
| basic | 2 | dada2 - vsearch | 6.626 | 0.000 |
| basic | 2 | deblur - vsearch | 1.312 | 0.189 |
| standard | 2 | dada2 - deblur | 4.907 | 0.000 |
| standard | 2 | dada2 - vsearch | 5.867 | 0.000 |
| standard | 2 | deblur - vsearch | 0.793 | 0.428 |
| extra | 2 | dada2 - deblur | 4.520 | 0.000 |
| extra | 2 | dada2 - vsearch | 6.201 | 0.000 |
| extra | 2 | deblur - vsearch | 1.517 | 0.129 |

**Table S5**. Permutational Multivariate Analysis of Variance tests using guano samples collected from one location (Fox State Forest, Hillsborough NH) between April through October 2016. PERMANOVA run using three distance inputs: unweighted abundance metric (Dice-Sorensen) and weighted abundance metrics (Bray-Curtis and Morisita-Horn) on rarefied samples testing effects of denoising method (Method), filtering parameter (Filt), and date of sample collection (MonthStart).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Source** | **df** | **SS** | **MS** | **F.Model** | **R2** | **P(perm)** |
| **Dice-Sorensen model** |  |  |  |  |  |  |
| Method | 2 | 0.3904 | 0.1952 | 0.5300 | 0.0013 | 1 |
| Filt | 2 | 1.2350 | 0.6175 | 1.6766 | 0.0041 | 0.002 |
| MonthStart | 3 | 42.7005 | 14.2335 | 38.6446 | 0.1416 | 0.001 |
| Method:Filt | 4 | 0.5143 | 0.1286 | 0.3491 | 0.0017 | 1 |
| Method:MonthStart | 6 | 0.8599 | 0.1433 | 0.3891 | 0.0029 | 1 |
| Filt:MonthStart | 6 | 1.4150 | 0.2358 | 0.6403 | 0.0047 | 1 |
| Method:Filt:MonthStart | 12 | 1.6909 | 0.1409 | 0.3826 | 0.0056 | 1 |
| Residuals | 686 | 252.6656 | 0.3683 |  | 0.8381 |  |
| Total | 721 | 301.4715 |  |  | 1.0000 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Bray-Curtis model** |  |  |  |  |  |  |
| Method | 2 | 0.0992 | 0.0496 | 0.1128 | 0.0003 | 1 |
| Filt | 2 | 0.5497 | 0.2748 | 0.6255 | 0.0017 | 0.995 |
| MonthStart | 3 | 28.1871 | 9.3957 | 21.3842 | 0.0850 | 0.001 |
| Method:Filt | 4 | 0.1183 | 0.0296 | 0.0673 | 0.0004 | 1 |
| Method:MonthStart | 6 | 0.2469 | 0.0412 | 0.0937 | 0.0007 | 1 |
| Filt:MonthStart | 6 | 0.6192 | 0.1032 | 0.2349 | 0.0019 | 1 |
| Method:Filt:MonthStart | 12 | 0.2627 | 0.0219 | 0.0498 | 0.0008 | 1 |
| Residuals | 686 | 301.4112 | 0.4394 |  | 0.9093 |  |
| Total | 721 | 331.4943 |  |  | 1.0000 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Morisita-Horn model** |  |  |  |  |  |  |
| Method | 2 | 0.0586 | 0.0293 | 0.0674 | 0.0002 | 1 |
| Filt | 2 | 0.3378 | 0.1689 | 0.3888 | 0.0010 | 1 |
| MonthStart | 3 | 29.5051 | 9.8350 | 22.6365 | 0.0897 | 0.001 |
| Method:Filt | 4 | 0.0914 | 0.0229 | 0.0526 | 0.0003 | 1 |
| Method:MonthStart | 6 | 0.1866 | 0.0311 | 0.0716 | 0.0006 | 1 |
| Filt:MonthStart | 6 | 0.4741 | 0.0790 | 0.1819 | 0.0014 | 1 |
| Method:Filt:MonthStart | 12 | 0.1894 | 0.0158 | 0.0363 | 0.0006 | 1 |
| Residuals | 686 | 298.0513 | 0.4345 |  | 0.9062 |  |
| Total | 721 | 328.8944 |  |  | 1.0000 |  |

**Table S6**. Summary of per-library shared ASVs among denoising programs. Shared ASVs within a library (orange, left to right) reflect those sequences common across all filtering regimes. Shared ASVs among denoising programs (blue, top to bottom) reflect sequences common across all mock community samples/libraries. Tables are grouped by vertical faceting reflecting filtering parameters: “Basic” represents default parameters for each filtering method; “Standard” requires a sample to have > 5000 reads, and an OTU to be present in > 1 sample; “Extra” includes “Standard” filters in addition to subtracting a fixed number of reads from all observations. Horizontal faceting describes how detected sequence variants in a mock sample aligned to expected mock community reference sequences: “Exact” are 100% identity matches; “Partial” are 97-99.9% identical; “Miss” are less than 97% identical to a reference.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | **Basic** | | | | | |  |  | **Standard** | | | | |  |  | | | **Extra** | | | | |  | | |
|  | Sample | DADA2 | | Deblur | VSEARCH |  | | DADA2 | | | | Deblur | VSEARCH |  |  | | | | DADA2 | | Deblur | VSEARCH |  | | | |
| **Exact** | libA | 22 | | 23 | 19 | 19 |  | 22 | | | | 22 | 19 | 19 |  | | | | 22 | | 22 | 19 | 19 |
| libB | 22 | | 22 | 21 | 21 |  | 22 | | | | 22 | 21 | 21 |  | | | | 22 | | 22 | 21 | 21 |
| libC | 22 | | 22 | 18 | 18 |  | 22 | | | | 22 | 18 | 18 |  | | | | 22 | | 22 | 18 | 18 |
| libD | 23 | | 23 | 19 | 19 |  | 22 | | | | 22 | 19 | 19 |  | | | | 22 | | 22 | 19 | 19 |
|  |  | 22 | | 22 | 16 |  | | 22 | | | | 22 | 16 |  | | | | 22 | | | 22 | 16 |  |
|  |  |  | |  |  |  |  |  | | | |  |  |  |  | | | |  | |  |  |  |
|  | Sample | DADA2 | | Deblur | VSEARCH |  |  | DADA2 | | | | Deblur | VSEARCH |  |  | | | | DADA2 | | Deblur | VSEARCH |  |
| **Partial** | libA | 3 | | 4 | 28 | 2 |  | 2 | | | | 3 | 4 | 2 |  | | | | 2 | | 4 | 4 | 2 |
| libB | 2 | | 2 | 35 | 2 |  | 2 | | | | 2 | 6 | 2 |  | | | | 2 | | 2 | 3 | 2 |
| libC | 2 | | 2 | 59 | 2 |  | 2 | | | | 2 | 20 | 2 |  | | | | 2 | | 2 | 8 | 2 |
| libD | 2 | | 138 | 173 | 2 |  | 2 | | | | 7 | 10 | 2 |  | | | | 2 | | 4 | 6 | 2 |
|  |  | 2 | | 2 | 2 |  | | 2 | | | | 2 | 2 |  | | | | 2 | | | 2 | 2 |  | | |
|  |  |  | |  |  |  |  |  | | | |  |  |  |  | | | |  | |  |  |  | | | |
|  | Sample | DADA2 | | Deblur | VSEARCH |  |  | DADA2 | | | | Deblur | VSEARCH |  |  | | | |  | |  |  |  | | | |
| **Miss** | libA | 5 | | 60 | 67 | 1 |  | 2 | | | | 11 | 26 | 1 |  | | | |  | |  |  |  | | | |
| libB | 8 | | 96 | 184 | 2 |  | 6 | | | | 16 | 62 | 2 |  | | | |  | |  |  |  | | | |
| libC | 7 | | 45 | 90 | 5 |  | 6 | | | | 12 | 44 | 5 |  | | | |  | |  |  |  | | | |
| libD | 20 | | 364 | 412 | 8 |  | 12 | | | | 23 | 92 | 7 |  | | | |  | |  |  |  | | | |
|  |  | | 0 | 0 | 3 |  | | 0 | | | | 0 | 3 |  | | | |  | | |  |  |  | | |

**Figure S1**. Locations for New England guano sampling. Guano samples included in this project constitute a subset of a broader project that included collections from 21 locations across New Hampshire and Maine in 2015 and 2016. The samples used in this study were from sites collected in 2016 in towns listed in **Table S1**, and shown in the map of New Hampshire below. Samples collected from Fairfield ME are not shown in this map.

