

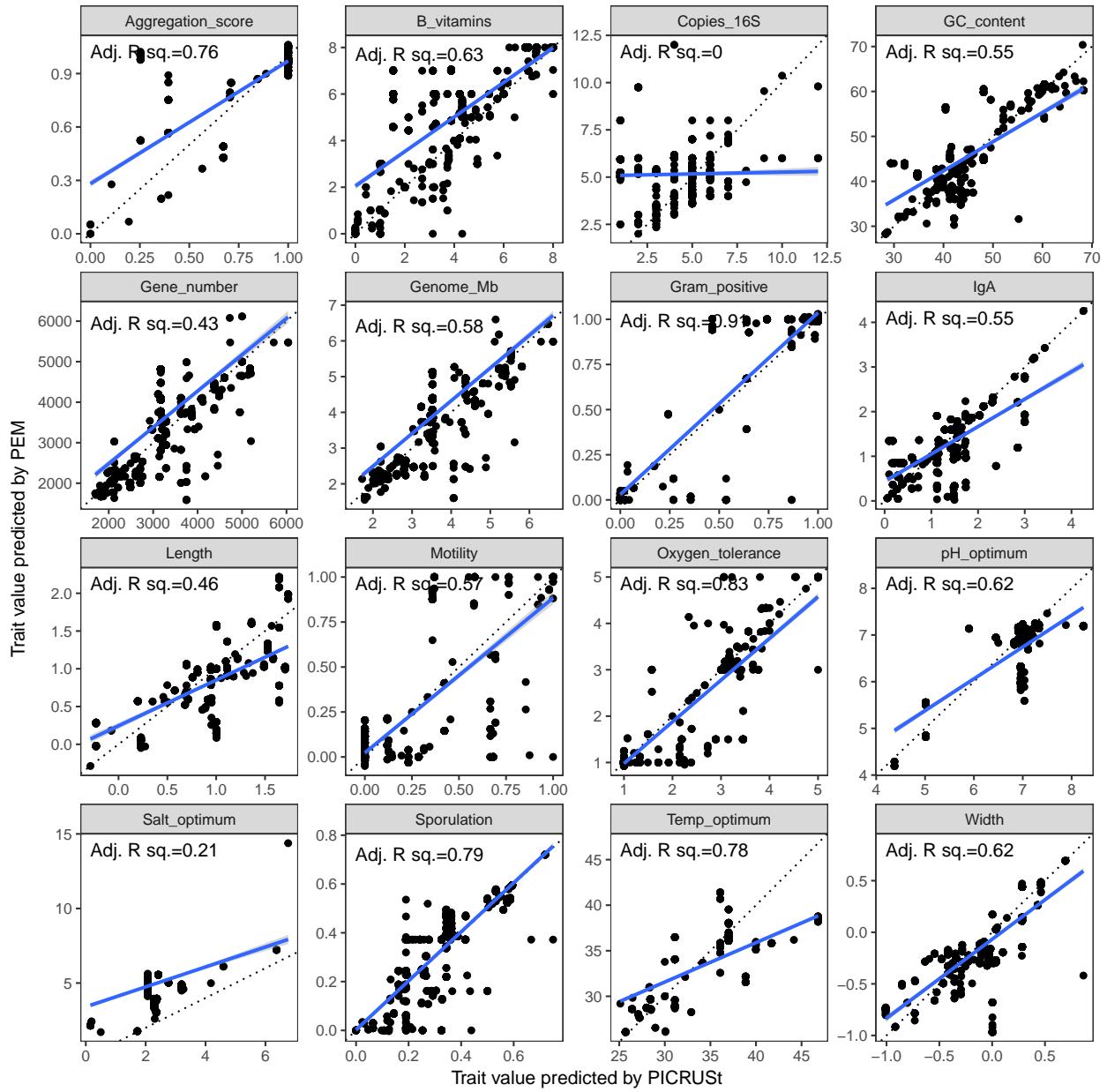
# Appendix A: Exploring methods of inferring trait values, and the effect of normalizing abundances by 16S copy number.

*John Guittar*

*March 29, 2018*

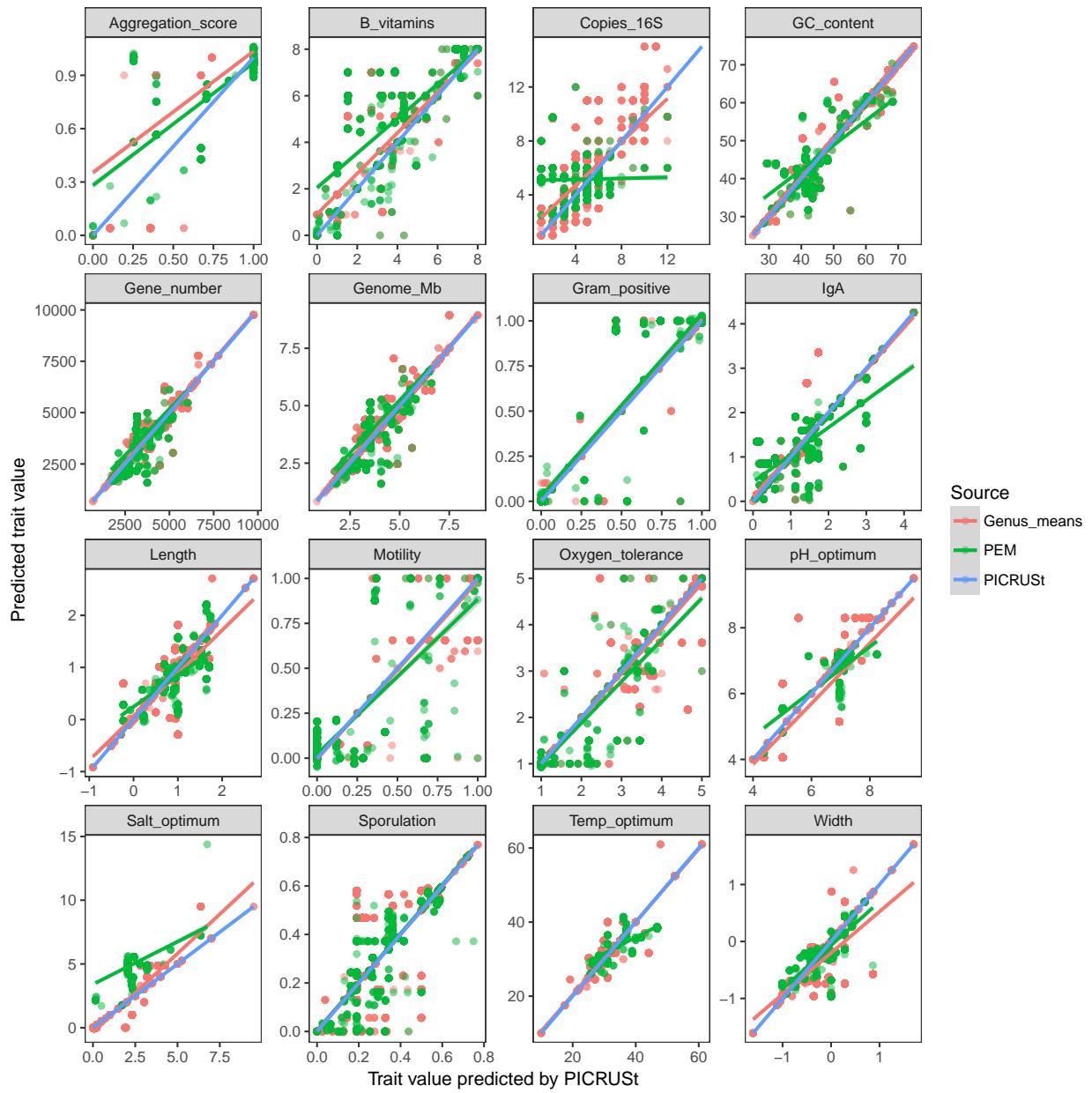
In this Appendix, we explore two questions:

1. How do different methods of inferring missing trait values compare? How do they change our conclusions, if at all?
2. To what degree does normalizing by 16S copy number alter our conclusions, if at all?



**Figure A-1. Comparison of trait predictions based on the PICRUSt algorithm and Phylogenetic Eigenvector Maps (PEM).**

R-squared values are printed in the top left of each panel. Blue lines show linear regressions, whereas black dotted lines show a 1:1 abline. Note that all known trait values are removed; i.e., only *predicted* trait values are shown here. Also note that there is something way off with our 16S copy number estimates (I think I know what is going on...).



**Figure A-2. Comparisons of trait predictions derived using genus means, PEM, and PICRUSt, plotted by PICRUSt predictions.**

Colored lines show linear regressions.

Trait	GenusMeans_vs_PEM	GenusMeans_vs_PICRUSt	PICRUSt_vs_PEM	Mean
Gram_positive	1.00	0.99	0.96	0.98
Oxygen_tolerance	0.99	0.98	0.91	0.96
Temp_optimum	0.93	0.96	0.88	0.92
Sporulation	0.98	0.89	0.89	0.92
Motility	0.99	0.98	0.76	0.91
GC_content	0.98	0.99	0.74	0.90
Genome_Mb	0.98	0.93	0.76	0.89
IgA	0.99	0.92	0.74	0.89
B_vitamins	0.99	0.86	0.80	0.88
Gene_number	0.98	0.92	0.66	0.85
Width	0.93	0.78	0.79	0.83
Length	0.89	0.80	0.68	0.79
Salt_optimum	0.95	0.93	0.46	0.78
Aggregation_score	0.65	0.77	0.87	0.76
pH_optimum	0.85	0.64	0.79	0.76
Copies_16S	0.95	0.66	0.04	0.55

**Table 1. Correlation statistics among trait prediction methods**

Numbers show pearson correlation coefficients (r). The far right column shows the mean among all three predictions.

## How does prediction method alter the qualitative outcomes of our trait-based analysis?

Again, note that the 16S predictions from PICRUSt are oddly off...

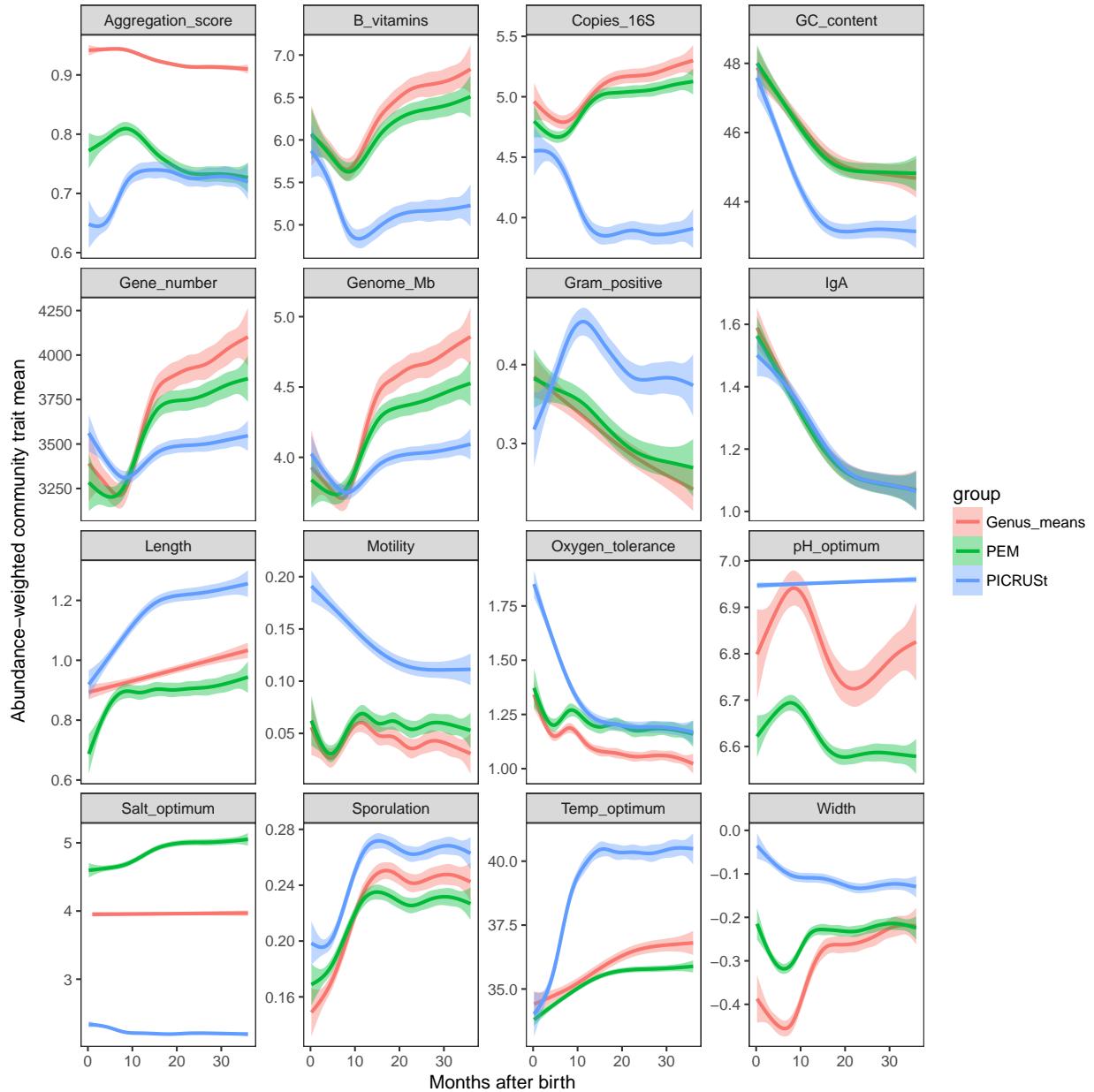


Figure A-3. Abundance-weighted trait-based patterns over time in infant gut microbiomes, using different methods to infer unknown traits.

2. To what degree does normalizing by 16S copy number alter our conclusions, if at all?

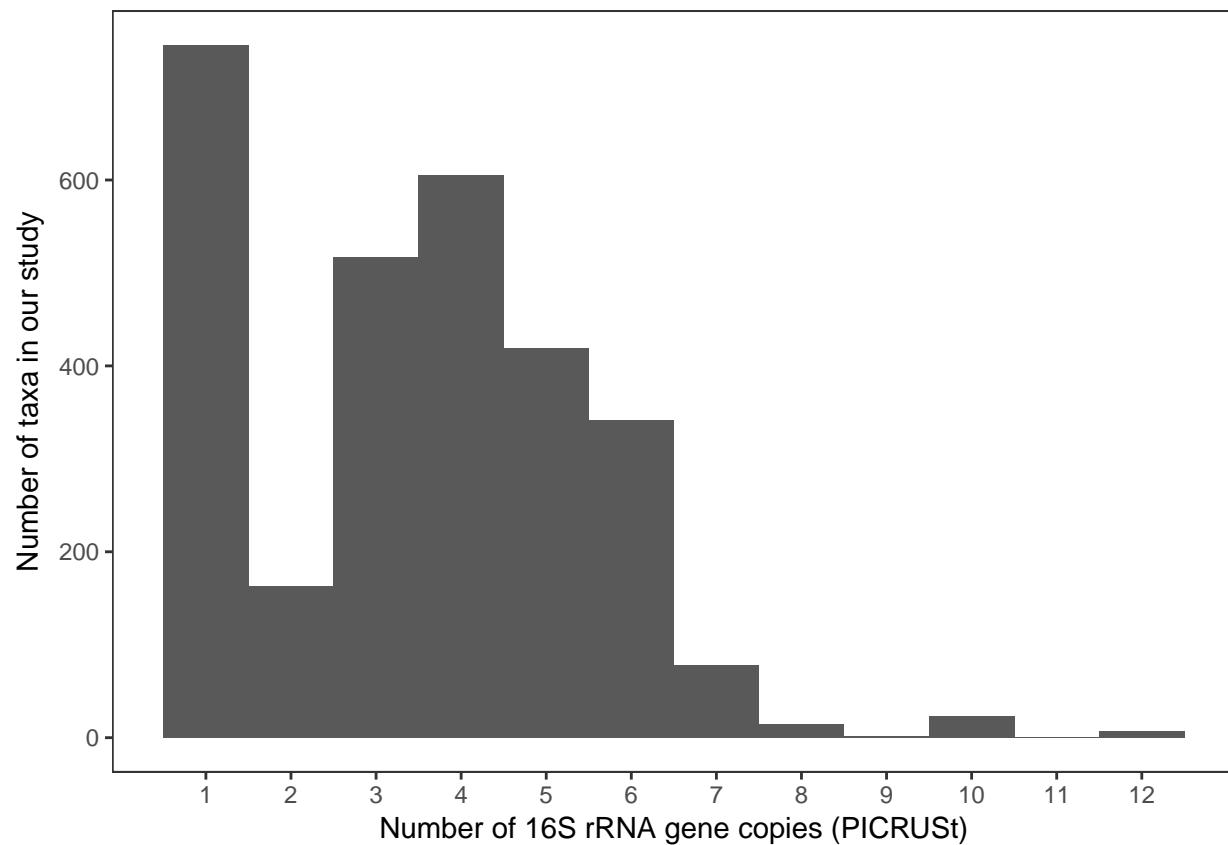
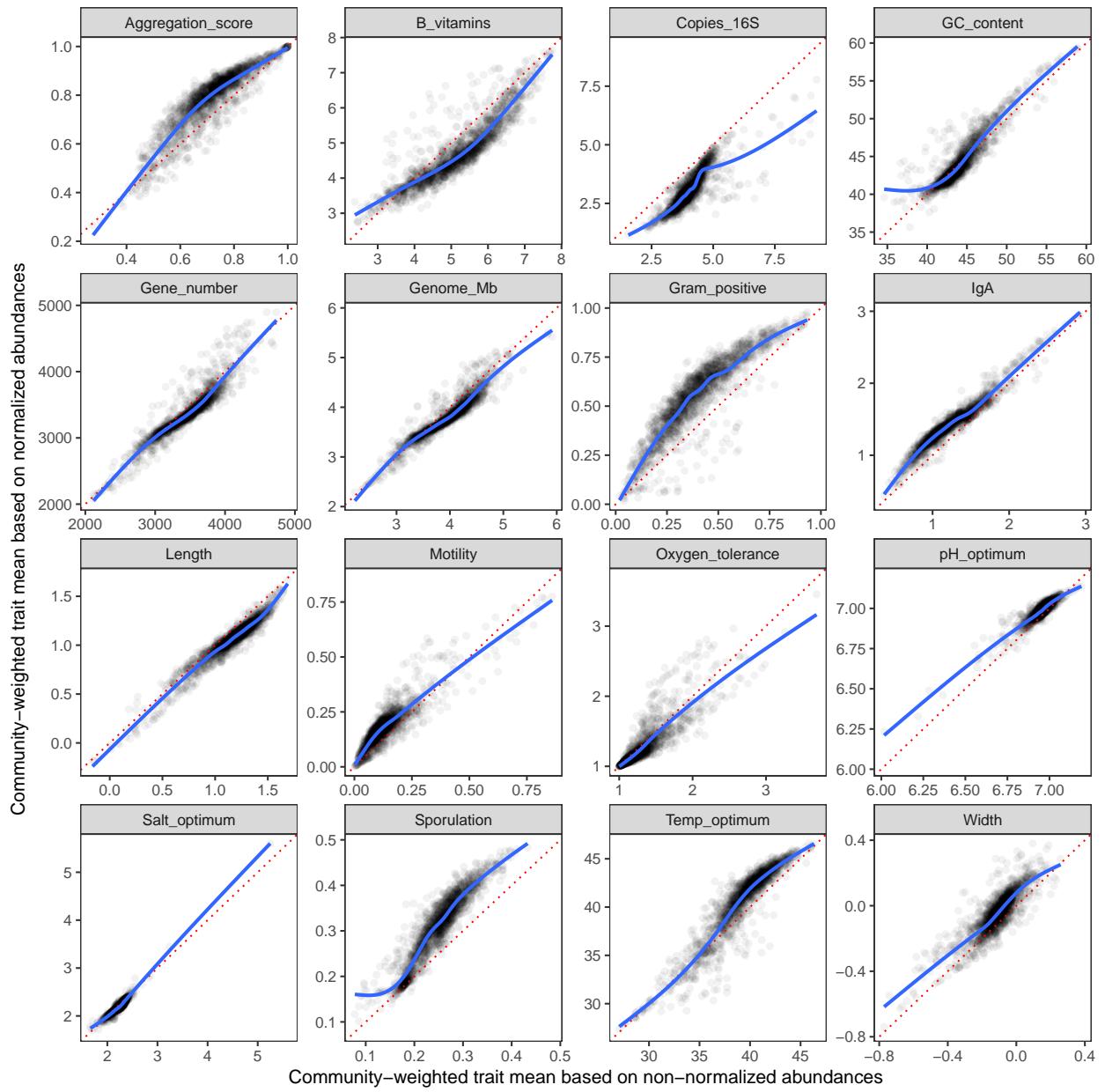
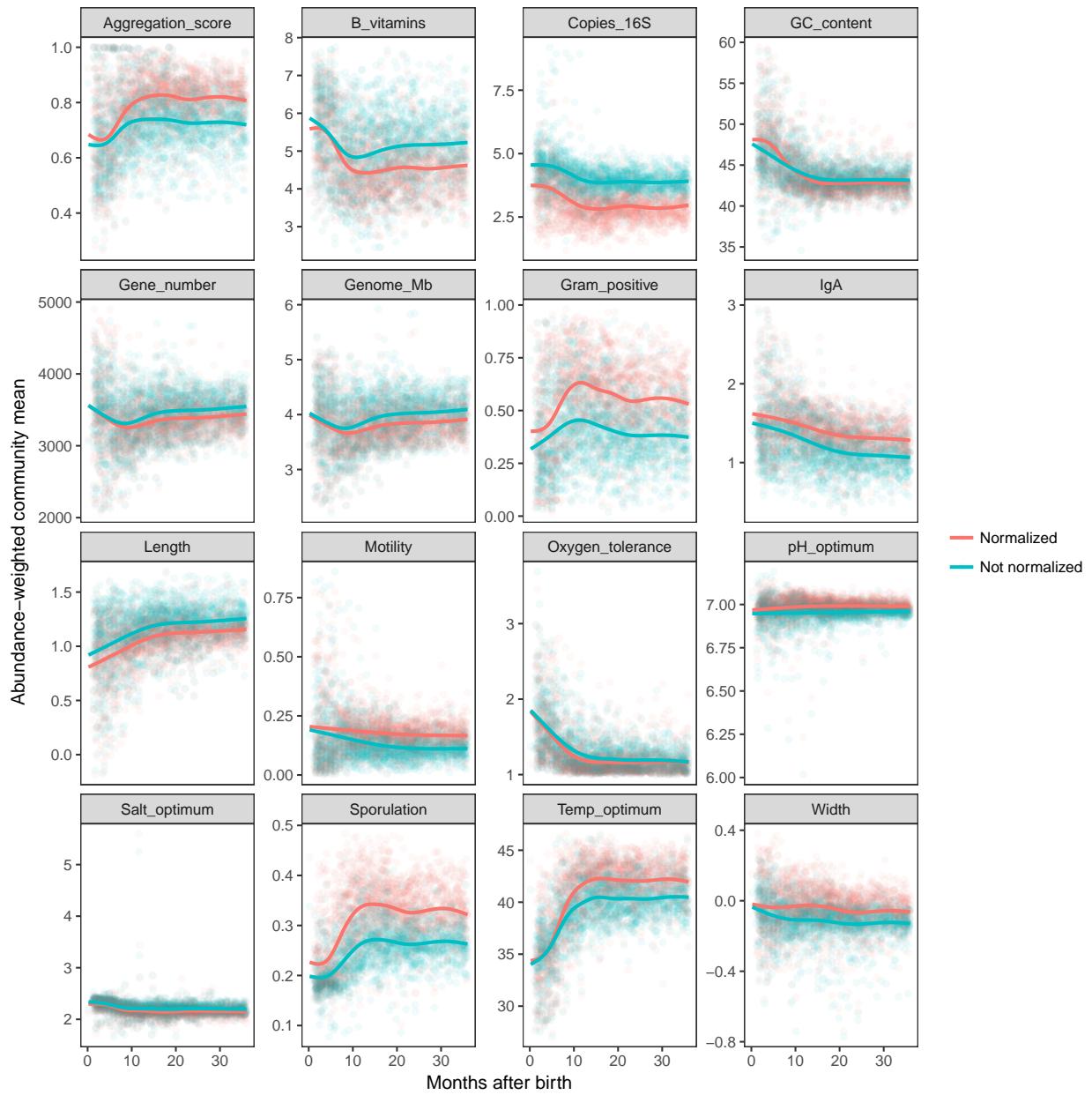


Figure A-4. A histogram of 16S Gene copy numbers for taxa in our study



**Figure A-5. A comparison of community-weighted trait means using abundances normalized and non-normalized by 16S gene copy number.**

Blue lines represent LOESS curves; dotted red lines represent 1:1 ablines.



**Figure A-6. A comparison of community-weighted trait means using abundances normalized and non-normalized by 16S gene copy number, plotted over time.**

Colored lines represent LOESS curves.