

Global analysis of gene expression dynamics within the marine microbial community during the VAHINE mesocosm experiment in the South West Pacific

U. Pfreundt¹, D. Spungin², I. Berman-Frank², S. Bonnet³, W. R. Hess¹

Supplementary Figures

Figure S1. The dominating KEGG categories in mesocosm M1. Sequencing fragments were mapped to categories (A) higher up in the BRITE hierarchy (Level 1) and (B) one level further down (Level 2). All fragments mapped to the overarching top-categories “Metabolism”, “Genetic Information Processing”, “Environmental Information Processing”, and “Cellular processes” were counted as 100% to get the relative importance of each category in each sample. Note that the functional diversity is high, with the 20 top categories on the lower level only making up around 50% of all assignments.

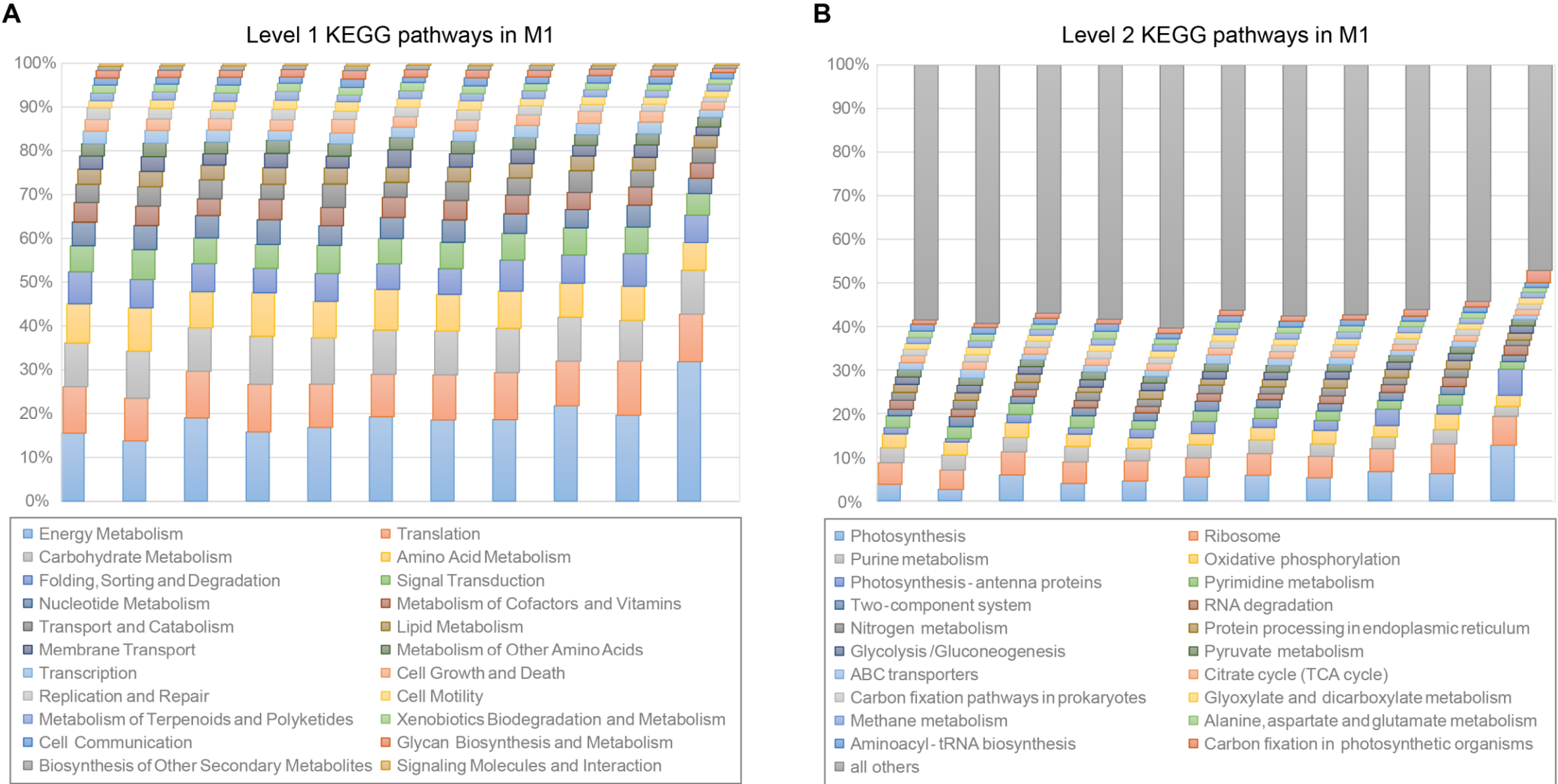
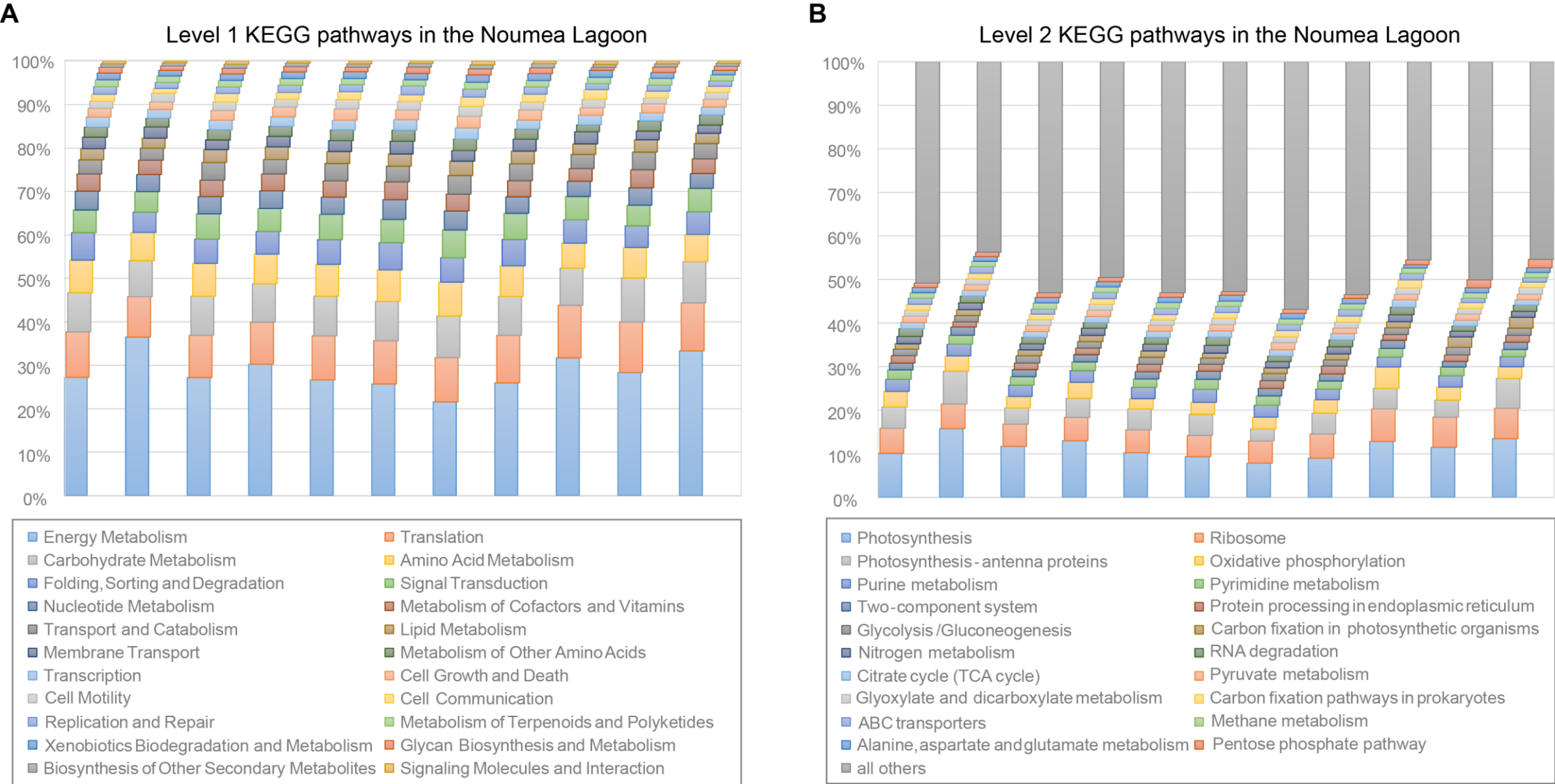


Figure S2. The dominating KEGG categories in the Noumea lagoon. Sequencing fragments were mapped to categories (A) higher up in the BRITE hierarchy (Level 1) and (B) one level further down (Level 2). All fragments mapped to the overarching top-categories “Metabolism”, “Genetic Information Processing”, “Environmental Information Processing”, and “Cellular processes” were counted as 100% to get the relative importance of each category in each sample.

Note that the functional diversity is high, with the 20 top categories on the lower level only making up around 50% of all assignments.



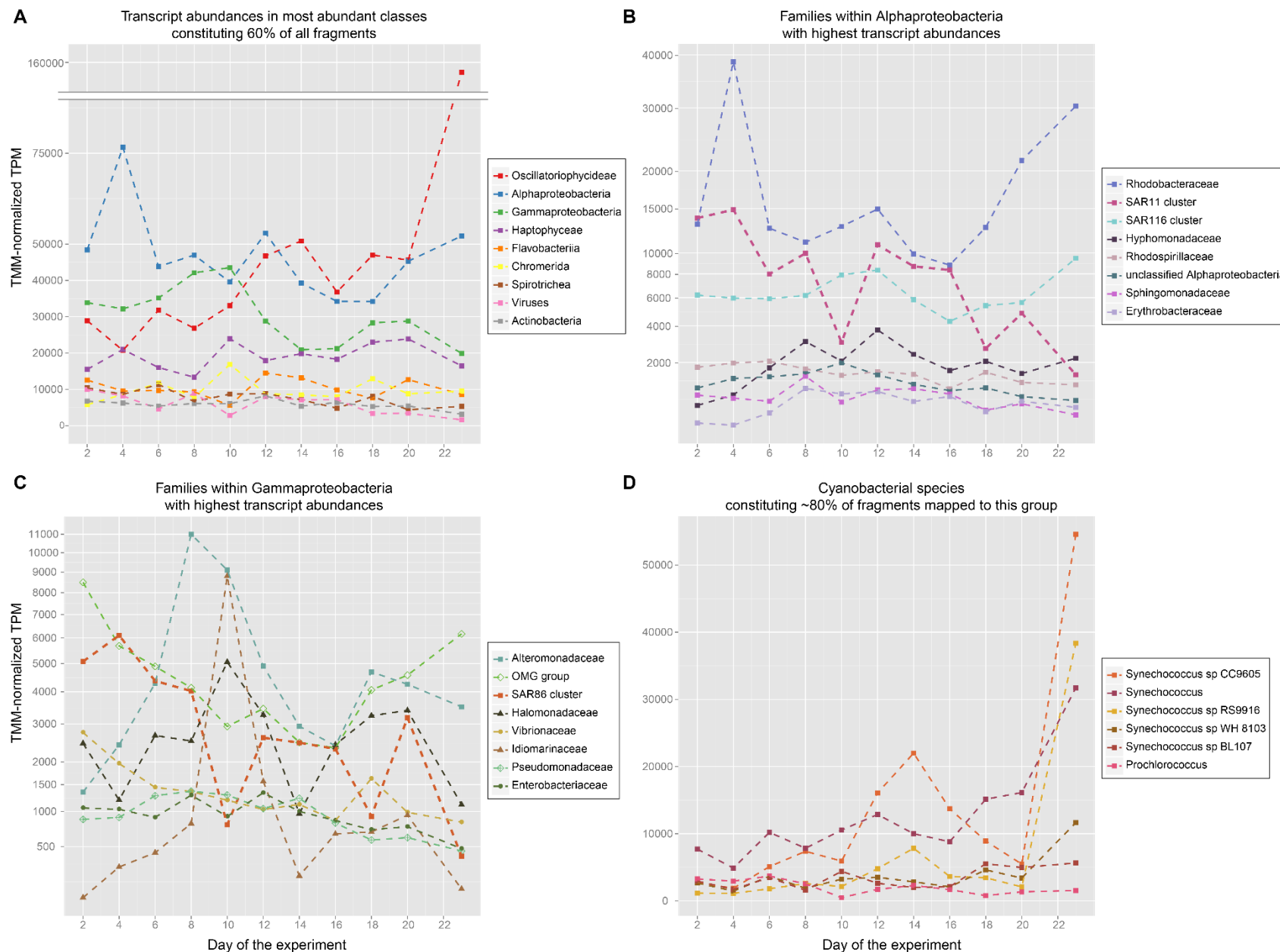


Figure S3. Taxonomic affiliation of transcripts from **mesocosm M1**. (A) Transcriptionally most active taxa (class level), sorted by mean expression values over all M1 samples. These classes together constituted 60 % of all mapped sequence fragments within M1 (sum of all samples). (B, C) Transcriptionally most active families within alpha- and gammaproteobacteria, respectively. Note the square-root scale for the y-axis. (D) Transcriptionally most active cyanobacterial species. These species together accounted for 80 % of all fragments mapped to cyanobacteria. Gene expression is presented in TPM (transcripts per million transcripts sequenced), normalized in between samples by TMM normalization (edgeR). Thus, plots can be directly compared, but values are relative.

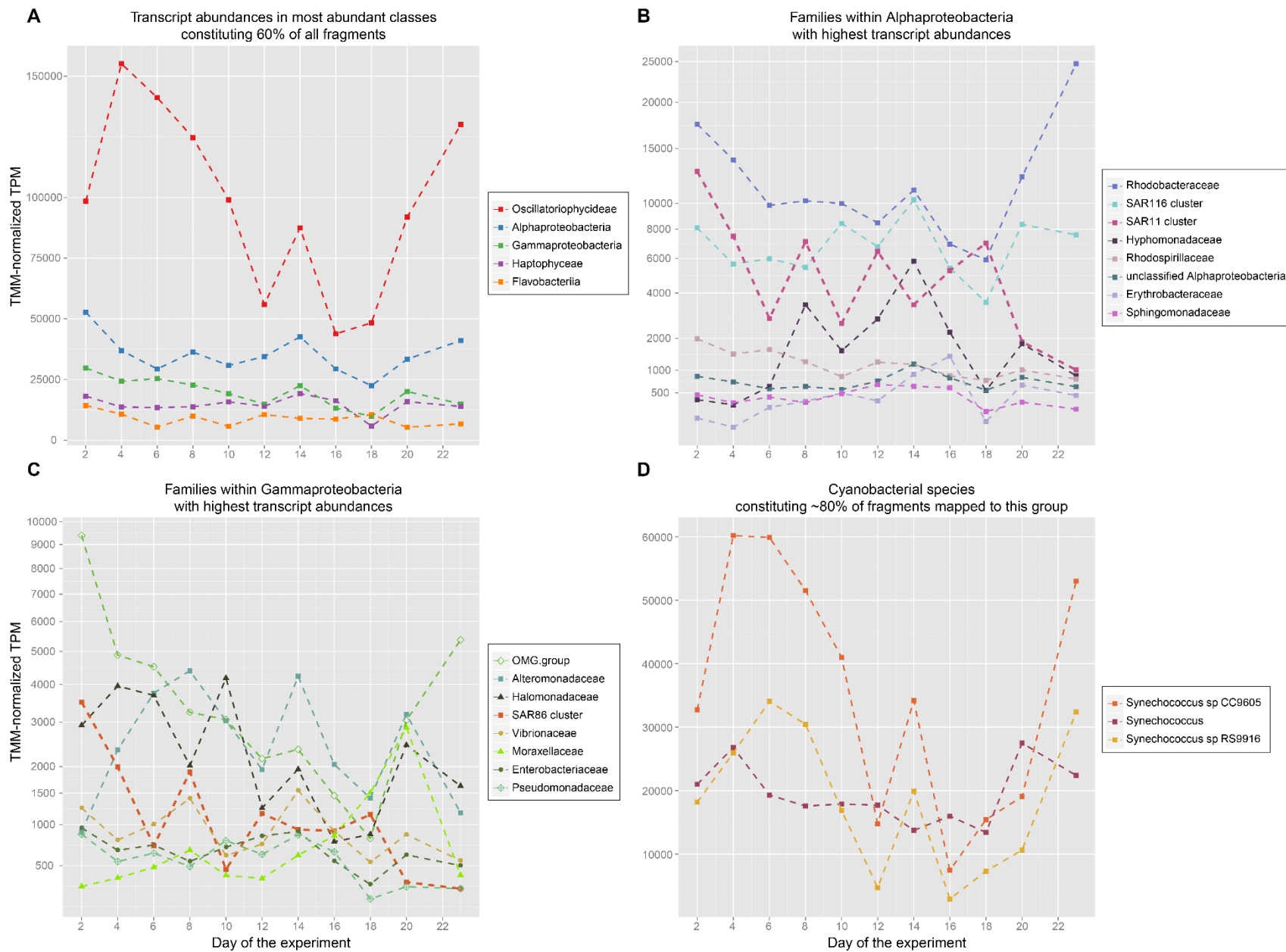
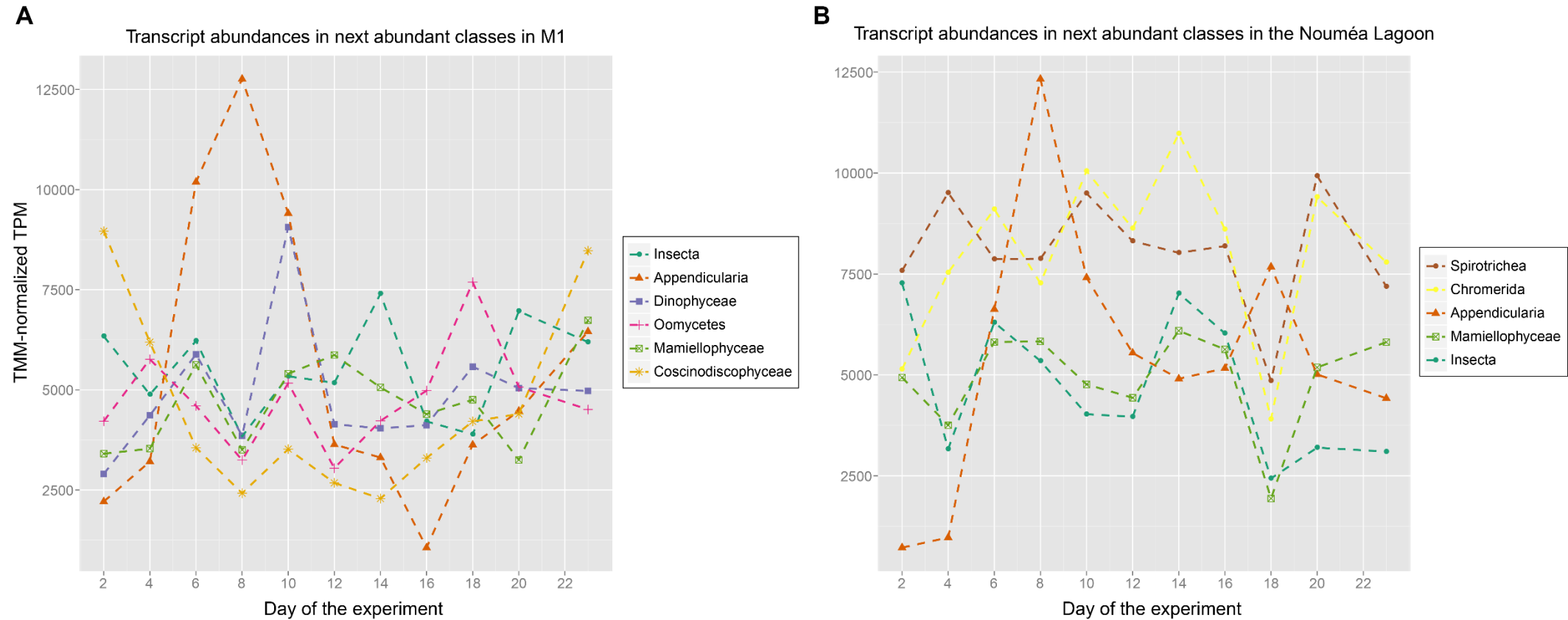


Figure S4. Taxonomic affiliation of transcripts inside the Nouméa lagoon. (A) Transcriptionally most active taxa (class level), sorted by mean expression values over all Nouméa lagoon samples. These classes together constituted 60 % of all mapped sequence fragments from all Nouméa lagoon samples. (B, C) Transcriptionally most active families within the alpha- and gammaproteobacteria, respectively. Note the square-root scale for the y-axis. (D) Transcriptionally most active cyanobacterial species. These species together accounted for 80 % of all fragments mapped to cyanobacteria. Gene expression is presented in TPM (transcripts per million transcripts sequenced), normalized in between samples by TMM normalization (edgeR). Thus, plots can be directly compared, but values are relative.

Figure S5. Taxonomic affiliation of transcripts on class level inside mesocosm M1 (A) and in the Nouméa Lagoon (B). Both plots show those taxonomic classes that rank (mean TPM over all M1 or Nouméa Lagoon samples) right after those shown in Fig. S3A and S4A and constitute the next 10% of all mapped sequence fragments. Together with those shown in Fig. S3A and S4A these classes constitute 70% of all mapped fragments. Note that diversity is generally higher inside M1. Transcriptional activity is presented in TPM (transcripts per million transcripts sequenced), normalized in between samples by TMM normalization (edgeR). Thus, plots can be directly compared, but values are relative.



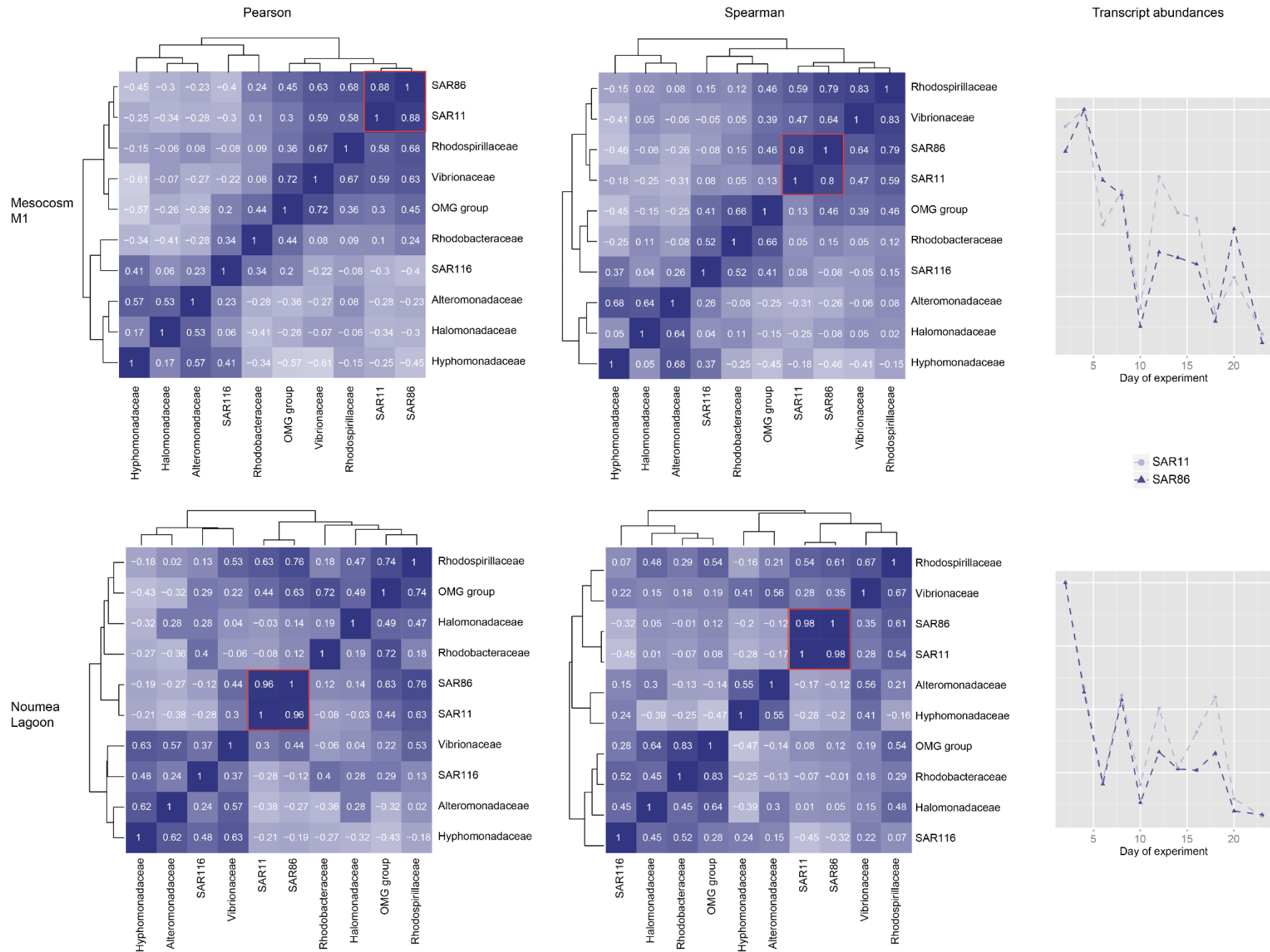


Figure S6. Pearson and Spearman rank correlation analysis of the top 5 ranked alpha- and gammaproteobacterial families with regard to their transcript abundances in M1 and the Noumea lagoon. Values in the cells of the heatmap represent the correlation values. SAR11 and SAR86 are boxed together in red.

On the right side, the transcript abundances of SAR11 and SAR86 are plotted over time for M1 (top) and Noumea lagoon (bottom), respectively.

The y-axis was scaled to the maximum value for each data series. The strong correlation of SAR86 and SAR11 transcripts seen in Fig. S3 and S4 was confirmed.

Figure S7. All 14 different Yfr103 transcripts from the list of top 100 transcripts (ranked after mean abundance; Table S1), ranked according to their mean transcript abundance in M1. Twelve could be assigned to *Synechococcus*, one to *Prochlorococcus* and one to picocyanobacteria, because it was too similar to both *Prochlorococcus* and *Synechococcus* homologs. Note the sqrt-scale and that Yfr103 transcript abundance followed the same trends as total *Synechococcus* transcript abundance.

