Supporting Information Appendix:

Localized coevolution between microbial predator and prey alters community-wide gene expression and ecosystem function

Authors: ⓘ Shane L. Hogle[1,*], Liisa Ruusulehto[2], ⓘ Johannes Cairns[3,4], ⓘ Jenni Hultman[2,5], ⓘ Teppo Hiltunen[1,2*]

[1]Department of Biology, University of Turku, Turku, Finland
[2]Department of Microbiology, University of Helsinki, Helsinki, Finland
[3]Department of Computer Science, University of Helsinki, Helsinki, Finland
[4]Organismal & Evolutionary Biology Research Programme, University of Helsinki, Finland
[5]Natural Resources Institute Finland, Helsinki, Finland

[*]**Corresponding Authors:** Shane Hogle and Teppo Hiltunen

Email: shane.hogle at utu.fi and teppo.hiltunen at utu.fi

This PDF file includes:
SI Text
SI Figures S1 to S8
SI Tables S1 to S5

Contents

List of Supplementary Figures

List of Supplementary Tables

1   Rendering coevolved ciliates axenic

After two years of coculture, 5 ml of the ciliate and SBW25 coculture was pipetted into 100 ml Proteose Peptone Yeast Extract (PPY) broth containing 50, 50, 42, and 17 µg ml^{-1} streptomycin, rifampicin, kanamycin, and tetracycline, respectively. Cocultures with antibiotics were grown at 28°C with continuous shaking at 70 rpm for five days. Next, 10 ml of these cocultures were inoculated again into 100 ml of antibiotic cocktail and grown at 28°C with shaking for two days. This procedure was repeated two times (21 days total), after which cultures were centrifuged (1700 rcf, 8 minutes, 4°C) and supernatant was decanted. Cell pellets were resuspended and added to 100 ml of PPY medium and grown for 3 days to enrich for ciliates. Axenicity of the purified ciliate cultures was checked by plating 100 µl of culture directly on 50% PPY agar plates, incubating at 28°C for 7 days, and checking for visible bacterial growth.

2   Calculating generation times

Under the conditions used in our experiment *Tetrahymena* generally reaches stationary phase after approximately 90 hours while *Pseudomonas* SBW25 reaches stationary phase after less than 24 hours. Each 100-fold dilution per serial transfer represents at least 6.6 ciliate generations ($T_d = 168\,\text{hrs} \times \frac{log(2)}{log(100)} = 25.3\,\text{hrs}$, $G_t = 168\,\text{hrs}/25.3\,\text{hrs} = 6.6\,\text{generations}$). It is more challenging to estimate the number of elapsed bacterial generations because SBW25 is in stationary phase for much of the serial transfer interval. If we assume synchronized growth between the ciliate and SBW25 then SBW25 would also have 6.6 generations per transfer interval. Thus, we expect that over the 111 transfer intervals of the coevolutionary experiment **at least** 737 *Tetrahymena* and *Pseudomonas* generations would have elapsed. However, the number of elapsed generations for SBW25 is likely much greater.

3   Gene and nucleotide parallelism from SBW25 long-term coevolution lines

We define parallelism as the degree to which distributions of mutations across genomes from independent *Pseudomonas fluorescens* SBW25 coevolution lines deviate from the null expectation following the general statistical framework of Good *et. al* [1]. Briefly, our null expectation under neutral evolution with a constant mutation rate across genes and intragenic regions is that mutations should be randomly distributed across the genome. At the gene level we would expect that the number of mutations per gene should be proportional to gene length.

3.1   Nucleotide level

We define nucleotide parallelism as the number of mutations occurring at the same site in the genome in independent *Pseudomonas fluorescens* SBW25 coevolution lines. For each site, we defined the multiplicity, m_i, as the number of coevolution lines with a mutation detected at that site, so that the multiplicity in this experiment could range from one to three.

We compare these observations to a naive expectation that mutations should be scattered randomly across the genome. The expected number of mutations with $m_i \geq m$ in a sample of total mutations size n_{tot} is:

$$S(m) \approx \sum_{n \geq m} \frac{n}{n_{tot}} \cdot L_{tot} \cdot \frac{\left(\frac{n_{tot}}{L_{tot}}\right)^n}{n!} e^{-n_{tot}/L_{tot}} \tag{1}$$

where L_{tot} is the total number of bases in the SBW25 genome. Nucleotide parallelism results are presented in Figure S1A, and the observed data show an excess of nucleotide parallelism relative to our simple null expectation. Specifically, we identify five base positions where the same mutation occurred in three independent coevolution lines and 11 base positions where the same mutation occurred in two independent coevolution lines.

3.2   Gene level

If selection pressures and mutation rates did not vary between genes, the number of mutations in each gene should be proportional to the target gene size. While it is difficult to estimate the local target size for beneficial, deleterious, and neutral mutations in any particular gene, we assume that gene length is a primary driver of the target size. Similar to our nucleotide-level analysis above, we then define a multiplicity for each gene according to:

$$m_i = n_i \cdot \frac{\overline{L}}{L_i} \tag{2}$$

where n_i is the number of mutations in $gene_i$ across all coevolution lines, L_i is the total number of sites in $gene_i$, and $\overline{L}$ is the average value of L_i across all genes in the genome. This definition ensures that under the null hypothesis, all genes have the same expected multiplicity $\overline{m} = \frac{n_{tot}}{n_{genes}}$.

We next define a null model for gene multiplicity that assumes mutations are assigned to genes with probability:

$$p_i \propto L_i r_i \tag{3}$$

for some set of enrichment factors r_i. In the alternate model, the maximum likelihood estimator of the enrichment factors r_i is the ratio of observed to expected gene multiplicities, $r_i = m_i/m$. The net increase of the log-likelihood relative to the null model ($r_i = 1$) is given by:

$$\Delta \ell = \sum_i n_i \log\left(\frac{m_i}{\overline{m}}\right) \tag{4}$$

This likelihood ratio estimator is equivalent to the total G-score introduced in earlier work [2]. The number of parallel mutated genes across the three coevolution lines exceeded our expectation ($p = 0.01$) under the null hypothesis (Figure S1B), where significance was assessed using permutation tests following Shoemaker *et. al* [3]. For permutations tests (n = 10,000) we randomly subsample mutations to an n_{tot} set size of 50 using the multinomial distribution, where the probability of sampling a mutation at $gene_i$ was given by $p_i = n_i/n_{tot}$.

We next identify specific genes enriched for mutations following the criteria of Good *et. al* [1]. We calculate a p value for each gene using the equation:

$$P_i = \sum_{n \geq n_i} \frac{\left(\frac{n_{tot}L_i}{\overline{L}N_{genes}}\right)^n}{n!} e^{-\frac{n_{tot}L_i}{\overline{L}N_{genes}}} \qquad (5)$$

Here we only consider genes with $n_i \geq 3$ to exclude low p values driven primarily by gene length. Under the null hypothesis, the expected number of genes with $P_i \geq p$ can be found using a Poisson survival curve given by:

$$\overline{N}(P) \approx \sum_{i=1}^{N_{genes}} \sum_{n=3}^{\infty} \theta(P - P_i(n, L_i)) \cdot \frac{\left(\frac{n_{tot}L_i}{\overline{L}N_{genes}}\right)^n}{n!} e^{-\left(\frac{n_{tot}L_i}{\overline{L}N_{genes}}\right)} \qquad (6)$$

We can compare this expected number to the observed number of genes $N(P)$ using a critical p value (P^*) such that

$$\frac{\overline{N}(P^*)}{N(P^*)} \leq \alpha \qquad (7)$$

for a given FDR α value of 0.05 (Figure S1C). For this value of $P^*(\alpha) = 1.6 \times 10^{-4}$, we then define the set of significantly enriched genes as:

$$I = \{i : P_i \leq P^*(\alpha)\} \qquad (8)$$

The list of these genes is available in Table S2 and plotted in Figure 2 from the main text.

4 Gaussian process regression

We modeled the log-transformed cell concentrations, optical density, and ATP concentrations using additive Gaussian Processes (GP), a nonparametric, supervised learning method for solving probabilistic regression or classification models. Gaussian processes are arbitrary collections of variables, each normally distributed, any finite number of which are jointly distributed as multivariate normal. Because a Gaussian process can fit any arbitrary functional form, they are useful for modeling nonlinear and nonstationary covariate effects and are widely used in time-series modeling [4]. Using the R package lpgr v1.04 [5], we modeled microcosm identifier, the presence of coevolved or ancestral SBW25, the presence of the coevolved ciliate, and the interaction between SBW25 and the ciliate each as time-dependent deviations from a shared time effect. The rapid spike in ATP concentrations and cell density upon transfer to new media at day 41 was modeled independently as a non-stationary variance-masked onset effect. We used the zero-sum kernel for categorical covariates, the variance-masked kernel for the onset effect, and a standard Gaussian process with exponentiated quadratic kernel for the time [5]. We ran four independent MCMC chains, each with 4000 iterations, and discarded the first half of each chain as burn-in. The mixing of chains was sufficient ($\hat{R} < 1.05$) [6]. To infer whether covariates were informative for the models, we used a probabilistic covariate selection method which estimates the proportion of variance explained for each covariate and noise and selects those covariates that, alongside noise, explain 95% of the variance [5].

5 Differential abundance of bacterial species

We modeled bacterial abundances from 16S amplicon count data using a hierarchical beta-binomial regression model [7] with experimental parameters (predation, SBW25 evolution, and their interaction) included as covariates. The replicate number per condition was not sufficient for the χ_r^2 approximation in a conventional likelihood ratio test comparing nested models for covariate significance. We instead used a parametric bootstrap procedure with 1000 bootstrap replicates to compute P-values for each covariate using the likelihood ratio test with nested models. P-values from the 30 individual statistical tests were adjusted to a false discovery rate of 5% ($q \leq 0.05$). These models all account for different sample sizes/depths, which eliminates the need for discarding information by rarefaction.

6 Transcriptomic sequence processing

Community total cDNA read pairs were also processed using BBDuk (version 38.61b (https://sourceforge.net/projects/bbmap/) to remove contaminants, trim reads that contained adapter sequence, right quality trim reads where quality drops below Q10, and exclude reads with more than 2 Ns. Optical duplicates and tile-edge duplicates (identical reads within 12000 pixels on the S4 flow cell) were also removed using clumpify.sh from BBMap. BBMap was then used to exclude reads mapping to 16S, 5S, 23S or 18S rRNAs from the 30 bacterial species or *Tetrahymena* based on exact matches. The remaining reads were then mapped against the 30 bacterial genomes and the assembled *Tetrahymena thermophila* SB210 macronuclear genome (NCBI assembly GCF_000189635.1) [8] using bbsplit.sh. Reads were discarded if they mapped ambiguously to multiple species genomes (multiple top-sites scoring above a penalty threshold, see bbmap documentation). Only the best scoring site was retained for reads mapping ambiguously within the same genome sequence. Transcript counts were estimated using featureCounts [9] with bacterial species gene coordinates predicted using prodigal [10]. Gene annotations in bacterial species were generated using Prokka [11] and Gene Ontologies and KEGG pathways were assigned using EggNOG-mapper 2.0.0 [12] and EggNOG 5.0 [13]. Other noncoding RNAs were excluded from all subsequent analyses by excluding reads mapping to anything other than the coding sequences annotated by Prokka. The median number of reads mapping to HAMBI species or *Tetrahymena* genome coding sequences was 1.29×10^6 (0.97%) with a range of 0.35×10^6 (0.35%) to 2.69×10^6 (3.53%).

7 Within-taxon sum scaling normalization

To account for potential confounding variation in underlying gene copy number (i.e., due to shifts in species abundance) between samples, we included a customized normalization factor for within-taxon sum scaling and an amplicon-based estimate of the source taxon's relative abundance as a covariate in the DESeq2 model [14] following general recommendations from *Zhang et al.* [15]. Thus, any variation in gene expression due to the variation in underlying DNA template is controlled for in our results.

To detect differentially abundant transcripts DESeq2 uses a generalized linear model of the form:

$$K_{ij} \sim \text{NB}(\mu_{ij}, \alpha_i) \tag{9}$$

$$\mu_{ij} = s_j q_{ij} \tag{10}$$

$$\log_2(q_{ij}) = x_j \beta_i \tag{11}$$

where counts K_{ij} for gene i, sample j are modeled using a negative binomial distribution with fitted mean μ_{ij} and a gene-specific dispersion parameter α_i. The process of fitting the model involves:

1. Estimating sample-specific size factors s_j

2. Estimating the dispersion parameter α_i which defines the relationship between the observed transcript count and its mean across samples

3. Fitting a negative binomial generalized linear model to estimate the coefficients β_i which gives the log2 fold changes for gene i for each column of the model matrix X. The fitted mean μ_{ij} is composed of a sample-specific size factor s_j and a parameter q_{ij} proportional to the expected true concentration of fragments for sample j

To perform sample specific within-taxon sum scaling we generalize equation 10 above to:

$$\mu_{ij} = NF_{ij}q_{ij} \tag{12}$$

where normalization factor matrix NF is of the same dimensions as the count matrix K. We used the relative abundance of each species derived from amplicon data for the normalization matrix, and divided out the gene-wise geometric means of the normalization matrix. This ensures that NF has row-wise geometric means of 1 so that the mean of normalized counts for a gene is close to the mean of the raw counts. The custom normalization matrix (NF_{ij}) was also used for the regularized log-transform of count data to the $\log_2$ scale.

Supplementary Figures:
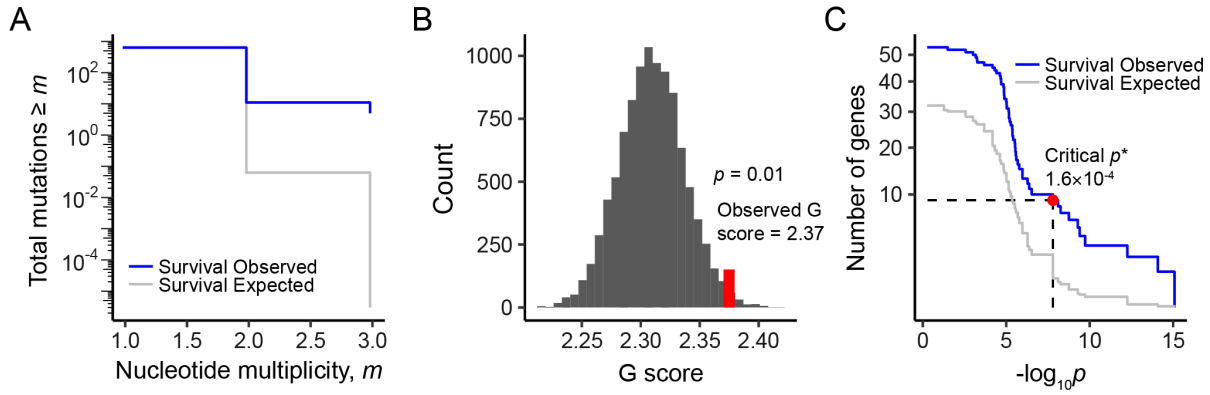
Supplementary Figure S1. Genetic parallelism at the nucleotide and gene-level in SBW25

A) The distribution of nucleotide multiplicity across the three independent *Pseudomonas fluorescens* SBW25 coevolution lines. The observed data is shown in blue where the expectation under a null model is in grey. **B)** Distribution of G scores from 10,000 null model simulations (histogram) and the observed value in red. The G score is the ratio of observed and expected multiplicities at the level of genes. **C)** The observed and expected number of genes with $n_i \geq 3$ and $P_i \leq P$ as a function of the p value. The red point denotes the false discovery rate-controlled significance threshold (P^*) for $\alpha = 0.05$ $\left(P^* = 1.6 \times 10^{-4}\right)$. At this significance level nine genes are significantly enriched for mutations across independent long-term coevolution lines (Table S2).

Supplementary Figure S2. Additive modeling of predator/prey abundances and community metabolic potential

Gaussian process regression models for bacterial prey density (OD_{600} or colony forming units ml^{-1}), ciliate predator density (cells ml^{-1}), and community metabolic potential (ATP assay fluorescence normalized to optical density). Results for each measured parameter are in rows. **A)** shows the modeled sum of the additive effects with $2\times$ standard deviation (gray bands) overlaid with observations (as in Figure 2 from the main text). **B)** shows the mean ($\pm$ stdv) for the relevances of each covariate in the regression model. Terms in black are significant for the final model. **C)** shows the covariate-specific effect for each covariate in the model. Covariates explaining over 95% of variance with noise are considered to be significant for the predictive performance of the model (black). Covariates in gray do not add to the predictive performance of the model.

Supplementary Figure S3. Relative abundance of bacterial strains across treatments and replicates

Species composition assessed via 16S rRNA gene amplicon sequencing at days 4, 41 and 45. Species recruiting less than 0.1% of reads are collapsed into the "Others" category.
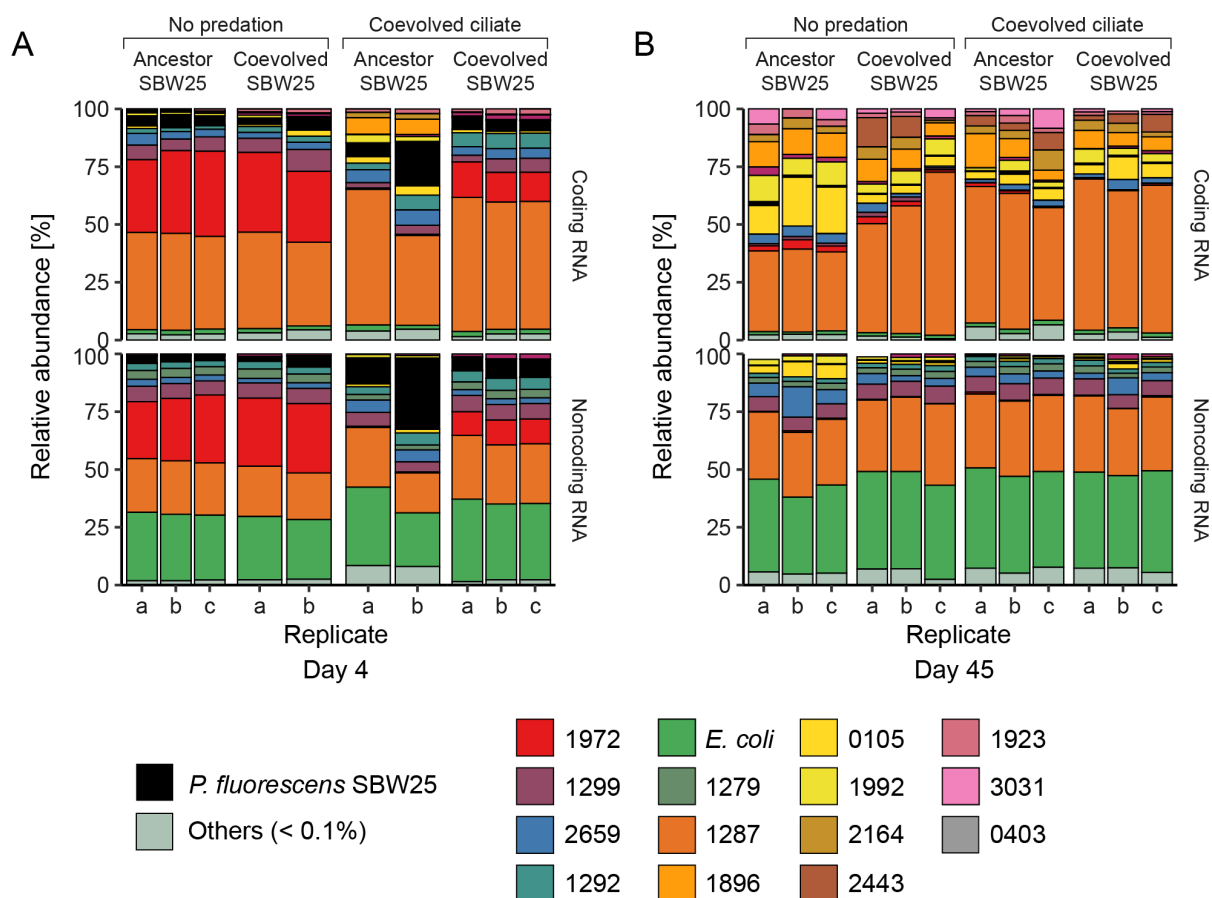
Supplementary Figure S4. Taxonomic composition of coding and noncoding RNA

A) Metatranscriptomic results from day 4 and **B)** results from day 45. RNA type is separated by protein coding potential (top row) and noncoding (bottom row). The percentage of RNA (coding or noncoding) assigned to a species is scaled relative to the total number of coding sequences and noncoding sequences (5S, 16S, 23S, and tRNAs) in each genome to account for variable genome sizes and variable rRNA operon number.
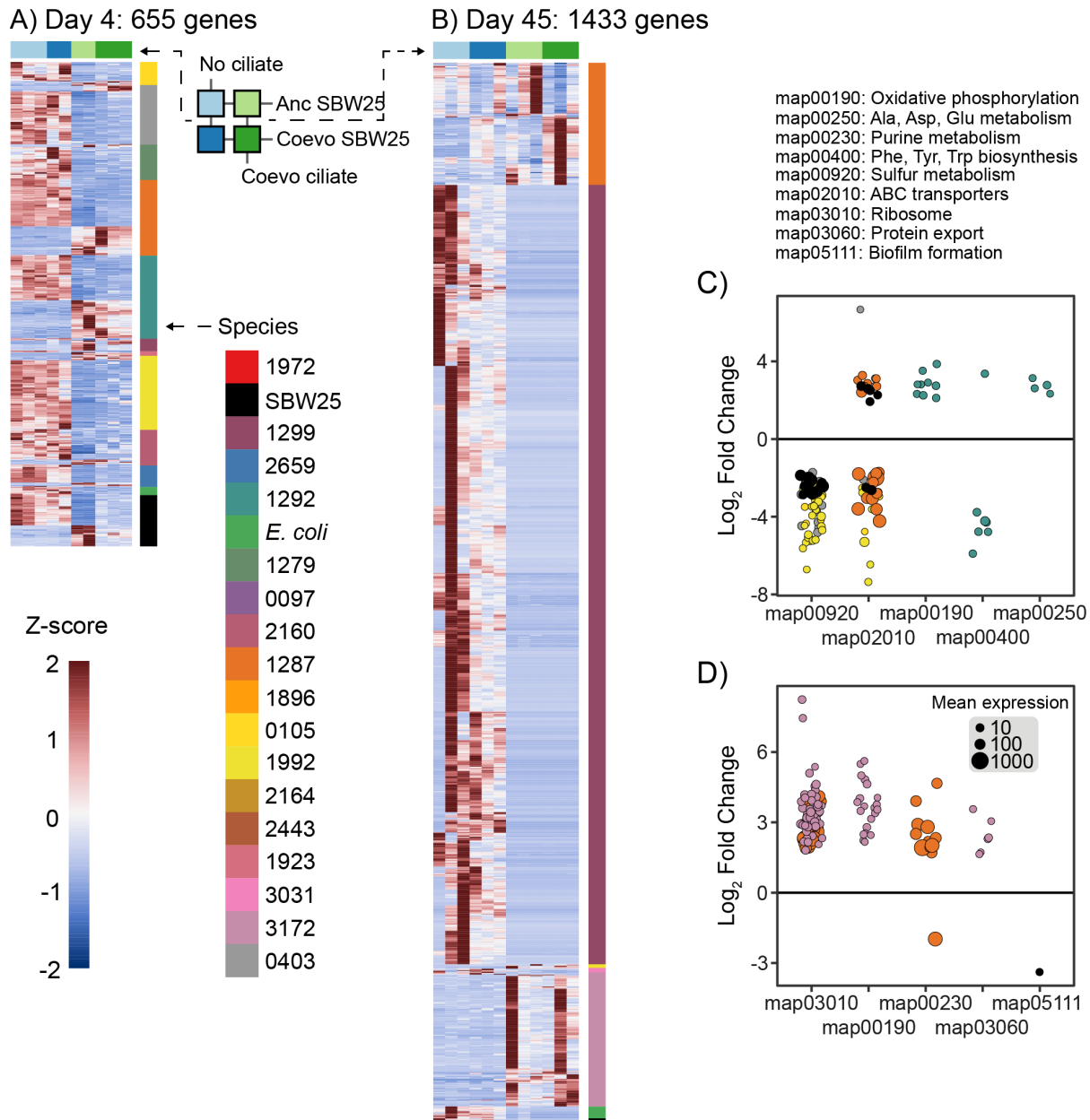
Supplementary Figure S5. Overall community functional response to the coevolved ciliate

Heapmaps for genes identified as differentially expressed due to the presence of the coevolved ciliate predator controlling for the effect of *Pseudomonas* SBW25 coevolution on **A)** day 4 and **B)** day 45. Each row is a gene colored by species identity and each column is a sample, and columns are colored according to treatment (see Figure 1 in the main text). Heatmap color shows the Z-score (i.e. the sample's relationship to the mean) for each gene across treatment categories. Functional enrichment of KEGG pathways (maps) in differentially expressed genes on **C)** day 4 and **D)** day 45. Points represent individual genes from a pathway (horizontal axis) and size is proportional to the normalized mean expression across all treatment conditions. Vertical axis depicts the log fold-change from predator-free to predation treatments (positive values show upregulation in the presence of ciliates). Colors are species. KEGG pathway abbreviations are listed above subplot C.

A) Day 4: 37 genes   B) Day 45: 228 genes

No ciliate

Anc SBW25

Coevo SBW25

Coevo ciliate

map00040: Pentose interconversions
map00364: Fluorobenzoate degradation
map00500: Starch and sucrose metabolism
map01503: CAMP resistance
map03010: Ribosome
map03070: Bacterial secretion system

Species

1972
SBW25
1299
2659
1292
E. coli
1279
0097
2160
1287
1896
0105
1992
2164
2443
1923
3031
3172
0403

Z-score

2
1
0
-1
-2

C)

D)

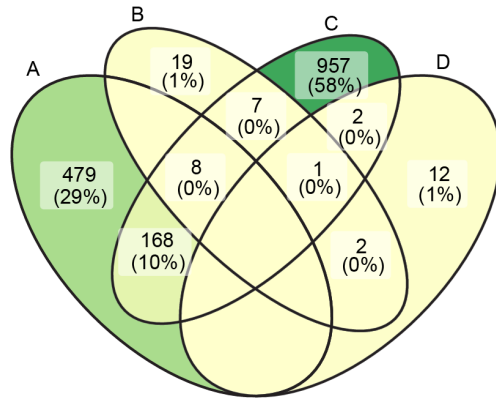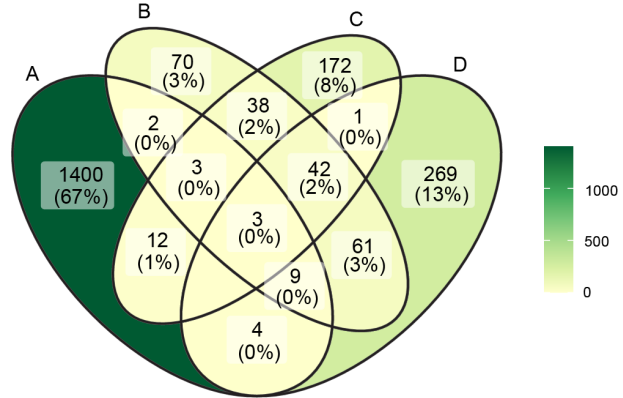**Supplementary Figure S6.** Overall community functional response to SBW25 coevolution

Heapmaps for genes identified as differentially expressed due to the presence of the coevolved *Pseudomonas* SBW25 controlling for the effect of predation. Plot organization is the same as in Figure S5.

Day 4

Day 45

A) Overall coevolved ciliate effect
B) Overall coevolved SBW25 effect
C) Specific coevolved SBW25 + coevolved ciliate effect
D) Specific coevolved SBW25 + ciliate-free effect

Supplementary Figure S7. Comparison of genes responding to different experimental treatments

Venn diagrams identifying the number of differentially regulated genes shared between treatment effects on day 4 and day 45 as identified by DESeq2's generalized linear models. Numbers in parentheses are percentages of the total number of genes identified across all treatments in each set, numbers above percentages are the total number genes unique or common to the intersection of the four sets, set colors are proportional to the total number of genes in each set intersection.
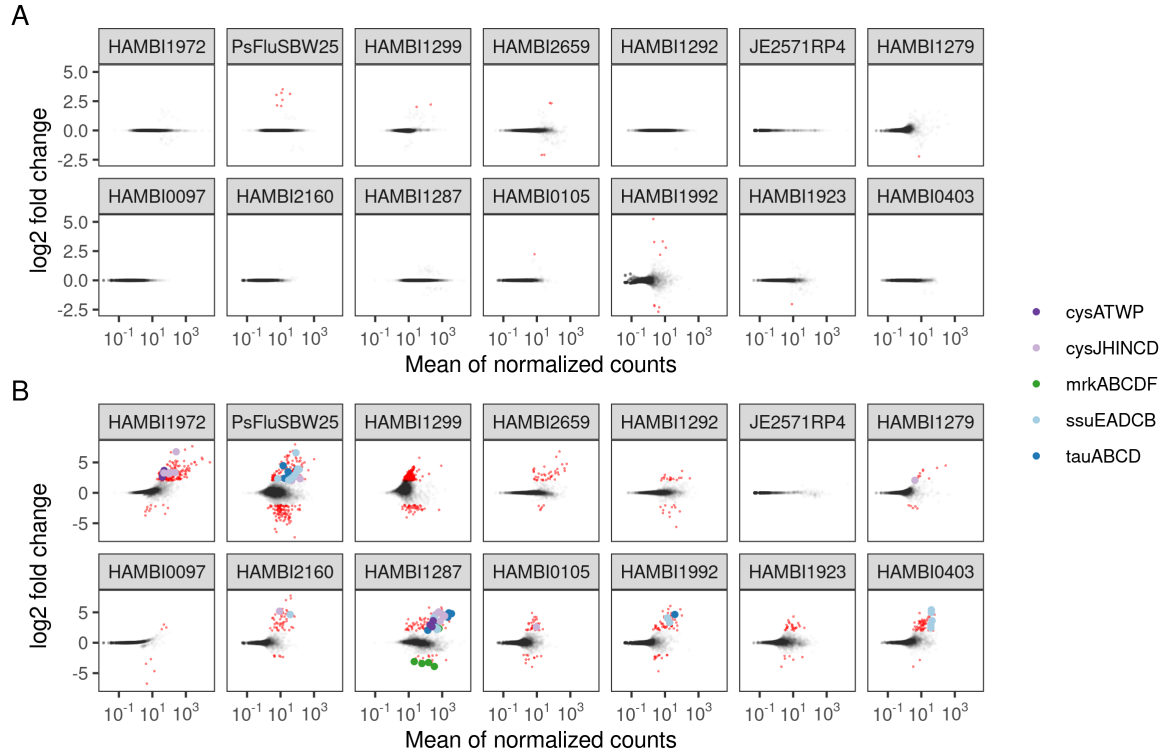
Supplementary Figure S8. Genes differentially expressed in response to *Pseudomonas* SBW25 coevolution

Plots of the estimated $\log_2$ fold change over average normalized expression strength for the coevolved *Pseudomonas* SBW25 versus ancestor *Pseudomonas* SBW25 comparison on day 4. Plot A compares SBW25 coevolution in predator-free treatments, and plot B compares SBW25 coevolution in the presence of the coevolved ciliate. Points highlighted in red are FDR adjusted p values < 0.1 with an absolute $\log_2$ fold change > 2. Effect sizes (i.e., fold changes) are moderated using the adaptive t prior shrinkage estimator from the apeglm package [16]. Larger, colored points indicate gene families of interest frmentioned in the main text. *cysATWP*: assimilatory sulfate transport system, *cysJHINCD*: assimilatory sulfate reduction to hydrogen sulfide, *mrkABCDF*: Type 3 fimbrial biosynthesis operon, *ssuEADCB*: aliphatic sulfonate utilization operon, *tauABCD*: taurine and sulfonate utilization operon.

Supplementary Tables:

Supplementary Table S1. Bacterial species used in the experiment

HAMBI collection strain ID	Species	NCBI RefSeq Assembly ID
0097	*Acinetobacter johnsonii*	GCF_003350215.1
0105	*Agrobacterium tumefaciens*	GCF_003350285.1
0216	*Azorhizobium caulinodans*	GCF_000010525.1
0262	*Brevundimonas bullata*	GCF_003350205.1
0403	*Comamonas testosterone*	GCF_000241525.1
1279	*Hafnia alvei*	GCF_000735375.1
1287	*Citrobacter koseri*	GCF_003350185.1
1292	*Morganella morganii*	GCF_001598895.1
1299	*Kluyvera intermedia*	GCF_001598315.1
1842	*Sphingobium yanoikuyae*	GCF_000315525.1
1874	*Sphingobacterium multivorum*	GCF_900457465.1
1896	*Sphingobacterium spiritivorum*	GCF_000143765.1
1923	*Myroides odoratus*	GCF_000243275.1
1966	*Chitinophaga filiformis*	GCF_900102545.1
1972	*Aeromonas caviae*	GCF_003350165.1
1988	*Chitinophaga sancti*	GCF_900119105.1
1992	*Phyllobacterium myrsinacearum*	GCF_003350115.1
2159	*Paraburkholderia caryophylli*	GCF_003350265.1
2160	*Bordetella avium*	GCF_003350095.1
2164	*Cupriavidus necator*	GCF_001592245.1
2443	*Paracoccus denitrificans*	GCF_900100045.1
2467	*Thermomonas haemolytica*	GCF_003350065.1
2494	*Paraburkholderia kururiensis*	GCF_003350035.1
2659	*Stenotrophomonas maltophilia*	GCF_000742995.1
2792	*Moraxella canis*	GCF_003350015.1
3031	*Niabella yanshanensis*	GCF_003349965.1
3172	*Azospirillum brasilense*	GCF_003349955.1
3237	*Microvirga lotononidis*	GCF_000262405.1
JE2571RP4	*Escherichia coli* str. K12 JE2571 (RP4)	GCF_000005845.1
SBW25	*Pseudomonas fluorescens* str. SBW25	GCF_000009225.1

HAMBI = University of Helsinki Microbial Domain Biological Resource Centre

Supplementary Table S2. Significant gene parallelism in evolved *Pseudomonas* populations

Locus	Gene	Product	Length	Num. pops	Obs. muts.	Exp. muts.	Obs. Mult.	$-\log_{10} p$
PFLU_4201		hypothetical protein	1578	2	3	0.50	1.87	7.80
PFLU_0691	*atoE*	Putative short-chain fatty acid transporter	1419	3	3	0.43	2.07	8.11
PFLU_0088	*algB*	Alginate biosynthesis transcriptional regulatory protein AlgB	1347	3	3	0.25	2.19	8.26
PFLU_4422	*flhB*	Flagellar biosynthetic protein FlhB	1137	3	3	0.15	2.59	8.75
PFLU_3304		hypothetical protein	948	1	3	0.13	3.10	9.28
PFLU_0821	*dppC*	Dipeptide transport system permease protein DppC	912	2	3	0.09	3.23	9.40
PFLU_1143	*glpF*	Glycerol uptake facilitator protein	816	2	3	7e-3	3.61	9.73
PFLU_1976		hypothetical protein	1191	2	4	1e-3	3.29	12.26
PFLU_0119	*lgrD*	Linear gramicidin synthase subunit D	3372	3	6	4e-5	1.75	14.08

Significant genes are listed by their locus (order in linear genome), name, putative functional product, gene length, the number of independent populations where the gene was mutated, the observed and expected number of mutations across the three populations, the corresponding multiplicity score, and the p value describing the probability of observing an equal or larger of mutations under the null model.

Supplementary Table S3. Covariates associated with Shannon index of species diversity

Parameter	Estimates	Standard Errors	p values
(Intercept)	1.39	0.02	0.00
Day 41	0.27	0.04	0.00
Day 45	0.00	0.04	0.98
SBW25 coevolution	0.02	0.03	0.49
Predation	-0.24	0.03	0.00
SBW25 coevolution:predation	0.05	0.04	0.23
σ_u^2	0.011		
$H0_1 : \sigma_u^2 = 0$	176.253		0.00
$H0_2 : \beta_1...\beta_p = 0$	4255.658		0.00
$R_{wls}^2 = 0.65,\ loglikelihood = 25.09,\ AIC = -36.18$			

Summary of results from the *betta* testing procedure. σ_u^2 is the estimate of the heterogeneity of variance. We show results for test of two different null hypotheses. $H0_1$ posits that the variation in the true Shannon diversity across populations is wholly attributable to the covariates with no unexplained random variation. $H0_2$ posits that none of the covariates explains the variation in richness across populations.

Supplementary Table S4. Permutational analysis of variance of bacterial species composition

Parameter	Df	Sum of Sqs	R^2|Mean Sq.	F	Pr(>F)
Day	2	0.60	0.33	26.10	0.0010
SBW25 coevolution	1	0.01	0.01	0.84	0.4060
Predation	1	0.84	0.47	73.18	0.0010
SBW25 coevolution:predation	1	0.01	0.01	1.03	0.3480
Residual	30	0.34	0.19		
Total	35	1.80	1.00		
Groups	11	0.06	0.01	1.59	0.1660
Residuals	24	0.08	0.00		

Permutation test for adonis (PERMANOVA) under reduced model. Model terms are added sequentially (first to last). Using 999 free permutations. Function call: `adonis2(formula = as.data.frame(comp) ~ days + coevolution * predation, data = metadf, method = "bray")` The lower table shows the results of betadisper (PERMDISP2) for testing multivariate homogeneity and an ANOVA of the distances to group centroids. The null hypothesis is that homogeneity of variances are the same in within the treatment groups. Function call: `betadisper(d = vegdist(comp, method = "bray"), group = paste(metadf[, 7], metadf[, 2], sep = "-"))`

Supplementary Table S5. Permutational analysis of community gene expression

A) Day 4

Parameter	Df	Sum of Sqs	R^2|Mean Sq.	F	Pr(>F)
SBW25 coevolution	1	0.42	0.18	4.12	0.0130
Predation	1	0.89	0.38	8.65	0.0010
SBW25 coevolution:predation	1	0.39	0.17	3.78	0.0180
Residual	6	0.61	0.27		
Total	9	2.31	1.00		
Groups	3	0.16	0.05	1.12	0.4140
Residuals	6	0.30	0.05		

B) Day 45

Parameter	Df	Sum of Sqs	R^2|Mean Sq.	F	Pr(>F)
SBW25 coevolution	1	0.56	0.17	2.39	0.0030
Predation	1	0.60	0.18	2.55	0.0020
SBW25 coevolution:predation	1	0.32	0.09	1.36	0.1260
Residual	8	1.88	0.56		
Total	11	3.35	1.00		
Groups	3	0.10	0.03	0.64	0.6128
Residuals	8	0.43	0.05		

Permutation tests for adonis (PERMANOVA) under reduced model. Model terms are added sequentially (first to last). Using 999 free permutations. Function call: `adonis2(formula = distatis_res_04$res4Splus$F[, 1: ...] ~ coevolution * predation, data = metadf, method = "euclidean")` The lower table shows the results of betadisper (PERMDISP2) for testing multivariate homogeneity and an ANOVA of the distances to group centroids. The null hypothesis is that homogeneity of variances are the same in within the treatment groups. Function call: `betadisper(d = vegdist(distatis_res_04$res4Splus$F[, 1: ... ], method = "euclidean"), group = ...`

References

1. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. en. *Nature* **551,** 45–50. ISSN: 0028-0836, 1476-4687 (Nov. 2017).

2. Tenaillon, O. *et al.* Tempo and mode of genome evolution in a 50,000-generation experiment. en. *Nature* **536,** 165–170. ISSN: 0028-0836, 1476-4687 (Aug. 2016).

3. Shoemaker, W. R., Polezhaeva, E., Givens, K. B. & Lennon, J. T. Molecular Evolutionary Dynamics of Energy Limited Microorganisms. en. *Mol. Biol. Evol.* ISSN: 0737-4038, 1537-1719 (July 2021).

4. Roberts, S. *et al.* Gaussian processes for time-series modelling. en. *Philos. Trans. A Math. Phys. Eng. Sci.* **371,** 20110550. ISSN: 1364-503X (Feb. 2013).

5. Timonen, J., Mannerström, H., Vehtari, A. & Lähdesmäki, H. lgpr: An interpretable nonparametric method for inferring covariate effects from longitudinal Data. *Bioinformatics* **37,** 1860–1867 (2021).

6. Gelman, A. & Rubin, D. B. Inference from Iterative Simulation Using Multiple Sequences. en. *Stat. Sci.* **7,** 457–472. ISSN: 0883-4237, 2168-8745 (Nov. 1992).

7. Martin, B. D., Witten, D. & Willis, A. D. Modeling microbial abundances and dysbiosis with beta-binomial regression. en. *Ann. Appl. Stat.* **14,** 94–115. ISSN: 1932-6157, 1941-7330 (Mar. 2020).

8. Eisen, J. A. *et al.* Macronuclear genome sequence of the ciliate Tetrahymena thermophila, a model eukaryote. en. *PLoS Biol.* **4,** e286. ISSN: 1544-9173, 1545-7885 (Sept. 2006).

9. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. en. *Bioinformatics* **30,** 923–930. ISSN: 1367-4803, 1367-4811 (Apr. 2014).

10. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. en. *BMC Bioinformatics* **11,** 119 (Mar. 2010).

11. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30,** 2068–2069. ISSN: 1367-4803 (July 2014).

12. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. en. *Mol. Biol. Evol.* **38,** 5825–5829. ISSN: 0737-4038, 1537-1719 (Dec. 2021).

13. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. en. *Nucleic Acids Res.* **47,** D309–D314. ISSN: 0305-1048, 1362-4962 (Jan. 2019).

14. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550. ISSN: 1465-6906, 1474-760X (2014).

15. Zhang, Y., Thompson, K. N., Huttenhower, C. & Franzosa, E. A. Statistical approaches for differential expression analysis in metatranscriptomics. en. *Bioinformatics* **37,** i34–i41. ISSN: 1367-4803, 1367-4811 (July 2021).

16. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. en. *Bioinformatics* **35,** 2084–2092. ISSN: 1367-4803, 1367-4811 (June 2019).