Environmental shaping of codon usage and functional adaptation across microbial communities

Maša Roller, Vedran Lucić, István Nagy, Tina Perica & Kristian Vlahoviček Supplementary information

TABLE OF CONTENTS

Supplementary Figures4

 Figure S6: Construction of artificial metagenomes. Artificial metagenomes were constructed by randomly selecting equivalent number of orthologous sequences for each COG group in the real metagenome from the NCBI bacterial genomes dataset. CU distance plots were prepared according to the Methods in the main manuscript. Distance ratios of genes to two axes (the MELP value) were calculated and the results are presented for all metagenomes in Figure S7.9

SUPPLEMENTARY FIGURES



Figure S1: The distance (MILC, outlined in Methods) and distance ratios (MELP, inset for each figure) of each gene's CU frequency to overall CU frequencies of six microbial metagenomes. The separation between two compared genomes is clearly visible in the MELP distributions (inset).



Figure S2: The frequency of all synonymous codons normalised per amino acid in the whole Sargasso Sea sample (N= 155,865,864) compared to A) Sargasso removed for all sequences belonging to the Alphaproteobacteria class (N=10,0594,880), B) Sargasso Alphaproteobacteria only (N=55,270,984), C) Sargasso sample where equal sequence samples per represented phyla were taken (n=32 sequences from 32 phyla, N=1024); D) Waseca soil (N= 16,224,742), E) US EBPR sludge (N= 4,639,388), F) Santa Cruz whale fall carcass bone (N= 4,730,048), G) obese mouse gut (N= 1,510,482), H) lean mouse gut (N= 2,444,470) and I) human gut (N=11,410,113). The intraclass correlation coefficient (ICC, see Methods) for each comparison is also shown.



Figure S3: Metagenomes show codon usage distribution similar to single genomes. The distance of each gene's codon usage (CU) frequency form the overall CU of the (meta)genome and ribosomal reference set, displayed as a Karlin B-plot for A) a single microbial genome (*Escherichia coli*, N=4,358) and B) a metagenome (whale carcass, N=33,422). The metagenome shows the same characteristic distribution as the genome with ribosomal genes closer to the CU of the ribosomal set than the overall CU of the whole (meta)genome. MELP – the measure of expression is derived by dividing the gene's distance to the whole genome with that of the distance to the ribosomal protein CU.



Figure S4: The distribution of CU distances (MILC) from the ribosomal reference set of all genes in the Sargasso Sea metagenome (meta, N=688,539), all ribosomal protein genes (ribo, N=1,3049), and 6 most abundant species: *Candidatus pelagibacter sp.* HTCC7211 (cp, N=214), uncultured marine gamma proteobacterium EBAC20E09 (eb, N=122), uncultured marine microorganism HF4000_005D21 80 (hf, N=80), uncultured marine alpha proteobacterium HOT2C01 (ho, N=70), *Prochlorococcus marinus* (pm, N=347) and *Psychroflexus torquis* (pt, N=181). The red line marks median distance of all ribosomal genes from the ribosomal reference set.



Figure S5. The distance of each gene's codon usage (CU) frequency form the overall CU of the metagenome and ribosomal reference set, displayed as a Karlin B-plot for non-randomised (left panel) and (right randomised panel) metagenomes for A) the Sargasso Sea (N=688,539), B) Santa Cruz whale carcass bone (N=33,422), C) US EBPR sludge (N=20,175) and D) acid mine biofilm (N=79,257) samples. When the amino acid content of a metagenome is kept constant but codons randomly chosen, the metagenome loses its characteristic shape.



Figure S6: Construction of artificial metagenomes. Artificial metagenomes were constructed by randomly selecting equivalent number of orthologous sequences for each COG group in the real metagenome from the NCBI bacterial genomes dataset. CU distance plots were prepared according to the Methods in the main manuscript. Distance ratios of genes to two axes (the MELP value) were calculated and the results are presented for all metagenomes in Figure S7.



Figure S7: Construction of artificial metagenomes. MELP (See methods in the main manuscript) values were calculated for each real and artificial metagenome, based on the distances in Figure S6. Distributions generally exhibit more variability in real than in artificial metagenomes. Differences were quantified in Figure S8.

MELP

MELP



Artificial vs. Artificial Metagenome Q–Q Plot



Figure S8: Construction of the artificial metagenomes from the NCBI bacterial genome datasets. Real metagenomes were decomposed into respective COG functional categories and the artificial metagenomes were generated by sampling the equivalent number of orthologous sequences for each COG group of the real metagenome from the bacterial genomes section of the NCBI, regardless of the phyletic composition. MELP distributions were calculated for each metagenome and the distributions (shown in Figure S7) were compared on quantile-quantile plots and evaluated statistically. On overall, the real metagenomes exhibit statistically significant difference in CU distribution variability, while the artificially generated metagenomes tend to adopt more similar and uniform codon usage distribution.

Transition Collutatingood Collutating		COG Supercategory	Santa Cruz Whale Fall Microbial Mat	Antarctic Whale Fall Bone	US EBPR Sludge	OZ EBPR Sludge	Waseca Farm Soil
RNA procession Information Information Information Information RNA procession An accession An acces	ճ թւ ս	Translation, ribosomal structure and biogenesis [J]					6
Ponomation Collidial processes Intercription (G) Replication, scontination and cipation Conventination and cipation Conventination and cipation Chormation structure and dynamics (E) Conventination and cipation Conventination and cipation Chormation structure and dynamics (E) Conventination and cipation Conventination Conventination Market structure and dynamics (E) Coll vector control, cell division, circonanisms (T) Coll vector <	oit ns nis	RNA processing and modification [A]	_]		J	-
Replication, recombination and repair [s] Replication Replicat	ទេខះ ទសិទ ខណ	Transcription [K]	-				٩.
Columatin structure and dynamics [8] Columatin structure and dynamics [8] Columatin structure and dynamics [8] Call cycle control, call division, termonol (10) Nuclear structure) Signal transduction (10) Signal transduction (10) Call cycle control, call division, termonol (10) Signal transduction (10) Signal transduction (10) Signal transduction (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll motily (10) Coll mot	for sors roc	Replication, recombination and repair [L]	ſ			ď	7
Cell cycle control, cell division, chromosone partitioning () Cell cycle control, cell division, chromosone partitioning () Cell cycle control, cell division, chromosone partitioning () Nuclease mechanisms () Cell methanisms () Cell cycle control, cell division, chromosone partitioning () Russian Control, cell division, chromosone partitioning () Cell methanisms ()	d Is U	Chromatin structure and dynamics [B]		-	-		
Description Mutchast structure (Motebabilism) Mutch		Cell cycle control, cell division, chromosome partitioning [D]	•	•	٩,		
Definition Cellulatification Cellulatification Signal transformed and signalisms (1) Cell wallimentarenervelopo messinsms (1) Cell wallimentarenervelopo metabolism (1) Carbo dramsport and metabolism (1) Undrarenervelopo metabolism (1) Undrarenervelopo metabolism (1) Centrarenervelopo metabolism (1) Undrarenervelopo metabolism (1) Undrarenervelopo metabolism (1) Centrarenervelopo metabolism (1) Undrarenervelopo metabolism (1) Centrarenervelopo metabolism (1) C		Nuclear structure [Y]					
Signal transduction mechanisme [1] Collume transport Cal valitmentation (c) Cal valitmentation (c) Pastransistion (c) Cal valitmentation (c) Cal valit Cal valitmentation (c) Pastransistion (c) Cal valitmentation (c) Cal valit Cal valit Mucle colde Can valit Valit Cal valit Valit Cal valit Valit Cal valit Cal valit Cal valit Pastransist Cal valit Pastransist Cal valit Pastransist Cal valit Cal valit Cal valit Pastransist Cal valit Cal valit Cal valit Pastransis Cal valit	bi ss	Defense mechanisms [V]	₽	₽	-0	₽	┛
Cell wallmembranelenvelope biogenesis (M) Cell modify (N) Cell modify (N) Cell modify (N) Cell modify (N) Cell modify (N) Cell modify (N) Contractelluar trituctures (M) Intractelluar trituctures (M) Intractelluar trituctures (M) Destranslational modification, and vesicular transport and metholosism (H) Contractelluar trituctures (M) Destranslational modification, and vesicular transport Minio acid transport and metabolism (H) Carbohydrate transport and metabolism (H) Unclearite tra	900 Uliu	Signal transduction mechanisms [T]		₽	ſ	J	-0
Coll motify (N) Coll motify (N) Coll motify (N) Coll motify (N) Coll motify (N) Coll motify (N) Cyckester (Z) Cyckester (Z) Cyckester (Z) Cyckester (Z) Cyckester (Z) Cyckester (Z) Cyckester (Z) Cyckester (Z) Coll motify (N) Cyckester (Z) Cyckester (Z) Cyckester (Z) Intracellular trafficking, secretion, and vestellular transport (N) Cyckester (Z) Cyckester (Z) Posttranslation and conversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Contraversion (C) Nucleotide transport and metabolism (F) Contraversion (C) Contraversion (C) Contraversion (C) Nucleotide transport and metabolism (F) Contraversion (C) Contraversion (C) Contraversion (C) Nucleotide transport and metabolism (F) Contraversion (C) Contraversion (C) Contraversion (C) Nucleotide transport and metabolism (F) Contraversion (C) Contraversion (C) Contraversion (C) </td <td>gna pro</td> <td>Cell wall/membrane/envelope biogenesis [M]</td> <td>ſ</td> <td></td> <td>92</td> <td>•</td> <td>ď</td>	gna pro	Cell wall/membrane/envelope biogenesis [M]	ſ		9 2	•	ď
Cytosketeon (2) Cytosketeon (2) <thc< td=""><td>lar sig</td><td>Cell motility [N]</td><td>~</td><td>-</td><td>₽</td><td>=</td><td>٩.</td></thc<>	lar sig	Cell motility [N]	~	-	₽	=	٩.
Extracellular structures (M) Extracellular structures (M) Extracellular structures (M) Intracellular trafficking, secretion, and vesicular transport (U) Destitansition and vesicular transport (U) Intracellular trafficking, secretion, and vesicular transport (U) Destitansition and vesicular transport (U) Intracellular trafficking, secretion, and vesicular transport and metabolism (E) Destitansition and vesicular transport and metabolism (E) Carbohydrate transport and metabolism (E) Carbohydrate transport and metabolism (E) Destination and vesicular transport and metabolism (E) Amino acid transport and metabolism (E) Carbohydrate transport and metabolism (E) Destination acid transport and metabolism (E) Amino acid transport and metabolism (E) Carbohydrate transport and metabolism (E) Destination acid transport and metabolism (E) Amino acid transport and metabolism (E) Destination acid transport and metabolism (E) Destination acid transport and metabolism (E) Secondary metabolism (E) Destination (E) Destination acid transport and metabolism (E) Destination (E) Secondary metabolism (E) Destination (E) Destination (E) Destination (E) Destination (E) Amino acid transport and metabolism (E) Destination (E) Destination (E) Destination (E) Destination (E) Secondary metr	pu njja	Cytoskeleton [Z]				_0	
Intracellular trafficking, secretion, and vesicular transport (U) Posttranslational modification, protein turnwer, chapterones (O) Posttranslational modification, protein turnwer, chapterones (O) Posttranslational modification, protein turnwer, chapterones (O) Posttranslational modification, protein turnwer, chapterones (O) Posttranslational metabolism (G) Restranslational modification, protein turnwer, chapterones (O) Posttranslational metabolism (G) Anniho add transport and metabolism (E) Posttranslational metabolism (E) Posttranslational metabolism (E) Anniho add transport and metabolism (E) Posttranslational metabolism (E) Posttranslational metabolism (E) Posttranslational metabolism (E) Anniho add transport and metabolism (E) Posttransport and metabolism (E) Posttransport and metabolism (E) Posttransport and metabolism (E) Posttransport and metabolism (E) Anniho add transport and metabolism (E) Posttransport and metabolism (9J B	Extracellular structures [W]	-0	-	_0	J	J
Posttranslational modification, protein turnover, chaperones (0) Posttranslational modification, protein turnover, chaperones (0) Energy production and conversion (C) Energy production and conversion (C) Energy production and conversion (C) Carbohydrate transport and metabolism (E) Amino acid transport and metabolism (E) Poeta- Amino acid transport and metabolism (E) Poeta- Poeta- Amino acid transport and metabolism (E) Poeta- Poeta- Indication (transport and metabolism (E) Poeta- Poeta- Indication (transport and metabolism (E) Poeta- Poeta- Indication (transport and metabolism (E) Poeta- Poeta- Indication prediction only (R) Poeta- Poeta- Poeta- Indication prediction only (R) Poeta- Poeta- Poeta- Poeta- Indication prediction only (R) Poeta- Poeta- Poeta- Poeta- Poeta- Indication prediction nolice (ransport and metabolism (E) Poeta- Poeta- <th></th> <th>Intracellular trafficking, secretion, and vesicular transport [U]</th> <th>٩.</th> <th>٩.</th> <th>-0</th> <th>٩.</th> <th>٩,</th>		Intracellular trafficking, secretion, and vesicular transport [U]	٩.	٩.	-0	٩.	٩,
Energy production and conversion (C) Energy production and conversion (C) Carbohydrate transport and metabolism (G) Amino acid transport and metabolism (G) Amino acid transport and metabolism (G) Amino acid transport and metabolism (G) Amino acid transport and metabolism (G) Amino acid transport and metabolism (G) Nucleotide transport and metabolism (H) Implementation Nucleotide transport and metabolism (H) Implementation Construme transport and metabolism (H) Implementation Secondary metabolism (H) Implementation Implementation Implementation <		Posttranslational modification, protein turnover, chaperones [O]	•	L	٩.		٩,
Carbohydrate transport and metabolism (6) Carbohydrate transport and metabolism (7) Carbohydrate transport and metabolism (7) Amino acid transport and metabolism (7) Nucleotide transport and metabolism (7) Image: Construent and metabolism (7) Nucleotide transport and metabolism (7) Nucleotide transport and metabolism (7) Image: Construent and metabolism (7) Image: Construent and metabolism (7) Nucleotide transport and metabolism (7) Image: Construent and metabolism (7) Imag		Energy production and conversion [C]	٩,	[٩,	Ļ	٩.
Amino acid transport and metabolism [E] Amino acid transport and metabolism [E] Image: Comparison of transport and metabolism [E] Nucleotide transport and metabolism [H] Lipid transport and metabolism [H] Image: Comparison of transport and metabolism [H] Nucleotide transport and metabolism [H] Image: Comparison of transport and metabolism [H] Image: Comparison of transport and metabolism [H] Nucleotide transport and metabolism [H] Image: Comparison of transport and metabolism [H] Image: Comparison of transport and metabolism [H] Secondary metabolites biosynthesis, transport and metabolism [H] Image: Comparison of transport and metabolism [H] Secondary metabolites biosynthesis, transport and metabolism [H] Image: Comparison of transport and metabolism [H] Image		Carbohydrate transport and metabolism [G]	┛	S	₽	•••	٦
Nucleotide transport and metabolism [H] Nucleotide transport and metabolism [H] Image: Second symplex is a second symplex is	ws	Amino acid transport and metabolism [E]	٦	₽		٦	٦
Coenzyme transport and metabolism [H] Lipid transport and metabolism [H] Lipid transport and metabolism [H] Inorganic ion transport and metabolism [H] Inorganic ion transport and metabolism [H] Inorganic ion transport and metabolism [H] Secondary metabolism [H] Inorganic ion transport and metabolism [H] Inorganic ion transport and metabolism [H] Secondary metabolism [H] Inorganic ion transport and metabolism [H] Inorganic ion transport and metabolism [H] Secondary metabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and metabolism [H] Secondary metabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Secondary metabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Secondary metabolites biosynthesis, transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H] Inorganic ion transport and catabolism [H]	ilo	Nucleotide transport and metabolism [F]	•	•	•	•	-0
Lipid transport and metabolism [1] Increased transport and metabolism [2] Increased transport and metabolism [2] Increased transport and metabolism [2] Increased transport and metabolism [2] Increased transport and metabolism [2] Secondary metabolism [2] Secondary metabolism [2] Increased transport and catabolism [2] Increased transport and catabolism [2] Secondary metabolism [2] Function unknown [5] Increased transport and catabolism [2] Increased transport and	təb	Coenzyme transport and metabolism [H]		₽	P	•••	-0
Inorganic ion transport and metabolism [P] Image: Comparise internation transport and metabolism [Q] Image: Comparise internation transport and catabolism [Q] Image: Comparis internation transport and catabolism [Q] </th <th>эM</th> <th>Lipid transport and metabolism [I]</th> <th></th> <th>•</th> <th>•••</th> <th>-0</th> <th>70</th>	эM	Lipid transport and metabolism [I]		•	•••	-0	70
Secondary metabolites biosynthesis, transport and catabolism [Q] Image: Condery metabolites biosynthesis, transport and catabolism [Q] Secondary metabolites biosynthesis, transport and catabolism [Q] Image: Condery metabolites biosynthesis, transport and catabolism [Q] Recondary metabolites biosynthesis, transport and catabolism [Q] Image: Condery metabolism [Q] General function unknown [S] Image: Condery metabolism [Q] Uncharacterized [X] Image: Condery metabolism [Q] Metabolism [Metabolism [Inorganic ion transport and metabolism [P]		₽		٩,	•
Control Ceneral function only [R] Ceneral function unknown [S] Poonerse Function unknown [S] Image: Control function function only [R] Image: Control function unknown [S] Image: Control function function only [R] Image: Control function		Secondary metabolites biosynthesis, transport and catabolism [Q]	•	∎₽	ſ	=	٦
Meter Top 3% Meter Bentom 3% (FDR corrected by values > 0.05) Meter Top 3% Meter Bentom 3% (FDR corrected by values > 0.05) Meter Top 3% (FDR corrected by values > 0.05) Meter Top 3% Meter Bentom 3% (FDR corrected by values > 0.05) Meter Top 3% (FDR corrected by values > 0.05) Meter Top 3% (FDR corrected by values > 0.05)	sq c- J	General function prediction only [R]					•
• 5 2 • 10% • 10% • 10% • 10% • 5% • 10% • 10% • 5% • 10% • 10% • 5% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10% • 10%	oor ara rize	Function unknown [S]	P	P	P		-
MELP Top 3% MELP Bottom 3% Expressivity enrichment -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5% 10% -10% -5% 0% 5	bq te bq	Uncharacterized [X]	D	-0	I.		
MELP Top 3% MELP Bottom 3% (FDR corrected p-values > 0.05) (FDR corrected p-values > 0.05)		MELP Top 3% MELP Bottom 3% Expressivity enrichment					
		MELP Top 3% MELP Bottom 3% (FDR corrected p-values > 0.05)	-10% -5% 0% 5% 10%	-10% -5% 0% 5% 10%	-10% -5% 0% 5% 10%	-10% -5% 0% 5% 10%	-10% 0% 10% 20% 30%

Figure S9: Comparison of gene enrichment for each functional COG supercategory. The top and low 3% of genes by gene expressivity values are shown for 4 metagenomes: Santa Cruz whale fall microbial mat (N=40,916), Antarctica whale fall bone (N=30,503), US (N=20,175) and OZ EBPR sludge (N=29,754).



Figure S10: Lean vs. obese gut microbiomes. Comparison of gene enrichment for each functional COG supercategory. The top and low 3% of genes by gene expressivity values are shown for all 3 gut metagenomes: lean human (N=47,765), lean (N=4,955) and obese mouse (N=4,058). Categories highlighted in yellow show the loss of optimisation for two metabolic functions in obese mouse fauna.

A	COG Supercategory	Sargasso Sea	Santa Cruz Whale Fall Bone	Acid Mine Biofilm
uo pu	Translation, ribosomal structure and biogenesis [J]			
natio ge a sssi	RNA processing and modification [A]		ц П	
forn	Replication, recombination and repair [L]			
pi st	Chromatin structure and dynamics [B]			¢.
	Cell cycle control, cell division, chromosome partitioning [D]	0	0	
ŝ	Nuclear structure [Y]			
ces	Defense mechanisms [V] Signal transduction mechanisms [T]			
pro	Cell wall/membrane/envelope biogenesis [M]			
sig	Cell motility [N]		ġ	8
ellu and	Cytoskeleton [Z]	I	0	<u> </u>
0	Extracellular structures [W]		U n	
	Posttranslational modification, protein turnover, chaperones [0]		ч Полони (страни)	
	Energy production and conversion [C]			
_	Carbohydrate transport and metabolism [G]			d
ism	Amino acid transport and metabolism [E]		_	
lodi	Nucleotide transport and metabolism [F]			
leta	Linid transport and metabolism [H]		n n	
2	Inorganic ion transport and metabolism [P]			
	Secondary metabolites biosynthesis, transport and catabolism [Q]			P
ac-ac-	General function prediction only [R]			
200 har eriz	Function unknown [S]			
ш 0 Ф	Uncharacterized [X]		ų	P
	Relative representation of genes with high expressivity after normalisation with occurrence of COG category in whole sample (statistically not significant marked in gray)	-5% 0% 5% 10% fraction of total sample Depletion Enrichment	-5% 0% 5% 10% fraction of total sample Depletion Enrichment	-5% 0% 5% 10% fraction of total sample Depletion Enrichment
В	COG Supercategory	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
B	COG Supercategory Translation, ribosomal structure and biogenesis [J]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
nation ge and essing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
formation orage and ocessing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Information storage and processing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Information storage and processing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
s Information storage and processing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
cess Information storage and ling Processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [T] Signal transduction mechanisms [T]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
process Information storage and storage and processing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [V] Signal transduction mechanisms [T] Cell wall/membrane/envelope biogenesis [M]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
llar process Information storage and t signalling Processing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [Y] Signal transduction mechanisms [Y] Cell wall/membrane/envelope biogenesis [M] Cell motility [N]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Jellular process Information and signalling processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [V] Signal transduction mechanisms [V] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Cellular process Information and signalling processing Brond	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [T] Signal transduction mechanisms [T] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [II]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Cellular process and signalling processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Cell wall/membrane/envelope biogenesis [M] Cell wall/membrane/envelope biogenesis [M] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Cellular process and signalling processing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Cell cycle control, cell division, chromosome partitioning [D] Cell cycle control, cell division, chromosome partitioning [D] Cell wall/membrane/envelope biogenesis [M] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
n Cellular process Information and signalling processing and processing	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [T] Cell wall/membrane/envelope biogenesis [M] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cotoxellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C] Carbohydrate transport and metabolism [G]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
nism Cellular process in triormation and signalling processing and processing the second seco	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [T] Cell wall/membrane/envelope biogenesis [M] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Carbohydrate transport and metabolism [G] Amino acid transport and metabolism [G]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
abolism Cellular process Information and signalling processing and processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [V] Cell wall/membrane/envelope biogenesis [M] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C] Carbohydrate transport and metabolism [E] Nucleotide transport and metabolism [E] Nucleotide transport and metabolism [E]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Metabolism Cellular process Information and signalling processing and processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [V] Cell wall/membrane/envelope biogenesis [M] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C] Carbohydrate transport and metabolism [F] Coenzyme transport and metabolism [F] Coenzyme transport and metabolism [F] Lipid transport and metabolism [F]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Metabolism Cellular process Information and signalling processing and processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [V] Signal transduction mechanisms [T] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C] Amino acid transport and metabolism [E] Nucleotide transport and metabolism [F] Coenzyme transport and metabolism [I] Lipid transport and metabolism [I]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Metabolism Cellular process Information and signalling processing and processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] RNA processing and modification [A] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Nuclear structure [Y] Defense mechanisms [V] Signal transduction mechanisms [V] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C] Amino acid transport and metabolism [E] Nucleotide transport and metabolism [F] Ccenzyme transport and metabolism [F] Lipid transport and metabolism [F] Secondary metabolites biosynthesis, transport and metabolism [P]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
orty rac- zed Metabolism Cellular process and signalling processing B	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Cell cycle control, cell division, chromosome partitioning [D] Cell cycle control, cell division, chromosome partitioning [D] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cytoskeleton [Z] Extracellular structure [W] Intracellular trafficking, secreton, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C] Amino acid transport and metabolism [E] Nucleotide transport and metabolism [F] Coenzyme transport and metabolism [F] Lipid transport and metabolism [F] Secondary metabolites biosynthesis, transport and metabolism [F] Secondary metabolites biosynthesis, transport and metabolism [F]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine
Poorly Metabolism Cellular process Information charac- terized metabolism Cellular processing and terized metabolism Cellular processing Cellular	COG Supercategory Translation, ribosomal structure and biogenesis [J] RNA processing and modification [A] Transcription [K] Replication, recombination and repair [L] Chromatin structure and dynamics [B] Cell cycle control, cell division, chromosome partitioning [D] Cell cycle control, cell division, chromosome partitioning [D] Cell cycle control, cell division, chromesome partitioning [D] Cell wall/membrane/envelope biogenesis [M] Cell motility [N] Cell motility [N] Cell motility [N] Cytoskeleton [Z] Extracellular structures [W] Intracellular trafficking, secretion, and vesicular transport [U] Posttranslational modification, protein turnover, chaperones [O] Energy production and conversion [C] Carbohydrate transport and metabolism [F] Nucleotide transport and metabolism [F] Coenzyme transport and metabolism [F] Lipid transport and metabolism [P] Secondary metabolites biosynthesis, transport and metabolism [Q] General function prediction only [R] Function unknown [S]	Artificial Sargasso	Artificial Santa Cruz Whale Fall Bone	Artificial Acid Mine

Figure S11: Comparison between enrichment profiles in real (A) and artificially assembled metagenomes (B). Artificial metagenomes were assembled from a random selection of NCBI bacterial genomes by maintaining constant function distribution of genes (COG categories) for the original metagenome and with the same total number of genes: (artificial) Sargasso Sea N=688,539, (artificial) acid mine biofilm N=79,257 and (artificial) Santa Cruz whale fall bone N=33,422 . The enrichment patterns visible in real metagenomes are substantially diminished in artificially assembled metagenomes.



Figure S12: Comparison of the fraction of GC content in ribosomal genes and genes of whole metagenomes. The resulting fractions were tested with the binomial test for each metagenome. All metagenomes show significant differences in GC content between ribosomal genes and all genes of the metagenome.

SUPPLEMENTARY TABLES

Table SI: Names of metagenomes used in this project, their NCBI Project IDs and references for the original sequencing projects.

* In-house assembly of trace data.

⁺Metagenomes downloaded preassembled.

Metagenome	NCBI Project ID	Reference
[†] Global Ocean Sampling Expedition Metagenome, the Sargasso Sea version 1	13694	(1)
[†] Waseca County Farm Soil Metagenome	13699	(2)
[†] Whale Fall Metagenomes	13700	
*5-Way (CG) Acid Mine Drainage Biofilm Metagenome	13696	(3)
*Human Distal Gut Biome	16729	(4)
*Lean Mouse 1 Gut Metagenome	17391	(5)
*Obese Mouse 1 Gut Metagenome	17397	
*US EBPR Sludge Metagenome	17657	(6)
*OZ EBPR Sludge Metagenome	17659	

Table SII: Number and length of metagenomic sequences used in the study, number of ORFs assigned through homology to the STRING/COG database and using the 3-nearest neighbour consensus rule and the number of ORFs used in MILC and MELP calculations.

	Total number	Total length of	Number of ORFs
	of sequences used	sequences used (bp)	assigned
Sargasso Sea	100,1987	585,970,413	688,539
Whale fall Santa Cruz	28,151	29,783,639	33,422
bone			
Whale fall Santa Cruz	29,934	30,862,440	40,916
microbial mat			
Whale fall Antarctic	26,232	28,868,943	30,503
bone			
Waseca farm soil	139,341	144,897,582	88,696
Acid mine biofilm	154,736	206,216,325	79,257
US EBPR sludge	36,222	41,320,987	20,175
OZ EBPR sludge	63,760	81,189,385	29,754
Human gut	79,613	85,004,190	47,765
Lean mouse gut	822,800	65,615,892	4,955
Obese mouse gut	573,519	62,721,309	4,058

Genome	NCBI Reference Sequence	Note	Reference
<i>P. acnes</i> KPA171202	NC_006085.1	Type IB	(7)
<i>P. acnes</i> 6609	CP002815	Type IB	(8)
<i>P. acnes</i> 51318	-	Type IB	in preparation
P. acnes PRP60	-	Type IA	
P. acnes 226	-	Type IA	
P. acnes 434	-	Type IA	
P. acnes PRP38	AIJP00000000	Type IC	(9)
P. acnes ATCC11828	CP003084	Type II	(10)
P. acnes PRP47	-	Type II	in preparation
<i>P. acnes</i> 35934	-	Type II	
<i>P. acnes</i> 9880	-	NC*	
<i>P. acnes</i> 33810	-	NC*	
<i>P. acnes</i> 440671	-	NC*	
<i>R. palustris</i> BisA53	NC_008435.1	none	
<i>R. palustris</i> BisB18	NC_007925	none	(11)
<i>R. palustris</i> BisB5	NC_007958	none	
R. palustris HaA2	NC_007778	none	-
R. palustris CGA009	chromosome: NC_005296	none	(12)
	plasmid: NC_005297		
<i>R. palustris</i> TIE-1	NC_011004	none	none

Table SIII: Genomes of *P. acnes* and *R. palustris* strains used in this project and references for the original sequencing projects.

*not clustered

Table SIV: (Additional File): List of species, the number of genes and their total codon counts per metagenome for species present in at least two metagenomes for species with at least 2,000 codons per metagenome classified with the MEtaGenome Analyzer (MEGAN).

Table SV (Additional File): Phylogenetic clade counts of the ribosomal reference set of the Sargasso Sea

 metagenome and the whole metagenome classified with the MEtaGenome Analyzer (MEGAN).

Table SVI: The number of genes per COG supercategory in the 6 *R. palustris* strains that fall within 10% of COGs with the smallest variation of MILC median and those that fall within 10% of COGs within the largest variation. The difference of counts per COG category is tested with the binomial test with FDR correction for greater occurrence in the tight 10% than the whole set and for greater occurrence in the wide 10% than the whole set and for greater occurrence in the wide 10% than the whole set of genes regardless of COG MILC median.

		number of	number of	FDR corrected	number of	FDR corrected
	COG supercategory	genes in	genes with	p value for	genes with	p value for
		whole sample	tight MILC	tight vs. all	wide MILC	wide vs. all
g	[J] Translation, ribosomal					
e ar	structure and biogenesis	967	282	0.00	58	1.00
86 98 86	[A] RNA processing and					
sin	modification	0	0	1.00	0	1.00
on s ces	[K] Transcription	1396	79	1.00	188	0.03
atic pro	[L] Replication,					
ů –	recombination and repair	900	95	0.29	111	0.51
lo	[B] Chromatin structure					
=	and dynamics	10	0	1.00	0	1.00
	[D] Cell cycle control, cell					
	division, chromosome					
	partitioning	164	12	1.00	26	0.18
	[Y] Nuclear structure	0	0	1.00	0	1.00
8	[V] Defence mechanisms	378	10	1.00	174	0.00
llin	[T] Signal transduction					
gna	mechanisms	1655	48	1.00	124	1.00
I si	[M] Cell					
anc	wall/membrane/envelop					
es	e biogenesis	1311	57	1.00	189	0.00
ess	[N] Cell motility	586	39	1.00	49	1.00
2 2	[Z] Cytoskeleton	3	3	0.01	0	1.00
r p	[W] Extracellular					
Iula	structures	0	0	1.00	0	1.00
Cell	[U] Intracellular					
-	trafficking, secretion, and					
	vesicular transport	448	81	0.00	44	1.00
	[O] Posttranslational					
	modification, protein					
	turnover, chaperones	965	120	0.00	61	1.00
	[C] Energy production					
	and conversion	1768	163	1.00	157	1.00
	[G] Carbohydrate					
	transport and metabolism	1085	51	1.00	149	0.03
	[E] Amino acid transport					
	and metabolism	2338	104	1.00	81	1.00
E	[F] Nucleotide transport		64	0.00		4.00
olis	and metabolism	401	61	0.00	35	1.00
tab	[H] Coenzyme transport	1021	70	1.00		1.00
Me	and metabolism	1031	76	1.00	57	1.00
_	[I] Lipid transport and	1420	40	1.00	202	0.00
	metabolism [B] la sussais is a	1438	40	1.00	283	0.00
	[P] Inorganic ion	1525	70	1.00	110	1.00
	transport and metabolism	1525	76	1.00	110	1.00
	[U] Secondary					
	metabolites biosynthesis,	001	24	1.00	202	0.00
2		901	24	1.00	262	0.00
ا√ eri:	[K] General function	2072	200	0.07	A A 7	0.00
oor act ed		28/2	280	0.27	44/	0.00
har		1929	499	0.00	107	1.00
U U	[x] Uncharacterised	0	0	1.00	0	1.00

Table SVII: The number of genes per COG supercategory in the 12 *P. acnes* that fall within 10% of COGs with the smallest variation of MILC median and those that fall within 10% of COGs within the largest variation. The

difference of counts per COG category is tested with the binomial test with FDR correction for greater occurrence in the tight 10% than the whole set and for greater occurrence in the wide 10% than the whole sample. COG supercategories with p values below 0.05, marked in yellow, show significant difference in count for the whole set of genes regardless of COG MILC median.

		number of	number of	FDR corrected	number of	FDR corrected
	COG supercategory	genes in	genes with	p-value for	genes with	p-value for
		whole sample	tight MILC	tight vs. all	wide MILC	wide vs. all
P	[J] Translation, ribosomal					
e ar	structure and biogenesis	1167	339	0.00	54	1.00
80 90	[A] RNA processing and					
sin	modification	10	0	1.00	0	1.00
on s ces	[K] Transcription	1113	87	1.00	132	0.33
atic pro	[L] Replication,					
Ë	recombination and repair	750	51	1.00	86	0.80
Jeo	[B] Chromatin structure					
-	and dynamics	0	0	1.00	0	1.00
	[D] Cell cycle control, cell					
	division, chromosome					
	partitioning	138	0	1.00	0	1.00
	[Y] Nuclear structure	0	0	1.00	0	1.00
ing	[V] Defense mechanisms	269	3	1.00	12	1.00
llar	[T] Signal transduction					
sign	mechanisms	468	36	1.00	100	0.00
p	[M] Cell					
sai	wall/membrane/envelope					
sse	biogenesis	704	27	1.00	0	1.00
Cee	[N] Cell motility	0	0	1.00	0	1.00
brc	[Z] Cytoskeleton	0	0	1.00	0	1.00
lar	[W] Extracellular structures	0	0	1.00	0	1.00
nlla	[U] Intracellular trafficking,					
Ŭ	secretion, and vesicular					
	transport	186	39	0.00	0	1.00
	[O] Posttranslational					
	modification, protein					
	turnover, chaperones	578	65	0.56	35	1.00
	[C] Energy production and					
	conversion	1104	/1	1.00	66	1.00
	[G] Carbohydrate transport	1016	60	1.00		0.00
	and metabolism	1846	68	1.00	537	0.00
	[E] Amino acid transport	4454	64	1.00	102	1.00
	and metabolism	1454	61	1.00	103	1.00
Sm	[F] Nucleotide transport	653		1.00	0	1.00
ilo		652	66	1.00	0	1.00
tak	[H] Coenzyme transport	0.94	00	1.00	53	1.00
Me	III Lipid transport and	984	63	1.00	52	1.00
	[1] Lipiu transport and	161	22	1.00	101	0.00
		401	22	1.00	101	0.00
	[P] morganic ion transport	806	70	1.00	171	0.01
		890	78	1.00	121	0.01
	[Q] Secondary metabolites					
	catabolism	156	۵	1 00	10	0 00
z	[R] General function	130	9	1.00	19	0.90
'ly teri	nrediction only	1529	102	1 00	79	1 00
oor rac	[S] Function unknown	963	248	0.00	106	1.00 0 90
P	[X] Uncharacterized	0	0	1 00	0	1 00
-		Ŭ	Ŭ	1.50	0	1.50

References

- 1. Venter, J.C., Remington, K., Heidelberg, J.F., Halpern, A.L., Rusch, D., Eisen, J.A., Wu, D.Y., Paulsen, I., Nelson, K.E., Nelson, W. *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66-74.
- Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C. *et al.* (2005) Comparative metagenomics of microbial communities. *Science*, **308**, 554-557.
- 3. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M., Solovyev, V.V., Rubin, E.M., Rokhsar, D.S. and Banfield, J.F. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, **428**, 37-43.
- 4. Gill, S.R., Pop, M., DeBoy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett, C.M. and Nelson, K.E. (2006) Metagenomic analysis of the human distal gut microbiome. *Science*, **312**, 1355-1359.
- 5. Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, **444**, 1027-1031.
- 6. Martin, H.G., Ivanova, N., Kunin, V., Warnecke, F., Barry, K.W., McHardy, A.C., Yeates, C., He, S.M., Salamov, A.A., Szeto, E. *et al.* (2006) Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nat. Biotechnol.*, **24**, 1263-1269.
- 7. Bruggemann, H., Henne, A., Hoster, F., Liesegang, H., Wiezer, A., Strittmatter, A., Hujer, S., Durre, P. and Gottschalk, G. (2004) The complete genome sequence of Propionibacterium acnes, a commensal of human skin. *Science*, **305**, 671-673.
- 8. Hunyadkurti, J., Feltoti, Z., Horvath, B., Nagymihaly, M., Voros, A., McDowell, A., Patrick, S., Urban, E. and Nagy, I. (2011) Complete genome sequence of Propionibacterium acnes type IB strain 6609. *J Bacteriol*, **193**, 4561-4562.
- 9. McDowell, A., Hunyadkurti, J., Horvath, B., Voros, A., Barnard, E., Patrick, S. and Nagy, I. (2012) Draft genome sequence of an antibiotic-resistant Propionibacterium acnes strain, PRP-38, from the novel type IC cluster. *J Bacteriol*, **194**, 3260-3261.
- Horvath, B., Hunyadkurti, J., Voros, A., Fekete, C., Urban, E., Kemeny, L. and Nagy, I. (2012) Genome sequence of Propionibacterium acnes type II strain ATCC 11828. *J Bacteriol*, 194, 202-203.
- 11. Oda, Y., Larimer, F.W., Chain, P.S.G., Malfatti, S., Shin, M.V., Vergez, L.M., Hauser, L., Land, M.L., Braatsch, S., Beatty, J.T. *et al.* (2008) Multiple genome sequences reveal adaptations of a phototrophic bacterium to sediment microenvironments. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 18543-18548.
- 12. Larimer, F.W., Chain, P., Hauser, L., Lamerdin, J., Malfatti, S., Do, L., Land, M.L., Pelletier, D.A., Beatty, J.T., Lang, A.S. *et al.* (2004) Complete genome sequence of the metabolically versatile photosynthetic bacterium Rhodopseudomonas palustris. *Nat. Biotechnol.*, **22**, 55-61.