Supplementary Information

Cyanobacterial recycling of the extracellular matrix in microbial mats

Rhona K Stuart[1], Xavier Mayali[1], Jackson Z Lee[2], R Craig Everroad[2], Mona Hwang[1], Brad Bebout[2], Peter K Weber[1], Jennifer Pett-Ridge[1], and Michael P Thelen[1]

[1]Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA
[2]Exobiology Branch, NASA Ames Research Center, Moffett Field, CA, USA

Supplementary Materials and Methods:

LC-MS/MS Proteomics. In preparation for proteome analyses, samples (EPS-L, EPS-B and total fractions from triplicate biological replicates from each ES mats and ESFC-1 biofilms) were precipitated and concentrated according to methods from Jiao *et al* (2011). Trypsin digests were performed according to established methods, modified by removing the gel purification steps (Shevchenko *et al.*, 2007)**.** Digested peptides were analyzed by LC-MS/MS on a Thermo Scientific Q Exactive Orbitrap Mass spectrometer in conjunction Proxeon Easy-nLC II HPLC (Thermo Scientific) and Proxeon nanospray source at UC Davis Proteomics Core facility, as previously described (Yung *et al.*, 2014). Tandem mass spectra were extracted and charge state deconvoluted by Proteome Discoverer (Thermo Scientific). All MS/MS samples were analyzed using X! Tandem (The GPM, thegpm.org; version TORNADO (2013.02.01.1)). X! Tandem was set up to search either ESFC-1 genome (Everroad *et al.*, 2013) or an assembled Elkhorn Slough metagenome JGI IMG/M Project GM00369; Sample Gs0002659) and the cRAP database of common laboratory contaminants (www.thegpm.org/crap; 114 entries) plus an equal number of reverse protein sequences, with settings previously described (Yung *et al.*, 2014).

Scaffold (version Scaffold_4.3.2, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications and to calculate normalized weighted spectra values. Peptide identifications were accepted if they exceeded specific database search engine thresholds. X! Tandem identifications required at least −Log(Expect Scores) of greater than 1.2 with a mass accuracy of 5 ppm. Protein identifications required at least 2 identified peptides. Using the parameters above, the decoy false discovery rate (dFDR) was calculated to be 1.1% on the protein level and 0.0% on the spectrum level (Tabb, 2007). The dFDR is defined as the number of false hits divided by the number of positive hits multiplied by 100%. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony (Tabb, 2007).

To test for significant differences between sample fractions (EPS-L, EPS-B and Total), proteomic data was first clustered using a hierarchical clustering algorithm (Eisen *et al.*, 1998) and a bootstrapped support tree (Graur and Li, 2000) (100 iterations, Supplementary Figures S5a-b). Second, a correspondence analysis (Culhane *et al.*, 2002; Fellenberg *et al.*, 2001) was used to visualize all data points relative to sample type (Supplementary Figures S5c-d). Analyses were preformed in MeV (version 4_8 (Saeed *et al.*, 2003)). Exoproteins were defined as proteins significantly overrepresented in an extracellular fraction over the Total fraction. This was calculated by comparing normalized abundances of the extracellular fraction proteins (each EPS-L and EPS-B) to the paired Total fraction proteins with at Students t-test ($p<0.05$) and then calculating a $\log_2$ fold change enrichment for the significant proteins. All significantly enriched exoproteins were found in all three biological replicates. The proteomics data have been deposited to the ProteomeXchange Consortium (Vizcaino *et al.*, 2014) via the MassIVE partner repository with the dataset identifier MSV000079031/PXD001783. Protein localization based on genome derived predictions was

performed using CELLO (Yu *et al.*, 2006) for ESFC-1 and ES mats as well as MetaLocGramN for ESFC-1 (Magnus *et al.*, 2012). For ES mats, predictions were run on full proteins identified as the closest BLASTp hit in the IMG database (img.jgi.doe.gov).

Enzyme activity assays. Bulk enzyme activity was assayed in the EPS-L fraction following the protocol established in Bell *et al.* (2013). Briefly, diluted EPS-L samples were combined in 96-well plates with a 0.3 mM fluorescent substrate. Each substrate was tested with a range of concentrations (0.05-1.5 mM) to ensure adequate substrate was added. L-Leucine-7-amido-4-methylcoumarin hydrochloride, 4-methylumbelliferyl α-D-glucopyranoside, 4-Methylumbelliferyl β-D-glucopyranoside and 4-Methylumbelliferyl β-D-cellobioside (Sigma Aldrich, St. Louis, MO, USA) were the substrates for peptidase, α-glucosidase, β-glucosidase, and cellulase activity respectively. 10 μM of 4-Methylumbelliferyl acetate and 7-Amino-4-methylcoumarin were added to each sample as standards. Controls included each sample with water and sterile 10% NaCl (substrate control) instead of sample. Plates were read every 15 minutes for 3 hours at 380/440 nm excitation/emission. Concentration was calculated at each time point according to the following:

(Assay – Substrate Control) / ([(Standard – Water) / (10umol/L × 0.00005L)] × 0.0002L)

Enzyme activity was then calculated by the slope of the line from all the time points and normalized to protein content in each EPS-L sample (measured by 280 nm absorbance).

Secondary Ion Mass Spectrometry (NanoSIMS) isotope imaging and Scanning Electron Microscopy (SEM). Samples were coated with ~5 nm of gold and then imaged with a FEI Inspect F SEM (Hillsboro, OR) to identify regions of interest for SIMS imaging. SIMS imaging was performed with a Cameca NanoSIMS 50 at Lawrence Livermore National Laboratory. A focused 2 pA, approximately 150 nm, 16 keV $^{133}Cs^{+}$ primary ion beam was

rastered over [15 x 15 μm^2] to [25 x 25 μm^2] analysis areas with 256 x 256 pixels and a dwell time of 1 ms/pixel 19-30 scans (cycles). Before analysis, samples were pre-sputtered with 90 pA of Cs^+ current (equivalent to approximately 50 nm) in order to reach sputtering equilibrium and to make sure that the isotope analysis targeted intracellular material rather than the surface of the cells (Supplementary Figure S6). Serial quantitative secondary ion images (maps) were simultaneously collected for $^{12}C^-$, $^{13}C^-$, and $^{12}C^{14}N^-$ using electron multipliers in pulse counting mode. For the later supplemental experiment $^{12}C_2^-$, $^{13}C^{12}C^-$, and $^{12}C^{14}N^-$ were collected since previous work indicates that the dimer (e.g $^{12}C^{12}C$) has better yield than the monomer (Pett-Ridge and Weber, 2012). Secondary electrons were also simultaneously collected as previously described (Pett-Ridge and Weber, 2012).

NanoSIMS ion image data were processed as described in Woebken *et al*., 2012. For each raster, quantitative ion ratio images were generated from the summed $^{13}C^{12}C^-$ and $^{12}C_2^-$ ion images to generate ^{13}C-enrichment images, presented in atom percent excess (APE) (Pett-Ridge and Weber, 2012; Popa *et al.*, 2007). Regions of interest (ROIs) for quantification of isotopic ratios were selected based on secondary electron, ^{13}C-enrichment images, and $^{12}C^{14}N^-$ ion images, which allowed cells to be specifically selected and hotspots of residual EPS to be excluded. Isotopic ratios were extracted by cycle and averaged. There is evidence for a small reduction in ^{13}C enrichment in cells following fixation (~5%, (Woebken *et al.*, 2015)), but all cells in these experiments were fixed with 4% PFA, so this dilution should be consistent with respect to enrichment levels.

REFERENCES

Bell CW, Fricks BE, Rocca JD, Steinweg JM, McMahon SK, Wallenstein MD (2013). High-throughput fluorometric measurement of potential soil extracellular enzyme activities. *J Vis Exp***:** e50961.

Culhane AC, Perriere G, Considine EC, Cotter TG, Higgins DG (2002). Between-group analysis of microarray data. *Bioinformatics* **18:** 1600-1608.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95:** 14863-14868.

Everroad RC, Woebken D, Singer SW, Burow LC, Kyrpides N, Woyke T *et al.,* (2013). Draft genome sequence of an oscillatorian cyanobacterium, strain ESFC-1. *Genome Announc* **1:** e00527.

Fellenberg K, Hauser NC, Brors B, Neutzner A, Hoheisel JD, Vingron M (2001). Correspondence analysis applied to microarray data. *Proc Natl Acad Sci U S A* **98:** 10781-10786.

Graur D, Li W-H (2000). *Fundamentals of molecular evolution, Second edition*. Sinauer Associates, Inc. Sunderland, Massachusetts.

Jiao Y, D'Haeseleer P, Dill BD, Shah M, VerBerkmoes NC, Hettich RL *et al.,* (2011). Identification of Biofilm Matrix-Associated Proteins from an Acid Mine Drainage Microbial Community. *Applied and Environmental Microbiology* **77:** 5230-5237.

Magnus M, Pawlowski M, Bujnicki JM (2012). MetaLocGramN: a meta-predictor of protein subcellular localization for Gram-negative bacteria. *Biochim Biophys Acta* **1824:** 1425-1433.

Pett-Ridge J, Weber PK (2012). NanoSIP: NanoSIMS Applications for Microbial Biology. *Microbial Systems Biology*. Humana Press. pp 375-408.

Popa R, Weber PK, Pett-Ridge J, Finzi JA, Fallon SJ, Hutcheon ID *et al.,* (2007). Carbon and nitrogen fixation and metabolite exchange in and between individual cells of *Anabaena oscillarioides*. *ISME J* **1:** 354-360.

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N *et al.,* (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34:** 374-378.

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2007). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1:** 2856-2860.

Tabb DL (2007). What's Driving False Discovery Rates? *J Proteome Res* **7:** 45-46.

Vizcaino JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Rios D *et al.,* (2014). ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32:** 223-226.

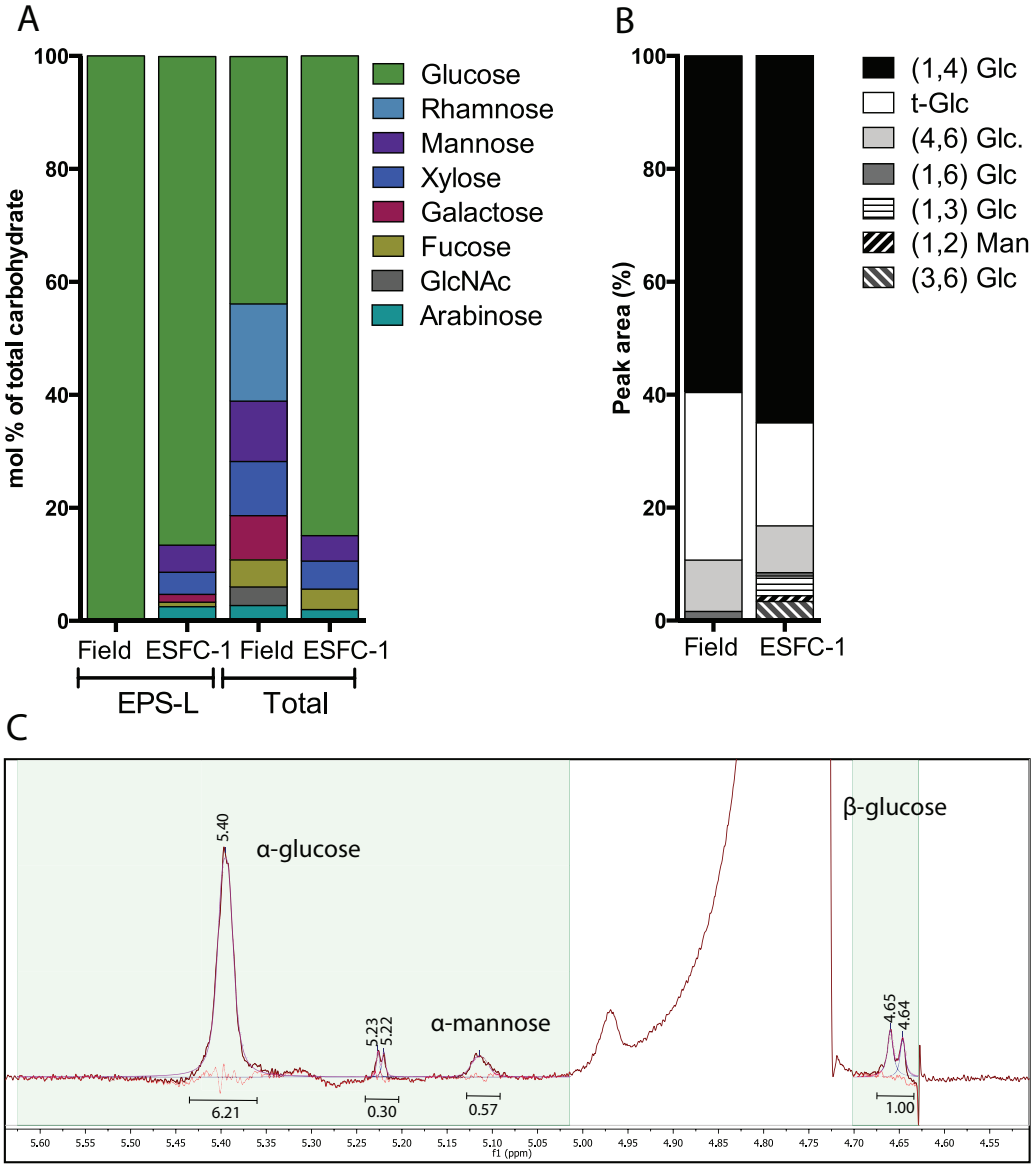
Woebken D, Burow LC, Prufert-Bebout L, Bebout BM, Hoehler TM, Pett-Ridge J *et al.,* (2012). Identification of a novel cyanobacterial group as active diazotrophs in a coastal microbial mat using NanoSIMS analysis. *ISME J* **6:** 1427-1439.

Woebken D, Burow LC, Behnam F, Mayali X, Schintlmeister A, Fleming ED *et al.,* (2015). Revisiting N(2) fixation in Guerrero Negro intertidal microbial mats with a functional single-cell approach. *ISME J* **9:** 485-496.

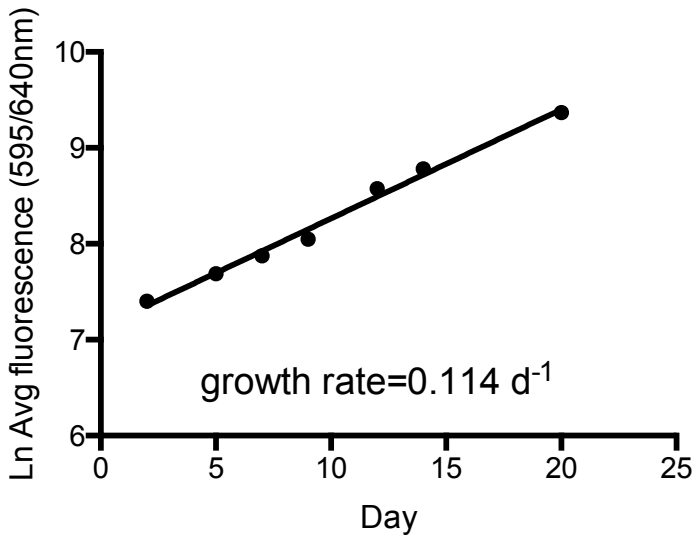
Yu CS, Chen YC, Lu CH, Hwang JK (2006). Prediction of protein subcellular localization. *Proteins* **64:** 643-651.

Yung MC, Ma J, Salemi MR, Phinney BS, Bowman GR, Jiao Y (2014). Shotgun proteomic analysis unveils survival and detoxification strategies by *Caulobacter crescentus* during exposure to uranium, chromium, and cadmium. *J Proteome Res* **13:** 1833-1847.

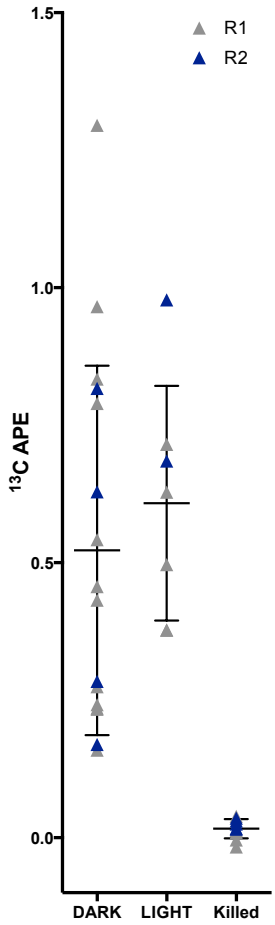
Supplementary Figures and Tables



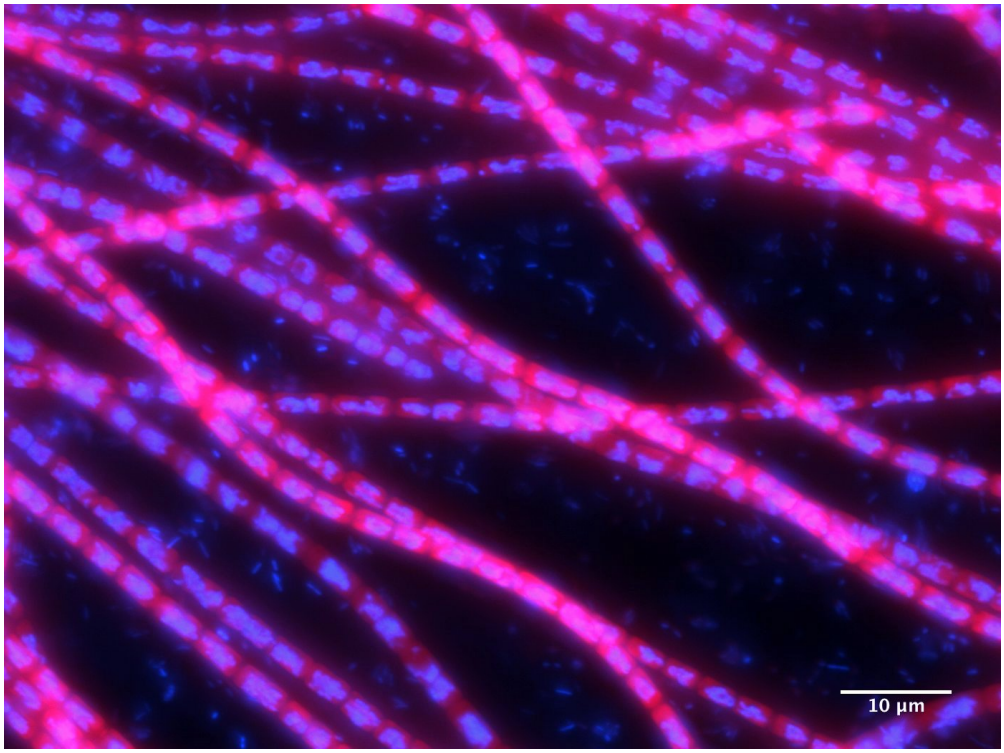
Supplementary Figure S1. Carbohydrate composition analyses for the EPS-L fraction of ESFC-1. (A) Glycosyl composition analysis of one sample of both EPS-L and Total fractions from ESFC-1 biofilm and ES mat (for comparison). (B) Linkage analysis of one EPS-L sample from ESFC-1 biofilm and ES mats. "Glc" is Glucose, "t-Glc" is terminal glucose and "Man" is mannose. (C) Anomeric portion of the 1D-1H NMR spectra of EPS-L from culture showing the regions used for line fitting. Integrated peak areas indicated under bracketed lines.



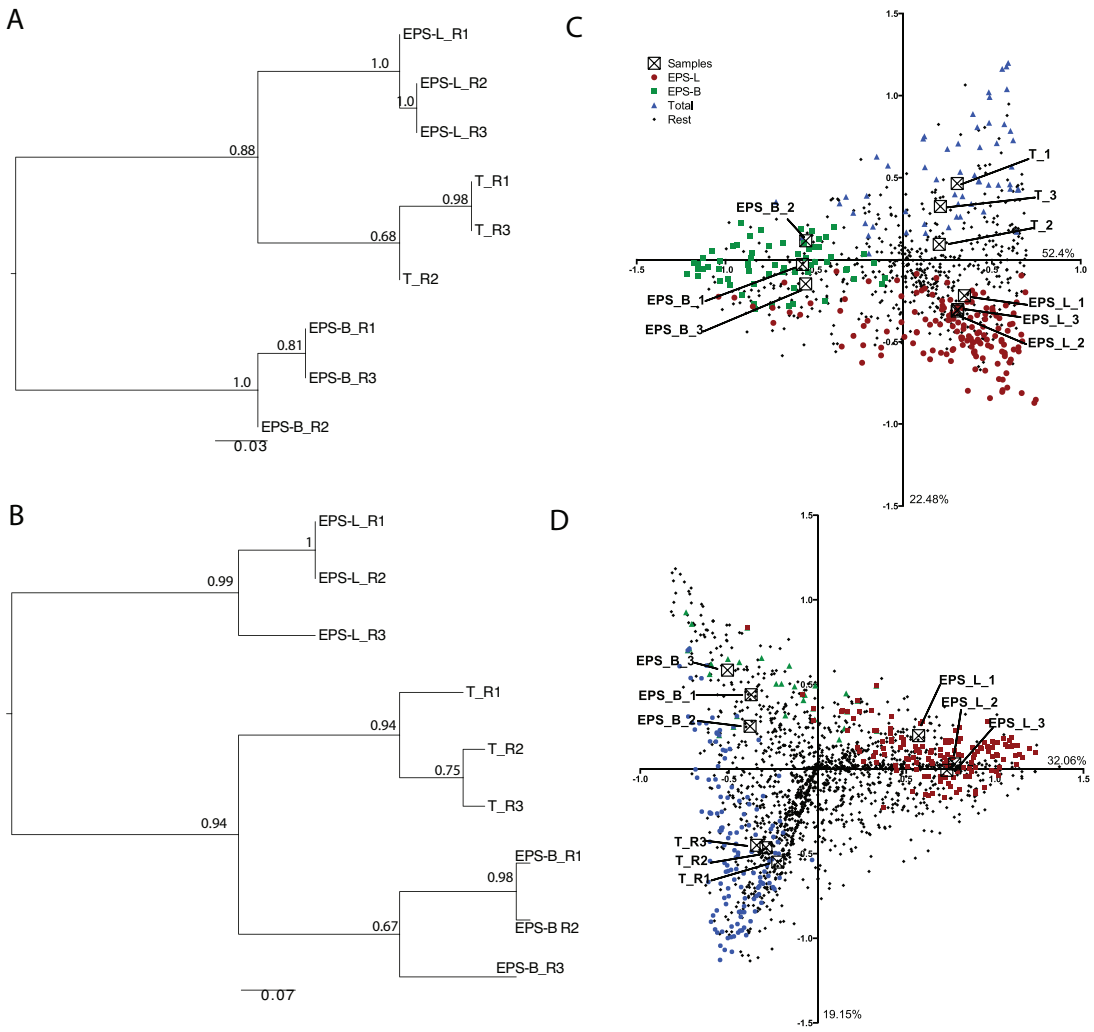
Supplementary Figure S2. ESFC-1 growth rate based on average fluorescence of 30 replicate wells.



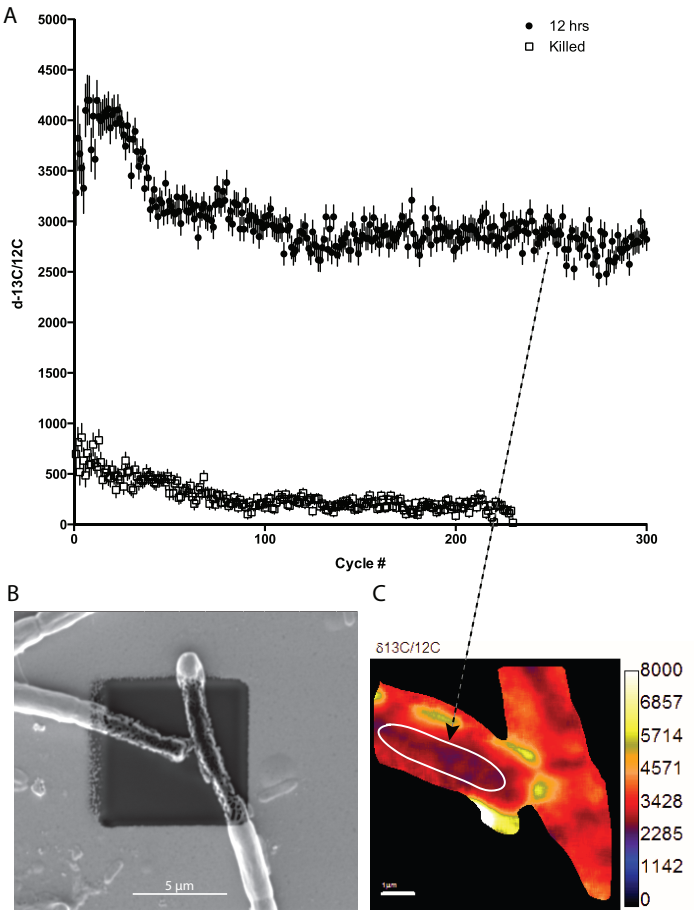
Supplementary Figure S3. Control experiments, NanoSIMS analysis of ESFC-1 trichomes incubated with ^{13}C EPS. (A) Plot of ^{13}C enrichment (APE, atom percent excess) of trichomes 6 hours after the addition of ^{13}C label, with points representing values from two biological replicates (R1or R2). All additions and treatments started at beginning of 12:12 light cycle. "DARK" is dark treatment (n=16), and "LIGHT" is light treatment (n=7). "Killed" (n=10) represents control cells that were fixed before incubation. ^{13}C EPS was from a different batch of labeled EPS (67.5 ^{13}C APE, 1:30 dilution into media), so enrichment values are not directly comparable with data presented in Figure 6.



Supplementary Figure S4: Sample microscopy image used to calculate relative biovolumes for ESFC-1 and contaminating microbes. Image is a calculated maximum intensity projection of a 90 µm z-stack. Red is autofluorescent ESFC-1 trichomes and blue is DAPI (4',6-diamidino-2-phenylindole) a general DNA stain.



Supplementary Figure S5. Hierarchical clustering support trees based on all proteins and abundances in each sample for ESFC-1 (A) and Elkhorn Slough (B). Numbers at nodes represent bootstrap values (100 iterations). Correspondence analysis (Culhane *et al.*, 2002; Fellenberg *et al.*, 2001) of all three cellular fractions (inertia values on each axis) for both ESFC-1 culture (C) and Elkhorn Slough (D). Colored points indicate significantly enriched proteins in each fraction (EPS-L Red; EPS-B Green; Total Blue).



Supplementary Figure S6. Initial NanoSIMS analyses of ESFC-1 trichomes (A) d-$^{13}C/^{12}C$ values for at least 200 raster cycles of a kill control and 12h sample. (B) SEM image after 200 cycles on corresponding 12h sample to indicate cellular material has been mostly sputtered away. (C) Corresponding d-$^{13}C/^{12}C$ image from 12h sample indicating the circled region of interest. Initial analyses show d-$^{13}C/^{12}C$ enrichments are elevated at cell surface. All further measurements were made after an initial pre-sputter to remove cell surface.

Supplementary Table S1 (Excel spreadsheet). All identified proteins and normalized spectral count abundances (average and standard deviation of three biological replicates), separate tabs for both ES mat and culture.

Supplementary Table S2: Homologous exoproteins in ES mat and ESFC-1

Meta-proteome closest genome locus	ESFC-1 locus	Functional Category	Description	ESFC-1 EPS_L	ESFC-1 EPS_B	ESFC-1 T	ES mat EPS_L	ES mat EPS_B	ES mat T	BlastP E-value	% ID	Localization prediction (ESFC-1/mat)
MC7420_1347	A3MYDRAFT_0167	Cell Wall, Capsule and Secreted	Bacterial pre-peptidase C-terminal domain.	**13.38**	23.67	8.46	**9.65**	**34.47**	2.90	0	65.3	EC/Peri
MC7420_5156	A3MYDRAFT_4337	Cell Wall, Capsule and Secreted	dTDP-4-dehydrorhamnose 3,5-epimerase	**4.74**	2.21	1.00	**1.94**	0.00	0.74	0	63.8	Cyto/Cyto
MC7420_7671	A3MYDRAFT_2567	Cell Wall, Capsule and Secreted	Secreted and surface protein containing fasciclin-like repeats**	9.24	**29.95**	4.42	**42.03**	2.48	3.78	0	71.4	Cyto/Cyto
Chr6712_2784	A3MYDRAFT_3398	Redox/Oxidative Stress Response	thioredoxin**	61.00	**180.68**	81.98	**210.91**	178.54	81.33	5E-67	87	Cyto/Cyto
MC7420_1916	A3MYDRAFT_4781	Redox/Oxidative Stress Response	Peroxiredoxin, PRX-like1**	1.00	**24.51**	1.38	**19.79**	0.00	3.71	0	70.1	Peri/Peri
MC7420_2640	A3MYDRAFT_2187	Redox/Oxidative Stress Response	Peroxiredoxin, PRX5 (EPS_B)	108.98	**150.35**	101.44	55.42	**68.01**	34.59	0	84.2	Peri/Peri
MC7420_8119	A3MYDRAFT_4267	Redox/Oxidative Stress Response	ferredoxin [2Fe-2S]	**69.74**	1.00	10.39	**521.76**	116.57	155.50	0	69.7	EC/Cyto
MC7420_4871	A3MYDRAFT_0260	Protein Metabolism	C-terminal peptidase (prc)	**32.70**	30.52	2.48	**27.63**	8.45	8.52	0	76.4	Cyto/Cyto
MC7420_7761	A3MYDRAFT_0778	Protein Metabolism	Predicted Zn-dependent peptidases	**16.83**	37.40	3.50	**38.25**	4.56	3.00	0	64.3	EC/Cyto
MC7420_7808	A3MYDRAFT_0777	Protein Metabolism	Predicted Zn-dependent peptidases	**16.33**	12.77	1.61	**4.74**	0.00	1.00	0	87.9	Cyto/OM
MC7420_3997	A3MYDRAFT_4772	Amino Acids and Derivatives	LL-diaminopimelate aminotransferase	**46.65**	36.61	37.00	**49.09**	3.76	11.27	0	85.6	Cyto/OM
MC7420_1588	A3MYDRAFT_1169	N,P,S Metabolism	sulfate adenylyltransferase	**68.82**	17.06	20.62	**25.19**	**17.30**	8.33	0	79	Cyto/Cyto
MC7420_2114	A3MYDRAFT_0160	N,P,S Metabolism	Inorganic pyrophosphatase	**49.88**	35.60	31.75	**142.49**	37.97	35.81	0	73.2	Cyto/Cyto
MC7420_6158	A3MYDRAFT_0759	N,P,S Metabolism and Amino Acids and Derivatives	cysteine synthase A	**9.64**	2.47	3.38	**10.89**	0.71	4.00	0	84	Cyto/Cyto
MC7420_6421	A3MYDRAFT_2192	DNA, Nucleosides and Nucleotides metabolism	IMP dehydrogenase family protein	**50.55**	29.26	23.40	**22.12**	**15.73**	3.21	0	70.3	Cyto/Cyto
MC7420_4416	A3MYDRAFT_3917	RNA Metabolism	Ribonuclease HI	3.429	**9.877**	4.176	**13.992**	1.122	8.465	0	61.1	Cyto/EC
MC7420_4292	A3MYDRAFT_2994	Carbon metabolism, other	Dienelactone hydrolase/carboxymethylenebutenolidase	**18.21**	8.93	4.34	**125.37**	1.06	10.12	0	71.1	EC/Cyto
MC7420_5787	A3MYDRAFT_2673	Central carbohydrate metabolism OR Cell Wall and Capsule	6-phosphogluconolactonase**	1.67	**14.98**	1.52	**4.89**	0.00	1.85	0	75.9	Cyto/Cyto
MC7420_5659	A3MYDRAFT_3274	Unknown or General	PPiase, Peptidyl-prolylcis-transisomerase (rotamase)-cyclophilin family**	44.89	**132.81**	68.76	**13.34**	0.88	0.99	0	64.9	Peri/Peri
MC7420_2220	A3MYDRAFT_3944	Unknown or General	Uncharacterized conserved protein**	1.00	**7.02**	1.38	**1.88**	1.08	1.00	0	90.1	Cyto/Cyto
MC7420_507	A3MYDRAFT_3184	Unknown or General	FOG: FHA domain	**12.78**	24.38	3.81	**3.72**	0.00	1.00	0	68.5	Cyto/Cyto
MC7420_6877	A3MYDRAFT_3669	Unknown or General	NAD-dependent aldehyde dehydrogenases	**14.03**	4.12	8.19	**12.59**	1.50	4.14	0	75.2	Cyto/Cyto

Supplementary Table S3: Bacterial counts and biovolumes (averages from 10 fields of view)

	Average # bacteria/27000 μm^3	Average bacterial μm^3/27000 μm^3	Average ESFC1 μm^3/27000 μm^3
Killed	212.17±85	50.9±22.1	1343.45±422.9
12 hour	145.9±62	35.9±17.6	1306.8±225.7
Average	179±47	43.4±10.6	1325.1±25.9