Supporting Information

Carini et al. 10.1073/pnas.1505034112

SI Methods

Harvest Methods for Fig. S4. To investigate whether str. HTCC7211 cells contained non-P lipids when grown exclusively with MPn as the sole P source, cells grown with excess MPn (100 μM) as a sole P source for more than three subcultures were harvested in midlogarithmic phase by centrifugation.

Harvest Methods for *P. ubique* Lipid Profiles (Fig. S5). P_i replete lipid profiles for *P. ubique* were obtained by harvesting cells grown with excess P_i (100 μM) in midlogarithmic growth phase (approximately 1.0×10^8 cells·mL^{-1}) by centrifugation. Half of the cell pellet was frozen immediately at −80 °C until lipid extraction. The remainder of the cell pellet was washed and resuspended in growth medium without added P_i for 4 d. After 4 d, the suspended cells were harvested by centrifugation for analysis of P_i starved lipids.

Polar Membrane Lipid Analysis. Lipids were extracted as described in ref. 36. Briefly, 2 mL of methanol, 1 mL of dichloromethane and 0.8 mL of PBS were added to a glass centrifuge tube. The internal recovery standard was added as 20 μL of 65.6 μM dinitrophenyl-phosphatidylethanolamine (DNP-PE) in methanol. Cells were disrupted in an ultrasonic bath for 10 min. An additional 1 mL of dichloromethane and 1 mL of phosphate buffer were added. The samples were centrifuged for 5 min at 5,159 × *g*. The organic phase (lower) was transferred to a new 2-mL HPLC vial (2-mL screw cap borosilicate glass) by using a combusted (450 °C for 4 h) glass pipette, and concentrated to 200 μL by drying under nitrogen. An aliquot of 20 μL of the total lipid extract was analyzed by high performance liquid chromatography mass spectrometry (HPLC/MS).

Chromatography conditions have been described (36). In brief, chromatographic separation was performed on an Agilent 1100 HPLC system (Agilent Technologies) consisting of a binary pump, autosampler, and a photodiode-array detector, leading into a Thermo Scientific LCQ Fleet ion trap mass spectrometer. A PrincetonSPHER DIOL column (150 × 2.1 mm inner diameter, 100 Å, 5 μm particle size; Princeton Chromatography) was used with a gradient of *n*-hexane:isopropanol:formic acid:aqueous ammonium hydroxide (25%; by volume) at a ratio of 800:200:1.2:04 (by volume) and isopropanol:water: formic acid:aqueous ammonium hydroxide (25%; by volume) at a ratio of 900:100:1.2:0.4 (by volume). The photodiode-array monitored 340 nm wavelength, which is the wavelength of maximum absorption of the internal recovery standard DNP-PE.

The ion-trap mass spectrometry methods used positive electrospray ionization and are loosely based on ref. 37. The heated electrospray source (HESI) was configured as follows; probe temperature 100 °C, capillary temperature 200 °C, spray voltage 4.5 kV, and sheath gas 30 and auxiliary gas 10 (both arbitrary units). Full scan data were collected in both positive and negative ion mode with data-dependent MS/MS spectra performed on the most abundant peak from each full scan. The normalized collision energy for the MS/MS spectra was 35.

MGDG, PE, and PG were identified and quantified directly against external curves of authentic standards. PE and PG were obtained from Avanti Polar Lipids and MGDG from Matreya LLC as described in ref. 37. Standards for ornithine lipids and GADG (described below) are not commercially available; thus, we quantified these molecules by using the MGDG standard curve. All analytes were normalized to the internal recovery standard DNP-PE.

To facilitate elucidation of GADG and ornithine lipids, data obtained by the ion-trap mass spectrometer was complemented by data from a high mass-resolution orbitrap (Exactive Plus; Thermo Scientific) mass spectrometer in conjunction with the chromatographic scheme described above. The GADG was first tentatively identified with the ion-trap mass spectrometer based on the MS^2 mass spectrum of a peak eluting immediately after PE; the nominal *m/z* was 760 with monoacyl glycerol (i.e., "lyso") fragments of *m/z* 311 and 313, and major diacylglycerol (i.e., DAG) fragments of *m/z* 549 representing head group neutral losses of 211. The samples were analyzed again by using the orbitrap in all-ion fragmentation mode (AIF; Fig. S6). The *m/z* of the tentative GADG peak was 760.55286 (Fig. S6) led to the assignment of the structural formula $C_{41}H_{78}NO_{11}(+)$ (δ = −5.4 ppm), which was an ammonium adduct of a molecule with the structural formula of $C_{41}H_{74}O_{11}(+)$. The ion-trap MS^2 fragments were also observed in the AIF mass spectrum, which contained a 549.48565 DAG fragment ($C_{35}H_{65}O_4(+)$; δ = −3.8 ppm), representing a concomitant neutral loss of 211.06721 consistent with glucuronic acid ($C_6H_6O_7(NH_4)$; δ = 0.0 ppm). In addition, lyso fragments of 311.25670 and 313.27242 were observed in the AIF mass spectrum, which are consistent with fragments containing 16:1 ($C_{19}H_{35}O_3(+)$;δ = −4.4 ppm) and 16:0 ($C_{19}H_{37}O_3(+)$;δ = −4.1 ppm) fatty acid moieties. This fragmentation pattern was effectively identical to that described recently for a 16:1/17:0 GADG (32). Although the GADG with the 16:0/16:1 fatty acids was predominant, species with 16:1/16:1, 16:0/18:1, and 16:1/18:1 were also common. Less common species were observed as molecular ions but not abundant enough to trigger MS^2 analysis in the ion trap; however, lyso fragments were observed in the AIF spectrum, consistent with 17:1 ($C_{20}H_{37}O_3(+)$;δ = −3.4 ppm), suggesting 16:0/17:1, 16:1/17:1, and 18:1/17:1.

An ornithine lipid eluting after GADG was tentatively identified on the ion trap as a molecular ion with a nominal *m/z* 625, which fragmented on MS^2 analysis to *m/z* 369 and 351, which are characteristic of 16:0 neutral ketene and neutral ketene-H_2O losses, respectively. The molecular ion at the same relative retention time yielded an *m/z* 625.54698 on the orbitrap (Fig. S7), which is consistent with a formula for an 3-OHornithine lipid ($C_{37}H_{73}N_2O_5(+)$; δ = −7.1 ppm). The AIF mass spectrum confirmed the formula of the neutral ketene losses with a fragments of *m/z* 369.30963 ($C_{21}H_{41}N_2O_3(+)$;δ = −4.2 ppm) and the 351.29913 ($C_{21}H_{39}N_2O_2(+)$; δ = −4.2 ppm). In addition, the "fingerprint" ornithine ion of ornithine lipids (45) was also found with *m/z* 115.08635 ($C_5H_{11}N_2O(+)$;δ = −2.1 ppm). The OL fragmentation pattern was exactly the same as reported for a confirmed OL (45).

The orbitrap mass spectrometry methods used positive electrospray ionization and are also based on ref. 37. The HESI source was configured as follows: probe temperature 100 °C, capillary temperature 200 °C, spray voltage 4.5 kV, and sheath gas 30 and auxiliary gas 10 (both arbitrary units). Full scan data were collected in positive ion mode with AIF spectra performed following each full scan. The normalized collision energy for the AIF spectra was 35. The resolution for both full scan and AIF spectra was set to 140,000. Mass error for our internal standard DNP-PE was −6.3 ppm for the GADG analytical run, and −7.8 ppm for the OL run.

Profile HMM Searches. Protein sequences from characterized *olsA* genes and best BLASTP hits (14 total sequences) were aligned with MUSCLE (41) and HMMs were generated by using the HMMER3 suite (46) using hmmbuild on default settings. The HMM was searched against a concatenated set of amino acid

sequences from all SAR11 genomes in this study by using hmmsearch on default settings.

OlsA HMM Search. The following OlsA amino acid sequences were used to train the HMM (listed as: IMG gene ID, accession, IMG product name, [scaffold source]):

1. 646741023 YP_003579131 phospholipid/glycerol acyltransferase [*Rhodobacter capsulatus* SB1003 chromosome: NC_014034].
2. 637181186 NP_384510 putative acyltransferase transmembrane protein [*Sinorhizobium meliloti* 1021: NC_003047].
3. 639295844 NP_531061 1-acyl-sn-glycerol-3-phosphate acyltransferase [*Agrobacterium tumefaciens* str. C58 chromosome circular: NC_003304].
4. 2579691166 1-acyl-sn-glycerol-3-phosphate acyltransferase [*Defluviimonas* sp. 20V17: AYXI01000172].
5. 640485362 YP_001169970 hypothetical protein [*Rhodobacter sphaeroides* ATCC 17025 plasmid pRSPA01: NC_009429].
6. 641468503 ZP_02153707 phospholipid/glycerol acyltransferase [*Oceanibulbus indolifex* HEL-45, unfinished sequence: NZ_ABID01000003].
7. 638006516 YP_613204 phospholipid/glycerol acyltransferase [*Silicibacter* sp. TM1040: NC_008044].
8. 639047591 ZP_01156799 acyltransferase, putative [*Oceanicola granulosus* HTCC2516, unfinished sequence: NZ_AAOT01000018].
9. 641157421 ZP_01903725 phospholipid/glycerol acyltransferase [*Roseobacter* sp. AzwK-3b, unfinished sequence: NZ_ABCR01000007]
10. 640639847 ZP_01742258 acyltransferase, putative [*Rhodobacterales bacterium* HTCC2150, unfinished sequence: NZ_AAXZ01000003]
11. 643824249 YP_002824597 predicted 1-acyl-sn-glycerol-3-phosphate acyltransferase [*Rhizobium* sp. NGR234: NC_012587]
12. 2509388809 1-acyl-sn-glycerol-3-phosphate acyltransferase [*Rhizobium leguminosarum* bv. trifolii WSM597: Rleg9-DRAFT_RLH.2]
13. 2571461232 1-acyl-sn-glycerol-3-phosphate acyltransferase [*Agrobacterium radiobacter* DSM 30147: ASXY01000442]
14. 637814933 YP_415473 Phospholipid/glycerol acyltransferase [*Brucella melitensis* biovar Abortus 2308 chromosome I: NC_007618].

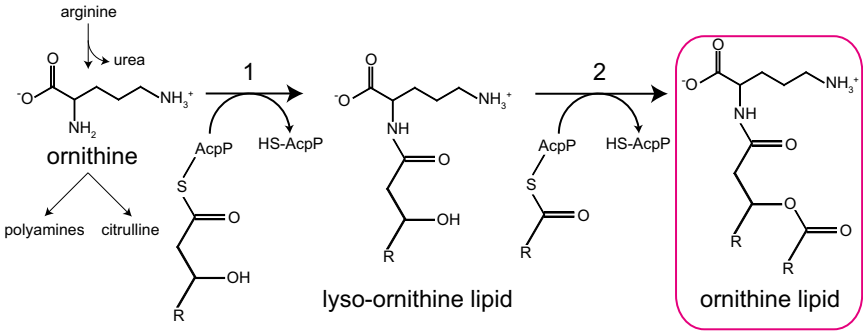

Fig. S1. Proposed ornithine lipid biosynthetic pathway in *Pelagibacter* sp. str HTCC7211. 1, lyso-ornithine synthase (*olsB*; HTCC7211_00011000); 2, OlsA in canonical ornithine lipid synthesis. We speculate the str. HTCC7211 gene HTCC7211_00011010 encodes a putative *olsA*-like *O*-acetyltransferase that catalyzes reaction 2.

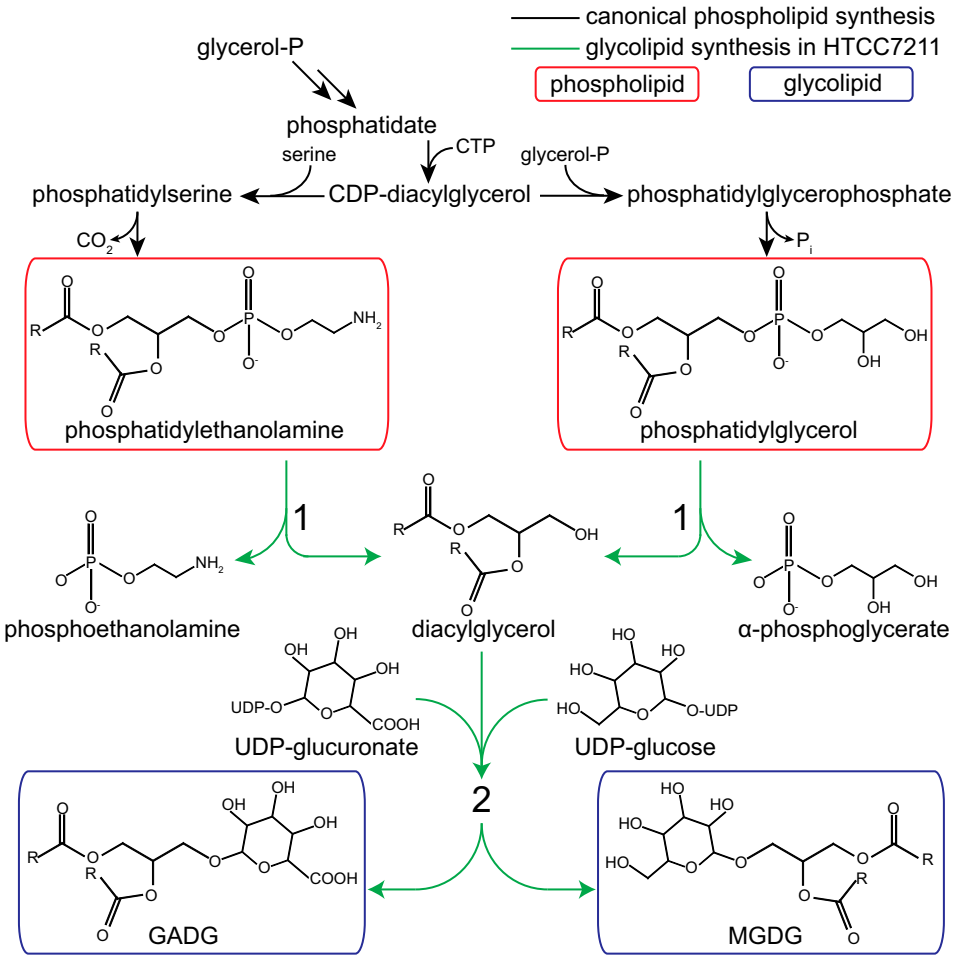

Fig. S2. Proposed glycolipid biosynthesis pathway in *Pelagibacter sp*. str. HTCC7211. 1, phospholipase C (*plsC*; HTCC7211_00011030) 2, bifunctional GADG/MGDG glycosyltransferase (HTCC7211_00011020).

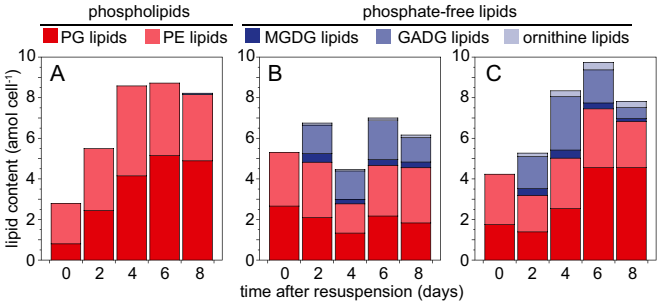

Fig. S3. Time course of molar lipid composition per str. HTCC7211 cell in AMS1 growth medium containing excess P_i (*A*); no P_i (*B*); or excess MPn (*C*). Because lipid content scales with cell size, greater total lipid content per cell likely reflects larger cells.

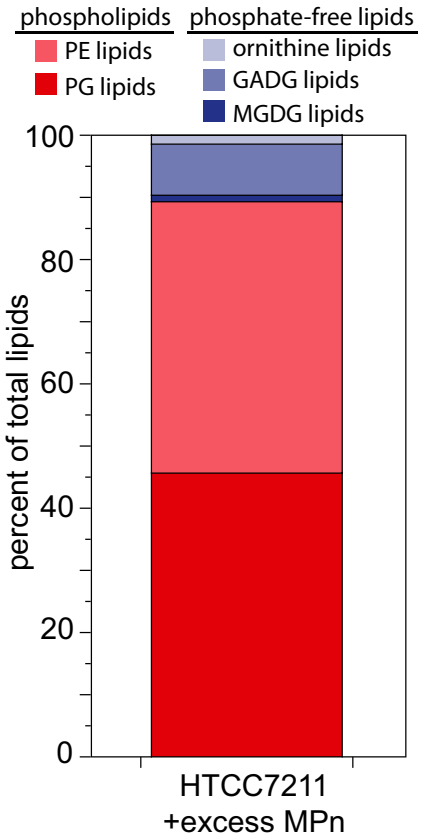

Fig. S4. Lipid composition of *Pelagibacter* sp. str. HTCC7211 grown with MPn as the sole P source for >20 generations.

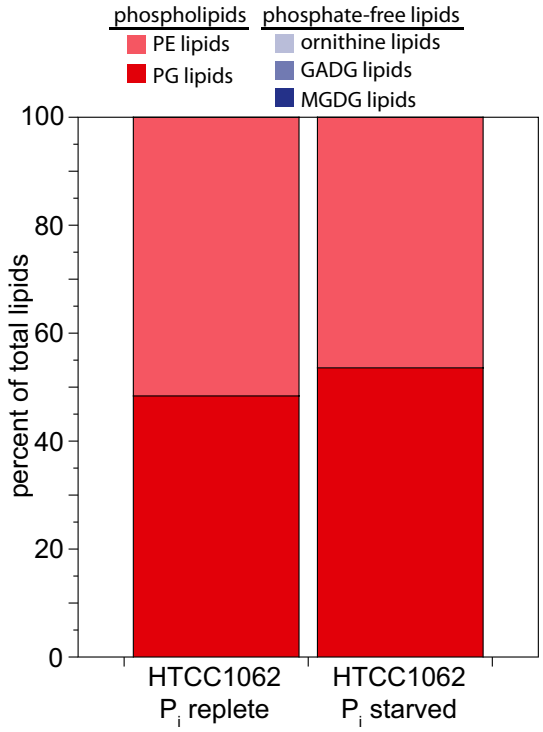

Fig. S5. Lipid profiles of *P. ubique* str. HTCC1062 in synthetic medium under P_i replete and P_i starved conditions.

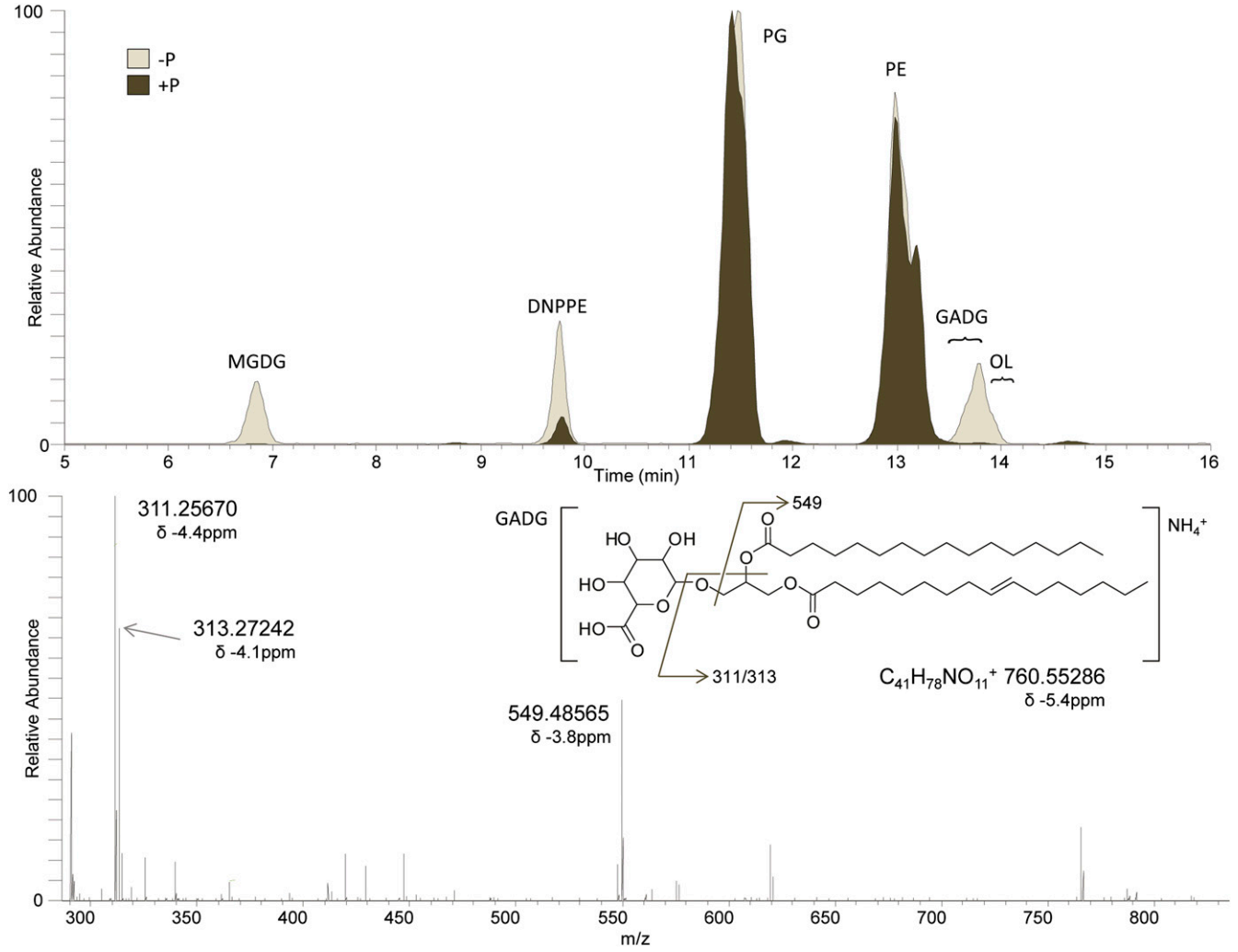

Fig. S6. Identification of GADG by high mass-resolution mass spectrometry. (*Top*) A total ion current chromatogram of lipid extracts of HTCC7211 grown under P_i replete (+P) and P_i deplete (-P) conditions. (*Bottom*) AIF mass spectrum averaged from 13.6 to 13.8 min, which shows the DAG and lyso fragments from GADG.

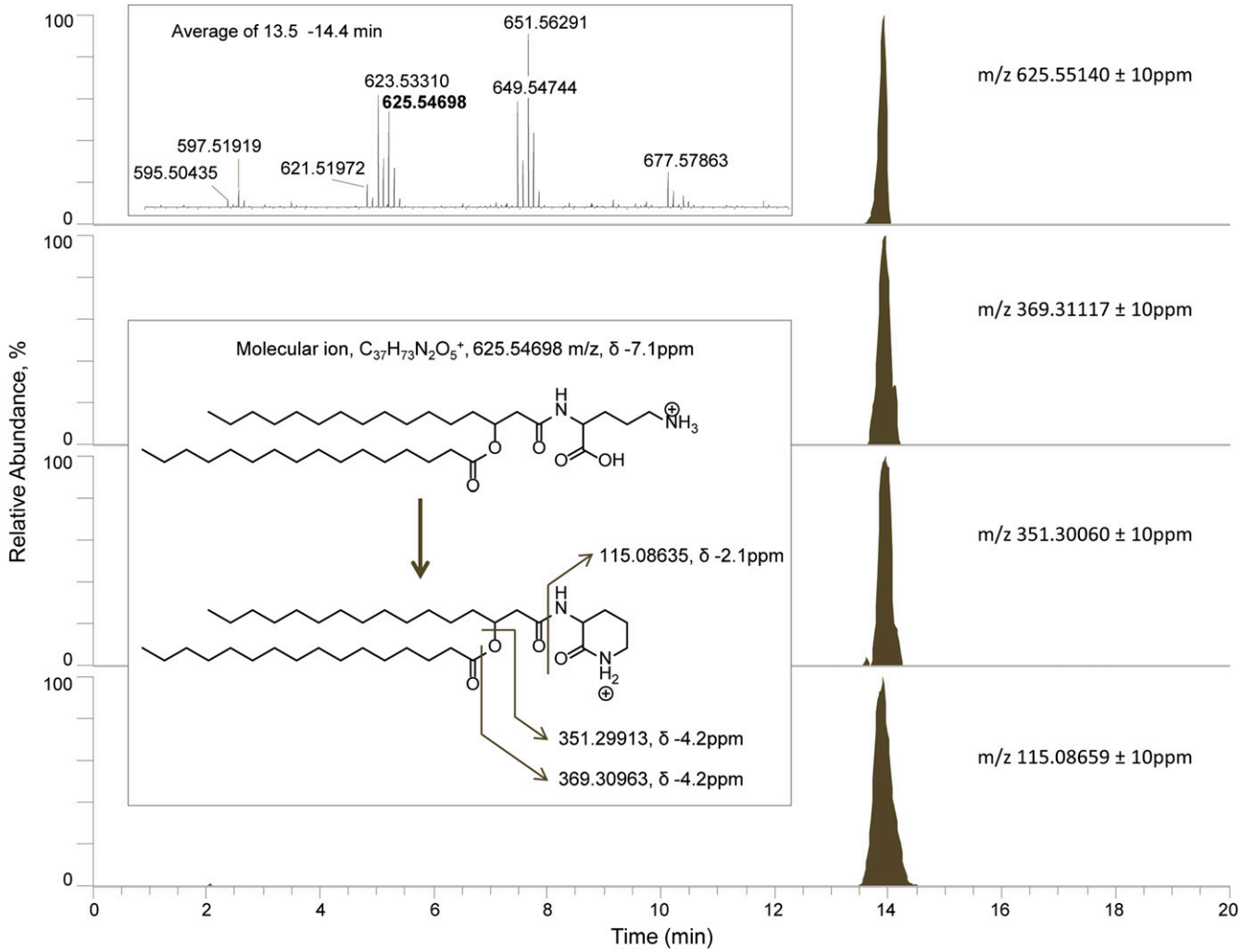

Fig. S7. Identification of ornithine lipids by high mass-resolution mass spectrometry. Chromatograms from AIF scans showing peaks centered (±10 ppm) on the predicted exact masses for the putative OL 3-OH 16:0/16:0, including from top to bottom, the OL molecular ion (*m/z* 625.55140), neutral ketene (*m/z* 369.31114), neutral ketene minus H_2O (*m/z* 351.30060), and ornithine fingerprint ions (*m/z* 115.08659). *Top Inset* shows a range of molecular ions eluting between 13.5 and 14.4 min and represent additional potential OL. *Bottom Inset* shows the structure of OL 3-OH 16:0/16:0, along with the observed *m/z* of the molecular ion and fragments.

Dataset S1. Data used to construct Fig. 1

Dataset S1

"All IPs" tab: Hal run orthology results for all four genes in Fig. 1 at all inflation parameters. "1.5 only" tab: Hal results for inflation parameter 1.5 used to construct the (*i*) open/filled bars in Fig. 1 and (*ii*) to determine gene synteny. "Synteny check all" tab shows adjacent genes (highlighted in green) for each organism. Tabs for 16S and 23S IDs are also noted.

Dataset S2. Carbon, nitrogen, and phosphorus composition of lipid polar head groups in SAR11 and other organisms

Dataset S2

Data were used to construct Fig. 4.