Supporting Information

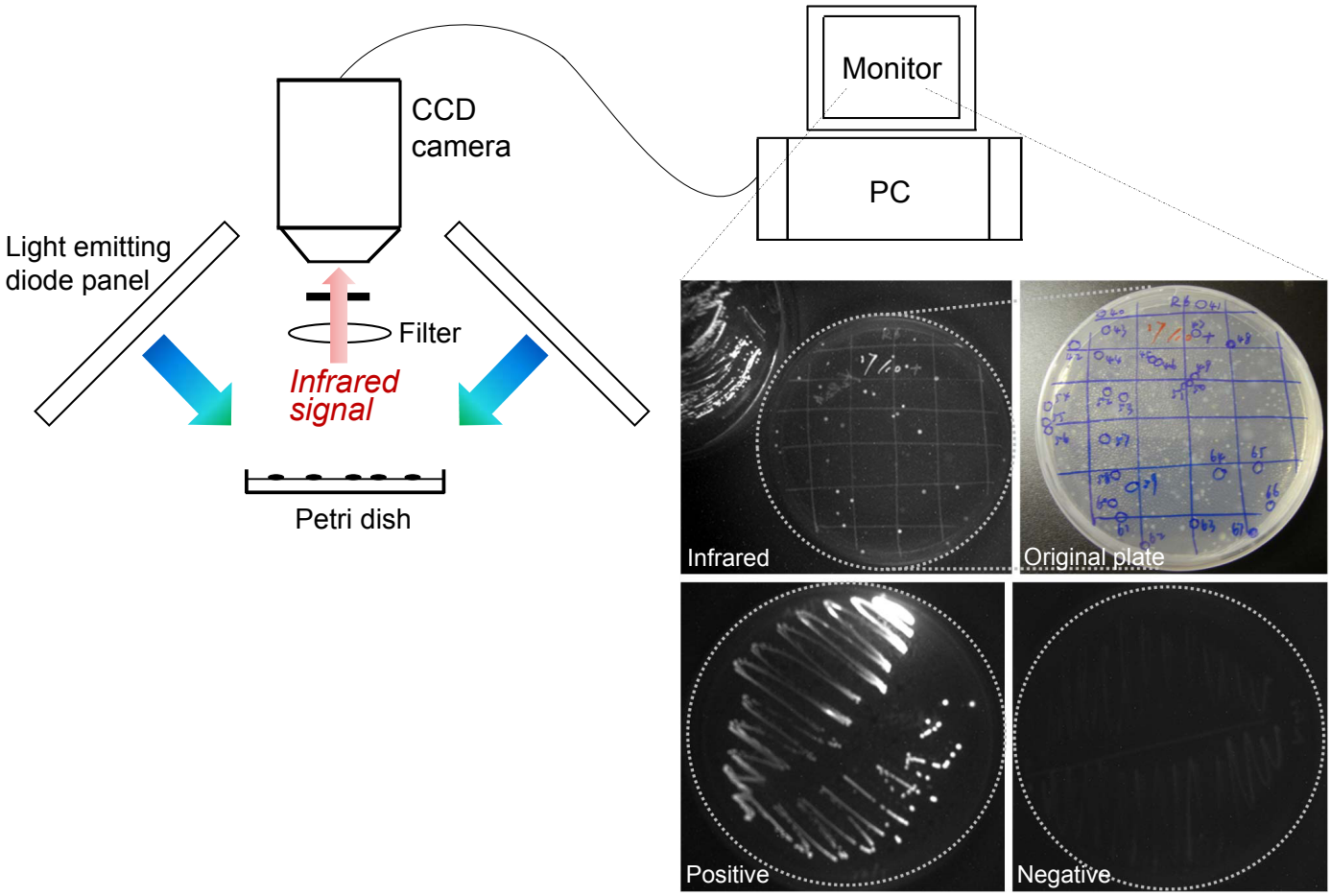

Fig. S1 A schematic diagram of the infra-red screening system. The principle of the detection system is based on the fact that light harvesting complexes of phototrophic organisms are excited in the blue-green spectral region (450 ~ 570 nm). The incoming light is captured by carotenoids, which instantly transfer the excitation to the bacteriochlorophyll molecules which release part of the absorbed energy in the form of infrared fluorescence. The BChl *a* autofluorescence is registered by a highly sensitive CCD camera to identify the positive colonies. Test agar plates are shown: *upper right*, a fresh ½ R_2A agar plate with a number of colonies formed; *upper left*, imaging of the signals passing the infrared filter and captured by CCD camera after this plate was illuminated with either blue or green light; *lower right* and *left*, depict images of a BChl *a*-positive and a BChl *a*-negative strain, respectively.

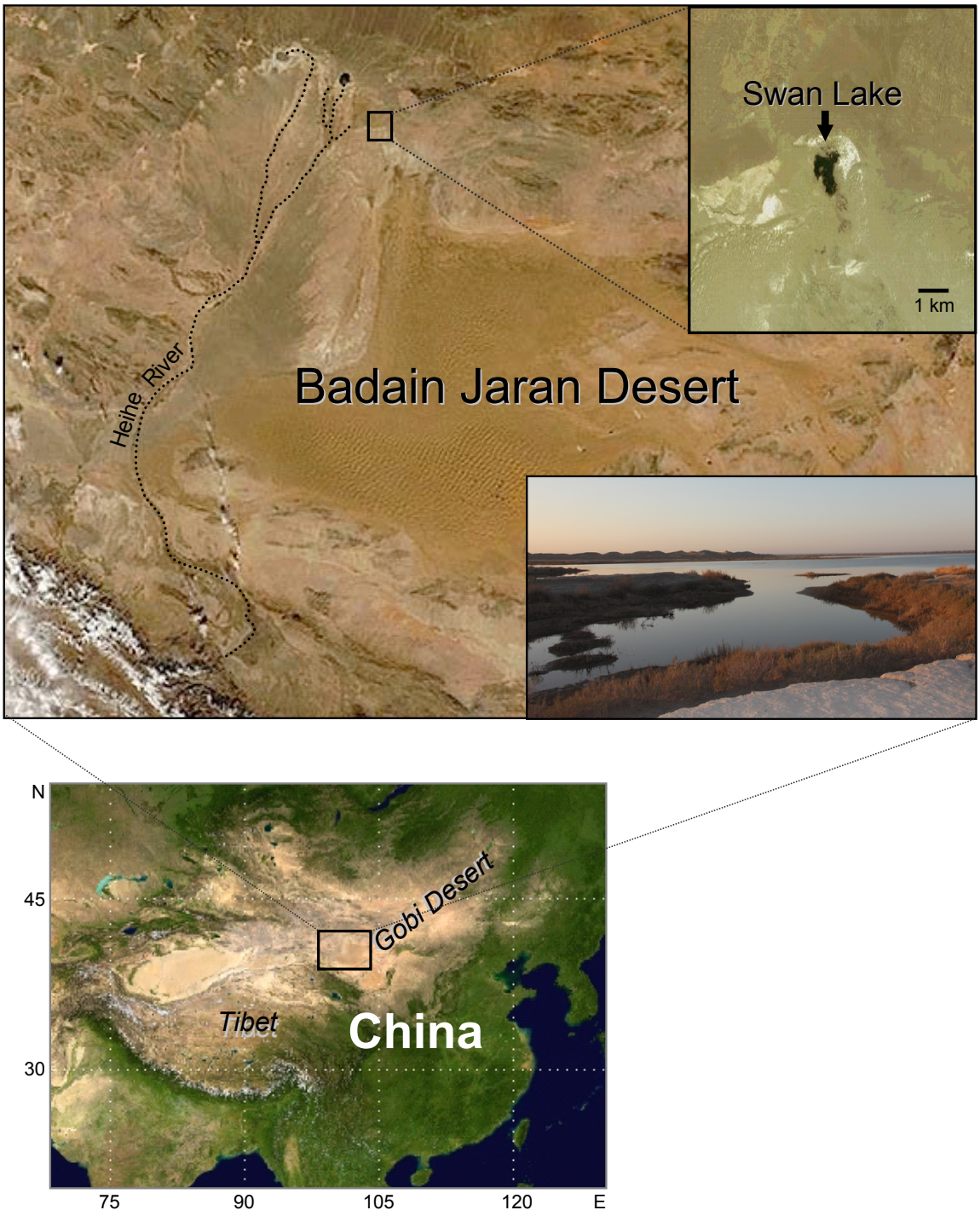

Fig. S2 Satellite images (from Google Earth) and a photograph of the Swan Lake in northern China.

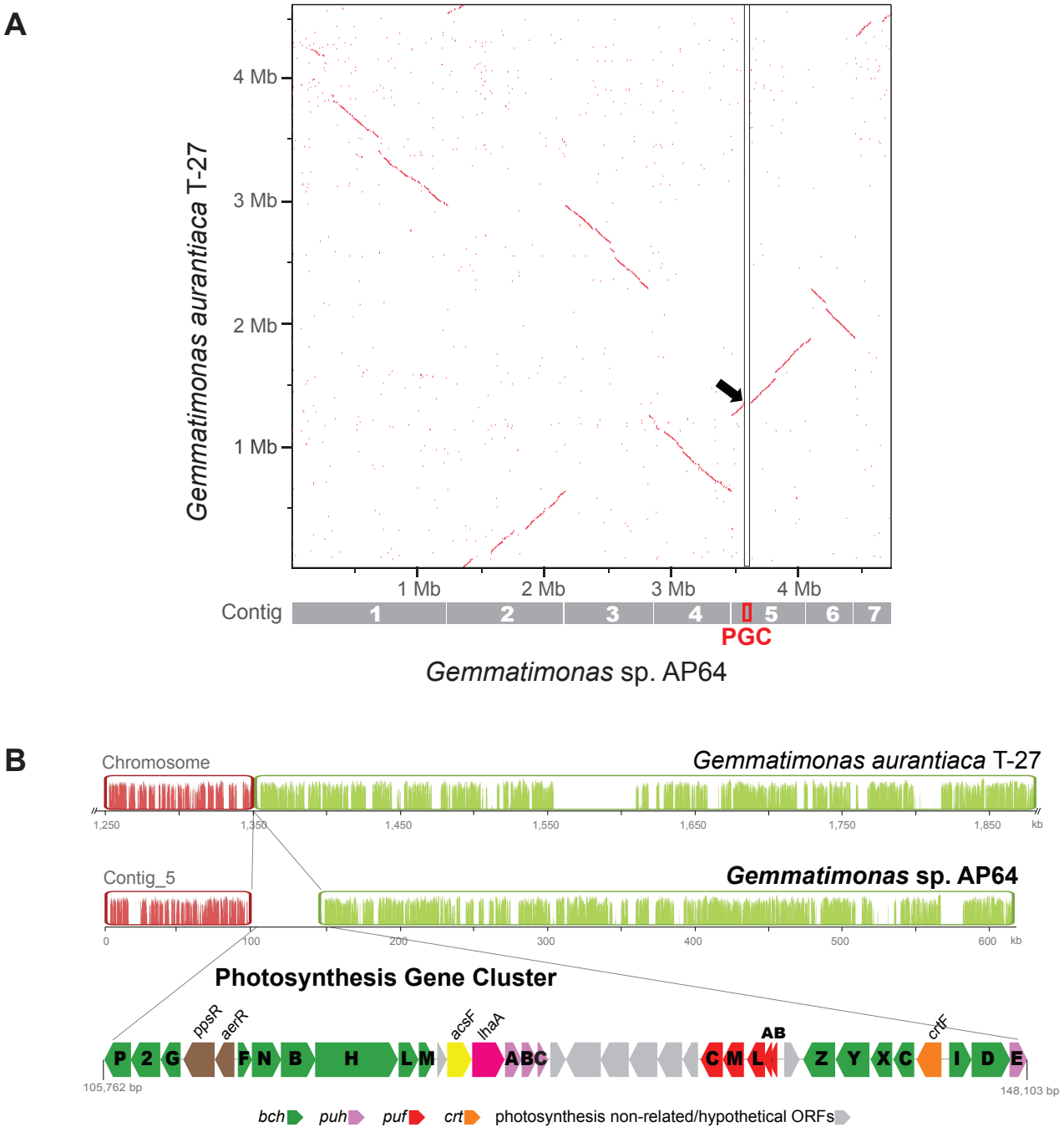

Fig. S3 (A) Genome blast dot plot for strain AP64 and *Gemmatimonas aurantiaca* T-27. The genome size is 4,706,869 bases for AP64 and 4,636,964 bases for T-27. Protein sequences from both genomes were compared by blastP against one another on the online RAST server and data were visualized through the SEED viewer. AP64 contigs are ordered according to their length (from the longest to the shortest). Highlighted region on the Contig_5 is the photosynthesis gene cluster (PGC) which corresponds to the gap region on the plot marked by an arrow. The blast dots for PGC proteins are shown in dashed box. **(B) Alignment of the PGC-containing contig of AP64 with the chromosome sequence of *Gemmatimonas aurantiaca* T-27** (NCBI accession number: NC_012489.1). Colored blocks represent syntenic regions that are presumably homologous and internally free from genomic rearrangement. Open areas within and between boxes represent the unique regions of the two organisms. Inside each box a similarity profile of the genome sequences is shown with the height corresponding to the average level of conservation in that region. *bch*, bacteriochlorophyll biosynthesis genes; *puh*, genes encoding reaction center assembly proteins; *puf*, genes encoding reaction center proteins; *crt*, carotenoid biosynthesis genes; *ppsR*, transcriptional repressor for photosynthesis genes; *aerR*, coenzyme B12-binding aerobic repressor; *acsF*, gene encoding oxygen dependent Mg-protoporphyrin IX monomethylester cyclase; *lhaA*, gene encoding light harvesting I complex assembly protein.

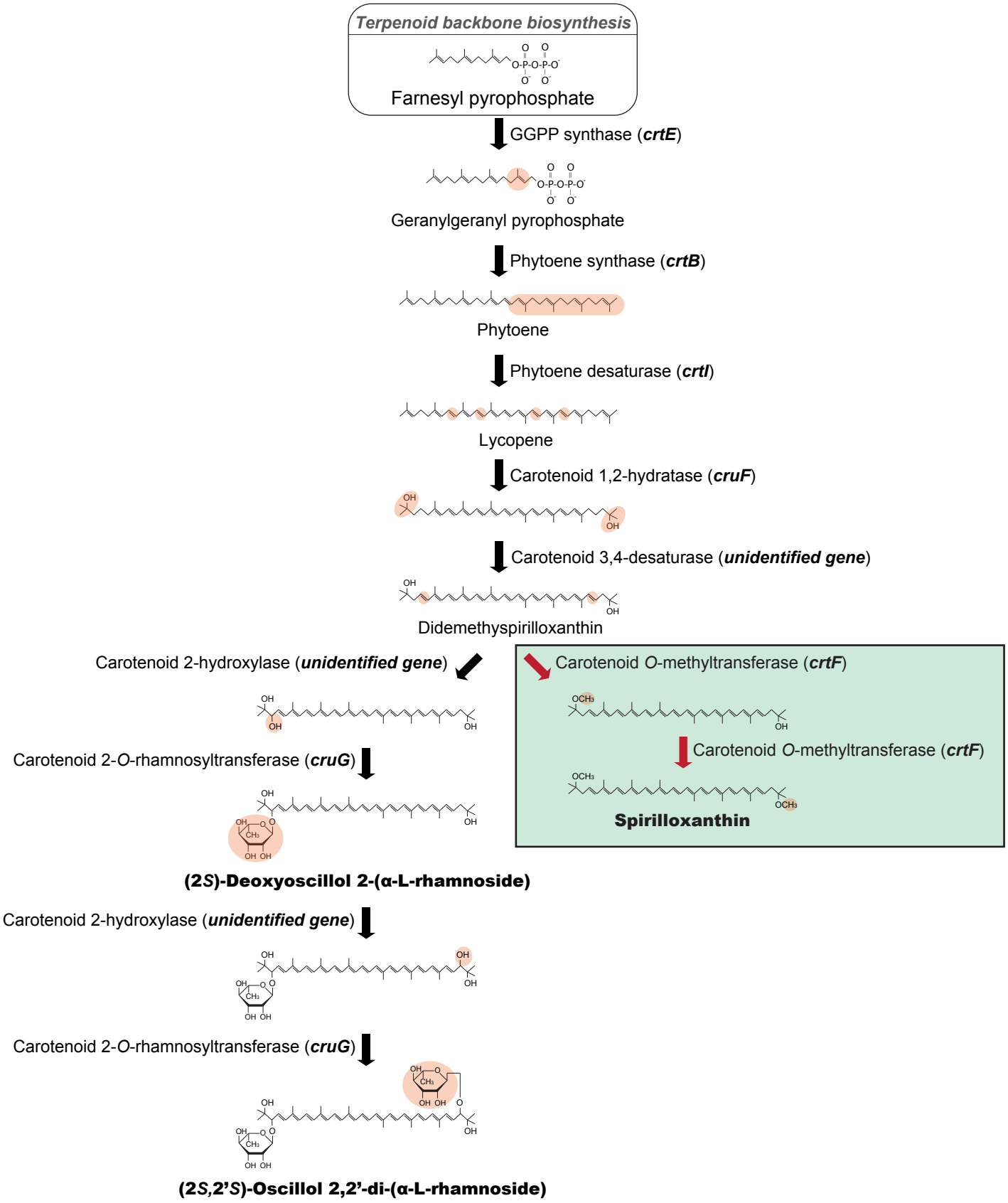

Fig. S4 Putative carotenoid biosynthesis pathway in AP64. The highlighted parts of each compound represent the modification sites of corresponding enzymes. The novel oscillol 2,2'-dirhamnoside biosynthesis pathway which contain two as yet unidentified genes was firstly proposed in *Gemmatimonas aurantiaca* T-27 by Takaichi *et al.* (Microbiolgy 2010, 156:757-763). Due to the very similar pigment (Fig. 3) and genome (Fig. S3) composition between strains AP64 and T-27, this pathway very likely exists in AP64 as well. In the highlighted box shows the proposed spirilloxanthin biosynthesis steps in AP64 catalyzed by the carotenoid *O*-methyltransferase (*crtF*) which is absent in T-27. For all carotenoid biosynthesis genes identified in the AP64 genome, see Table S6.

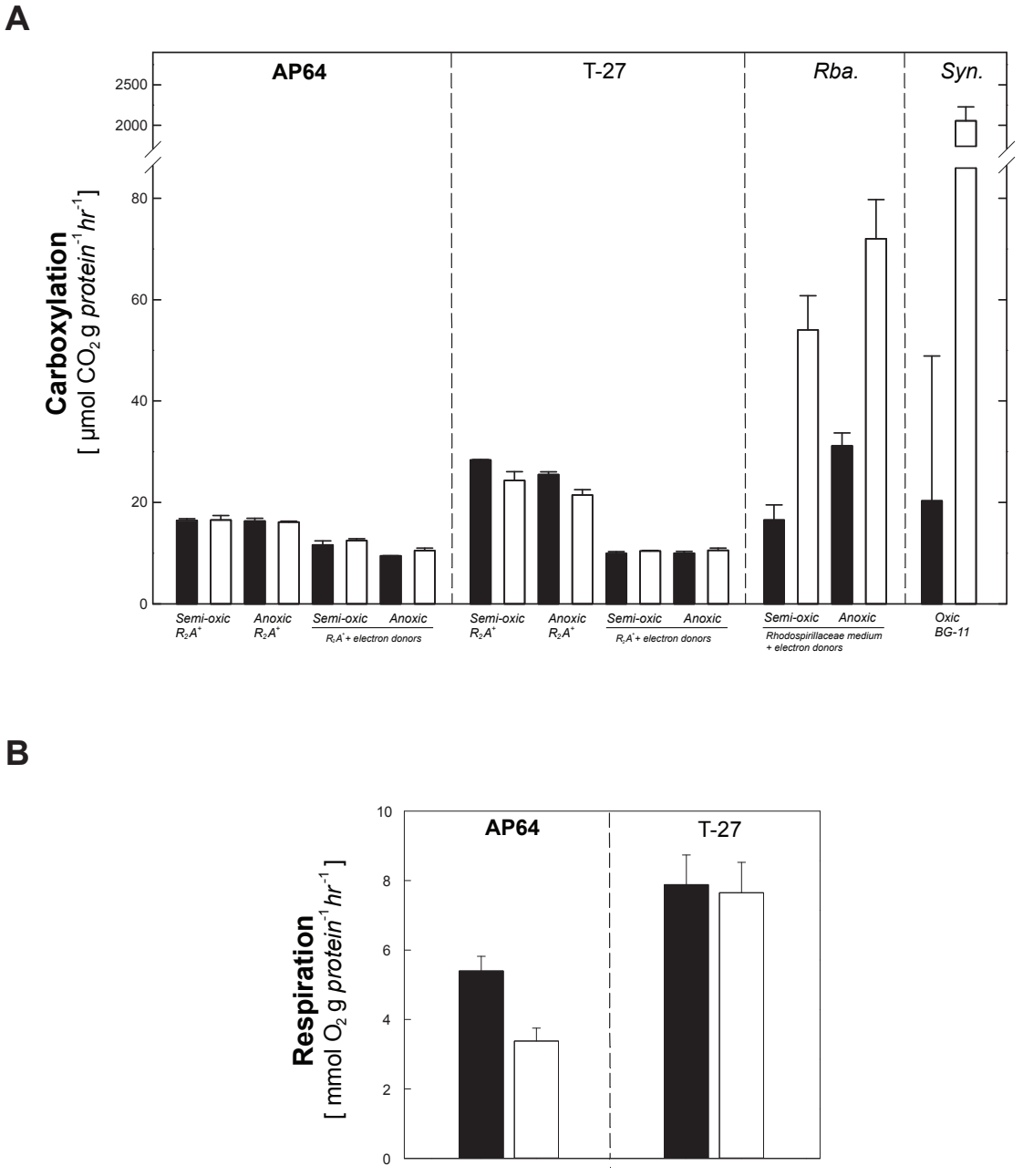

Fig. S5 Carbon incorporation and respiration activity in AP64 cells. (**A**),[14]C labeled bicarbonate incorporation determined in AP64 cells and in various control strains. *G. aurantiaca* T-27 was used as a heterotrophic control while strains *Rhodobacter sphaeroides* DSM158 (*Rba*.) and *Synechocystis* sp. PCC6803 (*Syn*.) were employed as photoautrophic controls. The tested growth environments were oxic for *Synechocystis* sp. PCC6803 and semi-oxic (~120 µM oxygen)/anoxic (>30 min bubbling of N_2, < 1 µM oxygen) for AP64, T-27 and *Rhodobacter sphaeroides* with or without electron donors supplemented in each specific medium which are shown on the graph. Media composition is given in Table S1. *Synechocystis* sp. PCC6803 was grown in the BG-11 medium. Each medium was supplemented with 3 mM sodium bicarbonate as a source of inorganic carbon. Electron donors are 1 mM Na_2S + 1 mM $Na_2S_2O_3$ + 1mM Na_2SO_3 for growth of AP64 and T-27 and 10 mM Na_2S + 1 mM $Na_2S_2O_3$ for *Rba*. Error bars represent st. dev. calculated from three replicates. (**B**), respiration activities in AP64 (left) and in T-27 (right). Black bars, in the dark; white bars, in the light (150 µmol photon m^2 s^{-1}). For more details see Materials and Methods.

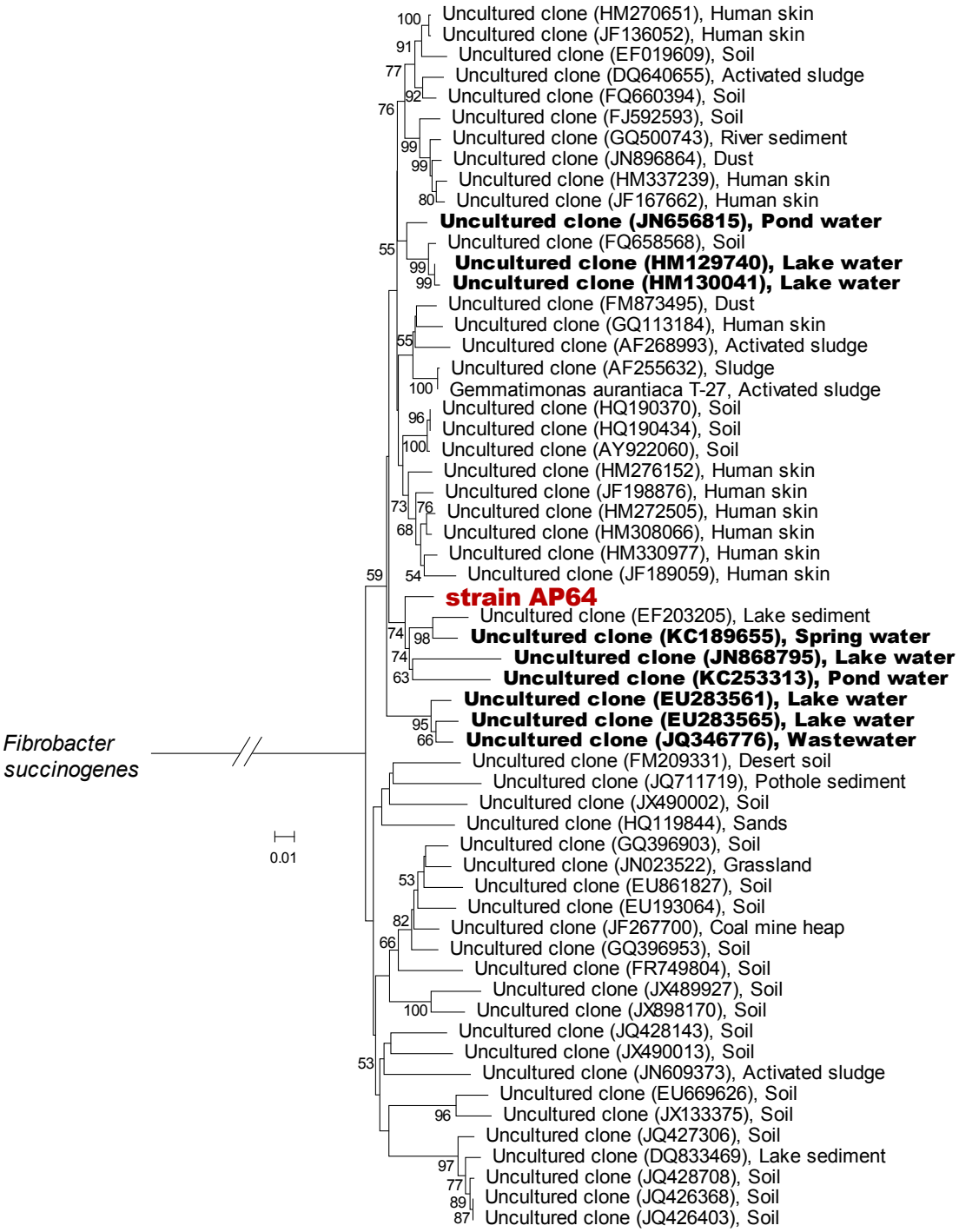

Fig. S6 A neighbor joining tree showing the relationship of AP64's 16S rRNA gene with environmental clones. Reference sequences and associated source information was retrieved from the NCBI GenBank database with the criteria set as length greater than 1,200 bp and sequence identity not less than 95%. The 16S rRNA gene sequence of *Fibrobacter succinogenes* was used as an outgroup. Aquatic environments related clones are highlighted in bold.

Table S1 List of media (A) and growth conditions (B) tested for strain AP64. For all media, pH was set to 8.0 because the source lake of this strain is alkaline (pH 8.76) and a test of pH 7.5 ~ 8.5 showed no significant influncence on AP64's growth. No growth was observed in a one month incubation with any of these media in liquid, and therefore AP64's growth was compared based on the amount and size of colonies formed on three-week incubated agar plates at 25 °C. -, no visible colonies formed; + to +++++, a few small colonies (diam. <0.5 mm) to a large number of normal sized colonies (diam. 1.0~2.0 mm). Each measurement was based on at least three plates. A light/dark cycle of 12 hrs/12 hrs was used with a light intensity of ~ 150 quanta m^{-2} s^{-1}.

(A)

Rhodospirillaceae medium* (Phototrophic growth by purple non sulfur bacteria)		**R_2A medium**# (Heterotrophic growth by freshwater bacteria)	1/2 strength	Full strength	2x strength	**R_2A^+ medium**	
Yeast extract	0.3 g	Yeast extract	0.25 g	0.5 g	1.0 g	Yeast extract	0.5 g
Na_2-succinate	1.0 g	Proteose peptone	0.25 g	0.5 g	1.0 g	Proteose peptone	0.5 g
NH_4-acetate	0.5 g	Glucose	0.25 g	0.5 g	1.0 g	Glucose	0.5 g
Fe(III) citrate solution (0.1% in H_2O)	5.0 mL	Soluble starch	0.25 g	0.5 g	1.0 g	Soluble starch	0.5 g
KH_2PO_4	0.5 g	Na-pyruvate	0.15 g	0.3 g	0.6 g	Na-pyruvate	0.3 g
$MgSO_4 \cdot 7H_2O$	0.4 g	K_2HPO_4	0.15 g	0.3 g	0.6 g	K_2HPO_4	0.3 g
NaCl	0.4 g	$MgSO_4 \cdot 7H_2O$	0.025 g	0.05 g	0.1 g	$MgSO_4 \cdot 7H_2O$	0.05 g
NH_4Cl	0.4 g	Distilled water	1.0 L	1.0 L	1.0 L	NH_4-acetate	0.05 g
$Ca_2Cl \cdot 2H_2O$	0.05 g					Vitamin solution (refer to DSMZ medium 462)	1.0 mL
Vitamin B12 solution (10 mg in 100 mL H_2O)	0.4 mL					Trace element solution SL-6 (refer to DSMZ medium 27)	0.05 mL
Trace element solution SL-6 (refer to DSMZ medium 27)	1.0 mL					Distilled water	1.0 L
Distilled water	1.0 L						

* formula same as the DSMZ medium 27 used for phototrophic growth of *Rhodobacter capsulatus, Rhodobacter sphaeroides, Rhodopseudomonas palustris, Rhodospirillum rubrum* and *Rubrivivax gelatinosus.*

\# based on the low nutrient DSMZ medium 830 used for slowly growing heterotrophic freshwater bacteria.

(B)

Growth medium (agar plates)		**Anoxic** (<0.2% O_2 atmosphere)		**Semi-oxic** (~10% O_2 atmosphere)		**Oxic** (normal atmosphere)	
		Dark	Light/Dark cycle	Dark	Light/Dark cycle	Dark	Light/Dark cycle
1. Rhodospirillaceae medium		-	-	-	-	-	-
2. supplemented with 10 mM $NaHCO_3$ and electron donors*		-	-	-	-	-	-
3. supplemented with organic carbon sources#		-	-	+++	+++	+	+
4. R_2A medium	1/2 strength	-	-	+++	+++	+	+
5.	Full strength	-	-	++++	++++	+	+
6.	2x strength	-	-	+	+	-	-
7. R_2A^+ medium		-	-	+++++	+++++	+	+
8. without organic carbon sources#		-	-	-	-	-	-
9. without main carbon sources but supplemented with 10 mM $NaHCO_3$ and eletron donors*		-	-	-	-	-	-

* electron donors: 10 mM Na_2S + 5 mM $Na_2S_2O_3$ + 5 mM Na_2SO_3. $Na_2S_2O_3$ and Na_2SO_3 were used because a sulfite dehydrogenase and a thiosulfate reductase were identified on Contig_1 (171948 to 173767) and on Contig_7 (226357 to 225716), respectively.

\# organic carbon sources: Glucose (0.5 g/L) + Proteose peptone (0.5 g/L)

Table S2 Next generation sequencing (NGS) reads and assembly statistics.

NGS Platform		
Illumina Hiseq 2000	PE reads number	12,655,514 x 2
	PE insert length	300 bp
	Read length	101 bp
	Total bases	2,556,413,828
454 GS-FLX plus	Total reads number	271,457
	average length	786.91
	Total bases	213,611,616
Final Assembly	Assembly size	4,706,869 bp
	Number of contigs (>1.5 kb)	7
	Largest contig	1,223,592 bp
	Smallest contig	287,047 bp
	N50	654,361 bp
	G+C content	64.4%
	Number of predicted genes	4,061
	5S rRNA	1
	16S rRNA	1
	23S rRNA	1
	tRNA genes	47
Contig Length	Contig_1	1,223,592 bp
	Contig_2	919,971 bp
	Contig_3	654,361 bp
	Contig_4	651,628 bp
	Contig_5	621,492 bp
	Contig_6	348,778 bp
	Contig_7	287,047 bp

Table S3 Top BLAST hits of 43 selected housekeeping genes from AP64 genome. For rRNA genes, the blastn algorithm was used; for protein encoding genes, the blastp algorithm was used.

Category	Gene encoding	Gene symbol	Located contig	Start Position	Stop Position	Gene Length (bp)	Organism	BLAST hit with the highest score			
								(sub-)Phylum	Score	E-value	Identity
rRNA	5S rRNA		Contig_6	5145	5263	119	*G. aurantiaca* T-27	Gemmatimonadetes	198	6.00E-48	97%
	16S rRNA	ssu	Contig_6	209	1,733	1,525	*G. aurantiaca* T-27	Gemmatimonadetes	2490	0	96%
	23S rRNA	lsu	Contig_6	2,053	4,995	2,943	*G. aurantiaca* T-27	Gemmatimonadetes	4726	0	96%
DNA replication	DNA gyrase subunit A	gyrA, copy1	Contig_1	1,030,210	1,032,642	2,433	*G. aurantiaca* T-27	Gemmatimonadetes	808	0	50%
		gyrA, copy2	Contig_2	859,102	861,657	2,556	*G. aurantiaca* T-27	Gemmatimonadetes	1451	0	88%
	DNA gyrase subunit B	gyrB	Contig_4	265,862	267,826	1,965	*G. aurantiaca* T-27	Gemmatimonadetes	1168	0	89%
Transcription	RNA polymerase sigma factor D	rpoD, copy1	Contig_4	394490	393603	888	*G. aurantiaca* T-27	Gemmatimonadetes	568	0	100%
		rpoD, copy2	Contig_4	523717	524625	909	*G. aurantiaca* T-27	Gemmatimonadetes	571	0	97%
		rpoD, copy3	Contig_5	194311	195171	861	*G. aurantiaca* T-27	Gemmatimonadetes	522	0	94%
	RNA polymerase sigma factor E	rpoE, copy1	Contig_1	941412	940897	516	*G. aurantiaca* T-27	Gemmatimonadetes	257	2.00E-84	72%
		rpoE, copy2	Contig_2	98375	97731	645	*G. aurantiaca* T-27	Gemmatimonadetes	371	3.00E-128	94%
		rpoE, copy3	Contig_2	175990	175487	504	*G. aurantiaca* T-27	Gemmatimonadetes	226	7.00E-72	67%
	Transcription elongation factor GreA		Contig_1	736360	736845	486	*G. aurantiaca* T-27	Gemmatimonadetes	267	8.00E-89	87%
Translation	Translation initiation factor 1	infA, copy1	Contig_1	449757	449554	204	*G. aurantiaca* T-27	Gemmatimonadetes	124	4.00E-35	90%
		infA, copy2	Contig_3	467349	467570	222	*Magnetospirillum* sp. SO-1	Alphaproteobacteria	110	6.00E-30	74%
	Translation initiation factor 2	infB	Contig_5	603786	606794	3009	*G. aurantiaca* T-27	Gemmatimonadetes	1539	0	80%
	Translation initiation factor 3	infC	Contig_6	97949	98404	456	*G. aurantiaca* T-27	Gemmatimonadetes	285	1.00E-95	95%
	Translation elongation factor G	-	Contig_2	832701	834797	2097	*G. aurantiaca* T-27	Gemmatimonadetes	1234	0	86%
	Translation elongation factor LepA	-	Contig_3	418061	416265	1797	*G. aurantiaca* T-27	Gemmatimonadetes	1124	0	95%
	Translation elongation factor P	-	Contig_3	475036	475569	534	*G. aurantiaca* T-27	Gemmatimonadetes	347	1.00E-119	94%
	Translation elongation factor Tu	-	Contig_4	206457	205255	1203	*G. aurantiaca* T-27	Gemmatimonadetes	768	0	96%
	Translation elongation factor G	-	Contig_4	208620	206500	2121	*G. aurantiaca* T-27	Gemmatimonadetes	1355	0	96%
	Translation elongation factor Ts	-	Contig_6	139582	140463	882	*G. aurantiaca* T-27	Gemmatimonadetes	495	2.00E-174	85%
	Aspartyl-tRNA synthetase	-	Contig_6	200702	202522	1821	*G. aurantiaca* T-27	Gemmatimonadetes	1057	0	88%
	Leucyl-tRNA synthetase	-	Contig_2	889433	891865	2433	*G. aurantiaca* T-27	Gemmatimonadetes	1438	0	85%
	Tryptophanyl-tRNA synthetase	-	Contig_4	188736	187684	1053	*G. aurantiaca* T-27	Gemmatimonadetes	569	0	86%
	Tyrosyl-tRNA synthetase	-	Contig_4	468195	469481	1287	*G. aurantiaca* T-27	Gemmatimonadetes	648	0	77%
	Valyl-tRNA synthetase	-	Contig_5	324199	321539	2661	*G. aurantiaca* T-27	Gemmatimonadetes	1571	0	86%
DNA repair	RecA protein	recA, copy1	Contig_1	937148	935892	1257	*G. aurantiaca* T-27	Gemmatimonadetes	302	8.00E-96	53%
		recA, copy2	Contig_6	57395	58471	1077	*G. aurantiaca* T-27	Gemmatimonadetes	617	0	92%
	ATP-dependent DNA helicase	recQ, copy1	Contig_4	72400	73965	1566	*G. aurantiaca* T-27	Gemmatimonadetes	873	0	85%
		recQ, copy2	Contig_6	217301	219097	1797	*G. aurantiaca* T-27	Gemmatimonadetes	790	0	67%
	DNA recombination and repair protein	recF	Contig_4	264250	265380	1131	*G. aurantiaca* T-27	Gemmatimonadetes	568	0	78%
	Recombination protein	recR	Contig_5	84314	84904	591	*G. aurantiaca* T-27	Gemmatimonadetes	322	2.00E-109	83%
	DNA recombination and repair protein	recO	Contig_5	516323	517045	723	*G. aurantiaca* T-27	Gemmatimonadetes	325	9.00E-109	67%
	Regulatory protein	recX	Contig_6	58587	59243	657	*G. aurantiaca* T-27	Gemmatimonadetes	210	3.00E-65	63%
	Single stranded DNA specific exonuclease	recJ	Contig_6	87719	89404	1686	*G. aurantiaca* T-27	Gemmatimonadetes	842	0	77%
	DNA repair protein	recN	Contig_6	111305	113047	1743	*G. aurantiaca* T-27	Gemmatimonadetes	761	0	70%
	Excinuclease ABC subunit A	uvrA, copy1	Contig_3	389193	386353	2841	*G. aurantiaca* T-27	Gemmatimonadetes	1612	0	84%
		uvrA, copy2	Contig_4	249824	246936	2889	*G. aurantiaca* T-27	Gemmatimonadetes	1737	0	90%
	Excinuclease ABC subunit B	uvrB	Contig_6	12496	14595	2100	*G. aurantiaca* T-27	Gemmatimonadetes	1258	0	92%
	Excinuclease ABC subunit C	uvrC, copy1	Contig_6	319342	317354	1989	*G. aurantiaca* T-27	Gemmatimonadetes	1035	0	82%
		uvrC, copy2	Contig_3	154524	153349	1176	*G. aurantiaca* T-27	Gemmatimonadetes	595	0	85%
		uvrC, copy3	Contig_7	232926	233978	1053	*G. aurantiaca* T-27	Gemmatimonadetes	540	0	78%
ATP production	ATP synthase A chain	atpA	Contig_3	79947	80903	957	*G. aurantiaca* T-27	Gemmatimonadetes	466	9.00E-162	74%
	ATP synthase C chain	atpC	Contig_3	81077	81355	279	*G. aurantiaca* T-27	Gemmatimonadetes	169	2.00E-52	96%
	ATP synthase delta chain	atpD	Contig_3	82018	82575	558	*G. aurantiaca* T-27	Gemmatimonadetes	279	1.00E-92	78%
	ATP synthase epsilon chain	atpE	Contig_4	387254	386994	261	*G. aurantiaca* T-27	Gemmatimonadetes	151	2.00E-45	90%
	ATP synthase beta chain	atpB	Contig_4	388688	387258	1431	*G. aurantiaca* T-27	Gemmatimonadetes	894	0	96%
	ATP synthase gamma chain	atpG	Contig_4	389587	388715	873	*G. aurantiaca* T-27	Gemmatimonadetes	442	2.00E-153	77%
	ATP synthase alpha chain	atpA	Contig_4	391170	389599	1572	*G. aurantiaca* T-27	Gemmatimonadetes	960	0	94%
others	Chaperone protein DnaK	dnaK	Contig_3	241319	239382	1938	*G. aurantiaca* T-27	Gemmatimonadetes	1164	0	92%
	Chaperone protein DnaJ	dnaJ, copy1	Contig_5	89990	91135	1146	*G. aurantiaca* T-27	Gemmatimonadetes	658	0	84%
		danJ, copy2	Contig_6	78359	79492	1134	*G. aurantiaca* T-27	Gemmatimonadetes	690	0	88%
	Heat shock protein, Hsp20 family	-	Contig_3	596899	596441	459	*G. aurantiaca* T-27	Gemmatimonadetes	119	7.00E-31	45%

Table S4 Top blastp hits of proteins annotated in the photosynthesis gene cluster of AP64. Genes are listed in the same order as their positions in the PGC.

Gene function	Gene symbol	Located contig	Start position	Stop position	Gene Length (bp)	BLASTP hit with the highest score				
						Organism	(sub-)Phylum	Score	E-value	Identity
Geranylgeranyl hydrogenase	bchP	Contig_5	107030	105762	1269	*Bradyrhizobium sp.* STM 3809	Alphaproteobacteria	533	0	66%
Bacteriochlorophyll synthase 44.5 kDa chain	bch2	Contig_5	108385	107027	1359	*Rubrivivax benzoatilyticus*	Betaproteobacteria	392	5.00E-129	50%
Bacteriochlorophyll a synthase	bchG	Contig_5	109296	108385	912	*Rubrivivax gelatinosus* IL144	Betaproteobacteria	371	2.00E-125	62%
Transcriptional regulator	ppsR	Contig_5	110882	109437	1446	*Ectothiorhodospira sp.* PHS-1	Gammaproteobacteria	442	1.00E-147	49%
Heme-binding SCHIC domain protein, putative oxygen sensor	aerR	Contig_5	111723	110848	876	*Rhodospirillum centenum* SW	Alphaproteobacteria	184	3.00E-53	43%
2-vinyl bacteriochlorophyllide hydratase	bchF	Contig_5	111998	112561	564	*Rhodomicrobium vannielii* ATCC 17100	Alphaproteobacteria	203	6.00E-63	67%
Light-independent protochlorophyllide reductase subunit N	bchN	Contig_5	112558	113811	1254	*Pseudanabaena biceps*	Cyanobacteria	645	0	75%
Light-independent protochlorophyllide reductase subunit B	bchB	Contig_5	113816	115558	1743	*Pseudanabaena biceps*	Cyanobacteria	758	0	67%
Protoporphyrin IX Mg-chelatase subunit H	bchH	Contig_5	115575	119375	3801	*Rhodospirillum centenum* SW	Alphaproteobacteria	1387	0	56%
Light-independent protochlorophyllide reductase iron-sulfur ATP-binding protein	bchL	Contig_5	119434	120321	888	*Synechococcus* sp. CB0205	Cyanobacteria	508	3.00E-179	87%
Mg-protoporphyrin O-methyltransferase	bchM	Contig_5	120402	121121	720	*Rubrivivax gelatinosus* IL144	Betaproteobacteria	278	1.00E-90	64%
hypothetical protein	-	Contig_5	121152	121484	333	*Rhodovulum sp.* PH10	Alphaproteobacteria	91.3	2.00E-21	49%
Mg protoporphyrin IX monomethyl ester oxidative cyclase (aerobic)	acsF	Contig_5	121530	122582	1053	*Methylobacterium* sp. 88A	Alphaproteobacteria	430	9.00E-147	57%
light-harvesting 1 complex assembly protein	lhaA	Contig_5	122611	124056	1446	*Thiorhodospira sibirica*	Gammaproteobacteria	488	2.00E-165	54%
Photosynthetic reaction center subunit	puhA	Contig_5	124086	124289	204	*Thermochromatium tepidum*	Gammaproteobacteria	76.3	2.00E-15	58%
Photosynthetic reaction center subunit	puhB	Contig_5	124289	124834	546	*Thiocapsa marina*	Gammaproteobacteria	186	2.00E-55	50%
Photosynthetic complex assembly protein	puhC	Contig_5	124837	125511	675	*Methylobacterium* sp. MB200	Alphaproteobacteria	156	2.00E-43	40%
hypothetical protein	-	Contig_5	125514	126059	546	*Rhodopseudomonas palustris* BisB18	Alphaproteobacteria	103	1.00E-24	36%
hypothetical protein	-	Contig_5	127958	126144	1815	*Mycobacterium tuberculosis*	Actinobacteria	115	1.00E-09	28%
hypothetical protein	-	Contig_5	129306	127969	1338	*Conexibacter woesei* DSM 14684	Actinobacteria	102	2.00E-07	28%
Putative kinase protein	-	Contig_5	130777	129356	1422	*Arthrospira* sp. PCC 8005	Cyanobacteria	340	5.00E-108	41%
hypothetical protein	-	Contig_5	131881	130784	1098	*Desulfotomaculum acetoxidans* DSM 771	Firmicutes	97.8	2.00E-19	28%
hypothetical protein	-	Contig_5	132526	131885	642	*Arthrospira platensis* NIES-39	Cyanobacteria	233	9.00E-74	53%
Photosynthetic reaction center cytochrome c subunit	pufC	Contig_5	133815	132718	1098	*Limnohabitans* sp. Rim47	Betaproteobacteria	281	3.00E-88	44%
Photosynthetic reaction center M subunit	pufM	Contig_5	134915	133812	1104	*Luminiphilus syltensis*	Gammaproteobacteria	495	2.00E-172	71%
Photosynthetic reaction center L subunit	pufL	Contig_5	135752	134928	825	*Lamprocystis purpurea*	Gammaproteobacteria	494	1.00E-174	79%
light-harvesting LHI alpha subunit	pufA	Contig_5	136119	135922	198	*Thiocapsa roseopersicina*	Gammaproteobacteria	67.8	2.00E-13	51%
light-harvesting LHI beta subunit	pufB	Contig_5	136473	136339	135	*Limnohabitans* sp. Rim28	Betaproteobacteria	60.1	7.00E-11	56%
Predicted hydrolases or acyltransferases	-	Contig_5	136851	137522	672	*Elioraea tepidiphila*	Alphaproteobacteria	160	3.00E-44	44%
Chlorophyllide reductase subunit Z	bchZ	Contig_5	139086	137629	1458	*Rhodospirillum centenum* SW	Alphaproteobacteria	816	0	83%
Chlorophyllide reductase subunit Y	bchY	Contig_5	140723	139086	1638	*Rhodopseudomonas palustris* BisB18	Alphaproteobacteria	829	0	80%
Chlorophyllide reductase subunit X	bchX	Contig_5	141742	140747	996	*Limnohabitans* sp. Rim47	Betaproteobacteria	568	0	88%
2-desacetyl-2-hydroxyethyl bacteriochlorophyllide *a* dehydrogenase	bchC	Contig_5	142692	141739	954	*Thiorhodovibrio* sp. 970	Gammaproteobacteria	390	5.00E-132	60%
Hydroxyneurosporene methyltransferase	crtF	Contig_5	144109	142880	1230	*Methyloversatilis universalis*	Betaproteobacteria	362	1.00E-118	52%
Protoporphyrin IX Mg-chelatase subunit I	bchI	Contig_5	144318	145343	1026	*Rhodopseudomonas palustris* BisA53	Alphaproteobacteria	471	3.00E-163	71%
Protoporphyrin IX Mg-chelatase subunit D	bchD	Contig_5	145324	147201	1878	*Thioflavicoccus mobilis* 8321	Gammaproteobacteria	479	2.00E-158	47%
Photosynthetic complex assembly protein 2	puhE	Contig_5	147198	148103	906	*Fulvimarina pelagi*	Alphaproteobacteria	113	8.00E-25	28%

Table S5 A complete list of enzymes involved in the bacteriochlorophyll biosynthesis pathway predicted from AP64 genome with *Gemmatimonas aurantiaca* T-27 as comparison.

Step in the pathway	Enzyme	Gene symbol	Located contig	inside (●) / outside (○) PGC	Start position	Stop position	Gene length (bp)	Homolog presence in *G. aurantiaca* T-27, protein identity
1. L-Glutamate → L-Glutamyl-tRNA	Glutamyl-tRNA synthetase (EC 6.1.1.17)	gltX	Contig_5	○	303631	302168	1464	+, 78% (378/484)
2. → Glutamate-1-semialdehyde	Glutamyl-tRNA reductase (EC 1.2.1.70)	hemA	Contig_5	○	105736	104273	1464	+, 28% (82/291)
3. → 5-Amino-levulinate	Glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8)	hemL, copy1	Contig_3	○	572761	573999	1239	+, 83% (355/425)
		hemL, copy2	Contig_5	○	197160	198473	1314	+, 83% (355/425)
4. → Porphobilinogen	Porphobilinogen synthase (EC 4.2.1.24)	hemB	Contig_1	○	1127978	1128973	996	+, 82% (265/321)
5. → Hydroxymethyl-bilane	Hydroxymethylbilane synthase (EC 2.5.1.61)	hemC	Contig_1	○	1126983	1127981	999	+, 77% (238/308)
6. → Uroporphyrinogen III	Uroporphyrinogen III synthase (EC 4.2.1.75)	hemD	Contig_1	○	1126029	1126808	780	+, 74% (186/252)
7. → Coproporphyrinogen III	Uroporphyrinogen III decarboxylase (EC 4.1.1.37)	hemE	Contig_5	○	104158	103064	1095	+, 82% (290/351)
8. → Protoporphyrinogen IX	Coproporphyrinogen oxidase (***Aerobic,*** EC 1.3.3.3)	hemF	Contig_5	○	103067	102072	996	+, 78% (240/307)
	Coproporphyrinogen dehydrogenase (***Anaerobic,*** EC 1.3.99.22)	hemN	Contig_7	○	243281	244645	1365	+, 62% (275/442)
9. → Protoporphyrin IX	Protoporphyrinogen oxidase (***Aerobic,*** EC 1.3.3.4)	hemY	Contig_5	○	102062	100596	1467	-
	Protoporphyrinogen IX dehydrogenase (***Anaerobic,*** EC 1.3.5.3)	hemG	Contig_7	○	245270	245704	435	+, 71% (98/138)
10. → Mg-protoporphyrin IX	Magnesium chelatase subunit I (EC 6.6.1.1)	bchI, copy1	Contig_5	●	144318	145343	1026	+, 86% (418/482)
		bchI, copy2	Contig_4	○	543911	545365	1455	+, 86% (418/482)
	Magnesium chelatase subunit D (EC 6.6.1.1)	bchD	Contig_5	●	145324	147201	1878	-
	Magnesium chelatase subunit H (EC 6.6.1.1)	bchH	Contig_5	●	115575	119375	3801	-
11. → Mg-protoporphyrin IX 13-methyl ester	Mg-protoporphyrin IX methyltransferase (EC 2.1.1.11)	bchM	Contig_5	●	120402	121121	720	-
12. → Divinylprotochlorophyllide	Mg-protoporphyrin IX monomethyl ester oxidative cyclase (***Aerobic,*** EC 1.14.13.81)	acsF	Contig_5	●	121530	122582	1053	-
	Mg-protoporphyrin IX monomethyl ester oxidative cyclase (***Anaerobic***)	bchE	Contig_7	○	241690	243291	1602	-
13. → Protochlorophyllide	Divinyl chlorophyllide ***a*** 8-vinyl-reductase (EC 1.3.1.75)	bciA*	Contig_2	○	338640	339923	1284	-
14. → Chlorophyllide *a*	Protochlorophyllide reductase (Light-dependent, EC 1.3.1.33)	-	Contig_1	○	675051	674083	969	-
	Protochlorophyllide reductase iron-sulfur ATP-binding protein (Light-independent, EC 1.3.7.7)	bchL	Contig_5	●	119434	120321	888	-
	Protochlorophyllide reductase subunit N (Light-independent, EC 1.3.7.7)	bchN	Contig_5	●	112558	113811	1254	-
	Protochlorophyllide reductase subunit B (Light-independent, EC 1.3.7.7)	bchB	Contig_5	●	113816	115558	1743	-
15. → 3-Vinyl-bacteriochlorophyllide *a*	Chlorophyllide ***a*** reductase subunit X	bchX	Contig_5	●	141742	140747	996	-
	Chlorophyllide ***a*** reductase subunit Y	bchY	Contig_5	●	140723	139086	1638	-
	Chlorophyllide ***a*** reductase subunit Z	bchZ	Contig_5	●	139086	137629	1458	-
16. → 3-Hydroxyethyl-bacteriochlorophyllide *a*	3-vinyl bacteriochlorophyllide ***a*** hydratase	bchF	Contig_5	●	111998	112561	564	-
17. → Bacteriochlorophyllide *a*	3-Hydroxyethyl-bacteriochlorophyllide ***a*** dehydrogenase	bchC	Contig_5	●	142692	141739	954	-
18. → Bacteriochlorophyll *a* (esterified with geranylgeraniol)	Bacteriochlorophyll synthase (EC 2.5.1.62)	bchG	Contig_5	●	109296	108385	912	-
19. → Bacteriochlorophyll *a* (esterified with phytol)	Geranylgeranyl reductase	bchP	Contig_5	●	107030	105762	1269	-

* annotated as *bciA* rather than *bchJ* based on Saunders *et al.* Biochemistry (2013) 52, 8442-8451 and Canniffe *et al.* Biochem. J. (2013) 450, 397-405

Table S6 Carotenoid biosynthesis enzymes annotated in AP64 genome and their top blastp hits in GenBank.

Contig	Start position	Stop position	Gene length (bp)	Enzyme	Gene	inside (●) / outside (○) PGC	The top-score BlastP hit in GenBank database		
							Accession no.	Species	Protein identity
Contig_6	38622	39626	1005	Geranylgeranyl diphosphate synthase (EC 2.5.1.29)	*crtE*	○	YP_002761444	*G. aurantiaca* T-27	281/334 (84%)
Contig_2	841179	842276	1098	Phytoene synthase (EC 2.5.1.32), copy_1	*crtB*	○	YP_002759973	*G. aurantiaca* T-27	299/352 (85%)
Contig_3	189833	188781	1053	Phytoene synthase (EC 2.5.1.32), copy_2	*crtB*	○	YP_002761918	*G. aurantiaca* T-27	178/294 (61%)
Contig_3	187341	185827	1515	Phytoene dehydrogenase (EC 1.14.99.-)	*crtI*	○	YP_002761920	*G. aurantiaca* T-27	368/486 (76%)
Contig_3	188768	187806	963	Carotenoid 1,2-hydratase	*cruF*	○	YP_002761919	*G. aurantiaca* T-27	191/288 (66%)
Contig_5	403532	404725	1194	Carotenoid 2-*O*-rhamnosyltransferase	*cruG*	○	YP_002760890	*G. aurantiaca* T-27	251/382 (66%)
Contig_5	144109	142880	1230	Carotenoid *O*-methyltransferase	*crtF*	●	WP_020165693	*M. universalis*	194/379 (51%)

Table S7 Enzymes with CO_2 or HCO_3^- as substrate annotated in AP64 genome with *Gemmatimonnas aurantiaca* T-27 as comparison.

Contig	Start position	Stop position	Gene length (bp)	Enzyme	Reaction catalyzed	Reversibility of reaction	Homolog presence in *G. aurantiaca* T-27, protein identity
Contig_3	274975	274205	771	carbonic anhydrase (EC 4.2.1.1)	$CO_2+H_2O \rightleftharpoons HCO_3^-$	Yes	+, 191/249 (77%)
Contig_4	123359	122862	498	carbonic anhydrase (EC 4.2.1.1)			+, 124/160 (78%)
Contig_6	133104	134648	1545	acetyl-CoA carboxylase (EC 6.3.4.14)	acetyl-CoA+ATP+HCO_3^- $\rightleftharpoons$ malonyl-CoA+ADP+P_i	Yes	+, 443/511 (87%)
Contig_6	48243	49601	1359	biotin carboxylase (EC 6.3.4.14)	biotin-carboxyl-carrier protein+ATP+$CO_2 \rightleftharpoons$ carboxylbiotin-carboxyl-carrier protein+ADP+P_i	Yes	+, 432/452 (96%)
Contig_7	41425	38642	2784	phosphoenolpyruvate carboxylase (EC 4.1.1.31)	PEP+HCO_3^- +$H_2O \rightarrow$ oxaloacetate+ADP+P_i	No	+, 723/923 (78%)
Contig_6	136080	136592	513	propionyl-CoA carboxylase (EC 6.4.1.3)	propanoyl-CoA+ATP+HCO_3^- $\rightleftharpoons$ methylmalonyl-CoA+ADP+P_i	Yes	+, 132/170 (78%)
Contig_3	374162	372957	1206	phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21), ATPase subunit	5-aminoimidazole ribonucleotide+$CO_2 \rightleftharpoons$ 5'-phosphoribosyl-4-carboxy-5-aminoimidazole+$2H^+$	Yes	+, 326/394 (83%)
Contig_3	374686	374159	528	phosphoribosylaminoimidazole carboxylase (EC 4.1.1.21), catalytic subunit			+, 149/169 (88%)
Contig_6	243675	244787	1113	carbamoyl-phosphate synthase (EC 6.3.5.5), small chain	2ATP+NH_3+CO_2+$H_2O \rightarrow$ 2ADP+phosphate+carbamoyl-phosphate	No	+, 309/370 (84%)
Contig_3	641545	644772	3228	carbamoyl-phosphate synthase (EC 6.3.5.5), large chain			+, 967/1072 (90%)

Table S8 Gemmatimonadetes-AcsF like sequences in the metagenome database CAMERA and in the NCBI whole genome shotgun (WGS) sequence database. AP64's AcsF amino acid sequence was used as a query for tblastn searching. The threshold was set as >85% sequence identity.

BLAST hit	Discription of metagenome project	Sequence form	Total length (bp)	GenBank accession number	Query database	Score	E-value	Identity (amino acid positions)
Wastewater metagenome HPminus27240.1	Sequencing batch reactors (SBR) enriched microbial communities from a Danish wastwater treatment plant	Assembled contig	6892	APMI01036896	NCBI WGS	674	0	88% (318/362)
Yellowstone Lake southeast arm water metagenome CAM_READ_0305803473	Yellowstone Lake metagenome 355:Southeast Arm, YLAKE0355-4F-01-738 Experiment, size fraction 0.1-0.8 µm	454 read	521	SRA026894	CAMERA	346	3.52E-92	93% (162/173)
Yellowstone Lake southeast arm water metagenome CAM_READ_0306001467	s.a.a.	454 read	519	SRA026894	CAMERA	303	5.23E-83	94% (141/150)
Yellowstone Lake southeast arm water metagenome CAM_READ_0305923507	s.a.a.	454 read	520	SRA026894	CAMERA	300	8.00E-84	93% (140/149)
Yellowstone Lake southeast arm water metagenome CAM_READ_0307488505	Yellowstone Lake metagenome 355b:Southeast Arm, YLAKE0355BSM-4F-01-736 Experiment, size fraction 0.8-3 µm	454 read	532	SRA026894	CAMERA	260	6.24E-84	93% (121/129)
Yellowstone Lake southeast arm water metagenome CAM_READ_0314277619	Yellowstone Lake metagenome 350: Mary Bay Mixing, YLAKE350-4F-01-765 Experiment, size fraction 0.1-0.8 µm	454 read	506	SRA026871	CAMERA	286	9.40E-75	88% (135/152)