

Supplementary materials

Supplementary Table I: Target regions and real-time PCR results

PCR primers, primer read depth (calculated as the average read depth for bases comprising both forward and reverse primer) and real-time PCR results are listed for each selected contig. Also shown is the PCR efficiency, as determined by performing real-time PCRs on a 10-fold dilution series of template. Cp values refer to the 10^x diluted samples (used in figure 3). For *P. rubescens* contig 13983, three primer sets were designed, each in a region with a different local read depth. The best BLASTX hit for each contig is also shown.

P. rubescens

Contig	Contig read depth	Best blast hit	Primer names	Primer sequence	Read depth primers	Cp (1/10 dilution)	PCR efficiency
73	19	Alpha-glucosidase [Cyanotheca sp. PCC 7424]	c22F	TGGTGAAAATCCTGACTTACCG	18	16.78	1.762
			c22R	GCCCGCAATACCAATATTTGT			
88	20	6-phosphogluconate dehydrogenase [Microcystis aeruginosa NIES-843]	c159_F	TGGTGAAAATCCTGACTTACCG	27	16.58	1.80
			c159_R	GTTGGGCTTGGGTTAAATTCA			
170	40	gas vesicle protein GvpN [Arthrospira maxima CS-328]	c174_F	ATTGTGGAGTTTATGGGACTCAAG	34	15.72	1.80
			c174_R	TACCATTTTGAACGGGTTAATT			
72	57	Hyp. prot and gas vesicle protein GvpK [Nostoc punctiforme PCC 73102]	c17_F	CAATCCCGTATTCATTAACCG	56	15.54	1.74
			c17_R	GCGGCACAGAGTTTACAACAA			
13921	99	hypothetical protein MC7420_6431 [Microcoleus chthonoplastes PCC 7420]	c519_F	TATCCTGACTAAATGCACCGCT	139	13.76	1.76
			c519_R	AAACCAGAGAGAGTGGAAGCA			
13972	190	transposase IS4 family protein [Cyanotheca sp. PCC 8802]	98c285F	TTGATAGAACTCATTCAAGCTG	181	13.45	1.73
			98c285R	TCACTCACTCAAACCTGTAACC			
13462	77	O.agardhii 16S rRNA gene	CO_TBR	GCCTTTACGAGGTTAAGCCT	79	15.70	1.65
			CN_TBR	AAGGTTCTTGGATTGTCAACCC			
13983 region 1	56	ApnC (NRPS) [Planktothrix agardhii NIVA-CYA 126/8]	mic_rt_R	TTGAGCTATGGAAGATTGGGAA	55	15.92	1.81
			mic_rt_F	CCCAATATCAAGTGCCTCAAG			
13983 region 2	56	ApnC (NRPS) [Planktothrix agardhii NIVA-CYA 126/8]	c13900_mic_F_2x	ACTCCTCCTCAGCCTAGGGTTA	39	16.19	1.71
			c13900_mic_R_2x	ACAAACGGAATCCGACTGAGG			
13983 region 3	56	ApnC (NRPS) [Planktothrix agardhii NIVA-CYA 126/8]	c13900_mic_F_3x	CTTAGCGGATATTCAAAGCCTG	65	15.24	1.74

A. flos-aquae

Contig	Contig read depth	Best blast hit	Primer names	Primer sequence	Read depth primers	Cp (1/10 dilution)	PCR efficiency
525	161	hypothetical protein AM1_C0013 [Acaryochloris marina MBIC11017]	Apc1166F	GCATCTAAAGGCTATCATCAA	163	13.95	1.67
			Apc1166R	AATGAATCAATATGGCATCTG			
13349	22	hypothetical protein Npun_F0406 [Nostoc punctiforme PCC 73102]	Apc80F	ATTTTCAACGCCAACCTAC	17	19.76	2.47
			Apc80R	AATATCATTGTGGGGAATTG			
468	45	nad-dependent epimerase/dehydratase [Phaeodactylum tricornutum CCAP 1055/1]	Apc353F	ATCAACATCAGCAGACTCG	38	18.69	1.45
			Apc353R	CGATGATGTTGTTATGAGA			
12978	80	photosystem II reaction centre protein PsbA/D1 [Anabaena variabilis ATCC 29413]	Apc1337F	GTTCTTGTCTCCGCAAT	87	18.66	1.78
			Apc1337R	AAGTG TAGTTGACGGCTGTT			
214	157	conserved hypothetical protein [Microscilla marina ATCC 23134]	Apc680F	TGCGTTATTAATTCATTTCC	152	14.02	1.84

Supplementary Table II: Annotation of the high read depths contigs for the *E. coli* and *P. gingivalis* assemblies

For each contig, the read depth, number of BLAST hits (BLASTN program) to the reference genome (maximum E value 10^{-16} , 97% or more identity at the nucleotide level), estimate of copy number ('Est. copy number') with upper and lower Confidence interval Limits (CL) are shown. In addition, the annotated features in the region of the BLAST hits are shown. Concurrent annotated features are separated by comma's. For contigs that had different annotations between the different BLAST hits, each unique annotation is preceded by a dash. Contig numbers marked with an asterisk had the known copy number outside of the confidence interval of the estimated copy number

E. coli

Contig	Length (bp)	Read depth	Blast hits	Est. copy number	Lower CL	Upper CL	Features
96	1124	266.8	10.0	9.7	8.2	11.5	IS5 transposase and trans-activator
120	3139	204.0	7.0	7.5	5.5	9.2	23S ribosomal RNA, 5S of <i>rrnA</i> , B, C, D, E, G, H operons
73	1148	196.5	7.0	7.2	5.5	8.6	rRNA-16S ribosomal RNA of <i>rrnA</i> , B, C, D, E, G, H operons
4	548	195.2	7.0	7.2	6.1	8.3	rRNA-16S ribosomal RNA of <i>rrnA</i> , B, C, D, E, G, H operons
58	705	192.6	7.0	7.0	5.8	8.4	KpLE2 phage-like element: IS2 insertion element transposase <i>InsAB</i> '
71	626	161.2	6.0	6.0	4.8	6.9	KpLE2 phage-like elements: IS2 insertion element repressor <i>InsA</i> , IS2 insertion element transposase <i>InsAB</i> '
67	1254	128.6	5.0	4.6	4.0	5.6	IS3 element protein <i>InsF</i> , IS3 element protein <i>InsF</i>
70	1196	85.6	3.1	3.1	2.4	4.0	KpLE2 phage-like element: IS30 transposase
81	1345	74.6	3.0	2.6	2.2	3.7	IS186/IS421 transposase
105	528	62.4	2.0	2.3	1.6	2.9	DLP12 prophage: DNA packaging protein/Qin prophage: predicted side tail fibre assembly protein
45*	734	60.0	3.0	2.1	1.7	2.8	<i>rhsA</i> , <i>rhsB</i> , <i>rhsC</i> element core protein
16	2072	57.5	2.0	2.1	1.4	3.0	Qin prophage: predicted DNA-binding, predicted site-specific, predicted tail fibre assembly, predicted side tail fibre assembly/Rac prophage: predicted tail fiber protein, predicted tail fiber assembly, predicted site-specific, predicted DNA-binding
76	1132	56.8	2.0	2.0	1.6	2.7	glutamate decarboxylase A, PLP-dependent/glutamate decarboxylase B, PLP-dependent
103	616	55.8	2.0	1.9	1.5	2.8	<i>rhsA</i> , <i>rhsB</i> element core protein
61	1030	54.6	2.0	2.0	1.5	2.6	protein chain elongation factor EF-Tu (<i>tufA</i> , <i>tufB</i>)
82	725	49.5	3.0	1.8	1.4	2.5	<i>rhsA</i> , <i>rhsB</i> element core protein
106*	848	41.3	2.0	1.4	1.1	2.0	<i>rhsD</i> element protein/pseudogene - <i>rhsE</i> element core protein <i>RhsE</i>
88*	1912	32.4	1.6	1.2	0.7	1.6	<i>rhsA</i> element core protein, lyase containing HEAT-repeat/ <i>rhsB</i> element core protein <i>RshB</i> , predicted protein, predicted transposase/conserved protein

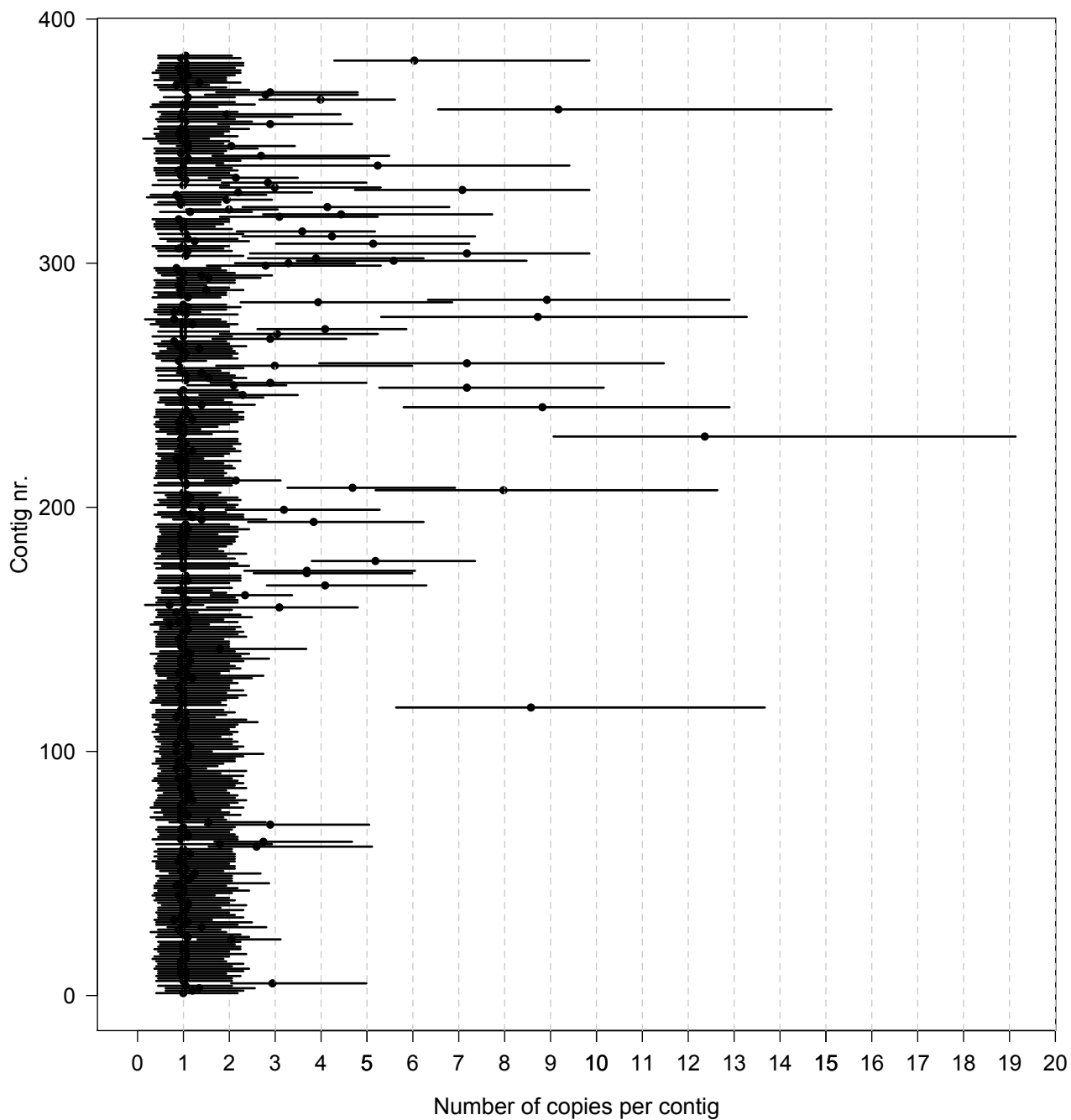
P. gingivalis

Contig	Length (bp)	Read depth	Blast hits	Est. copy number	Lower CL	Upper CL	Features
125	754	537.7	11.0	10.8	7.4	15.7	ISPg5 transposase Orf2
62	1143	515.1	10.0	10.3	7.7	14.6	ISPg4 transposase
152	720	293.6	5.1	5.8	3.7	9.0	hemagglutinin protein HagA (three copies in the same CDS)/HagD/HagE
64	1207	256.6	5.0	5.0	3.7	7.8	ISPg2 transposase
65	5824	213.3	4.0	4.2	2.5	6.6	5S ribosomal RNA, 23S ribosomal RNA, tRNA-Ala, tRNA-Ile, hypothetical protein, hypothetical protein, 16S ribosomal RNA, hypothetical protein, hypothetical protein
116	852	110.9	2.0	2.2	1.5	3.4	hemagglutinin protein HagB/HagC
163	592	100.9	2.0	2.0	1.4	3.0	von Willebrand factor type A domain protein/putative lipoprotein
112	3709	99	2.0	2.0	1.3	3.2	transcriptional regulator AraC family, MATE efflux family protein, hypothetical protein
122	804	98	2.0	1.9	1.3	3.1	histone-like family DNA-binding protein
41	4826	96.6	2.0	1.9	1.2	3.3	conserved hypothetical protein, conserved hypothetical protein, conserved hypothetical protein, putative lipoprotein, integrase, integrase
7	820	92.6	2.0	1.9	1.1	2.7	intergenic
159	673	63.7	2.0	1.3	0.5	2.5	translation elongation factor P

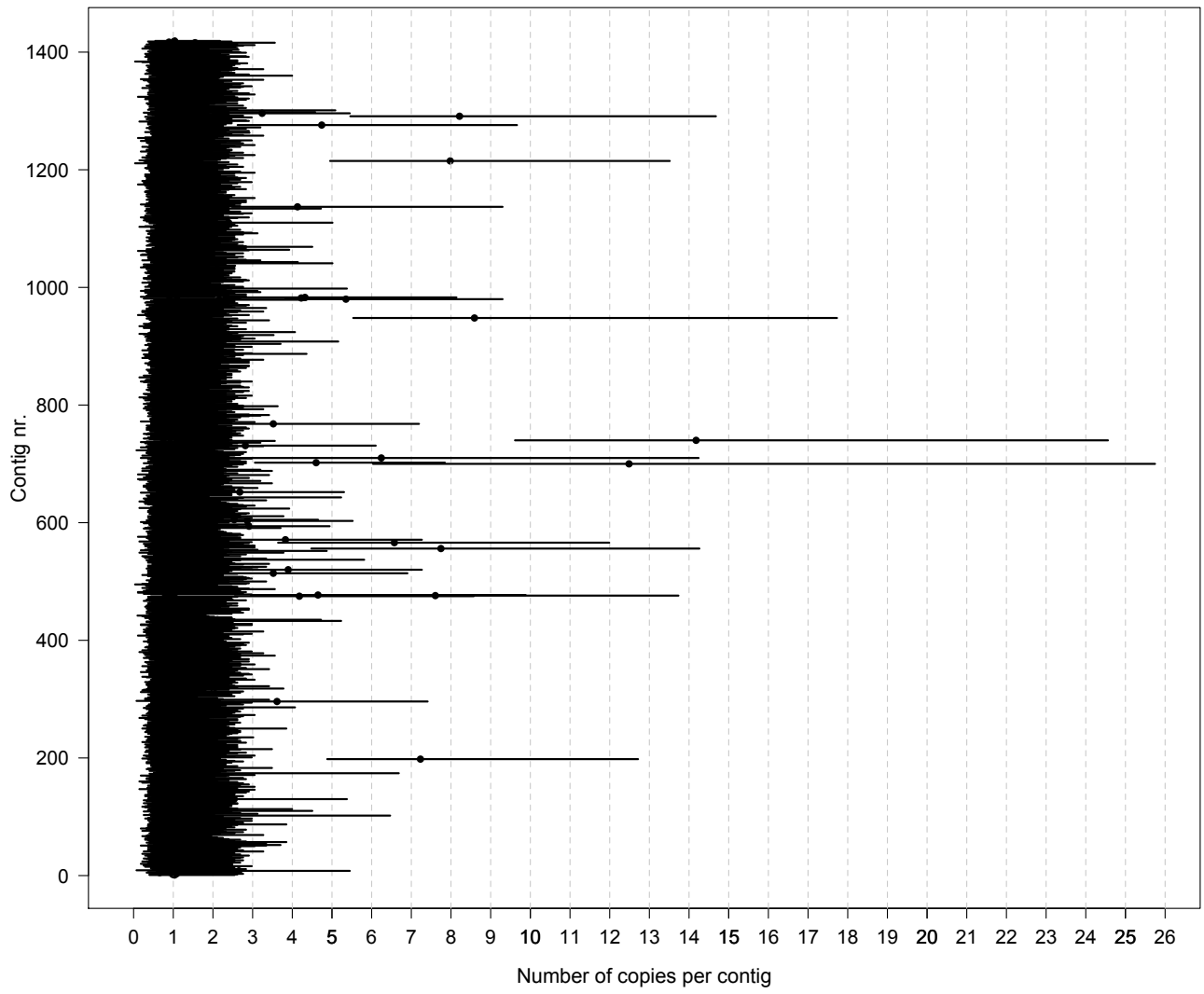
Supplementary Figure 2

Per-contig copy number estimates for *P. rubescens* and *A. flos-aqua* contigs. For each contig (vertical axis), the copy number is indicated with the corresponding confidence interval. A) *P. rubescens* B) *A. flos-aqua*.

A



B



Supplementary Table III: Annotation of all high read depths contigs for *P. rubescens* and *A. flos-aquae* assemblies

For each contig determined to be from a repeated region of the genome (estimated copy number of at least 1.5x), the length, read depth and estimate of copy number (‘Est. copy number’) with upper and lower Confidence interval Limits (CL) are shown. In addition, BLASTX results are shown (maximum E value 10^{-16}). When a contig had hits in multiple different regions, these are separated by a comma. The species to which the BLAST hit belongs are shown in between square brackets.

P. rubescens

Contig	Length (bp)	Read depth	Est. copy number	Lower CL	Upper CL	Features
13664	1309	259.7	12.4	9.1	19.1	transposase, IS4 family protein [Nostoc punctiforme PCC 73102]
13972	937	190.3	9.2	6.5	15.1	transposase IS4 family protein [Cyanotheca sp. PCC 8802]
13823	1190	180.2	8.9	6.3	12.9	transposase [Microcystis aeruginosa NIES-843]
13688	1109	176.3	8.8	5.8	12.9	transposase [Trichodesmium erythraeum IMS101]
136	3509	173.3	8.6	5.6	13.7	No hits
13792	9424	173.1	8.7	5.3	13.3	hypothetical protein Npun_R2618 [Nostoc punctiforme PCC 73102], DnaB domain protein helicase domain protein [Cyanotheca sp. PCC 7822]
13610	852	163.2	8.0	5.2	12.6	No hits
13711	1051	145.0	7.2	5.3	10.2	No hits
13735	2163	144.0	7.2	4.0	11.5	conserved hypothetical protein [Cyanotheca sp. PCC 7425]
13901	611	140.9	7.1	4.7	9.8	transposase, IS605 OrfB family [Cyanotheca sp. PCC 8801]
13846	902	132.8	7.2	2.5	9.8	hypothetical protein L8106_22791 [Lyngbya sp. PCC 8106]
14014	770	123.8	6.0	4.3	9.8	Histone-like DNA-binding protein [Lyngbya sp. PCC 8106]
13843	1712	111.4	5.6	3.5	8.5	RNA-directed DNA polymerase [Microcystis aeruginosa NIES-843]
13469	669	104.2	5.2	3.8	7.4	No hits
13858	641	99.5	5.1	3.0	7.2	hypothetical protein L8106_22631 [Lyngbya sp. PCC 8106]
13921	2057	99.3	5.2	1.7	9.4	transposase [Microcystis aeruginosa NIES-843]
13613	623	95.8	4.7	3.3	6.9	No hits
13885	2475	90.2	4.4	2.7	7.7	hypothetical protein L8106_13565 [Lyngbya sp. PCC 8106], hypothetical protein RCCS2_13854 [Roseobacter sp. CCS2], hypothetical protein Ava_A0028 [Anabaena variabilis ATCC 29413]
13868	1002	86.8	4.2	2.3	7.4	hypothetical protein L8106_13515 [Lyngbya sp. PCC 8106], hypothetical protein L8106_13520 [Lyngbya sp. PCC 8106]
13457	820	84.1	4.1	2.8	6.3	putative transposase [Lyngbya sp. PCC 8106]
13891	14027	82.6	4.1	2.3	6.8	hypothetical protein gl13812 [Gloeobacter violaceus PCC 7421], conserved hypothetical protein [Geobacter sp. M21], ORF013 [Staphylococcus phage EW], protochlorophyllide oxidoreductase [Nodularia spumigena CCY9414], hypothetical protein Mrad2831_6458 [Methylobacterium radiotolerans JCM 2831], putative GTP-binding protein [Marinobacter algicola DG893], putative ribose ABC transporter, ATP-binding protein [Clostridium difficile ATCC 43255], hypothetical protein Npun_R2618 [Nostoc punctiforme PCC 73102], GA25805 [Drosophila pseudoobscura pseudoobscura], hypothetical protein gl13812 [Gloeobacter violaceus PCC 7421], protochlorophyllide oxidoreductase [Nodularia spumigena CCY9414]

13821	23143	80.4	3.9	2.2	6.9	hypothetical protein CY0110_31940 [Cyanotheca sp. CCY0110], hypothetical protein L8106_07736 [Lyngbya sp. PCC 8106], hypothetical protein Tery_4236 [Trichodesmium erythraeum IMS101], helicase-like [Trichodesmium erythraeum IMS101], hypothetical protein PFL_4984 [Pseudomonas fluorescens Pf-5], Phytanoyl-CoA dioxygenase [Burkholderia graminis C4D1M], short-chain dehydrogenase [Azorhizobium caulinodans ORS 571], hypothetical protein SGR_180 [Streptomyces griseus subsp. griseus NBRC 13350], hypothetical protein SGR_179 [Streptomyces griseus subsp. griseus NBRC 13350], phosphoserine phosphatase, putative [Stigmatella aurantiaca DW4/3-1], aminotransferase [Azorhizobium caulinodans ORS 571], hypothetical protein PFL_4984 [Pseudomonas fluorescens Pf-5]
13763	693	80.2	4.1	2.6	5.9	transposase [Microcystis aeruginosa NIES-843]
13981	2361	79.9	4.0	2.7	5.6	RNA-directed DNA polymerase [Microcystis aeruginosa NIES-843]
13844	1318	78.8	3.9	2.4	6.2	No hits
13492	4639	77.5	3.8	2.4	6.2	hypothetical protein L8106_06319 [Lyngbya sp. PCC 8106]/NAD-dependent epimerase/dehydratase [Cyanotheca sp. PCC 8801]
13462	1575	76.9	3.7	2.5	6.0	16S rRNA
13463	2891	75.4	3.7	2.3	6.0	23S rRNA
13870	1456	70.4	3.6	2.2	5.2	putative bacteriophage-related protein [Ralstonia solanacearum GMI1000]
13842	640	65.0	3.3	2.1	4.7	hypothetical protein L8106_13465 [Lyngbya sp. PCC 8106]
13600	1696	63.4	3.2	1.9	5.3	phage Gp37Gp68 family protein [Nostoc punctiforme PCC 73102]
13884	1762	62.1	3.1	1.8	5.2	AAA ATPase, central region [Lyngbya sp. PCC 8106]
13761	3257	62.0	3.0	1.8	5.2	hypothetical protein L8106_13565 [Lyngbya sp. PCC 8106]
13734	5108	60.9	3.0	1.7	6.0	type III restriction enzyme, res subunit [Acidovorax avenae subsp. citrulli AAC00-1], unnamed protein product [Microcystis aeruginosa PCC 7806]
13903	9762	60.8	3.0	1.8	5.3	peptide synthetase [Anabaena sp. 90]
5	1182	60.7	2.9	2.0	5.0	hypothetical protein L8106_13520 [Lyngbya sp. PCC 8106]
608	1763	60.1	3.1	1.5	4.8	AAA ATPase, central region [Lyngbya sp. PCC 8106]
13907	2227	60.0	2.8	1.8	5.0	hypothetical protein L8106_13965 [Lyngbya sp. PCC 8106], hypothetical protein L8106_13620 [Lyngbya sp. PCC 8106]
13715	10768	58.7	2.9	1.6	5.0	Beta-ketoacyl synthase [bacterium Ellin514], NADH-dependent butanol dehydrogenase [Clostridium botulinum B1 str. Okra], Phytanoyl-CoA dioxygenase family [Microcoleus chthonoplastes PCC 7420], putative decarboxylase [Streptomyces griseus subsp. griseus NBRC 13350]
13984	1907	58.2	2.9	1.7	4.8	hypothetical protein L8106_13500 [Lyngbya sp. PCC 8106]
13931	1607	58.1	2.7	1.6	5.5	hypothetical protein L8106_13515 [Lyngbya sp. PCC 8106], hypothetical protein L8106_13520 [Lyngbya sp. PCC 8106]
79	26894	58.0	2.9	1.5	5.0	Phosphoglycerate mutase [Cyanotheca sp. PCC 7822],N-acetylglutamate synthase [Nitrococcus mobilis Nb-231],hypothetical protein N9414_06739 [Nodularia spumigena CCY9414],GNAT family acetyltransferase [Flavobacterium psychrophilum JIP02/86],KAP P-loop [Rhodospirillum rubrum T118],hypothetical protein Cyan8802DRAFT_4573 [Cyanotheca sp. PCC 8802],RecF/RecN/SMC N terminal domain, putative [Stigmatella aurantiaca DW4/3-1],putative transcriptional regulator [Nostoc punctiforme PCC 73102],helicase domain-containing protein [Nostoc punctiforme PCC 73102],hypothetical protein S7335_905 [Synechococcus sp. PCC 7335],integrase family protein [Cyanotheca sp. PCC 8802],
13758	1316	57.7	2.9	1.6	4.5	No hits
13951	2026	57.5	2.9	1.8	4.7	hypothetical protein L8106_13620 [Lyngbya sp. PCC 8106], hypothetical protein L8106_13590 [Lyngbya sp. PCC 8106]
13841	4628	57.3	2.8	1.5	5.3	putative helicase [Arthrobacter aurescens TC1]
72	1006	57.1	2.7	1.7	4.7	gas vesicle protein GvpK [Nostoc punctiforme PCC 73102]/hypothetical protein L8106_17822 [Lyngbya sp. PCC 8106]

13983	31787	55.8	2.8	1.5	4.8	hypothetical protein L8106_22726 [Lyngbya sp. PCC 8106],hypothetical protein Aave_1398 [Acidovorax avenae subsp. citrulli AAC00-1],hypothetical protein alr1877 [Nostoc sp. PCC 7120],Tn5044 transposase [Lyngbya sp. PCC 8106],AMP-dependent synthetase and ligase [Nostoc punctiforme PCC 73102],hypothetical protein [Candidatus Kuenenia stuttgartiensis],hypothetical protein Npun_F3358 [Nostoc punctiforme PCC 73102],aminotransferase, class III family [Stigmatella aurantiaca DW4/3-1],McnC protein [Microcystis aeruginosa NIES-843],peptide synthetase [Anabaena circinalis 90],ABC transporter related [Nostoc punctiforme PCC 73102],
70	2659	53.6	2.6	1.6	5.1	phage Gp37Gp68 family protein [Nostoc punctiforme PCC 73102]
12995	739	46.7	2.3	1.6	3.4	No hits
13708	1082	45.3	2.3	1.3	3.5	heat shock protein [Lyngbya sp. PCC 8106]
13910	714	44.0	2.1	1.6	3.5	glycosyl transferase family 2 [Arthrospira maxima CS-328]
13616	805	42.7	2.1	1.5	3.1	transposase [Microcystis aeruginosa NIES-843]
13900	739	41.6	2.2	0.9	3.8	No hits
13970	2544	41.6	1.9	0.5	4.4	hypothetical protein L8106_13440 [Lyngbya sp. PCC 8106], hypothetical protein L8106_13445 [Lyngbya sp. PCC 8106], hypothetical protein L8106_13465 [Lyngbya sp. PCC 8106]
24	2214	41.4	2.0	1.3	3.1	OciA [Planktothrix agardhii NIVA-CYA 116]
13713	1145	39.7	2.1	1.1	3.2	photosystem q(b) protein [Arthrospira maxima CS-328]
170	1520	39.7	1.8	1.0	3.7	gas vesicle protein GvpN [Arthrospira maxima CS-328]
13938	794	38.9	2.0	0.9	3.4	RNA-directed DNA polymerase [Lyngbya sp. PCC 8106]
13888	516	38.9	2.0	1.1	3.1	transposase [Lyngbya sp. PCC 8106]
13896	587	36.2	1.9	0.9	2.9	PilT protein domain protein [Arthrospira maxima CS-328], hypothetical protein AmaxDRAFT_3136 [Arthrospira maxima CS-328]
71	1057	34.1	1.8	0.4	2.9	ATP-binding region ATPase domain protein [Arthrospira maxima CS-328]/heat shock protein 90 [Trichodesmium erythraeum IMS101]
13717	520	31.8	1.5	1.1	2.4	No hits
81	2330	31.0	1.5	0.7	2.8	AAA ATPase, central region [Lyngbya sp. PCC 8106]
13834	778	30.8	1.5	0.9	2.7	No hits
13827	811	29.6	1.5	0.9	2.3	No hits

A. flos-aquae

Contig	Length (bp)	Read depth	Est. copy number	Lower CL	Upper CL	Features
13355	911	301.0	14.2	9.6	24.6	transposase [Microcystis aeruginosa NIES-843]
13273	748	257.2	12.5	6.0	25.7	transposase [Nodularia spumigena CCY9414], transposase [Nodularia spumigena CCY9414]
13683	1256	189.6	8.6	5.5	17.7	transposase [Nostoc sp. PCC 7120]
14262	560	174.4	8.2	5.5	14.7	transposase and inactivated derivatives [Syntrophus aciditrophicus]
14128	1185	165.6	8.0	5.0	13.5	unnamed protein product [Microcystis aeruginosa PCC 7806]
12918	1870	163.3	7.7	4.5	14.3	transposase [Microcystis aeruginosa NIES-843]
525	1566	160.6	7.6	4.4	13.7	hypothetical protein AM1_C0013 [Acaryochloris marina MBIC11017], hypothetical protein AM1_C0013 [Acaryochloris marina MBIC11017]
214	567	156.5	7.2	4.9	12.7	conserved hypothetical protein [Microscilla marina ATCC 23134], conserved hypothetical protein [Microscilla marina ATCC 23134]
12934	575	137.3	6.6	3.6	12.0	IS1 transposase subfamily, putative [Synechococcus sp. PCC 7335]
13740	502	115.3	5.4	0.2	14.2	No hits

13286	1782	109.0	6.2	3.8	9.3	transposase [Cyanotheca sp. ATCC 51142]
526	526	103.6	4.6	2.3	8.1	16S
14235	971	98.6	4.7			
13275	839	96.9	4.6	2.7	7.8	23S
13746	1585	91.0	4.3	2.6	9.7	transposase [Microcystis aeruginosa NIES-843]
13996	761	90.1	4.1	2.5	8.6	transposase [Nodularia spumigena CCY9414]
13744	2169	89.9	4.2	3.0	9.9	No hits
524	1381	88.8	4.2	3.1	7.8	hypothetical protein Npun_AF174 [Nostoc punctiforme PCC 73102]hypothetical protein MC7420_8266 [Microcoleus chthonoplastes PCC 7420]
12978	976	80.4	3.8	2.2	7.3	transposase [Cyanotheca sp. ATCC 51142]
12875	576	80.3	3.9	2.5	9.3	cell wall-associated hydrolase [Burkholderia multivorans ATCC 17616]
317	12570	77.6	3.6	2.2	7.3	photosystem II protein D1 [Nostoc sp. PCC 7120]
13411	2188	76.6	3.5	2.1	7.4	hypothetical protein all3502 [Nostoc sp. PCC 7120] , tRNA synthetase class II core domain-containing protein [Streptococcus gordonii str. Challis substr. CH1] , chromosome segregation ATPase-like protein [Clostridium phytofermentans ISDg] , hypothetical protein THERM_00393070 [Tetrahymena thermophila SB210] , hypothetical protein N9414_22178 [Nodularia spumigena CCY9414] , gp29 [Enterobacteria phage N15]
12868	1736	74.9	3.5	2.1	7.2	RNA-directed DNA polymerase (Reverse transcriptase):HNH endonuclease [Crocospaera watsonii WH 8501]
14272	634	68.6	3.2	1.8	6.9	similar to Transposase [Crocospaera watsonii WH 8501]
13344	609	64.1	2.8	1.6	5.5	hypothetical protein RB886 [Rhodospirellula baltica SH 1]
13076	616	61.6	2.9	2.0	5.4	hypothetical protein Npun_F3986 [Nostoc punctiforme PCC 73102]
13109	1408	60.5	2.9	2.0	4.9	transposase [Nostoc sp. PCC 7120]
13210	1897	58.6	2.7	1.6	6.1	hypothetical protein N9414_19552 [Nodularia spumigena CCY9414]
9	1496	55.6	2.7	1.8	5.3	hypothetical protein CY0110_01714 [Cyanotheca sp. CCY0110]
141	7089	54.1	2.5	1.5	5.4	hypothetical protein N9414_06264 [Nodularia spumigena CCY9414]
13881	1138	54.0	2.5	1.4	5.4	hypothetical protein alr9014 [Nostoc sp. PCC 7120], hypothetical protein slr6006 [Synechocystis sp. PCC 6803]
13111	606	53.2	2.5	1.7	4.5	IS605 family transposase OrfB [Nostoc punctiforme PCC 73102], transposase [Nostoc sp. PCC 7120]
13954	1117	53.2	2.4	1.3	5.2	similar to Transposase [Crocospaera watsonii WH 8501]
466	6990	51.4	2.4	1.1	4.4	primase P4 [Cyanotheca sp. PCC 7424]
13600	588	49.7	2.5	1.4	4.7	hyalin repeat-containing protein [Microcystis aeruginosa NIES-843]
14283	10638	47.7	2.2	1.5	5.0	putative AAA ATPase [Cyanotheca sp. ATCC 51142]
13730	766	47.6	2.2	1.3	3.9	conserved hypothetical protein [Cyanotheca sp. PCC 7425]
13839	3248	47.1	2.1	0.8	4.6	lanthionine synthetase C family protein [Nostoc punctiforme PCC 73102], WD-40 repeat-containing protein [Nostoc punctiforme PCC 73102]
13773	8122	46.9	2.2	1.4	4.4	transposase [Anabaena circinalis AWQC131C]
13139	500	45.7	2.2	1.0	5.1	Cobyrinic acid ac-diamide synthase [Cyanotheca sp. PCC 7425] , hypothetical protein Ava_0235 [Anabaena variabilis ATCC 29413] , Phage integrase [Anabaena variabilis ATCC 29413] , hypothetical protein MAE_18850 [Microcystis aeruginosa NIES-843] , conserved hypothetical protein [Cyanotheca sp. PCC 7424] , putative phage repressor [Shewanella baltica OS195] , hypothetical protein Ava_A0028 [Anabaena variabilis ATCC 29413] , midasin [Saccharomyces cerevisiae YJM789]
14390	518	45.3	2.1	1.1	5.0	WD-40 repeat protein [Cyanotheca sp. CCY0110]
14275	5756	45.1	2.2	1.1	5.4	chaperone protein DnaK (Hsp 70) [Mycobacterium abscessus]
468	1337	44.9	2.1	1.3	4.0	glycosyl transferase, group I [Nostoc punctiforme PCC 73102]
340	683	43.6	2.0	1.0	4.7	nucleotide sugar epimerase [Rhodospirellula baltica SH 1]

13841	1113	43.4	2.0	1.3	3.8	hypothetical protein N9414_08590 [Nodularia spumigena CCY9414]
13643	818	39.7	1.9	1.2	4.1	IS891/IS1136/IS1341 transposase [Anabaena variabilis ATCC 29413]
13353	714	38.7	1.8	1.1	3.5	photosystem II reaction centre protein PsbD/D2 [Anabaena variabilis ATCC 29413]
13872	1857	38.1	1.8	1.1	3.6	hypothetical protein MC7420_4812 [Microcoleus chthonoplastes PCC 7420], hypothetical protein N9414_05429 [Nodularia spumigena CCY9414]
320	575	34.0	1.5	0.2	3.1	No hits
14493	903	33.6	1.5	0.9	3.4	putative lipoprotein [Shewanella oneidensis MR-1], hypothetical protein MAE_58460 [Microcystis aeruginosa NIES-843]
14394	763	31.9	1.5	0.4	2.8	hypothetical protein Npun_R5163 [Nostoc punctiforme PCC 73102]
13765	631	31.1	1.6	0.9	3.9	transposase [Microcystis aeruginosa NIES-843]
14293	537	30.9	1.5	0.8	2.9	hypothetical protein Npun_R5163 [Nostoc punctiforme PCC 73102]
14287	502	30.7	1.5	0.8	3.6	transposase, IS4 family protein [Nostoc punctiforme PCC 73102]
13126	659	27.7	1.5	0.7	2.8	hypothetical protein alr0554 [Nostoc sp. PCC 7120]