

# Subsystem: Carotenoid Biosynthesis

Olga Vassieva, Fellowship for Interpretation of Genomes

## Introduction

- Carotenoids are synthesised in plants, photosynthetic bacteria (carotenes, xanthins) and archaea (bacteriorhodopsin ) as components of light harvesting complexes and UV-protectors, as well as in nonphotosynthetic organisms - where their main function is UV-protection.
- Carotenoids differ by length of carbon skeleton (C-30,C-40, C-50 carotenoids), number and type of cyclic structures in a molecule (one or two rings of beta or epsilon conformation), presence of ketogroups, and types of carbohydrate side chains.
- About 30 genes involved in these pathways are currently known. Nevertheless, many steps of carotenoid biosynthesis are still a mystery, and genes encoding several enzymes of central importance have not been identified in a number of organisms, including cyanobacteria (“missing genes”).

***Subsystem: Carotenoid biosynthesis (general list of known functions)***

Column	Abbrev	Functional Role
1	PS	Phytoene synthase (EC 2.5.1.32)
2	PDS	Phytoene desaturase, pro-zeta-carotene producing (EC 1.-.-.-)
3	NDS	Neurosporene desaturase (EC 1.-.-.-)
4	GPS	Geranylgeranyl pyrophosphate synthetase (EC 2.5.1.29)
5	CI	Carotenoid cis-trans isomerase (EC 5.2.-.-)
6	ZDS	Pro-zeta-carotene desaturase, polycopene producing (EC 1.-.-.-)
7	BCH	Beta-carotene hydroxylase
8	BLCY	Lycopene beta cyclase (EC 1.14.-.-)
9	BCK	Beta-carotene ketolase (EC 1.14.-.-)
10	ELCY	Lycopene epsilon cyclase
11	ZGT	Zeaxanthin glucosyl transferase
12	PCS	"phi-Carotenoid synthase" (EC 1.3.-.- and EC 2.1.1.-)
13	HNS	Hydroxyneurosporene synthase
14	HNM	Hydroxyneurosporene methyltransferase (EC 2.1.1.-)
15	PD	Phytoene dehydrogenase and related proteins
16	VM	CrtV-methyltransferase-like protein
17	TM	CrtT-methyltransferase-like protein
18	BCHF	2-vinyl bacteriochlorophyllide hydratase (EC 4.2.1.-)
19	PDSB	Phytoene desaturase, neurosporene or lycopene producing (EC 1.3.-.-)
20	DSS	Dehydrosqualene synthase (EC 2.5.1.-)
21	SD	Squalene desaturase
22	SM	Spheroidene/spirilloxanthin monooxygenase
23	LE	Lycopene elongase
24	SRCB	sigma-factor-like regulator of carotenoid biosynthesis
25	RCB	Regulator of carotenoid biosynthesis

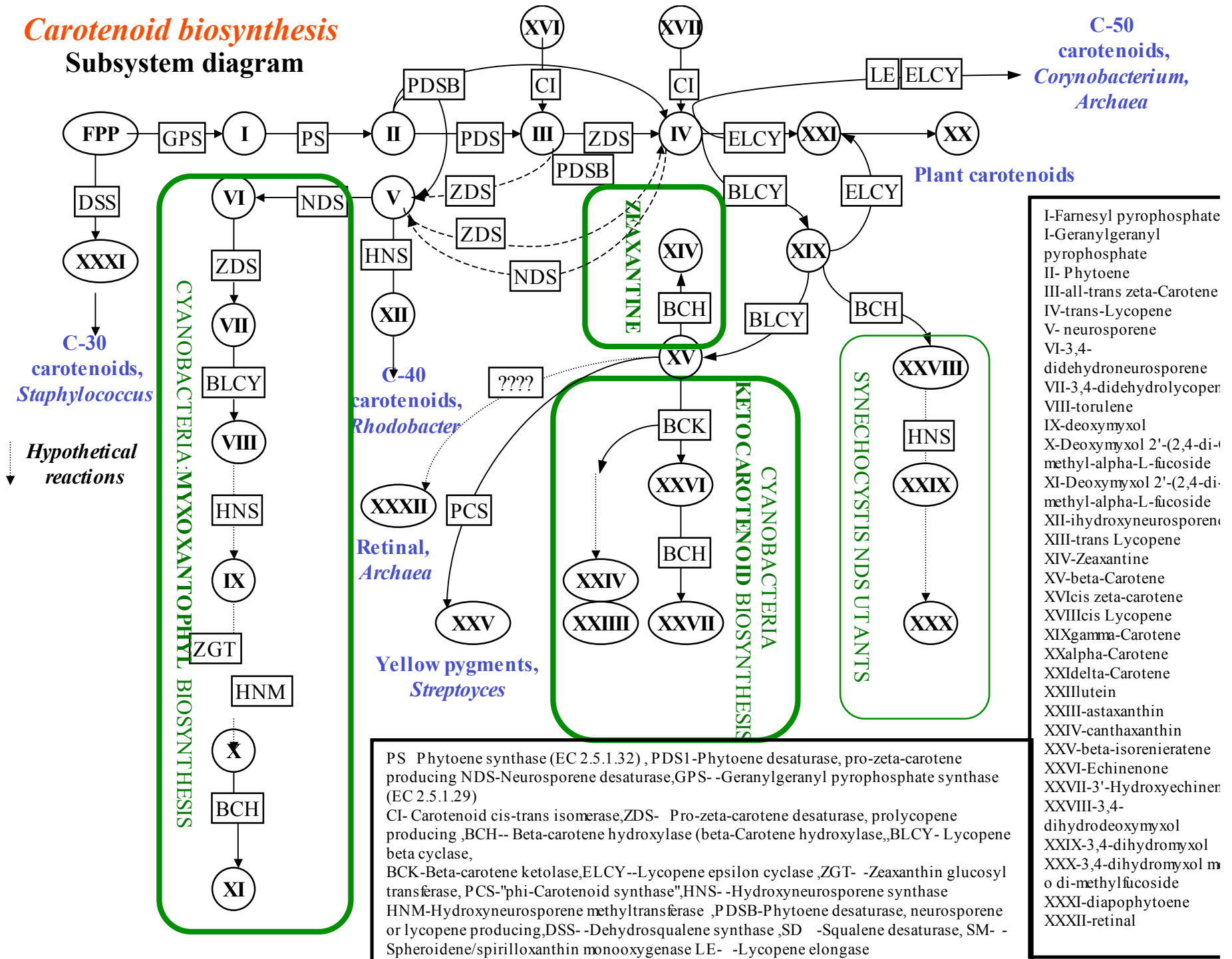
Organism	Variant Code	PS	PDS	NDS	GPS	CI	ZDS	BCH	BLCY	BCK	ELCY	ZGT	PCS
Synechocystis sp. PCC 6803 [B]	0	<a href="#">1269</a>	<a href="#">1268</a>	<a href="#">261</a>	111, 3167	<a href="#">2828</a>	<a href="#">1821</a>	<a href="#">889</a>		<a href="#">2592</a>			<a href="#">73</a>
	0	<a href="#">1367</a>	<a href="#">795</a>		<a href="#">254</a>	<a href="#">640</a>	<a href="#">1394</a>		<a href="#">180</a>				
Thermosynechococcus elongatus BP-1 [B]	0	<a href="#">1559</a>	<a href="#">1560</a>	<a href="#">232</a>	<a href="#">20</a>		<a href="#">336</a>	<a href="#">1899</a>					
Synechococcus sp. WH 8102 [B]	0	<a href="#">2247</a>	<a href="#">2248</a>		<a href="#">737</a>	<a href="#">896</a>	<a href="#">2206</a>	<a href="#">288</a>	<a href="#">723</a>	<a href="#">1689</a>			
Prochlorococcus marinus MED4 [B]	0	<a href="#">305</a>	<a href="#">607</a>		<a href="#">1185</a>	<a href="#">1138</a>	<a href="#">333</a>	<a href="#">212</a>	<a href="#">1191</a>				
Prochlorococcus marinus str. MIT 9313 [B]	0	<a href="#">1995</a>	<a href="#">1996</a>		<a href="#">1107</a>	<a href="#">1049</a>	1960, 953	<a href="#">1810</a>	<a href="#">1121</a>	<a href="#">1552</a>	<a href="#">1767</a>		
Prochlorococcus marinus subsp. marinus str. CCMP1375 [B]	0	<a href="#">166</a>	<a href="#">167</a>		<a href="#">1127</a>	<a href="#">583</a>	<a href="#">136</a>	<a href="#">266</a>	<a href="#">1134</a>		<a href="#">788</a>		
Synechococcus elongatus PCC 7942 [B]	0	<a href="#">1293</a>	<a href="#">1292</a>	<a href="#">313</a>	<a href="#">125</a>	<a href="#">577</a>	<a href="#">836</a>	<a href="#">1734</a>	<a href="#">1370</a>				
Nostoc sp. PCC 7120 [B]	0	<a href="#">2140</a>	<a href="#">2139</a>	<a href="#">5430</a>	<a href="#">524</a>	<a href="#">2371</a>	2095, 2689	<a href="#">4316</a>	<a href="#">5430</a>	<a href="#">4051</a>			
Nostoc punctiforme [B]	0	<a href="#">2875</a>	<a href="#">2874</a>	<a href="#">5675</a>	5170, 745	<a href="#">3633</a>	464, 859	<a href="#">5347</a>	<a href="#">5675</a>	184, 772		<a href="#">2699</a>	
Organism	Variant Code	PS	PDS	NDS	GPS	CI	ZDS	BCH	BLCY	BCK	ELCY	ZGT	PCS
Anabaena variabilis ATCC 29413 [B]	0	<a href="#">3722</a>	<a href="#">3721</a>	<a href="#">4341</a>	<a href="#">5421</a>	<a href="#">4633</a>	4853, 6105	<a href="#">1577</a>	<a href="#">4341</a>	<a href="#">1964</a>			
Arabidopsis thaliana [E]	0	<a href="#">23888</a>	19124, 19125		16359, 13592, 13594, 14243, 15259, 15387, 2024, 21784, 21966, 4426, 7401, 8858, 8860, 9405	<a href="#">860</a>	12461, 12462	20484, 27141	<a href="#">27645</a>	<a href="#">26798</a>	<a href="#">13046</a>	<a href="#">23869</a>	
Crocospaera watsonii WH 8501 [B]	0	<a href="#">762</a>	<a href="#">761</a>	<a href="#">3206</a>	<a href="#">2154</a>	<a href="#">2857</a>	2233, 4102	1038, 4133		<a href="#">1968</a>		<a href="#">411</a>	
Gloeobacter violaceus PCC 7421 [B]	0	<a href="#">1744</a>			416, 753	<a href="#">2133</a>			<a href="#">2874</a>	1728, 394			
Oryza sativa (japonica cultivar-group) [E]	0				<a href="#">5744</a>								
Prochlorococcus marinus subsp. pastoris str. CCMP1986 [B]	0	<a href="#">143</a>	<a href="#">1562</a>		<a href="#">1066</a>	<a href="#">1111</a>	<a href="#">115</a>		<a href="#">1060</a>		<a href="#">630</a>		
Trichodesmium erythraeum IMS101 [B]	0	<a href="#">5153</a>	<a href="#">5154</a>	<a href="#">3089</a>	<a href="#">1721</a>	<a href="#">6180</a>	<a href="#">5085</a>	<a href="#">5965</a>				<a href="#">6281</a>	
Rhodobacter capsulatus SB1003 [B]	0	<a href="#">2832</a>	<a href="#">2829</a>		2828, 2896								
Staphylococcus aureus NCTC 8325 [B]	0				<a href="#">1286</a>								
Organism	Variant Code	PS	PDS	NDS	GPS	CI	ZDS	BCH	BLCY	BCK	ELCY	ZGT	PCS
Streptomyces avermitilis MA-4680 [B]	0	<a href="#">1025</a>	1024, 1654		1130, 1653, 2999, 3008				<a href="#">1022</a>	<a href="#">1002</a>		<a href="#">7360</a>	<a href="#">1020</a>

*Subsystem:  
Carotenoid  
biosynthesis*

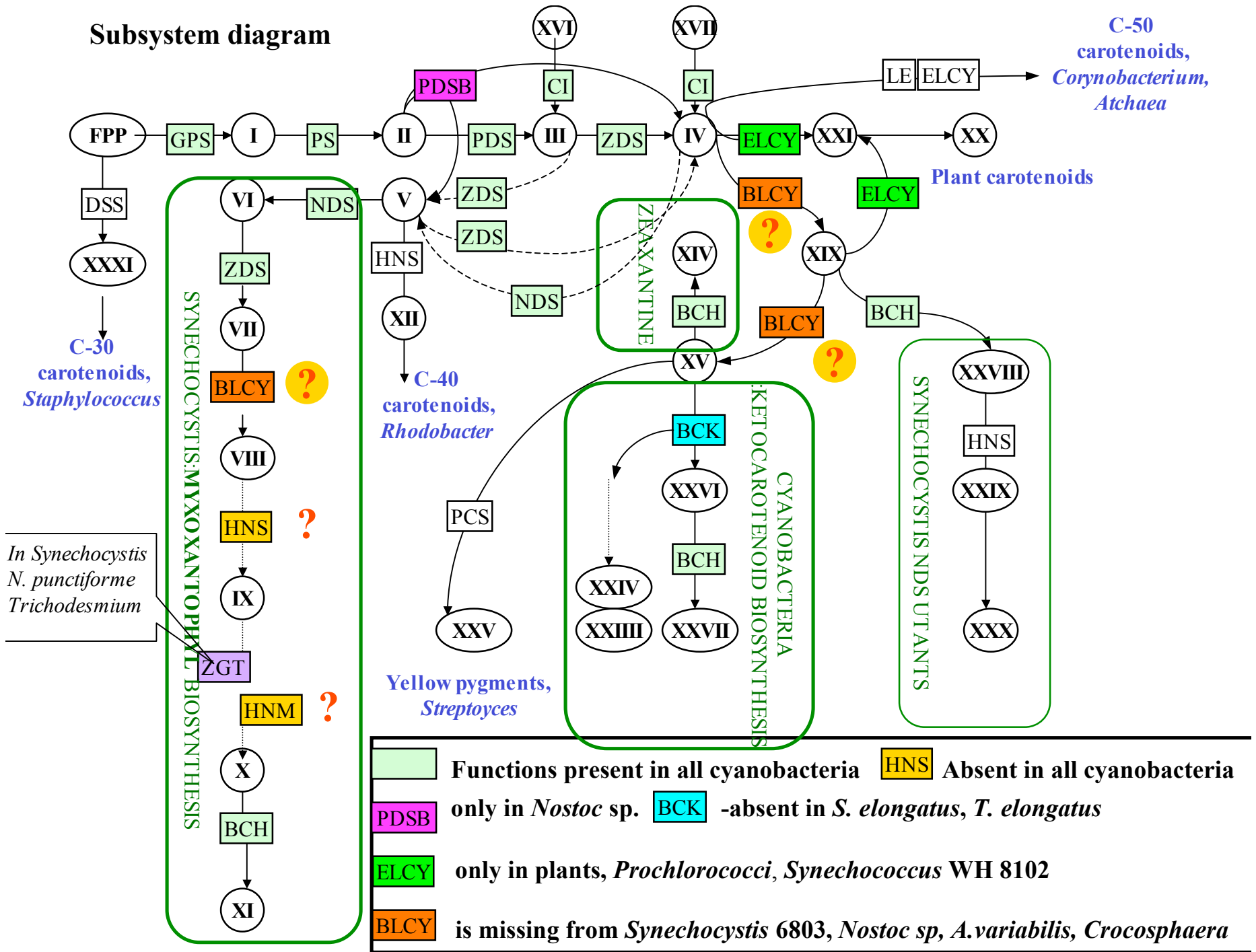
**CYANOBACTERIA**

For notes and [comments](#)  
see subsystem (hyperlinked)

# Carotenoid biosynthesis Subsystem diagram

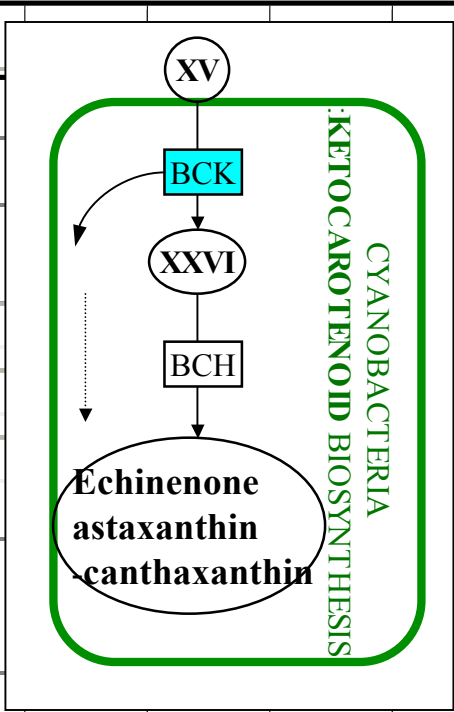


# Subsystem diagram



# Open questions, comments, conjectures. 1. The case of “missing” beta-carotene ketolase

Organism	Variant Code	S	CI	ZDS	BCH	BLCY	BCK	ELCY
Synechrocystis sp. PCC 6803 [B]	0	167	2828	1821	889		2592	
Chlorobium tepidum TLS [B]	0	254	640	1394		180		
Thermosynechococcus elongatus BP-1 [B]	0	20		336	1899			
Synechococcus sp. WH 8102 [B]	0	737	896	2206	288	723	1689	
Prochlorococcus marinus MED4 [B]	0	1185	1138	333	212	1191		1767
Prochlorococcus marinus str. MIT 9313 [B]	0	1107	1049	1960, 953	1810	1121	1552	788
Prochlorococcus marinus subsp. marinus str. CCMP1375 [B]	0	1127	583	136	266	1134		
Synechococcus elongatus PCC 7942 [B]	0							
Nostoc sp. PCC 7120 [B]	0	1			836	1734		1370
Nostoc punctiforme [B]	0	2			4316		4051	
Anabaena variabilis ATCC 29413 [B]	0	2			359	5347	184, 772	
Crocospaera watsonii WH 8501 [B]	0	3			1577		1964	
	0				1038, 4133		1968	



*Biosynthesis of keto-carotenoids depends on beta-carotene ketolase – known in bacteria as the product of crtW or crtO genes*

*crtO*

? *Our annotation*

? *Our annotation 30% homology to carotene isomerase*

*crtO*

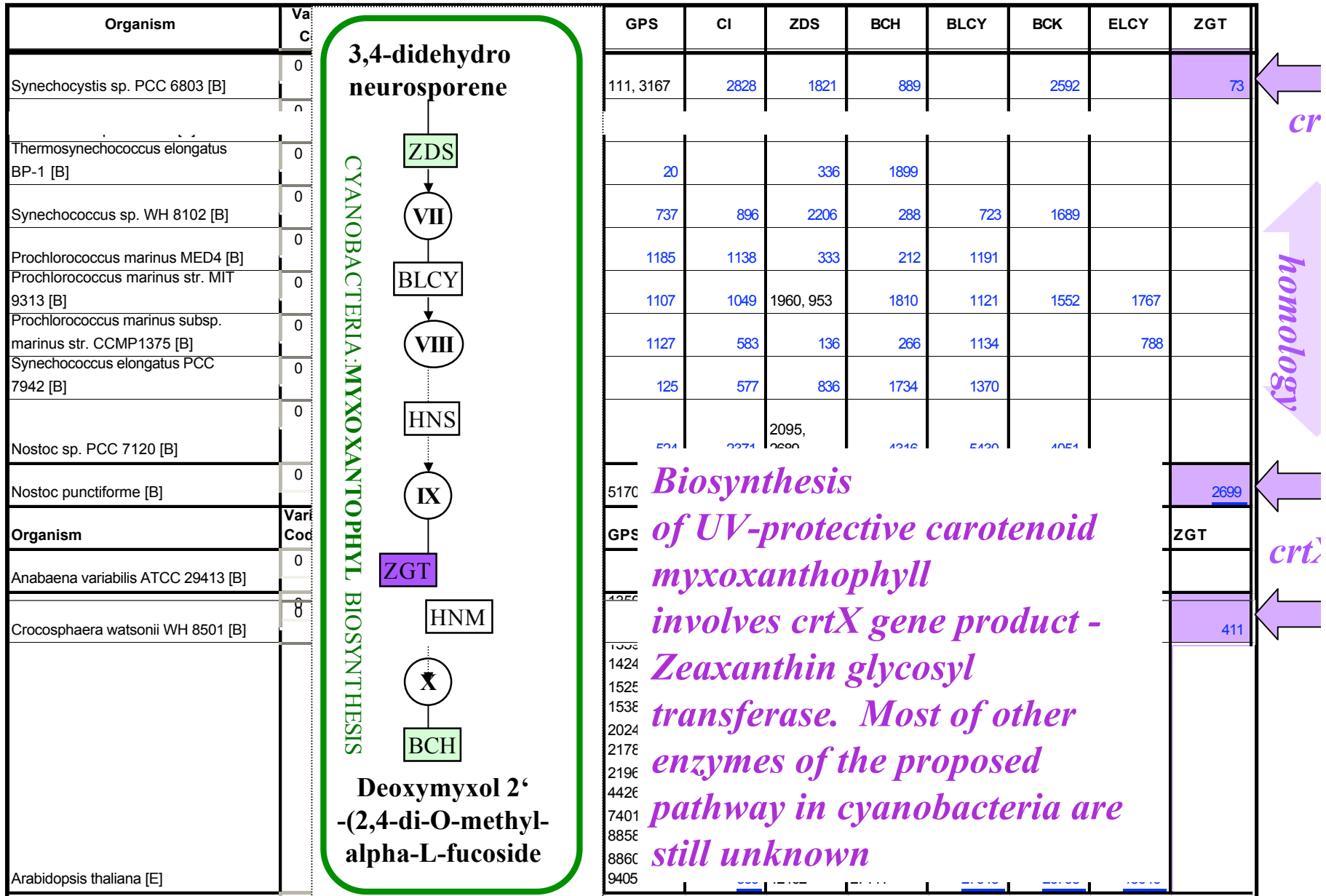
*crtO*

*crtO*

?

*Our annotation: 30% homology to fatty acid desaturases (characteristic for plant carotene ketolases/hydroxylases). Circadian clock cluster. Homologs in all Prochlorococci*

## Open questions, comments, conjectures. 2. The case of “missing” glycosyl transferase



## Case 3. The case of “missing” beta-lycopene cyclase

Organism	Variant Code	PS	PDS	NDS	GPS	CI	ZDS	BCH	BLCY	BCK	ELCY
Synechocystis sp. PCC 6803 [B]	0	1269	1268	261	111, 3167	2828	1821	889	?	2592	
Chlorobium tepidum TLS [B]	0	1367	795		254	640	156		180		
Thermosynechococcus elongatus BP-1 [B]	0	1559	1560	232	20		336	1899	?		
Synechococcus sp. WH 8102 [B]	0	2247	2248		737	896	2206	288	723	1689	
Prochlorococcus marinus MED4 [B]	0	305	607		1185	1138	333	212	1191		1767
Prochlorococcus marinus str. MIT 9313 [B]	0	1995	1996		1107	1049	1960, 953	1810	1121	1552	788
Prochlorococcus marinus subsp. marinus str. CCMP1375 [B]	0	166	167		1127	583	136	266	1134		
Synechococcus elongatus PCC 7942 [B]	0	1293	1292	313	125	577	836	1734	?		1370
Nostoc sp. PCC 7120 [B]	0	2140	2139	5430	524	2371	2095, 2689	4316	?	4051	
Nostoc punctiforme [B]	0	2875	2874	5675	5170, 745	3633	464, 859	5347	?	184, 772	
Anabaena variabilis ATCC 29413 [B]	0	3722	3721	4341	5421	4633	4853, 6105	1577	?	1964	
Crocospaera watsonii WH 8501 [B]	0	762	761	3206	2154	2857	2233, 4102	1038, 4133	?	1968	

30%homolo



## *Search for gene candidates for the role of “missing” lycopene cyclase*

Gene candidates were selected on the basis of (i) functional clustering with known genes of carotenoid biosynthesis pathway (7 candidates), and/or (ii) sequence similarity with known forms of lycopene cyclase (6 candidates).

Hypothetical proteins as well as hypothetical cyclases, synthases, and dehydrogenases were considered. Since known lycopene cyclases are membrane proteins, hypothetical proteins with predicted transmembrane domains were preferentially selected.

Occurrence profile was used as an additional filter: gene candidates were ranked based on the presence of their clear homologs in all genomes lacking known lycopene cyclase (several cyanobacteria and *Chlorobium tepidum* - see previous slide) and their absence from the genomes where lycopene cyclase homolog(s) have been identified (e.g. *Synechococcus* sp. WH 8102 and *Prochlorococci*).

## Gene candidates for the role of “missing” lycopene cyclase identified in cyanobacter

### **I: Oligoketide cyclase/lipid transport protein homolog (example: Slr0941 in *Synechocystis* 6803)**

**Supporting evidences:** (i) homologous to oligoketide cyclase/lipid transport protein, (ii) is a membrane protein, (iii) very strong chromosomal clustering: is co-localized with orthologs of pro-zeta-carotene desaturase, prolycopene producing (ZDS) in nearly all analyzed cyanobacterial genomes. **Negative features:** (i) orthologs are present in *Prochlorococci*, and (ii) absent in the genome of *Chlorobium tepidum*.

### **II: Two closely related hypothetical dehydrogenases (examples: Sll0659 and Sll0147)**

**Supporting evidences:** (i) nearly perfect occurrence profile, (ii) proposed function – dehydrogenase, (iii) presence a UbiH domain (Ubiquinone biosynthesis reductase), hence potential activity towards related isoprenoid derivatives, (iv) distant homology with lycopene cyclase (CrtL) from *Prochlorococci*.

### **III: Putative farnesyl diphosphate farnesyl transferase (example: Sll0513 in *Synechocystis* 6803)**

**Supporting evidences:** (i) highly homologous to phytoene/squalene synthetase and/or farnesyl diphosphate farnesyl-transferase (EC2.5.1.21), (ii) distant homology to lycopene cyclase / phytoene synthase from *Blakeslea trispora*, (iii) satisfactory occurrence profile, (iv) is a membrane protein. **Negative feature:** absent in the genomes of *Chlorobium tepidum* and *T. elongates*

### **IV: Hypothetical protein (example: Slr1579 in *Synechocystis* 6803)**

**Supporting evidence:** (i) homology, albeit distant, with CrtYm (monocyclic)-lycopene cyclase, functionally analogous to the one inferred in *Synechocystis* 6803 (monocyclic lycopene cyclase). This low homology may be still meaningful, since CrtYm has the same low homology to *Prochlorococci* lycopene cyclase (CrtY). **Negative feature:** absent in *Chlorobium tepidum*

### **V: Hypothetical protein (example: Slr1394 in *Synechocystis* 6803)**

**Supporting evidence:** (i) strong chromosomal clustering with geranylgeranyl pyrophosphate synthase (GPS) in cyanobacteria (excluding *Synechocystis* 6803), (ii) integral membrane protein, (iii) orthologs are evident in plants, (iv) satisfactory occurrence. **Negative features:** ortholog is absent in *Chlorobium tepidum*, and present (rather distant however) in *Prochlorococci* and *Synechococcus* sp. WH 8102.

### **Cyclase or carotene p450 oxygenase/hydroxylase**

**Supporting evidence:** (i) chromosomal coupling with beta-caroten ketolase (EC 1.14.-.-). Distant homolog of lycopene cyclase/phytoene synthase from *Blakeslea trispora*

*Gene candidates  
for “missing”  
functional roles in  
carotenoid  
biosynthesis:  
occurrence profile  
in cyanobacterial  
genomes.*

Organism/protein	I	II	III	IV	V	p450
Synechocystis sp. PCC 6803	slr0941	sll0659/ sll0147	sll0513	slr1579	slr1394	slr0574
Chlorobium tepidum TLS		gjl21673292				
Thermosynechococcus elongatus BP-1	gjl22297879,	/gjl22298682	gjl22298640	gjl22299456	gjl22297563	
Synechococcus sp. WH 8102	gjl33866744			gjl33866807	gjl33865277	gjl33865042
Prochlorococcus marinus MED4	<25% homology					
Prochlorococcus marinus str. MIT 9313	gjl33864232				gjl33863379 <50% homology	gjl33863832
Prochlorococcus marinus subsp. marinus str. CCMP1375	gjl33239587			gjl33239638 (only 18% homology)	gjl33240578	
Synechococcus elongatus PCC 7942	gjl46129933	gjl45511722		gjl45511757	gjl46129485	
Nostoc sp. PCC 7120	gjl17229873	gjl17228415/ gjl17231016	gjl17229297	gjl17231534	gjl172277	gjl17231238
Nostoc punctiforme	gjl23130102	gjl23126269/ gjl23130367	gjl23125690	gjl23127750	gjl23129765	gjl23127990
Anabaena variabilis ATCC 29413		gjl46135210	Alr1805 protein	gjl45508370	gjl45510320	gjl105304
Crocospaera w, WH8501	ref ZP_00178762.1	gjl46118345/ gjl45528290	gjl46119198	gjl46119504	gjl46118491	gjl45526212
Trihodesmium IMS101	gjl48893657	gjl48894952	gjl48891505	gjl48890958	gjl48893862	gjl48894521

## Case 4. Predicted carotenoid glycosyl transferase

Synechocystis sp. PCC 6803 [B]	
Chlorobium tepidum TLS [B]	
Thermosynechococcus elongatus BP-1 [B]	
Synechococcus sp. WH 8102 [B]	
Prochlorococcus marinus MED4 [B]	
Prochlorococcus marinus str. MIT 9313 [B]	
Prochlorococcus marinus subsp. marinus str. CCMP1375 [B]	
Synechococcus elongatus PCC 7942 [B]	
Nostoc sp. PCC 7120 [B]	
Nostoc punctiforme [B]	
Anabaena variabilis ATCC 29413 [B]	
Crocospaera w, WH8501	
Trihodesmium IMS101	

- Homologs present in the all cyanobacterial genomes
- Example: gi|33865263 (SYNW0729) in *Synechococcus* sp. WH 8102
- Current annotation in this organism: Glycosyltransferase (EC 2.4.1.-)
- Clustering with lycopene cyclase in *Synechococcus* sp. WH 8102 and *Prochlorococci*
- Annotations in other organisms: hypothetical protein

**Case 5. Strong functional coupling of hypothetical ORFs with known genes of isoprenoid or carotenoid biosynthesis may be indicative of their potential role in these pathways**

**Genes related to isoprenoid/carotenoid biosynthesis (open arrows):**

- 1: Carotene isomerase
- 2: Phytoene desaturase, z-carotene producing
- 3: 1-deoxy-D-xylulose-5-P reductoisomerase
- 4: Ubiquinone monooxygenase UbiB

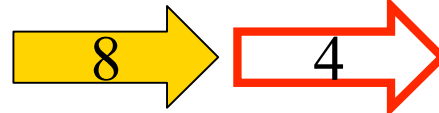
**Gene candidates (solid arrows), annotated currently as:**

- 5: hypothetical P-loop-ATPase nucleotidyltransferase
- 6: motif: prenyl binding site
- 7: Oridoketide cyclase/lipid transport protein
- 8: sensory transduction histidine kinase

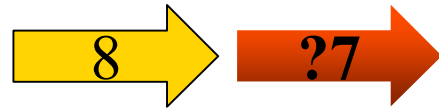
*Crocospaera watsonii* WH 8501



*Synechococcus* sp. WH 8102

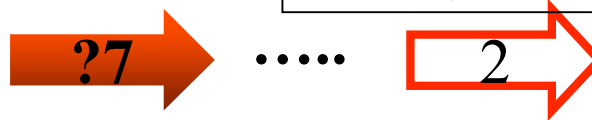


*Nostoc punctiforme*



*Prochlorococcus marinus*

*Synechococcus* sp. WH 8102



*Synechococcus elongatus*



*Pyrococcus horikoshii*



*Aquifex aeolicus* VF5

