



**Table S1.** Morphological description of the sequencing strains.

Sequencing Strains	Morphological Description
<b><i>Nostoc sphaeroides</i> CHAB2801</b>	Colonies are spherical, soft, colloid, with smooth surface and solid periphery. Solitary colonies are usually up to 1-4 cm in diameter. Filaments flexuous, irregularly aggregated. Vegetative cells are short barrel-shaped or nearly spherical, about 4.55 $\mu\text{m}$ long and about 4.7 $\mu\text{m}$ wide, terminal cells often slightly elongated and narrowed up to 2.5 $\mu\text{m}$ . Heterocytes are nearly spherical, 6.2-7.3 $\mu\text{m}$ in length and 5.4-6.6 $\mu\text{m}$ in width. The akinetes cell wall is smooth and spherical.
<b><i>Minunostoc cylindricum</i> CHAB5844</b>	Colonies macroscopic, gelatinous, irregularly clustered, free-living on wet rocky wall in the form of irregular, amorphous, blue-green gelatinous clusters. Filaments flexuous, freely entangled. Sheath usually present, colourless. Vegetative cells barrel-shaped, 3.1- (4.3)-5.4 $\mu\text{m}$ long, 2.5- (3.2)-3.8 $\mu\text{m}$ wide, length: width ratio 1.3. Heterocytes and akinetes were not observed.
<b><i>Nostoc sp.</i> XA013</b>	Colonies microscopic. In the early stage of life, the filaments were very densely aggregated, could not be seen clearly. In the later stage, the filaments arranged loosely, and the filament broke up into short filaments, and then the spheres burst to release shorter filaments. The end cells of filaments are longer than wide, 2.26-3.34 $\mu\text{m}$ in length, short-barrel to long-barrel vegetative cells, 1.46-2.15-2.80 $\mu\text{m}$ in length and 2.07-2.42-2.80 $\mu\text{m}$ in width. Heterocytes are short barrel-shaped, larger than vegetative cells, about 3.97 $\mu\text{m}$ long and 4.47 $\mu\text{m}$ wide. No akinete was found.
<b><i>Nostoc sp.</i> XA010</b>	Colonies are gelatinous, irregularly clustered. Filaments are usually enveloped by gelatinous sheath. Vegetative cells are short barrel-shaped, 3.15 $\mu\text{m}$ long and 3.51 $\mu\text{m}$ wide. Heterocytes are larger than vegetative cells, long barrel-shaped, 4.81 $\mu\text{m}$ long and 4.34 $\mu\text{m}$ wide. No akinete was found.
<b><i>Nostoc sp.</i> CHAB5714</b>	Several small spherical colonies are enveloped by a large colony, and the large colony surface are smooth with firm periderm. Filaments are densely entangled especially in young colonies. Vegetative cells are short barrel-shaped to nearly spherical, or long oval, about 3.09 $\mu\text{m}$ in length and 4.43 $\mu\text{m}$ in width. Heterocytes are nearly spherical, similar in size to vegetative cells, with a diameter of 4.3 $\mu\text{m}$ . No akinete was observed.
<b><i>Nostoc sp.</i> CHAB5715</b>	Colonies are spherical, microscopic. In the early stage of life, small colonies are aggregated into a long filamentous colony. Old colony are solitary. Trichome are enveloped by a thick, colorless sheath. Vegetative cells are short barrel-shaped, 4.19 $\mu\text{m}$ long and 5.29 $\mu\text{m}$ wide. Heterocytes are nearly spherical, slightly larger than vegetative cells, with a diameter of 5.99 $\mu\text{m}$ . No akinete was observed.
<b><i>Nostoc sp.</i> CHAB5784</b>	Colonies are spherical, microscopic. Young colonies are connected with each other by the heterocyte, filaments are densely entangled in colonies. Single trichome are enveloped by sheath or not. Vegetative cells are oval or spherical, about 3.36 $\mu\text{m}$ in length and 3.50 $\mu\text{m}$ in width, and the ratio of length to width is 0.96. Heterocytes are slightly larger than vegetative cells, usually short barrel-shaped, about 3.89 $\mu\text{m}$ in length and 4.06 $\mu\text{m}$ in width. No akinete was observed.
<b><i>Nostoc sp.</i> CHAB5836</b>	In young colonies, filaments are closely wound, filaments adhere to the firm sheath, and then to form a large spherical colony. The vegetative cells of a single filament are cylindrical, and the vegetative cells in the colony are short barrel-shaped, 3.48 $\mu\text{m}$ long and 3.68 $\mu\text{m}$ wide. Heterocytes are larger than vegetative cells, nearly spherical or short barrel-shaped, 4.10 $\mu\text{m}$ long and 4.50 $\mu\text{m}$ wide, with a length-width ratio of 0.91. No akinete was observed.
<b><i>Nostoc sp.</i> CHAB5824</b>	Colonies are spherical, microscopic. heterocyte are at both end of the young, oval colony. Filaments are densely entangled in colonies. Vegetative cells are oval, 1.82-4.11-5.14 $\mu\text{m}$ long and 1.82-2.56-3.58 $\mu\text{m}$ wide. Heterocytes and vegetative cells are similar in size, short barrel-shaped, 2.44-3.04-3.95 $\mu\text{m}$ long and 2.33-2.99-3.73 $\mu\text{m}$ wide. No akinete was observed.
<b><i>Nostoc sp.</i> CHAB5834</b>	Colonies are spherical, macroscopic, with loose and irregular filaments. Vegetative cells are spherical to short barrel-shaped, 1.77-2.35-2.95 $\mu\text{m}$ long and 2.03-2.63-3.43 $\mu\text{m}$ wide; heterocytic,

spherical or long elliptic, slightly larger than vegetative cells, about 3.58  $\mu\text{m}$  in length and 3.66  $\mu\text{m}$  in width, no akinete was observed.

**Table S2.** Genome assembly and binning results.

Organism name	Length	Contigs	GC (%)	Proteins	Completeness	Contamination
<i>N. favosum</i> CHAB5714	8,514,377	245	42.28	9637	100	1.89
<i>N. exigua</i> CHAB5715	7,054,926	1201	42.52	9732	87.62	3.41
<i>N. mirabile</i> CHAB5784	9,892,412	297	42.18	11341	99.92	4.19
<i>N. tenue</i> CHAB5824	8,117,724	138	42.27	8296	99.7	2.42
<i>N. rigida</i> CHAB5834	9,307,052	563	43.22	10676	100	7.82
<i>N. pachydermaticum</i> CHAB5836	7,461,718	384	42.55	8549	99.11	2.78
<i>N. minutum</i> XA010	8,210,661	137	41.80	8536	100	1.56
<i>N. xinjiangensis</i> XA013	8,483,714	211	41.71	8865	100	6.9
<i>N. sphaeroides</i> CHAB2801	9,217,956	11	41.45	10585	Complete	Complete
<i>Minunostoc cylindricum</i> CHAB5844	7,868,568	208	41.40	8248	99.04	1.41

**Table S3.** Basic information of downloaded sequences in this research.

Organism name	GenBank number	Length	contigs	GC (%)	Proteins	Completeness	Contamination
<i>N.sp.</i> 213	1168001	8,331,440	380	41.4	7021	99	0.74
<i>N.sp.</i> 232	1167991	9,158,766	398	41.6	8097	98.78	1.37
<i>N. commune</i> NIES-4072	1693891	8,198,700	44	41.40	7489	99.77	0.22
<i>N.sp.</i> ATCC 53789	1822731	8,536,834	220	41.32	7370	99.56	0.44
<i>N. calcicola</i> FACHB-389	916171	9,049,214	354	41.5	7401	99.22	1.11
<i>N. piscinale</i> CENA21	488611	7,094,556	11	40.5	6172	96.99	0.44
<i>N.sp.</i> PCC 7120	31208	7,211,789	7	41.32	6931	99.19	0
<i>N.sp.</i> PCC 7107	502788	6,329,823	1	41.32	5456	99.26	0.33
<i>N.sp.</i> PCC 7524	502768	6,718,869	3	41.32	5655	99.33	0
<i>N.sp.</i> CENA543	1532881	7,207,640	6	41.32	6329	99.78	0.22
<i>N.sp.</i> 5183	1592541	7,339,908	5	41.32	6769	99.55	0.44
<i>N. linckia</i> NIES-25	1232421	8,590,040	21	41	7042	99.89	0.89
<i>N.sp.</i> NIES-2111	1232511	7,876,755	11	41.32	7945	99.3	0.11
<i>N.sp.</i> KVJ20	786451	9,184,886	425	41.32	7813	99.67	0.89
<i>N.sp.</i> 210A	1168011	8,325,913	73	41.32	7130	99.22	1.33
<i>N.sp.</i> NIES-3756	635661	6,987,571	3	41.32	6451	99.3	0.11
<i>N.sp.</i> DB3992	1368581	8,521,473	1,037	41.32	7860	94.54	4.6
<i>N.sp.</i> RF31Y	1110281	9,205,737	629	42	7759	99.11	2.61
<i>N. cycadae</i> WK-1	1533001	6,994,461	150	40.6	5803	99.56	0.89
<i>N.sp.</i> NIES-4103	1232571	8,590,607	50	41.18	6994	99.26	0.44
<i>N. carneum</i> NIES-2107	1232481	9,359,223	16	40.70	7541	99.33	0.19
<i>N.sp.</i> 106C	1110051	8,662,725	241	41.3	7184	99.11	1.44
<i>N.sp.</i> T09	1110041	8,251,853	314	41.32	6969	99.11	0.33
<i>N. punctiforme</i> PCC73102	41068	9,059,191	6	41.42	7548	99.56	0.44
<i>N. sphaeroides</i> Kutzing En	1912231	6,684,490	5	41.58	5998	99.48	0
<i>N.sp.</i> N6	1592491	8,898,590	11	41.32	8353	99.44	1.19
<i>N. flagelliforme</i> CCNUN1	1460651	10,231,233	9	41.84	9655	100	3.07

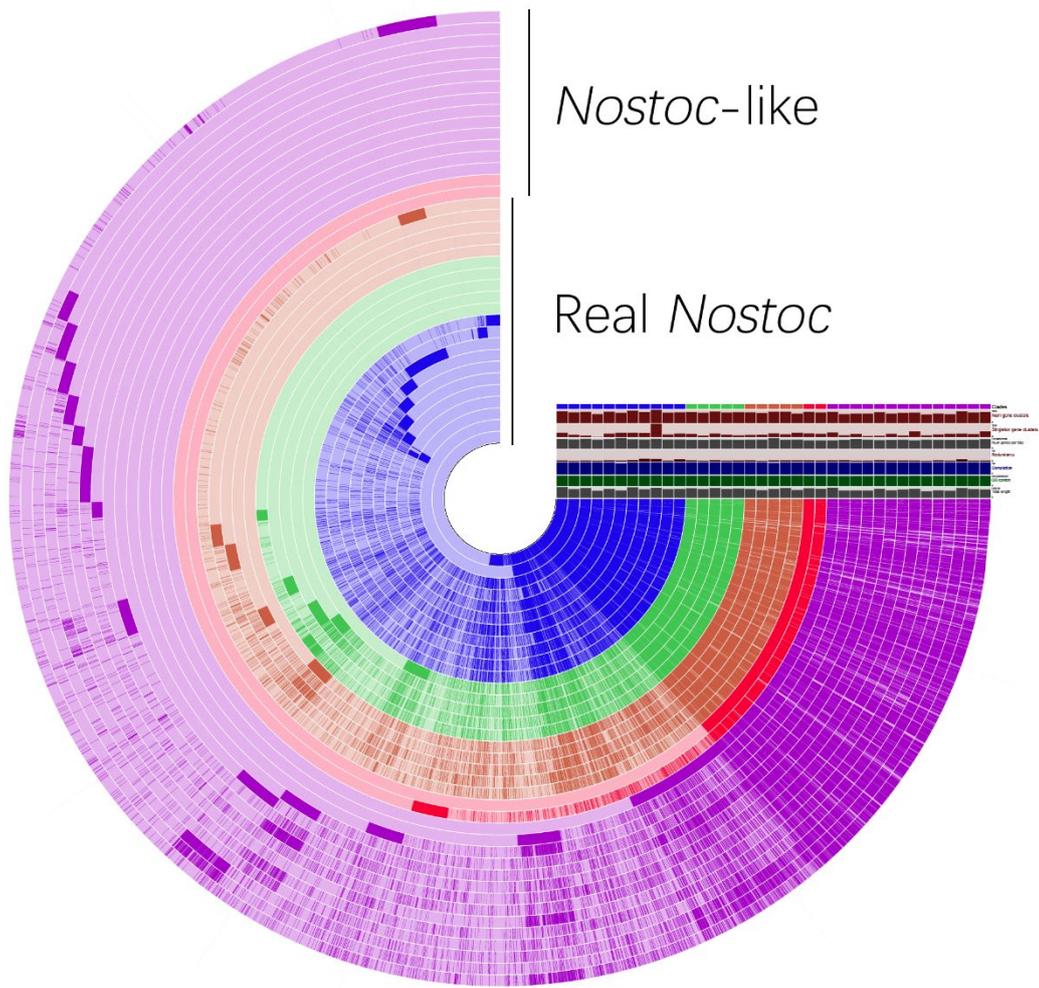
**Table S4.** The genes present in the genome of the NSS closest common ancestor but missing in the previous ancestor.

Gene Name	COG	Function
<i>alkB</i>	L	Alkylated DNA repair protein
<i>ALL0363</i>	L	Transposase
<i>RC1_0998</i>	L	Transposase
<i>AVA_2181</i>	L	Inherit from COG: transposase
<i>pre</i>	S	plasmid recombination enzyme
<i>wzm</i>	V	Transporter
<i>PLAV_1218</i>	S	methylase helicase
<i>BSL1696</i>	S	WGR domain-containing protein
<i>nifV</i>	E	Homocitrate synthase
<i>psaB</i>	C	PsaA and PsaB bind P700, the primary electron donor of photosystem I (PSI), as well as the electron acceptors A0, A1 and FX. PSI is a plastocyanin cytochrome c6-ferredoxin oxidoreductase, converting photonic excitation into a charge separation, which transfers an electron from the donor P700 chlorophyll pair to the spectroscopically characterized acceptors A0, A1, FX, FA and FB in turn. Oxidized P700 is reduced on the lumenal side of the thylakoid membrane by plastocyanin or cytochrome c6
<i>PilL</i>	T	CheA signal transduction histidine kinase
<i>PsaJ</i>	C	May help in the organization of the PsaE and PsaF subunits (By similarity)
<i>YfdZ</i>	E	Aminotransferase
<i>OCAR_5339</i>	S	oxidoreductase
<i>folE</i>	H	GTP cyclohydrolase I
<i>mtnD</i>	S	Catalyzes 2 different reactions between oxygene and the acireductone 1,2-dihydroxy-3-keto-5-methylthiopentene (DHK-MTPene) depending upon the metal bound in the active site. Fe-containing acireductone dioxygenase (Fe-ARD) produces formate and 2-keto-4- methylthiobutyrate (KMTB), the alpha-ketoacid precursor of methionine in the methionine recycle pathway. Ni-containing acireductone dioxygenase (Ni-ARD) produces methylthiopropionate, carbon monoxide and formate, and does not lie on the methionine recycle pathway (By similarity)
<i>cheB</i>	T	catalyzes the demethylation of specific methylglutamate residues introduced into the chemoreceptors (methyl-accepting chemotaxis proteins) by CheR (By similarity)
<i>PDEN_1306</i>	D	Cobyrinic acid ac-diamide synthase
<i>ceaA</i>	P	efflux system protein
<i>AZL_008610</i>	S	Phage protein Gp37/Gp68
<i>ssuD</i>	S	Monoxygenase
<i>ndh</i>	C	FAD-dependent pyridine nucleotide-disulfide oxidoreductase
<i>gstA</i>	O	Glutathione S-transferase N-terminal domain protein
<i>metX</i>	E	homoserine O-acetyltransferase
<i>mutT</i>	F	Nudix hydrolase
<i>SCLAV_5486</i>	S	Inherit from bactNOG: Cupin 2, conserved barrel domain protein
<i>BL01679</i>	E	SAM-dependent methyltransferase
<i>SCLAV_2440</i>	S	ABC transporter
<i>ycgL</i>	S	Predicted nucleotidyltransferase
<i>lon</i>	O	Inherit from bactNOG: ATP-dependent Lon protease
<i>yaiS</i>	S	GlcNAc-PI de-N-acetylase
<i>rfbF</i>	M	Glucose-1-phosphate cytidyltransferase
<i>umaA</i>	M	cyclopropane-fatty-acyl-phospholipid synthase
<i>SSCG_04708</i>	S	protein containing aminopeptidase domain
<i>rfbC</i>	M	dTDP-4-dehydrorhamnose 3,5-epimerase

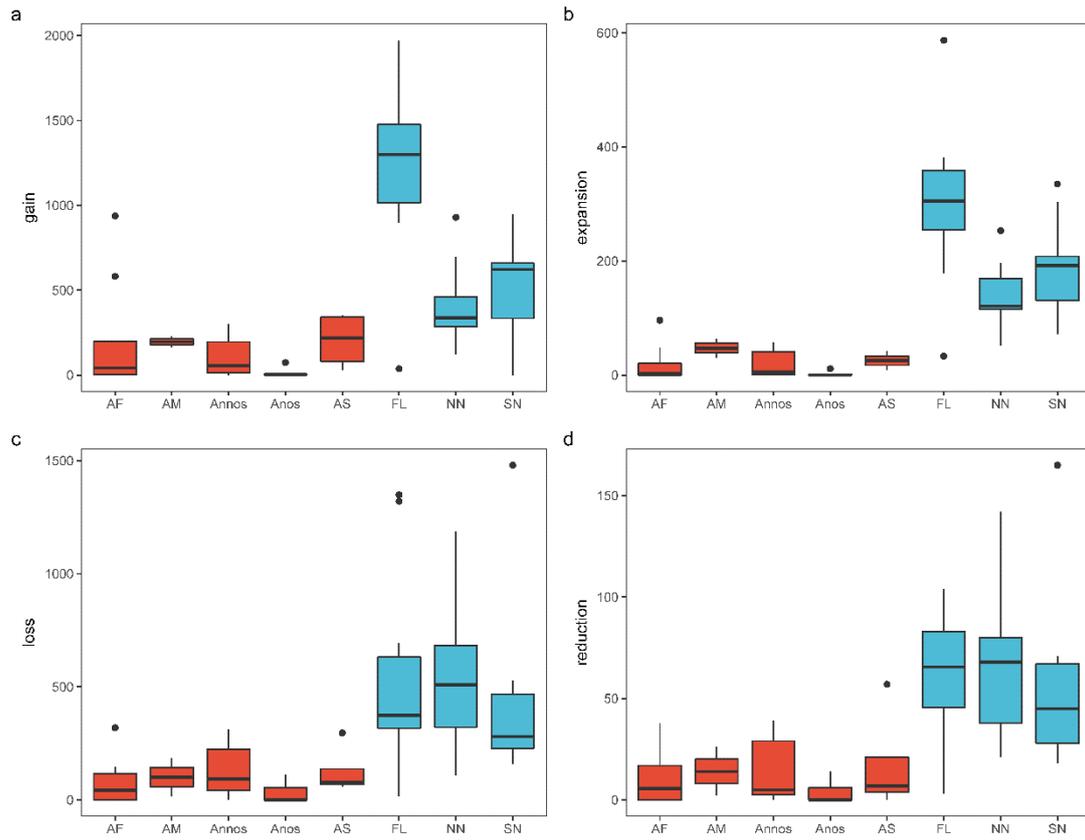
## Supplementary material S6

**Table S5.** The genes present in the genome of the common ancestor of NSS and CRC but missing in the NSS ancestor.

Gene Name	COG	Function
<i>GLL1611</i>	L	Transposase
<i>dnaB</i>	L	Replicative DNA helicase
<i>AVA_1381</i>	L	Transposase IS66
<i>NblA</i>	S	phycobilisome degradation protein
<i>CheB2</i>	T	CheB methylesterase
<i>AppA</i>	E	Extracellular solute-binding protein, family 5
<i>AqpZ</i>	M	Channel that permits osmotically driven movement of water in both directions. It is involved in the osmoregulation and in the maintenance of cell turgor during volume expansion in rapidly growing cells. It mediates rapid entry or exit of water in response to abrupt changes in osmolarity (By similarity)
<i>BMUL_4716</i>	S	ATP-binding protein
<i>gumM</i>	M	Glycosyl transferase WecB TagA CpsF
<i>yokB</i>	K	Transcriptional regulator, TetR family
<i>spoT</i>	K, T	HDc
<i>igiC</i>	Q	Asp Glu hydantoin racemase
<i>lpcA</i>	M	Glycosyl transferase family 8
<i>wcaF</i>	M	Inherit from bactNOG: colanic acid biosynthesis acetyltransferase wcaF
<i>SCLAV_1551</i>	S	Methyltransferase type 11
<i>HTH_0578</i>	S	Phosphatidylethanolamine-binding protein
<i>mtnB</i>	G	Catalyzes the dehydration of methylthioribulose-1- phosphate (MTRu-1-P) into 2,3-diketo-5-methylthiopentyl-1- phosphate (DK-MTP-1-P) (By similarity)
<i>dmpI</i>	S	4-oxalocrotonate tautomerase
BL05203	E	Amino acid permease
<i>PolB</i>	L	DNA polymerase
<i>OCAR_6107</i>	M	Glycosyl transferases group 1
<i>rsgA</i>	S	May play a role in 30S ribosomal subunit biogenesis. Unusual circularly permuted GTPase that catalyzes rapid hydrolysis of GTP with a slow catalytic turnover (By similarity)
<i>rbpD</i>	S	RNA-binding protein
<i>purM</i>	F	phosphoribosylaminoimidazole synthetase
<i>bchE</i>	C	hopanoid biosynthesis associated radical SAM protein HpnJ
<i>ypfJ</i>	S	Putative neutral zinc metallopeptidase
<i>tauB</i>	P	ABC transporter
PPS	G	pyruvate phosphate dikinase PEP pyruvate-binding
<i>yccU</i>	S	CoA_binding
<i>rplL</i>	J	Seems to be the binding site for several of the factors involved in protein synthesis and appears to be essential for accurate translation (By similarity)
<i>sbP</i>	S	56kDa selenium binding protein (SBP56)
<i>OCAR_5127</i>	S	Sulfite exporter TauE/SafE
<i>pyrE</i>	F	Catalyzes the transfer of a ribosyl phosphate group from 5-phosphoribose 1-diphosphate to orotate, leading to the formation of orotidine monophosphate (OMP) (By similarity)
<i>modA</i>	P	ABC transporter, periplasmic molybdate-binding protein
<i>MC7420_3283</i>	T	RecA-superfamily ATPases implicated in signal transduction
<i>SCLAV_3941</i>	O	PHB
<i>cpsF</i>	M	Oligosaccharide biosynthesis protein Alg14 like
<i>pssE</i>	S	Glycosyltransferase 28 domain protein



**Figure S2.** Pangenome analysis of all available genomes.



**Figure S2.** Evolutionary events happened along the phylogeny of *Nostoc* and other *Nostoc*-like species. **AF:** Ancestor of Free-living *Nostoc* **AM:** Ancestor of mixture clades **Annos:** Ancestor of Non-*Nostoc* species (ie., *Nostoc*-like species) **AS:** Ancestor of Real *Nostoc* species **FL:** Free-living *nostoc* **NN:** Non-Symbiosis *Nostoc* **SN:** Symbiosis *nostoc*.