

Table 1 General features of the Nostocales genomes of the newly recognized ADA clade, including eight genomes newly reported in this work. Shading indicates membership of the phylogroups ADA-1 through -4 delineated in Fig. 1.

Genome	Abbrev. name	Isolation Site & Date	Genome Size (Mbp)	¹ Source/Status/Contig No.	NCBI Accession Number	² Estimated % Genome Completeness/Contamination	GC%	No. Protein Coding Sequences	³ No. rRNA operons (5S-16S-23S)	Reference
<i>Anabaena</i> sp. CRKS33	Ana CRKS33	Cheney Reservoir, KS, USA; 30-Aug-2013	4.95	EM / D 1109	LJOT00000000	99.44 1.78	37.6	4638	2	This work
<i>Dolichospermum circinale</i> AWQC131C	Dol 131C	Lake Cargelligo, NSW, Australia	4.45	C / D 121	NZ_KE384588	99.56 0.00	37.01	3750	1	D'Agostino et al., 2014
<i>Dolichospermum circinale</i> AWQC310F	Dol 310F	Farm Dam, Milawa, VIC, Australia; 1995	4.41	C / D 82	NZ_KE384663	99.56 0.00	37.33	3676	2	D'Agostino et al., 2014
<i>Anabaena</i> sp. AL09	Ana AL09	Lake Ontario, Canada; 1-Aug-2005	4.66	C / D 109	LJOQ00000000	98.11 0.00	38.1	3988	-	This work
<i>Anabaena</i> sp. LE011-02	Ana LE011	Lake Erie, USA; 12-Jul-2011	4.74	C / D 122	LJOP00000000	99.22 0.11	38.06	4072	-	This work
<i>Anabaena</i> sp. MDT14b	Ana MDT14b	Upper Klamath Lake, OR, USA; 4-Jun-2014	4.96	EM / D 1227	LJOV00000000	97.17 4.22	38.9	4546	4	This work
<i>Anabaena</i> sp. 90	Ana 90	Lake Vesijärvi, Finland; 30-Jul-1986	5.31	C / F 5(2 + 3)	NC_019427	- -	38.1	4444	5	Wang et al., 2012
<i>Anabaena</i> sp. AL93	Ana AL93	American Lake, WA, USA; 1993	5.66	C / D 217	LJOU00000000	99.67 0.52	38.4	4693	5	Brown et al., 2016
<i>Anabaena</i> sp. WA102	Ana WA102	Anderson Lake, WA, USA; 20-May-2013	5.78	C / F 2(1 + 1)	NZ_CP011456	- -	38.4	4880	5	Brown et al., 2016
<i>Aphanizomenon flos-aquae</i> 2012/KM1/D3	AFA KM1D3	Curonian Lagoon, Lithuania, 2012	5.74	C / D 325	NZ_JSDP01000254	87.52 7.22	38.22	4601	5	Sulcius et al., 2015
<i>Aphanizomenon flos-aquae</i> NIES-81	AFA NIES-81	Lake Kasumigaura, Japan; 1978	5.85	C / D 103	NZ_KI928192	99.67 0.56	37.37	4744	5	Cao et al., 2014
<i>Aphanizomenon flos-aquae</i> MDT14a	AFA MDT14a	Upper Klamath Lake, OR, USA; 4-Jun-2014	4.63	EM / D 193	LJOX00000000	99.00 1.00	37.11	3936	6	This work
<i>Aphanizomenon flos-aquae</i> LD13	AFA LD13	Upper Klamath Lake, OR, USA; 6-Aug-2013	4.44	C / D 307	LJOY00000000	99.67 0.37	37.05	3787	3	This work
<i>Anabaena</i> sp. WA113	Ana WA113	Cranberry Lake, WA, USA; 11-Aug-2014	4.70	EM / D 279	LJOS00000000	99.89 0.44	37.22	4002	7	This work
<i>Aphanizomenon flos-aquae</i> WA102	AFA WA102	Anderson Lake, WA, USA; 20-May-2013	5.48 ⁴	EM / D 1160	LJOW00000000	99.89 3.60	39.12	4685 ⁴	7	This work

¹ EM, environmental metagenome-assembled genome; C, culture; D, draft genome; F, finished genome. No plasmids were identified for the draft genomes. Plasmid(s) were associated with both finished genomes; number of chromosomes and plasmids indicated in parentheses.

² % Genome completeness and proportion of contaminating sequences estimated by CheckM (Parks et al., 2015).

³ Multiple copy ribosomal RNA genes are susceptible to absence or incomplete representation in shotgun metagenomes; thus, rRNA operon numbers are provisional for draft genomes (See Supplementary Table 3 for more details). tRNAs are likewise commonly incompletely represented in draft genomes: see Supplementary Table 4 for tRNA genes.

⁴ Reflects removal of 464,126 bp of likely viral sequences in 20 contigs of original draft genome cluster, including 611 coding sequences and 13 tRNA genes.

Table 2 Genome fragmentation in ADA clade genomes

	ADA-1			ADA-2				ADA-3				ADA-4			
	Ana CRKS33	Dol 131C	Dol 310F	Ana AL09	Ana LE011	Ana MDT14b	Ana 90	Ana AL93	Ana WA102	AFA KM1D3	AFA NIES-81	AFA MDT14a	AFA LD13	Ana WA113	AFA WA102
Av. within-group LCB length (kb)	4.89	4.89	4.88	4.51	4.54	4.61	4.14	9.32	9.11	7.47	9.6	8.42	8.33	8.47	8.39
% genome in LCBs >10 kb	45.0	49.2	48.8	26.6	27.5	25.8	22.8	60.7	58.7	43.7	63.2	60.1	60.9	58.4	49.2
Total Tn genes (kb)	28.7	13.9	12.7	25.9	21.1	47.7	69.3	30.9	104.1	92.3	43.2	21.0	17.3	31.3	59.0
Total HNH genes (kb)	6.3	1.7	2.7	3.3	3.8	6.1	7.8	6.2	8.0	7.7	6.7	2.1	3.4	3.2	4.4
% genome in Tn + HNH genes	0.71	0.35	0.35	0.63	0.53	1.08	1.45	0.66	1.94	1.74	0.85	0.50	0.47	0.73	1.16
% genome as repeats >100 bp ¹	(0.14)	(0.53)	(0.77)	—	(0.05)	(0.04)	5.11	(0.09)	8.34	(7.84)	(5.55)	(0.11)	(0.11)	(0.08)	(0.07)
No. regions of possible phage origin ²	0 (4)	0	1 (3)	(0) 2	0	1 (4)	3 (5)	1 (2)	3 (4)	0 (1)	0 (2)	2 (5)	0 (1)	0	0 (8)
Prophage-like regions (kb, #) ³	—	—	8.1	—	—	—	60.3 (4x)	—	26.5 (3x)	—	—	87	—	—	—

LCB, locally collinear block (Darling et al., 2010); Tn, transposon; HNH, HNH homing endonuclease

Tn and HNH-HE genes and repeat regions were identified by annotations

¹Numbers in parentheses refer to values for draft genomes, for which Illumina short-read sequencing can severely underestimate the presence of repeat sequences

²Number of prophage fragments identified by PHAST; no intact prophages were found. The numbers in parentheses refer to the number of phage-like fragments identified by VirSorter, some of which may be contaminating contigs representing phage fragments.

³Prophage fragments identified by PHAST

See Suppl Table 5 for more details