

Supporting Fig. S6: Upstream sequences of DIF1-motif-containing transcriptional start sites in *Anabaena* PCC 7120

Coord ^a	Gene ^b	Pos ^b	Sequence ^c	Log ₂ (TS ₁₀) ^d	Reads _{8h} ^d	Log ₂ (8h/0h) ^d
94186	F <i>alr0091</i>	+1127	TTATGGCAAGTT TCC TGATCATGGTCGCTATCTTCGGTAAAGTAATCACAGGCTTTAG	4.62	2742	4.64
217724	F <i>alr0201</i>	+59	TATCGCTGTCCA TCC TGATAATCCCCAAGAGTCGCACAATAGAGGAAATAAGGTCAGC	-1.69	127	6.50
238644	R <i>alr0222 (gltA)</i>	+473	GCCGCCAAAGCA TCCGGT TCTTTCTCGTTTGGAGCATGTAGAGAAAATTGGCGGAATAA	0.04	254	5.91
333637	F <i>all0293</i>	+1468	TTCAACGGAACT TCCGGC TCTATGGGTAGATCAGTTGATTGTGTTAGTACAACCGGGT	0.88	55	5.31
518833	R <i>all0438</i>	-184	TCCCCAATTCAA TT CGGAATATTCTATTAATCTGGAATAATAAAATTATCTAGCTCT	3.19	1019	7.91
580704	F <i>asr0485</i>	-23	GAGGTATTATTG TCCGGA TATTTGTCTTTCACTGCGAAAAAAATTATAGATTCTCCT	0.06	335	5.57
1088088	R <i>all0935</i>	-23	ATCATGATTTAT TG CGGATAAAGCATTATATCTGTGGAAAGAACTATAGATTTAGC	-2.21	220	3.97
1110714	R <i>all0955</i>	-28	GTTATGCAGAAA TCC AGATTTTATTATTCCTGGTGTAAAAGCTAATGAGAGCAAAC	2.23	75	3.43
1300124	R <i>all1110</i>	+566	GCCGCCTACTTTG CCGGA TGGAATGAAATTTTGTCTCTAATTTTGTAGGTATCGCC	-0.66	49	5.14
1443442	F <i>alr1222</i>	+1174	AAACCGTTGAAG TCC TGATTGAGCAAGAAAATCCTGAAAGTGGGAAATTAATTGGTC	-2.98	69	3.63
1590546	R <i>all1339</i>	+389	TTTAAAAGAAGAA CCGGA TGCCTCTATTGGTGACGTGTTTTTGTATTAAGTTATATA	-1.41	77	5.79
1612522	R <i>alr1358</i>	+524	TAAAACGGTTT TCCGGT GTCTTTTCTAAATGTCTGAAAAATAAATGTAACGCCAG	3.95	170	5.33
1688255	R <i>all1425</i>	+895	CGGTTAGACGTT TCCGGA AAGAGTTCAATTTTCGCAAGAACATTACTTATATGTCTTGC	2.33	237	7.39
1904547	R <i>all1617</i>	+1206	TATTAGCGGTTT TCC TGATGGAACCTTCCGCCCATGCAGAACTTAAGCAAAGTGCAG	-2.09	93	4.05
1999834	F <i>alr1677</i>	-214	ATTTGCTAAATT TCC CGATTGATCCAGTAATTAAGTTAATTAACCCTGGGTAAAA	-0.31	122	4.86
2022661	R <i>all1692 (sigC)</i>	-328	ATTAGAAAATCG T CGGAATTACTTATATACCCATGTAGATGTGACTATGCCGATAA	-0.52	4412	4.40
2084891	R <i>all1731</i>	-394	AGTTTGCTTTCC TG CGGACGCCACTCACACCTGTAGGGTATTACTAAGGATGAATT	2.59	61	5.45
2101843	R <i>all1747</i>	-42	GGCAACATAATC TCC TGATGGACAATCAGTATCACTGTAAGATTTAATCGGAGTATTC	5.25	715	8.98
2134180	F <i>asr1775</i>	-65	ATAAAACCTTTT T CGGATCTCTCTCAAGATTCTTGAAGAAGTGTACTAGTGTGTCT	-1.28	164	3.87
2155896	F <i>alr1793</i>	-21	ATGTCAATTTTC TCC AGATAAAAAATAATTCGCCTATCATAGGGTGAGTTTTGGGAAT	5.15	131	3.96
2821366	F <i>alr2339 (hetR)</i>	-272	GAGCAGATAAGT TCCGGA TAATAGGGAAGATCCTGTAGGTTACTTATTAGAACCAT	2.50	1898	3.55
2881051	R <i>all2396</i>	-337	GGTATTTCTGTA TCCGGA TAATTAACATGCCAAGTGAATTTCTTAACAGTAGTGTGT	1.56	50	3.17
2913386	F <i>alr2428</i>	+1405	TGCCCATACTCT TCC TGACGGTTCACAAGATGGGAAATTTGTCAATCAGGTATTTA	1.89	386	4.51
2975615	F <i>alr2478</i>	-187	GTAATCAAAAAT TG CGGAATTTGCCCAACCATTGAGAACTAATAAAGCAGCGAACAG	-2.70	41	3.31
2984123	F <i>alr2483</i>	+212	GCGGCTGTTGAT TG CGGATATAGAAGCGCCTCGGCAGAATATAGCTTTAGAACAGGAT	2.67	51	5.20
3072950	R <i>all2571</i>	-127	TAAGAAAAAGTT TG CGGATTTGAGTCCCAATGAGTATCTAAGTTGTTAGTCTTACC	-2.38	1581	6.22
3449710	F <i>alr2833</i>	-110	GCATTATTAGAG TCCGGA GAAATCACTGGGATTATGAAAAATTCGTAATCATGCTG	3.95	467	6.79
3453831	F <i>alr2835 (hepA)</i>	-407	GAATTAGGTTTA TCC TGAAGAGTAAAATAAATCCGAATATCCTAAATTAATGAGGA	4.02	94	6.07
3457231	F <i>alr2837</i>	-105	TTCCGATGTTCA T CGGAAGCACCACAATTTAGCCGTAAGTATGTTTAGCCGTTATC	-0.70	800	3.42
3569154	F <i>alr2933</i>	-27	TTGTGCGCCCTT CCGGA TGTAGCGATCCGAGTAGAACCTGTTTTTAGTTCTAGTG	1.20	194	3.30
4066829	F <i>alr3361</i>	-298	TTGATGAAAATGC CCGGA AGTTAATACATTTAAAGAGCTAATTGACAGTAAGTATAAA	-0.66	119	6.41
4113609	F <i>all3402</i>	-233	ACACCCATTTCT TCC TGACCTCTGTGTTTCACCCACGAATAAACAATTTAACGCATTTC	-0.69	289	3.68
4273242	R (<i>nsiR1</i>)	0	TCTACAACCTTA TCCGGA TGTGTTAGCATCTGCCCAAGACAGTTTAAACGGTAGATGCA	0.50	368	4.71
4354143	F <i>alr3602</i>	+1407	TTTATCGGGACA TCCGGA CAACCAATTCGCGCACAGAGTTAAAATCTAGACCCTG	0.55	112	4.00

4408600	F		CAAAAGTATTTA	TCCGGA	TATAGGATGGGCGCGAAGCTCTTATAGTTCA	AGTCCATACA	0.70	71	5.67	
4431362	F	<i>all3675</i>	+42	TTAACGGAAATG	TCCGGA	TTTTTTTTGTCAATTTGGCTAATAACTTTAAAGGGAATGAT		1.42	60	5.43
4601709	F	<i>alr3808 (dps)</i>	-317	TCTGATAATTTT	TCCGGA	GAACACCATTATTTACAAGTAGAGTGTGATGTCTAGGAAA		3.93	252	3.90
4616158	F	<i>alr3817</i>	-118	AGAACTCCTGAA	TCCGGA	AAATCTTCTGGGGCTTTTTGCTGTCTAGTGGTATAGGGGTGG		-2.42	46	5.05
4792909	F	<i>alr3979</i>	-831	TCCAGCCTGTCA	TCCGGA	ACGTCTGGGCAATTGTCCGGAGATTAGAAGAAGATGATGG		1.20	76	5.77
5007675	R	<i>alr4161</i>	+48	AAATTTGACGTG	TCCGGA	CATTTCAGTCCTTCCCCTTGAATAGCAGAAAAGACCTCAAC		-1.69	172	3.13
5056900	R	<i>all4220</i>	-38	AAATCAACATTA	TCCGGA	ATTTTAGTGTAGATAATAGAAATGTGAGGTAAAGAACGGA		-0.23	1059	6.74
5150808	R	<i>all4296</i>	+1251	AATGTTTCAATTA	TCCGGA	CTGCGCCCCCACAAATTTGAAAAATTATGACATTGAGCGA		3.02	641	6.83
5199210	R	<i>all4343</i>	-484	GGCTAGAAATAA	TCCGGA	TTTAGCAGCAAAATCAGCGTAGTACAATATATATACCCCTTT		-0.58	67	3.59
5248692	R	<i>all4379</i>	+998	GGTTTGCCTTTA	TCCGGA	TCTGCACTGTGTAAGTTTGAAGATATGACAGTTAATC		1.68	2138	6.98
5742628	R	<i>all4822</i>	-34	CCCACAGCACCT	TCCGGA	AAATTAATAAA-AAACGCCGAAAAATATTGCAATAATCTCT		3.95	95	3.50
5774872	F	<i>alr4847</i>	+609	TGGATGAAAATG	TCCGGA	AGGTTAGCTA-AAAATGCAGATATTCTAATTTACGATTCCA		4.36	59	3.41
5953754	F	<i>alr4984</i>	-387	TACACAATTTTA	TCCGGA	TATATAACCGCTTCAGGTGGAATAAATTTCTTATTATCTAG		-0.69	58	5.38
5954131	F	<i>alr4984</i>	-25	AGTAGACGTTAT	TCCGGA	ATAATTGAT-TTTTTTCTGTCAAAGGAATTATGGGTCATA		-2.56	235	7.38
6031913	R	<i>alr5059</i>	+286	GTTTGGCGTTCT	TCCGGA	ACGCTTGTTAGGATTTGTGAAGGTTTATCAAGAATGCGGG		3.02	110	6.29
6190304	F	<i>all5192</i>	+254	ATCATCACCAAA	TCCGGA	ATCCCAGAG-AGATTTCTGTAAGTCACTAATACTATGGAAATC		3.02	83	3.89
6271762	F	<i>alr5256</i>	+612	CTAGGAATAGAG	TCCGGA	TATGGCGTGCCTTTTCCAAGAAGGAATTAGCATACTCTTGT		3.02	161	6.84
6279444	R	<i>all5263 (sigA)</i>	-605	CCTATTCAATTC	TCCGGA	CTAATTTAT-TTATGAAGGAAATGTAAAAATGGGTGCGTT		3.02	53	3.67
6393337	R	<i>alr5354 (hglD)</i>	+134	AAATAAAATTTT	TCCGGA	CTTGCCTTTT-ACTGGGTCGTAATAATATTGTGGGATCTACA		4.91	116	6.37
6411861	R	<i>all5369</i>	-21	TGGGAATTATGT	TCCGGA	TGACACTCGTGCAAGCAAGTCTATGTATAAAAATAGAACC		3.02	115	3.04
β:96684	F	<i>all7605</i>	+335	TGCAATTAACCG	TCCGGA	TACCATCGACAAAACGGATGTAATATTGCTTATCTTTGTAT		6.03	48	5.51

Excluded DIF⁺ regions

942986	F	<i>alr0819*</i>	-266	GTAAAGCATAGT	TCCGGA	AAACAATAATGGAACGACAGATAATCTAAAAATATATCTTT		3.02	113	4.75
3747939	F	<i>alr3096*</i>	-81	ACAACGTGGCAT	TCCGGA	CAACCGTATGCGATAAACTGCTGCTAAACTAGCAACTGTA		5.46	200	3.24
3990504	R	<i>alr3307*</i>	+327	ATATTCTGCTGCA	TCCGGA	TACAGATATTGTGGCTGGAGTGTGCCAATATCGATAGAC		2.51	80	5.84

patS and *patX*

2771633	R	<i>asl2301</i>	-633	TGCATTTAACTA	TCCGGA	TTTTTTACC-AAATAACTGTAGAGAAAAGTATTAAGTAAA		0.04	47	2.08
2811207	R	<i>asl2332</i>	-24	TTTTTAGATTCT	TCCGGA	CTCGTTATATAAACTCAGTAGCAATAGATAAGAGAATTT		0.40	528	1.95

- ^a Coordinates of the transcriptional start sites are all from the *Anabaena* PCC 7120 chromosome, except for one from its beta plasmid. The sequence is rightward, either forwards from the coordinate (F) or backwards (B) to the coordinate.
- ^b The position of the coordinate is relative to the translational start site of the given gene. Positive positions indicate that the sequence lies within the gene at the given position. Needless to say, a promoter lying within a gene cannot drive the transcription of that gene. The genes are given to identify only the positions of the sites. *nsiR1* determines a non-coding RNA (Ionescu *et al.*, 2010). Where no gene is given, no downstream gene parallel to the DIF⁺ region has been identified. The genes shown are from Mitschke *et al* (2011), Suppl. Table 7, those with upstream DIF⁺ motifs and expression levels 8-hrs after nitrogen stepdown at least 8x higher than at 0-hrs. In addition, *patS* and *patX* are shown for comparison.
- ^c The DIF1 motif is highlighted in blue. Nucleotides highlighted in red are taken from the coordinates given by Mitschke *et al* (2011) as transcriptional initiation sites. Nucleotides highlighted in green are the most common in a column with the information content greater than or equal to 1. Those highlighted in cyan are the most common in a column with the information content between 0.4 and 1. The nucleotide frequencies are shown graphically in Fig. 7D.
- ^d Data taken from Mitschke *et al* (2011), Suppl. Table 1, indicates for each region (**Log₂(TS₋₁₀)**) the quality of the -10 region judged by a PSSM trained on -10 regions from Ana7120 tRNA genes, on a log base 2 scale; (**Reads_{8h}**) the abundance of transcripts with the indicated 5' end at 8 hr after N-stepdown, given as the number of reads detected; (**Log₂(8h/0h)**) the induction of transcription at the indicated site, comparing RNA abundance at 8-hr after N-stepdown to abundance before N-stepdown, on a log base 2 scale.

References

Ionescu, D.I., Voß, B., Oren, A., Hess, W.R., Muro-Pastor, A.M. (2010) Heterocyst-specific transcription of NsiR1, a non-coding RNA encoded in a tandem array of direct repeats in cyanobacteria. *J Mol Biol* **398**: 177-88.