

Comprehensive Analysis of Tandem Repeat Sequences in Cyanobacteria Genome

Toshiaki Katayama¹

katayama@kuicr.kyoto-u.ac.jp

Shinobu Okamoto²

okamoto@bio.c.u-tokyo.ac.jp

Rei Narikawa²

cc16733@mail.ecc.u-tokyo.ac.jp

Takatomo Fujisawa²

cc97724@mail.ecc.u-tokyo.ac.jp

Shuichi Kawashima¹

shuichi@kuicr.kyoto-u.ac.jp

Masumi Itoh¹

itoh@kuicr.kyoto-u.ac.jp

Masayuki Ohmori²

cohmori@mail.ecc.u-tokyo.ac.jp

Minoru Kanehisa¹

kanehisa@kuicr.kyoto-u.ac.jp

¹ Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

² Department of Life Sciences, University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153-8902, Japan

Keywords: genome analysis, repeat sequence, cyanobacteria

1 Introduction

Over the past year, two complete genome sequences of the cyanobacteria (filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. PCC7120 [3] and thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1 [6]) became available in addition to the *Synechocystis* sp. PCC6803 genome [4]. Additionally, five species are in draft to finishing phase and other eight genome projects are also in progress. Among these genomes, some repeat sequences have been identified, however, their functions are not yet characterized well [5].

In this research, we have applied comprehensive search for tandem repeat sequences on each cyanobacterium genome. From this screening, we have found new tandem repeat units in addition to previously reported repeat sequences such as STRRs (short tandemly repeated repetitive) and LTRRs (long tandemly repeated repetitive) sequences [5].

	Tandem repeat sites	Non-redundant repeat units
<i>Anabaena</i> sp. PCC7120	942	294
<i>Synechocystis</i> sp. PCC6803	49	30
<i>Thermosynechococcus elongatus</i> BP-1	97	62
<i>Synechococcus</i> sp. WH8102	60	35
<i>Prochlorococcus marinus</i> sp. MED4	47	29

Table 1: Number of tandem repeats and identical repeat units found in cyanobacteria genomes.

2 Method and Results

To identify tandem repeats in the cyanobacteria genome sequences, we used TRF (Tandem Repeat Finder) ver 3.01 [1] and then the results were filtered and clustered into non-redundant repeat units by merging circularly shifted patterns and/or considering complementary sequences as the same unit.

Table 2: Highly represented repeat units in *Anabaena* sp. PCC7120.

Times	avg	len	bp	STRU	Consensus
77	4.5	7	2156	STRR7	AAAATTC
53	7.9	7	2968	STRR1a	AATCCCC
36	6.4	7	1512	STRR1g	ACTGGGG
35	6.7	7	1715	STRR3g	AACAGTC
26	5.4	7	910	STRU11	ACAACCTG
21	6.0	7	882	STRR2t	AATGACT
17	4.9	7	595	STRU12	ACCAATT
16	5.7	8	768	STRU49	ACACCCTT
14	4.8	8	560	STRU48	ACACCCTT
14	3.6	14	784	STRU129	CCCAATCCCCAGTC
13	5.2	7	455	STRU13	ACCCATT
13	4.8	8	520	STRU54	AGGGGTAT

As a result, only *Anabaena* has an excessive amount of tandem repeats in the genome as shown in Table 1. Partial list of non-redundant repeat units sorted by the occurrence frequency in the *Anabaena* sp. PCC7120 genome is shown in Table 2. Among the known repeat units (STRR1–7), we determined new highly represented oligomers named STRU (short tandemly repeated units). These units are useful for genotyping novel *Anabaena* and *Nostoc* strains by PCR fingerprinting [2].

Acknowledgments

This work was supported by the grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

- [1] Benson, G., Tandem repeats finder: A program to analyze DNA sequences, *Nucleic Acids Res.*, 27:573–580, 1999.
- [2] Guevara, R., Armesto, J.J., and Caru, M., Genetic diversity of *Nostoc* microsymbionts from *Gunnera tinctoria* revealed by PCR-STRR fingerprinting, *Microb. Ecol.*, 44:127–136, 2002.
- [3] Kaneko, T. *et al.*, Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120, *DNA Res.*, 8:205–113, 2001.
- [4] Kaneko, T. *et al.*, Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions, *DNA Res.*, 3:109–136, 1996.
- [5] Meeks, J. C., Elhai, J., Thiel, T. Potts, M., Larimer, F. Lamerdin, J., Predki, P., and Atlas, R., An overview of the genome of *Nostoc punctiforme*, a multicellular, symbiotic cyanobacterium, *Photosynth. Res.*, 70:85–106, 2001.
- [6] Nakamura, Y. *et al.*, Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1, *DNA Res.*, 9:123–130, 2002.