Excess Information at T7-like Promoters and Classification of T7-like Phages

Zehua Chen, [1] **Thomas D. Schneider,** [2]

Keywords: T7-like promoters, T7 group, excess information, individual information.

1 Introduction.

Transcription plays a key role in the expression of genetic information. Most double-stranded DNA bacteriophages utilize the transcription system of their hosts, while the T7 group of phages mainly utilize their own transcription system, which consists of a single subunit RNA polymerase (ssRNAP) and a set of promoters located on the phage genome. Molecular information theory can be used to precisely characterize the sequence conservation at DNA binding sites, and has been widely applied to many genetic systems. As early as 1984, molecular information theory was applied to analyze the T7 promoters. It was noted that a roughly two fold excess information exists at T7 promoters [1]. In this work, we extended analysis to the other T7-like phage promoters and built promoter models for seven phages. The results show that excess information exists for all seven models.

Besides the ICTV taxonomic system and a genome-based classification which was recently proposed [2], some unique features could also be used for classification for certain group of phages. We propose that the T7-like promoters and the phage specific RNAP are two key features which can be used to classify a phage as a member of the T7 group or not.

2 Software and Methods.

Most programs used in this work are available at `http://www.lecb.ncifcrf.gov/toms/`. Promoter models were built with the programs **delila**, **alist**, **encode**, **rseq**, **dalvec** and **makelogo**. The programs **scan** and **lister** were used for genome scanning. **Genhis** and **genpic** were used to plot individual information distribution. Phylogenetic analyses were conducted with the programs **diffribl**, **neighbor** and **drawtree**.

3 Results and Discussion.

A total of seven promoter models were created for phages T7, ϕA1122, T3, YeO3-12, gh-1, K11 and SP6, and different combined models were also built. The ϕA1122 model is almost the same as the T7 model and the YeO3-12 model is almost the same as the T3 model. The other three models are diversified except for the DNA region from -7 to -4. Information analysis shows that excess information exists for all seven promoters, with excess ratios ranging from 1.52 to 1.85 (Table 1).

The seven models and the combined 93-sites model were used for genome scanning. When a model was used to scan its own genome or a closely related genome, a significant gap of individual information distribution was observed between real promoters and background. When the combined model was used to scan the seven phage genomes and other possible T7-like phages, ϕKMV, P60, VpV262, SIO1 and PaP3, 92 of the 94 promoters (including

[1]LECB, National Cancer Institute at Frederick, Frederick, MD. E-mail: `chenze@ncifcrf.gov`
[2]LECB, National Cancer Institute at Frederick, Frederick, MD. E-mail: `toms@ncifcrf.gov`

one T7 promoter in the T3 genome) above 12 bits can be picked up, while no sites above 12 bits can be picked up for the other phages, indicating that they may not belong to the T7 group. Also, a total of 17 host genomes and closely related genomes were scanned with each of the models, and 15 T7-like promoters were identified in 6 pathogens.

Phylogenetic analysis with the T7-like models show that the eight T7-like phages can be classified into five subgroups. Combining with some data from literature, we propose that T7-like phages can be classified as shown in Figure 1.

Phage	Number of sites	Rs (bits)	Range	Host	Genome size(Mb)	Rf (bits)	Rs-Rf (bits)	Rs/Rf
T7	17	34.9	-20,+5	*E. coli*	4.64	19.1	15.8	1.83
ϕA1122	17	35.2	-20,+5	*Y. pestis*	4.60	19.1	16.1	1.85
T3	14	33.6	-18,+5	*E. coli*	4.64	19.3	14.3	1.74
YeO3-12	15	34.2	-18,+5	*Y. enterocolitica*	4.62	19.2	15.0	1.78
gh-1	10	33.5	-18,+5	*P. putida*	6.18	20.2	13.3	1.65
SP6	11	32.6	-18,+5	*S. typhimurium*	4.86	19.8	12.8	1.65
K11	9	31.0	-17,+4	*Klebsiella*	6.0	20.4	10.6	1.52

Table 1: Excess information at T7-like promoters.

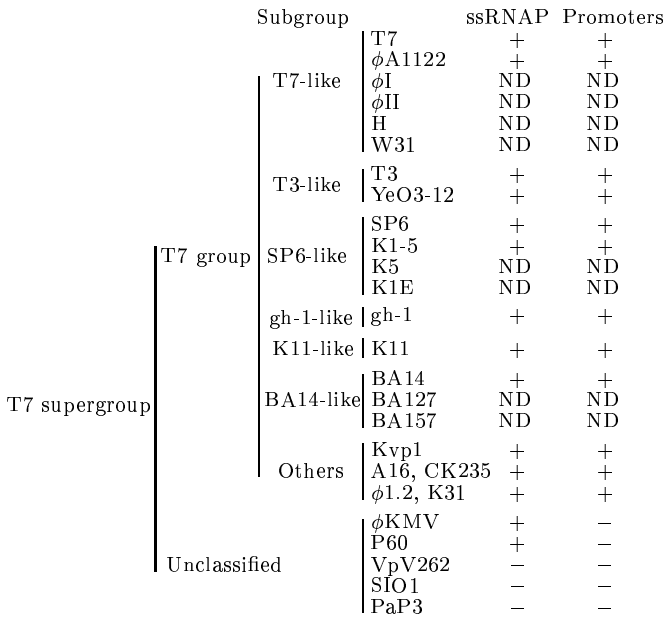

Figure 1: Classification of T7-like phages

References

[1] Schneider, T. D., Stormo, G. D., Gold, L. and Ehrenfeucht, A. 1986. Information content of binding sites on nucleotide sequences. *J Mol Biol*, 188: 415-431.

[2] Rohwer, F. and Edwards, R. 2002. The phage proteomic tree: a genome-based taxonomy for phage. *J Bacteriol*, 184:4529-4535.