

# Evaluation of Standardized Methods in the Analysis of Difficult DNA Repeat Sequences: A follow up to the ABRF 2003 Sequencing Survey.

Glenis. J. Wiebe<sup>a</sup>, Rashmi Pershad<sup>b</sup>, Helaman Escobar<sup>c</sup>, John W. Hawes<sup>d</sup>, Timothy Hunter<sup>e</sup>, Emily Jackson-Machelski<sup>f</sup>, Kevin L. Knudtson<sup>g</sup>, Heather Lin<sup>h</sup>, David S. Needleman<sup>i</sup>, and Margaret Robertson<sup>j</sup>

<sup>a</sup>Max Plack Institute of Molecular Cell Biology & Genetics, Dresden, Germany, <sup>b</sup>UT MD Anderson Cancer Center, Houston, TX, <sup>c</sup>Brigham Young University, Provo, UT, <sup>d</sup>Miami University, Oxford, OH, <sup>e</sup>Vermont Cancer Center, Burlington, VT, <sup>f</sup>Washington University School of Medicine, St. Louis, MO, <sup>g</sup>University of Iowa, Iowa City, IA, <sup>h</sup>University of Utah, Salt Lake City, UT, <sup>i</sup>UT Houston Medical School, Houston, TX, <sup>j</sup>University of California San Francisco, Emeryville, CA

## ABSTRACT

The ability to successfully sequence through regions containing long nucleotide repeats continues to challenge many DNA sequencing facilities. Consequently, the DNA Sequencing Research Committee (DSRG) of the Association of Biomolecular Resources Association (ABRF) recently conducted a survey in which DNA sequencing laboratories were invited to sequence three highly repetitive templates. Participants were asked to submit their sequencing results along with information regarding reaction conditions, chemistries, additives, and instruments. Many interesting trends emerged, which were presented at the ABRF 2003 meeting in Denver, Colorado (1). However, since such a wide variety of protocols were used, and since the number of participants using any given protocol was small, it was not possible to attribute any statistical significance to any particular parameter. Moreover, it was not completely clear to what extent success was related to the protocol used or the experience of the laboratory itself. Therefore, the members of the DSRG designed a standardized protocol and conducted an internal study with the same three difficult templates, in order to focus on lab-to-lab variability, and how this relates to previously obtained results.

1. Hawes, J.W., et al. JBT 14: 101 (abstract).

## METHOD

The following standardized protocol used in this internal study was based on successful sequencing of the same templates by participants in the ABRF 2003 study. To minimize variation, plasmid templates were prepared at a single location using Qiagen Maxi-Prep kits, analyzed by absorbance measurements at 260 and 280 nm, and distributed along with a common primer to the seven participating labs.

### Reaction:

300 ng template A, B or C  
5 pmol M13 forward primer  
4 µl BigDye v3.1 (Applied Biosystems)  
2 µl 5x Sequencing Buffer (Applied Biosystems)  
H<sub>2</sub>O to 20 µl

### Cycling:

initial denaturation: 95°C for 5 minutes  
30 cycles: 96°C 10 sec, 50°C 5 sec, 60°C for 4 min.  
rapid ramp and hold at 4°C

Post-reaction clean up was by gel filtration or ethanol precipitation. Samples were resuspended in water or formamide, and run on an Applied Biosystems 3100 DNA Sequencer with either 36, 50, or 80 cm capillary arrays, using instrument defaults for injection and run conditions. Phred q20 scores (the number of bases with a quality value >20) were used for data analysis. Results were then assigned to three of four categories (bins) based on the positioning of different repetitive motifs within the templates (see electropherograms below and Table 3), as a way of monitoring the success at negotiating each of the difficult regions. For example, a sequence for Template A, with a q20 of >425, has read through the entire repeat. On the other hand, a sequence with a q20 between 301-425 drops off or fails at some point due to the repetitive nature of this region.

**Table 1.** Top 5 sequences from the ABRF 2003 study.

Template	q20	Chemistry	Mixed?	Instrument
A	837	BDT v3.1	no	377
	815	BDT v3.1	no	3100
	767	BDT v3.1	no	3700
	765	BDT v3.1	no	3100
	765	BDT v1.1	no	3100
B	951	dGTP v3.0	no	377
	879	BDT v3.0	dGTP v3.0, 4:1	3100
	871	BDT v3.0	dGTP v2.0, 2:1	3100
	856	BDT v3.1	no	377
	843	BDT v3.0	no	3730
C	908	dGTP v3.0	no	377
	881	dGTP v3.0	no	3100
	835	BDT v3.0	dGTP v2.0, 2:1	3100
	820	BDT v3.1	no	377
	811	dGTP v3.0	no	3730

**Table 2.** Sequencing results from the internal study. Templates A, B, and C were prepared according to the standardized protocol and run on a 3100.

Template	Precipitation Method	Loading Media	Array Length	q20
A	CentriSep-8	water	80	830
A	G50 Sephadex	water	80	765
A	Millipore	water	50	613
A	Edge	Hi-Di	50	526
A	Edge	water	50	501
A	EtOH	water	50	440
A	EtOH	Hi-Di	50	429
A	Edge	water	50	376
A	EtOH	water	36	316
B	G50 Sephadex	water	80	920
	CentriSep-8	water	80	703
	EtOH	water	50	641
	EtOH	Hi-Di	50	637
	Millipore	water	50	579
	EtOH	water	36	546
	Edge	Hi-Di	50	395
C	G50 Sephadex	water	80	904
	CentriSep-8	water	80	849
	Millipore	water	50	635
	Edge	water	50	629
	Edge	Hi-Di	50	624
	EtOH	Hi-Di	50	617
	EtOH	water	50	593
C	Edge	water	50	354
	Edge	water	50	169
	EtOH	water	36	32

Two important trends were noted in the ABRF 2003 study: 1) the top two results from each template were achieved with a 377 and 3100, respectively, and 2) the BigDye v3.1 or BigDye dGTP kits (alone or in a mixture) were generally better at dealing with repetitive samples (Table 1). Buffers, additives, purification methods, loading media and cycling conditions, among others, did not seem to play a significant role. However, since each sample submitted used a slightly different protocol, it was difficult to attach much significance to each variable. The study presented here was designed to minimize the protocol differences as much as possible, in order to focus on lab-to-lab variability.

The results are summarized in Table 2. Perhaps not surprisingly, the sequences from the longer 80 cm capillaries, which allow for higher resolution, occupy the top position in each set. The 36 cm capillaries, providing the lowest resolution, actually worked remarkably well in the case of template B, but otherwise did not perform as well as the others. Please note that for samples run on the 36 cm array, the instrument default injection time was not suitable, and needed to be lowered from 30 secs to 15 secs. As per the ABRF 2003 study, purification methods and loading media were not found to affect sequence outcome. Overall, there is a higher success rate from the internal study compared to the ABRF study (Table 3), and we suggest that the standardized protocol presented here is a good starting point for sequencing through highly repetitive regions.

Despite our attempts to control the experimental parameters, it is apparent that there is still quite a variation in the quality of the sequences obtained, indicating that success is determined to a large extent by the laboratory conditions. Further sources of variation to consider in future studies are instrument fidelity, capillary array usage, reagent quality, and technical experience.

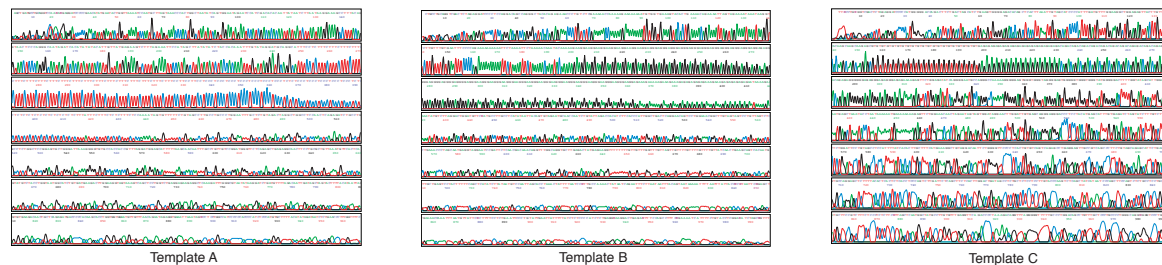
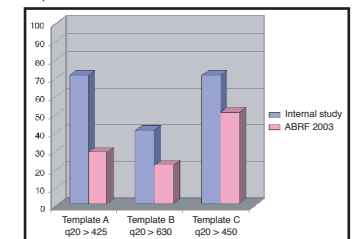
For more information on the ABRF 2003 study, please download our report "Sequencing through Difficult Repetitive Sequence" from the ABRF web-site:

<http://www.abrf.org/ResearchGroups/DNASequencing/EPosters/DSRG2003Study.pdf>

**Table 3.** A comparison of binned sequencing results obtained from the internal and ABRF 2003 studies.

# Bases > q20	Number of Samples (Percentage)	
	Internal Study	ABRF2003 Study
Template A		
> 425	7 (70%)	33 (28%)
301 - 425	2 (20%)	69 (58%)
< 300	1 (10%)	16 (14%)
Template B		
> 630	4 (40%)	25 (21%)
421 - 630	2 (20%)	21 (18%)
301 - 420	2 (20%)	15 (13%)
< 300	1 (10%)	58 (48%)
Template C		
> 450	7 (70%)	62 (50%)
450 - 321	1 (10%)	38 (30%)
200 - 320	0	9 (7%)
< 200	2 (20%)	16 (13%)

**Figure 1.** Percentage of successful samples for each of the three templates from both the internal and ABRF 2003 studies.



## ACKNOWLEDGEMENTS

Thanks to all who participated in the ABRF 2003 Study, to George Grills for providing the clones, and to James Van Ee for Phred analysis.