

# Results from the 2008 DSRG Difficult Template Sequencing Study

Bintzler, D.<sup>1</sup>, Adam, D.<sup>2</sup>, Detwiler, M.<sup>3</sup>, Kieleczawa, J.<sup>4</sup>, Needleman, D.<sup>5</sup>, Schweitzer, P.<sup>6</sup>, Singh, S.<sup>7</sup>, Steen, R.<sup>8</sup>, Zianni, M.<sup>9</sup>

<sup>1</sup> University of Cincinnati, <sup>2</sup> University of British Columbia, <sup>3</sup> Roswell Park Cancer Institute,

<sup>4</sup>Wyeth Research, Cambridge, <sup>5</sup> USDA-ARS-Eastern Regional Research Center, <sup>6</sup>Cornell University, <sup>7</sup>University of Minnesota, <sup>8</sup> Harvard Medical School, <sup>9</sup>The Ohio State University

## Abstract

In the past, the DNA Sequencing Research Group (DSRG) has conducted research studies on DNA templates that were moderately GC rich (DSRG 1997 study) or contained repeat elements (DSRG 2003 study). Though improvements in DNA sequencing chemistries have helped with difficult templates, one of the remaining challenges for classical Sanger DNA sequencing is the ability to effectively sequence through various types of difficult regions in DNA templates. In the 2008 DSRG difficult template sequencing study, we have expanded the number of templates and nature of the difficult sequences, and have designed the study to identify and develop optimal protocols for such templates. The templates in this study include a moderate GC-rich region (up to 72%), a very GC-rich region (up to 95%), a long non-repeat di-nucleotide region, an Alu-repeat, a 19 base long G/C homopolymer, and a hairpin-containing template. The same set of difficult templates (prep method/concentration/primer) were distributed to all willing participants and the study will be carried out in 2 phases:

**Phase I:** Participants will apply a protocol of their choice, and each condition will be performed in triplicate, and the data will be returned to the DSRG for analysis.

**Phase II:** The DSRG will recommend optimal protocols for each template and participants will re-sequence all templates in triplicate utilizing the 3 best protocols selected from phase I.

The DSRG has a long and successful track record of conducting inter-laboratory research studies to optimize protocols for DNA sequencing, providing benchmarking opportunities, as well as resources for troubleshooting DNA sequencing reactions and instrumentation. Here we present the results of the DSRG 2008 difficult template study with the goal of defining optimal sequencing protocols for these types of difficult templates.

## Methods

### Phase I

DSRG selected 8 difficult templates for the study as follows...

- DNA 1, 2 - GC rich regions
- DNA 3 - region containing 24 bp hairpin
- DNA 4 - 18 C base region separated by 7 bases
- DNA 5 - 456 base non-repeating TC (forward) GA (reverse)
- DNA 6 - 147 base non-repeating GA (forward) TC (reverse)
- DNA 7 - 19 base and 15 base inverted repeats and 19 base poly G
- DNA 8 - Alu repeat with a 22 base inverted repeat and 84 base loop
- DNA 9 - pGem, as a control

30 sets of samples were sent to ABRF members. The samples were sequenced using in-house methods. Chromatograms were uploaded to an FTP site at <ftp://ftp.genetics.utah.edu>, and Q >20 results were analyzed using Sequence Scanner v1.0. PCR cycles, reaction additives, instrumentation and cleanup methods were compared to determine the protocols most successful in sequencing through the difficult regions.

### Phase II

Participants were provided the protocols determined to best sequence through each difficult region. Participants were asked to repeat the samples using these protocols and resubmit the new results for comparison. Results from Phase I and Phase II were compared.

## Results

**Table 1:** Optimal protocols for DSRG study samples identified from Phase I. These protocols were determined to provide the best sequencing results.

Prot #	DNA #	Primer	BD	dGTP	5X ABI	Additive	Preferred	Cycling Protocol
		μl/5μM	μl	μl	μl	μl	Cleanup*	
1	200	1.0	1.5	0.5	0	Betaime <sup>2</sup> / <sub>2.0</sub>	10	CleanSeq [(96°C/10sec)(50°C/5sec)(60°C/4min)] x35 → 4°C∞
2	100	1.0	2.0	1.0	0	Betaime <sup>2</sup> / <sub>2.0</sub>	10	BDX 96°C/1min → [(96°C/10sec)(50°C/5sec)(60°C/4min)] x25 → 4°C∞
3	200	0.5	1.0	0.1	0	Betaime <sup>2</sup> / <sub>1.6</sub>	10	EtOH 95°C/5min → 98°C/40sec → 60°C/4min → [(98°C/10sec)(60°C/4min)] x24 → 4°C∞
4	300	1.0	3.0	1.0	0	DNIS <sup>2</sup> / <sub>1.0</sub>	20	G-50 B-house # 96°C/1min → [(96°C/10sec)(50°C/5sec)(60°C/4min)] x30 → 4°C∞
5	200	2.0	1.0	0	0	None	15	CleanSeq 95°C/1min → [(98°C/4sec)(50°C/15sec)(60°C/2.5min)] x39 → 4°C∞
6	200	1.0	1.0	0	1.5	Betaime <sup>2</sup> / <sub>2.0</sub>	10	CleanSeq [(96°C/10sec)(50°C/5sec)(60°C/4min)] x35 → 4°C∞
7	200	1.0	1.5	0.5	0	None	10	CleanSeq [(96°C/10sec)(50°C/5sec)(60°C/4min)] x35 → 4°C∞
8	100	1.0	3.0	1.0	0	None	10	BDX 98°C/5min → [(96°C/10sec)(50°C/5sec)(60°C/4min)] x25 → 4°C∞
9	200	1.0	0.75	0.25	1.5	None	10	CleanSeq 95°C/1min → [(95°C/10sec)(50°C/5sec)(60°C/2min)] x35 → 4°C∞
10	100	1.0	2.0	0	0	None	20	Edge DTR v3 100°C/2min → [(96°C/30sec)(50°C/15sec)(60°C/4min)] x26 → 4°C∞

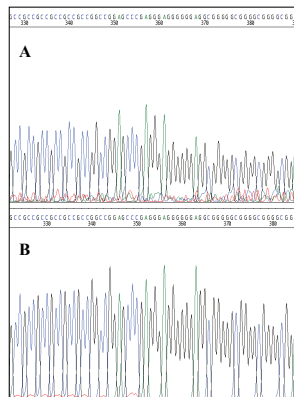
\* dGTP v1.0 may be substituted for v3.0.

\* Cleanup used for the protocol giving best results.

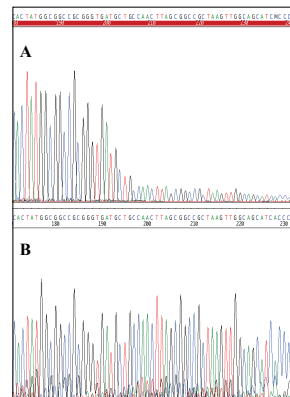
# G-50 Sephadex and Millipore MAHVN4550 plate.

**Table 2:** Assignment of protocols to specific DNA templates.

DNA #	Primer	Characteristics	Protocol 1	Protocol 2
1	F	94% GC over 200 bases/101 base non-repeat G/C	1	2
	R	90% GC over 200 bases/73% GC over next 400 bases	1	3
2	F	70% GC over 300 bases	6	8
	R	78% GC over 150 bases	6	1
3	F	24 base hairpin with T <sub>m</sub> >95°C	1	4
	R	24 base hairpin with T <sub>m</sub> >95°C	5	10
4	F	18 Cs/10 Cs, separated by 7 bases	6	1
	R	18 Gs/10 Gs, separated by 7 bases	1	10
5	F	456 base non-repeat T/C	1	7
	R	456 base non-repeat G/A	1	7
6	F	147 base non-repeat G/A	9	6
	R	147 base non-repeat T/C	6	1
7	F	19 and 15 bases inverted repeats, followed by 19 Cs and 41 base non-repeat T/A	1	3
	R	19 and 15 bases inverted repeats, followed by 19 Gs and 41 base non-repeat A/T	1	3
8	F	Alu repeat + 22 base inverted repeat/84 base loop	1	7
	R	pGem3zf	9	10



**Figure 1:** DNA sample 1 with the forward primer. Sample (A) shows an average result submitted and sample (B) shows the result for the best sample.



**Figure 2:** DNA sample 3 with the forward primer. Sample (A) shows an average result submitted and sample (B) shows the result for the best sample.

**Table 3:** A Comparison of the Q>20 Results from Phase I and Phase II for 12 facilities. Selected examples of samples with data showing better results in Phase I (blue) and data showing better results in Phase II (red).

P1	DNA 1		DNA 2		DNA 3		DNA 4		DNA 5		DNA 6		DNA 7		DNA 8																
	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2															
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0														
2	442	247	243	624	852	254	953	871	251	141	138	26	131	719	574	455	401	215	141	250	165	960	574	600	533	607	823	403			
3	168	454	117	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	551	990	921	938	938	104	104	761	1007	891	56	889	0	789	830	531	0	569	778	575	948	1083	1011	1021	853	961	819	825	748	1020	878
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	147	433	90	162	643	589	829	1058	783	29	783	0	84	25	43	102	152	510	541	892	1070	522	1053	953	487	588	801	738	782	888	
8	372	440	34	742	923	828	868	938	718	160	839	0	121	109	32	0	601	811	479	0	869	898	841	920	738	309	565	825	964	598	
9	191	552	0	1819	1804	284	203	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	218	441	188	521	518	582	889	918	814	0	787	0	84	0	86	79	569	252	588	128	1088	881	987	838	611	574	512	572	948	943	
11	152	299	16	314	533	603	679	619	438	0	331	0	118	0	211	94	547	880	0	200	874	744	70	817	688	495	417	417	287	12	
12	202	0	738	581	290	0	859	871	288	0	0	0	150	0	34	0	313	0	308	148	722	0	980	813	280	0	881	582	284	721	

## Discussion

The DSRG had designed this study as a learning experience for participants and associated laboratories. Difficult templates remain a challenge for many sequencing laboratories, and the design of the study combined the experience of multiple laboratories in determination of an optimized protocol for difficult templates.

Phase I of the study provided an opportunity to compare how the difficult samples would work under in-house protocols. For phase II of the study, participants were provided with the best protocols in Phase I (Tables 1 and 2) and allowed to resequence the samples. Table 3 compares whether a participant was able to show improvement in sequencing data. In 18 of the results the laboratory produced a better result using the in-house protocol and in 36 of the results the laboratory produced a better result using a protocol provided in Phase II. Therefore, each participant could measure the effectiveness of the in-house protocols.

The sample that proved most difficult for the participants was the 24 base pair hairpin using the reverse primer. This sample was not successfully sequenced by the participants and hairpin characteristic of the sample have proven most difficult in sequencing laboratories. However, a protocol can be designed for sequencing this type of difficult template as shown in Figure 2.

## Acknowledgements

DSRG would like to express our appreciation to the study participants whose efforts made the study possible. We would also like to acknowledge Michelle Mader, Kim Marquetten and Erica Mazaika from Wyeth Research.