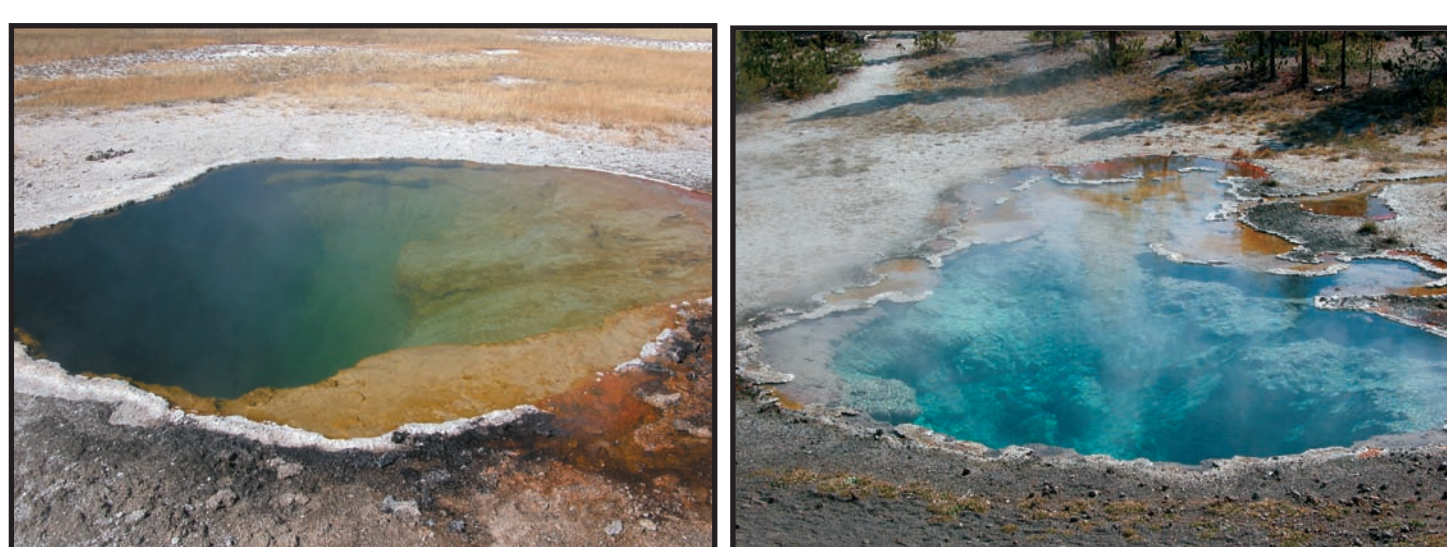


Improved DNA Polymerases for DNA Amplification and Sequencing

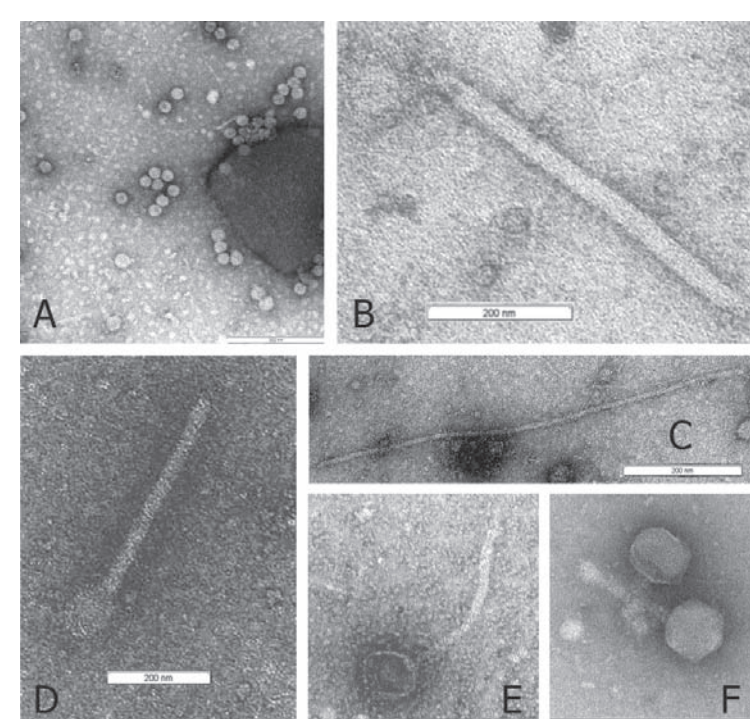
Tom Schoenfeld, Vinay Dhodda, Melodee Patterson, Nick Hermersmann, David Mead
Lucigen Corporation, Middleton, WI

Abstract

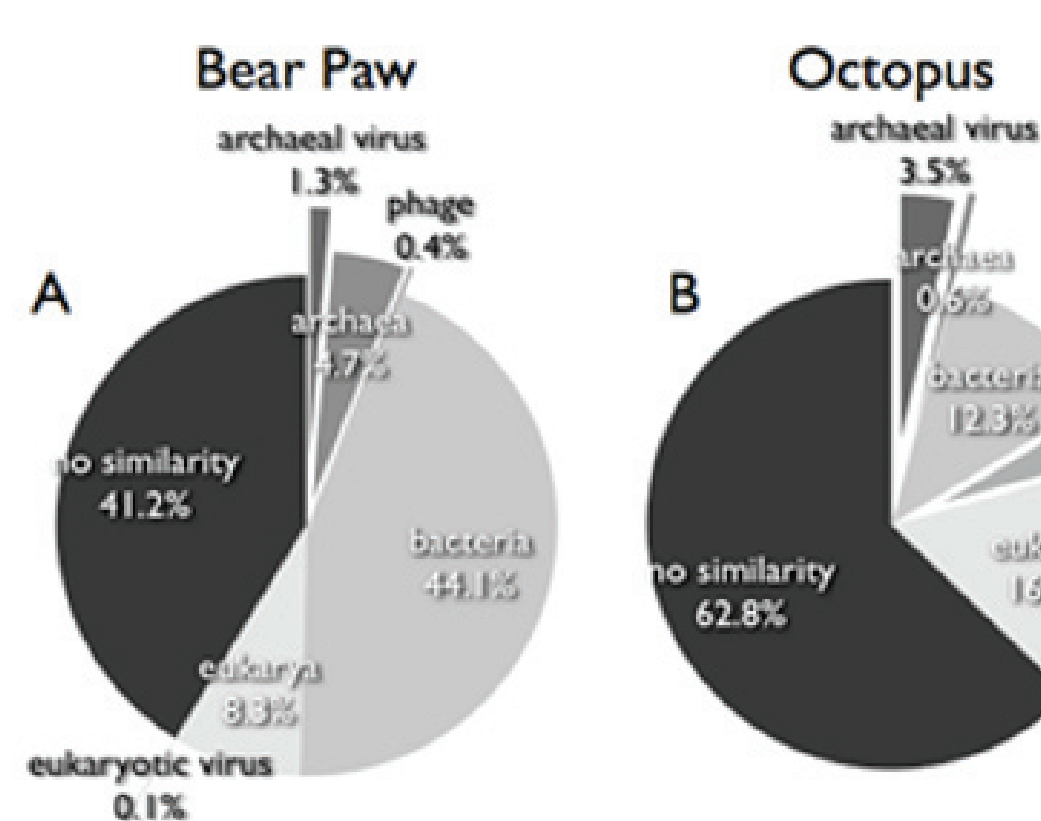
DNA polymerases are key components of most DNA amplification, sequencing and analysis platforms. As methods are developed and optimized for speed, throughput, accuracy and reliability, the need for improved enzymes is increasingly evident. Lucigen uses high throughput screening for novel enzymes as a starting point for engineering and evolution to address these needs. Screening of 37 Mb of sequence from genomic libraries constructed from viruses directly isolated from thermal springs resulted in the discovery of hundreds of thermostable viral DNA polymerases, as well as their respective accessory proteins. Ten pol genes were expressed to produce thermostable DNA polymerase activity. Data mining of other thermophile genomes in our collection has provided additional DNA polymerase activities. Directed mutagenesis allowed modification of the activities of these enzymes to improve their utility. These efforts have provided enzymes with unique properties. One of these allows unusually high-fidelity, high-efficiency PCR amplification and amplification of otherwise refractory sequences. Its thermostability and inherent strand displacement activity allows isothermal amplification at elevated temperatures, which results in greater specificity and lower background than is possible using conventional polymerases. Its inherent reverse transcriptase activity allows single tube, single enzyme RT PCR detection of mRNA transcripts. The enzyme synthesizes through regions of secondary structure which has resulted in reduced miscalls in chain-termination SNP detection and DNA sequencing. Additional improvements due to other DNA polymerases will also be discussed.



Bear Paw Hot Spring (74°C) left and Octopus Hot Spring (93°C) right



TEM Images of virus-like particles directly isolated from YNP hot springs.

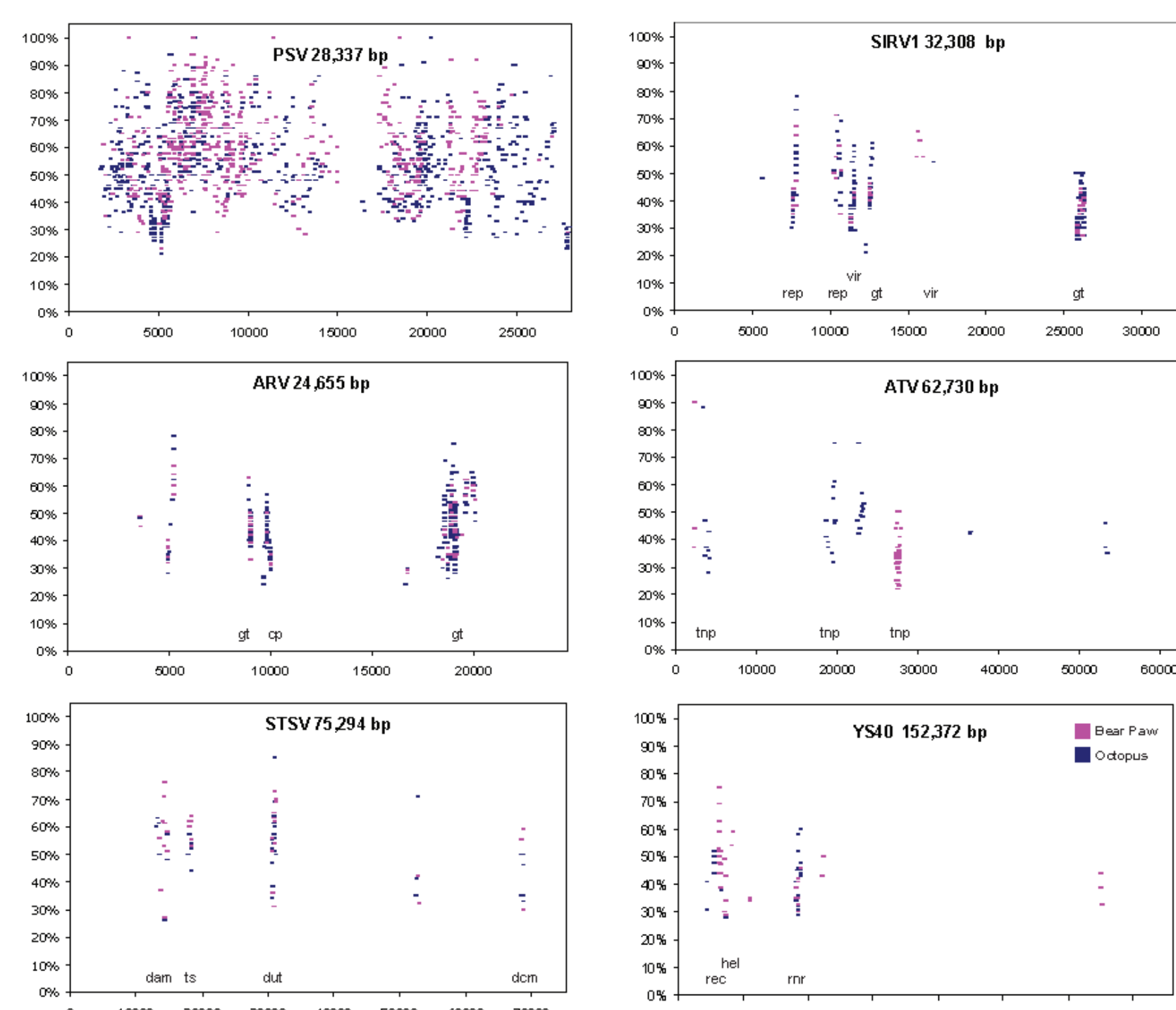


Broad classification of viral metagenomic contigs based on tBLASTx similarities. Contigs assembled at 95% identity from Bear Paw and Octopus reads (Panel A and B, respectively) were compared to sequences in GenBank to infer phylogeny. Shown are frequencies of contigs with no significant sequence similarity in GenBank ($E < 0.001$) and those with sequence similarity to Bacteria, Archaea, Eukarya and their respective viruses.

Numbers of tBLASTx similarities to cultivated viruses in metagenomic libraries.

Virus (Acc. No.)	Bearpaw	Octopus
ARV, <i>Acidianus</i> rod-shaped virus	36	228
SIRV, <i>Sulfolobus islandicus</i> rod-shaped virus	30	217
PSV, <i>Pyrobaculum</i> spherical virus	44	152
SIFV, <i>S. islandicus</i> filamentous virus	7	46
STSV1, <i>Sulfolobus tengchongensis</i> spindle-shaped virus 1	26	22
ATV, <i>Acidianus</i> two-tailed virus	8	17
TTSV1, <i>Thermoproteus tenax</i> spherical virus 1	6	12
YS40, <i>Thermus thermophilus</i> YS40 phage	15	41
Twort, <i>Staphylococcus</i> phage Twort	4	21

Alignment of Octopus and Bear Paw viral metagenomic library contigs with cultured virus genomes.



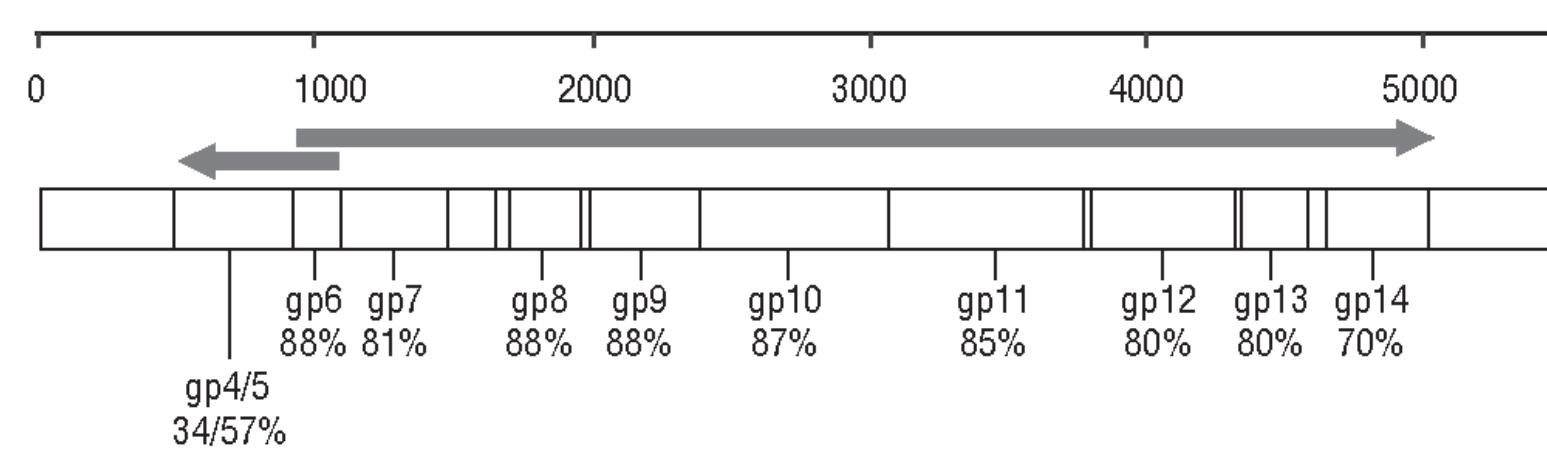
Contigs from the viral metagenomic libraries were compared by tBLASTx to the genomes of PSV, SRV1, ARV, ATV, STSV1 and YS40. Each bar represents a unique alignment of the metagenomic sequence to the indicated location on the cultivated viral genome, shown on the horizontal axis. Percent coding sequence identities are shown in the vertical axis. Red bars indicate Bear Paw alignments; blue bars indicate Octopus alignments. Also shown are the known or predicted functions of the conserved coding sequences (*rep*, replication related, *vir*, virion component, *gp*, glycosyltransferase; *trp*, transposase; *cp*, coat protein; *dam*, adenine DNA methylase; *ts*, thymidylate synthase; *dut*, dUTPase; *dcm*, cytosine DNA methylase; *hel*, helicase; *rec*, recombinase; *rrn*, ribonucleotide reductase).

Sequence assembly data and estimation of viral diversity.

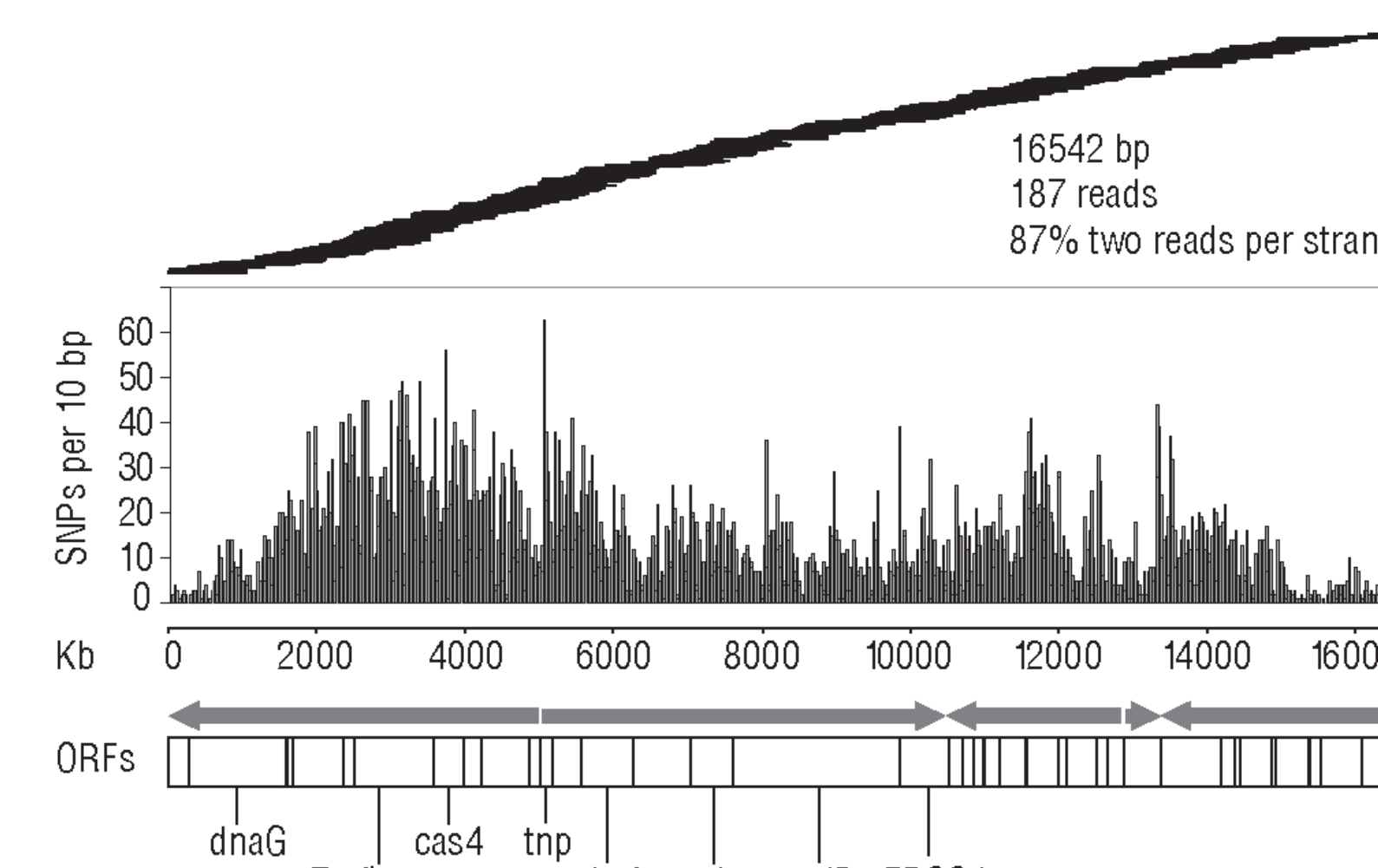
	Bearpaw	Octopus	Totals
Sequence reads	7,685	21,198	28,883
	Bear Paw 95%	Octopus 95%	Bear Paw 50%
Contigs assembled	6,191	13,543	4,850
Avg. reads per contig	1.239	3.129	1.587
Largest contig (nt)	3,503	4,554	8,007
Power law richness	1,440	1,310	548
Evenness score	0.946	0.954	0.933
Most abundant virus	2.14 %	1.88 %	3.93 %
Shannon-Wiener score	6.88	6.85	5.88

Largest contigs from 50% assemblies.

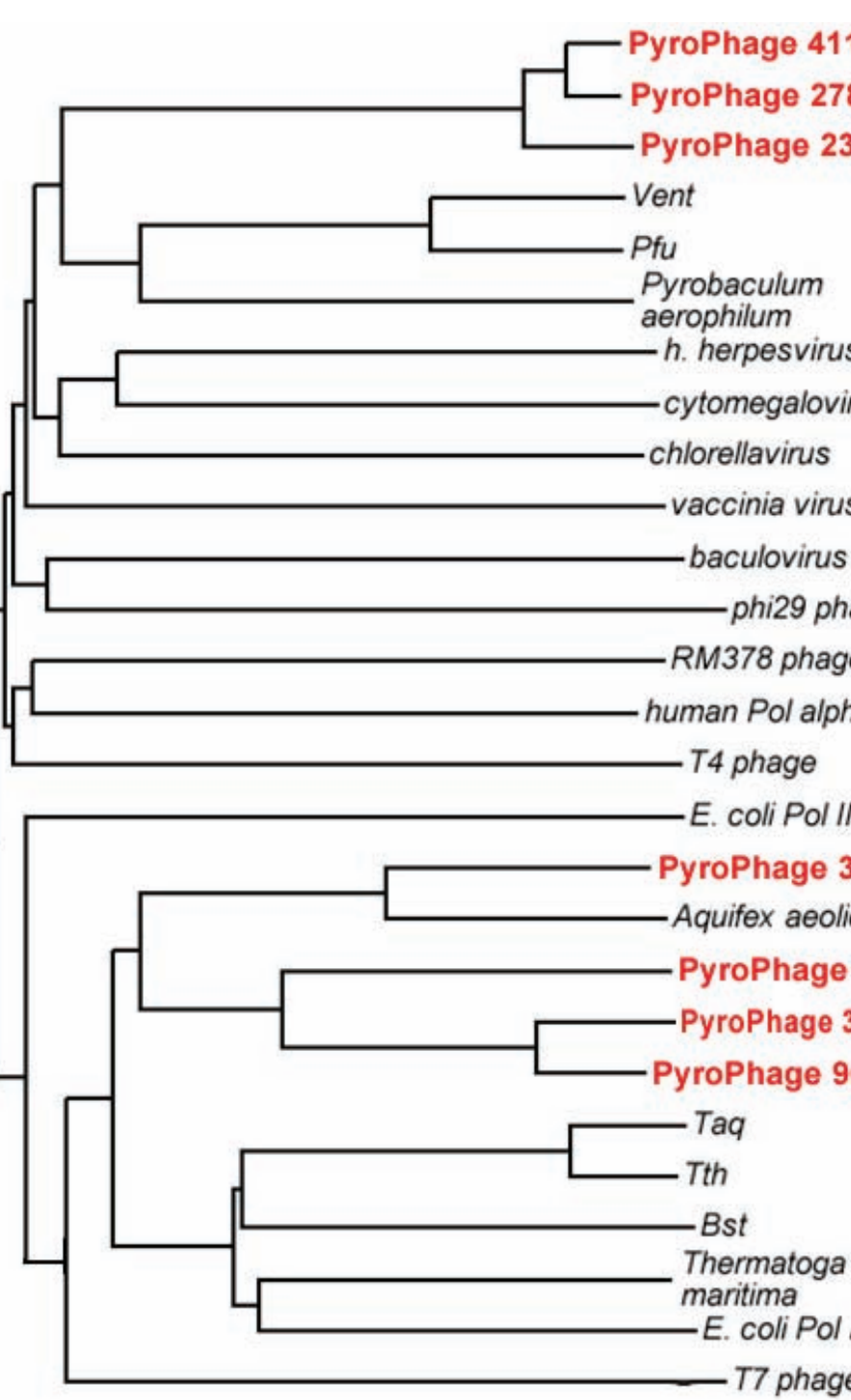
Contig	Length	Total Seq Length	No. of Seq.	Average coverage	Cumulative length
Octopus					
1496	35089	1814653	1601	51.72	35089
591	18974	1421079	1115	74.9	54063
1596	18451	360922	349	19.56	72514
426	17755	453426	414	25.54	90269
722	16542	182178	187	11.01	108811
1539	16465	383575	369	23.3	123276
Bear Paw					
480	8007	43088	47	5.38	8007
687	6276	75303	81	12	14283
1151	5749	44544	49	7.75	20032
372	5492	66055	71	12.03	25524
793	5219	27159	29	5.2	30743
754	5027	21341	23	4.25	35770



Genes and gene order are highly conserved between *Pyrobaculum* spherical virus and a consensus contig from the Bear Paw library. Contig 372 (5492 bp, 71 reads) was assembled at $\geq 50\%$ identity from the Bear Paw library. Open reading frames identified by GeneMark algorithm were compared by BLASTp to proteins in GenBank. Similarities to *Pyrobaculum* spherical virus proteins are shown with percent coding identity. The gene names are based on the annotation in GenBank and are named in order of their location on the viral chromosome. Direction of transcription is indicated by the arrows.



Alignment of nucleotide polymorphisms with coding sequences in a 16.5 kb consensus contig from Octopus Hot Spring. Contig 722 was assembled at $\geq 50\%$ identity from the Octopus library. Sequence coverage is shown on the top, with each line representing a separate read. Single nucleotide polymorphisms per 10 base pairs were normalized to the number of reads covering the respective nucleotide (middle) and are aligned with predicted open reading frames from the consensus sequence in the contig (bottom). Direction of transcription is shown by the arrows. Similarities to known genes were identified by BLASTp.



Dendrogram based on ClustalW analysis comparing amino sequences of the thermophilic viral DNA polymerases (red) with various family A and B DNA polymerases.

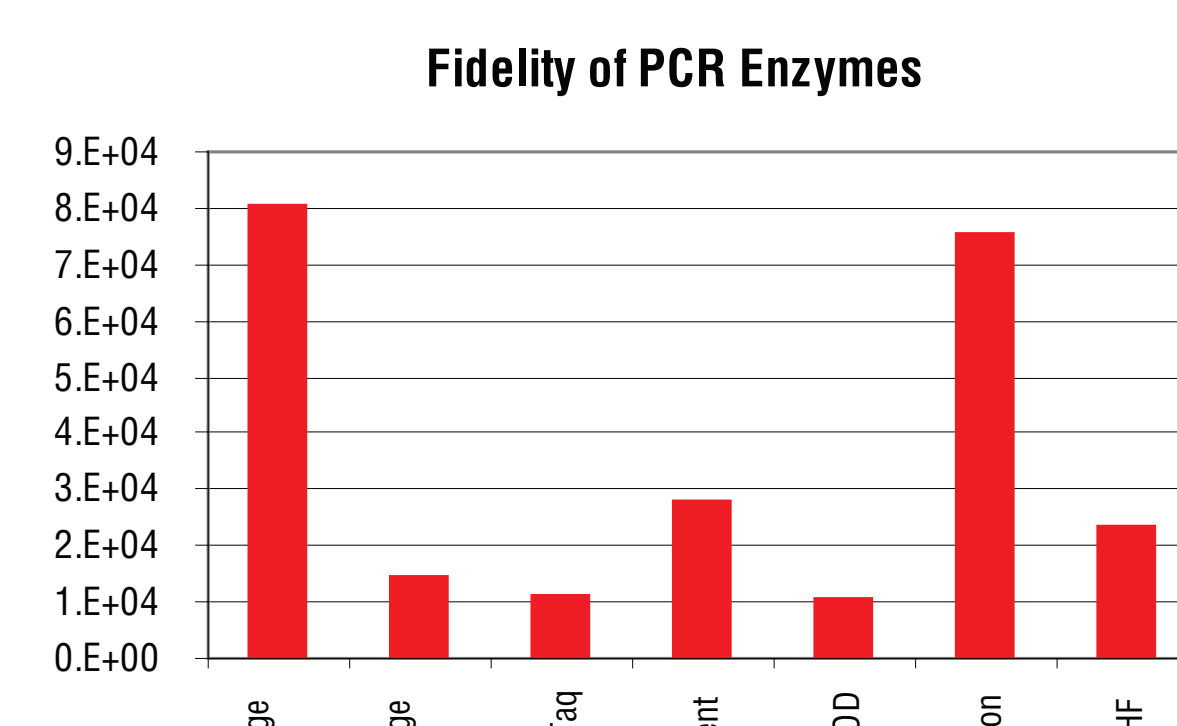
These samples were collected under Research Permit YELL-05240. This work was supported by NSF Grants 0109756 and 0215988 and NIH-NHGRI grant 1 R43 HG002714-01 to TS and DOE DE-FG02-02ER83484 to DAM and the assistance of the Delaware NSF EpSCOR program for computational support.

Lucigen Advanced Products for Molecular Biology

Lucigen Corporation 2120 W. Greenview Dr., Ste 9 Middleton, WI 53562 www.lucigen.com

PyroPhage 3173 DNA Polymerase

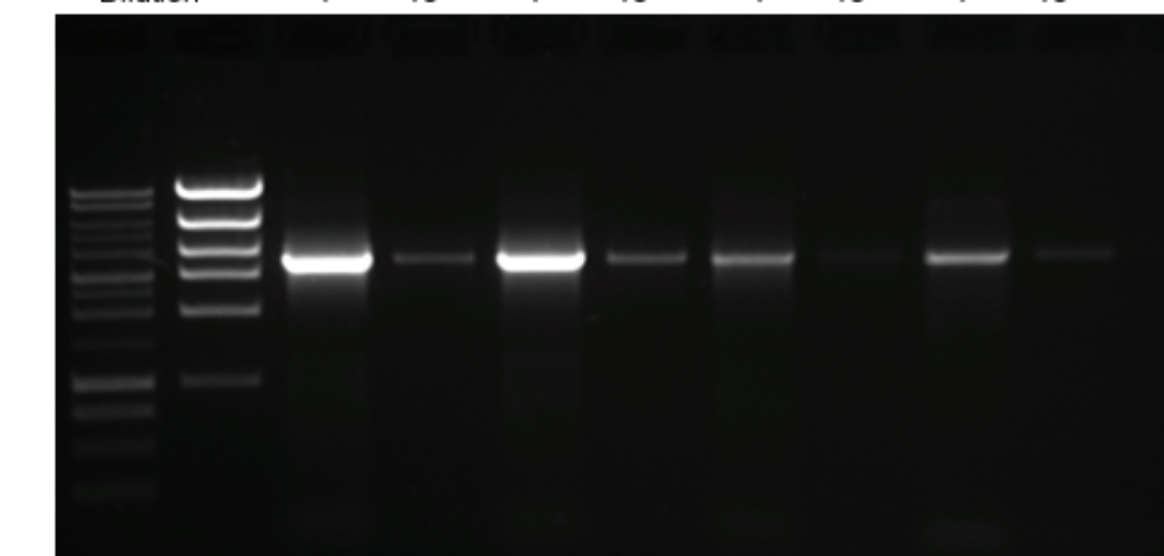
Wild Type	
5'-3' exo	-
3-5 exo (inactivated in the exo- mutant)	strong
Strand displacement	strong
Extension from nicks	strong
Thermostability (T _{1/2} @95°)	10 min.
Km dNTPs	40 μM
Km DNA	5.3 nM
Processivity	47 nt
Fidelity	8 X 10 ⁴



PCR Fidelity
Fidelity measurements based on the LacIq forward mutation assay compare PyroPhage 3173 DNA polymerase (wild type and exo minus) to various commercially available enzymes.

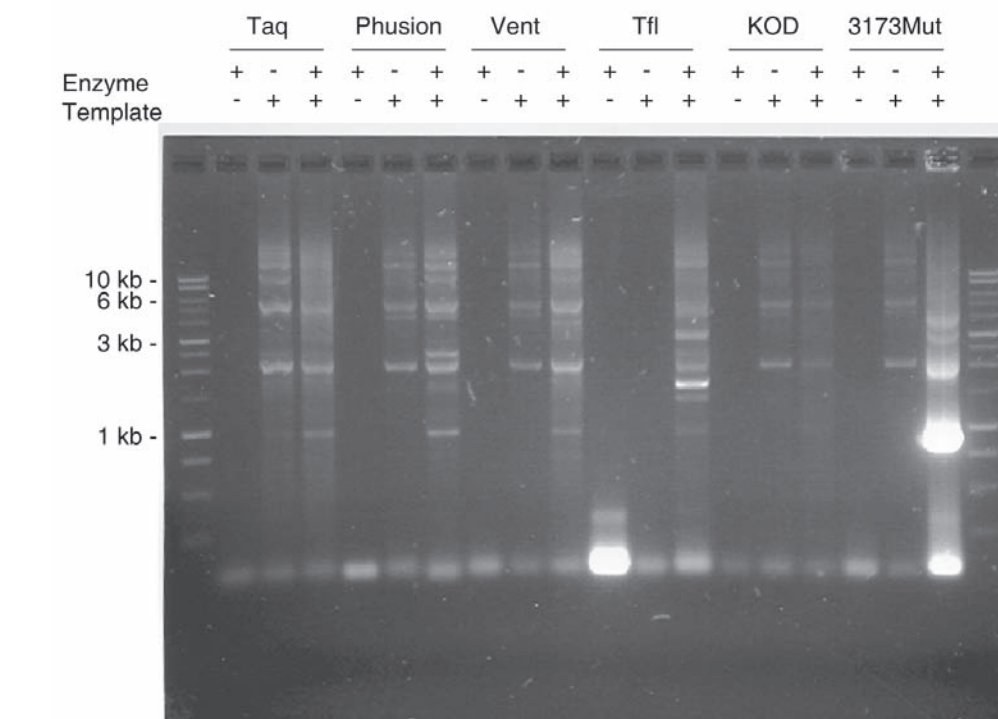
Amplification efficiency

Enzyme	Cycles	Yield	Efficiency
PyroPhage exo- E51A	21	18 μg	43%
PyroPhage exo- D49A	21	18 μg	43%
Taq	25	3.2 μg	28%
Vent	25	3.2 μg	28%



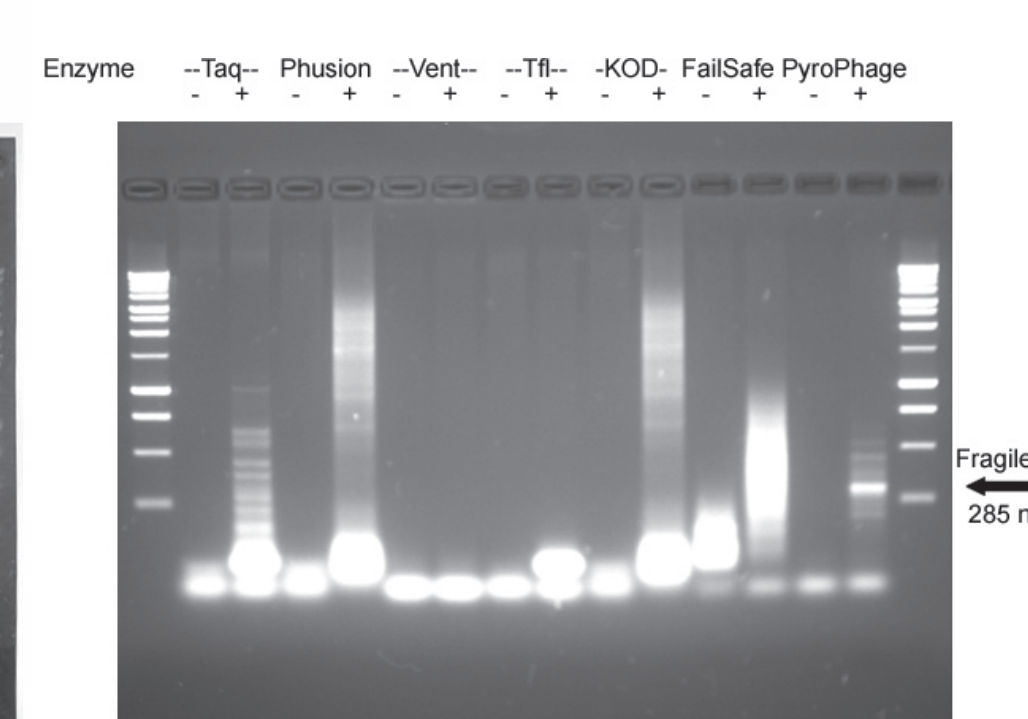
Amplification efficiency of various DNAPs. PCR reactions were quantified by gel electrophoresis. Cycle numbers for the PCRs are shown.

PCR amplification of *cobA*

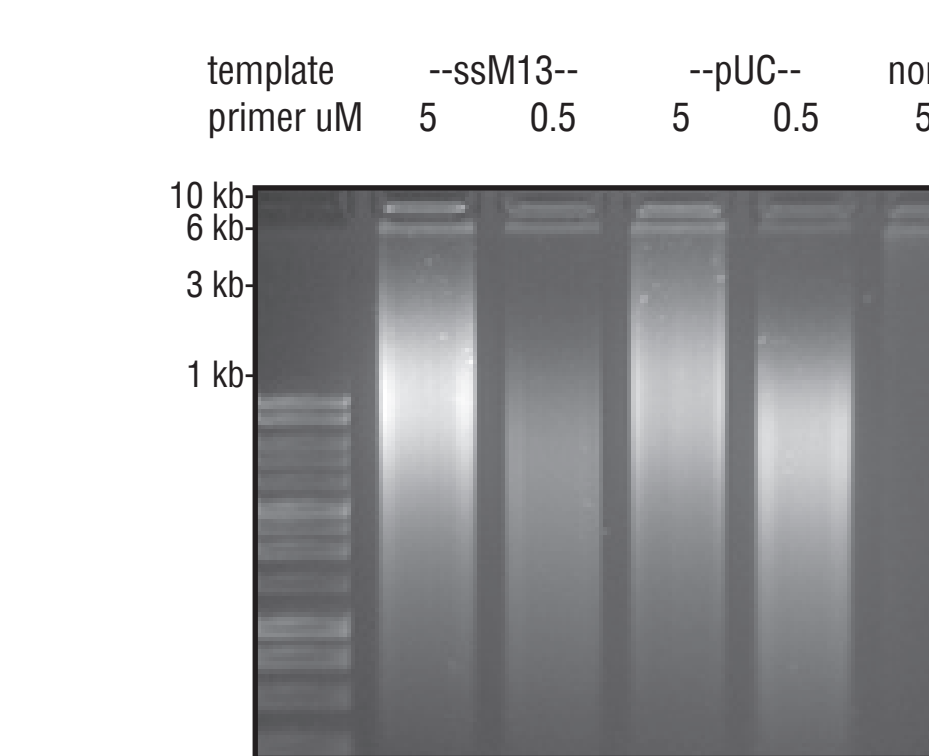


PCR amplification of difficult templates. PCR amplification of the *Bacillus* *cobA* gene (left) and the human Fragile X gene (right) using the indicated DNAPs.

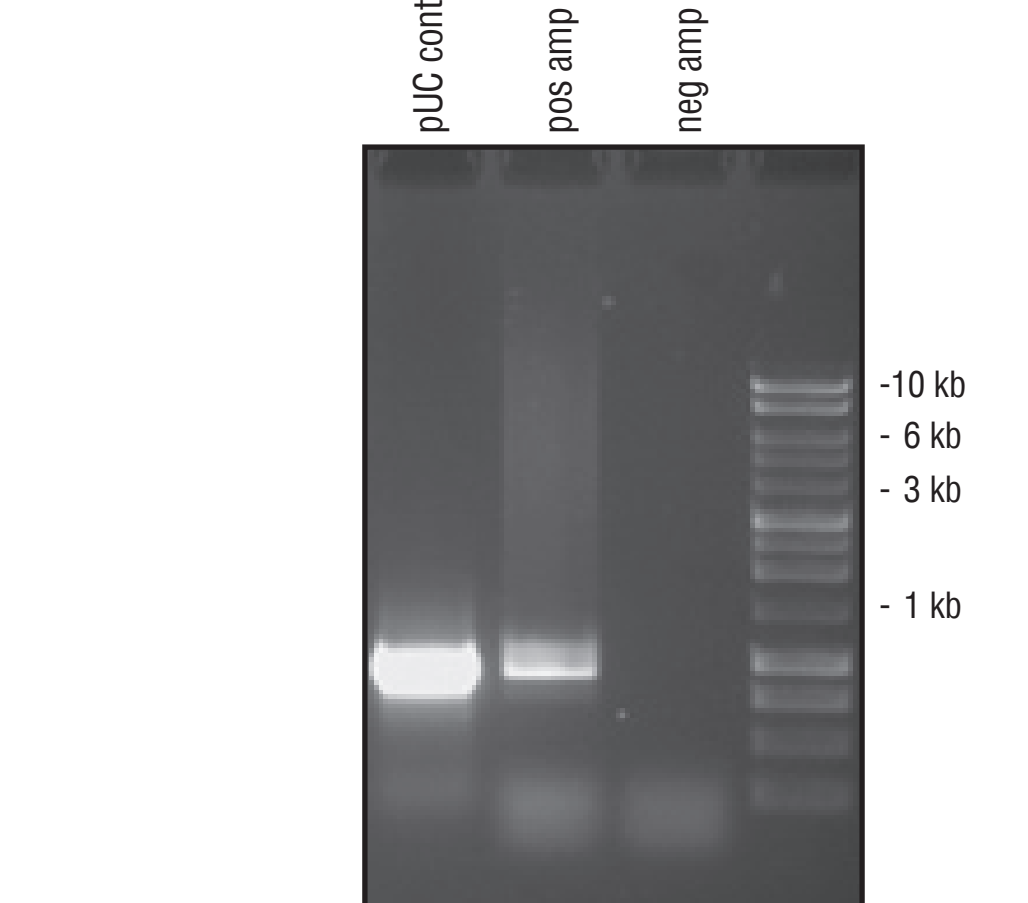
PCR amplification of Fragile X



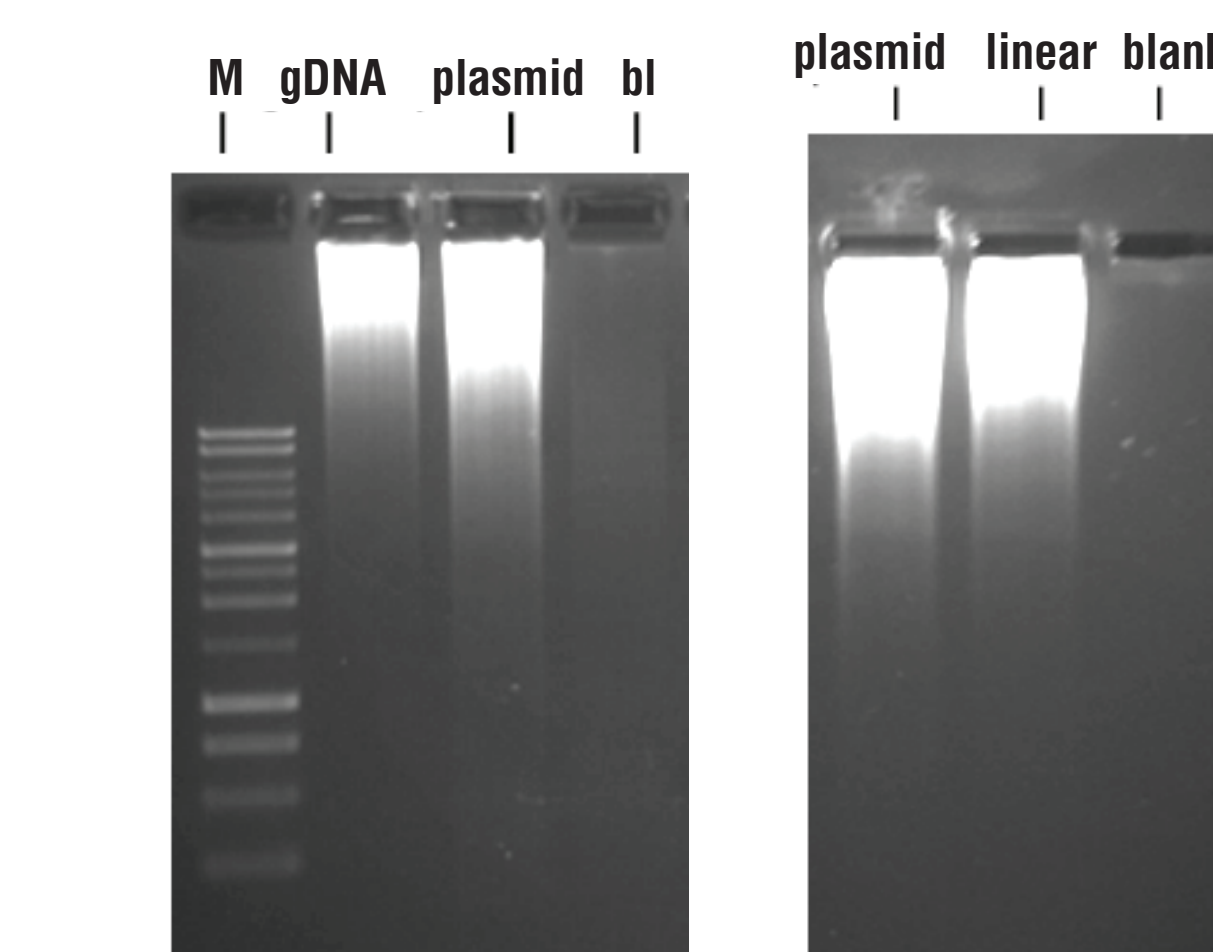
A) Isothermal amplification of circular templates using PyroPhage 3173 POL



B) Verification by PCR of amplification specificity



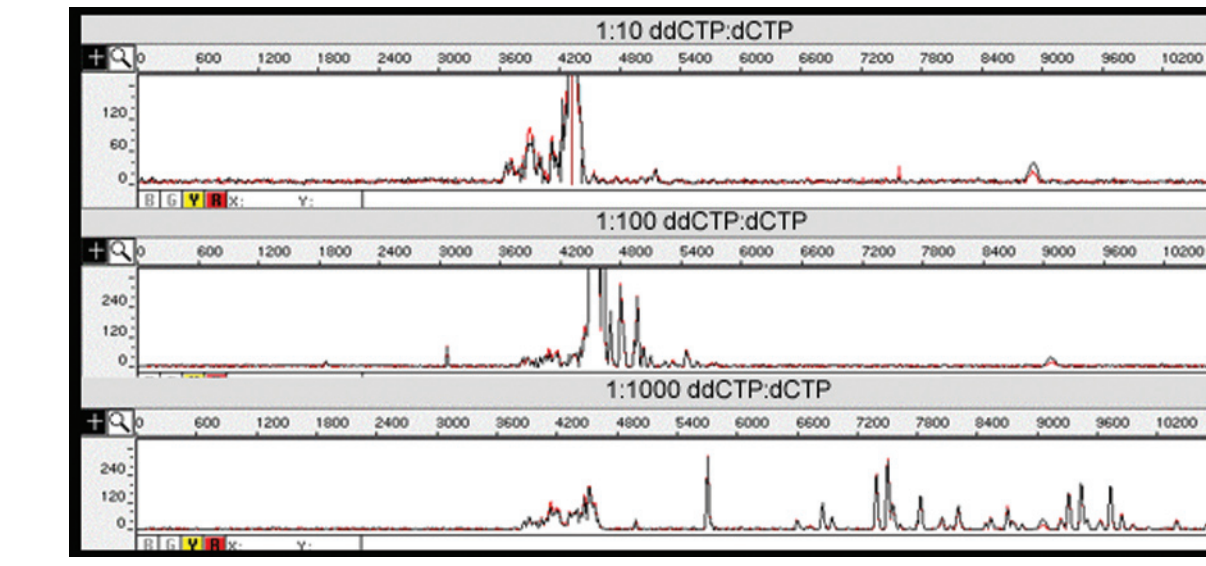
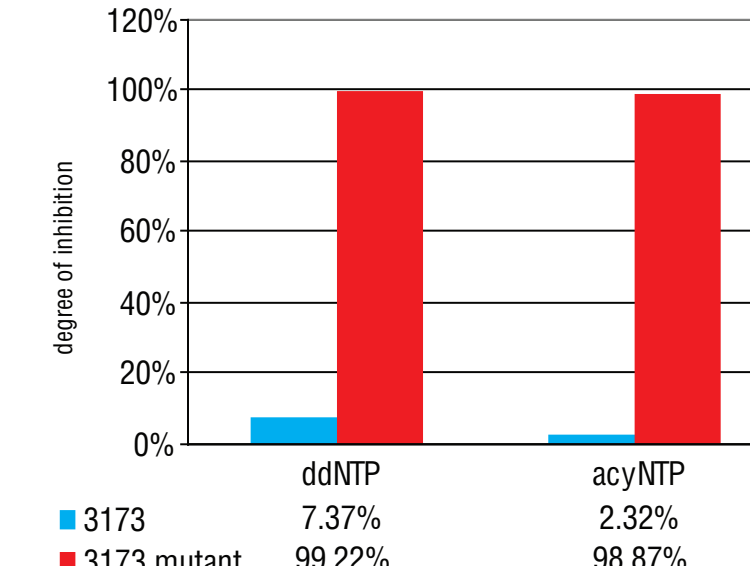
Rolling circle amplification using 3173 DNAP. PyroPhage 3173 wt DNA polymerase was used to amplify M13, pUC19 DNA, or no template. Primers were added at concentrations of 5 or 0.5 μM (lanes 3 and 5). Panel B shows PCR amplification of a 1 kb sequence of pUC19, RCA product, and a negative control.



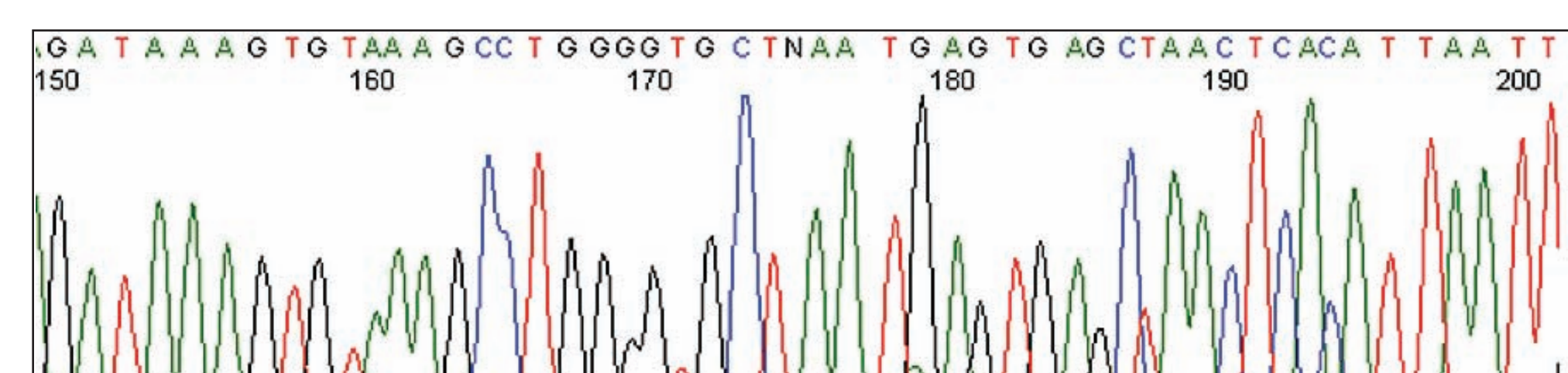
Primer free isothermal amplification. PyroPhage 3173 exo minus DNA polymerase was used to amplify *E. coli* gDNA, supercoiled plasmid and pJAZZ linear vector. Reactions were incubated two hours at 55°C. In the absence of added template (blank), no product is seen.

Reverse-transcribed PCR
PyroPhage 3173 DNA polymerase and actin primers were used to reverse transcribe and amplify β actin cDNA from mouse liver RNA. Reactions were incubated at 60°C for 60 minutes, followed by the 35 standard PCR cycles.

Inhibition by Chain Terminators



Incorporation of chain terminating nucleotides. The D49A exo- mutant of 3173 and its derivative were tested for discrimination against dideoxy and acyclo chain terminators based on inhibition of DNAP activity (left panel). Rox ddCTP dye terminator was included at ratios between 1:10 and 1:1000 relative to each dNTP. Termination of extension was detected by the ABI 310 Genetic Analyzer (right panel).



Sequencing with PyroPhage 3173. PyroPhage 3173 derivative was used to sequence pUC19 plasmid using fluorescent dye terminator chemistry.