

Viral community genomics of thermal aquifers and improved DNA polymerases for DNA amplification and sequencing

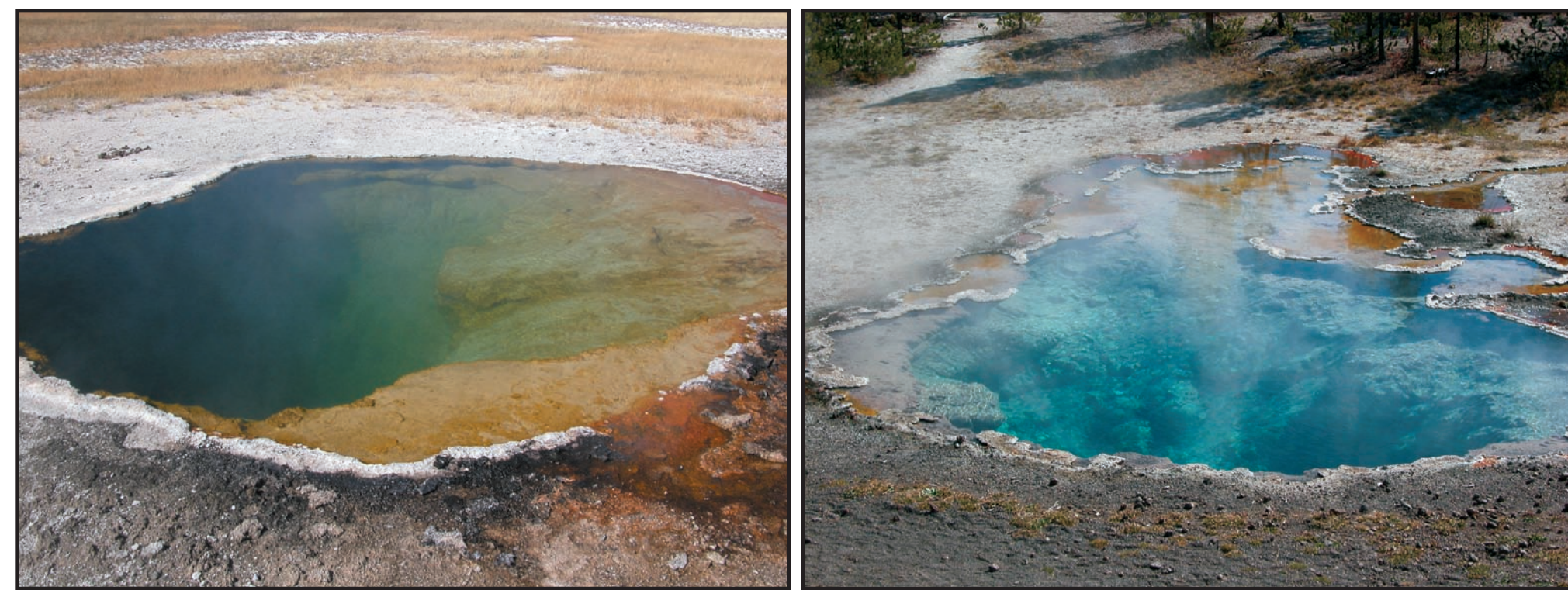
T. Schoenfeld¹, V. Dhodda¹, R. DiFrancesco¹, M. Patterson¹, P. Richardson², Mark Young³, D. Mead¹;

¹Lucigen Corporation, Middleton, WI, ²Joint Genome Institute, Walnut Creek, CA., ³Montana State University, Bozeman, MT

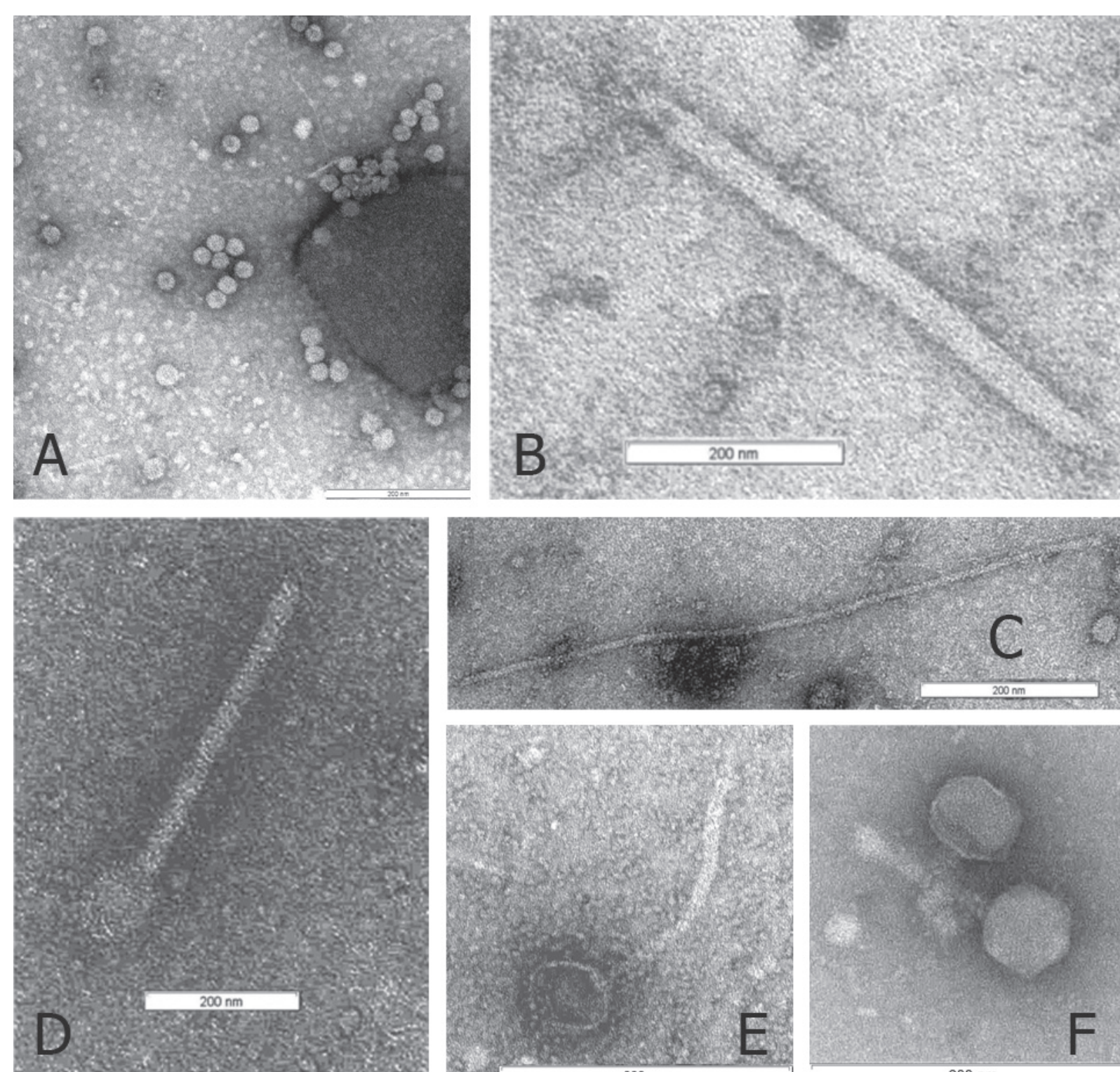
Abstract

Research on thermophilic viruses (phage) has relied almost exclusively on samples isolated by culturing. This study examined viruses directly isolated from five 74 to 93°C alkaline siliceous hot springs in Yellowstone National Park. Viral abundances ranged from 1E5 to 1E6 virus-like particles per milliliter and a variety of bacteriophage and crenarchaeal virus-like morphologies were observed. Viral metagenomic libraries were constructed from two hot springs and nearly 30 Mb of sequence was determined. An estimated 400 to 1400 different viral species inhabit each hot spring and hundreds of virus-like coding sequences were detected including predicted replication, recombination, lysis, transcription, lysogeny, and structural genes. Although the two hot springs are separated by 5 km, 21 to 28% of the sequences of one hot spring were highly similar to the other, suggesting substantial overlap of the populations. Remarkably, almost the entire 28 kb genome of the cultivated *Pyrobaculum* spherical virus has an apparent homolog in the two hot springs. Similarities to crenarchaeal viruses cultivated from acid hot springs on other continents were limited to specific open reading frames, but were equally as strong, indicating conservation of these coding sequences between viruses inhabiting very distinct biomes. Thermophilic viruses were examined as sources of improved enzymes for DNA amplification and sequencing, resulting in the discovery of over 200 DNA polymerase genes, 58 of which were full length. The encoded thermostable viral DNA polymerases are much more diverse than known thermostable enzymes. The fact that no pol gene was re-isolated suggests this diversity is the tip of a very large iceberg. We have expressed 10 of these DNA polymerase genes. One allows isothermal whole genome amplification at elevated temperatures with greater specificity and lower background than conventional polymerases. It also improves PCR fidelity and allows amplification of otherwise refractory templates. Its inherent reverse transcriptase activity suggests utility in RT-PCR. This enzyme has also shown potential as an improved sequencing enzyme.

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.



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

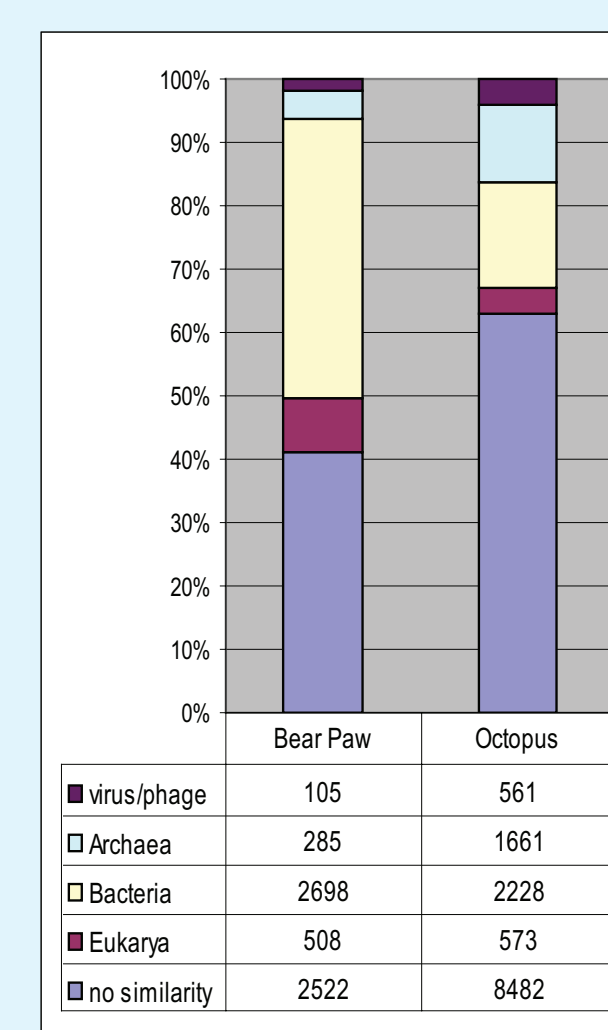


TEM Images of virus-like particles directly isolated from YNP hot springs. Images from Azure (Panels A, B, D), Cavern (Panel C), Bear Paw (Panel E) and Octopus (Panel F) hot springs are shown.

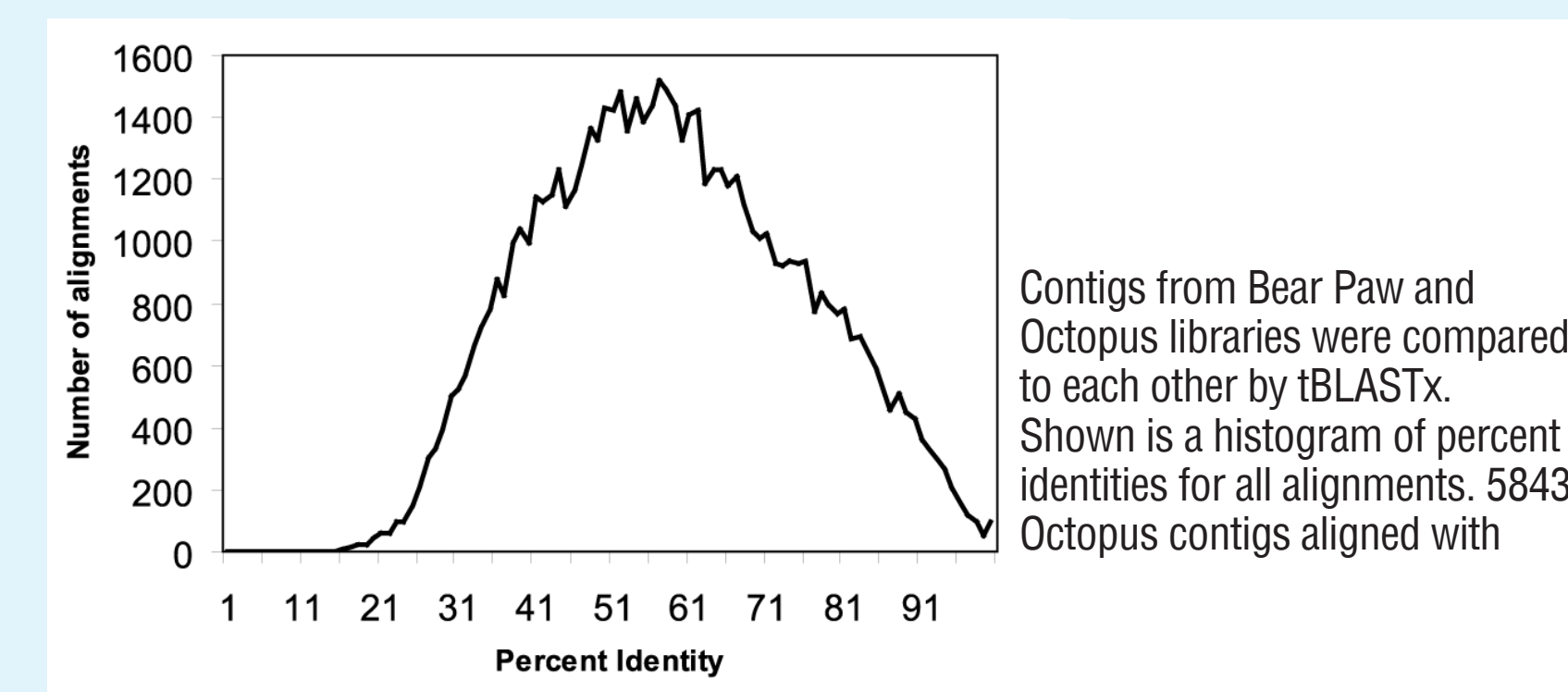
Sequence assembly data and estimation of viral diversity.

	Bearpaw	Octopus	Totals
Sequence reads	7,685	21,198	28,883
Contigs assembled	6,191	13,543	19,734
Ave. reads per contig	1.239	3.129	2.184
Largest contig (nt)	3,503	4,554	
Chao1 species richness	428	524	
Power law richness	1,440	1,310	
Evenness score	0.946	0.954	
Most abundant virus	2.14 %	1.88 %	
Shannon-Wiener score	6.88	6.85	

Sources of the Strongest BLASTx similarity



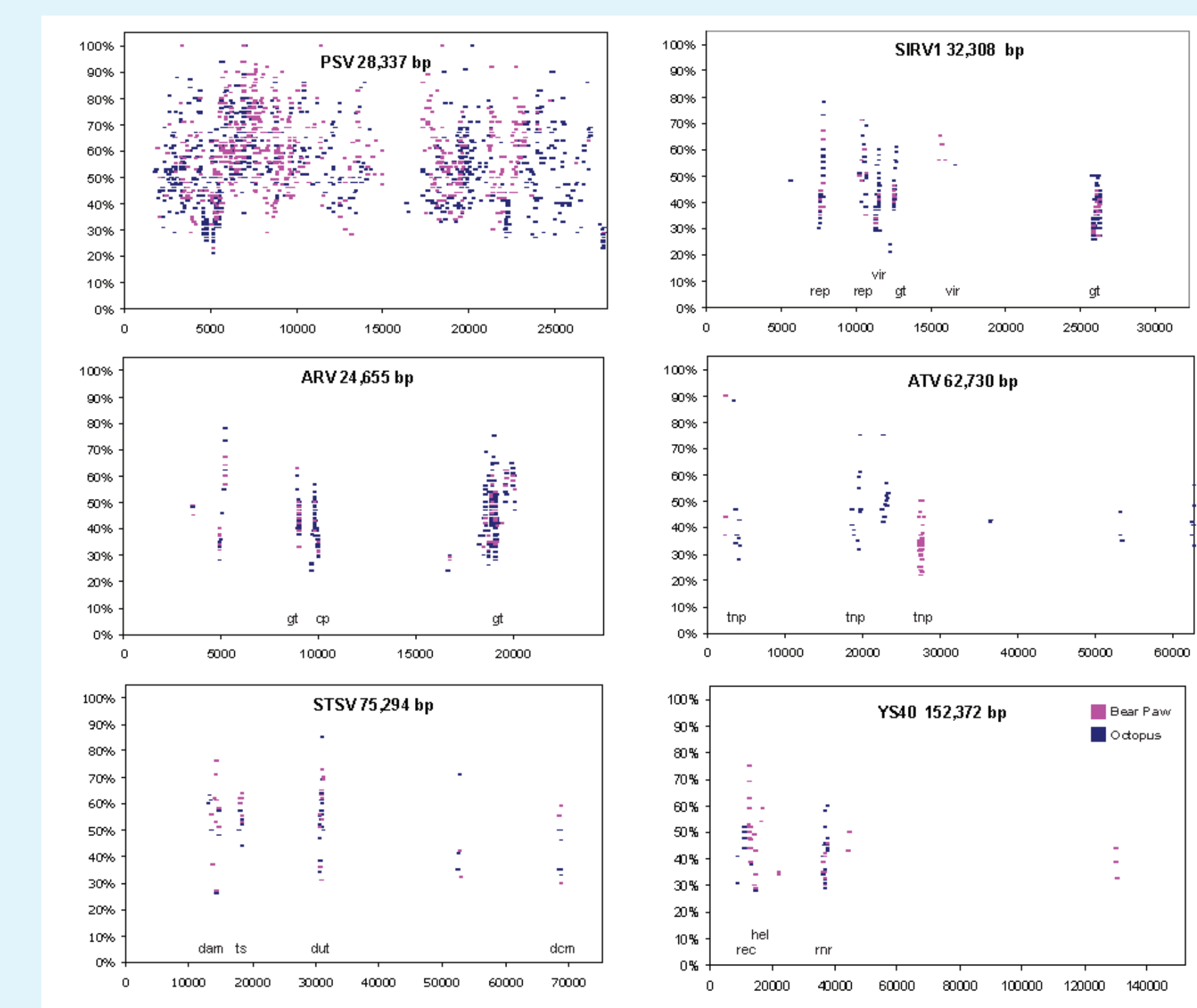
Interlibrary comparison by tBLASTx alignment.



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.

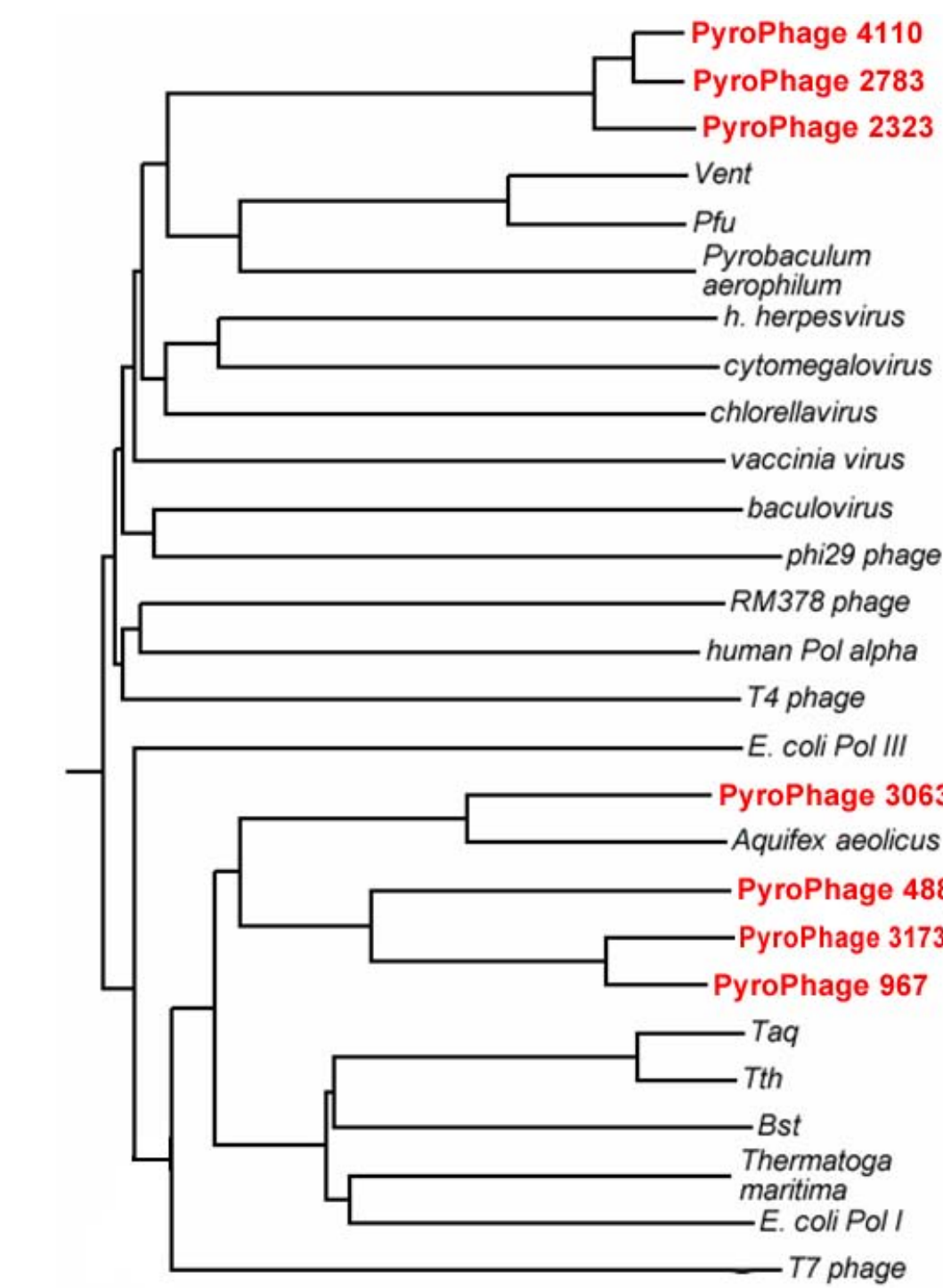


Contigs from the viral metagenomic libraries were compared by tBLASTx to the genomes of PSV, SIRV1, ARV, ATV, STSV and YS40 SIFV TTSV. Each bar represents a unique alignment of the metagenomic sequence to the indicated location on the cultivated viral genome, shown on the horizontal axis. Percent coding 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; *gt*, glycosyltransferase; *tnp*, transposase; *cp*, coat protein; *dam*, adenine DNA methylase; *ts*, thymidylate synthase; *dut*, dUTPase; *dcm*, cytosine DNA methylase; *hel*, helicase; *rec*, recombinase; *mrr*, ribonucleotide reductase).

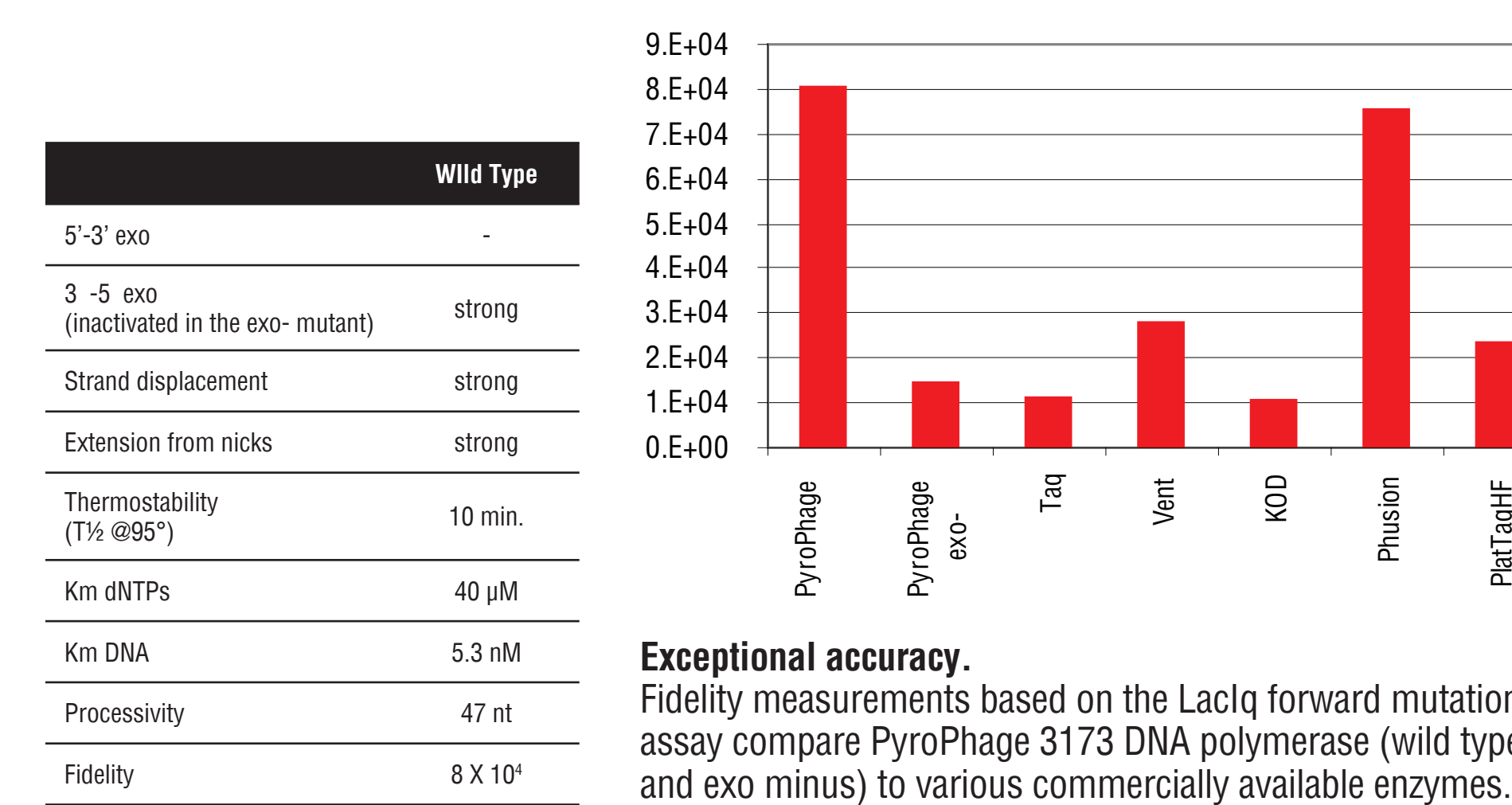
Numbers of apparent virus-like genetic elements in the hot springs viral metagenomes based on BLASTx sequence similarities.

Functional Grouping	Bear Paw	Octopus	Bear Paw	Octopus
Total Reads	6572	13974		
No Similarity	2545	8469	38.7%	60.6%
assembly/lysis	254	282	3.9%	2.0%
infection	181	213	2.8%	1.5%
lysogeny	52	34	0.8%	0.2%
NA metabolism	1445	2130	22.0%	15.2%
recombination	170	171	2.6%	1.2%
replication/repair	466	784	7.1%	5.6%
structural	96	314	1.5%	2.2%
transcription	187	196	2.8%	1.4%
translation	221	336	3.4%	2.4%
no keyword match	955	1045	14.5%	7.5%

ClustalW analysis of the PyroPhage DNA polymerases.



Fidelity of PCR Enzymes



Exceptional accuracy.

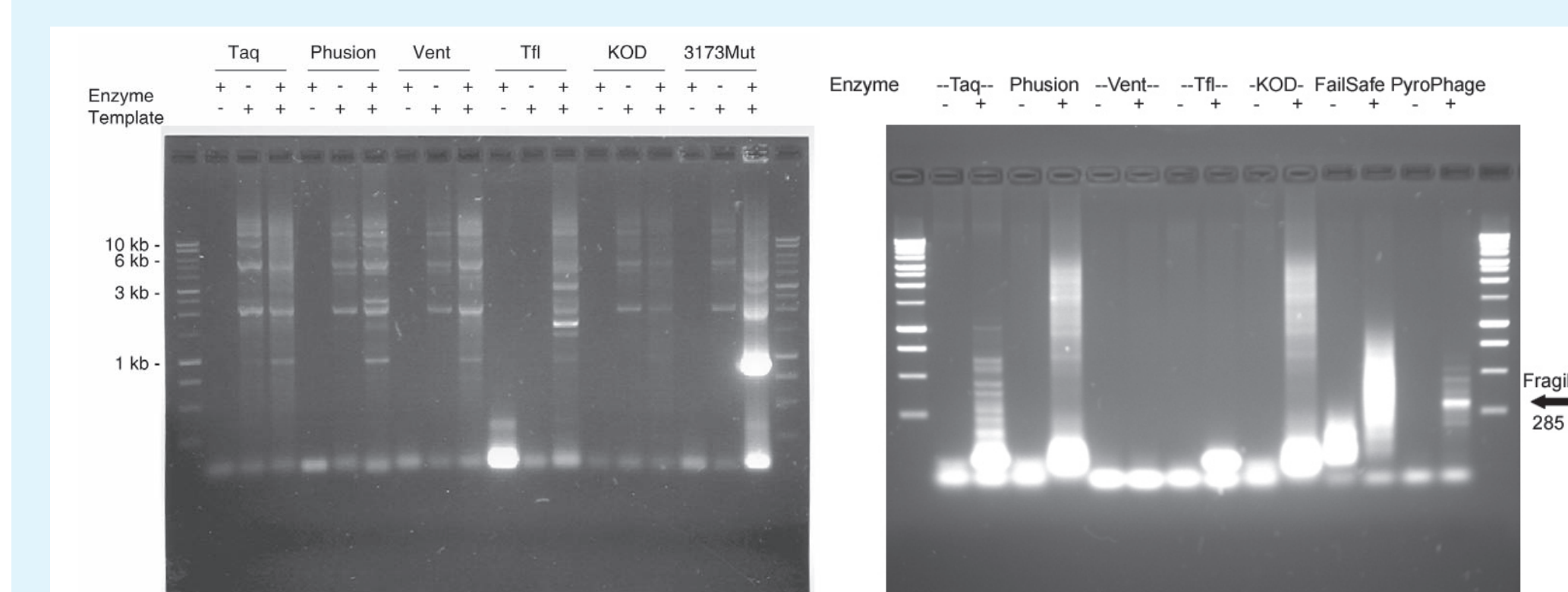
Fidelity measurements based on the LaqI forward mutation assay compare PyroPhage 3173 DNA polymerase (wild type and exo minus) to various commercially available enzymes.

Amplification efficiency

Enzyme	Cycles	Yield	Efficiency
PyroPhage exo- E51A	21	18 µg	43%
PyroPhage exo- D48A	21	18 µg	43%
Taq	25	3.2 µg	26%
Vent	25	3.2 µg	26%

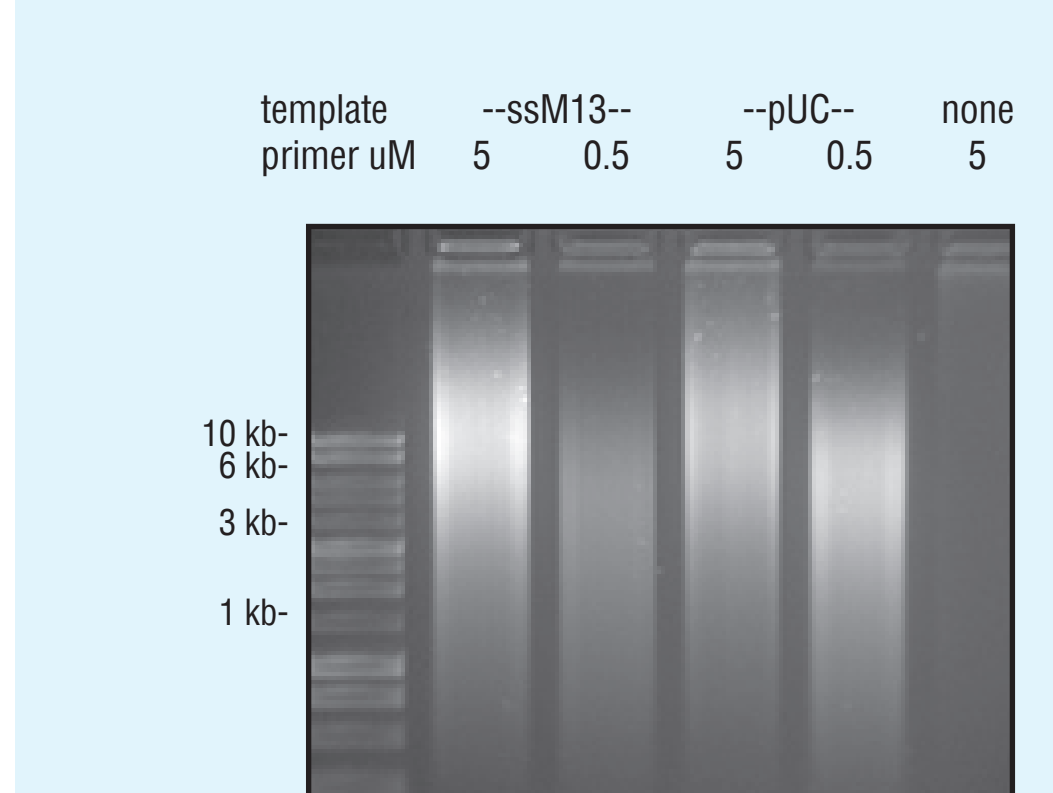
Amplification efficiency of various DNAPs. PCR reactions were quantified by gel electrophoresis. Cycle numbers for the PCRs are shown. Efficiency was calculated by published methods (5).

PCR amplification of *cobA* and PCR amplification of Fragile X

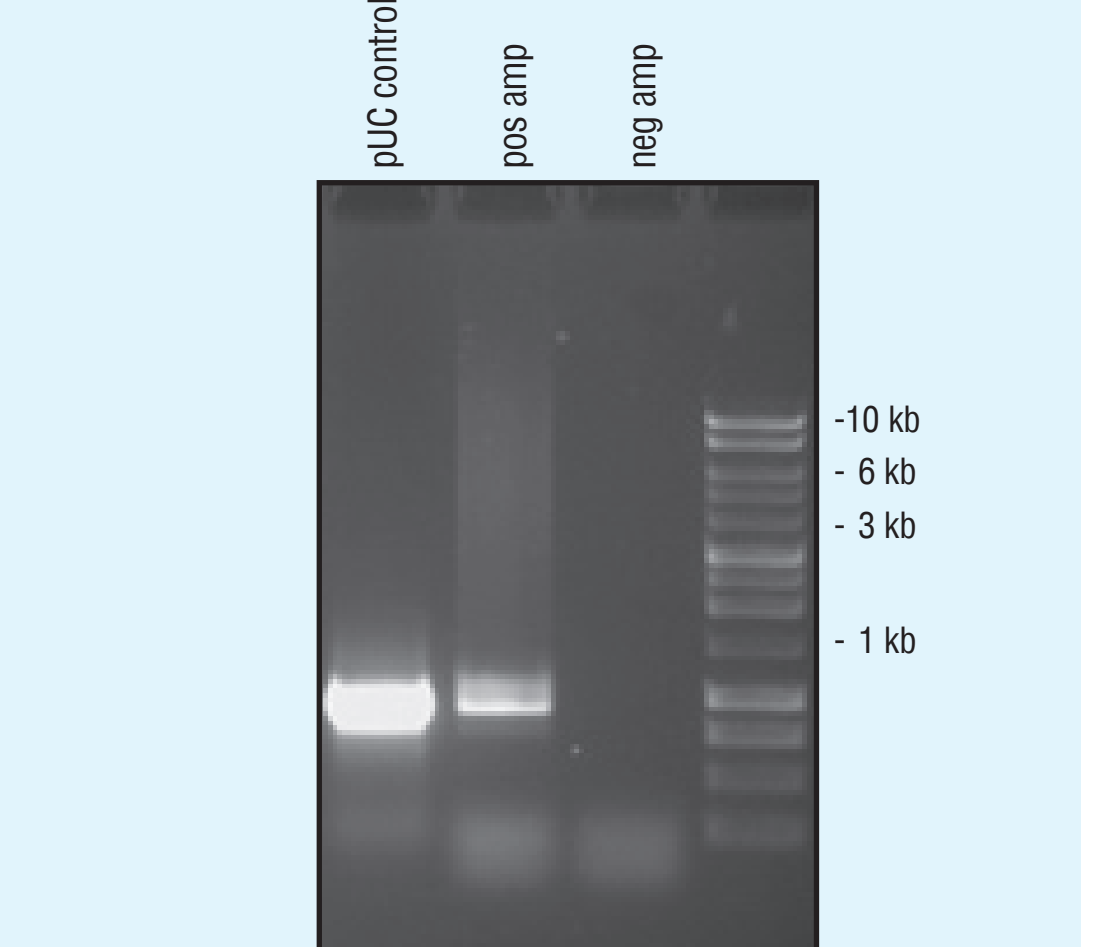


PCR amplification of difficult templates. PCR amplification of the *Bacillus* *cobA* gene (left) and the human Fragile X gene (right) using the indicated DNAPs.

A) Isothermal amplification of circular templates using PyroPhage 3173 POL

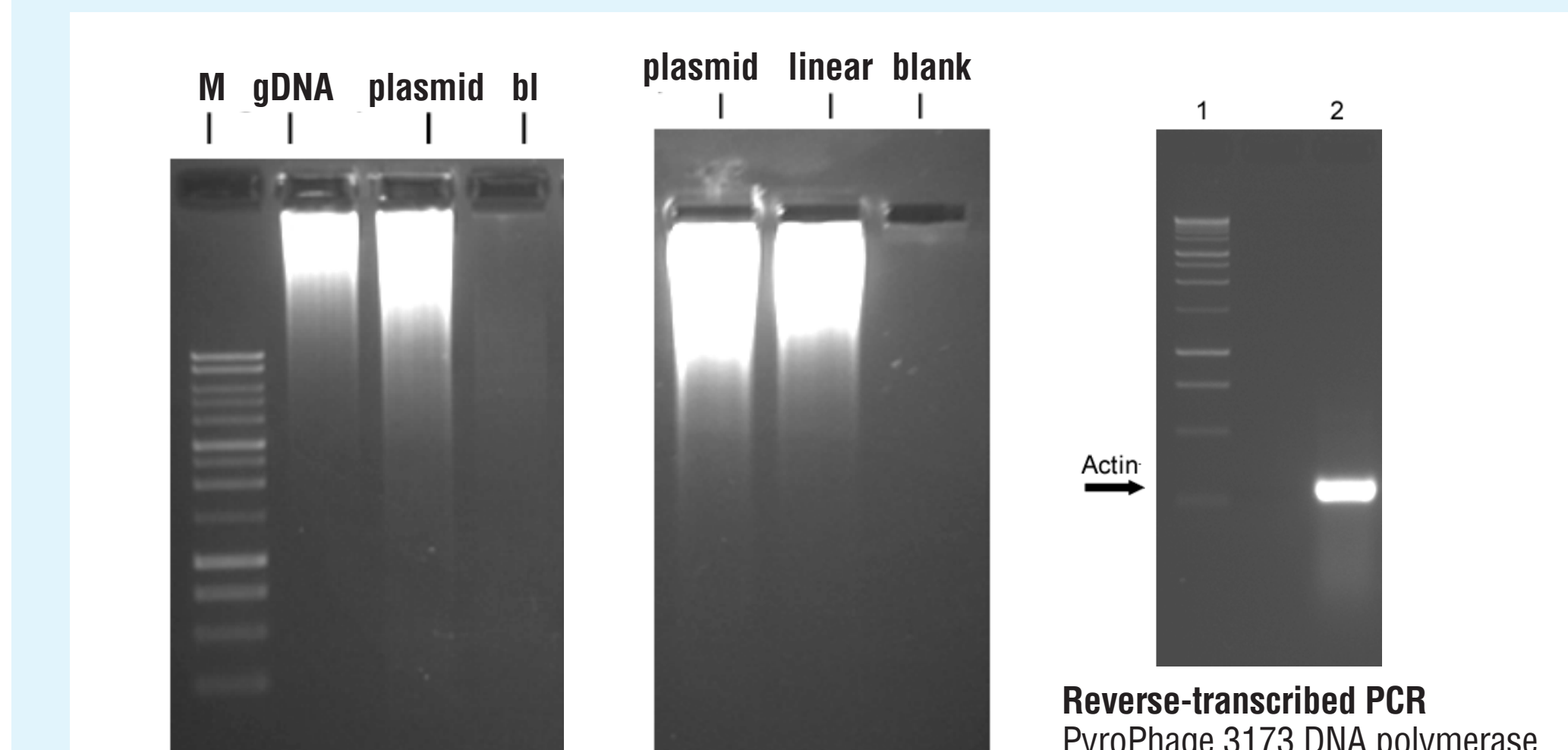


B) Verification by PCR of amplification specificity



Rolling circle amplification using 3173 DNAP.

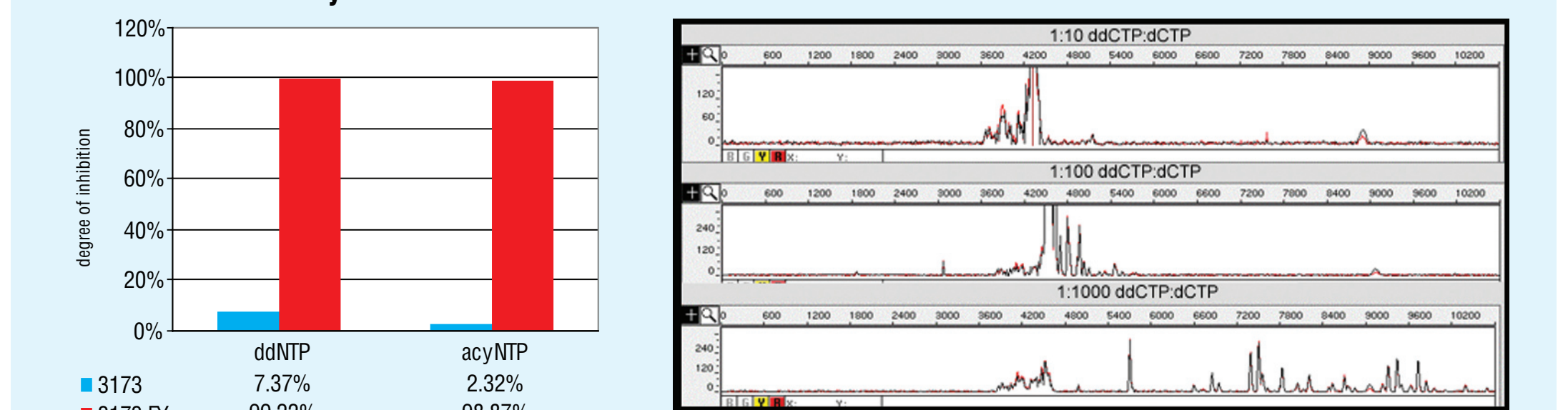
PyroPhage 3173 wt DNA polymerase was used to amplify M13, pUC19 DNA, or no template. Primers were added at concentrations of 5 or 0.5 µM (lanes 3 and 5). Panel B shows PCR amplification of a 1 kb sequence of pUC19, RCA product, and a negative control.



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.

Inhibition by Chain Terminators



Incorporation of chain terminating nucleotides.

The D49A exo- mutant of 3173 and its (FY) 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).

Lucigen Advanced Products for Molecular Biology

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