

ABSTRACT

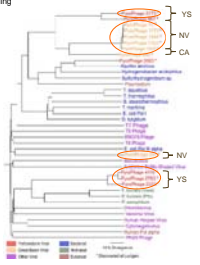
Comparison of the metagenomic profiles of viruses in three circumnuclear hot springs of similar temperature and chemistry shows significant differences in gene content and structure associated with geographical and temporal separation. Viral metagenomes were isolated from Octopus and Black Pool (Yellowstone) and Great Boiling Spring (Nevada). Roche 454 and Sanger sequence data were used to assemble contigs up to 33.5 kb, some of which may contain nearly complete viral genomes. Putative replication operons in these assemblies suggest novel replication strategies and provide new biological reagents. Viral replicases were identified by sequence similarity and by functional activity. More than twenty novel thermostable viral DNA polymerases (Pol)s were expressed and characterized, all of which are molecularly and biochemically divergent from known Pol's. Based on amino acid sequences, these Pol's fall into four clades. Clade 1 includes family A-like Pol's and is the most diverse, geographically dispersed, and persistent. Clade 2 includes family B-like Pol's and was relatively conserved and abundant in one Octopus metagenome, but was not detected in later screens of this or other springs. Clade 3 was exclusive to Black Pool and includes primer-polymerase enzymes similar to those seen in archaeal plasmids, but never before reported in viral genomes. Clade 4 is comprised of a single Pol that was discovered by functional screening of the Great Basin metagenome and has weak but significant similarity (E = 0.73) to a single open reading frame of unknown function in the crenarchaeal virus, SSV1. Assembly of these pol genes with other sequences in the metagenomes reveals the genomic context of replication genes. Both Clade 1 and 2 Pol's appear to be encoded in operons that include genes for likely subunit helicases and recB-like endonucleases. The gene for one Clade 1 Pol is contained in the 33.5 kb contig allowing identification of adjacent genes, several of which appear to be replicase subunit genes. This Pol appears to be expressed as a polyprotein of 1608 amino acids that is processed *in vitro* and presumably *in vivo* to generate 70 kD Pol's. A truncation product of this protein has proven highly useful for single enzyme RT-PCR and isothermal RNA and DNA detection. Sequence similarity and the polyprotein structure are shared by the other representatives of this clade. These Pol's also share sequence similarity, but not the polyprotein structure, with Pol's of Aquificales order bacteria. Remarkably, they also have strong similarities in sequence, gene structure, and thermostability with Pol's encoded by the plasmids of several parasitic protists, e.g., *Plasmodium*, suggesting unusual patterns of gene transfer. Isolation of sequence variants of both the Clade 1 and 2 Pol's allows mapping of coding variation associated with measurable biochemical differences. The high conservation, relative abundance, and wide geographic distribution of certain replication-associated genes, especially *pol* and *hel* genes, suggests they may be valuable as signature genes for several types of thermophilic viruses. Biochemical characterization of selected viral replicases will be discussed.

Thermophilic Viral Metagenomics



TEM images of the viruses directly isolated from the springs.

PyroPhage Polymerases were isolated from near-boiling hot springs.

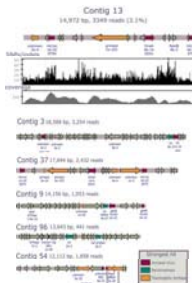


A phylogram of selected PyroPhage viral polymerases (circled) based on ClustalW alignment of the amino acid sequences. Pol's were discovered by sequence similarity and functional screening.

454 Sequencing of Viral Metagenomes

Spring	Genome Center	Number of Reads	Assembled @75% NAID	Unassembled	Largest contig	Contigs >2 kb	Contigs >2 kb
Black Pool	Roche 454	117,653	107,119	10,534	18,588	7,058	986
Octopus	Broad	229,553	164,800	64,753	33,529	31,888	1,580
Great Boiling Spring	JGI	258,950	254,533	4,417	9,547	38,333	429

Viral metagenomes were sequenced by 454 technology by the indicated Genome Center.



Viral Replicases Assembled from 454 Reads



Sequences from the Black Pool Library were assembled and genes identified by GeneMark. Predicted ORFs were annotated by BLASTp analysis. The six largest contigs are shown.

Family A DNA Polymerases

74-like Pol

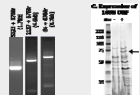


The family of highly related clones identified by functional screening encode an apparent polyprotein with polymerase activity in the C-terminal half and potential accessory functions in the N-terminal half.

3173 Pol Assembled from Sanger Metagenome



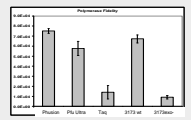
A contig assembled from Sanger data includes the 3173 Pol polyprotein and putative accessory proteins.



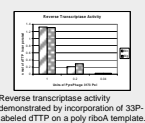
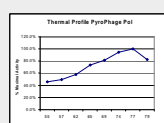
PCR confirms the assemblies.

Expression of the 1608 aa ORF.

5'-3' exonuclease	-
3'-5' exonuclease	Strong
Strand displacement	Strong
Extension from nick	Strong
Thermostability (1% @85°C)	10 min.
Km dNTPs	40 μM
Km DNA	5.3 nM
Processivity	47 nt
Fidelity	7 X 10 ⁴
Template	DNA / RNA

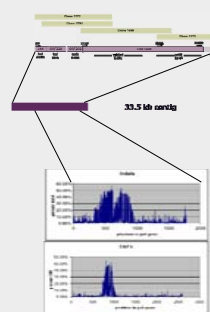


PCR fidelity shown by the Laclq reversion assay.



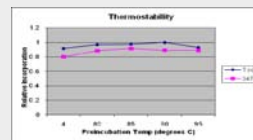
Assembly of 3173 Phage Genome from NextGen Metagenome Sequence

- Octopus Metagenome**
- 164,800 reads
 - 75% NAID assembly
 - Largest contig
 - 68,774 reads
 - 33.5 kb
 - Average 775 reads per nt
 - 37 ORFs
 - Contains the 3173 replication operon.



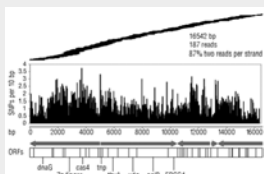
347 Pol – a Novel Replicase Discovered by Functional Metagenomics

- DNA polymerase activity by radioactive incorporation and primer extension assay
- Strongest E value = 0.55 to SSV1 protein
- NTP binding domain
- 35 kD
- Highly thermostable



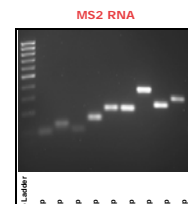
Family B Replicase Genes in a 16.5 kb Contig

- Sanger reads assembled at 50% NAID
- 16.5 kb contig
- Encodes an apparent replication operon including likely replicase subunits.

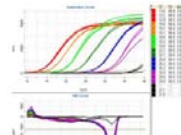


Applications Using PyroPhage 3173 Pol

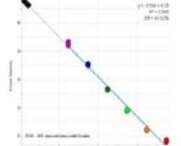
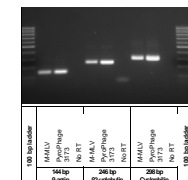
RT-PCR



Real-time RT-qPCR



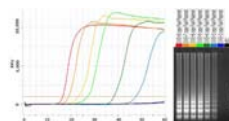
Human transcripts



RT-PCR using PyroScript Mix

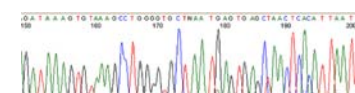
Top left, viral RNA from phage MS2 was amplified by 40 cycles of RT-PCR. Targets of 89 to 362 bp were amplified. Bottom left, total human Liver RNA (1 μg) was reverse transcribed by Moloney Murine Leukemia Virus or by PyroPhage 3173 Pol, then PCR amplified using Lucigen EconoTaq[®] PLUS Master Mix. Top right, the 160 bp primer set used in a 10X dilution series with amplification detected by real-time RT-qPCR. Center right, the melt curve shows high specificity across the dilution series. Bottom right, the Ct values from the real-time data is graphed (log₁₀) vs. the dilutions.

Isothermal Amplification



PyroPhage 3173 Pol in isothermal detection of a serial dilution of a plasmid containing influenza A target sequence

Sanger sequencing



An engineered variant of the PyroPhage 3173 Pol was used in dye terminator Sanger sequencing.

Conclusions

- Viral metagenomes represent a vast, untapped resource for enzyme discovery
- Thermostable PyroPhage 3173 Pol directly RT-PCR amplifies viral RNA and human transcripts.
- Effective for real-time and endpoint RT-PCR analyses.
- PyroPhage RT is also effective in RT-LAMP and Sequencing.



Acknowledgements

This work was supported by NSF and NIH (NAID and NHRI) grants to TS and the Moore foundation. Sequencing of viral metagenomes was performed by JGI, Broad and Roche 454.