Supplemental Data.

Table S1. Thermostable DNA polymerase activity in large insert clones.

Pol clone	Insert length (aa)	Activity (counts per minute per 5 µl)
Empty vector	N/A	578
OCT-1608-4B9	1602	36194
OCT-2H9	1465	18419
OCT-3G9	1423	17320
OCT-3D11	993	20004
OCT-4E10	971	27467
OCT-3D1	807	14020

Figure S1. The sample sites spanned large geographical distances in the western U.S. (Panel A), including Yellowstone National Park, WY (Octopus spring (OCT), Panel B), the Long Valley Caldera, CA (Little Hot Creek (LHC), Panel C), and Gerlach, NV in the northwest Great Basin (Great Boiling Spring (GBS), Panel D).

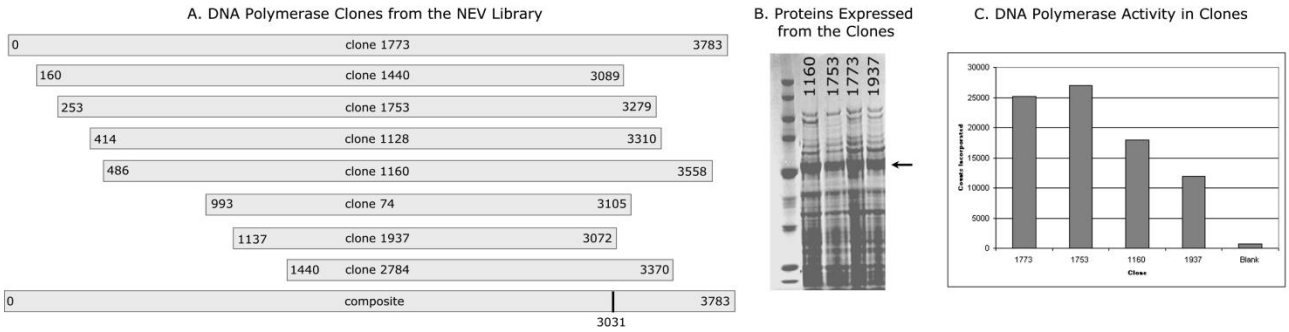

Figure S2. Alignment of the Family A Pols from GBS. The inserts of eight of the shotgun library clones from the functional screen of the GBS library were sequenced in their entirety and aligned using ClusalW (Panel A). Shown are the regions of overlap between the clones and a consensus. Four of the GBS clones were expressed in *E. coli* and the lysates incubated at 70°C for ten minutes to inactivate host activity, and soluble proteins were resolved by SDS PAGE. The arrow indicates the expressed Pol of ~55 KDa (Panel B). The heat-stable cell lysates were assayed for DNA polymerase activity at 70°C (Panel C).

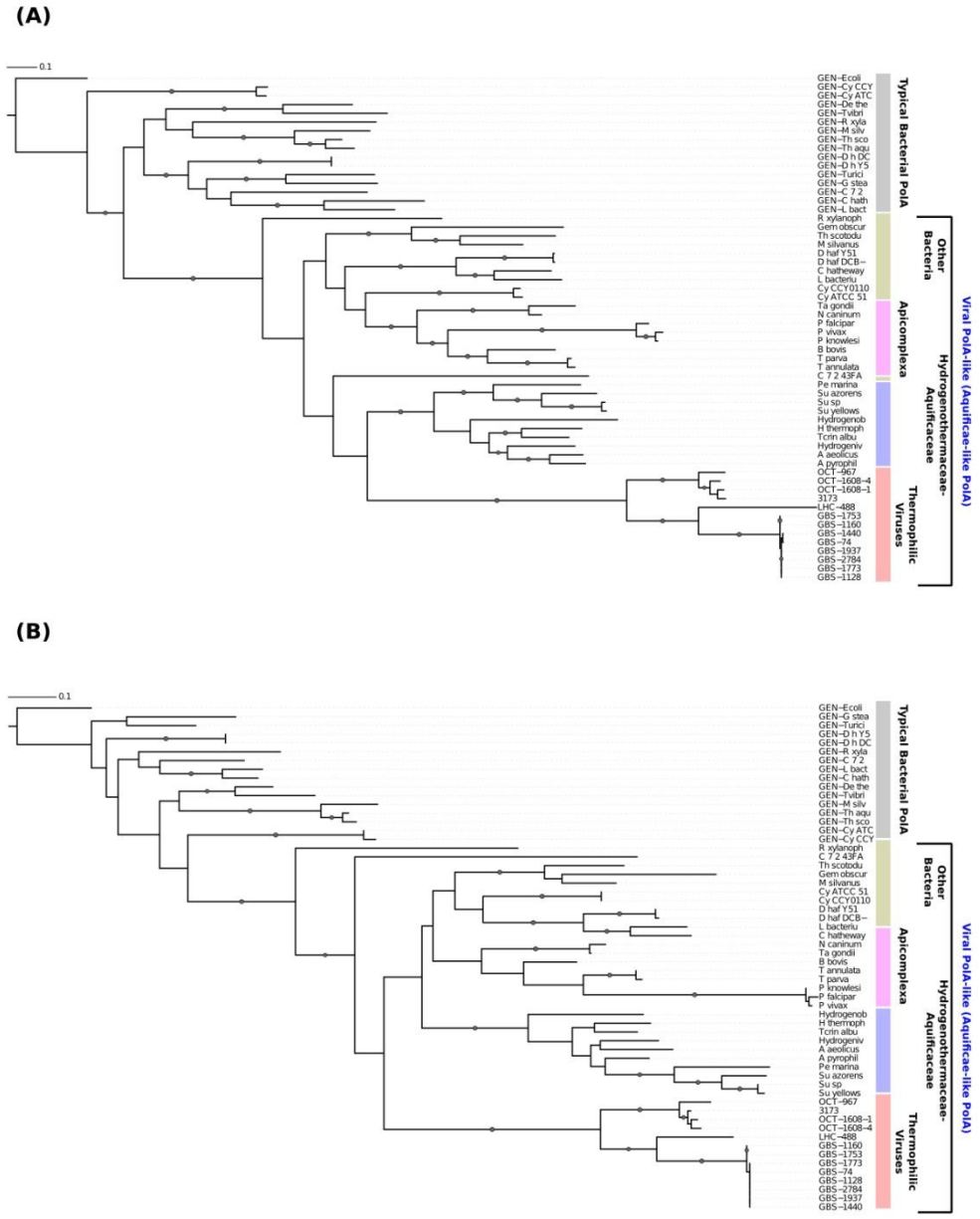

Figure S3. Maximum-likelihood phylogeny of carboxy- terminal region of PolAs without (A) and with (B) a 50% mask. The region used for the analyses consisted of the 588 carboxy-terminal amino acids corresponding to clone OCT-3173 (Figure 2). Branches with >80% bootstrap values (1000 replicates) are indicated. The mask was created using Gblocks (ver 0.91b, (Castresana 2000) using default command line parameters except the minimum percentage of sequences for a flank position (`-b2=' option) was changed from the default 85% to 50%. The Gblocks masked alignment contained 139 amino acid positions, while the unmasked alignment contained 905.

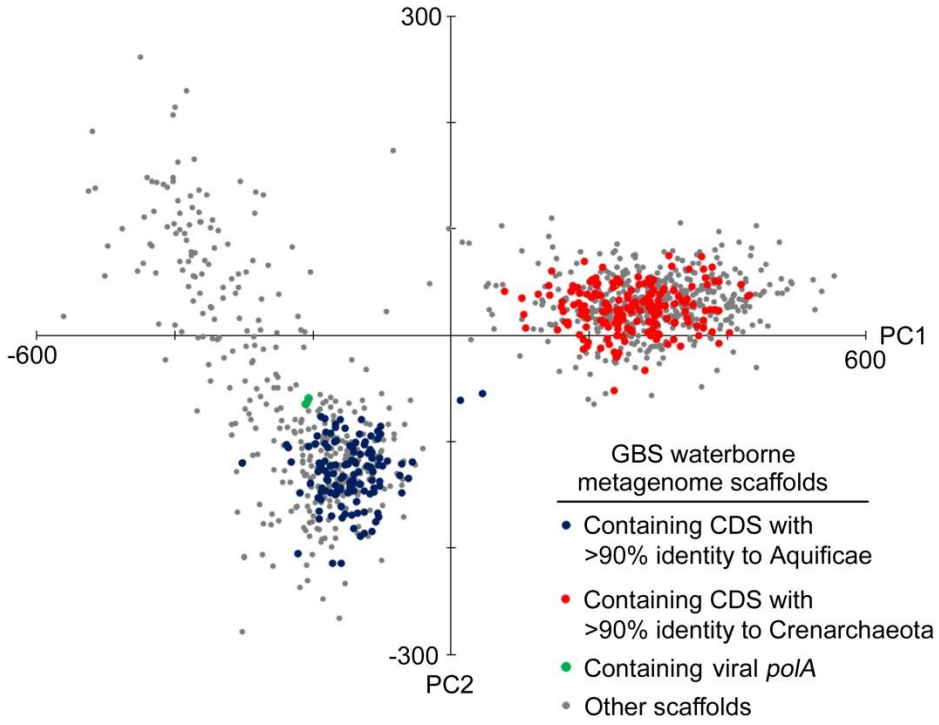

Figure S4. Clustering of contigs in the GBS water metagenome. Scaffolds >1 kb in length from the GBS waterborne cellular metagenome (Taxon Object ID 2084038020, available at http://img.jgi.doe.gov/m) were plotted based on the first two dimensions (PC1 and PC2) of a normalized principle components analysis of their tetranucleotide frequency (using the JCVI Multi-Dimensional Scatter Plot Viewer tool (http://gos.jcvi.org/openAccess/scatterPlotViewer.html) with a chop length of 5 kb and overlap length of 1 kb). Scaffolds containing predicted coding sequences with >90% amino acid identity to predicted proteins in genomes of cultured Crenarchaeota (red; *Pyrobaculum*) and Aquificae (blue; *Thermocrinis*) are indicted, and scaffolds in both the GBS waterborne cellular and viral (Taxon Object ID 2058419004) metagenomes containing viral PolAs are shown in green.