Uracil-DNA Glycosylase-DNA Polymerase Fusion Enzymes Found Primarily in Viruses or Extrachromosomal Elements Diverged before Extant DNA Polymerase A

Ryan Doss[1], Marike Palmer[1,2], Simon Roux[3], Gözde Shipman[1], Joy MacDonald[4], Phil Brumm[4], Alyssa Hassinger[4], Kurt Throckmorton[4], Aaron Lomax[4], David A. Mead[4], & Brian P. Hedlund[1,5]

[1]School of Life Sciences, University of Nevada, Las Vegas, Las Vegas, Nevada, U.S.A.
[2]University of Manitoba, Manitoba, Canada
[3]Department of Energy Joint Genome Institute,
Berkeley, California, U.S.A.
[4]Varizymes, Middleton, Wisconsin, U.S.A
[5]Nevada Institute of Personalized Medicine, Las Vegas, Nevada, U.S.A.

DNA polymerases have complex evolutionary histories that include frequent swapping and/or acquisition of auxiliary domains. A few *Bacillus* phages encode DNA polymerase A (PolA) enzymes fused to an Uracil-DNA Glycosylase (UDG) domain, but these are poorly studied. Here, we describe 9,432 multifunctional UDG-PolA enzymes from public data sampled from diverse environments. Analysis of scaffolds encoding these enzymes with VirSorter2 revealed that 42.3% are classified as viral with high confidence, with most others likely representing some kind of mobile genetic element. To assess functionality, four of these genes were expressed in *Escherichia coli* and revealed DNA polymerase activity, and half of those have exonuclease activity, and currently one has been found to contain UDG activity. Phylogenetic analyses with diverse polymerases in the DNA/RNA polymerase superfamily revealed that these UDG-PolA enzymes and the less common UDG-PolB enzymes diversified at basal nodes within the PolA/PolB phylogeny. This would suggest that the UDG-PolA enzymes were the evolutionary precursors of extant canonical PolAs. We also structurally analyzed the four best-performing enzymes by *in silico* structural predictions using Colabfold to reveal the differences between the UDG-family IV active site and the active sites of these UDG-PolAs. Although the current study exponentially expands the known diversity and environmental distribution of these UDG-PolA enzymes, their roles in extant viruses and mobile genetic elements are yet to be defined.