# Uracil-DNA Glycosylase-DNA Polymerase Fusion Enzymes Found Primarily in Mobile Genetic 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]Department of Microbiology, University of Manitoba, Manitoba, Canada [3]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, U.S.A. [4]Varizymes, Middleton, Wisconsin, U.S.A [5]Nevada Institute of Personalized Medicine, Las Vegas, Nevada, U.S.A.

## Overview

- Enzymes that contain both a uracil-DNA glycosylase and a polymerase domain are ubiquitous in mobile genetic elements, and the evolutionary role these enzymes play is unknown.
- This work conducts a phylogenetic analysis of these enzymes with other polymerases to determine if these fusions occur in extant DNA polymerase A genes, or before the divergence of DNA polymerase A genes.
- Even though these enzymes are ubiquitous, this work is also the first functional analysis of the domains within one of these enzymes.

## Introduction

Uracil DNA-glycosylases (UDGs) aid in repairing the deamination of cytosine to uracil, or uracil misincorporation post-synthesis by inverting uracil from single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) and cleaving the N-glycosidic bond, leaving an abasic site. Metagenomic studies of high-temperature geothermal springs revealed a unique fusion between a UDG domain and DNA polymerase A (PolA) in viral or mobile genetic elements (UDG-PolA). Interrogation of the Joint Genome Institute (JGI) databases revealed that UDG-PolA genes pervade throughout all environments, but only two out of 4898 UDG-PolA sequences are from cultivated microbes, and 11 out of the 4898 UDG-PolA sequences are from cultivated viruses. This interrogation also revealed putative genes encoding a DNA polymerase B (PolB) domain fused to a UDG (UDG-PolB). To date no functional analysis has been conducted on UDG-PolA or UDG-PolB even though they are prevalent in metagenomes.

Here, phylogenetic analyses of the polymerase domains of the RNA/DNA polymerase superfamily were conducted to examine the evolutionary history of UDG-Pol fusions. Assays were also performed to determine the functionality of the UDG and the PolA domains of a UDG-PolA from one of the two cultivated microbes to contain these fusions, the hyperthermophilic bacterium *Fervidibacter sacchari*.

## Results

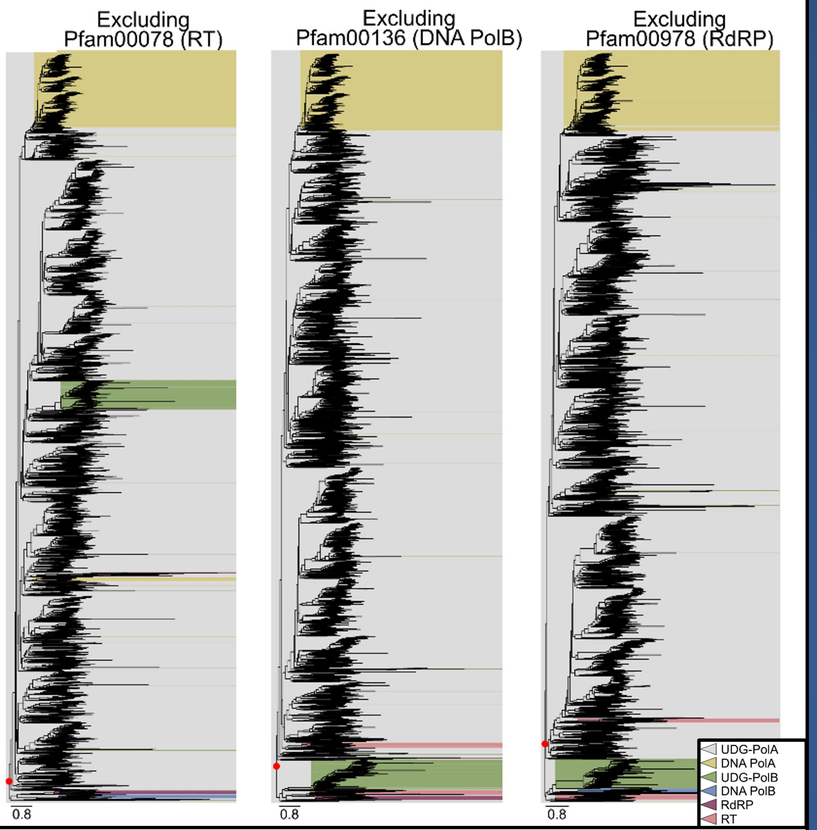



**Figure 1: Phylogeny of polymerase domains**
Previous research indicates that the RNA/DNA polymerase superfamily (PolA, PolB, RNA-dependent RNA polymerase (RdRP), and reverse transcriptase (RT)) evolved from a common β-barrel protein[3]. All UDG-PolA sequences and additional UDG-PolB sequences identified from the JGI IMG/M database, alongside the pfam seed alignment groups of PolA, PolB, RdRP, and RT were aligned using MAFFT-DASH FFT-NS-1, trimmed at a gappyness threshold of 0.95 with MSA-Trimmer, and analyzed through FastTree v. 2. In total, 643 PolA, 29 PolB, 102 RT, 36 RdRP, 4,898 UDG-PolA, and 245 UDG-PolB sequences were included. To verify that the patterns observed are not an artifact of long-branch attraction, each outgroup from the RNA/DNA polymerase superfamily was successively removed before alignment, and the analysis was repeated.

## Results

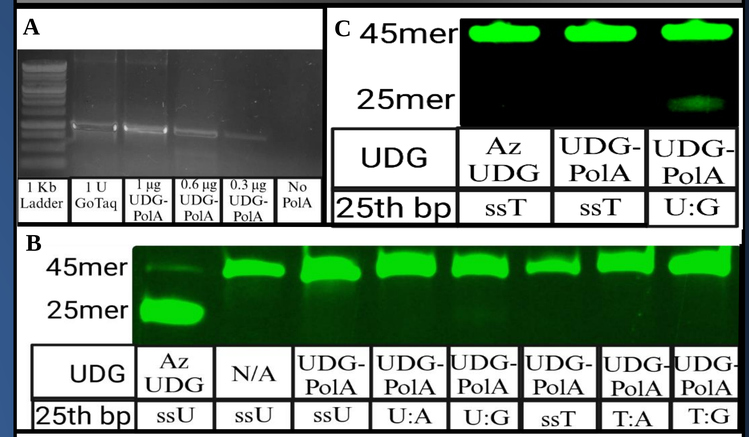
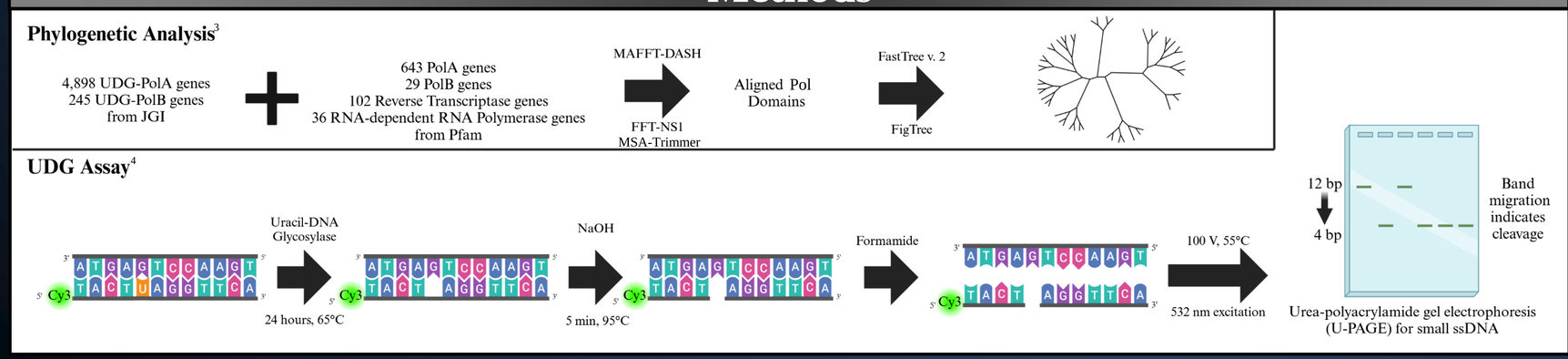



**Figure 2: UDG-PolA domain activity**
All assays used the following buffer: 50 mM Tris-HCl pH 9.0, 2.5 mM MgCl, 20 mM $(NH_4)_2SO_4$, 30% w/v trehalose. A) Polymerase activity was assessed by PCR amplification of 1 Kb of λ DNA. 0.2 mM dNTPs, 0.2 mM forward and reverse primer, and 100 ng λ DNA was added to 25 µL of buffer. Denaturation occurred at 90 °C for 10 seconds, annealing occurred at 57 °C for 30 seconds, extension occurred at 72 °C for 2 minutes, and twenty-five cycles were used in total. Taq polymerase was used as a positive control. B) UDG activity on various DNA substrates was assessed using a custom UDG assay after 24 hours. 100 ng of an oligonucleotide with a Cy3 fluorophore on the 5' end and 3 phosphorothioate bonds at the 3' end was incubated in 25 µL of buffer. ArticZymes cod UDG (AzUDG) was used as a positive control and incubated at 37 °C for 30 minutes instead of 24 hours. There are trace amounts of cleavage occurring at the U:G nucleotide pairing. C) UDG Assay repeated with 0.2% v/v Triton-X for 48 hours.

## Conclusion

- Most UDG-PolA sequences cluster outside of extant PolA sequences, suggesting that UDG-PolA enzymes diverged before extant DNA PolA enzymes. Even when outgroups from the RNA/DNA superfamily are removed, this pattern remains the same, indicating this is not an artifact of long branch attractions.
- The polymerase domain of the UDG-PolA from *F. sacchari* can amplify DNA and is therefore functional.
- The UDG-PolA from *F. sacchari* contains very weak UDG activity, and is seemingly specific for uracil-guanine mismatches, indicating a preference for uracils originating from cytosine deamination repair.
- This is the first look into the evolutionary history and functional activity of one of the ubiquitous UDG-PolA enzymes.
- Further work will be done to optimize the UDG assay for UDG-PoA, and more UDG-PolA enzymes will be analyzed

## Methods




## Acknowledgements & References

*This material is based upon work supported in part by the Nevada NASA Space Grant Consortium under Grant No. 80NSSC20M0043 and in part by a grant from the National Institute of General Medical Sciences (GM103440).*

References:

1. Palmer, M., Doss, R.K., Roux, S. *et al.* 2024. Multifunctional viral UDG-DNA polymerase A enzymes are the ancestor of family A polymerases. *in prep.*
2. Hölz, K., Pavlic, A., Lietard, J. *et al.* 2019. Specificity and efficiency of the uracil DNA glycosylase-mediated strand cleavage surveyed on large sequence libraries. *Sci. Rep.* 9.17822
3. Heli A.M. Mönttinen, H.A.M., Ravantti, J.J, & Poranen, M.M. 2016. Common structural core of three-dozen residues reveals intersuperfamily relationships, *Mol. Biol. and Evol.* 33.7
4. Zhang, L., Jiang, D., Gan, Q. *et al.* 2021. Identification of a novel bifunctional uracil DNA glycosylase from *Thermococcus barophilus* Ch5. *Appl. Microbiol. Biotechnol.* 105.5449–5460