Metagenome mining and functional analysis reveal oxidized guanine DNA repair at the Lost City Hydrothermal Field

Payton H. Utzman¶, Vincent P. Mays¶, Briggs C. Miller, Mary C. Fairbanks, William J. Brazelton, Martin P. Horvath*

School of Biological Sciences, University of Utah, 257 S 1400 E, Salt Lake City, Utah, USA

*Corresponding author: martin.horvath@utah.edu

¶These authors contributed equally to this work.

Introduction

The Lost City Hydrothermal Field (LCHF) resembles conditions of a younger planet and thus provides a window to study the origin of life on Earth and other planets (1–3). Located 15 kilometers from the Mid-Atlantic Ridge, the LCHF comprises a series of carbonate chimneys at depths ranging from 700-800 meters below the ocean surface (3). The temperature and pH of the LCHF are both elevated, with temperature ranging from 40°C to 90°C and pH ranging from 9 to 11 (1). At this depth, light does not penetrate, magmatic energy sources are unavailable, and dissolved carbon dioxide is scarce (1). Despite these environmental constraints, archaea and bacteria inhabit the chimneys and hydrothermal fluids venting from the subsurface (1,4). The chemoautotrophic microbes take advantage of chemical compounds generated by subseafloor geochemical reactions such as serpentinization, the aqueous alteration of ultramafic rocks (1). Serpentinization produces hydrogen gas and low-molecular-weight hydrocarbons, which fuel modern microbial communities and also would have been needed to fuel self-replicating molecules and the emergence of primitive metabolic pathways as an antecedent to cellular life (1,2,5). Hydrothermal circulation underneath the LCHF depletes seawater oxygen, leading to an anoxic hydrothermal environment very different from the nearby oxygen-rich seawater (1,3,6–8). As such, the subsurface microbial communities may offer a glimpse into how life emerged and existed before the Great Oxidation Event that occurred over two billion years ago.

The unusual environmental conditions of the LCHF present several biochemical challenges to the survival of microbes (4,7,8). For example, high temperatures and alkaline pH conditions present at the LCHF potentiate chemistry to generate DNA-damaging, reactive oxygen species (9). However, it remains unclear whether reactive oxygen species (ROS) are a major threat to resident microbes given that the subseafloor underneath the LCHF is largely

41 devoid of molecular oxygen. We reasoned that the prevalence or absence of DNA repair
42 pathways that cope with oxidative damage would provide insight to the question: Are ROS a real
43 and present danger to life at the LCHF?

44 The guanine oxidation (GO) DNA repair system addresses the most common type of
45 DNA damage caused by reactive oxygen species, the 8-oxo-7,8-dihydroguanine (OG)
46 promutagen (**Figure 1**) (10). The OG base differs from guanine by addition of only two atoms,
47 but these change the hydrogen bonding properties so that OG pairs equally well with cytosine
48 and adenine during DNA replication. The resulting OG:A lesions fuel G:C → T:A transversion
49 mutations if not intercepted by the GO DNA repair system (9). The GO system comprises
50 enzymes encoded by mutT, mutM, and mutY, first discovered through genetic analyses of
51 *Escherichia coli* that demonstrated specific protection from G:C → T:A mutations by these three
52 genes (11–14). Homologs or functional equivalents of these GO system components are found
53 throughout all three kingdoms of life (15–17), underscoring the importance of the system, yet
54 there are several instances where particular bacteria (18,19) or eukaryotes (20) make do without
55 one or more of these genes.

56 **Fig 1. Overview of the GO System.** The gene products of mutT, mutM, and mutY (tan bubbles) prevent
57 or repair oxidized guanine DNA damage caused by ROS. DNA glycosylases Fpg and MutY remove OG
58 from OG:C and A from OG:A, respectively to create AP sites with no base. Additional enzymes (AP
59 nucleases, pink; DNA polymerase, orange; and DNA ligase, teal) cooperate with the GO pathway,
60 process these AP sites and ultimately restore the GC base pair. MutT neutralizes OG nucleotide
61 triphosphates to prevent incorporation of the OG nucleotide during DNA replication, thereby ensuring
62 that OG found in DNA is on the parent strand, not the daughter strand.
63

64 Biochemical analyses of the OG system gene products have provided a complete
65 mechanistic picture for the GO repair system. MutT hydrolyzes the OG nucleotide triphosphate
66 to sanitize the nucleotide pool, thus limiting incorporation of the promutagen into DNA by DNA
67 polymerase (13,21). The enzyme encoded by mutM, called formamidopyrimidine-DNA
68 glycosylase (Fpg), locates OG:C base pairs and excises the OG base to initiate base excision

69 repair (BER) (12,22). MutY locates OG:A lesions and excises the A base to initiate BER (11,14).
70 Fpg and MutY thus act separately on two different intermediates to prevent G:C → T:A
71 mutations. These DNA glycosylases generate abasic (apurinic/apyrimidinic; AP) sites, which are
72 themselves mutagenic if not processed by downstream, general BER enzymes, particularly AP
73 nucleases (e.g. exonuclease III and endonuclease IV), DNA polymerase and DNA ligase (17,23–
74 25) as shown in **Figure 1**.

75 MutY is the final safeguard of the GO system. If left uncorrected, replication of OG:A
76 lesions results in permanent G:C → T:A transversion mutations as demonstrated by mutY loss of
77 function mutants (26,27). Underperformance of the mammalian homolog, MUTYH, leads to
78 early onset cancer in humans, first discovered for a class of colon cancers now recognized as
79 MUTYH Associated Polyposis (28). MutY is made up of two domains that both contribute to
80 DNA binding and biochemical functions. The N-terminal catalytic domain shares structural
81 homology with EndoIII and other members of the Helix-hairpin-Helix (HhH) protein
82 superfamily (17). The C-terminal OG-recognition domain shares structural homology with MutT
83 and other NUDIX hydrolase family members (17,29). Functionally important and highly
84 conserved residues define chemical motifs in both domains (**Figure 2**). These chemical motifs
85 interact with the OG:A lesion and chelate the iron-sulfur cluster cofactor as revealed by x-ray
86 structural analysis (**Figure 2**) (30–33). For example, residues in the N-terminal domain establish
87 the catalytic mechanism for adenine excision (**Figure 2A** and **Figure 2B**) (32,34). Residues
88 found in a beta loop of the C-terminal domain recognize the OG base and thus direct adenine
89 removal from OG:A lesions (**Figure 2C**) (33). Some motifs are shared among other DNA
90 glycosylases, such as the residues that chelate the 4Fe4S iron-sulfur cluster cofactor (**Figure 2D**)
91 (17). Chemical motifs particular to MutY, especially the OG-recognition residue Ser 308 (Figure

92 2C) and supporting residues in the C-terminal domain, are conserved across organisms and are
93 not found in other DNA glycosylases and therefore can be used to identify MutY genes (17).

94 **Fig 2. MutY chemical motifs.** Panels A-C show interactions between MutY residues and DNA,
95 with DNA highlighted in blue. Panel A was made from pdb 3g0q, a structure which describes
96 interaction with the competitive inhibitor fluoro-A (31); panels B-D were made from pdb 6u7t, a
97 structure which describes interaction with a transition state analog (1N) of the oxacarbenium ion
98 formed during catalysis of adenine excision (33). MutY catalytic residues (highlighted in green)
99 interact with an adenine base (panel A) and the 1N moiety which mimics the charge and shape of
100 the transition state (panel B). The OG-recognition residues (highlighted in orange) provide
101 hydrogen bonding interactions with the OG base (panel C). Four cysteine residues chelate the
102 iron-sulfur metal cofactor (panel D). All of these interactions are important for MutY activity.
103
104 Our study investigated whether microbes in the anoxic LCHF environment use the GO
105 DNA repair system to mitigate damage caused by reactive oxygen species. It is important to note
106 that not all organisms have an intact GO repair system; examples are missing one, some or all
107 components. MutY in particular was absent frequently in a survey of 699 bacterial genomes (19),
108 and its absence may indicate relaxed evolutionary selection from oxidized guanine damage (35).
109 We mined for homologous genes within the LCHF microbial community and recombinantly
110 expressed candidate MutY enzymes to characterize function. We found genes encoding GO
111 system components and general base excision repair enzymes at all LCHF sites. MutY homologs
112 from the LCHF suppressed mutations when expressed in mutY deficient *Escherichia coli* (*E.*
113 *coli)* strains indicating these function similarly to authentic MutY. These Lost City MutY
114 homologs could be assigned confidently to metagenome assembled genomes (MAG)s, allowing
115 for additional gene inventory analyses that revealed metabolic strategies involving sulfur
116 oxidation and nitrogen reduction. These results have important implications for understanding
117 the repair of oxidative guanine damage in low-oxygen environments, similar to those that existed
118 on a younger Earth, as well as those that may exist on other planets and moons.
119

Results

Identification of the GO DNA repair system in LCHF microbes

To investigate the potential for LCHF microbes to endure DNA damage caused by ROS despite inhabiting a low oxygen environment, we searched for the GO DNA repair system in metagenomes obtained from LCHF hydrothermal fluids (36). We identified gene homologs for MutT, MutM, and MutY, which constitute the complete GO system (**Figure 3)**. The relative abundances of these GO system gene homologs were similar to that of two other DNA repair enzymes that were also frequently found in LCHF metagenomes. Exonuclease III and endonuclease IV work in conjunction with the GO system and perform general functions necessary for all base excision repair pathways, namely the processing of AP sites (24). MutY was underrepresented in each of two samples from a chimney named "Marker 3", indicating that this GO system component is not encoded by some of the LCHF residents (**Figure 3**, pink).

Fig 3. Abundance of GO system gene homologs. Listed on the vertical axis are genes encoding DNA repair enzymes. Genes xthA and nfo genes are generally necessary for DNA repair involving base excision repair in bacteria, including the particular GO system investigated here. Together, mutT, mutM and mutY constitute the GO system that deals specifically with oxidized guanine. Across the horizontal axis are the various LCHF sites, coded by color, from which samples were collected in duplicate, along with the reported temperature and pH. The normalized coverage of each gene is reported as a proportional unit suitable for cross-sample comparisons, the transcripts/fragments per million (TPM).

141 ## Metagenomic mining for MutY genes

142 Having determined that GO system gene homologs are abundant at the LCHF, we
143 focused our efforts on the final safeguard of the pathway, MutY. A BLASTP search against the
144 LCHF metagenome with query MutY sequences from *Geobacillus stearothermophilus (Gs*
145 MutY) and *E. coli* (*Ec* MutY) preliminarily identified 649 putative MutY candidates on the basis
146 of sequence identity (see **Figure 4A**). Structure-guided alignments of these preliminary hits were
147 examined for presence and absence of MutY-defining chemical motifs. We paid particular
148 attention to the chemical motif associated with OG recognition as these residues in the C-
149 terminal domain establish OG:A specificity, which is the hallmark of MutY (37,29,33). This
150 approach authenticated 160 LCHF MutYs (**Figure 4B**). Four representative LCHF MutYs were
151 selected for further analyses described below (red branches in **Figure 4A and Figure 4B**).

152 **Fig 4. Phylogeny for LCHF MutYs**. (A) 649 sequences were identified as LCHF MutY
153 candidates due to sequence similarity to existing MutYs (green branches) and aligned to
154 reconstruct evolutionary relationships. (B) A subset of 160 members contained all necessary
155 MutY-defining chemical motifs. Alignment of these authenticated LCHF MutYs revealed
156 varying evolutionary distances from familiar MutYs and provided a basis for selecting four
157 representative members (red branches).
158

159 **Figure 5** highlights conservation and diversity for the MutY-defining chemical motifs
160 found in the 160 LCHF MutYs. All of the LCHF MutYs retain the chemical motif to coordinate
161 the iron-sulfur cluster cofactor comprising 4 invariant Cys residues (4 Cs, yellow in **Figure 5**), a
162 feature that is also found in other HhH family members such as EndoIII (16), but which is absent
163 for some "clusterless" MutYs (38). Other invariant and highly conserved motifs make critical
164 interaction with the DNA and provide key catalytic functions for adenine base excision,
165 explaining the high degree of sequence conservation at these positions. For example, all LCHF
166 MutYs use a Glu acid which provides acid base catalysis for the mechanism (first E, green in

167 **Figure 5**). Also, all LCHF MutYs use a Gln (first Q, red in **Figure 5**) and a Tyr at position 88
168 (first Y, red in **Figure 5**), to wedge between base pairs to distort the DNA for access to the
169 adenine as seen in x-ray crystal structures of *Gs* MutY (30,32). Structures of *Gs* MutY
170 interacting with a transition state analog revealed close contact with Tyr 126 (second Y, green in
171 Figure 5), Asp 144 (D, green in Figure 5), and supported by Asn 146 (N, green in Figure 5)
172 indicating these chemical motifs stabilize the transition state during catalysis (32–34). For the
173 LCHF MutYs, the Asn residue is invariant, the catalytic Asp is nearly invariant, replaced by
174 chemically similar Glu for five LCHF MutYs, and the catalytic Tyr residue is often replace by
175 Ser and sometimes by Thr or Asn. The residue found between the catalytic Asp and Asn is
176 always a small residue, most often Gly (G, red in Figure 5) but sometimes Ala, Val or Thr.

177 **Fig 5. LCHF MutY chemical motifs.** (A) Conservation and diversity of MutY-defining
178 chemical motifs are depicted with a sequence logo for the 160 LCHF MutYs. These motifs are
179 associated with biochemical functions including DNA binding, enzyme catalysis, attachment of
180 the iron-sulfur cofactor, and recognition of the damaged OG base. Sequence logo generated by
181 Weblogo (39,40). See **Supporting Information** for a complete alignment and a percent identity
182 matrix for the representative LCHF MutYs. (B) Chemical motifs located in MutY as shown with
183 color-coded positions for Gs MutY. (C) Alignment for select chemical motifs highlights
184 conservation among the four representative LCHF MutYs and MutYs from *Ec* and *Gs*.
185
186 By contrast with these numerous and highly conserved motifs for the N-terminal domain,
187 fewer motifs with greater sequence divergence are found in the C-terminal domain. The MutY
188 ancestor is thought to have resulted from a gene fusion event that attached a MutT-like domain to
189 the C-terminus of a general adenine glycosylase enzyme, and the C-terminal domain of modern
190 *E. coli* MutY confers OG:A specificity (29). X-ray crystal structures of *Gs* MutY interacting with
191 OG:A and with G:A highlighted conformational difference for a Ser residue in the C-terminal
192 domain (Ser 308 in *Gs* MutY) (33) and mutational analysis showed this Ser residue and its close
193 neighbors (Phe 307 and His 309 in *Gs* MutY) establish OG specificity (33). Informed by these
194 insights from structure, we eliminated LCHF MutY candidates that lacked a C-terminal domain

195 and its OG-recognition motif. Alignment of the 160 LCHF MutYs that passed this test showed
196 that a second His residue is also well conserved within the H-x-FSH sequence motif (**Figure**
197 **5C**). As is evident from the alignment, there are many variations with residues replaced by close
198 analogs at each position. Ser, which makes the key contact with N7 of OG and no contact with
199 G, is often replaced by Thr, which can make the same H-bond interactions. Likewise the two His
200 residues are each often replaced by polar residues (e.g. Q, N, R or K) that can also H-bond to the
201 DNA phosphate backbone as observed for His 305 and His 309 in *Gs* MutY.

202 Two other positions with high conservation were revealed for the C-terminal domain in
203 this analysis of the 160 LCHF MutYs. These define a L-xxx-P motif. These residues are replaced
204 by other residues with comparable chemical properties. The Leu position is often another
205 hydrophobic residue such as Met or Phe, and the Pro position is most frequently replaced by Glu,
206 a residue that can present aliphatic methylene groups and thus resemble Pro if the polar group
207 hydrogen bonds with the peptide amide. In the structure of *Gs* MutY, the Pro at position 269
208 nucleates a hydrophobic core for the C-terminal domain. The Leu at position 265 makes a strong
209 VDW contact with Tyr89 in the N-terminal domain to support stacking of Tyr88 between bases
210 of the DNA, a molecular contact that suggests communication between the OG-recognition
211 domain and the catalytic domain. Other evolutionary analyses have highlighted the motifs
212 important for DNA contacts, catalysis and OG recognition (17), but the L-xxx-P motif has not
213 been identified previously.

214 Four representative LCHF MutYs were examined for physical properties. **Table 1** reports
215 these physical properties including predicted protein size, isoelectric point (pI), and stability
216 (Tm). Generally the physical characteristics measured for LCHF MutY representatives were
217 comparable to each other and to predicted properties of *Ec* MutY and *Gs* MutY. The predicted

218 Tm for LCHF MutY 3 was above 65℃, distinguishing it as the most stable enzyme (**Table 1**),
219 which may reflect adaptation to a high temperature environment. The isoelectric point predicted
220 for each of the LCHF MutY representatives is 3 pH units above the pI predicted for *Gs* MutY
221 and between 0.1 - 0.5 pH units above the pI predicted for *Ec* MutY, indicating that more
222 numerous positively charged residues have been recruited, possibly as an adaptation to alkaline
223 pH conditions of the LCHF.
224
225 **Table 1.** Protein Properties.

MutY	Length (residues)	MW (kDa)	pI	Predicted Tm (°C) (N-domain; C-domain)
Gs MutY	372	41.8	5.3	55 - 65 (55 - 65 ; < 55)
Ec MutY	355	39.1	8.6	< 55 (< 55 ; 55 - 65)
LCHF MutY 1	358	39.0	9.1	55 - 65 (55 - 65 ; < 55)
LCHF MutY 2	352	38.3	8.8	< 55 (55 - 65 ; < 55)
LCHF MutY 3	370	42.0	8.7	> 65 (>65 ; 55 - 65)
LCHF MutY 4	376	44.0	9.0	55 - 65 (55 - 65 ; 55 - 65)

226
227 **Identification of LCHF MutY organisms, gene neighbors,**
228 **environmental conditions, and metabolic strategies**

229 Our next objective was to identify the organisms from which these LCHF MutY enzymes
230 originated. Each of the four representative LCHF MutY sequences were derived from contiguous
231 DNA sequences (contigs) belonging to a metagenome-assembled genome (MAG) representing a

232 LCHF microbe. The taxonomic classification of these MAGs indicated that LCHF MutY 1
233 originated from a species of *Marinosulfonomonas,* LCHF MutY 2 from the family
234 *Rhodobacteraceae*, LCHF MutY 3 from the family *Thiotrichaceae*, and LCHF MutY 4 from the
235 family *Flavobacteriaceae* (**Figure 6**). The taxonomic classification of each contig was consistent
236 with the classification of the MAG to which it belonged, supporting the idea that the MutY gene
237 is a long term resident and not a recent arrival through phage infection or some other horizontal
238 gene transfer mechanism. For the remainder of this work we will refer to the MutY-encoding
239 organisms as the lowest-level classification that was determined for each LCHF MutY (e.g.
240 LCHF MutY 3 will now be referred to as *Thiotrichaceae* MutY).

Figure 6. Taxonomic classification. LCHF MutY-encoding contigs were found in several branches of bacteria. The classification places these in relation to MutY of *G. stearothermophilus* and *E. coli* (accession IDs included). LCHF MutYs were mapped to their respective microbes by two methods indicated by the two arrows (see text for details).

241

242 The inclusion of MutY contigs in MAGs provided an opportunity to examine gene
243 neighbors for the representative LCHF MutYs. The GO repair genes are located at distant loci in
244 *E. coli* (12), and belong to separate, complex operons (41). However, MutY is the immediate 5'-
245 neighbor to YggX within gammaproteobacteria, and homologs of YggX are present outside this
246 lineage, occasionally nearby to MutY (e.g. *Bacillus subtilis*). As gene neighbors, MutY and
247 YggX are part of a SoxRS regulated operon in *E. coli* (41,42). YggX provides oxidative stress
248 protection and iron transport function with a critical Cys residue close to the N-terminus of this
249 small protein (43,44). A protein matching these features is encoded by a gene partly overlapping
250 with and the nearest 3' neighbor to *Thiotrichaceae* MutY (see **Supporting Information**).
251 To reveal the environmental conditions of these MutY-encoding organisms, we analyzed
252 the sequence coverage of each LCHF MutY contig at each of the sampling sites.

253 *Marinosulfonomonas* MutY, *Thiotrichaceae* MutY, and *Flavobacteriaceae* MutY were identified
254 at all sampling sites, ranging from 21°C to 58°C and pH 8.1 to 9.5. *Rhodobacteraceae* MutY was
255 present at all sampling sites excluding Marker 3 and was therefore found in temperatures ranging
256 from 24°C to 52°C and pH 8.1 to 9.5.

257 We further investigated the metabolic strategies utilized by MutY-encoding microbes by
258 examining the inventory of predicted protein functions in each MAG (**Table 2,** see also
259 **Supporting Information**). Each LCHF MutY-containing MAG possessed at least two forms of
260 cytochrome oxidase, with the exception of the *Flavobacteriaceae* MAG. The *Flavobacteriaceae*
261 MAG is only 44% complete, however, so no strong conclusions can be made regarding the
262 absence of genes. Cytochrome oxidases commonly provide sources of free radicals and are
263 essential to aerobic metabolism. Predicted proteins indicative of dissimilatory nitrate and nitrite
264 reduction were found in the *Marinosulfonomonas, Rhodobacteraceae,* and *Thiotrichaceae*
265 MAGs*,* suggesting that these organisms may be capable of using nitrate or nitrite as alternative
266 electron acceptors when oxygen is not available. Furthermore, the *Marinosulfonomonas* and
267 *Rhodobacteraceae* MAGs include predicted protein functions associated with the oxidation of
268 reduced sulfur compounds, though it is important to note that the directionality of these reactions
269 cannot be fully determined from bioinformatics alone. These patterns speak to the potential
270 origins of oxidants within the MutY-encoding organisms as discussed below.
271
272
273
274
275

276 **Table 2.** Metabolic genes identified in LCHF MutY organisms

	Symbol	KEGG ID	*Marinosulfonomonas*[a] MAG 1	*Marinosulfonomonas*[a] MAG 2	*Rhodobacteraceae*	*Thiotrichaceae*	*Flavobacteriaceae*
Cytochrome Oxidases	UQCRFS1	K00411	X	X	X	X	
	coxA	K02274			X	X	
	ccoN	K00404	X	X	X		
	cydA	K00425		X			
	cyoB	K02298			X		
Sulfur Oxidation	soxA	K17222	X	X	X		
	soxX	K17223	X	X	X		
	soxB	K17224	X	X	X		
	soxC	K17225	X	X	X		
	soxY	K17226	X		X		
	soxZ	K17227	X		X		
Nitrogen Reduction	narG	K00370	X	X	X		
	narH	K00371	X	X	X		
	nirB	K00362		X	X	X	
	nirD	K00363		X		X	
	nirK	K00368			X		
	norB	K04561			X		
	norC	K02305			X		
	nosZ	K00376			X		
	MAG Completeness[b]		88.37	88.21	93.69	66.12	44.27
	MAG Contamination[b]		16.43	0.61	1.36	11.75	1.65

277
278 [a]*Marinosulfonomonas* MutY belongs to two separate MAGs and genes for each are reported separately.
279 [b]Completeness and contamination scores generated by CheckM v1.0.5 as described in Brazelton *et al.* 2022 (36).
280

281 # Predicted protein structures and virtual docking experiments

282 To assess the likelihood of the LCHF MutY sequences folding into enzymes capable of
283 activity on the OG:A substrate, protein structures were predicted using AlphaFold2 (45) (**Figure**
284 **7**). These predicted structures were associated with high confidence (see Supporting Information
285 figures for pLDDT scores). Superpositions revealed that the predicted structures for the LCHF

286 MutYs are each highly comparable with the experimentally determined structure for *Gs* MutY as
287 indicated by visual inspection (**Figure 7**) and by low, pairwise root mean square deviation
288 (RMSD) values (**Table 3**). The whole protein superpositions were dominated by the larger, more
289 structurally conserved N-terminal domain. Breaking the analysis into two separate domains
290 showed that the C-terminal domain, although more plastic, retained core structural features that
291 could be superimposed. The MutY-defining chemical motifs are positioned in locations similar
292 to those seen for the *Gs* MutY reference structure, providing evidence these LCHF MutY
293 enzymes are capable of recognizing OG:A lesions and excising the adenine base. Concisely, the
294 LCHF MutY structure predictions resemble a functional MutY enzyme from a thermophilic
295 bacterium.

296 **Figure 7**. **Structure predictions and virtual docking of MutY ligands**. (A) The x-ray crystal
297 structure of *Gs* MutY (blue ribbon; pdb id 3g0q) in complex with DNA (white helix) highlights
298 the positions of the adenine nucleotide (red) and the OG nucleotide (yellow). (B-F) Virtual
299 docking of ligands. Adenine and OG were separately docked to identify binding surfaces for
300 these ligands in the structures of *Gs* MutY (B) which served as a positive control and the four
301 representative LCHF MutYs: *Marinosulfonomonas* MutY (C); *Rhodobacteraceae* MutY (D);
302 *Thiotrichaceae* MutY (E); and *Flavobacteriaceae* MutY (F).
303
304

305 **Table 3. Superposition and virtual docking.**

MutY	RMSD to *Gs* MutY (Å)[a] (number of residues)			Binding affinity (kcal/ mol)[b]	
	Whole protein	NTD	CTD	Adenine	OG
Gs MutY (6U7T)[c]	0.53 (338)	0.39 (211)	0.54 (122)	-5.8	-5.7
Marinosulfonomonas MutY	0.95 (180)	0.95 (192)	0.95 (65)	-5.1	-5.5
Rhodobacteraceae MutY	0.96 (199)	0.95 (192)	0.83 (47)	-5.4	-5.3
Thiotrichaceae MutY	1.11 (243)	0.99 (198)	1.00 (64)	-5.0	-5.7
Flavobacteriaceae MutY	1.01 (260)	0.87 (200)	0.95 (48)	-5.4	-5.8

306 [a] Superposition with *Gs* MutY (pdb id 3G0Q) was calculated for all residues and separately for N-terminal domain
307 (NTD) and for C-terminal domain (CTD) by *ChimeraX* (46).
308
309 [b]Binding affinity for the best outcome from docking adenine to the enzyme active site and OG to the OG-
310 recognition calculated by *Autodock VINA* (47)
311
312 [c] Reference superposition values provided by comparing a second structure of the same *Gs* MutY (pdb id 6u7t)
313

314 We performed virtual docking experiments to examine the potential for molecular
315 interaction with adenine and OG ligands. MutY scans DNA looking for the OG:A base-pair by
316 sensing the major-groove disposition of the exocyclic amine of the OG base in its syn
317 conformation (48,49). After this initial encounter, the enzyme bends the DNA, flips the adenine
318 base from the DNA double helix into the active site pocket, and positions OG in its anti
319 conformation as seen in structures of the enzyme complexed to DNA (30,31). Thus, multiple
320 conformations and orientations for the OG and adenine ligands were anticipated. Representative
321 docking outcomes are shown in **Figure 7** with binding affinities reported in **Table 3**. The
322 adenine and OG ligands docked to surfaces found in the active site of the N-terminal domain
323 and close to the OG-recognition motif found in the C-terminal domain of each of the
324 representative LCHF MutY enzymes, similar to docking outcomes obtained with *Gs* MutY

325 (**Figure 7**). As anticipated, the precise orientation and position for these docked ligands varied,
326 and none exactly match the disposition of the adenine base or OG base as presented in the
327 context of double stranded DNA. Nevertheless, binding affinities for the ligand-LCHF MutY
328 complexes ranged from -5.0 to -5.8 kcal/mol indicating favorable interactions are attainable and
329 similar to the binding affinities measured for *Gs* MutY by the same virtual docking method.
330

331 Testing mutation suppression activity of LCHF MutY enzymes 332 by recombinant expression

333 With *in silico* data suggesting that the LCHF MutY representatives encoded functional
334 MutY enzymes capable of proper folding and interaction with OG:A lesions, we recombinantly
335 expressed the genes in *E. coli* and measured mutation suppression activity *in vivo*. Three of the
336 representative LCHF MutYs were successfully cloned into the pKK223 expression plasmid as
337 verified by Sanger sequencing. The *Flavobacteriaceae* MutY appeared to be toxic to *E. coli* as
338 only mutant versions of the gene were obtained from multiple cloning attempts, a situation that is
339 reminiscent of *Gs* MutY, which is also apparently toxic to *E. coli* and could not be cloned into
340 pKK223 (33).
341 To test the mutation suppression activity of *Marinosulfonomonas (Ms),*
342 *Rhodobacteraceae (Rb), and Thiotrichaceae (Tt)* MutYs, we measured mutation rates with a
343 rifampicin resistance assay (50). Several, independent, single-point mutations within the gene
344 encoding RNA polymerase beta-subunit (*rpoB*) prevent the antibiotic rifampicin from binding to
345 RpoB. Thus, spontaneous Rif^R mutants arising in overnight cultures can be counted by the
346 colonies that emerge on rifampicin containing plates. Cultures expressing functional MutY
347 delivered by plasmid DNA transformation have low Rif^R frequency compared to the high Rif^R

348 frequency characterizing the reporter strain that lacks MutY and MutM (see **Materials and**
349 **Methods**).

350 Cultures with an empty plasmid (*null*) and cultures with a plasmid encoding *Ec* MutY
351 showed significant differences in the frequency of Rif^R mutants, with median values of 101 and
352 12, respectively, indicating the assay was fit for use (significance determined by non-overlap of
353 median 95% confidence intervals) (**Figure 8** and **Supporting information**). *Ms* MutY*, Rb*
354 MutY*,* and *Tt* MutY each demonstrated significant mutation suppression activity when compared
355 to the *null*. Indeed, *Rb* MutY showed mutation suppression performance equivalent to that
356 measured for *Ec* MutY, and *Ms* MutY was apparently better at suppressing mutations than *Ec*
357 MutY, a remarkable outcome given the evolutionary time separating these species (**Figure 8**). *Tt*
358 MutY showed partial function. Cultures expressing *Tt* MutY suppressed Rif^R mutants to 50% of
359 the rate observed for *null* cultures, but allowed a mutation rate about 4-fold greater than that
360 measured for cultures expressing *Ec* MutY (**Figure 8** and **Supporting information**).

361 **Figure 8**. **Functional analysis.** Bars represent median Rif^R colony counts for *E. coli* cultures
362 expressing MutY, MutY variants, or no MutY (*null*) from a plasmid DNA. Error bars represent
363 95% confidence intervals as determined by bootstrap sampling (see Supporting Information for
364 tabulated values). *Ms* MutY, *Rb* MutY, and *Tt* MutY each suppressed mutations as evidenced by
365 non-overlap of Rif^R confidence intervals compared to *null* cultures. Altered versions of each
366 LCHF MutY tested the importance of residues for catalysis and OG-recognition. Designated
367 *catalysis-* and *recognition-* along the X-axis, these alterations severely impacted mutation
368 suppression function indicating the LCHF MutYs share mechanistic features with the extensively
369 studied enzymes *Ec* MutY and *Gs* MutY.
370

371 To investigate the biochemical mechanism employed by LCHF MutY enzymes, we
372 altered residues essential for OG:A recognition and catalysis, then repeated the mutation
373 suppression assay (10, 12). Two mutants of each LCHF MutY were constructed through site-
374 directed substitution of residues. One set of substitutions was designed to disable the OG-
375 recognition motif by replacing F(S/T)H residues (**Figure 2 and Figure 5**) with alanine residues

376 (designated *recognition*-); the other set of substitutions was designed to disable catalysis by
377 replacing the active site Asp and Glu residues with structurally similar, but chemically inert Asn
378 and Gln residues (*catalysis*-). For all three LCHF MutYs, these targeted substitutions disabled
379 mutation suppression function *in vivo* as shown by elevated Rif^R frequencies for cultures
380 expressing the *recognition*- and *catalysis*- versions. The mutation frequencies for cultures
381 expressing these site-specific substitution variants were comparable to the Rif^R frequencies
382 measured for *null* cultures as judged by overlapping 95% confidence intervals (**Figure 8** and
383 **Supporting Information**). These results indicate that the representative LCHF MutYs suppress
384 mutations by a mechanism that is highly similar to the strategy executed by *Ec* MutY and *Gs*
385 MutY.
386

387 # Discussion

388 To gain insight into DNA repair strategies in early Earth-like environments, we
389 investigated the status of the GO DNA repair system within microbes inhabiting the LCHF. Our
390 approach included mining of metagenomic data, bioinformatic comparisons informed by
391 structure and mechanistic understanding, predictive molecular modeling, and functional
392 analyses. The degree to which this approach succeeded was dependent on the assembly of
393 metagenomic sequences into contigs long enough to contain full-length genes (36). Earlier
394 attempts to search for MutY genes within older LCHF metagenomes with shorter contigs yielded
395 a number of hits, but these were truncated and therefore missing critical motifs, explaining weak
396 mutation suppression function (unpublished results). The longer contigs utilized in this study
397 allowed us to capture entire MutY genes, bin these MutY-encoding contigs into MAGs to assess

398 associated gene inventories, and thereby infer metabolic strategies for the microbes expressing
399 the GO DNA repair components.

400 Within the initial set of 649 LCHF MutY candidates identified by sequence identity, 160
401 genes encoded proteins with all of the chemical motifs known to be important for MutY function
402 from X-ray crystal structures and mechanistic studies available for *Ec* MutY and *Gs* MutY
403 (37,29–32,51,33,48). Indeed, leveraging this extensive body of knowledge allowed us to select
404 features such as sequence length, presence of MutY motifs, and structural prediction to
405 distinguish LCHF MutYs from other members of the helix-hairpin-helix superfamily.
406 Recombinant expression in *E. coli* revealed that LCHF MutY representatives suppress mutations
407 *in vivo* by a mechanism that depends on the catalytic and OG-recognition motifs (**Figure 8**),
408 strongly suggesting these are functional enzymes that actively seek and initiate repair of OG:A
409 lesions within their respective LCHF microbes. Toxicity observed for one LCHF gene encoding
410 *Flavobacteriaceae* MutY, which could not be cloned except with disabling nonsense mutations,
411 underscores the risks and dangers posed by MutY and DNA glycosylases in general, which
412 initiate DNA repair by damaging the DNA further, creating AP sites that are themselves
413 destabilizing (52). The potential for lethal outcomes makes cross-species function observed for
414 *Ms* MutY*, Rb* MutY*,* and *Tt* MutY across vast evolutionary time all the more remarkable.

415 Retained function across evolutionary and species barriers strongly suggests that MutY
416 interacts with the base excision repair apparatus through some well-preserved mechanism that
417 relies on a universal language understood by all organisms. Most critically, the AP sites
418 generated by MutY should be recognizable to downstream AP nucleases. Protein-protein
419 interaction between AP nucleases and MutY has been discussed as a possible mechanism (53–
420 56), but on its own such a mechanism would rely on coevolution of protein partners. Our results

421 and those reported by others for complementation with the eukaryote homolog MUTYH (57)
422 speaks for a mechanism that is less sensitive to sequence divergence. Therefore, we favor a
423 model where the distorted DNA structure created by MutY signals the location of the AP site for
424 handoff to the BER apparatus, as has been suggested previously (55).

425 *Thiotrichaceae* MutY underperformed in our functional evaluation despite coming from a
426 gammaproteobacterium most closely related to *E. coli,* compared with the other recombinantly
427 expressed LCHF MutYs. *Thiotrichaceae* MutY may be adapted to physical or biochemical
428 conditions different from those that can support growth of the *E. coli* reporter strain. In support
429 of this idea, the *Thiotrichaceae* MutY is predicted to be ultrastable with respect to temperature as
430 reflected in the theoretical T_m greater than 65℃ (**Table 1**, LCHF MutY 3). Indeed, *Geobacillus*
431 *stearothermophilus* MutY, from an extreme thermophile, has been shown to perform poorly at
432 37 °C and optimally at a much higher temperature of 60 ℃ (33). Thus, lower mutation
433 suppression performance observed for *Tt* MutY may simply be due to the difference in
434 temperature needed for the *in vivo* experiment and the temperature of the habitat where
435 *Thiotrichaceae* thrives at the LCHF. In future work, biochemical characterization of purified
436 LCHF MutY enzymes at high temperatures and high pH could address this question directly.

437 Our metagenomic analysis revealed that gene homologs encoding the GO DNA repair
438 system are abundant in basement microbes inhabiting the LCHF (**Figure 3**). This observation is
439 surprising given that the basement of the LCHF is expected to be anoxic (1,3). The chemical
440 agents commonly thought of for producing oxidized guanine (OG) are ROS derived from
441 molecular oxygen *via* aerobic metabolism. In an anoxic environment, what chemical agents are
442 producing OG and how are these generated? Models of hydrothermal field chemistry predict
443 abiotic production of ROS which the microbial residents may encounter, although these would

444 probably react with cell protective structures before encountering DNA (58). Continual mixing
445 of seawater with the anoxic hydrothermal fluid could provide molecular oxygen at the interface
446 where hydrothermal fluids vent into ambient seawater at the seafloor (1,3). Facultative anaerobes
447 at this interface would inevitably generate ROS (61–66), and therefore benefit from the GO
448 DNA repair system.

449 Another explanation for the source of OG in basement microbes of the LCHF, which is
450 suggested by our gene inventory analysis (**Table 2** and **Supporting Information**), involves
451 reactive sulfur species (RSS) and reactive nitrogen species (RNS). Many of the basement-
452 dwelling microbes within the LCHF appear to metabolize sulfur and nitrogen for energy
453 conservation (4,36), strategies that generate RSS and RNS as metabolic byproducts (67,68).
454 Indeed, mechanisms for the oxidation of guanine by both RSS and RNS have been described,
455 including the formation of 8-oxoguanine (OG) and chemically similar 8-nitroguanine, which
456 templates for adenine in a fashion similar to OG (69–72). The oxidation of guanine by RSS and
457 RNS generated from microbial metabolism would produce OG independent of molecular oxygen
458 and thereby necessitate the GO DNA repair system for both facultative and obligate anaerobes
459 inhabiting the LCHF.

460 Whether organisms developed biochemical systems to deal with oxidative damage before
461 or after the Great Oxidation Event (GOE) remains an open question (73,74). It is reasonable to
462 think of these systems arising in response to the selective pressure of oxidative damage from
463 rising O_2 levels. However, it is also possible that these systems were already in place as a coping
464 mechanism for other oxidants and were repurposed to deal with the new source of oxidizing
465 agents when O_2 became readily available. Indeed, obligate anaerobes contain many of the same
466 pathways to deal with oxidative damage as aerobes (73,74). Our discovery of the GO DNA

467 repair system in basement dwellers of the LCHF adds to this body of evidence and supports the
468 hypothesis that oxidative damage repair systems were established before the GOE. We
469 considered the caveat of possible phage-mediated gene transfer – modern microbes adapted to
470 oxygen-rich regions elsewhere may be the source of LCHF MutYs. However, the
471 correspondence of taxonomic assignments based on MutY sequence and based on MAGs and the
472 high degree of sequence diversity seen for LCHF MutY enzymes are inconsistent with expansion
473 of GO DNA repair in the LCHF by horizontal gene transfer. Thus, it seems more likely that
474 LCHF microbes inherited the GO DNA repair system from a common ancestor and retained it
475 through necessity, even in the absence of extrinsic O_2.
476

477 # Conclusion

478 Performing empirical studies on how life may have evolved on Earth and other planets is
479 inherently difficult due to time and spatial barriers. Unique sites such as the LCHF serve as
480 representatives of these theorized environments (3,75). By discovering the GO DNA repair
481 system at the LCHF and validating mutation suppression function by LCHF MutYs, we infer that
482 microbes within the anoxic environment of the LCHF basement are under evolutionary pressure
483 to repair OG lesions. The evolutionary pressure and source of OG appear to be driven by
484 nitrogen reactive species or sulfur reactive species as supported by the metabolic survey of the
485 MutY-encoding organisms. These results highlight the need for DNA-based life to manage
486 oxidized guanine damage even in anoxic environments. Moreover, this work adds evidence for
487 the more general hypothesis that life established biochemical systems to deal with oxidative
488 damage early, well before the GOE, and should be considered when developing an evolutionary
489 model for early life.

Materials and Methods

Metagenomic Sequencing and Analysis of LCHF Fluid Samples

Generation, assembly, and annotation of metagenomes from the Lost City Hydrothermal Field (LCHF) were described in Brazelton et al. 2022 (36) and are briefly summarized here. In 2018, the remotely operated vehicle (ROV) Jason collected samples of fluids venting from chimneys at the LCHF, which is located near the Mid-Atlantic Ridge at 30 ºN latitude and a depth of ~800 m. Whole-genome community sequences ("metagenomes") were generated from the fluid samples, and assembled metagenomic contigs were binned into metagenome-assembled genomes (MAGs). Potential gene homologs encoding enzymes involved in the GO DNA repair system were identified by conducting KEGG (76) orthology assignment using the BlastKOALA v2.2 program (77). The selected genes that were identified include: MutT (KEGG ID: K03574), MutM (K10563), and MutY (K03575) along with the genes xthA (K01142) and nfo (K01151), which encode exonuclease III and endonuclease IV, respectively.

The relative abundance of each GO repair pathway gene homolog at each LCHF chimney location was calculated as the normalized metagenomic sequence coverage, determined by mapping of reads from each fluid sample against the pooled assembly. Coverages are reported as transcripts (or metagenomic fragments) per million (TPM), which is a proportional unit suitable for comparisons of relative abundances between samples (36,78).

510 ## Identification of LCHF MutYs

511 Candidate MutY genes in the LCHF metagenomes were identified with a BLASTP (79)
512 search against predicted protein sequences from the LCHF pooled metagenomic assembly using
513 MutY queries from *Geobacillus stearothermophilus* (*Gs* MutY) (NCBI Accession ID: P83847.2)
514 and *Escherichia coli* (*Ec* MutY) (NCBI Accession ID: CDP76921.1). The diversity of these
515 candidates, visualized by aligning the sequences along with *Gs* MutY and *Ec* MutY with Clustal
516 Omega (80) and iTOL (81), was used to build an initial phylogeny. LCHF MutY candidate
517 sequences were then aligned by PROMALS3D (82), guided by the structure of *Gs* MutY (pdb id
518 6u7). Sequence diversity in the C-terminal domain prevented reliable alignment of this region in
519 this first pass at structure-guided alignment. To overcome this challenge, sequences were split
520 into two parts: one part with all residues before position V147 in the *Gs* MutY protein which
521 were reliably aligned, and the second part with all residues following position N146 in the *Gs*
522 MutY protein which were aligned inconsistently in the first pass. These two parts were
523 separately resubmitted for alignment by PROMALS3D guided by the corresponding portions of
524 the crystal structure. For inclusion in this alignment, the C-terminal part was required to pass a
525 minimum length criteria of 160 residues. The resulting alignment was inspected for the MutY-
526 defining chemical motifs described in the text, and a phylogeny was constructed for the 160
527 authenticated LCHF MutYs with iTOL (81). Selection of the four representative LCHF MutYs
528 was guided by this phylogeny and by the completeness of each associated MAG.

529 ## Taxonomic classification

530 Contiguous DNA sequences containing the LCHF MutY representatives the contig DNA
531 were assigned taxonomic classifications using the program MMseqs2 (83) and the Genome
532 Taxonomy Database (GTDB) as described in Brazelton et al., 2022 (36). Taxonomic

533 classification of each MAG that included a contig of interest was performed with GTDB-Tk
534 v1.5.1 (84). The environmental distributions of MutY-encoding MAGs were inspected for
535 potential signs of contamination from ambient seawater during sample collection. This
536 possibility was ruled out by the absence of all MutY-encoding taxa reported in this study in the
537 background seawater samples collected by Brazelton et al (36). MAG completeness and
538 contamination scores were generated by CheckM v1.0.5 as described in Brazelton et al. 2022
539 (36).

540 Prediction of physical parameters

541 The theoretical molecular weights and pIs of the LCHF MutY representatives and known
542 MutY sequences were generated via ExPASY (85). The theoretical melting temperature of the
543 representatives was calculated with the Tm predictor from the Institute of Bioinformatics and
544 Structural Biology, National Tsing-Hua University (86).

545 Molecular modeling

546 Protein structures for the LCHF MutY representatives were predicted by the AlphaFold2
547 plugin within UCSF ChimeraX (33, 34). Predicted structures were superimposed with crystal
548 structure of *Gs* MutY (pdb 3G0Q) using the MatchMaker tool in ChimeraX to generate RMSD
549 (Å) for pruned atom pairs. Initial superpositions were dominated by residues in the N-terminal
550 domain. To fairly compare structures for the more diverse C-terminal domains, the linker region
551 between domains was identified by inspection, and superposition with *Gs* MutY was repeated
552 using the MatchMaker function with appropriate selection of residues in the N-terminal domain
553 and in the C-terminal domain.

554 Ligand docking experiments were executed with the program AutoDock Vina (47).
555 Ligand structures representing adenine and OG were prepared with the ligand preparation tools

556 implemented with Autodock Tools, and receptors structures were prepared from the structures
557 predicted by AlphaFold or from the crystal structure of *Gs* MutY with the receptor preparation
558 tools, also as implemented with Autodock Tools (87,88). Receptor and ligand structures were
559 treated as rigid objects. Separate 20 x 20 x 20 $Å^3$ search volumes were defined for adenine and
560 for OG. The adenine search volume was centered on the position of residue A5L:18 in chain C of
561 the *Gs* MutY crystal structure, and the OG search volume was centered on the position of residue
562 8OG:6 in chain B.

563 **Recombinant DNA cloning**

564 Synthetic genes encoding the LCHF MutYs were codon optimized for expression in *E.*
565 *coli* except that pause sites with rare codons were engineered so as to retain pause sites found in
566 the gene encoding *Ec* MutY. Gene Blocks were ordered from Integrated DNA Technologies
567 (IDT). LCHF MutY genes designed in this way were cloned into the low-expression pKK223
568 vector by ligation-independent cloning (LIC) as follows. PCR reactions with the high-fidelity
569 Phusion polymerase amplified the synthetic Gene Blocks and two overlapping fragments derived
570 from approximate halves of the pKK223 plasmid. PCR products were separated by
571 electrophoresis in 0.8% agarose x1 TAE gels containing 1 μg/mL ethidium bromide. DNA was
572 visualized by long-wavelength UV shadowing to allow dissection of gel bands, and the DNA
573 was purified with the GeneJet gel extraction system, treated with Dpn1 at 37 °C for 45 min, and
574 heat shock transformed directly into DH5α competent cells. Clones were selected on ampicillin
575 media plates. The plasmid DNA was purified from 4-mL overnight cultures by use of the Wizard
576 Plus MiniPrep kit according to the manufacturer's instructions. The sequence of the LCHF MutY
577 encoding gene was verified by Sanger sequencing with UpTac and TacTerm primers. Genes
578 encoding site-directed substitution variants were created by amplifying two overlapping

579 fragments of the LCHF MutY-pKK223 plasmid with mutagenic PCR primers followed by
580 similar gel purification and transformation procedures. In our hands the LIC cloning efficiency
581 was close to 95% except for the *Flavobacteriaceae* MutY gene which could not be cloned intact.

582 ## Mutation suppression assay

583 Mutation rates were measured by the method outlined previously (50). The CC104
584 *mutm::KAN muty::TET* cells (89) were heat-shock transformed with a pKK223 plasmid encoding
585 the *Ec* MutY gene, LCHF MutY genes, or no gene (*null*). Transformants selected from
586 kanamycin-ampicillin-tetracycline (KAT) media plates were diluted prior to inoculation of 2-mL
587 KAT liquid media, and these cultures were grown overnight for 18 hours at 37°C with shaking at
588 180 rpm. Cultures were kept cold on ice or at +4 °C prior to further processing. Cells were
589 collected by centrifugation, the media was removed by aspiration, and cells were resuspended in
590 an equal volume of 0.85% sodium chloride before seeding 100 µL aliquots to kanamycin-
591 ampicillin-rifampicin (KAR) media plates. Dilutions of the washed cells were also seeded to
592 KAR plates (10^{-1} dilution) and kanamycin-ampicillin (KA) plates (10^{-7} dilution), and allowed to
593 overnight at 37°C for 18 hours. The number of Rif^R mutants were counted by counting the colony
594 forming units (CFU).

595 Statistical analysis was performed in *R* as previously described (33). Confidence intervals
596 were obtained by bootstrap resampling of 10,000 trials as implemented with the boot package.
597 Data and *R* programs, including the CFU data from the Rif^R assay, the statistical analyses, the
598 code to generate the bar graph shown in **Figure 8**, and the values reported in **Supporting Table**
599 **S7**, have been deposited at the GitHub repository https://github.com/paytonutzman/Lost-City-
600 MutY-Discovery**.**

Acknowledgements

This work was supported by NSF awards to MPH (CHE:CLP- 1905249, 2204229) and to W.J.B. (OCE-1536405) and by the NASA Astrobiology Institute Rock-Powered Life team. We thank Markel Kolendrianos, Peyton Russelberg, and Sonia Sehgal for technical expertise. We thank the University of Utah undergraduate students enrolled in Molecular Biology of DNA Lab (BIOL3525 fall 2022) for their contribution to measuring Rif^R mutant frequency data. In particular, we would like to thank the following students for their extra efforts and contributions in the lab: Tieker Duncan, Shilpi Kharidia, Jackson Munn, Kenzie Fleming, Alex Ballinger, Andrew Petersen, Brook Miller, Tom Christensen, Madison Haught, Emi Wickens, Spencer Sonntag, Abigail Johnston, Sam Aamodt, Jasmine Jacobo, Alyssa Le, Sam Hendry, Saydra Galloway, Hiroshi Aoki, Peyton Merchant, Kaliece Carter, Annie Joseph, Kathleen Brabb, Natalie Morgan, Sophia Khalaji, Helena Haddadin, Hadlee Young, Brenden Roberts, Mason Hansen, Mackenzie Montzingo, and Quyen Tran. We especially acknowledge Emily Dart who first searched LCHF metagenomes for MutY homologs which provided the impetus for development of this project.

References

1. Kelley DS, Karson JA, Früh-Green GL, Yoerger DR, Shank TM, Butterfield DA, et al. A Serpentinite-Hosted Ecosystem: The Lost City Hydrothermal Field. Science. 2005 Mar 4;307(5714):1428–34.
2. Amador ES, Bandfield JL, Brazelton WJ, Kelley D. The Lost City Hydrothermal Field: A Spectroscopic and Astrobiological Analogue for Nili Fossae, Mars. Astrobiology. 2017 Nov;17(11):1138–60.
3. Kelley DS, Karson JA, Blackman DK, Früh-Green GL, Butterfield DA, Lilley MD, et al. An off-axis hydrothermal vent field near the Mid-Atlantic Ridge at 30° N. Nature. 2001 Jul;412(6843):145–9.
4. Brazelton WJ, Schrenk MO, Kelley DS, Baross JA. Methane- and Sulfur-Metabolizing Microbial Communities Dominate the Lost City Hydrothermal Field Ecosystem. Appl

Environ Microbiol. 2006 Sep;72(9):6257–70.
5. Russell MJ, Hall AJ, Martin W. Serpentinization as a source of energy at the origin of life. Geobiology. 2010;8(5):355–71.
6. Proskurowski G, Lilley MD, Seewald JS, Früh-Green GL, Olson EJ, Lupton JE, et al. Abiogenic Hydrocarbon Production at Lost City Hydrothermal Field. Science. 2008 Feb;319(5863):604–7.
7. Lang SQ, Butterfield DA, Lilley MD, Paul Johnson H, Hedges JI. Dissolved organic carbon in ridge-axis and ridge-flank hydrothermal systems. Geochimica et Cosmochimica Acta. 2006 Aug 1;70(15):3830–42.
8. Lang SQ, Brazelton WJ. Habitability of the marine serpentinite subsurface: a case study of the Lost City hydrothermal field. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2020 Feb 21;378(2165):20180429.
9. Bruskov VI, Malakhova LV, Masalimov ZK, Chernikov AV. Heat-induced formation of reactive oxygen species and 8-oxoguanine, a biomarker of damage to DNA. Nucleic Acids Res. 2002 Mar 15;30(6):1354–63.
10. Kino K, Hirao-Suzuki M, Morikawa M, Sakaga A, Miyazawa H. Generation, repair and replication of guanine oxidation products. Genes Environ. 2017 Aug 1;39:21.
11. Nghiem Y, Cabrera M, Cupples CG, Miller JH. The mutY gene: a mutator locus in Escherichia coli that generates G.C----T.A transversions. Proceedings of the National Academy of Sciences. 1988 Apr;85(8):2709–13.
12. Cabrera M, Nghiem Y, Miller JH. mutM, a second mutator locus in Escherichia coli that generates G.C----T.A transversions. J Bacteriol. 1988 Nov;170(11):5405–7.
13. Maki H, Sekiguchi M. MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis. Nature. 1992 Jan;355(6357):273–5.
14. Michaels ML, Cruz C, Grollman AP, Miller JH. Evidence that MutY and MutM combine to prevent mutations by an oxidatively damaged form of guanine in DNA. Proc Natl Acad Sci U S A. 1992 Aug 1;89(15):7022–5.
15. Prakash A, Doublié S, Wallace SS. Chapter 4 - The Fpg/Nei Family of DNA Glycosylases: Substrates, Structures, and Search for Damage. In: Doetsch PW, editor. Progress in Molecular Biology and Translational Science [Internet]. Academic Press; 2012 [cited 2023 Mar 20]. p. 71–91. (Mechanisms of DNA Repair; vol. 110). Available from: https://www.sciencedirect.com/science/article/pii/B9780123876652000043
16. Denver DR, Swenson SL, Lynch M. An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases. Mol Biol Evol. 2003 Oct;20(10):1603–11.
17. Trasviña-Arenas CH, Demir M, Lin WJ, David SS. Structure, function and evolution of the Helix-hairpin-Helix DNA glycosylase superfamily: Piecing together the evolutionary puzzle of DNA base damage repair mechanisms. DNA Repair. 2021 Dec 1;108:103231.
18. Kuwahara H, Takaki Y, Shimamura S, Yoshida T, Maeda T, Kunieda T, et al. Loss of genes for DNA recombination and repair in the reductive genome evolution of thioautotrophic symbionts of Calyptogena clams. BMC Evol Biol. 2011 Oct 3;11(1):285.
19. Garcia-Gonzalez A, Rivera-Rivera R, Massey S. The Presence of the DNA Repair Genes mutM, mutY, mutL, and mutS is Related to Proteome Size in Bacterial Genomes. Frontiers in Genetics. 2012;3:3.
20. Jansson K, Blomberg A, Sunnerhagen P, Alm Rosenblad M. Evolutionary loss of 8-oxo-G repair components among eukaryotes. Genome Integrity. 2010 Sep 1;1(1):12.
21. Fowler RG, Schaaper RM. The role of the mutT gene of Escherichia coli in maintaining

674 replication fidelity. FEMS Microbiology Reviews. 1997 Aug 1;21(1):43–54.
675 22. Boiteux S, O'Connor TR, Laval J. Formamidopyrimidine-DNA glycosylase of Escherichia
676 coli: cloning and sequencing of the fpg structural gene and overproduction of the protein.
677 The EMBO Journal. 1987 Oct;6(10):3177–83.
678 23. Shida T, Noda M, Sekiguchi J. Cleavage of single- and double-stranded DNAs containing an
679 abasic residue by Escherichia coli exonuclease III (AP endonuclease VI). Nucleic Acids Res.
680 1996 Nov 15;24(22):4572–6.
681 24. Parikh SS, Mol CD, Tainer JA. Base excision repair enzyme family portrait: integrating the
682 structure and chemistry of an entire DNA repair pathway. Structure. 1997 Dec
683 15;5(12):1543–50.
684 25. Ljungquist S, Lindahl T, Howard-Flanders P. Methyl methane sulfonate-sensitive mutant of
685 Escherichia coli deficient in an endonuclease specific for apurinic sites in deoxyribonucleic
686 acid. J Bacteriol. 1976 May;126(2):646–53.
687 26. Lind PA, Andersson DI. Whole-genome mutational biases in bacteria. Proceedings of the
688 National Academy of Sciences. 2008 Nov 18;105(46):17878–83.
689 27. Foster PL, Lee H, Popodi E, Townes JP, Tang H. Determinants of spontaneous mutation in
690 the bacterium Escherichia coli as revealed by whole-genome sequencing. Proceedings of the
691 National Academy of Sciences. 2015 Nov 3;112(44):E5990–9.
692 28. Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, et al.
693 Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal
694 tumors. Nat Genet. 2002 Feb;30(2):227–32.
695 29. Noll DM, Gogos A, Granek JA, Clarke ND. The C-Terminal Domain of the Adenine-DNA
696 Glycosylase MutY Confers Specificity for 8-Oxoguanine·Adenine Mispairs and May Have
697 Evolved from MutT, an 8-Oxo-dGTPase. Biochemistry. 1999 May 1;38(20):6374–9.
698 30. Fromme JC, Banerjee A, Huang SJ, Verdine GL. Structural basis for removal of adenine
699 mispaired with 8-oxoguanine by MutY adenine DNA glycosylase. Nature. 2004
700 Feb;427(6975):652–6.
701 31. Lee S, Verdine GL. Atomic substitution reveals the structural basis for substrate adenine
702 recognition and removal by adenine DNA glycosylase. Proceedings of the National
703 Academy of Sciences. 2009 Nov 3;106(44):18497–502.
704 32. Woods RD, O'Shea VL, Chu A, Cao S, Richards JL, Horvath MP, et al. Structure and
705 stereochemistry of the base excision repair glycosylase MutY reveal a mechanism similar to
706 retaining glycosidases. Nucleic Acids Res. 2016 Jan 29;44(2):801–10.
707 33. Russelburg LP, O'Shea Murray VL, Demir M, Knutsen KR, Sehgal SL, Cao S, et al.
708 Structural Basis for Finding OG Lesions and Avoiding Undamaged G by the DNA
709 Glycosylase MutY. ACS Chem Biol. 2020 Jan 17;15(1):93–102.
710 34. Demir M, Russelburg LP, Lin WJ, Trasviña-Arenas CH, Huang B, Yuen PK, et al. Structural
711 snapshots of base excision by the cancer-associated variant MutY N146S reveal a retaining
712 mechanism. Nucleic Acids Research. 2023 Jan 12;gkac1246.
713 35. Kuwahara H, Takaki Y, Shimamura S, Yoshida T, Maeda T, Kunieda T, et al. Loss of genes
714 for DNA recombination and repair in the reductive genome evolution of thioautotrophic
715 symbionts of Calyptogena clams. BMC Evol Biol. 2011 Oct 3;11(1):285.
716 36. Brazelton WJ, McGonigle JM, Motamedi S, Pendleton HL, Twing KI, Miller BC, et al.
717 Metabolic Strategies Shared by Basement Residents of the Lost City Hydrothermal Field.
718 Applied and Environmental Microbiology. 2022 Aug 11;88(17):e00929-22.
719 37. Porello SL, Leyes AE, David SS. Single-turnover and pre-steady-state kinetics of the

720 reaction of the adenine glycosylase MutY with mismatch-containing DNA substrates.
721 Biochemistry. 1998 Oct 20;37(42):14756–64.
722 38. Trasviña-Arenas CH, Lopez-Castillo LM, Sanchez-Sandoval E, Brieba LG. Dispensability of
723 the [4Fe-4S] cluster in novel homologues of adenine glycosylase MutY. FEBS J. 2016
724 Feb;283(3):521–40.
725 39. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator.
726 Genome Res. 2004 Jun;14(6):1188–90.
727 40. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences.
728 Nucleic Acids Res. 1990 Oct 25;18(20):6097–100.
729 41. Gifford CM, Wallace SS. The Genes Encoding Formamidopyrimidine and MutY DNA
730 Glycosylases in Escherichia coli Are Transcribed as Part of Complex Operons. J Bacteriol.
731 1999 Jul;181(14):4223–36.
732 42. Pomposiello PJ, Koutsolioutsou A, Carrasco D, Demple B. SoxRS-regulated expression and
733 genetic analysis of the yggX gene of Escherichia coli. J Bacteriol. 2003 Nov;185(22):6624–
734 32.
735 43. Gralnick J, Downs D. Protection from superoxide damage associated with an increased level
736 of the YggX protein in Salmonella enterica. Proc Natl Acad Sci U S A. 2001 Jul
737 3;98(14):8030–5.
738 44. Gralnick JA, Downs DM. The YggX Protein of Salmonella enterica Is Involved in Fe(II)
739 Trafficking and Minimizes the DNA Damage Caused by Hydroxyl Radicals: RESIDUE
740 CYS-7 IS ESSENTIAL FOR YggX FUNCTION*. Journal of Biological Chemistry. 2003
741 Jun 6;278(23):20708–15.
742 45. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold:
743 making protein folding accessible to all. Nat Methods. 2022 Jun;19(6):679–82.
744 46. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF
745 ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci.
746 2021 Jan;30(1):70–82.
747 47. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking
748 Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model. 2021 Aug
749 23;61(8):3891–8.
750 48. Manlove AH, McKibbin PL, Doyle EL, Majumdar C, Hamm ML, David SS. Structure–
751 Activity Relationships Reveal Key Features of 8-Oxoguanine: A Mismatch Detection by the
752 MutY Glycosylase. ACS Chem Biol. 2017 Sep 15;12(9):2335–44.
753 49. Nelson SR, Kathe SD, Hilzinger TS, Averill AM, Warshaw DM, Wallace SS, et al. Single
754 molecule glycosylase studies with engineered 8-oxoguanine DNA damage sites show
755 functional defects of a MUTYH polyposis variant. Nucleic Acids Res. 2019 Apr
756 8;47(6):3058–71.
757 50. Majumdar C, Nuñez NN, Raetz AG, Khuu C, David SS. Cellular assays for studying the Fe-
758 S cluster containing base excision repair glycosylase MUTYH and homologs. Methods
759 Enzymol. 2018;599:69–99.
760 51. Berti PJ, McCann JAB. Toward a detailed understanding of base excision repair enzymes:
761 transition state and mechanistic analyses of N-glycoside hydrolysis and N-glycoside transfer.
762 Chem Rev. 2006 Feb;106(2):506–55.
763 52. Loeb LA, Preston BD. Mutagenesis by Apurinic/Apyrimidinic Sites. Annual Review of
764 Genetics. 1986;20(1):201–30.
765 53. Michaels ML, Tchou J, Grollman AP, Miller JH. A repair system for 8-oxo-7,8-

dihydrodeoxyguanine. Biochemistry. 1992 Nov 17;31(45):10964–8.
54. van der Veen S, Tang CM. The BER necessities: the repair of DNA damage in human-adapted bacterial pathogens. Nat Rev Microbiol. 2015 Feb;13(2):83–94.
55. Pope MA, Porello SL, David SS. Escherichia coli Apurinic-Apyrimidinic Endonucleases Enhance the Turnover of the Adenine Glycosylase MutY with G:A Substrates *. Journal of Biological Chemistry. 2002 Jun 21;277(25):22605–15.
56. Lu AL, Lee CY, Li L, Li X. Physical and functional interactions between Escherichia coli MutY and endonuclease VIII. Biochem J. 2006 Jan 1;393(Pt 1):381–7.
57. Komine K, Shimodaira H, Takao M, Soeda H, Zhang X, Takahashi M, et al. Functional Complementation Assay for 47 MUTYH Variants in a MutY-Disrupted Escherichia Coli Strain. Hum Mutat. 2015 Jul;36(7):704–11.
58. Shaw TJ, Luther GW, Rosas R, Oldham VE, Coffey NR, Ferry JL, et al. Fe-catalyzed sulfide oxidation in hydrothermal plumes is a source of reactive oxygen species to the ocean. Proceedings of the National Academy of Sciences. 2021 Oct 5;118(40):e2026654118.
59. Brazelton WJ, Nelson B, Schrenk MO. Metagenomic Evidence for H2 Oxidation and H2 Production by Serpentinite-Hosted Subsurface Microbial Communities. Front Microbiol. 2012 Jan 6;2:268.
60. Brazelton WJ, Ludwig KA, Sogin ML, Andreishcheva EN, Kelley DS, Shen CC, et al. Archaea and bacteria with surprising microdiversity show shifts in dominance over 1,000-year time scales in hydrothermal chimneys. Proceedings of the National Academy of Sciences. 2010 Jan 26;107(4):1612–7.
61. Hayyan M, Hashim MA, AlNashef IM. Superoxide Ion: Generation and Chemical Implications. Chem Rev. 2016 Mar 9;116(5):3029–85.
62. Imlay JA. A Metabolic Enzyme That Rapidly Produces Superoxide, Fumarate Reductase of Escherichia coli(∗). Journal of Biological Chemistry. 1995 Aug 25;270(34):19767–77.
63. Messner KR, Imlay JA. The Identification of Primary Sites of Superoxide and Hydrogen Peroxide Formation in the Aerobic Respiratory Chain and Sulfite Reductase Complex of Escherichia coli *. Journal of Biological Chemistry. 1999 Apr 9;274(15):10119–28.
64. Kussmaul L, Hirst J. The mechanism of superoxide production by NADH:ubiquinone oxidoreductase (complex I) from bovine heart mitochondria. Proceedings of the National Academy of Sciences. 2006 May 16;103(20):7607–12.
65. Korshunov S, Imlay JA. Two sources of endogenous H2O2 in Escherichia coli. Mol Microbiol. 2010 Mar;75(6):1389–401.
66. Imlay JA. The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat Rev Microbiol. 2013 Jul;11(7):443–54.
67. Rodionov DA, Dubchak IL, Arkin AP, Alm EJ, Gelfand MS. Dissimilatory Metabolism of Nitrogen Oxides in Bacteria: Comparative Reconstruction of Transcriptional Networks. PLOS Computational Biology. 2005 Oct 28;1(5):e55.
68. Han S, Li Y, Gao H. Generation and Physiology of Hydrogen Sulfide and Reactive Sulfur Species in Bacteria. Antioxidants. 2022 Dec;11(12):2487.
69. Attene-Ramos MS, Wagner ED, Gaskins HR, Plewa MJ. Hydrogen Sulfide Induces Direct Radical-Associated DNA Damage. Molecular Cancer Research. 2007 May 1;5(5):455–9.
70. Joyner-Matos J, Predmore BL, Stein JR, Leeuwenburgh C, Julian D. Hydrogen Sulfide Induces Oxidative Damage to RNA and DNA in a Sulfide-Tolerant Marine Invertebrate. Physiological and Biochemical Zoology. 2010 Mar;83(2):356–65.
71. Bhamra I, Compagnone-Post P, O'Neil I, Iwanejko L, Bates A, Cosstick R. Base-pairing

812 preferences, physicochemical properties and mutational behaviour of the DNA lesion 8-
813 nitroguanine†. Nucleic acids research. 2012 Sep 10;40.
814 72. Suzuki N, Yasui M, Geacintov NE, Shafirovich V, Shibutani S. Miscoding Events during
815 DNA Synthesis Past the Nitration-Damaged Base 8-Nitroguanine. Biochemistry. 2005 Jun
816 1;44(25):9238–45.
817 73. Lu Z, Imlay JA. When anaerobes encounter oxygen: mechanisms of oxygen toxicity,
818 tolerance and defence. Nat Rev Microbiol. 2021 Dec;19(12):774–85.
819 74. Ślesak I, Ślesak H, Kruk J. Oxygen and Hydrogen Peroxide in the Early Evolution of Life on
820 Earth: In silico Comparative Analysis of Biochemical Pathways. Astrobiology. 2012
821 Aug;12(8):775–84.
822 75. Shock EL, Schulte MD. Organic synthesis during fluid mixing in hydrothermal systems.
823 Journal of Geophysical Research: Planets. 1998;103(E12):28513–27.
824 76. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids
825 Research. 2000 Jan 1;28(1):27–30.
826 77. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for
827 Functional Characterization of Genome and Metagenome Sequences. Journal of Molecular
828 Biology. 2016 Feb 22;428(4):726–31.
829 78. Thornton CN, Tanner WD, VanDerslice JA, Brazelton WJ. Localized effect of treated
830 wastewater effluent on the resistome of an urban watershed. GigaScience. 2020 Nov
831 10;9(11):giaa125.
832 79. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J
833 Mol Biol. 1990 Oct 5;215(3):403–10.
834 80. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of
835 high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems
836 Biology. 2011 Jan;7(1):539.
837 81. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree
838 display and annotation. Nucleic Acids Research. 2021 Jul 2;49(W1):W293–6.
839 82. Pei J, Kim BH, Grishin NV. PROMALS3D: a tool for multiple protein sequence and
840 structure alignments. Nucleic Acids Res. 2008 Apr;36(7):2295–300.
841 83. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the
842 analysis of massive data sets. Nat Biotechnol. 2017 Nov;35(11):1026–8.
843 84. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes
844 with the Genome Taxonomy Database. Bioinformatics. 2020 Mar 1;36(6):1925–7.
845 85. Walker J, Bairoch A, Wilkins MR, Appel RD, Duvaud S, Gattiker A, et al. Protein
846 Identification and Analysis Tools on the Expasy Server. In: The Proteomics Protocols
847 Handbook. 2005. p. 571–607. (Springer Protocols Handbooks (SPH)).
848 86. Ku T, Lu P, Chan C, Wang T, Lai S, Lyu P, et al. Predicting melting temperature directly
849 from protein sequences. Computational Biology and Chemistry. 2009 Dec 1;33(6):445–50.
850 87. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4
851 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J Comput
852 Chem. 2009 Dec;30(16):2785–91.
853 88. Forli S, Huey R, Pique ME, Sanner M, Goodsell DS, Olson AJ. Computational protein-
854 ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc. 2016
855 May;11(5):905–19.
856 89. Cupples CG, Miller JH. A set of lacZ mutations in Escherichia coli that allow rapid detection
857 of each of the six base substitutions. Proc Natl Acad Sci U S A. 1989 Jul;86(14):5345–9.

Supporting Information

Data and code availability

The results of the rifampicin resistance assay and R code to produce the reported statistics can be found at https://github.com/paytonutzman/Lost-City-MutY-Discovery. Additionally, FASTA files of the discovered MutY enzymes and MutYcandidates can be found within this repository.

S1 Dataset. Alignment of Lost City MutY homologs

Chemical motifs are highlighted in columns. Alignment was generated by Promals3D (81) guided by the structure of *Gs* MutY. It was necessary to align sequences in the first block including up to N146 separately from the second block and third block because otherwise the C-terminal domain residues were aligned inconsistently. The homologs flagged with dark red highlighting were eliminated because of missing chemical motifs. The homolog flagged with light pink highlighting required manual adjustment so as to align the H-x-FSH motif.

S2 Table. Percent Identity Matrix.

S3 Fig. MutY gene neighbors. Structural homology for *Ec* YggX (left) and nearest neighbor to *Thiotrichaceae* MutY (right). The structure solution NMR structure for *Ec* YggX (pdb id 1yhd) is superimposable to the structure predicted by AlphaFold2 for the nearest neighbor to *Thiotrichaceae* MutY, with RMSD of 0.99 Å for 69 pruned pairs selected from 88 possible pairs. The Cys residue critical for function is highlighted with a yellow sphere for the sulfhydryl group.

S4 Table. Metabolic gene identification

[a]*Marinosulfonomonas* MutY contig belongs to two separate MAGs and each are reported separately as MAG 1 and 2 respectively.
[b]KEGG ID gene not identified in any MAG and not reported in Table 2.
[c]Completeness and contamination scores generated by CheckM v1.0.5 as described in Brazelton et al 2022 (36).

A KEGG ID analysis was used to identify the potential metabolic strategies of the MutY encoding organisms at the LCHF. The full metabolic KEGG ID search is shown above.

S5 Fig. AlphaFold Structure Prediction pLDDT Scores

pLDDT scores represent the confidence in the prediction calculated by AlphaFold2.

S6 Table. Ligand binding affinity (kcal / mol)

*Binding affinities for each of the nine output modes generated by AutoDock Vina for molecular docking of the ligand to a protein structure. Each mode represents a predicted ligand pose, which differs by a combination of position, orientation, and rotamer conformation. The receptor structure was obtained from pdb id 3g0q for *Gs* MutY and through structure prediction for the LCHF *Marinosulfonomas* MutY, *Rhodobacteraceae* MutY, *Thiotrichaceae* MutY, and *Flavobacteriaceae* MutY**.**

903

904 **S7 Table. Rifampicin resistance assay.**

905 [a]95% Confidence intervals determined by bootstrap method, see Materials and Methods for details.
906 [b]Mutation Frequency reported as median number of resistant colonies per 10^8 viable colonies. Fold change was
907 calculated by dividing Rif^R frequency by the frequency measured for cultures expressing *Ec* MutY.
908
909

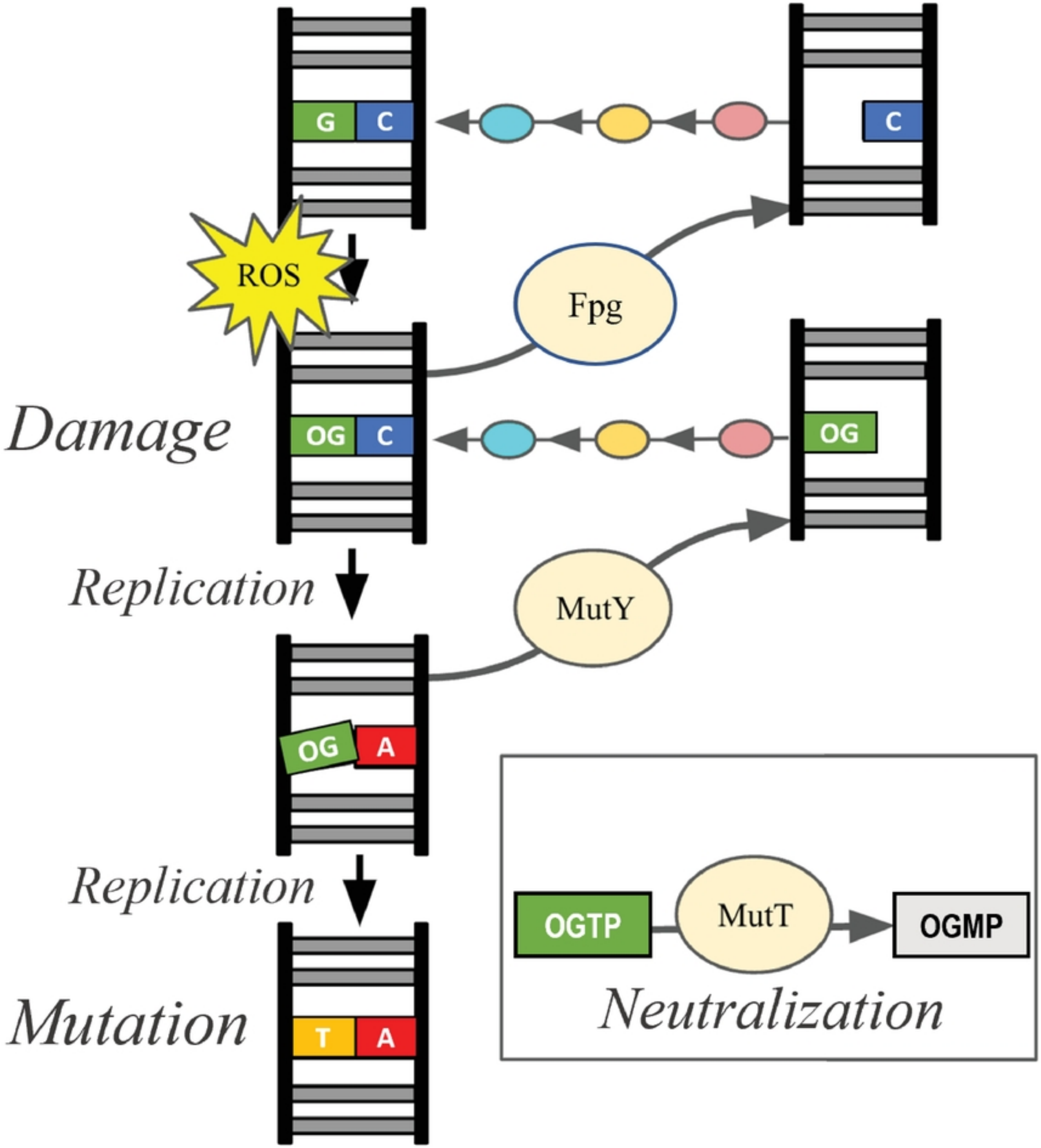

G
C
C
ROS
Fpg
Damage
OG
C
OG
Replication
MutY
OG
A
Replication
OGTP
MutT
OGMP
Mutation
T
A
Neutralization

Figure 1

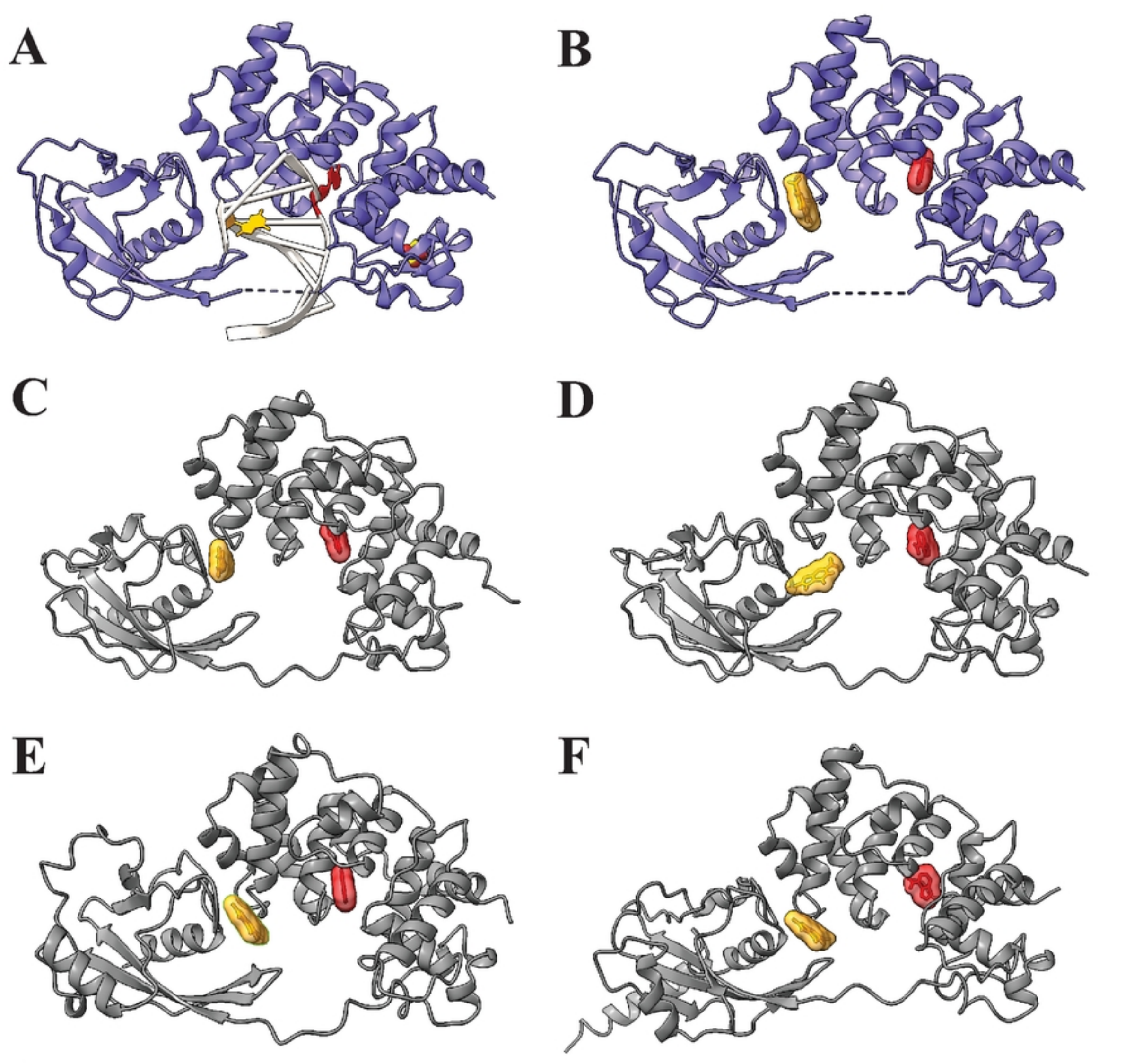
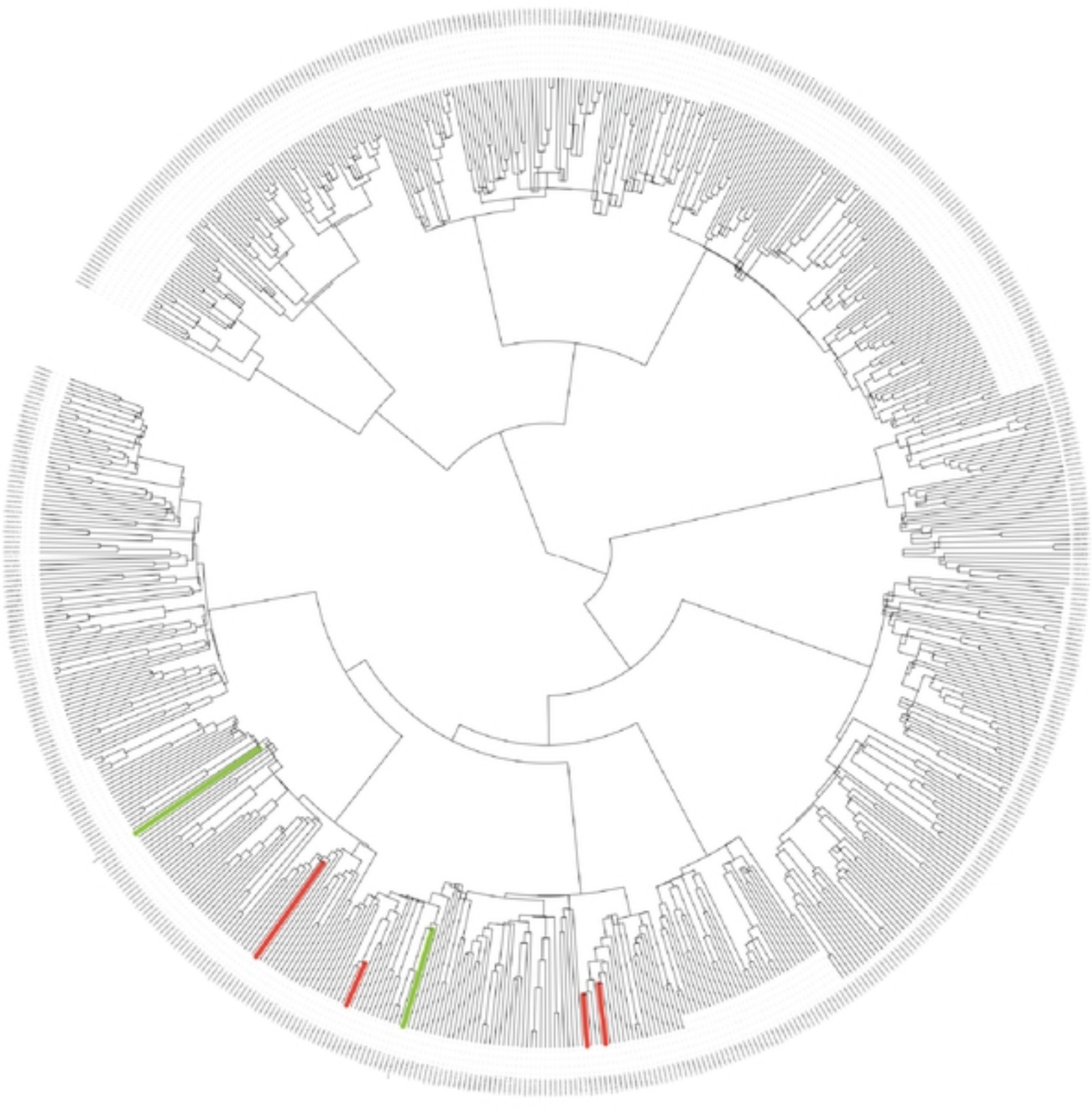


Figure 7

A



B

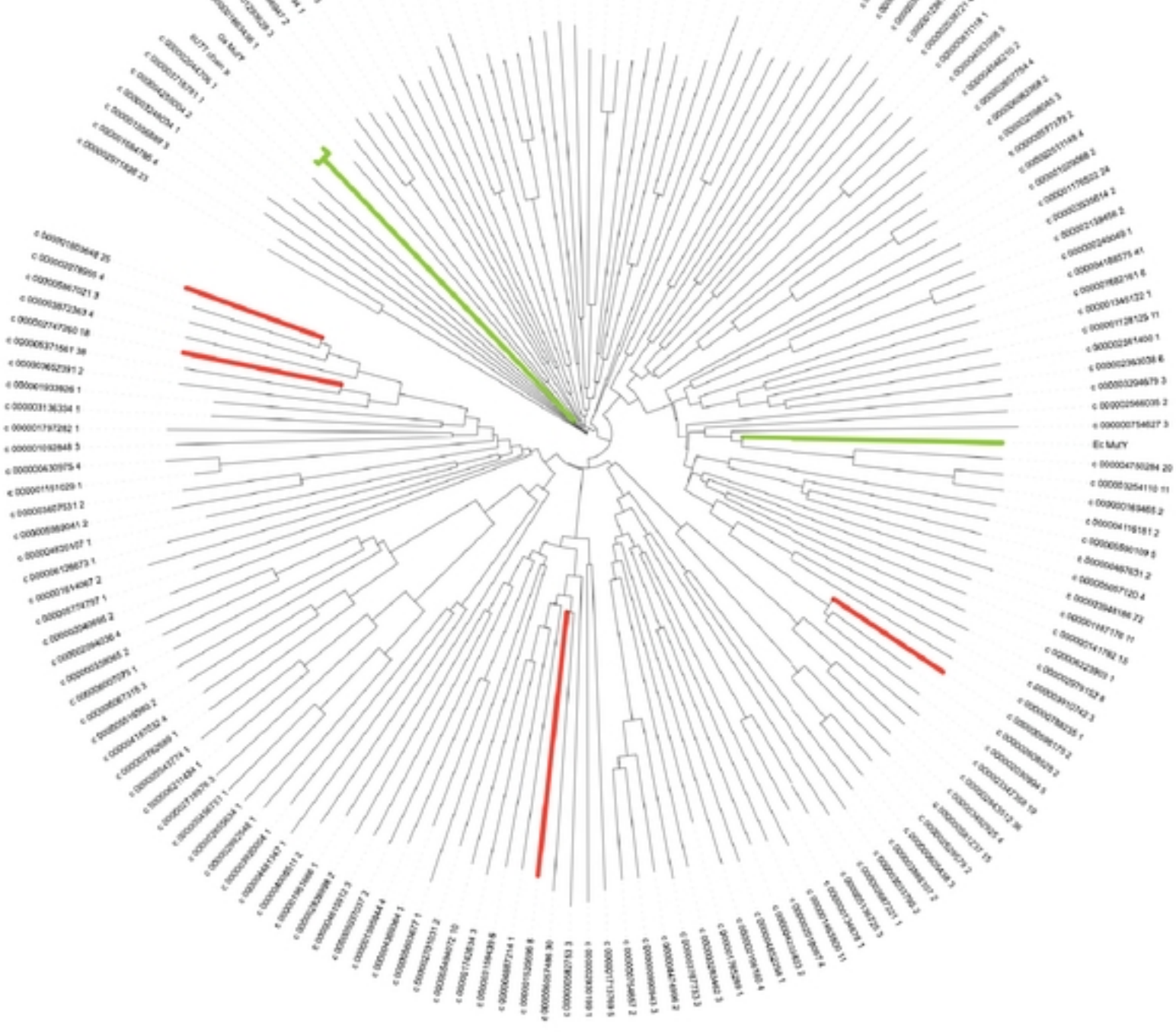


Figure 4

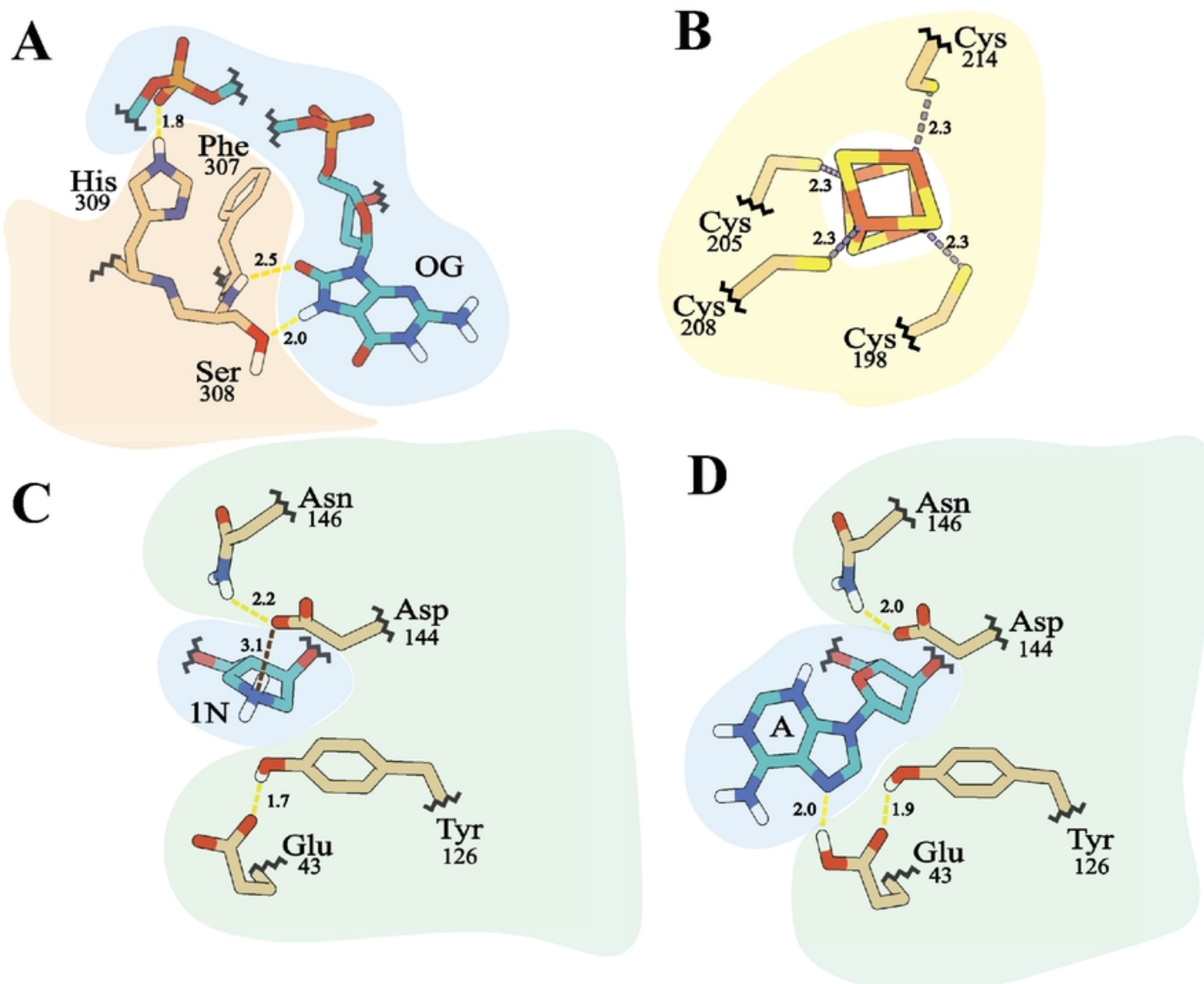

A
1.8
Phe
307
His
309
2.5
OG
2.0
Ser
308
B
Cys
214
2.3
2.3
Cys
205
2.3
2.3
Cys
208
Cys
198
C
Asn
146
2.2
Asp
144
3.1
1N
1.7
Glu
43
Tyr
126
D
Asn
146
2.0
Asp
144
A
2.0
1.9
Glu
43
Tyr
126

Figure 2

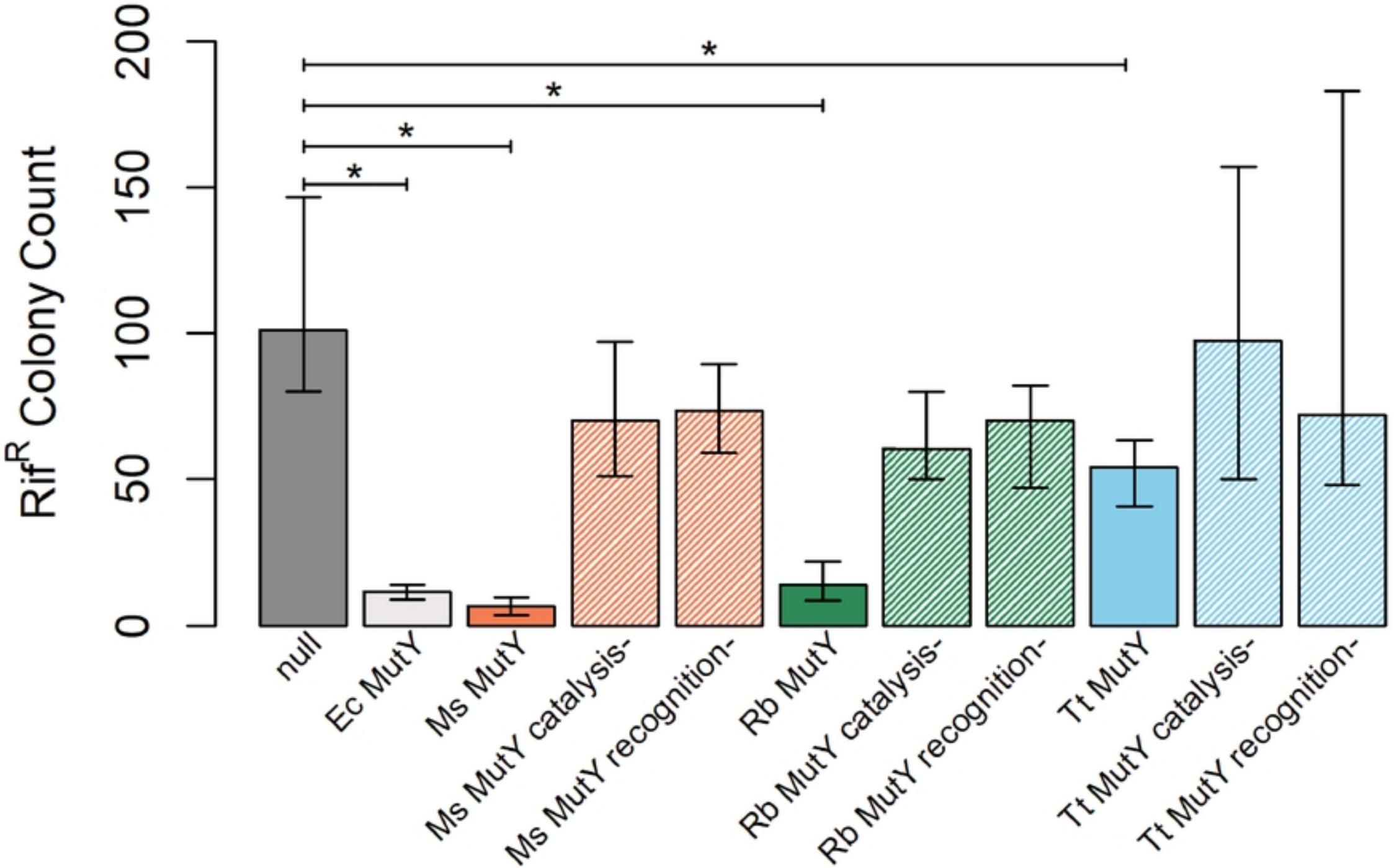

Rif^R Colony Count
0
50
100
150
200
*
*
*
*
null
Ec MutY
Ms MutY
Ms MutY catalysis-
Ms MutY recognition-
Rb MutY
Rb MutY catalysis-
Rb MutY recognition-
Tt MutY
Tt MutY catalysis-
Tt MutY recognition-

Figure 8

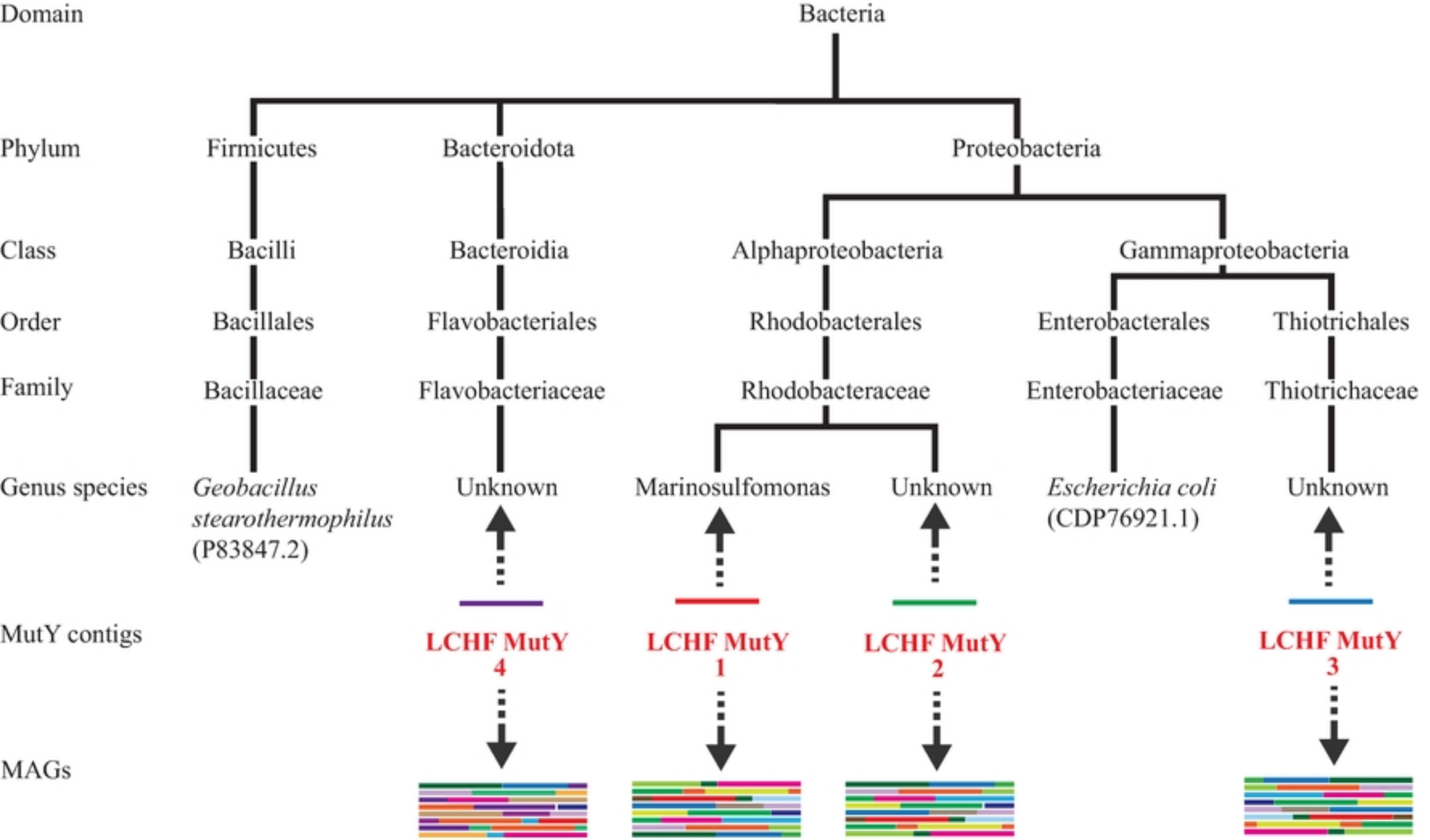

Domain
Bacteria
Phylum
Firmicutes
Bacteroidota
Proteobacteria
Class
Bacilli
Bacteroidia
Alphaproteobacteria
Gammaproteobacteria
Order
Bacillales
Flavobacteriales
Rhodobacterales
Enterobacterales
Thiotrichales
Family
Bacillaceae
Flavobacteriaceae
Rhodobacteraceae
Enterobacteriaceae
Thiotrichaceae
Genus species
Geobacillus stearothermophilus (P83847.2)
Unknown
Marinosulfomonas
Unknown
Escherichia coli (CDP76921.1)
Unknown
MutY contigs
LCHF MutY 4
LCHF MutY 1
LCHF MutY 2
LCHF MutY 3
MAGs

Figure 6

xthA (exonuclease III)
nfo (endonuclease IV)
mutT
mutM
mutY
Max T (°C) 58 21 24 52 41
pH 8.6 9.5 9.3 8.7 8.1
Chimney
Marker 3
Marker 2
Calypso
Sombreo
Camel Humps
Normalized Coverage (TPM)
1
10
100
1000

Figure 3

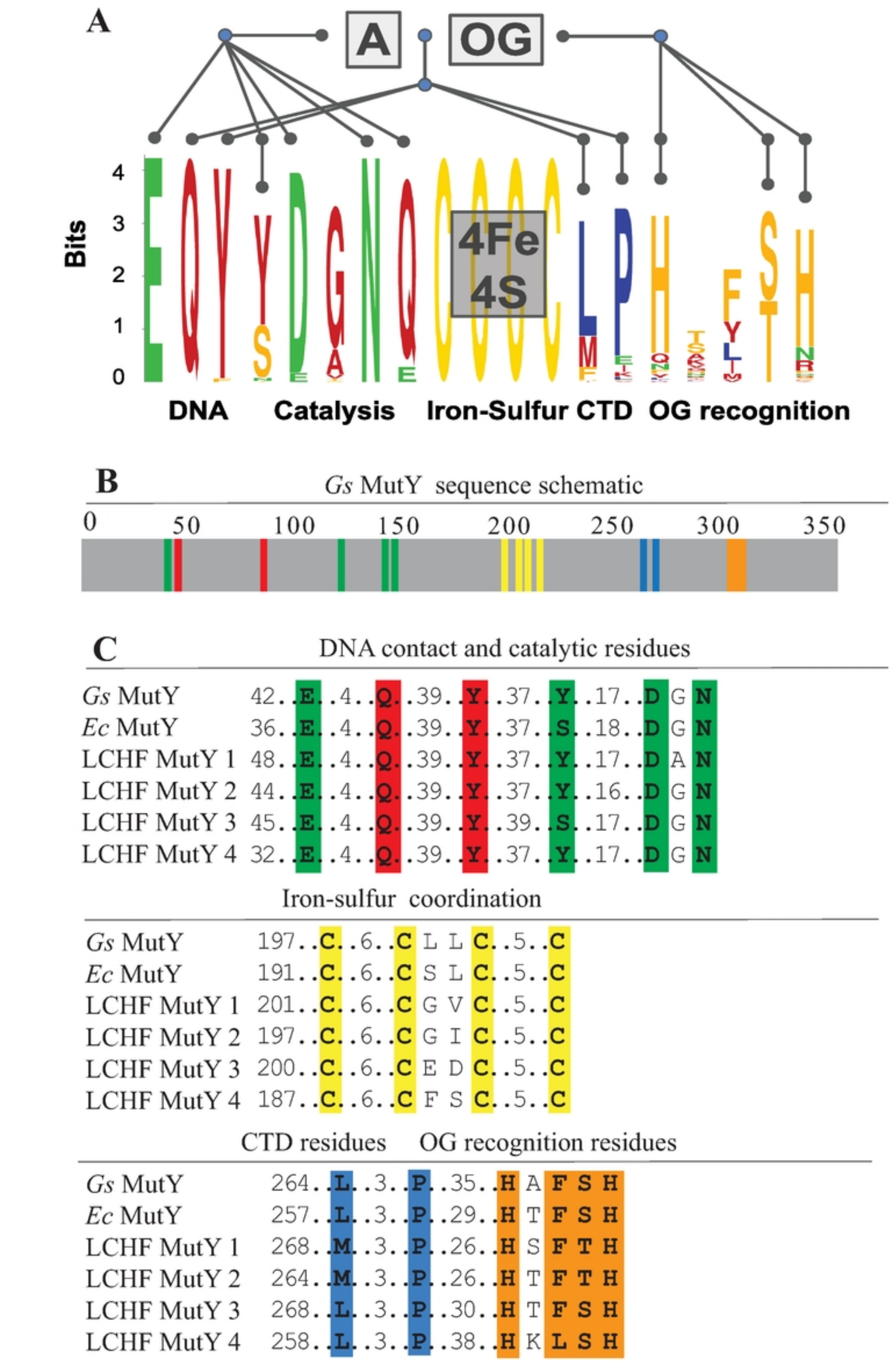

A
A
OG
Bits
4
3
2
1
0
4Fe
4S
DNA
Catalysis
Iron-Sulfur CTD
OG recognition
B
Gs MutY sequence schematic
0
50
100
150
200
250
300
350
C
DNA contact and catalytic residues
Gs MutY 42..E..4..Q..39..Y..37..Y..17..D G N
Ec MutY 36..E..4..Q..39..Y..37..S..18..D G N
LCHF MutY 1 48..E..4..Q..39..Y..37..Y..17..D A N
LCHF MutY 2 44..E..4..Q..39..Y..37..Y..16..D G N
LCHF MutY 3 45..E..4..Q..39..Y..39..S..17..D G N
LCHF MutY 4 32..E..4..Q..39..Y..37..Y..17..D G N
Iron-sulfur coordination
Gs MutY 197..C..6..C L L C..5..C
Ec MutY 191..C..6..C S L C..5..C
LCHF MutY 1 201..C..6..C G V C..5..C
LCHF MutY 2 197..C..6..C G I C..5..C
LCHF MutY 3 200..C..6..C E D C..5..C
LCHF MutY 4 187..C..6..C F S C..5..C
CTD residues
OG recognition residues
Gs MutY 264..L..3..P..35..H A F S H
Ec MutY 257..L..3..P..29..H T F S H
LCHF MutY 1 268..M..3..P..26..H S F T H
LCHF MutY 2 264..M..3..P..26..H T F T H
LCHF MutY 3 268..L..3..P..30..H T F S H
LCHF MutY 4 258..L..3..P..38..H K L S H

Figure 5