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

Abstract

The GO DNA repair system protects against GC → TC mutations by finding and removing oxidized guanine. The system is mechanistically well understood but its origins are unknown. We searched metagenomes and abundantly found the genes encoding GO DNA repair at the Lost City Hydrothermal Field (LCHF). We recombinantly expressed the final enzyme in the system to show MutY homologs function to suppress mutations. Microbes at the LCHF thrive without sunlight, fueled by the products of geochemical transformations of seafloor rocks, under conditions believed to resemble a young Earth. High levels of the reductant H2 and low levels of O2 in this environment raise the question, why are resident microbes equipped to repair damage caused by oxidative stress? MutY genes could be assigned to metagenome assembled genomes (MAGs), and thereby associate GO DNA repair with metabolic pathways that generate reactive oxygen, nitrogen and sulfur species. Our results indicate that cell-based life was under evolutionary pressure to cope with oxidized guanine well before O2 levels rose following the great oxidation event.

[*]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

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

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

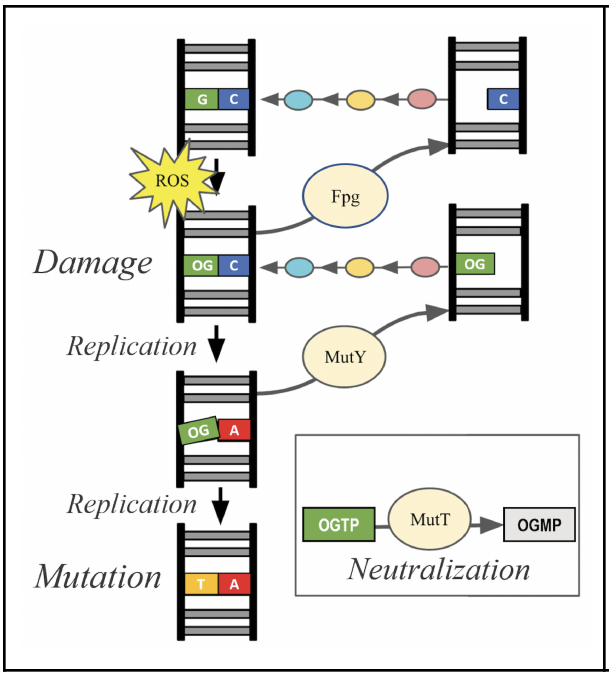


Figure 1. Overview of the GO System. The gene products of mutT, mutM, and mutY (tan bubbles) prevent or repair oxidized guanine DNA damage caused by ROS. DNA glycosylases Fpg and MutY remove OG from OG:C and A from OG:A, respectively to create AP sites with no base. Additional enzymes (AP nucleases,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

Biochemical analyses of the OG system gene products have provided a complete mechanistic picture for the GO repair system. MutT hydrolyzes the OG nucleotide triphosphate to sanitize the nucleotide pool, thus limiting incorporation of the promutagen into DNA by DNA polymerase (13,21). The enzyme encoded by mutM, called formamidopyrimidine-DNA glycosylase (Fpg), locates OG:C base pairs and excises the OG base to initiate base excision repair (BER) (12,22). MutY locates OG:A lesions and excises the A base to initiate BER (11,14). Fpg and MutY thus act separately on two different intermediates to prevent G:C → T:A mutations. These DNA glycosylases generate abasic (apurinic/apyrimidinic; AP) sites, which are themselves mutagenic if not processed by downstream, general BER enzymes, particularly AP nucleases (e.g. exonuclease III and endonuclease IV), DNA polymerase and DNA ligase (17,23–25) as shown in **Figure 1**.

MutY is the final safeguard of the GO system. If left uncorrected, replication of OG:A lesions results in permanent G:C → T:A transversion mutations as demonstrated by mutY loss of function mutants (26,27). Underperformance of the mammalian homolog, MUTYH, leads to

early onset cancer in humans, first discovered for a class of colon cancers now recognized as MUTYH Associated Polyposis (28). MutY is made up of two domains that both contribute to DNA binding and biochemical functions. The N-terminal catalytic domain shares structural homology with EndoIII and other members of the Helix-hairpin-Helix (HhH) protein superfamily (17). The C-terminal OG-recognition domain shares structural homology with MutT and other NUDIX hydrolase family members (17,29). Functionally important and highly conserved residues define chemical motifs in both domains (**Figure 2**). These chemical motifs interact with the OG:A lesion and chelate the iron-sulfur cluster cofactor as revealed by x-ray structural analysis (**Figure 2**) (30–33). For example, residues in the N-terminal domain establish the catalytic mechanism for adenine excision (**Figure 2A** and **Figure 2B**) (32,34). Residues found in a beta loop of the C-terminal domain recognize the OG base and thus direct adenine removal from OG:A lesions (**Figure 2C**) (33). Some motifs are shared among other DNA glycosylases, such as the residues that chelate the 4Fe4S iron-sulfur cluster cofactor (**Figure 2D**) (17). Chemical motifs particular to MutY, especially the OG-recognition residue Ser 308 (Figure 2C) and supporting residues in the C-terminal domain, are conserved across organisms and are not found in other DNA glycosylases and therefore can be used to identify MutY genes (17).

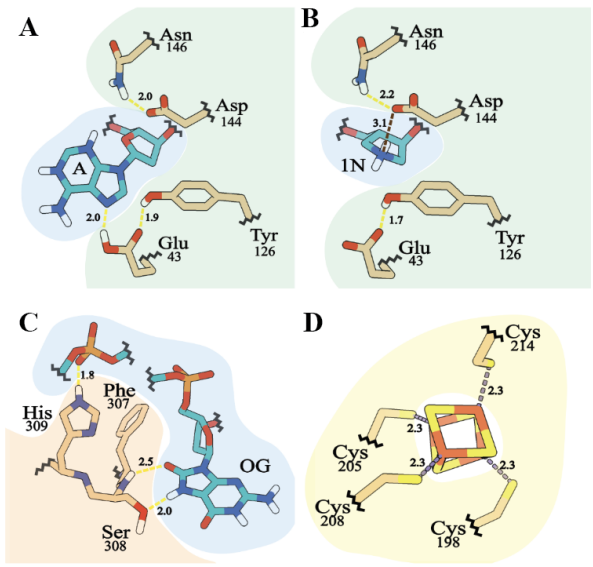


Figure 2. MutY chemical motifs. Panels A-C show interactions between MutY residues and DNA, with DNA highlighted in blue. Panel A was made from pdb 3g0q, a structure which describes interaction with the competitive inhibitor fluoro-A (31); panels B-D were made from pdb 6u7t, a structure which describes interaction with a transition state analog (1N) of the oxacarbenium ion formed during catalysis of adenine excision (33). MutY catalytic residues (highlighted in green) interact with an adenine base (panel A) and the 1N moiety which mimics the charge and shape of the transition state (panel B). The OG-recognition residues (highlighted in orange) provide hydrogen bonding interactions with the OG base (panel C). Four cysteine residues chelate the iron-sulfur metal cofactor (panel D). All of these interactions are important for MutY activity.

Our study investigated whether microbes in the anoxic LCHF environment use the GO DNA repair system to mitigate damage caused by reactive oxygen species. It is important to note that not all organisms have an intact GO repair system; examples are missing one, some or all components. MutY in particular was absent frequently in a survey of 699 bacterial genomes (19), and its absence may indicate relaxed evolutionary selection from oxidized guanine damage (35). We mined for homologous genes within the LCHF microbial community and recombinantly expressed candidate MutY enzymes to characterize function. We found genes encoding GO system components and general base excision repair enzymes at all LCHF sites. MutY homologs from the LCHF suppressed mutations when expressed in mutY deficient *Escherichia coli* (*E. coli)* strains indicating these function similarly to authentic MutY. These Lost City MutY homologs could be assigned confidently to metagenome assembled genomes (MAG)s, allowing for additional gene inventory analyses that revealed metabolic strategies involving sulfur

oxidation and nitrogen reduction. These results have important implications for understanding the repair of oxidative guanine damage in low-oxygen environments, similar to those that existed on a younger Earth, as well as those that may exist on other planets and moons.

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).

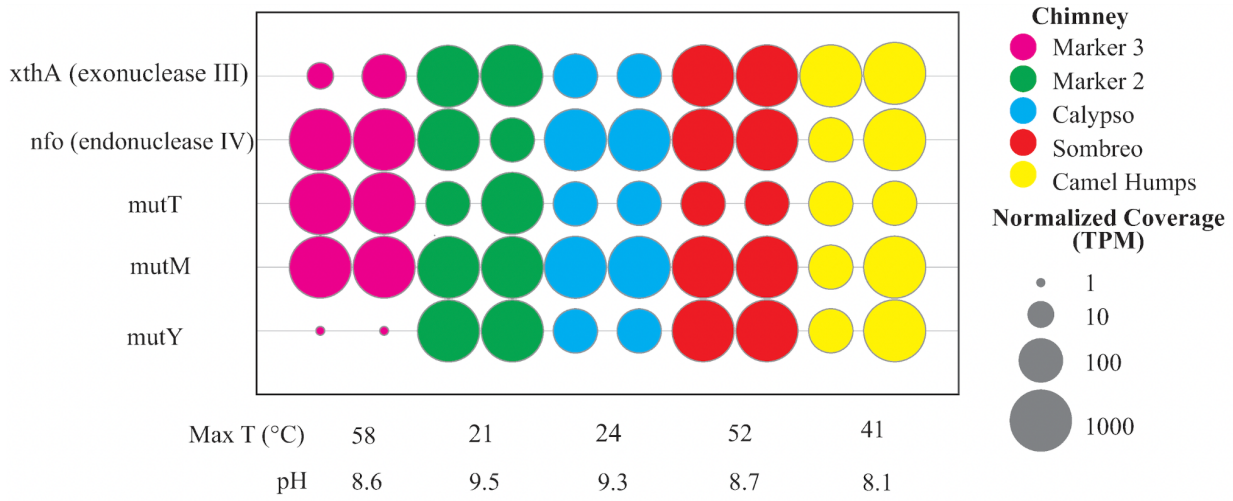


Figure 3. Abundance of GO system gene homologs. Listed on the vertical axis are genes encoding DNA repair enzymes. Genes xthA and nfo genes are generally necessary for DNA repair involving base excision repair in bacteria, including the particular GO system investigated here. Together, mutT, mutM and 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).

Metagenomic mining for MutY genes

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

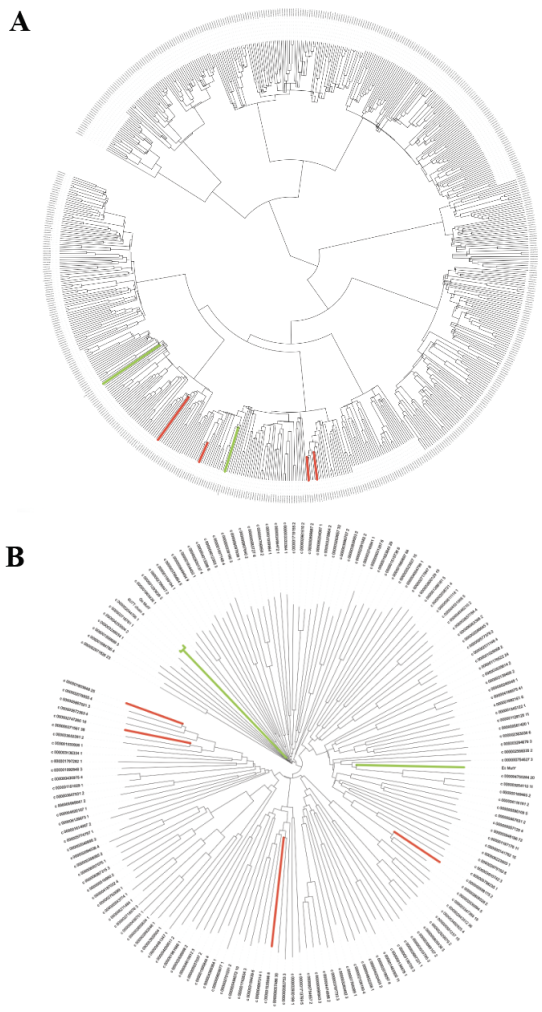


Figure 4. Phylogeny for LCHF MutYs. (A) 649 sequences were identified as LCHF MutY candidates due to sequence similarity to existing MutYs (green branches) and aligned to reconstruct evolutionary relationships. (B) A subset of 160 members contained all necessary MutY-defining chemical motifs. Alignment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

Figure 5 highlights conservation and diversity for the MutY-defining chemical motifs found in the 160 LCHF MutYs. All of the LCHF MutYs retain the chemical motif to coordinate the iron-sulfur cluster cofactor comprising 4 invariant Cys residues (4 Cs, yellow in **Figure 5**), a feature that is also found in other HhH family members such as EndoIII (16), but which is absent for some "clusterless" MutYs (38). Other invariant and highly conserved motifs make critical interaction with the DNA and provide key catalytic functions for adenine base excision,

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

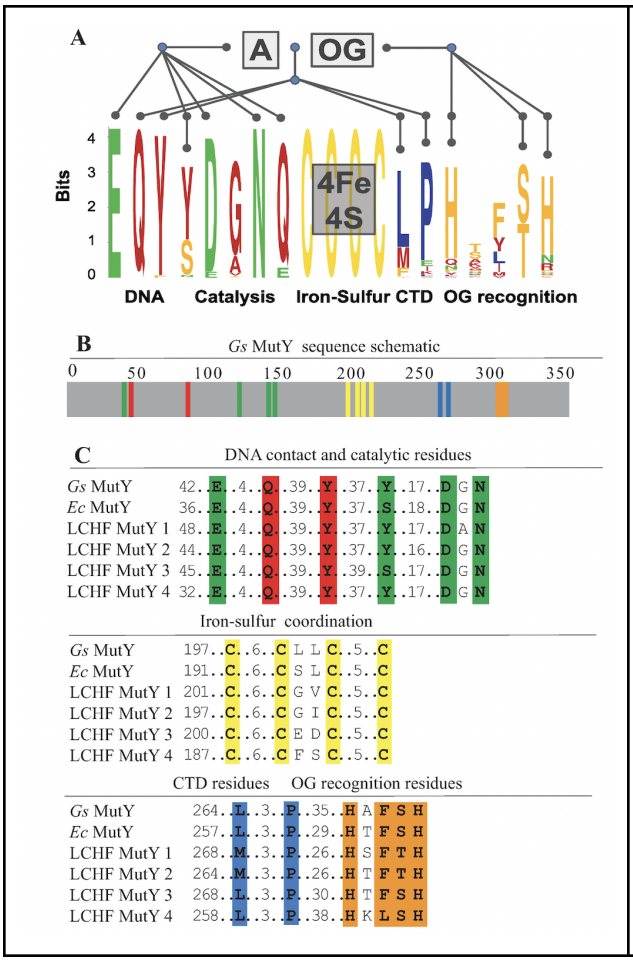


Figure 5. LCHF MutY chemical motifs. (A) Conservation and diversity of MutY-defining chemical motifs are depicted with a sequence logo for the 160 LCHF MutYs. These motifs are associated with biochemical functions including DNA binding, enzyme catalysis, attachment of the iron-sulfur cofactor, and recognition of the damaged OG base. Sequence logo generated by Weblogo (39,40). See **Supporting Information** for a complete alignment and a percent identity matrix for the representative LCHF MutYs. (B) Chemical motifs located in MutY as shown with color-coded positions for Gs MutY. (C) Alignment for select chemical motifs highlights conservation among the four representative LCHF MutYs and MutYs from *Ec* and *Gs*.

By contrast with these numerous and highly conserved motifs for the N-terminal domain, fewer motifs with greater sequence divergence are found in the C-terminal domain. The MutY ancestor is thought to have resulted from a gene fusion event that attached a MutT-like domain to the C-terminus of a general adenine glycosylase enzyme, and the C-terminal domain of modern *E. coli* MutY confers OG:A specificity (29). X-ray crystal structures of *Gs* MutY interacting with OG:A and with G:A highlighted conformational difference for a Ser residue in the C-terminal domain (Ser 308 in *Gs* MutY) (33) and mutational analysis showed this Ser residue and its close neighbors (Phe 307 and His 309 in *Gs* MutY) establish OG specificity (33). Informed by these insights from structure, we eliminated LCHF MutY candidates that lacked a C-terminal domain and its OG-recognition motif. Alignment of the 160 LCHF MutYs that passed this test showed

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

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

Four representative LCHF MutYs were examined for physical properties. **Table 1** reports these physical properties including predicted protein size, isoelectric point (pI), and stability (Tm). Generally the physical characteristics measured for LCHF MutY representatives were comparable to each other and to predicted properties of *Ec* MutY and *Gs* MutY. The predicted Tm for LCHF MutY 3 was above 65℃, distinguishing it as the most stable enzyme (**Table 1**),

which may reflect adaptation to a high temperature environment. The isoelectric point predicted for each of the LCHF MutY representatives is 3 pH units above the pI predicted for *Gs* MutY and between 0.1 - 0.5 pH units above the pI predicted for *Ec* MutY, indicating that more numerous positively charged residues have been recruited, possibly as an adaptation to alkaline pH conditions of the LCHF.

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)

Identification of LCHF MutY organisms, gene neighbors, environmental conditions, and metabolic strategies

Our next objective was to identify the organisms from which these LCHF MutY enzymes originated. Each of the four representative LCHF MutY sequences were derived from contiguous DNA sequences (contigs) belonging to a metagenome-assembled genome (MAG) representing a LCHF microbe. The taxonomic classification of these MAGs indicated that LCHF MutY 1 originated from a species of *Marinosulfonomonas,* LCHF MutY 2 from the family

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

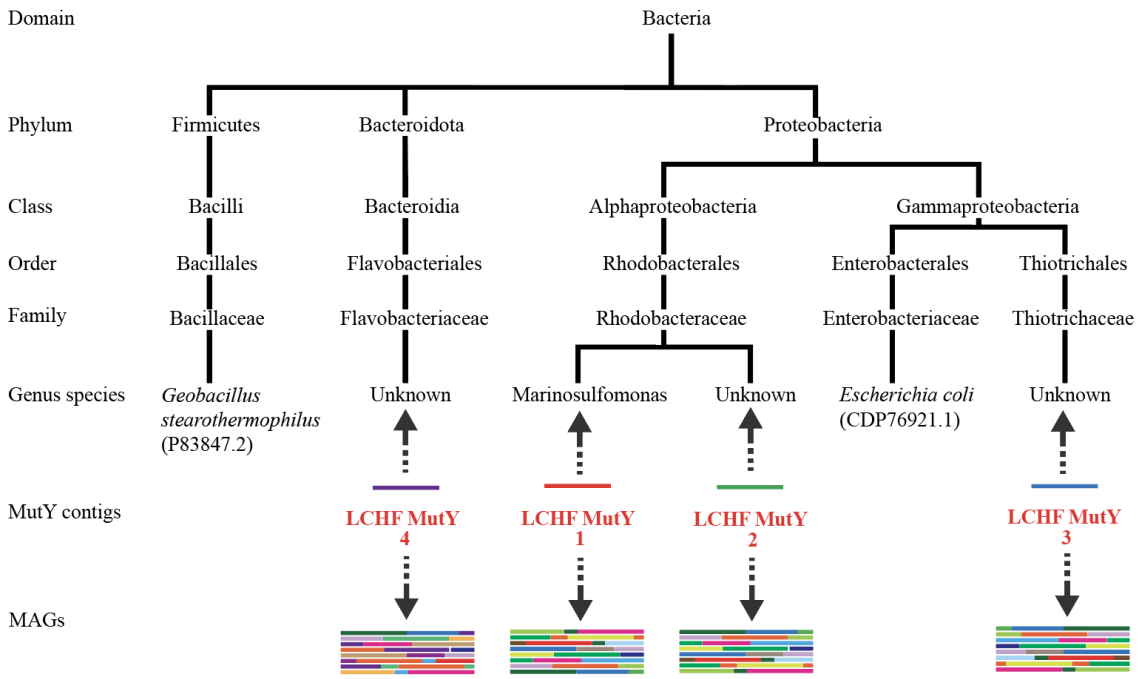


Figure 6. Taxonomic classification. LCHF MutY-encoding contigs were found in several branches of bacteria. The classification places these in relation to MutY of *G. stearothermophilus* and *E. coli* (accession IDs included). LCHF MutYs were mapped to their respective microbes by two methods indicated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

The inclusion of MutY contigs in MAGs provided an opportunity to examine gene neighbors for the representative LCHF MutYs. The GO repair genes are located at distant loci in *E. coli* (12), and belong to separate, complex operons (41). However, MutY is the immediate

5'-neighbor to YggX within gammaproteobacteria, and homologs of YggX are present outside this lineage, occasionally nearby to MutY (e.g. *Bacillus subtilis*). As gene neighbors, MutY and YggX are part of a SoxRS regulated operon in *E. coli* (41,42). YggX provides oxidative stress protection and iron transport function with a critical Cys residue close to the N-terminus of this small protein (43,44). A protein matching these features is encoded by a gene partly overlapping with and the nearest 3' neighbor to *Thiotrichaceae* MutY (see **Supporting Information**).

To reveal the environmental conditions of these MutY-encoding organisms, we analyzed the sequence coverage of each LCHF MutY contig at each of the sampling sites. *Marinosulfonomonas* MutY, *Thiotrichaceae* MutY, and *Flavobacteriaceae* MutY were identified at all sampling sites, ranging from 21℃ to 58℃ and pH 8.1 to 9.5. *Rhodobacteraceae* MutY was present at all sampling sites excluding Marker 3 and was therefore found in temperatures ranging from 24℃ to 52℃ and pH 8.1 to 9.5.

We further investigated the metabolic strategies utilized by MutY-encoding microbes by examining the inventory of predicted protein functions in each MAG (**Table 2,** see also **Supporting Information**). Each LCHF MutY-containing MAG possessed at least two forms of cytochrome oxidase, with the exception of the *Flavobacteriaceae* MAG. The *Flavobacteriaceae* MAG is only 44% complete, however, so no strong conclusions can be made regarding the absence of genes. Cytochrome oxidases commonly provide sources of free radicals and are essential to aerobic metabolism. Predicted proteins indicative of dissimilatory nitrate and nitrite reduction were found in the *Marinosulfonomonas, Rhodobacteraceae,* and *Thiotrichaceae* MAGs*,* suggesting that these organisms may be capable of using nitrate or nitrite as alternative electron acceptors when oxygen is not available. Furthermore, the *Marinosulfonomonas* and *Rhodobacteraceae* MAGs include predicted protein functions associated with the oxidation of

reduced sulfur compounds, though it is important to note that the directionality of these reactions cannot be fully determined from bioinformatics alone. These patterns speak to the potential origins of oxidants within the MutY-encoding organisms as discussed below.

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

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

Predicted protein structures and virtual docking experiments

To assess the likelihood of the LCHF MutY sequences folding into enzymes capable of activity on the OG:A substrate, protein structures were predicted using AlphaFold2 (45) (**Figure 7**). These predicted structures were associated with high confidence (see Supporting Information figures for pLDDT scores). Superpositions revealed that the predicted structures for the LCHF MutYs are each highly comparable with the experimentally determined structure for *Gs* MutY as indicated by visual inspection (**Figure 7**) and by low, pairwise root mean square deviation (RMSD) values (**Table 3**). The whole protein superpositions were dominated by the larger, more structurally conserved N-terminal domain. Breaking the analysis into two separate domains showed that the C-terminal domain, although more plastic, retained core structural features that could be superimposed. The MutY-defining chemical motifs are positioned in locations similar to those seen for the *Gs* MutY reference structure, providing evidence these LCHF MutY enzymes are capable of recognizing OG:A lesions and excising the adenine base. Concisely, the LCHF MutY structure predictions resemble a functional MutY enzyme from a thermophilic bacterium.

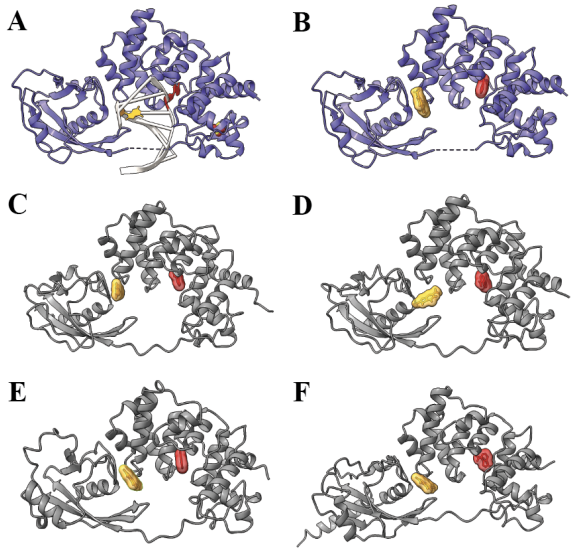


Figure 7. **Structure predictions and virtual docking of MutY ligands**. (A) The x-ray crystal structure of *Gs* MutY (blue ribbon; pdb id 3g0q) in complex with DNA (white helix) highlights the positions of the adenine nucleotide (red) and the OG nucleotide (yellow). (B-F) Virtual docking of ligands.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

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

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

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

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

Testing mutation suppression activity of LCHF MutY enzymes by recombinant expression

With *in silico* data suggesting that the LCHF MutY representatives encoded functional MutY enzymes capable of proper folding and interaction with OG:A lesions, we recombinantly expressed the genes in *E. coli* and measured mutation suppression activity *in vivo*. Three of the representative LCHF MutYs were successfully cloned into the pKK223 expression plasmid as verified by Sanger sequencing. The *Flavobacteriaceae* MutY appeared to be toxic to *E. coli* as only mutant versions of the gene were obtained from multiple cloning attempts, a situation that is reminiscent of *Gs* MutY, which is also apparently toxic to *E. coli* and could not be cloned into pKK223 (33).

To test the mutation suppression activity of *Marinosulfonomonas (Ms), Rhodobacteraceae (Rb), and Thiotrichaceae (Tt)* MutYs, we measured mutation rates with a rifampicin resistance assay (50). Several, independent, single-point mutations within the gene encoding RNA polymerase beta-subunit (*rpoB*) prevent the antibiotic rifampicin from binding to RpoB. Thus, spontaneous Rif^R mutants arising in overnight cultures can be counted by the colonies that emerge on rifampicin containing plates. Cultures expressing functional MutY delivered by plasmid DNA transformation have low Rif^R frequency compared to the high Rif^R

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

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

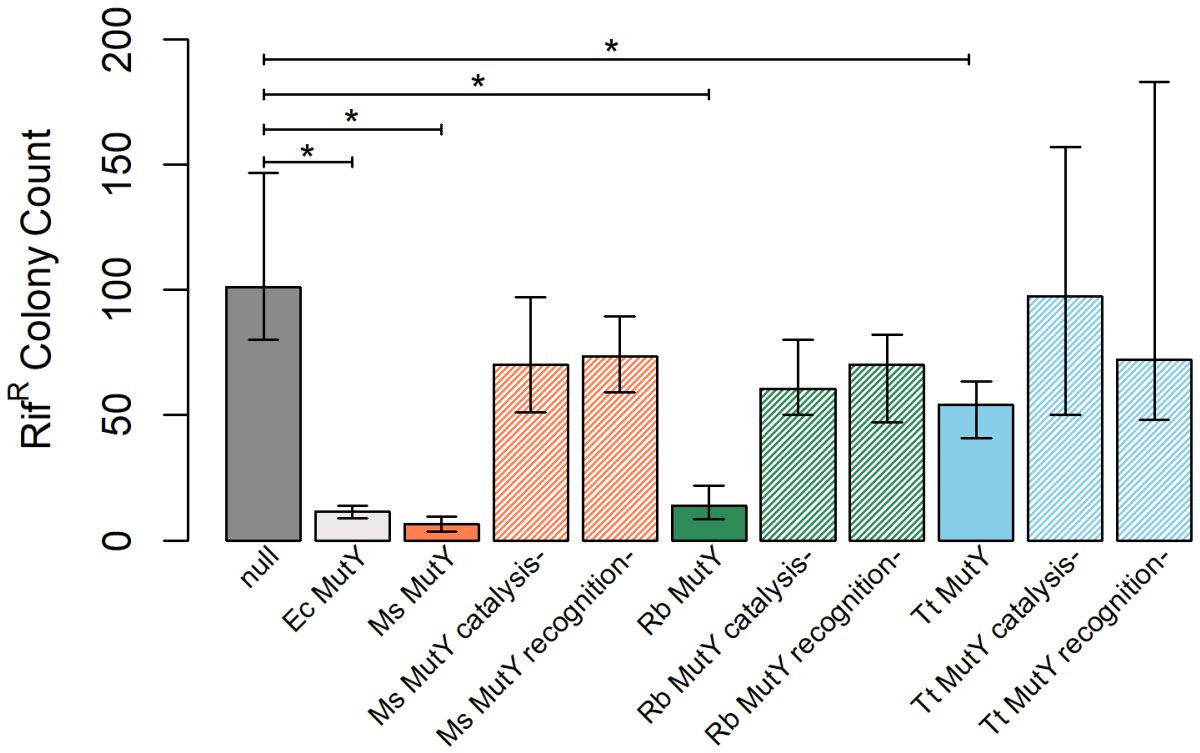


Figure 8. **Functional analysis.** Bars represent median Rif^R colony counts for *E. coli* cultures expressing MutY, MutY variants, or no MutY (*null*) from a plasmid DNA. Error bars represent 95% confidence intervals as determined by bootstrap sampling (see Supporting Information for tabulated values).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

To investigate the biochemical mechanism employed by LCHF MutY enzymes, we altered residues essential for OG:A recognition and catalysis, then repeated the mutation suppression assay (10, 12). Two mutants of each LCHF MutY were constructed through site-directed substitution of residues. One set of substitutions was designed to disable the OG-recognition motif by replacing F(S/T)H residues (**Figure 2 and Figure 5**) with alanine residues (designated *recognition*-); the other set of substitutions was designed to disable catalysis by replacing the active site Asp and Glu residues with structurally similar, but chemically inert

Asn and Gln residues (*catalysis*-). For all three LCHF MutYs, these targeted substitutions disabled mutation suppression function *in vivo* as shown by elevated Rif^R frequencies for cultures expressing the *recognition*- and *catalysis*- versions. The mutation frequencies for cultures expressing these site-specific substitution variants were comparable to the Rif^R frequencies measured for *null* cultures as judged by overlapping 95% confidence intervals (**Figure 8** and **Supporting Information**). These results indicate that the representative LCHF MutYs suppress mutations by a mechanism that is highly similar to the strategy executed by *Ec* MutY and *Gs* MutY.

Discussion

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

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

Retained function across evolutionary and species barriers strongly suggests that MutY interacts with the base excision repair apparatus through some well-preserved mechanism that relies on a universal language understood by all organisms. Most critically, the AP sites generated by MutY should be recognizable to downstream AP nucleases. Protein-protein interaction between AP nucleases and MutY has been discussed as a possible mechanism (53–56), but on its own such a mechanism would rely on coevolution of protein partners. Our results and those reported by others for complementation with the eukaryote homolog MUTYH (57) speaks for a mechanism that is less sensitive to sequence divergence. Therefore, we favor a

model where the distorted DNA structure created by MutY signals the location of the AP site for handoff to the BER apparatus, as has been suggested previously (55).

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

Our metagenomic analysis revealed that gene homologs encoding the GO DNA repair system are abundant in basement microbes inhabiting the LCHF (**Figure 3**). This observation is surprising given that the basement of the LCHF is expected to be anoxic (1,3). The chemical agents commonly thought of for producing oxidized guanine (OG) are ROS derived from molecular oxygen *via* aerobic metabolism. In an anoxic environment, what chemical agents are producing OG and how are these generated? Models of hydrothermal field chemistry predict abiotic production of ROS which the microbial residents may encounter, although these would probably react with cell protective structures before encountering DNA (58). Continual mixing of seawater with the anoxic hydrothermal fluid could provide molecular oxygen at the interface

where hydrothermal fluids vent into ambient seawater at the seafloor (1,3). Facultative anaerobes at this interface would inevitably generate ROS (61–66), and therefore benefit from the GO DNA repair system.

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

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

considered the caveat of possible phage-mediated gene transfer – modern microbes adapted to oxygen-rich regions elsewhere may be the source of LCHF MutYs. However, the correspondence of taxonomic assignments based on MutY sequence and based on MAGs and the high degree of sequence diversity seen for LCHF MutY enzymes are inconsistent with expansion of GO DNA repair in the LCHF by horizontal gene transfer. Thus, it seems more likely that LCHF microbes inherited the GO DNA repair system from a common ancestor and retained it through necessity, even in the absence of extrinsic O_2.

Conclusion

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

Identification of LCHF MutYs

Candidate MutY genes in the LCHF metagenomes were identified with a BLASTP (79) search against predicted protein sequences from the LCHF pooled metagenomic assembly using

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

Taxonomic classification

Contiguous DNA sequences containing the LCHF MutY representatives the contig DNA were assigned taxonomic classifications using the program MMseqs2 (83) and the Genome Taxonomy Database (GTDB) as described in Brazelton et al., 2022 (36). Taxonomic classification of each MAG that included a contig of interest was performed with GTDB-Tk v1.5.1 (84). The environmental distributions of MutY-encoding MAGs were inspected for potential signs of contamination from ambient seawater during sample collection. This

possibility was ruled out by the absence of all MutY-encoding taxa reported in this study in the background seawater samples collected by Brazelton et al (36). MAG completeness and contamination scores were generated by CheckM v1.0.5 as described in Brazelton et al. 2022 (36).

Prediction of physical parameters

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

Molecular modeling

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

Ligand docking experiments were executed with the program AutoDock Vina (47). Ligand structures representing adenine and OG were prepared with the ligand preparation tools implemented with Autodock Tools, and receptors structures were prepared from the structures predicted by AlphaFold or from the crystal structure of *Gs* MutY with the receptor preparation tools, also as implemented with Autodock Tools (87,88). Receptor and ligand structures were

treated as rigid objects. Separate 20 x 20 x 20 $Å^3$ search volumes were defined for adenine and for OG. The adenine search volume was centered on the position of residue A5L:18 in chain C of the *Gs* MutY crystal structure, and the OG search volume was centered on the position of residue 8OG:6 in chain B.

Recombinant DNA cloning

Synthetic genes encoding the LCHF MutYs were codon optimized for expression in *E. coli* except that pause sites with rare codons were engineered so as to retain pause sites found in the gene encoding *Ec* MutY. Gene Blocks were ordered from Integrated DNA Technologies (IDT). LCHF MutY genes designed in this way were cloned into the low-expression pKK223 vector by ligation-independent cloning (LIC) as follows. PCR reactions with the high-fidelity Phusion polymerase amplified the synthetic Gene Blocks and two overlapping fragments derived from approximate halves of the pKK223 plasmid. PCR products were separated by electrophoresis in 0.8% agarose x1 TAE gels containing 1 μg/mL ethidium bromide. DNA was visualized by long-wavelength UV shadowing to allow dissection of gel bands, and the DNA was purified with the GeneJet gel extraction system, treated with Dpn1 at 37 °C for 45 min, and heat shock transformed directly into DH5α competent cells. Clones were selected on ampicillin media plates. The plasmid DNA was purified from 4-mL overnight cultures by use of the Wizard Plus MiniPrep kit according to the manufacturer's instructions. The sequence of the LCHF MutY encoding gene was verified by Sanger sequencing with UpTac and TacTerm primers. Genes encoding site-directed substitution variants were created by amplifying two overlapping fragments of the LCHF MutY-pKK223 plasmid with mutagenic PCR primers followed by similar gel purification and transformation procedures. In our hands the LIC cloning efficiency was close to 95% except for the *Flavobacteriaceae* MutY gene which could not be cloned intact.

Mutation suppression assay

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

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

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

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

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

```
Conservation:                  9 5 579 67999 5  5   65  7         55      6   9 98989 9   5      5        8        65969 67  66 5           859
sp_P17802_MUTY_ECOL     28  KTPYKVWLSEVMLQQTQVATVIP-YFERFMARFPTVTDLANAPLDEVLHLWTGLGYYARARNLHKAAQQVATLHGGKFPETFEEVAALPGVGRETAGAILSLSLGKHFPILDGN  140
sp_P83847_MUTY_GEOS     34  RDPYKVWVSEVMLQQTRVETVIP-YFEQFIDRFPTLEALADADEDEVLKAWEGLGYYSRVRNLHAAVKEVKTRYGGKVPDDPDEFSRLKGVGPYTVGAVLSLAYGVPEPAVDGN  146
c_000003652391_2        37  PDPYTVWLSEVMLQQTTVAAVTG-YFRRFTTRWPTVQALAAAADDGDVMGAWAGLGYYARARNLLKCARVVCETHAGAFPNTYAGLIALPGIGPYTAAAIAAIAFDRAEVVVDGN  149
c_000000031207_6        26  EDPYAIWLSEVMLQQTRVTTVIP-YWHRFLEKYPDPKSLARAPLEQVLELWAGLGYYSRGRNLHAAACQIVNDHDGIFPRDPELIGKLPGIGEYTTAAICSIAYGEPLAVIDGN  138
c_000000430975_4        38  PNVYHVWLSEIMLQQTKVATVLS-YYLNFLHKWPTLNDLATATRQNVLQAWAGLGYYARARNLHACAQIVMEKYGGTFPKTEKELLRLPGIGHYTAAAIVAIAHGQPSVVVDGN  150
c_000001007148_4        35  RDPYVVWVREIMLQQTTVTAVVP-YLDRFLKRFPTVESLAGAKSAEVLRVWEGLGYYSRARNLSRGAKLIVEEWGGQWRNTATELQSLPGVGPYTAGAIASFAFDQSAGIVEAN  147
c_000001092848_3        48  PDPYRVWLSEIMLQQTTVGAVKP-YCERFLERWPTVEALARADLQEVMKAWAGLGYYSRARNLKKCAEEVAHRHGGVFPDTEAGLKALPGIGDYTAAVAAIAFGRRAAVVDGN  160
c_000001151029_1         1  ------WLSEIMLQQTKVATVLS-YYSNFLHKWPTLADLAMATRHNVLQAWAGLGYYARARNLHACSQIVMEKYGGTFPKTEKELLKLPGIGHYTAAAIVAIAHGQPSVVIDGN  107
c_000001797282_1        58  PDPYRVWLSEVMLQQTTIAAVRP-YFERFLTRFPDIFALAEAPEEAVMSAWAGLGYYSRARNLHACARTVAA-AGGRFPDTAEGLRKLPGIGAYTAGAIAAIAFDRQEAAVDGN  169
c_000001803648_25       40  PDPYRIWMSEVMLQQTTVAAVAK-YFHAFTTRWPTIIALASANDDDVMGQWAGLGYYARARNLLKCARVVVADYGGNFPDTRDELLKLPGIGPYTAAAIASIAFGRSETVMDAN  152
c_000002078955_4        40  AEPYRVWMSEVMLQQTTVAAVAK-YFHAFTKRWPTVDALAAAKDQDVMGQWAGLGYYARARNLLKCARMVVSDYNSEFPETRDELLKLPGIGPYTAAAIASIAFGRPECVMDGN  152
c_000002106160_4        28  KEPYKIWLSETMLQQTQVKTVIP-FYNSWIEKYPDFESVAGARLDSLLKSWEGLGYYNRCHNFHKAVKIIVNKYNSSLPINIKDFMVLPGVGSYTSAAVYSIVFLHPIPVVDGN  140
c_000003607531_2        38  KVPYKIWLSEIMLQQTTVTTVGP-YFQHFLKTWPTIDELANASLDDVLKVWQGLGYYSRARNLHKCAKLICNNFDGRFPNTEEQLEQLPGIGPYTAAAILAIAFDKPAIVVDGN  150
c_000003872363_4        36  PDPYHIWLSEVMLQQTTVAAVVK-YFNTFTKRWPNIQALATAQDADVMAEWAGLGYYARARNLLKCARMIVNDHGGVFPNDRAALLALPGIGPYTAAAISAIAFGRSACVVDGN  148
c_000004820107_1        43  PNPYYVWLSEIMLQQTTVVTVGP-YFTKFVERWPAIHDLAKADRDDVMHEWAGLGYYARARNLHKCAQIVSKELNGIFPEEQKDLEKLPGIGGYTSAAIRAIAFSKPANVVDGN  155
c_000005774797_1        29  PNPYHVLLSEFMLQQTTVATVKT-YFEQFTKKWPSVEHFSHATDDEIMAQWAGLGYYARARNLVKTIHTIH--NQGHFPASAYELQKLPGIGPYTSAAIAAIAFNEPILPRDGN  139
c_000005867021_3        40  PDPYRIWLSEVMLQQTTVVAVAK-YFHVFTKRWPNITALANAKDEDVMGEWAGLGYYARARNLLKCARMIVADFGGVFPDTRDDLLKLPGIGPYTAAAIASIAFGRSECVMDGN  152
c_000004546210_2        31  PSKYKTWISEIMLQQTQVNTVEP-YFHRFMERFPTVEELASSSPDEVLQIWSGLGYYARARNIHRSAQLIIKQHKGELPSDYESLLHLPGIGEETAGAILSLSGIEPKPILDGN  143
c_000004551008_5        31  PSKYKTWISEIMLQQTQVATVEP-YFIQFIQRFPEVTELASSSLDEVLQLWSGLGYYARARNIHRSARIIVNQHHSELPSDYESLLTLPGIGKETAGAILSLSGIEPKPILDAN  143
c_000002511148_4        26  KTPYRVWVSEIMLQQTQVSTAIP-YFEKFIKEIPDVKSLSEASLDLVLSLWSGLGYYARARNLHSAKIINLEHGGKIPSTRDELMALPGIGRETAGAILALGFNKKAPILDGN  138
c_000000577378_2        14  KTPYRVWVSEIMLQQTQVSTVIP-YFKKFVREIPNIKSLSEASLDLVLSLWSGLGYYARARNLHKSAKIINLDHGGKIPSTGGELMALPGIGRETAGAILALGFNMKAPILDGN  126
c_000000598175_2        37  INPYRVWVSEIMLQQTQVKTVIP-YFEKFMATFPDVQTLAHADQDLVLKHWSGLGYYARARNMHKAAKMICGDFKGQFPDNLEAMQSLSGIGRETAAAILSIASNQPQAILDGN  149
c_000000754627_3        34  KTPYKVWVSEVMLQQTQVVTVIP-YFEKFMQSFPDIIALANADEDQVLHHWTGLGYYARARNLHKAAKIVRDKYQGQFPSTLEEVMDLPGIGRETAGAVLSLSLGQHHPILDGN  146
c_000000811118_1        25  KDPYKVWISEIMLQQTVTTVIP-YFNKFISEYPTIKTLASTSLDEVLSLWSGLGYYSRARNIHKTAKILKKDFDCKLPNEIEALMSLPGIGRETAGAILSLGFEQSGVILDGN  137
c_000000899687_2        29  NDPYSIWVSEIMLQQTQVKTVIP-YYERWIKTLPTIDKLANAPEQKILKLWEGLGYYSRAKNLKKSAKIICKEMNGKLPNTVKNLQNLPGIGRYTAGAISSIAFGLKAPVLDGN  141
c_000001176522_24       50  RTPYRVWVAEVMLQQTQVKTVVP-YFERFMEAFPAVDVLADAEMDAVLSQWAGLGYYARGRNLHKAARQIVAQ--GGFPETLAGWQALPGVGRETAAAIMAQAYGQRAAILDGN  160
c_000001345122_1         4  ISPYRVWVSEIMLQQTQVKTVIP-YYQRFMKSFPSIEQLASATQDEVLAHWAGLGYYARGRNLHKSAQIIQTEYYGKFPQIFDDIIALPGIGKETAGAIMSIALKQRMPILDGN  116
c_000001515736_6        18  RDPYRVWLSEIMLQQTQVSTVLG-YYQRFLDAFPDVASLAGAPQDAVLALWSGLGYYSRARNLHKCAQTVMQQWGGAFPQTAEELATLPGIGRETAGAISSFCFSERVPILDAN  130
c_000001682161_6        25  PTPYRTWVSEIMLQQTQVKTVIP-YYLRFMESFPTIQSLAKASQEQVLAHWAGLGYYARGRNLHKTAQTIVEQHKGQFPNRFEEVVALTGIGRETAGAILSIAMQQRMPILDGN  137
c_000001923643_29       28  RDPYRVWLSEIMLQQTQVSTVLG-YYQRFLDAFPDVASLAAAPQDAVLALWSGLGYYSRARNLHRCAQAVVQEWGGAFPRRAEDLATLPGIGRETAGAISAFCFSERVPILDAN  140
c_000002030994_5        26  INPYRVWVSEIMLQQTQVKTVIP-YYEKFMTSFPDVQTLASADQDTVLKHWSGLGYYARARNMHKAAQMVCDDFAGVFPDNLESMQSLSGIGRETAAAILSIASNQQEAILDGN  138
c_000002038721_4        28  KSPYHIWISEIMLQQTQVKTVIP-YFHKFINSFPSIKSLAKSDLDKVLANWSGLGYYTRAKNIHQASKIIHQKYHGSIPKNYEQLIELPGIGESTAGALLTLAFDKPGIILDGN  140
c_000002523527_15       43  RDPYRIWLSEIMLQQTQVAAVIE-YFQRFVAALPTVAALAAASADEVMALWAGLGYYSRARNLHRCAKTVVDEHAGVFPTDPEVLVTLPGIGRETAAAIAAFSAGVRSPILDGN  155
c_000002608528_2        26  INPYRVWVSEIMLQQTQVKTVIP-YYQKFMTTFPDVQTLANADQDLVLKHWSGLGYYARARNMHKAAKMVCDDFSGILPDNLEDMQSLSGIGRETAAAILSIASNQSQAILDGN  138
c_000002961510_2        29  KDPYSIWVSEIMLQQTQVKTVIP-YYERWIKTLPTIGKLANAPEQKILKLWEGLGYYSRAKNLKKSAQIICKEMNGELPKTVNDLKNLPGVGRYTAGAISSIAFGLKAPVLDGN  141
c_000002979152_8        34  RTSYRVWVSEIMLQQTQVVTVIP-YYERFMQSFPAVEGLAKATQNEVLDHWTGLGYYSRARNLHKAAQFVVEQCNGVLPESLDGLIELPGIGRETAAAILSLAGNKSEAILDGN  146
c_000003248034_1        46  KNPYHIWLSEVMLQQTRVTTVVP-YYERFLDRFPTLESLATASEQSVLQAWSGLGYYRRAKYLRASAQQIINKWNGQLPKEYDKLRALPGFGAYTAAAVSSIAFGRKHAAVDGN  158
c_000003294679_3        31  KTPYSIWVSEIMLQQTQVKTVIP-YFKKFIVSFPTLRVLAQASLDDVLHHWSGLGYYARAHNLHKTACLVNEKFSAQLPLDMKELCLLPGIGRETAGAILALSCNQKHPILDGN  143
c_000003333364_1        28  YDPYEVWISEVILQQTQMERGVA-YIERFLERFPTMESLAAAEEDEVLKVWEGLGYYARARNLRKGAQHMVEEHGGTFPADVEAARALPGVGRYTAGAVLSQAYNLEEPIVDAN  140
c_000003402697_5        28  REPYYVWVSEIMSQQTQIDRVADQFFPRFIETFPTVEALAAAADWETVYPCWDGLGYYNRGKNLLKAAKVLVDKHEGELPQTVAGLEALPGVGKYTAHAVLAFAFDKKVPAIDTN  141
c_000003492925_4        25  DDPYHIYLSEVMLQQTQVKTVLERYYFPFLKKFPTLKALGDANLDDVLKMWEGLGYYSRAKNLHKTAGLIEV-----LPSDIDELIRLPGIGKNTAHAVATFVYKKPVPIMEAN  133
c_000003766664_5        41  PTPYKVWVSEIMLQQTRVETVIP-YFQRFLSSFPDVHALAAASLDEVLSRWSGLGYYSRARNLHKASQFVVESWDGIFPSDAKMLEEMPGVGRYTAGAIASIAFGMPAPIVDGN  153
c_000003948186_72       28  ITPYRVWLSEIMLQQTQVATVIS-YFERFVQTFPDVGALARAPLDEVLHLWTGLGYYARARNLHKAAGLIVDTHGSELPDSVQQLCELPGIGRETAGAICAIALRKHAVIMDGN  140
c_000005057120_4        28  ITPYRIWLSEVMLQQTQVATVIP-YFNQFIDYFPDIHKLATAEDDAVLHLWSGLGYYARARNLHKTAKILSKN-KGIFPDDLDSLMELPGIGRETAGAILSIAFNNSYPILDGN  139
c_000005590109_5        28  ISPYRVWLSEVMLQQTQVTTVIP-YFNHFIQKFPDVHQLANAPLDSVLHLWSGLGYYARARNLHKTAKIISET-DGEFPNDLNLLMQLPGIGRETAGAISSIAFNKSQPILDGN  139
c_000006063368_2.       26  KSPYKVWISEIMLQQTQVTTVIP-YFNKFLLRFPNIEKLAESELDEVMSYWSGLGYYSRARNLYKAALEITCKHKSKLPNSYNELVALPGIGKTTAGAILALGFQKKAAILDGN  138
c_000000467631_2         7  STPYQTWLSEIMLQQTQVTTVIP-YFNRFIQTFPNIATLADAPLDAVLQHWAGLGYYARARNLHKTA-IIINKNGGVFPDDVDGLLELPGIGRETAGAILSIAFNKSHPIILDGN  118
c_000000788235_1        27  ITPYRVWISEIMLQQTQVATVIP-YYQRFITSFPDVTTLANADEDLVLQHWAGLGYYSRARNLHKAAQQVRDEFSATVPADLDSLQSLVGIGRETAGAILSLAYEQRHAILDGN  139
c_000001187176_11       28  MTPYRVWVSEIMLQQTQVATVIP-YYERFMAALPDVHALAAAPADEVLHLWTGLGYYSRARNLHRAAKRVVEECAGQFPGDVQGLTDLPGIGPETAGAIASLSMGVRAPILDGN  140
c_000001660697_64       28  GDAYRTWLSEIMLQQTQVSAVLG-YYARFIERFPTVQALAAAPADDVMAAWAGLGYYTRARNLHRCAQIVVAEHGGIFPRDPEVLASLPGIGRETAAAIAAFSYGVRAAILDGN  140
c_000003378864_2        39  RDPYAIWLSEIMLQQTQVETVLP-YYDRFLTAFPTIADLAAAPLGRVLKMWEGLGYYARCRNLHRAAQKIVSDHNGRFPATFEDVLALPGIGQETAGAILTFAYDQRHPLLDGN  151
c_000003910742_3        28  PTPYRVWISEIMLQQTQVVTVIP-YYQNFMDSFPDVRTLAHAEEDLVLQHWAGLGYYSRARNLHKSAKLICDEFSGVIPDDIDQISLPGIGRETAGAILSLAYQQQQAILDGN  140
c_000004116181_2         1  -----VWVSEIMLQQTQVATVVP-YFNRFMAQFPDVKTLAQTSLDAVLQHWAGLGYYARARNLHKTAKIIAQQ--GCFPNTLETLMQLPGIGQETAGAIMSIAFQQSCPILDGN  106
c_000006223903_1        29  RNPYRIWISEIMLQQTQVTTVIP-YYERFMARFPHIVTLASAKLDTVLAEWTGLGYYARARHLHQAAGVIVARHGGVFPDQFHDVVALPGIGRETAGAILAFSHGQRHAILDGN  141
c_000001029068_2         2  -NVYYVWLSEVVLQQTQVASAVG-YFSNFIKRFPTIGALSAASEDDVLSAWSGLGYYARARNLHKTAQIINRQYSGVFPSTVEQVLALPGIGRETAGAILSIAYNQACPILDGN  113
c_000001128125_11       36  ISPYRVWVSEIMLQQTQVATVIP-YFKRFMLSFPNVEVLAKATQDEVLSHWAGLGYYARGRNLHKAARQIVEQHDGIFPSQFDDILALPGIGRETAGAILSISLNQRFAILDGN  148
c_000001595844_4        27  SDAYAIWLSEIILQQTRVAQGMP-YYIRFLEAFPTVFDLAKADEQEVMRLWQGLGYYSRARNLHFTAAQQVANEFGGKFPDNYADLLALKGIGPYTAAAIASFAFGLPHAVVDGN  139
c_000001719155_2        31  RDPYRILVSEFMLQQTQVKTVIP-YFERWIKSFPTLQKLAEARESTVLKHWEGLGYYSRARNLRKAAQKIQQEYSGKVPESMNEIMKLPGVGRYTAGAVLSIAFGQKVPVLDGN  143
c_000002363038_6        36  INPYRVWVSEIMLQQTQVVTVIP-YFERFMQRFPEVSDLAAASQDEVLSHWAGLGYYARGRNLHKAAQLIVEQHNGVFPKSFDEVLNLPGIGRETAGAILSISLQQRFPILDGN  148
c_000002561400_1         1  ----------EIMLQQTQVATVIP-YFERFMQRFPKVADLACADQEEVLTYWAGLGYYARGRNLHKAAQLIVEKHGGQFPSDFEEVLALPGIGRETAGAILSISLEQRFPILDGN  104
c_000002566035_2        28  VSPYKVWVSEIMLQQTQVTTVIP-YFKRFMKRFPDVQTLANSSLDEVLHYWTGLGYYARARNLYKTAKIITKEYQGNFPLSLNEIMSLPGIGRETAGAVLSLSYDKCFPILDGN  140
c_000002598045_3        26  ITPYKVWVSEIMLQQTQVVTAIP-YFKKFIQEFPTVKSLSKATPDQVLKLWSGLGYYARARNLHKAAKLICSNHEGKIPSSINELVELPGIGKETAGAILSLGFKKKAPILDGN  138
c_000004188575_41       28  PTPYQVWVSEIMLQQTQVATVIP-YYQRFMIRFPDNATLAAASQDDVLSYWTGLGYYARARNLHRCAQYVAEIHQGQLPAELDALIALPGIGRETAGAIMALAYHQRFSILDGN  140
c_000000240049_1        26  KSLYRVWVSEIMLQQTQVTTVIP-YFARFMLQFNDIEALSNATQDDVLLYWAGLGYYSRARNLHKAAQIIQQQHGGCFPRQYNDVLALPGIGPETAGAILAQSQGQRHAILDGN  138
c_000002139456_2        26  KSLYRVWVSEVMLQQTQVATVIP-YFARFMQQFSDIEALSNATQDEVLLYWAGLGYYSRARNLHKAAQIIQQQHGGCFPQQYDDVLALPGIGPETAGAILAQALGQRHAILDGN  138
c_000003535614_2        28  RSLYRVWVSEIMLQQTQVATVIP-YYQRFMQRFKNIAALAEATQDDVLLHWAGLGYYSRARNLHKAAQIITASHQGKFPENYHDVLTLPGIGPETAAAILAQATGQRHAILDGN  140
c_000002657784_4        25  KTPYKVWISEIMLQQTQVVTVIP-FYSAFINRFPNIQELATSSEEEVMSFWSGLGYYSRARNIYKTAQILQKKFESKLPSSLSQLISLPGIGPETAGAILSLGYKKRAPILDAN  137
c_000003800129_15       30  TTPYRVWISEMMLQQTQVNTAKD-YFLNFIEQYPDLESIKNASEEEILILWKGLGYYRRAKFIYQAKEKIFKDYNGLFPDEFEDIISLPGIGKETAGAILSIAYGKPFPILDAN  142
c_000003716781_1        17  DDAYSVWISESMLQQTRVETVIP-YFERFLKRFPTVEALAEADIDEVLARWSGLGYYSRARRLHAAAREIQDRHDGEFPQTRDEAEDLPGVGSYTAGAVLSIAYGLDEPLVDGN  129
c_000001556689_3        18  RDPYAILVSEIMLQQTRVETVIN-RYGSFLERFPTIADLAIASLDDVLAEWSGLGYYRRPRLLHRLATVVVDTHGGEIPDTLEELLELPGIGPYTSAAVGSIAFGIPALSIDGN  130
c_000001834452_1         1  ------WISEIMLQQTRVEVVVP-YFEAFIKRFPSIEVLANAQEEEVLSYWAGLGYYKRSRLLLKAARVIQSEMDGVLPCDSATLRRLPGIGDYTAAAIASISFQEQIPVVDGN  107
c_000004255004_2        35  RDPYAIWVSETMLQQTRVATVLP-YYYRFMKAFPSVEALDQAPLDAVRNVWSGLGYYQRAANLKKAASVLMNRHNGKLPADYQTLLGLPGVGQYTAGAVLSIAFDRRYPAPDGN  147
c_000000605438_3        31  QSAYEIYLSEIMLQQTQVKTVLERFYFQFLEKFPTLKAVANAPVDDILKAWEGLGYYTRARNLHKTAIVTKGI----LPKSAEALEQLSGIGKETAHAVACFAFNEPLPILDAN  140
c_000000581237_15       31  QSAYEIYLSEIMLQQTQVKTVLERFYFQFLDKFPTLKDVANAPVDDVLKAWEGLGYYTRARNLHKTAVDTKGI----LPNNAEKLEELSGIGKETAHAVACFAFDEPLPILDAN  140
```

```
c_000002529579_2        31  QNSYKIYLSEIMLQQTQVKTVLERFYFQFLEKFPTLEAVANAPVDDVLKAWEGLGYYTRARNLHKTAIATKGV----LPKRAEALEKLSGIGKSTAHAVACFAFNEPLPILDAN  140
c_000006097838_1         1  RDPYAIWISEVMLQQTQVKTVIP-YWRRWMEQLPDIASLAAADEDTVIKLWEGLGYYSRARNLQKAARLLCEELSGKFPRDLCGVLALPGVGRYTAGAICSIAYNQPEPLVDGN  113
c_000000583727_6        27  RDPYAIWVSEIMLQQTQVKTVIP-YWQRWMQLPDIASLAAADEDTVIKLWEGLGYYSRGRNLQKAARIYDELAGRFPRDLAGMLALPGVGRYTAGAICSIAYNQPEPLVDGN  139
c_000005807640_2        29  RDPYAIWVSEIMLQQTQVKTVIP-YWQRWMQRLPDIASLAAADEDTVIKLWEGLGYYSRARNLQKAARIRDALSDRFPRDLAGALALPGVGRYTGGAICSIAYNQPEPIVDGN  141
c_000000134878_1        12  NDPYKIWISEIMLQQTQVSTVQD-YYLRWIKRFPTIQSVADASIDEVLKHWEGLGYYARARNFYHSCKALSTTT-SCVPKGMDEFQKLKGVGPYIGAAVQSIAFNIPAGVVDGN  123
c_000005136725_3        27  SDPYKIWISEIMLQQTQVSTVRD-YYIRWIKQFPDIQSVADASIDKVLKYWEGLGYYSRVRNFHKSCQMLIRNK-QDIPKGLDEFQKLKGVGPYIASAVQSIAFNIPTGVIDGN  138
c_000001684786_4        24  PEPYRVWVSEIMLQQTRVETVIP-YFERFLDRFPTPAALAEASEDEVLTLWSGLGYYRRARSLLEGAKAVVRDHEGVFPRSYEAALDIPGVGPYTAAAILSIAYGEPWPVLDGN  136
c_000002971826_23       34  PEPYRVWVSEIMLQQTRVDTVIP-YFERFLDRFPTLSDLAAASEDEVLTLWSGLGYYRRARSLLAGAKTVIEHHAGVFPRDYEAALEIPGVGPYTAAAILSVAYGEPWPVLDGN  146
c_000002040695_2        37  ANPYVFPISEVMLQQTTVATVVP-YFKKWIKKWPTLHALAEASLDDVLHTWQGLGYYTRARNVHRCAQEVVQSYGGLFPQNVSQLRKLPGIGPYTAGALSAIAFEQNSVAVDGN  149
c_000003263657_32       41  NDPYSVWISEMMLQQTQVTQAEP-YYERFMAQFPSVETLAGATLDDVLKAWEGLGYYARARYLHRAAVQVVDQYGGRIPESLSEISILPGIGPYTAAAILSIAYHRDHAAVDGN  153
c_000005989041_2        38  KDPYKIWLSEIMLQQTTVKAVGP-YFQLFLKTWPTIDKLAKATLDDVLRVWQGLGYYSRARNMHKCAKLICNNFGGRFPNTEEQLERLPGIGPYTAAAISAIAFDKPANVVDGN  150
c_000002391082_2        36  DNPYHIWISEIMLQQTRVDQMGA-YFERFVGAFPTVEKLAAASEDQVLKVWEGLGYYARARNMHKAAVQIAGQLGGCIPDTYAGLIELPGIGEYTAAAVSSIAFDRDHPVLDGN  148
c_000003839553_2        43  SDPYHIWISEIMLQQTRVDQMGG-YFERFITAFPTLEALAGASGDEVLKVWEGLGYYARACNMHKAARIAFELEGQIPNTYDGLIDLPGVGEYTAAAVSSIAFDCDHPVLDGN  155
c_000003996707_2        34  GNPYHVWISEIMLQRTRVDQMQP-FFERFTAAFPTVGALAAASQEEVLKAWEGLGYYARARNLHRAAQRVV-EAGGNIPDRYEELVGLPGIGPYTAAAVSSIAFYDLPVLDGN  145
c_000002044706_1        16  DDPYRILVSEIMLQQTRVTTVLR-YYPAFLERFPTLADLAAADQEDVLAAWSGLGFYRRARNLHACARQVVDRYDGEVPRDPEELGSLPGIGRYTVGAILSSAHNARLPILDGN  128
c_000006007075_1        31  KREYYTLVSEFMLQQTQVSTVIP-YFKNFIKDIPDLNSJAKINEKKLLKYWEGLGYYSRVRNLKKTAKIVVSKYKKKLPTTLEELKTLPGIGDYTASAILSIAFRNFPIPLDGN  143
c_000002094036_4         7  QKEYYTLVSEFMLQQTKVSTVIP-YFNNFLKNIPNMQSLAKINEKKLLKYWQGLGYYSRAKNLKKTAKIIVDNYQGKLPSDFEELKKLPGIGDYTANAISAIVFNKPTIPLDGN  119
c_000002762689_1        31  KKQYYTLVSEFMLQQTQVATVIP-FFNRFIKNIPDLEALANFDNHKLIKLWEGLGYYSRVRNLKKTAQIVVKNFDKKLPRKYLDLKSLPGIGDYTASAISAIAFNKRIIPLDGN  143
c_000005516980_2        31  KREYYTLVSEFMLQQTQVSTVIP-YFKNFIKNIPNLKMLAKVQNKKLLIKLWEGLGYYSRARNLKKTSQVIIKNFKGKVPENFEDLKSLPGIGNYTASAILAIAFNKPYIPLDGN  143
c_000003920004_1        37  PNPYYVLVSEFMLQQTTVNTVIS-RFKDFIKKWPNLKKLSMINENQILQFWSGLGYYARAKNLLNSAKIISLKFNNIVPDNYNDLIDLPGVGDYTAKAVLGIGYNKSVMPVDAN  149
c_000004008511_2        31  PDPYFVFVSEYMLQQTTVGTVKT-RFEEFILKWPSINDLARISEKTILNFWSGLGYYSRARNLLKAAKIIKKEFNSKIPNNYDDLINLPGIGDYTTKAILGIAYNKSVMPLDVN  143
c_000004481347_1        38  TNAYYVLVSEFMLQQTTVNTVTK-RFNEFIKLWPSIDKLSRISENRILRFWSGLGYYNRATNLLKTIKIISRDFKFKVPQHYDQLIELPGVGDYTAKAILGIAYNQPVLPLDAN  150
c_000003787733_3        30  NDLYKIWLSEVMLQQTRVNTVVP-YYLLWVKKYKSIKSVAKADYQNLLKLWEGLGYYSRCKNFYKSCLIIINDYDGKIPLDYLTFRSLPGVGDYIASAVLSIGLNQPHAAIDSN  142
c_000000754657_2        30  HDLYKIWISEVMLQQTRVNTVVP-YYLRWVEKYESIKSVAKADELKLLKLWEGLGYYSRCKNFHKACKIIVNDYDGEVPSDYQKFLLLPGVGEYTASAVFSIGLKQKYAAIDGN  142
c_000000990943_3        30  NDLYKIWISEVMLQQTRVNTVVP-YYLRWVEKYKSIKSVAKADDLKLLKLWEGLGYYSRCRNFHKACKIIVNDYGGEFPSDYQKFRLLPGVGDYTASAVFSIGLKQTYAAIDGN  142
c_000004474996_2        29  NDLYKIWLSEVMLQQTRVETVIP-YYLKWVEKYKTIKSVAEADGQKLLKLWEGLGYYSRCRNFHKACKIVVNDFTGEIPSDYQNFRSLPGVGDYTASAVFSIGLNQTYPALDGN  141
c_000001463500_11       17  NNLYHVYLSEVMLQQTKVKTAIP-YYKRWLKKYPTIESVALSNLDDLLKLWEGLGYYSRCVNFYKACIILYSQYNSQLPNNMTDFLKLPGVGNYISAAVYSIVDKIPAPAVDVN  129
c_000001286181_5        24  NDPYKVWVSEIMLQQTQVSTVIP-FYENFMKSFPDVLTLSKANLDEVLNNWAGLGFYRRAKNLHQTAKIITDKYHQVFPNNYEDLVALPGIGESTAGAILALAFNKKGTIILDGN  136
c_000001293628_3        30  REPYRVWVSEIMLQQTRVETAIP-YYNRWMERFPTVDALAEGHEQAVLKAWEGLGYYSRARNLRRSALIVRERLGGVLPQDSRSLKTLPGIGEYSAGAIASIAYGEVVPAVDGN  142
c_000001535696_8        24  RDPYKIWLSEIMLQQTRVAQGAP-YYLKFIKHFATVYDLAKASEEQVLKLWQGLGYYSRARHLHATAKTIVNEYGGNFPNTYKELLKLKGIGDYTASAIASICFDRPEPVVDGN  136
c_000001614067_2        37  SNPYFVYLSEIMLQQTVVKTVIP-YFLKFVRKWPDINALAKAELHEINSYWAGLGYYSRAKNLHETAKIISNKYDGFLPTDKNSLMALPGIGEYTSSAIMAIAFDRKSNVVDGN  149
c_000001765289_1         3  DDPYKIWLSEIMLQQTQVKTVVP-YYNRWIKRYPSIKSVALADLGAVLKMWEGLGYYTRCRNFHTAAKIVVKKFNGIIPNNWENFSSLPGVGDYTAAAVLSIAFNKPYAVMDGN  115
c_000001961666_1        18  VDAYKVWVSEIMLQQTTVNAVIS-RYKKFISRFTNVKVLAKAPLEEVLEEWAGLGYYSRARNLHSSAKIIVNKYEGIIPSEENKLINLPGIGTYTSGAIRAIAFNRPSLAIDVN  130
c_000002018097_4        25  NEPYNIWLSEVMLQQTQVKTVVP-YYTKWIKKFPTLKSVAESDLNSLLKLWEGLGYYSRCRNFHKATQIVVEKYKSVVPNDYKLFRQLPGVGDYIASAVLSIAYNKKYPAVDAN  137
c_000002843512_36       26  DDPYHIYLSEVMLQQTQLKTVLERYYFPFLKRFPTLKTLGETPLDDVLKMWEGLGYYNRAKNLHKTATLIEE-----LPSNIDELIKLPGIGKYTAHAVATFAYKKPVPIMEAN  134
c_000002930199_1        22  NNPYKIWISEIILQQTQIQTGIR-YYHKFIRHFPDINSLALAEEVDVLNVWQGLGYYNRVLNMLYSAKIIVKQHKSTFPTKYDELIKLKGIGPYTAAAISSICKNEKRAVLDGN  134
c_000003033795_2        22  RTPYRVWVSEIMLQQTQVNTVIP-YYKKWIKKYPTLKSFKESNFDDVIKIWEGLGYYSRCHNMFNAAKLI----NSTFPNNYDDLINLPGIGDYTAKTILAIAFKKNFVGIDTN  130
c_000003888107_2        23  HDPYKIWISEIMLQQTQVSTVIP-YYNKWINKFPNATKLSKANYNDVLKLWEGLGYYSRCKNIYKTSHLV----QKNFPDSYSELIKLPGIGDYTAKTILAIAFDKNQVGIDTN  131
c_000004013286_6        44  RDPYAIWVSEIMLQQTQVNTVVP-FFERFIESFPTVEKLAAATEAAVLKRWEGLGYYRRARNLHRAAREVTDQYDGQLPRDPATLSRLPGIGQYTAGAILSIAFDCREPIVEAN  156
c_000004852258_1         6  SDPYKIWLSEVMLQQTQVETVIP-YYQKWINHYPTIKSVALENIDNLLKIWEGLGYYSRCRNFHEAIKDIFYNHNGVIPKQIEDFKSLKGVGIYISGAVMSIAFNQTHIAVDGN  118
c_000005254087_1        32  RDPYAVWVAEVMLQQTQIATVIP-YYERWLARFPTVQALAEASLDEVLKWEGLGYYGRARNLHAAVQMVMNEYDGRIPTTAADLLKLKGIGRYTANAIASIAYNEAVPVLDGN  144
c_000005603677_1        29  QDPYKIWLAEIIMQQTRLEQGLP-YYERFIAAYPSIKELAETEEQKLMKLWQGLGYYARARNLQATAKRIVAEYAGIFPDNYKAILDLKGVGEYTAAAIASAFYAFDLPYAVVDGN  141
c_000003283462_3        17  TGFYRIWLSEIMLQQTQVKRVSI-FYTKWIKRFPTLESVANAKEEHLLKVWEGLGYYNRCRNFYKATKIIIDDYNGKIPNDYQTFRLLPGVGDYTASAVLSIAFGMPYPAIDGN  129
c_000002687221_1        31  SNPYFIWLSEVMLQQTRVSTASP-YYQHWISALPDIHSVAEAPIDNILKLWEGLGYYGRARNFHKACRIIIEKHGGVIPSDPLEFSKLPGVGDYICAAVMSIAFNLPIPAADGN  143
c_000003347358_19       57  DEAYYIYLSEVMLQQTQVKTVLERYYIPFIERFPTLKHLGEAPLDDVLKMWEGLGYYNRAKNLHKTATLVDQ-----LPSNIDELIKLPGIGKYTAHAVATFAFHQPVPIMEAN  165
c_000002701031_2        24  SDSYKIWLSEIILQQTRVEQGLP-YFKEFITKFPKIELLAKASEDEVLKLWEGLGYYSRARNMRLTAKAIVEKFNGVFPKDYEKILKLKGIGPYTAAAICSFAYNLPYAVVDGN  136
c_000000582753_3        32  KSSYKIWLSEIILLQQTKVSQGLP-YYRSFIESYPRLEDLAHSSEEKVLKLWQGLGYYSRARNLHFTAQYIFNELSNEFPNNYKDLLKLKGVGDYTASAIASFAFDEVVPVLDGN  144
c_000001713769_5        24  NDPYSVWLSEIIILQQTRIAQGLP-YYLKFINKYPTINSLAKANENDILILWQGLGYYSRARNLLKTAKFIVDERNGKFPTTYIELIKLKGIGEYTAGAISSICFNERRAVLDGN  136
c_000004369364_1        27  SDPYAIWLSEIILQQTRVAQGLP-YYLKFLEAFPTVQDLAQAPEQDIMLLWQGLGYYSRARNLHFTAKQVSDELNGLFPNSYTGLLSLKGVGPYTAAAIASFAYKLPHAVVDGN  139
c_000006057486_30       28  KDPYLIWLSEIIILQQTKIAQGTP-YYLKFVEKFGNVSLLAKADEKQVLKLWQGLGYYSRARNLHKTAIYINNNLKGVFPNNYTDLKKLKGIGDYTASAIASFCFNLPEATVDGN  140
c_000005494072_10       24  KDPYKIWLSEIIILQQTKIEQGLP-YYMSFVENYPNIETLAIADEQEVLKLWQGLGYYSRARNLHFTAKSIVKNHNGKFPDNYNELLKLKGVGDYTASAISSICFDEKCAVVDGN  136
c_000001742634_3        27  KKPYFIWLSEIILQQTRTTQGLP-YYIAFIKHFPTIFDLAKTSETKVLKLWQGLGYYSRARNLHFTAKYIVDNYNGKFPENYKELIKLKGVGDYTASAIASICYNEKTAVVDGN  139
c_000002826998_2        27  KNPYLIWLSEIILQQTRVEQGLP-YFLKFKEKYPTVTDLANAEEDEIMKLWEGLGYYSRARNLHFSAKYITAELDGIFPKHTEILKLKGVGTYTAAAISSFAYDLPHAVLDGN  139
c_000003159439_6        24  SNPYHIWISEIILQQTRVDQGLD-YYKRFIATFPTINNLAKAKEEAVLKLWQGLGYYSRARNLHFSAQYIVNELDGKFPSSYKELLKLKGVGDYTASAIASICYNEVTAVVDGN  136
c_000004488714_1        29  REPYKIWLSEIILQQTRVAQGLP-YYQRFIESFPIQLNLAAASEQNVLKLWQGLGYYSRARNMHFTAQYVVNELEGEFPCSYKELMKLKGVGDYTASAVASFCYDEVQPVLDGN  141
c_000005037037_2        27  KNPYYIWLSEIILQQTRVEQGTK-YFLAFKKNYLTIFDLAQAKETEVLKLWEGLGYYSRARNLHFTAKYIANDLEGVFPRNYTDLLKLKGVGSYTASAIASFAYDIPKAVVDGN  139
c_000001059964_1        19  YEPYQVWISEIMLQQTQVQTMLP-YYHRWMERLPRLQDVAEASEDLLIKLWEGLGYYRRVHNIKKTAVIICRDYGGTFPGSYRQQIQLPGIGPYTAGAISAIAFNFSEPVLDGN  131
c_000002498472_1        28  YSPYQVWISEIMLQQTQMDRVTL-FFTRWMDTFPDVNTLASASETRVLKLWEGLGYYSRARNILKTAQLLVTEYQGQIPDDPKQLLSFPGIGPYTAGAIASISFNHDVPVVDAN  140
c_000004615912_3        17  KNPYLIWLSEVILQQTRTAQGLP-YYEKFKNRFPTIDLMANAHEDEIMKLWEGLGYYSRARNMHFTAKYITEELGGEFPDTLEGLKKLKGVGPYTAAAIASFAYDLPHAVLDGN  129
c_000004766858_2        30  YHPYEIWVSEIMLQQTQVKTMLP-FYTRWMSELPTVHSVAAASEEMVLKLWEGLGYYSRARNLQKAARVICDEHGGELPADFDHIRALPGIGPYTAGALASIAFNQPRPVVDGN  142
c_000002747260_18       36  PDAYAVWLSEVMLQQTTVAAVRD-YFRRFTTRWPTVADLAAAEDADVMGEWAGLGYYARARNLLKCARVVVAEHGGHFPDTRDGLLKLPGIGPYTAAAIAAIAFDEPAVVVDGN  148
c_000005371561_38       36  PDPYRVWLSEIMLQQTTVAAVRD-YFHRFTARWPTVADLAAAEDGDVMGEWAGLGYYARARNLLKCARVVVSEHGGRFPDDPALLALPGVGPYTAAAIAAIAFDKPETVMDGN  148
c_000000169465_2        32  VTPYKVWVSEIMLQQTQVTTVIP-YFKRFTAMASFPTVHDLAKASQDDVLHHWTGLGYYARARNLHKAANTLVEKYNGEFPYSLEEVMDLPGIGRSTAGAILSLSRNMRFPILDGN  144
c_000003254110_11       32  VTPYKVWVSEIMLQQTQVTTVIP-YFERFMASFPTVHDLAKASQDDVLHHWTGLGYYARARNLHKAANRLVDEYNGEFPFSLEEVMDLPGIGRSTAGAILSLSRNMRFAILDGN  144
c_000004750284_20       32  VTPYKVWVSEIMLQQTQVATVIP-YFNRFMASYPSVFELANASQDDVLHHWTGLGYYARARNLHKAAKTLVEDYEGVFPPTLDEVVSLPGIGRSTAGAILSLSRDERYPILDGN  144
c_000000141782_15       37  ATPYRVWISEIMLQQTQVATVIP-YFERFLQRFPDVATLAAADADSVLHLWSGLGYYARARNLHRAARVVVEAHGGEFPRRFEEVQALPGIGRSTAGAILSLGCGQRHAILDGN  149
c_000002717847_8        31  ITPYKVWISEMMLQQTQVVTAQS-YYSKFLKKYPSLTSMKMATEDDILSLWSGLGYYRRASFIYQAKEKIFLDFNGRFPATYDELITLPGIGRSTAGAILSIAYQKPFPILDGN  143
c_000004054799_1        18  ISPYRVWISEIMLQQTQVKTVIP-YFYKFIKKYPNLGALLQASEDEILAQWSGLGFYRRAKNIYKACRIISEDFNSKLPTNINDLESLPGIGRSTAGAIMSIAFNEGHPILDGN  130
c_000006126673_1        43  PNPYFVWLSEIMLQQTTVSAVIP-YFLKFTQRWPDVWALANADVDDIMSAWAGLGYYARARNMHKCAKTIVSDYDGIFPQEQAVLKTLAGIGDYTSAAITAIVFDKYATVVDGN  155
c_000000339186_3         1  -DPYSIWVSEIMLQQTQVSTVIA-YWERWMRALPTIDKLARAKPTQVLKLWEGLGYYSRARNLQVAAKEIVKNHDGQFPRDRPAIRALPGIGRYTAGAISSIAFGLPEPIVDGN  112
c_000003029168_2        32  RDPYAIWISEIMLQQTTVESVIP-IYERFMELFPTVDVLAVANPTDVLQAWAGLGYYKRARHLHEAAIEIVNRFGSRIPDHLSDLQSLPGIGRYTAGAIASIAFDQCVPVVDTN  144
c_000001863436_1        48  GDPYAIWVSEVMLQQTQVVTVID-YYTRWLARFPDVETLAAAEIDEVLELWAGLGYYRRARLLHRAAQHVVEECGGALPQTAGELEDLPGIGPYTAGAIASIAFGEAAPLVDGN  160
c_000002830137_4        41  RDLYHIWVSETMLQQTQVATVVP-YFQRFVERFPNITSLAAADEEDVLRLWEGLGYYRRARNLHRAAQQVVQQHHAQFPADFDTLLSLPGIGRYTAGAILSIALDQRLPILEAN  153
c_000004612302_3        35  RDPYAVWIREIMLQQTTVKAVVP-YLERFLDQFPTIEVLAASETDEVLRVWEGLGYYSRARNLHRAARMIVEEWGGQWRETAVELQALPGVGPYTAGAIASFAFDQPAAIVEAN  147
c_000001933926_1        21  TDPYRVWLSEIMLQQTTVATVTP-YFLRFVERWPTVHDLAAADLDRVLHAWQGLGYYARARNMHACAKRISGDMAGVFPDQENELRRLPGIGPYTAAAIAAIAFDRPATPVDGN  133
c_000003136334_1        33  MDPYRVWLSEIMLQQTTTITVRG-YFEDFVSRWPSIEDLAAAELDAVLHAWQGLGYYARARNLHKCAKLITKEYGGKFPEREEALVKLPGIGPYTAAAIAAIAFGHIATPVDGN  145
c_000005849454_9        30  RDPYRTWVSEIMLQQTRVEAVIP-YFEAFLKRFPDVHALATATEEEVLQRWAGLGYYRRARFLHAAAQKVVEDWDGRFPETADTLRAMPGVGAYSAAIASLAFGEAVPVVDGN  142
c_000001169194_1        41  TDPYRILVSEIMLQQTRVETVLR-YYESWLKFPTLGALASANSREVLKAWEGLGYYRRARNLHSAARRIRELPGGHFPTSYRDLRKLPGVGQYTAGAVASIGFGEIVPAVDGN  153
c_000002786947_2        41  TDPYRILVSEIMLQQTRVETVLR-YYESWLKQFPTLGTLASADSTEVLKAWEGLGYYRRAKNLHRAAKHIRNVLGGRFPTSSRDLRQLPGVGQYTAGAVASIGFGEIVPAVDGN  153
c_000001279808_52       28  RSPYYVWISEIMSQQTQIDRVAEKFFPRFIATFPTVEALAAASWDEVYPVWEGLGYYNRGKNLLKAAQIIVEKHDGEIPQTAVELEALPGVGKYTAHAVLAFAFDQKVPAIDTN  141
c_000000358065_2        31  QKQYFILISEFMLQQTQVKTVIP-YFENFINKIPNLKSLAKVNNRKLMKCWEGLGYYSRAKNLKKSAKIIIDNYKGRLPNDFEELKKLPGIGDYTASAISAIAFNKPIIPLDGN  143
c_000003745941_1        22  AALYPTVVSEFMLQQTQVKTVLP-YFHRWLKRFPDFKSLAKAKEETVLKHWEGLGYYSRARNLHRLAREILALN--SIPEDPESWQQFKGIGAYTAAAITSISFNHPVACVDGN  132
c_000004187032_4        31  KKHYYTLVSEFMLQQTQVVTVIP-YFDRFINKIPNLKKLSTIQDNELILKWEGLGYYSRARNLKKTAQIIISKHHGKIPNNYEDLISLPGIGNYTANAILAIAFNKSYIPLDGN  143
c_000006067315_3        31  QTEYFTLVSEFMLQQTQVKTVIP-YFNNFINKIPNLKKLANVNETRLMKCWEGLGYYSRARNLKKTAKIIISEFNSNLPNTIEELKNLPGIGDYTSRAVLAIAFNKPIIPLDGN  143
c_000005543774_1         1  KKLYYRILSEFMLQQTQVKTVIP-YFNNFVKKVPSLKALSKSNEKKILKWEGLGYYSRAKNLHKTSKILVKKYDSQIPDSFEKLKELPGIGDYTANILLALIYNHPRIGIDGN  113
c_000006211484_1        23  KKQYYRILSEFMLQQTQVKTVIP-YFNNFVKKIPSLKALSNSDEKTVLKLWEGLGYYARAKNLHKTSKILIKKYNGQIPKSFVKLKELPGIGDYTANILLALIYNQPCIALDVN  135
c_000002718976_3        32  KKLYYRLLSEFMLQQTQVKTVIP-YFNKFTKKFKTIKALSKSNEREILKMWEGLGYYRRARNLLACSKILVKNYKSRLPRSIIEIKKLPGIGDYTANALLGLVYNEPRIAVDGN  144
c_000002992548_1        33  KEQYYRILSEFMLQQTQVKTVIP-YFKNFINKIPNLKVLSSSNEKTILKLWEGLGYYRRAKNLHKTAKILTKNYKGKLPKEFDEIKKLPGVGEYTANALLALVYNQPRIPFDGN  145
c_000000456751_1         1  ----YVLVSEFMLQQTQVKTVIP-YFTKFVKKIPNLKSLSTSNEKKILKLWEGLGYYRRARNLHKTAKILIKNCKGKLPKKFAEIIKLPGVGEYTANSLLALVHNKPSIPIDGN  109
c_000002655634_1        28  KKEYYVLVSEFMLQQTQVKTVIP-YFKKFVTKIPNLKTLSTSNEKQVLKLWEGLGYYRRVRNLHKTAKILLKQYKAKLPKKFQEIIKLPGVGEYTANSLSALIHNKPCIPIDGN  140
c_000004232403_2        27  NDPYKIWLSEVMLQQTRVETVIP-YYNTWLENYKTIQSVAQANLDDLLKIWEGMGYYSRCKNFYYAAKEIICKYKGTIPNQFDVFKSLPGVGDYIAGAVMSIAFNQPHPSIDGN  139
6u7t.before.V147.pdb    28  RDPYKVWVSEVMLQQTRVETVIP-YFEQFIDRFPTLEALADADEDEVLKAWEGLGYYSRVRNLHAAVKEVKTRYGGKVPDDPDEFSRLKGVGPYTVGAVLSLAYGVPEPAVDGN  140
Consensus_aa:                .ssY.l@lSElMLQQTpV.hhhs.Y@.p@hpp@PslpsLA.tpbppllphWpGLGYYsRt+Nlhphtp.l.pph.t.hPpshp.l..LPGlG..TAtAlhtht@sp..shlDtN
```

```
Conservation:                   95  9                                                  75 87 85  9    5 9      8     9 5
sp_P17802_MUTY_ECOL     141  VKRVLARCYAVSGWPG------KKEVENKLWSL----SEQ-VTP---------AVGVERFNQAMMDLGAMICTRSKPKCS----LCPLQNGCIAAANNSWALYPGKKPK-----  225
sp_P83847_MUTY_GEOS     147  VMRVLSRLFLVTDDIA------KPSTRKRFEQI----VRE-IMA---------YENPGAFNEALIELGALVCTPRRPSCL----LCPVQAYCQAFAEGVAEELPVKMKK-----  231
c_000003652391_2        150  VERVMARLFDISTPLP--------GAKPALKAR----AAS-LTP---------KLRPGDYAQAVMDLGATICTARAPTCG----VCPLNDCCHARRAGTAAELPKRTPK-----  232
c_000000031207_6        139  VERVLSRYLKIDGNPK------SGTARKRIQQA----ADD-ALD---------QENPGDHNQALMDLGRSICTPRSPDCQ----RCPLASGCGARASGEPTRWPTRRQK-----  223
c_000000430975_4        151  VERVVSRIFQITQPIP--------LSRPTLRQL----AAG-LTP---------SKRPGDYAQAIMDLGSLVCTPRKPKCN----VCPLHDICLANISGTPSDFPKRTPR-----  233
c_000001007148_4        148  TARFYARLLALRSDLH------SGEGRRQLWSL----AER-IVP---------RDRPGEFNQALIDLGATVCGV-TPRCQ----ECPVRGSCRSLELGEVERIPARASR-----  231
c_000001092848_3        161  VERVITRLRAIGTPLP--------AAKPQVRKA----VAA-LVP---------AERPGDFAQAMMDLGATICTPRRPSCI----LCPLQAFCAARGAAAQESFPVKASK-----  243
c_000001151029_1        108  VERVVSRIFQITQPIP--------LSRPTLRQL----AAG-LTP---------SKRPGDYAQAMMDLGSLVCTPRNPKCN----VCPWHDICLANISGTPSDFPKRTPR-----  190
c_000001797282_1        170  VERVLSRAYAVEAPLP--------GSRPEIRRL----TQA-LVP---------PDRPGDFAQALMDLGATICTPKRPACA----LCPWMRPCRARSLGTQESFPRKIKV-----  252
c_000001803648_25       153  IERVMARMHDIHSPLP--------SAKPELLAC----AAA-LTP---------DFRAGDYAQAVMDLGATICTPKSASCG----VCPWSFACLARINGTAAVLPKRTPK-----  235
c_000002078955_4        153  VERVMARLHNIHTPLP--------AAKPELFSK----ATA-LTP---------DLRAGDYAQAVMDLGATICTPKSPTCG----ICPWSCACVALISGTAAELPKKTPK-----  235
c_000002106160_4        141  VNRVMSRLLRLLRFTP--------YNKKRVLST----LIR-WLD---------RSRPGDFNQALMELGSTVCLPKNPKCE----NCPLKLSCGGTLSGSPESYPSPKPK-----  223
c_000003607531_2        151  VERVITRVFGIFKPLP-----EARPIIKDCAAK----LAP-KTT---------RGRPGDYAQAVMDLGATVCRPRKPLCD----LCPWEKYCIANQQNFVDQVPRKAPK-----  236
c_000003872363_4        149  VERVMARLFNVITPLP--------AAKPELTAH----AAA-LTP---------VDRAGDYAQAVMDLGATICTPKSPGCG----ICPCTEYCAALPEGDAANLPLRPPK-----  231
c_000004820107_1        156  IERIMARYLAITEPLP-------DSKANLKNLAA----GLS-EER---------KDRPGDYAQAMMDLGATICTPTSPKCS----LCPVNDSCTAKRQGIAESLPIKKKK-----  240
c_000005774797_1        140  INRVLSRVYNIDDAIN--------RPSARLARI----AET-LAH---------PKRSGDVAQAMMDLGAAICKPKNPNCE----TCPICDQCKAHSLGRTHELPVKAPQ-----  222
c_000005867021_3        153  VERVMARLYNIHSPLP--------AVKPMLLTH----ATA-LTP---------EFRAGDYAQAVMDLGATVCTPKSPTCG----ICPWSFDCIAHAKGTAAQLPKKTPK-----  235
```

```
c_000004546210_2      144  VKRVLSRFFGIKGWSG------ESKVSKELWDL----SAK-SLP---------VDNFQVYTQGIMDLGATICLSRNPNCS----ICPIQRRCYAHQNHLVDEIPSKKPK-----  228
c_000004551008_5      144  VKRVLSRFFGVKGWSG------ESKVSKELWDL----SAK-SLP---------VDNFEIYTQGIMDLGATVCLPKNPNCS----NCPIVNKCYSYLNQLVDEIPSKKLA-----  228
c_000002511148_4      139  VKRVLARHFKIEGKLE------TSSKRKELWQI----SED-LLP---------DNKIEVYTQAIMDLGATLCIKSSPACN----LCPLNQDCLALKEDLVTVLPSKTKP-----  223
c_000000577378_2      127  VKRVLARYFKIEGKLE------TSSKRKELWQV----SEG-LLP---------DNKIEIYTQAIMDLGATVCTKSNPVCN----LCPLNQDCLALKEDLVTVLPSKTKP-----  211
c_000000598175_2      150  VKRVLTRVFVVDGWPG------KADVLKKLWLL----AEQ-TTP---------ASRNADYTQAIMDLGATLCTRSKPRCE----DCPLKADCLAFQQGNQTVYPTSKPK-----  234
c_000000754627_3      147  VKRVLARFFMVEGWYG------VKKVENQLWHL----SEQ-LTP---------KNNVTEFNQAMMDLGASLCSRSRFDCE----ACPLNSRCGAFNAGKVKEFPHSKPK-----  231
c_000000811118_1      138  VKRVLRISGSKSPMN-------KYKTEKSLKEF----ANI-LLP---------NSKHKEYSQGLMDLGAIICRPNNPNCS----ECPVDEDCAANQKEIQDEIPIKITK-----  222
c_000000899687_2      142  VKRVLSRLFCINKNGT------TSAFENCLWGE----AED-LLP---------VRRPGDFNQALMELGATVCIPASPICQ----QCPLRTICKAFLKNKVNEFPPSKNK-----  226
c_000001176522_24     161  VKRVLCRLFGWGVEG-------ERQVEAHLWTL----AEC-LLPA--------QSEVMPDYTQAQMDLGATLCTRTKPRCE----ACPFQAECVAYRQGRVAELPAPRK-----  247
c_000001345122_1      117  VKRVLCRFDAVTSWSG------NKKTEKVLWDR----ADE-FTP---------QERFDDYTQAIMDLGATLCTRSKPKCD----VCPVEADCDAKKMDKVIQFPYSNPK-----  201
c_000001515736_6      131  VRRVLTRVLAFDGDLA------QSRNERQLWEY----AQQ-LCPTE-----NLHEAMPRYTQGMMDLGASVCTSRKPTCL----VCPLDSECRAARAGNPENYPVRTRR-----  219
c_000001682161_6      138  VKRVLCRFDAINSWSG------NKQTESQLWQR----AEQ-LTP---------SKRVKDYTQAIMDLGATVCTRSKPSCE----ICPVAENCQARQLNQVAKYPFSKPK-----  222
c_000001923643_29     141  VRRVLTRVLAFDADLA------VARNERELWDL----AQQ-LCPTE-----DLQQAMPRYTQGLMDLGATICTPRKPSCL----VCPLQPQCRAARFGNPENYPVRTRK-----  229
c_000002030994_5      139  VKRVLSRVFTVVGWPG------KADVLKKFWVL----AEQ-TTP---------KTRNADYTQAIMDLGATLCTRSKPRCE----DCPLKTDCLAFQQGNQTEYPTKKPK-----  223
c_000002038721_4      141  VKRVLLRVLGNKSPIN------HSKTQKELKKF----AAE-LLP---------KNDFRTYSQGIMDLGSMICTPINPTCK----NCPIQNKCKSFQLNLQNMIPVKTSK-----  225
c_000002523527_15     156  VKRVFARVFGIHGYPG------ERAIENRMWAL----AEQALPPAG----RHQADHMVATQGLMDLGATICSRGKPACLADAEACPLVADCVARKDNLTAVLPTPRKP-----  250
c_000002608528_2      139  VKRVLTRVFTVEGWPG------KSGILKKLWLL----AEQ-TTP---------QTRNADYTQAIMDLGATLCTRSKPRCE----DCPLRAGCLAFQQGNQTDYPTKKPK-----  223
c_000002961510_2      142  VKRVLSRLFCINENGT------TSASENRLWQQ----AED-LLP---------VRRPGDFNQALMELGATVCIPKSPICQ----QCPMRTLCKAFLKNKITEFPPSKKK-----  226
c_000002979152_8      147  VKRVLARYHAIEGWPG------KSAVLKELWEK----AEL-HLP---------AERNADYTQAMMDLGATLCSRSKPACE----RCPVVDGCAAARKSSWALYPGKKPR-----  231
c_000003248034_1      159  VLRVLTRVLDEKGNVA------STAIQQKLTAA----ARL-LVES------VGPHSAGTWNQALMEVGATVCTPQSPSCD----RCPINRWCSAWKAGTQKTRPVKARP-----  246
c_000003294679_3      144  VRRVFTRHKAIKGWPG------TTMVQNVLWKI----AED-NTP---------SNRVADYTQAIMDLGATICTRSHPRCE----LCPVNKDCIAHSRGQEAEFPHKNIK-----  228
c_000003333364_1      141  VERVLARLFDVDAPIK------SSQAQSFLWET----ARS-LIP---------TGQAREFNQAIMEFGNLVCGR-KPSCM----FCPLTEECEAHYLGIVEHRPVPAER-----  226
c_000003402687_3      142  ISKIITVLWPG----------------QDTVEV----ANE-LVS--------LSQSGYAWNSAMMDLASSLRAGNPVE------GALGETYFSDEAICRLFKPKRKKPEKKIK  219
c_000003492925_4      134  VKRILCRLHKLKTP-----------SDKELWRI----AYN-LVD---------KKNPFDYNQAMMDIGATLCTPKKPDCE----TCPLSDICQGQNDPTLYPTKKKR-------  211
c_000003766664_5      154  VIRVLSRLCDLEEEVG------ETKAKERIWDW----AER-LVD---------PQNASAFNQGMMELGAMLCTPTSPGCD----QCPWREDCRARAAGTVLERPIKKAK-----  238
c_000003948186_72     141  VKRVLSRVHRVAGWAG------RASVNKELWQL----AER-YTP---------KRRVERYTQAIMDLGATVCVRGQPRCH----ECPVAKHCEALRHDCVQSFPQPKAR-----  225
c_000005057120_4      140  VKRVLTRYFAISGWTG------NSKISNELWEI----SSR-YTP---------KIRVADYTQAMMDLGATLCTRSKPQCT----ICPIQADCLAKKQNKTAELPTPKPA-----  224
c_000005590109_5      140  VRRVLARFHVISGWTG------GTKVSNKLWQL----SSD-YTP---------LLRIADYTQAIMDLGATICIRSKPKCT----QCPINSACLAKIEGKISELPTPKPK-----  224
c_000006063368_2      139  VKRVLTRHRKIETDLS------KSSTLKKLWKL----SEL-LLP---------DVRIDTYTQAIMDLGAIVCKKKSPLCE----ICPVNKDCLALKANLVDCLPVKRKI-----  223
c_000000467631_2      119  VKRVLARYHAISGWTG------SAKISSELWKI----SAR-YTP---------AERCADYTQAMMDLGATLCTRSKPRCG----ECPISSDCVARIGDKVKLLPTPKPA-----  203
c_000000788235_1      140  VKRVLTRYHAIEGWSG------KGHILKQLWSI----AES-HLP---------DTRYANYTQAMMDLGATVCTRSKPCL----GCPIHADCSAFQAGTPTAYPTPKPK-----  224
c_000001187176_11     141  VKRVLARYHAVQGWPG------EKPVHDRLWAI----AEA-HTP---------HSRVAHYTQAMMDLGATLCTRSKPSCL----ICPLQAGCQARMLGEPTAYPQARPR-----  225
c_000001660697_64     141  VKRVFARVFGIDGFPG------DKRVEDTLWRI----AEAVLPP---------ADGIQPYTQGLMDMGATVCTRGKPACLTGERACPLESLCEARSTDRVMELPVPRPR-----  230
c_000003378864_2      152  VKRVLSRLYDVSDP-----------VGTAAATRTLWAYSTELLN--------QADDAYTHNQAIMELGATLCTPKKTACT----RCPVSVRCDALARDTVPLRPIKRPR-----  237
c_000003910742_3      141  VKRVLTRYHAVEGWTG------KSQVLKQLWLL----AES-HLP---------ATRYANYTQAMMDLGATVCSRTKPDCD----HCPLNSNCAALSKGNPTAYPTPKPK-----  225
c_000004116181_2      107  VKRVLARFNAVQGWPG------NASVNKQLWAI----SAR-YTP---------VERVADYTQAMMDLGATVCTRTRPGCQ----ICPVNENCLARKKSLTEIIPASRPV-----  191
c_000006223903_1      142  VKRVLSRYHAVEGWSG------RAEVARRLWSL----AEA-HTP---------TRRVAAYTQAIMDLGAMVC-RRRPDCA----TCPLAGGCKAHKQNQVHDYPASKPT-----  225
c_000001029068_2      114  VKRVLARYHQVDGHYG------QGKTIKQLWKL----SEK-HTP---------NTRNNDYTQAIMDLGSIVCTRLNPVCD----CCPVALRCLAKKTHTQHLYPASKPK-----  198
c_000001128125_11     149  VKRVLSRYFGIHGWPG------EKKIEADLWHK----ADL-LTP---------PKRFADYTQAIMDLGATTCKRSKPLCG----VCPVQKSCVALKREETHSIPASKPK-----  233
c_000001595844_4      140  VFRVLARYTANFIDIG------TPAGKKEFTLL----ANQ-LLD---------KKVPTNPDCS----NCPLQIKCQAFSQDTIAQLPVKEKK-----  224
c_000001719155_2      144  VKRVLSRLFLLKENGE------TRKSENILWET----MQQ-LLP---------ETGAGNFNQAIMELGATVCLPKNPLCL----LCPLKRNCQAQKSGKQNLYPPRKRA-----  228
c_000002363038_6      149  VKRVLSRYFAIEGWPG------EKRVENQMWGL----ADN-LTP---------ETDFADYTQAIMDLGATLCKRSKPNCN----LCPVKTDCRALKTDQVASFPNSKPK-----  233
c_000002561400_1      105  VKRVLTRFFGVHGWPG------EKQVENRLWHR----ADE-LTP---------SVRFADYTQAIMDLGATLCKRSKPNCI----TCPVHSGCQALKKGEVHLLPSFKPK-----  189
c_000002566035_2      141  VKRVLTRYHGIKGWPG------EKKIENKLWSL----AEL-YTP---------RKLTKQYTQAIMDLGATVCTRRKPDCD----ICPLTLTCFAREMGKQHDFPGKKPT-----  225
c_000002598045_3      139  VKRVLARHFKIEGNLQ------ASSTTKQLWQI----SES-LLP---------GDRIDIYTQAIMDLGATVCTKSNPLCS----NCPVNQDCSAFKEGLVDTLPLKVKR-----  223
c_000004188575_41     141  VKRVLARFYAIASWPG------NKATEQQMWQQ----ADN-LLP---------DQRIAAYIQAQMDLGATLCTRSKPNCD----CCPLQSNCQAYARGTPTDFPIAKPK-----  225
c_000000240049_1      139  VKRVLARHQAIEGWSG------QSPVSKKLWTI----AEQ-FTP---------ENQLADYTQAIMDLGATVCTRSSAKCG----VCPLSEDCLAYQQQRVADFPGKKPK-----  223
c_000002139456_2      139  VKRVLARYQAVEGWSG------QSQVSKKLWAI----AEQ-LTP---------KEQLADYTQAIMDLGATVCTRSTAKCG----GCPLSEDCLAHQQQTVADFPGKKPK-----  223
c_000003535614_2      141  VKRVLARYQAIEGWPG------KPAIEKQLWVF----AER-YTP---------TEELADYTQAIMDLGATLCTRASPKCG----SCPLNNGCKALANNCVAALPTRKPK-----  225
c_000002657784_4      138  VKRVLARYKKIEGDLS------KALNIKTLWRI----SES-LTP---------EEKVDLYNQAIMDLGALVCTKSSPKCR----ICPVSRDCIAFNKKLIQVLPRKKLT-----  222
c_000003800129_15     143  VKRVVGRLYNQESY----------QEKKFWEL----SKK-ILD---------KSNPFLFQQGIMDIGATICHIHNPQCD----ECPLGRWCLSKKNGSFYTYKKKK-------  220
c_000003716781_1      130  VRRVLARWFAVEGVV------TKAEPRREIEGL----AAA-LVG--------GVNDPGSWNQALMELGATVCLPKDPRCD----SCPVMTECRARARGLERAVPELPVR-----  215
c_000001556689_3      131  VARIMCRLAGIADDTR------RSKVRQRLEAM----AAD-AIE---------CHPPGTFNQALMELGARVCTPRSPRCD----QCPCSSYCEAHELGIEEQIPPRRSK-----  215
c_000001834452_1      108  VCRVAARFLGLKSPLG------SSALRGQARDW----GET-LHAG------VSAGLAGSLNEGLMELGATVCRPRAPLCG----TCPISEKCVARATNRVTEIPKKAKR-----  195
c_000004255004_2      148  ARRVYARLLSLTN-------------LRAINET----AEQ-MVS---------HSRPGDFNQAVMELGATVCLPGKPRCE----RCPLAYRCQARSVGAFKLQLRRPTR-----  225
c_000000605438_3      141  VKRILYRFFAVKEA------------NDKKLWEM----AYD-LYD---------KDNAYIYNQTMMDIGSAICTHKNPLCT----QCPFVSLCQGKDEPL--LYPTKKKK-----  218
c_000000581237_15     141  VKRILYRFFAMTSC------------NDKKLWEM----SYE-LYD---------RENSYIYNQTMMDIGSAICTHKNPACD----VCPFEPLCQGKESPL--LYPEKKVK-----  218
c_000002529579_2      141  VKRILYRFFAVKEA------------NEKKLWEM----AYD-LYD---------KENAYIYNQTMMDIGSLICTHKNPLCT----QCPFVSLCQGKDEPL--LYPTKKKK-----  218
c_000006097838_1      114  VARVLTRLLGIERPSK------QKQVVEQLWSV----TAD-LVKHAVDLQKTSRRNATAFNQAMMELGALVCTPRNPLCH----DCPLKGHCIARKQNKVNRIVSAKR-----  207
c_000000583727_6      140  VARVLTRLLGIEQSPK------DKPVTDRLWSV----ADE-LVKHTVTLQKPRQRNASAFNQAMMELGALVCTPRNPLCG----DCPLKGRCTARKQNKVETIPALAKR-----  233
c_000005807640_2      142  VARVLTRLLGIEESPK------ARPVADRLWSV----ADE-LVNHAVTLRKPSQQNASAFNQAMMELGALVCTPRNPLCG----DCPLKGRCIARKQNKVETIPALAKR-----  235
c_000000134878_1      124  VNRVVSRFYAYQDPPS--------KNKKKMEKF----MSL-IIN---------DNRPGDINQSVMDMGRYICKPRTPLCD----ECPIHNGCKALKLGIQSDLPIKIKK-----  206
c_000005136725_3      139  VNRVVSRFYGYRDPPL--------KNKKKIEKF----MNA-IIN---------DKRPGDINQSVMDMGRYICKPRTPLCD----RCPIQYGCKALKLGIQADLPVKIKK-----  221
c_000001684786_4      137  VERVMTRILRLAENPR------QVSTKKRLKEI----AGG-WMP---------ADKASSFNQAMMELGALICSPVSPDCP----SCPVREVCAAAEQGDPENFPRLPPK-----  221
c_000002971826_23     147  VERVLTRILRLADNPR------TAATKKYLKEI----AGG-WMP---------RGKASSFNQAMMELGALICVPVSPNCA----ACPIHDACAAAEHGDPENFPKLPPR-----  231
c_000002040695_2      150  IARVFSRVLQYPGSKE-------ALTTSLRQP----ISL-FLP---------FKRIGDYTQALMDLGALLCRPCNPQCS----LCPIKTLCKSAHHNTVHLFPHKTKK-----  232
c_000003263657_32     154  VIRVLSRLYRIHEDAT------KGPAKRRFQEL----ADG-LLQ---------KGQAADFNQAMMELGATVCLPQKPRCT----VCPVQCHCRAFHELSDPSGLPYKPP-----  238
c_000005989041_2      151  IERVITRIFGICTPFP-----EAKTIVKDCAAK----LAP-KTT---------RGRPGDYAQAVMDLGAIVCRPKKPLCD----LCPWKKCCIANQRNFVDQVPQKVSK-----  236
c_000002391082_2      149  VTRVLCRLLRIEEDPR------RTAIEAELIAA----GEA-LLA---------RGEAGDFNQAMMELGARVCTPRNPLCP----TCPVGQWCRAQAELDDPSALPYKVP-----  233
c_000003839553_2      156  VIRVLCRHLRIEGDPR------RAQIKAELIAA----VEA-LLA---------RGQAGDFNQAMMELGARICTPRKPLCE----TCPVGATCRARAELDDPTALPYKAP-----  240
c_000003996707_2      146  VTRVLCRLLRLEGDPR------KAAVKTELIAA----GER-LLA---------RGRAGDFNQAMMELGARICKPTSPACG----QCAVVEMCRAHAELDDPAQLPTKVR-----  230
c_000002044706_1      129  VIRVLTRLFRIPGDPG------TAAVSRRLWQL----ATD-VLP---------QNRPGDFNQALMDLGATVCKPIAPLCE----QCPLAGLCAVRAAGNAEDFPNAGRR-----  213
c_000006007075_1      144  VERIIKRILNLKTEKE--------ISKENIIKK----KKI-LGM---------SDRSSDYAQALMELGALVCKPKNPYCK----KCPITKNCLSYKKKDFEI---ISKN-----  223
c_000002094036_4      120  IERVLKRILNLKSEVK--------TKKENLHKQ----KKV-FGQ---------TNRSNDYVQALMEIGALICKPKIPHCN----KCPITKNCLSYKKKDFKI---KPKN-----  199
c_000002762689_1      144  IERVLKRYLYLKKENE--------INKENLIKK----KKI-FGY---------SNRASDYAQALMELGALICKPNNPHCE----QCPISNKCIALKKKDFLL---TKIK-----  223
c_000005516980_2      144  IERVLKRYLYLKKENE--------IQKENLIEK----KSV-FGL---------SQRSSDYAQALMELGALICKPSNPLCN----QCPISKKCKSFRKKDFNL---IKNT-----  223
c_000003920004_1      150  IKRMLARLYGLDQSIN--------LINKKITSL----SKF-YES---------KKQSSNLIQAFMDYGSIICVPRNPKCG----ICIISRECIANQRKISNIIPKKIKS-----  232
c_000004008511_2      144  IERILVRLYGLKLPII--------KIKNDLRKK----SDN-FIS---------KNSSSDLIQAFMDFGSLICTPRNPDCV----NCLIQKNCNAFNYNLQDIIPVKLKS-----  226
c_000004481347_1      151  IKRIIARLYGIKTSLL--------LNKKTIEDI----ASK-YQS---------SNKASDLIQSFMDYGSVICLPRNPKCD----ECLIEKFCEARKKNIQHLIPFKNLS-----  233
c_000003787733_3      143  VKRVMARILKIKHITK--------RNNRRIRNR----LIK-WMD---------PERPGDINQALMDLANKICRVDHAYCN----TCPIDKVCMAEKMSIPESYPTKLNK-----  225
c_000004754657_2      143  VRRVMSRLLRIKNITK--------RNKKRIHNT----LIE-WMD---------RERPGDINQALMDLANKVCRVDHAQCN----SCPIDEICMANKMSIPESYPTRIKR-----  225
c_000000990943_3      143  VRRVMARVLRIKNITN--------RNKKRINNT----LIK-WMD---------PVRPGDINQALMDLANSICRVDHAHCS----ICPIDEICMAEKMPIPESYPIKVKK-----  225
c_000004474996_2      142  VIRVMARLLRIKNITK--------RNKNRIYNT----LIK-WMD---------KERPGDINQALMDLANKICCVDHALCS----TCPIDEICMANNMSIPESYPAKLEK-----  224
c_000001463500_11     130  VTRVLSRILAIRNLTP--------YNRRRIFNV----LEK-FID---------KERPGDVNQALMDLGSSICTPINTECI----ICPVSQFCKAKKTLLPFGYPIKINK-----  212
c_000001286181_5      137  VKRVVSRYLNVENNPR--------ELKKSIKPF----LYQ-NSP---------KANYQSFGQGMMDLGSLICVKEKPKCD----LCPIQKTCLSFKKQDFIKTKKNT-------  217
c_000001293628_3      143  VRRVLSRLFDLADP----------TDASLRTR----AGE-LVD---------PDRPGDWNQALMELGATVCTARSPRCS----VCPLTKQCRAETEGTQELRPAPKRR-----  222
c_000001535696_8      137  VYRVLSRYFGVDLPIN------TTEGVKYYKEL----AFK-VMD---------SQNIRDYNQGIMEFGAIQCAPKNPNCF----SCPLSDSCVALQENKIDTLPVKLKK-----  221
c_000001614067_2      150  VERVFSRFYAVEKPIK--------ESKIFIKNI----AEK-HLP---------DNRHGDYAQALMDLGSLICIPKSPRCK----MCPLLAICDVGGTTSAKQYPIKLPK-----  232
c_000001765289_1      116  AKRVMSRILGIKNL--------TSWNLSRINKT----LSN-IIP---------EHTPGNFNQSVMELGATLCRPRSPSCN----KCPLSFGCKA-FKTNKPDYYPKPAA-----  197
c_000001961666_1      131  VERVITRLFKIDNNK---------KVKNKIKKY----TQS-LIP---------QNKPGDFAEALMDLGSLICKRANPTCS----LCPLKLNCKGYISNQIKVSVKNKKL-----  212
c_000002018097_4      138  LKRVISRYLGIKKLTR--------RNTLRIHNQ----LNK-MLT---------NGRPGDINQALMDVGSLICKPNDVICI----KCPLINKCKGYSYGNPIMYPDKNKR-----  220
c_000002843512_36     135  VKRILCRLQKLKTP----------TDKELWKI----AYD-LVN---------KKNPFDYNQAMMDIGATLCTPKNPNCE----ACPLSDICQGQSDPALYPTKKKR-------  212
c_000002930199_1      135  VYRLISRLYNINTAIN------TNKGREEFQSI----ANN-LLP---------NKNTGLYNQAIMDFGSIQCKKYNPKCN----ICPLQKECQSAILGIVNERPVKTLS-----  219
c_000003033795_2      131  LERIGYRILGLKTKTK--------RNKKRVVEY----LEK-NQC---------ANNPGDYNEALMDLGSSLCKASITYCN----QCPLKKICKAYTSSSPILYPTPKVP-----  213
c_000003888107_2      132  LERIGFRLLGLKKRSK--------FNQNKVKKF----LEK-IQD---------KNNPGDFNQALMDLGSNICQSQSVNCR----VCPINDSCRAFLSLNPVDYPAPIKP-----  214
c_000004013286_6      157  SRRVLSRVFAIKGDLS------RQPMKHLLWDL----AKF-VLP---------KENVGDFNQAMMELGSQCCQKQYPRCM----QCPIKEWCRAFQEGLQEELPFTPIQ-----  241
c_000004852258_1      119  HSRIIARVLGVKNQTS--------RNLNRINNY----LKK-LVR---------EGNPGEINQALMDMGSNICKSNKVICV----NCPFHFSCKAFQSGHPLSYPKKILP-----  201
c_000005254087_1      145  VVRVLSRLTDFAEDVT------TTPAKKQLWQM----AAK-LVP---------ADRPGDYNEAIMELGQTLCLPTKTDCV----ACPLTAVCLARQRGTQMERPVRPPR-----  229
c_000005603677_1      142  VYRVLSRYFGQALPID------SSAGKKWFKAK----AQA-LLW---------EDDPANYNQAMMDFGALHCRPKNPSCG----DCPLAKNCQAKANKQQLELPKKGKK-----  226
c_000003283462_3      130  IKRIMARMLGVKKITN--------FNYKRIKKT----LLN-LIS---------SSRPGDFNQALMDLGSQVCRPNQAHCY----RCPLKLFCKAGNTLNPEKYPRPLSK-----  212
c_000002687221_1      144  AIRVISRMNSIDSPYP--------KSRKKIINF----LSE-HID---------TIRPGYFNQALMDLGRVICTYKSPSCY----ICPVNTYCYSYVNNTVDKFPTKIKK-----  226
c_000003347358_19     166  VKRILCRLEKLRTP-----------SDTELWDI----AYT-LVD---------KINPFDYNQAMMDIGATVCIPKKPQCD----ICPLNDICKGQEEPTLYPTKKRR-------  243
c_000002701031_2      137  VIRVLSRYFGIQTPFD------SSDGKKEFQLL----AEE-LLI---------KEKAGENNQAIMEFGALQCTPKSPNCN----VCPLQNTCFAFNNNLILELPKKAKK-----  221
c_000000582753_3      145  VYRFISRLYGVSTPIN------SGKANKEFKVI----LNA-IID---------EKNPATFNQAIMEFGSLQCTPKSPNCE----SCTFSEICFAYKNNVIKDLPVKLKA-----  229
c_000001713769_5      137  VYRVIARFYGIDVPVN------NHLGKKFYMDY----AQK-LAP---------KKSCGDYNQGIMDFGSLICKPKAPLCD----KCMLAKDICIASKMKNVNYFPVRLKK-----  221
c_000004369364_1      140  VFRVLSRYTANFSDIG------TPVGKKEFTLL----ANQ-FLD---------KKQPGRFNQALMDLGALVCKPVKPICP----DCPLHSDCLAFAQDIISSLPVKAKK-----  224
c_000006057486_30     141  VYRVLSRYFGVSTPIN------SAKGIKEFKQL----AQI-LIN---------PKNPGTHNQSIMEFGALMCKPQKPNCT----NCPLNDSCFALQNNQLIKTLPIKTNK-----  225
c_000005494072_10     137  VYRVLSRYFGIEYAIN------SSKGIKKFKTL----AQS-LLL---------NENFGLHNQAIMDFGATICTPKKAKCD----SCIFCESCSALQHNKVSELPVKTPK-----  221
c_000001742634_3      140  VYRVLARYFGIKTAIN------STNGIKEFKLL----AQQ-LLP---------NNKIGAYNQAIMDFGATICTPKKPKCD----ICIFNDSCIALQKNKITELPTKTNK-----  224
c_000002826998_2      140  VFRVLSRYFGIETPMD------TSEGKKQFTHL----ANE-LLL---------KKKPAIYNQAIMDFGAIQCTPKSPDCK----HCFLKKKCQALQENKISKLPVKSKK-----  224
c_000003159439_6      137  VYRVLARYFGIRTSTN------STKGIKEFKQL----AQE-LID---------TKDPATFNQAIMEFGAIQCKPKNPNCN----NCPLNTSCIALQKKLITILPIKDKK-----  221
c_000004887214_1      142  VYRVLARYFGIRTSIN------TSKAKKEFKEV----AFE-LID---------SKNPGLFNQAIMEFGSLQCTPKNPTCS----ICPLNDTCVALQTNQVEVLPKKQPK-----  226
c_000005037037_2      140  VIRVLSRFFGIENTFQ------TVKEKKEFQEI----ADV-LIS---------KKSAADYNQAIMDFGAQICTPKNYQCE----KCPISANCFAYKNKQQNNLPLKKKK-----  224
c_000001059964_1      132  MIRVISRMMNNQEEPR------LAKTRKRLWQL----ARE-WIP---------EGEARRFNQAMMEIGALICLPSNPKCL----LCPMNNHCRALQKGTVSLIPLKPPK-----  216
```

```
c_000002498472_1        141  VERVFARIFNIDLIPG------SPEARQLNWEK----AEE-LLP---------KGHARNFNQALMELGALICKPKKPLCT----SCPLVTHCLALKYDLVPERPVPKKS-----  225
c_000004615912_3        130  VYRVLSRFFGIKAAID------ISKNKKIFEVL----AQE-VLD---------ASQAGKYNQAIMDFGATMCKPKSPDCP----LCPLSKKCIALKEELVTTLPIKSKK-----  214
c_000004766858_2        143  IERVLCRIEDIGEPPK------STPVQKQLWSL----AAE-WIP---------HGHAREFNQGLMELGATLCTPRNPSCL----LCPANAWCVAFHTGRTDTIPAPRVR-----  227
c_000002747260_18       149  VERVMARLFQIETPLP--------GAKPELTEL----AGR-LTP---------DIRPGDHAQAVMDLGATICTPKSPACG----ICPVMTGCTARAAGAQADLPRKTPK-----  231
c_000005371561_38       149  VERVMARLFDIHEPLP--------AAKPVLKAQ----AAA-LTP---------ATRPGDHAQAVMDLGATICTPRNPACG----ICPLRAPCAARAAGTQADLPRKTPK-----  231
c_000000169465_2        145  VKRVLARYYSIGGWPG------QKKVENQLWEV----AEK-NTPTN-----SEGGRCANYTQVMMDLGAMICTRSKPKCD----ECPLQADCIAYAQGAQADYPGKKPK-----  233
c_000003254110_11       145  VKRVLARYYAISGWPG------QKKVENQLWEV----AEK-NTPTN-----PEGGRCANYTQVMMDLGAMICTRSKPKCD----ECPLQPDCIAYAQGAQTDYPGKKPK-----  233
c_000004750284_20       145  VKRVLARYYAIEGWPG------QKAVENALWDV----AEK-NTP---------QNRCANYTQVMMDLGAMVCTRSKPKCD----ECPLQRDCVAFAQGKQTDYPGKKPK-----  229
c_000000141782_15       150  VKRVLARYHAVAGWPG------ETAVSRRILWAL----AER-YTP---------DNRTADYNQAMMDLGATVCVRRRPRCG----VCPLADGCRARREGNPEAYPGSRPR-----  234
c_000002717847_8        144  VKRVLARLFLQQEF----------KEGLFWNL----SAQ-CLD---------EDDPYSYQQGIMDIGATLCTPKAPLCP----ECPLNKNCLAFKENDFQLPSSKKPI-----  223
c_000004054799_1        131  VKRVLGRYKKISFKT-------ENEKQKKLWLI----SEE-LTP---------NKESFEYTQGIMDLGATHCSKSRPRCD----LCPVSNDCDSAFKEVSNKVSLRKAI-----  214
c_000006126673_1        156  IERVIARYHAINTPLP------KAKKEIKERAT----IYT-DNL---------PERPGDYAQALMDLGATICRPKSPLCY----SCPLKEGCTAYSLGQAEEYPKRQKK-----  240
c_000000339186_3        113  VIRVLTRIFGIRKNPK------DKATREYLWSL----AET-LVA--------KAENSSALNQSLMELGATVCTQRRPDCA----ACPVARSCTARRERATEIIPAKGEP-----  198
c_000003029168_2        145  VARVVSRLFALEGDLK------DTALQKKLWEI----TED-LVP---------NKRPGDFNQAMMELGAIICSTDSPRCT----LCPLAKPCHARVTGEPWRFPSESVT-----  229
c_000001863436_1        161  VARVLARLFAVDGDPR------SGDNQRRFWRL----AGA-LID---------RGRPGDFNQSLMELGATVCTPQGPSCL----LCPVREYCQAFATGDPTRYPAVAKR-----  245
c_000002830137_4        154  TTRVLSRLLAFHGDPS------QRVGQQLLWAF----AES-LLP---------RKRVGEFNQSLMELGAIICIPRKPNCE----RCPVRMLCAAHRQGEQLSIPISRRR-----  238
c_000004612302_3        148  TARLFSRLLALSVDPS------SSEGRGRLWSL----AER-VVP---------RTRAGEFNQALIDLGATVCGV-EPRCQ----ECPVRSSCRSLEQGVVDRVPVRSPR-----  231
c_000001933926_1        134  IERVMSRLHRLETPLP--------AAKPEIAEV----AQR-LTP---------RQRAGDFAQAMDLGATVCRPRRPNCG----ACPWNASCLAHAAGIQETLPRKLPK-----  216
c_000003136334_1        146  VERVTARLHRISKPLP--------GAKKYLKEK----AAE-LTP---------SERSGDYAQAVMDLGATVCTPRKPNCP----LCPWSYICKAYLAGDAEKFPKRVPK-----  228
c_000005849454_9        143  VKRVAARWAGLELAADDLALEKAARALGEAWMT----DLP-LGK---------LSAPGELNEGLMELGATVCSPRNADCG----ACPLASGCVALADGMVDVLPLPKRA-----  233
c_000001169194_1        154  VRRVLARLFDKREP----------RARWLNRT----AAE-LVD---------KDRPGDWNQALMELGAMICSPKNPRCE----DCPVQLWCTAHEQGTQALSPAGVRR-----  233
c_000002786947_2        154  VRRVLARLFDKHEP----------GVRWLNHT----AAS-LVD---------KDRPGDWNQALMELGAMICAPKNPRCE----DCPVQRWCTAHEQGTQALSPAGIRR-----  233
c_000004278808_52       142  ISKIITVLWPN----------KGTVKV----AES-LVA--------AANSGYIWNSAMMDLASSLRAGNEIE------AELGSTYFADVEVRDQFKPKRKKPSRTAK  219
c_000000358065_2        144  IERVLKRILNLKTEKE--------TKKENLHKL----KKV-FGI---------TNRSCDYVQALMEIGALLCKPKNPYCE----NCPITKNCLSFKKKDFEI---KKKN-----  223
c_000003745941_1        133  VIRILCRLTADNTVYT-----DNTQAMKVFTPL----AEV-LLD---------TKTPGQHNEAMMEFGAMVCSKVNPTCD----ICPFNSVCLGRNHQPER-LPIKAAK-----  217
c_000004187032_4        144  IERVLKRYLHLKKDNE--------IQKNNLIKK----KSI-FGI---------SSRASDYAQALMELGAMICKPKNPECS----ACPILKSCKSYKKKDFDL---AKIR-----  223
c_000006067315_3        144  VERVLKRVFYLKNEKE--------ISKENLINK----KFF-FGK---------SNRSSDYAQAIMEIGALICKPANPLCF----ECPISKNCRSFKKNDFEI---TSKN-----  223
c_000005543774_1        114  VRRVLLRLFNLETKD-----------------------PAN-LFK---------TKRNGDLAEALMEFGAIICKPKDPLCY----KCLLKKNCLYYKSESKI-RFNKKIK-----  185
c_000006211484_1        136  VRRVLFRLDNININD-----------------------AVN-LFK---------TQRNGDLAEALMEFGAVICRPKDPRCH----ECKIKMCAYFKSKSKI-KFNRKIK-----  207
c_000002718976_3        145  VRRVFSRNLNIEENK--------INFDKFIKKN----KKK-FFI---------TKRNDDFVEALMEFGALICKPKDPNCL----TCCLNKTCKYFKSSKKI-KNIKNKM-----  226
c_000002992548_1        146  IKRVFSRLF--NANL--------EKNEKRIKRI----TKE-QLY---------TDRCGDLAEALMEFGAVICKPSNPMCN----ICNLKANCYFFKNKTSV-LVNKKRR-----  225
c_000000456751_1        110  VKRVFSRLLLINTSN--------QNIDNEIKKI----TYK-LPK---------IDRNADFAEALMEFGALVCKPQNPLCK----ICHLQKYCKFFNSENYI-SLNKKFS-----  191
c_000002655634_1        141  VKRFFSRLFLIDVSS---------QNFENEIKKI----TNK-LPK---------VNRNADLAEALMEFGAVVCKPKNPLCK----ICNLKKYCKFFNGENYT-LPNKKFF-----  222
c_000004232403_2        140  HRRVISRILGIKNLSR--------HNKNRINNY----LKQ-LVA---------IDRPGEINQAIMDIGSIICKPKEVFCD----FCPVQFSCKACVSSKPLEYPNKIYN-----  222
6u7t.after.N146.pdb     141  VMRVLSRLFLVTDDIA-----KPSTRKRFEQI----VRE-IMA---------YENPGAFNEALIELGALVCTPRRPSCL----LCPVQAYCQAFAEGVAEELPVKMKK-----  225
Consensus_aa:                l.RVhsRhh.lp.............ppph..h....s...hhs..........ppssp@sQAhM-lGthlC..ppP.C......CPl...C.t........hP.........
```

```
Conservation:                                                   5   5
sp_P17802_MUTY_ECOL     226  -----QTLPE-------RTGYF----LLLQ-----HEDEVLLAQRPPS-GLWGGLYCFP---QFADEESLRQ---------------------------WLAQRQIA------A  281
sp_P83847_MUTY_GEOS     232  -----TAVKQ-------VPLAV----AVLAD----DEGRVLIRKRDST-GLLANLWEFP---SCETDGADG----------------KEKLEQ----MVGEQYGLQ------V  294
c_000003652391_2        233  -----QPKPT-------RYGIV----YLARD----AHGAWLLERRPDK-GLLGGMLGWP---TSDWNDAPG-----------------------------DQPPFA------A  286
c_000000031207_6        224  -----RTTEQ-------QWWAA----AVVV------CGDQVLVWIGDG-ELLAGHRGTP---LARLSGSGL---------------DADKMIG----KEFKRLGLD-----GA  286
c_000000430975_4        234  -----RKKQR-------RYGVV----YWIER----RDGAISLRQRPEN-GIFGGMTEIP---GSLWKASPL--------------------SRH----SIHAQAPIP------A  293
c_000001007148_4        232  -----RDVTD-------VFELL----VVMR-----RGKRLLLIQHPDG-ERWGGMWGFP---ALSRDRALARS-------------RFDERLIE----SEVADWGCD------V  297
c_000001092848_3        244  -----AAKPS-------RRGAA----FVAVR----ADDAVLLRRRPES-GLLGGMSEVP---TSGWSAGSD---------------GEDGP---------SAASFP------A  302
c_000001151029_1        191  -----TKKRK-------RYGVV----YWIER----RDGAIVLRQRPEN-GIFGGMTEIP---GSLWQAAPL-----------------------SRH----SIHAQAPIP------A  250
c_000001797282_1        253  -----AKGAL-------RRGAA----FVAIR----SGDEAVLLRTRPPE-GLLGNMAEPP---GSAWEPDYD---------------VAAAL---------LDAPLD------A  312
c_000001803648_25       236  -----PVKPT-------RYGIA----YIARR----GDGAWLLERRPDK-GLLGGMLGWP---GAQWGDQAV-----------------------------DKPPVS------A  289
c_000002078955_4        236  -----PVKPI-------RYGIA----YIARR----TDGAWLLERRPDK-GLLGGMLGWP---GAEWGAKAV-----------------------------ENPPVS------A  289
c_000002106160_4        224  -----QSLPW-------QNVLT----GVIW-----KRDSFLIIKRAGY-KHLNNLWELP---GGRTLTKNN---------------PHPQLEE----LLKNKFKLE------V  286
c_000003607531_2        237  -----KVKAV-------RKGVV----FLALM----SDGSLFLRRRPEN-GLLGGMMEFP---STVWLERSL--------------------SPK----EICEAAPFK------T  296
c_000003872363_4        232  -----PAKPI-------RYGTA----YIARR----HDGAWLLERRPDN-GLLGGMLGWP---GSEWGEEAV-----------------------------ENPPFS------A  285
c_000004820107_1        241  -----KPKPQ-------KRGHV----FWISD----DSGNILFIRRPEN-EMLGGMLGLP---TTEWDIRP-----------------------------IKLP------Q  291
c_000005774797_1        223  -----RSKKK-------WNAFA----YVIYN----ADGNIGFVRRDKD-ALLGGMWALP---TSDWSTSPV--------------------FQP-----------PI------L  275
c_000005867021_3        236  -----PVKPT-------RYGIA----YVARR----PDGAWLLERRPDK-GLLGGMLGWP---GAEWGAQAV-----------------------------EKPPIV------A  289
c_000004546210_2        229  -----KKKRT-------ESMYF----LMIVF----DGKKLFLEKRKPK-GIWGGLWSFP---QIQLADDPQI-----------------------------WCEQQFTNN------I  286
c_000004551008_5        229  -----KKKRT-------ESTYF----LMIIF----DGKKLFLEKRKPK-GIWGGLWSFP---QIQLADDPKT-----------------------------WCEQQFTNK------I  286
c_000002511148_4        224  -----KSKPT-------KKVVW----LLPQG----PSGEVLLEKRKPK-GIWGGLWAFI---EAEKKNELEL-----------------------------ELSRRFKTD------C  281
c_000000577378_2        212  -----KSKPT-------KKVVW----LLPQG----PSGEVLLEKRKPK-GIWGGLWTFI---EAEKKNELEL-----------------------------ELSRRFKTD------C  269
c_000000598175_2        235  -----KVMPE-------KQAIM----VILKN----SKQEVFMQKRPPV-GIWGSLWCFP---QFEKRAFAEE-----------------------------WLKSNYGKL------I  292
c_000000754627_3        232  -----KAVPK-------KSCHQ----LIIK-----HNDKVLMEKRPNS-GIWGGLFGFF---EFNEYSELET-----------------------------FLAQQGLK-------  286
c_000000811118_1        223  -----KEKIE-------KKIDW----ILIK-----TKNKVLLKRNKEK-GIWQNLWVLP---EKDFFSSKKI-----------------------------SKI------L  273
c_000000899687_2        227  -----ISSKK-------IEVSA----GIII-----KNKKVYIQQRVKN-GLMGGLWEFP---GGKREQGES---------------PEECLKR----EIKEELRVN------V  289
c_000001176522_24       248  -----KVMPV-------RKVCW----QVSLS-----QGCVWLVQNPAR-GLWGGLWVLP---EGNR-----------------------------------------  288
c_000001345122_1        202  -----KIIPT-------KEKFF----LILVT----EAFQVALVKRPEK-GIWGGLWSLP---EFDSVQEIDL----------------------------FLDGNEMVAT----LK  261
c_000001515736_6        220  -----IKRSS-------ESWWL----LLAVD----ARRRVWLQKRPQA-GIWAGLYSPP---VFEDYAALQS----------------------------YAQAAWSQD-----GT  278
c_000001682161_6        223  -----KTIPT-------KERWF----LVLLT----KEGNIGLIKRPAK-GIWGGLWSLP---EFDEVKLLEN----------------------------FITDNNVKCL----LK  282
c_000001923643_29       230  -----LKRSS-------EAWWL----LIAVD----GQGRVWLERRPQQ-GIWAGLYAPP---VFNEREALEQ----------------------------AAARHWPAA-----RS  288
c_000002030994_5        224  -----KEMPE-------KHTVM----LILQN----DQQEVLMQKRPPV-GIWGGLWCFP---QFDKTALAEE----------------------------WLDDSFGMS------L  281
c_000002038721_4        226  -----SQKKE-------KNINW----LLIA-----SKDKVLLKRNVNK-GIWQNLWLFP---EANSLNYEK-----------------------------FNL------L  275
c_000002523527_15       251  -----AAIPE-------RSTVM----LIVRH-----GRDVLVQLRPSS-GIWGGLWSLP---EMPVDTVPFD--------------SEAAEESA----LELARAYGS------P  315
c_000002608528_2        224  -----KVMPE-------KQTVM----LILKN----GKQEVFMQKRPPA-GIWGGLWCFP---QFDNHENVLE----------------------------WLDNNYAVS------F  281
c_000002961510_2        227  -----VPSKK-------IEVSA----GIII-----KNKKIYIQQRSKN-ALMGGLWEFP---GGKREKGES---------------PEECLKR----EIKEELRVN------V  289
c_000002979152_8        232  -----KDKRI-------KQSVM----LYVSN----ENADVYLERRSST-GIWGGLWSLP---EFETNVKAND----------------------------WLVKTFALE------S  289
c_000003248034_1        247  -----RKKKR-------VMTSV----IACE-----QGGQILMKQRSVNARLMPGFWELP---EVEGPFLNQ----------------ADL-------------ARWS------L  301
c_000003294679_3        229  -----KTNPL-------KTTNM----LFLRF-----EDRIYLERRPSS-GVWGGLWSFP---EISKDCSVSD----------------------------WYQSRFGAY------P  285
c_000003333364_1        225  -----KPAVP-------VDVAS----GVLV-----HDGYIFIQKRPED-GVWPGLWEFP---GGTVEKGET----------------PAQCVVR----EFQEELEWD------V  287
c_000003402697_5        220  TEKPKKRKHR-------IEVGI----ACIW-----REGKYLVQARPKG-KSFEGSWEFP---GGKREKGEN----------------FRECVKR----EIQEEVGID------V  287
c_000003492925_4        212  ------VVPT-------REENI----MLYV-----YNDKLSLTQRKG--KFLHGLWGFE---NTEMPL-----------------------------------------  252
c_000003766664_5        239  -----RPPKK-------IALVAC----VIRD-----EQGAVLLTQRPEK-GLFGGLWEVP---LKEVDETPS----------------KAGVGR----LVKETVGLT------A  301
c_000003948186_72       226  -----NNIPT-------RHTQM----LLIEN----TAGEVWLEKRPAQ-GIWGGLWSLP---ELATGQNAT-----------------------------YCEQELQIA------V  283
c_000005057120_4        225  -----KVLAV-------KQLIF----LMLQD----QQTQFLLEKRPAS-GIWGGLWSFP---EFQSFAEIKS----------------------------WCLEN-DIA------M  281
c_000005590109_5        225  -----KSLPV-------KKITF----LMCQD----QQGRILLEQRPKT-GIWGGLWSFP---EYESLANIQS----------------------------CFLEK-GIP------I  281
c_000006063368_2        224  -----KTKTT-------KHVHW----LIPYS----HKGRFLLQKRLNK-GLWEGLWVFI---EKEKEEMLKK----------------------------ECKSIFQIK------D  281
c_000000467631_2        204  -----KTLPV-------KNIMF----LMLKD----ELGLVLLEKRPAT-GIWGGLWSLP---EFETLSEIQS----------------------------FCQQN-NFA------I  260
c_000000788235_1        225  -----KTLPE-------KKAVF----LLLIN----INHEIFIIKRPPS-GIWGGLWSLP---QFDTYLQAQQ----------------------------WYEQYFSTH------L  282
c_000001187176_11       226  -----KRIPV-------RQCLM----PLLVN----PQGDIWLQRRPDS-GLWGGLWCPP---QLDSREQLDA----------------------------LVIAQ-GWQ------A  282
c_000001660697_64       231  -----KAIPE-------RAATM----VIALH-----GETVLLQRRPQR-GIWGGLWSLP---LVGEMDDALDAHP----------------------------LDADTVRQAALAYGTV  296
c_000003378864_2        238  -----KPVPH-------HHIGV----GVIWK----DDQVLIQLRPPE-GLLGGLWEFP---GGKQEPAET----------------IEETVRREIREELGVD----------V  300
c_000003910742_3        226  -----KVIPE-------KNAVM----LMLIN----NKDEVFIMKRPSP-GIWGGLWSFP---QFDDYDQAQA----------------------------WHENYFGFP------C  283
c_000004116181_2        192  -----KKQAT-------RTCVF----ILIVN----AKQEIILLQKRSPV-GIWGGLWSLP---EVEDTIEAIH----------------------------WCSSR-KIT------V  248
c_000006223903_1        226  -----RPRPR-------REVVM----IMIKD----SRDRVLLERRASS-GIWGGLWSFP---ECPVEEALEP----------------------------WCRMRFGAG------I  283
c_000001029068_2        199  -----KPKPV-------KTTAM----LIFDN----GKGGVYLQKRPAK-GIWGGLWSFV---ECSATDKAIT----------------------------KTAHNFHPQ------A  256
c_000001128125_11       234  -----KDKPV-------KRAFF----LLLLN----SERQVLLRKRPQK-GVWAGLWSLP---EFDSLQGFKE----------------------------SIESQTTV------I  290
c_000001595844_4        225  -----IKRKT-------RYFYY----FVFR-----KGSDYYIQKRTKK-DIWQSLYEFP---NIESKTPLN----------------DNEVKIV----FQEKYPRIK------F  287
c_000001719155_2        229  -----APSTK-------IAVSA----GVLF-----RRNRIYIQKRKAE-GLMGGLWEFP---GGKFESGES----------------PEQCLHR----EIKEELGVT------I  291
c_000002363038_6        234  -----KNKPV-------KQKWL----LLQKN----SAEQVFLYKRPQK-GIWAGLWSLP---EFDSYDRAVD----------------------------FLNGEIHRH-----DA  292
c_000002561400_1        190  -----KEKPV-------KQAWF----VVQLN----AKNQIRLIKRPQK-GIWAGLWSFP---EFCSVTDAKT----------------------------YLQTQTPG------L  246
c_000002566035_2        226  -----AVIPK-------RDTVF----AIMEN----NNGEILLEQRPSS-GIWGGLWCFP---EFSSPLRIEE----------------------------MIKRKYGFN------I  283
c_000002598045_3        224  -----KSKLT-------KKVLW----LLPQG----PSGEILLEKRKEE-GIWGGLWSFI---ETEKKAELEL----------------------------ALSRNFDSN------I  281
c_000004188575_41       226  -----KTIPT-------RQSHW----LILLT----DDRRVLLSKRPEK-GIWGGLWCFP---SFDDTHTL----------------------------EQHIDNT------K  279
c_000000240049_1        224  -----KILPV-------REKRL----LIIRN----KQGHYLMEKRPPS-GIWGGMWSLP---ELTLDKSIAE----------------------------SIERNWQLS------V  281
c_000002139456_2        224  -----KILPV-------REKRL----LIIRN----KQGYYLMEKRPSS-GIWGGMWSLP---ELTLDESIAE----------------------------SVERNWQLN------V  281
c_000003535614_2        226  -----KILPI-------QQKRL----LIIRN----EQGAYLMEKRPPT-GIWGGLWSLP---ELALEAPIIT----------------------------EVKNTWQII------V  283
c_000002657784_4        223  -----KQKPF-------KTVFW----LVVMN----KNGKVLLKKRNNL-GVWKGLWSFL---ESENIQDLGK----------------------------ECLTMFEKR------K  280
c_000003800129_15       221  -----TPKKK-------VFLNY----EIYK-----RDGCIYMIKNMDL-GFWNELWMPP---YKIVE-------------------------------------------  262
c_000003716781_1        216  -----PTPID-------VELEM----FLVH-----DHGRVLLRRRDEG-GRMAGLWELP---TREVAAAELRLF----------PAELELEL-------------------  272
c_000001556689_3        216  -----LVKA--------VTEYA----AVIE-----DNGRLLLFRGQRP-SLVSDMWEFP---TLDSRLADTSSRSLTEKPASSHSRAQEAKLSR----YIKEQLGWS------V  293
c_000001834452_1        196  -----INWKE-------VHLLY----GVAS-----CPSGVLLEERKS--GWNQGLWEPP---SVPYDQKEE-----------------PDL----AWREFQPQR------G  253
c_000004255004_2        226  -------TKA-------VQWPL----TLAK-----WRSRILLHRRPDK-GILASLWELP---TPENLPAEL---------------IDETL--------------------  274
c_000000605438_3        219  -----IKKPI-------RKRAL----IIYA-----KNDKYALAQNEE--RLLSGLWGFK---QEEDFEF-----------------------------------------I  262
c_000000581237_15       219  -----KKKPI-------RKRTL----IIYT-----KDDKYALSQNSE--RLLSGLWGFK---QEEEFEF-----------------------------------------S  262
```

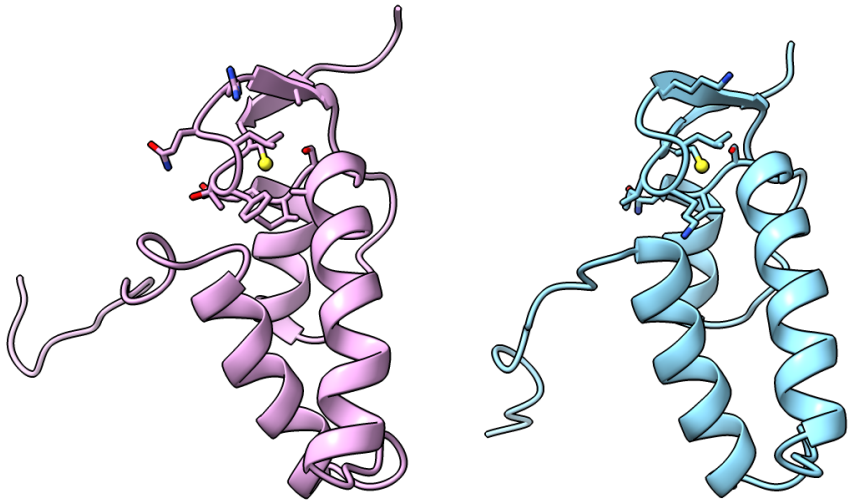
```
c_000002529579_2          219  -----IKKPI-------RKRTL----VIYH-----KANKYALAQNEE--RLLSGLWGFK---QEEDFEL----------------------------------------L  262
c_000006097838_1          208  -----APVSK-------RRFIG----FIIA-----WHGNVFVQRRPSG-EVNSGFWEFP---NWLVDDTQ------------------VQPAD----LASTKLALG------Q  267
c_000000583727_6          234  -----AATTN-------RRFIG----FLIA-----WRGKVFVQRRPSG-EVNAGFWEFP---NWQVNDAE------------------LKPAA----LAKRELKLG------Q  293
c_000005807640_2          236  -----AATTK-------RRFIG----FLIT-----WHGKVFVQRRPSG-EVNAGFWEFP---NWEV-DAA------------------ATADQ----LAKRELRLG------Q  294
c_000000134878_1          207  -----ETKPH-------YDVAV----GIIW-----DKKKLLITKRKEE-GLLGGLWEFP---GGKIKKNEK---------------IESAIKR----EIKEELSIN------I  269
c_000005136725_3          222  -----HTKPH-------YDVAV----GIIW-----DKQKLLITKRKEE-GLLGGLWEFP---GGKMKKNET---------------IKSAIKR----EIKEELSIN------I  284
c_000001684786_4          222  -----RKTIA-------VELEA----GIIR-----RGKRCLLERSESF-DFLEGLWLFP---LSLPGD--------------------GGIAA----RLGQALETD------I  279
c_000002971826_23         232  -----RKTVA-------VELEA----GIIR-----RGKRYLLERNEDL-DYLEGLWSFP---LARPDS--------------------GGIAA----RLRKRLATS------I  289
c_000002040695_2          233  -----TKKPC-------RWGHF----FFLVHIPSSGSPVSVLLEKEHG-PLLEGLWRPL---TQSWVHKKT---------------EAH--------------FPY------K  290
c_000003263657_32         239  -----KKERP------HHDHVA----GVIK-----LDDKFLIIQRPLD-GLLGGLWEFP---GGRRRESES---------------LTEGVTR----TILESTGLT-----V  302
c_000005989041_2          237  -----KVKPV-------RKGIV----FLAIM----SDGSLFLRRRPET-GLLGGMMEFP---STNWLERRI--------------------SPE----EICEAAPFK------S  296
c_000002391082_2          234  -----KKQKP------HYQVAC----GVIC-----KGDRLLLAQRPSA-GMLGGLWEFP---GGKQEEGET---------------LQQCLQR----EIAEELAIE------I  297
c_000003839553_2          241  -----KKKKP------HYQVAA----GVIS-----KGRQLLIAQRPAE-VMLGGLWEFP---GGKQEEGET---------------LEECLVR----EIREELGIE------I  304
c_000003996707_2          231  -----KPTRP------HYEVTA----GVIW-----KGSKLLVAQRPSD-GMLGGLWEFP---GGKREPGES---------------LQECLRR----EIREELTID------I  294
c_000002044706_1          214  -----TKVTF-------VQRVA----LFLEK----QDGSFLLVRRPAD-GLLAGMWELP---GRELGDGEAPE-------------RLASTLCG---EFGLTIRPD-------  270
c_000006007075_1          224  -----KKKID-------KYFLI----KIYK-----KNKKILLIRNTKF-NFLKNFKIFP---MEEIPKP-----------------------------------------  267
c_000002094036_4          200  -----KKIIN-------KFYLA----TLYK-----NQNQILLIKNDKF-KFLKNLLIFP---MREITYP-----------------------------------------D  244
c_000002762689_1          224  -----KKNNN-------KYYLL----KVYK-----DKNKYLLIKNTKF-NFLKNFSIFP---MQELSKP-----------------------------------------  267
c_000005516980_2          224  -----KKNIN-------KYFIL----KVYK-----KDQRYLLIKNTKF-NFLKNLRIFP---MEELFQP-----------------------------------------  267
c_000003920004_1          233  ----KKSRMK------KYTRAY----IIIN-----GSEEILVRRRASK-GMLPSMLEVP---NDKWVTEKK---------------LLVRD------SSVSLFK---------  292
c_000004008511_2          227  ----KSTKRK------KYSRAY----IFYN-----EKNEILIRKRPSK-GMLASMLEIP---NDNWVINKK---------------SLVTD------TIAIKLK---------  286
c_000004481347_1          234  ----LEQKKI------KFTRAY----IIMN-----KKNEVLVNRRKSK-GMLASMLEVP---NDEWVDVKK---------------KLKRH------DVYKDLS---------  293
c_000003787733_3          226  -----KIIPH-------KEIVA----GIIW-----QRDKFLITKRPEN-ALLGELWEFP---SAEIQFNET---------------PIGALRR----QISKDFNIY------I  288
c_000000754657_2          226  -----KVIPH-------KEIVA----GIIW-----QREKFLITKREEN-ALLGGLWEFP---GGEIASSET---------------PVGALRR----QIKEKWDID------I  288
c_000000990943_3          226  -----KVIPH------KDIVT----GIIW-----RGKKFLITKRSEN-ALLRGLWELP---GGEFESNET---------------PIEALRR----KIKEECDID------I  288
c_000004474996_2          225  -----KVIPH-------KEFVA----GMIC-----QGNKFLITKRNEN-ALLGGLWEFP---GWEIVSSET---------------PIGALRR----KIKEEWAID------I  287
c_000001463500_11         213  -----SIKNI-------KHFIA----AIIW-----YNKEFLIMKRPVN-SMLGGLWELP---NTQIDLKIS---------------DDIQLFN----YIKNSLNFT------I  275
c_000001286181_5          218  -----VTKTQ-------ESFHW----FIYQ-----KNNKVMLCKNPDE-GIWPNLWVFP---KRELFQT-----------------------------------------K  262
c_000001293628_3          223  -----PRVRE-------ASYAV----LVTLN----PVGELLMVRRPPD-GLLGGLWEFP---AVEMEAEERNPWGGELR-VGDSLHGRCSARLE----DLGVQLSAR------R  301
c_000001535696_8          222  -----TKVRN-------RYFNY----IIPIAEADDGSRKTLLNQRKGK-DIWQNLWQFP---LIETKREID---------------LEEVKK-----RLCEVMHLAD---DPQ  291
c_000001614067_2          233  -----KEKEE-------RYGLF----FYLQN----KDGAVLFETNKSS-GLLANMDVLP---SIGWYEESN---------------RFKKSPK----FNKKKPNFL-----GM  297
c_000001765289_1          198  -----KKRKP------HYTIVA----GIIW-----RDNTFFIQRRPEK-AMLGGLWEFP---GGKVEEGES---------------LEAALKR----EIKEECGVV------P  261
c_000001961666_1          213  -----KTIRK-------GKCY----IKRL----NDSKFLFIRNPAK-GLLGSMLLFP---TYGWLHSDH---------------DDYVK----SMIEKFIKN------I  273
c_000002018097_4          221  -----KPIPT-------KEYVA----VFIS-----YNDHIFINQRSED-GLLGGLWELP---MIEVFENSN---------------KIKALKT----YAKNKLGLS------I  283
c_000002843512_36         213  ------VVPT-------REQNI----MLYF-----YNDKLSLTQRTG--KFLHGLWGFE---AIEVPL------------------------------------------  253
c_000002930199_1          220  -----RKPKI-------RYFNY----LFIK-----KNQQFLIQQRGTN-DIWKKLYELP---LIESKEKLN---------------REKIMA------HQYFKPFK------I  280
c_000003033795_2          214  -----KKIPI-------YDIAV----SVVE-----YQNKILITKRQNK-NFLPGLWEFP---GGKIEKKET---------------ATQAIIR----EVKEETNLT------V  276
c_000003888107_2          215  -----KTIPT-------IKVST----CIIR-----KQNKLLILQRPID-TMLGGLWEFP---GGKIQNNES---------------RKKAVIR----EVMEETNLA------I  277
c_000004013286_6          242  -----PKIQA-------VREVA----VIVR-----KGKRVLVVQRPSD-GRWAGMYDFP---RFECSFRSKE--------------KIRQLMVR----GVGQQTGVP------I  306
c_000004852258_1          202  -----KTIPT-------KTIAA----ALVN-----HGDNIFITKRPLK-GLLGGLWELP---NIELVNGEI---------------PEDLLKI----KFADQFGLT------I  264
c_000005254087_1          230  -----KRTPHYDVVAGIIWQNG----VSPG-----EGGRFLIAQRPLN-GLLGGLWEFP---GGKQEPDET---------------LPQALER----EIREEMDMM------I  299
c_000005603677_1          227  -----LKIKK-------RYFYY----FVWRG----TNGGYYLRKRQEK-DIWQGLYEFP---MIELDKPLK---------------TSELKT-----LWAHWELEE------L  289
c_000003283462_3          213  -----KNIPH-------YDVVV----GMIW-----KEGRFLIQKRENQ-KHLSGLWELP---GGKIENNER---------------SELALIR----EIKEECDFE------V  275
c_000002687221_1          227  -----VSRPH-------HNVAV----GLIW-----KDNRILISKRNAS-GLLGGLWEFP---GGKIRSGES---------------GSSCVVR----KTQEILNVL------V  289
c_000003347358_19         244  ------VVPT-------REENI----LVRV-----YDDKLSLTQREG--KFLHGLWGFE---SVEVPH------------------------------------------  284
c_000002701031_2          222  -----LKIRK-------RYFEF----LIIN-----ENEKILIQKREK--GIWKGLYQFP---LIENSVKKSE--------------RDIENSKS----WKKLFKHNS------I  285
c_000000582753_3          230  -----RKSTN-------RYFTY----YYIS-----DSNHLYIKKRVGK-GIWESLYEFP---LSENLIAFE---------------NEVLDK-----QIFKNSIND------F  291
c_000001713769_5          222  -----NAPKI-------IHFNY----LVLL-----DSDHMICMDRIKN-GIWKNLFQFP---MIESKKELN---------------KTQVLSN----EIFKSIAPIS---NSD  287
c_000004369364_1          225  -----IKKRT-------RYFYY----LVFK-----EKDLFLIQKRTQK-DIWQSLFEFP---NIEGMEELN---------------ESEIREA----IQKKYPNLK------Y  287
c_000006057486_30         226  -----IKIKK-------RHFNF----IVI-N----NNQHTYIHQRKDN-DIWKNLYQFP---LIESKHEIDTN------------KVIKSSIF-----QDFFKNNSY------T  291
c_000000140712_10         222  -----IKIKK-------RYFNY----LVVLS----KKNKITLNKRINK-GIWQNLFEFP---LIETVETITIK------------TLVEHEYL-----EKIINKKQL-----VD  289
c_000001742634_3          225  -----IKIKK-------RFFNY----LILIT----KDNKTIINKRINK-DIWLNLFEFP---LLETKTEINEN------------KLVKNDFF-----QKLILNDTT-----IN  292
c_000002826998_2          225  -----IKKKT-------RYFNY----LVLNF-----GKNTWVQKRTKK-DIWQNLYEFP---LLETDKIIS---------------QKELMKS----ELWKESVSDF----DF  289
c_000003159439_6          222  -----IKIKK-------RYFNY----LVIIN----SQNKIALQRRKHK-DIWKGLYEFP---LIETLTAIDIE------------QLLETEEF-----HQIINEDQA------I  288
c_000004887214_1          227  -----LKIKK-------RYFNY----LILKT----KNEETLVKQRKGK-GIWQNLYEFP---LLETEKENIYL------------ELIDHSIF-----KTEMKEENNF------D  293
c_000005037037_2          225  -----VKVKE-------RFFYF----LVLN-----IDNTTYIEQRQNN-DIWKGLFQFP---MIETQQLIN---------------KPKELEK----LITNKINVR------F  287
c_000001059964_1          217  -----KKVKI-------LYRGC----AVIL-----FQDGLILCKRNQP-GLLEGLWEFP---SVELGKGDN---------------LRKALQE----HLLEELGKP------F  279
c_000002498472_1          226  -----KKIIP-------IDMAT----GILI-----HNGMLFIQQRLPD-DVWGSLWEFP---GGRMELKET---------------PEQTVVR----EFLEETEFK------V  288
c_000004615912_3          215  -----IRKKK-------RYFHY----FISHF-----EGKILLRKRIQK-DIWQNLYEFP---MLETKDEELK--------------ISRVQKT----QLWQQLFHGNEKVKVL  284
c_000004766858_2          228  -----PEITR-------IEKLT----ALVQ-----NGDCWLVRKRPEG-GLMGGLWEFP---TWEELPSENS--------------AAAWLTQ----QLHEDYGVA------A  291
c_000002747260_18         232  -----KPKPT-------RHGIA----YLGRR----ADGAWLLERRPDK-GLLGGMLGWP---GGDWAETAT-----------------------------EAPPID------A  285
c_000005371561_38         232  -----KPKPT-------RRGIA----YLGRR----ADGAWLLERRPEK-GLLGGMLGWP---GSDWAESPP-----------------------------EASPPCP------G  286
c_000000169465_2          234  -----KALPE-------KSTYM----MVAQ-----FNSQVYLEQRPST-GLWGGLYGFI---EVSSIEEGIE---------------------------QLAKRGIS------V  289
c_000003254110_11         234  -----KALPE-------KATFM----MVAQ-----FNSQVYLEQRPST-GLWGGLYGFI---EVSSIEEGIE---------------------------QFKKRGVN------V  289
c_000004750284_20         230  -----KVTPV-------RQTYM----LIPM-----FGQQVFLEQRPQS-GIWGGLYGFL---EAQSAEEGIS---------------------------LLAQRGIE------V  285
c_000000141782_15         235  -----REKPL-------RRTRM----LLIVD-----GGRVLLERRPPS-GIWGGLWCPP---ECPPEVDVES---------------------------YCRHELGLE------V  291
c_000002717847_8          224  -------KKD-------VYLEF----SLYK-----DGPKYFLAQTEAL-GFWKKLWMPP---VKIVN------------------------------------------  263
c_000004054799_1          215  ------QKPT--------VKLN----FILPH----TKNEILMHKKQAS-EYWESLWIPI---DGDEVKIK-----------------------------------------G  259
c_000006126673_1          241  -----KARPQ-------RYGYV----YWISD----DDGNILVHKRPEK-GLLGGLYGLP---TSSWETDIE------------------TVEH----PIFISEQIK------P  301
c_000000339186_3          199  -----IRYED-------KQIDV----YLIT-----RGPRYLVQQRPEG-GVNAGFWEFP---NSDSGNKYS------------------APK--------------------  247
c_000003029168_2          230  -----VPLQSREE-VGIVIWAG---KPYN-----ADSLIFIVQRPPT-GLWAGLWEFP---HDEQTTDEP---------------HNAAALR----IAEATVGPT------M  298
c_000001863436_1          246  -----ARQKP-------VAARTAVVTSGGA-----EGRRFLMVKRPAD-GLLGGLWEFP---TVEAAGDVA---------------PTAQAVS----RLLARLTGGP---VPA  315
c_000002830137_4          239  -----TQTEH-------VVEAA----VVIR-----RGSRVLLRRHEDR-ERWAGLWDFP---RFPFDPDATN--------------NGHQQLVD----KSRELTGFE------I  303
c_000004612302_3          232  -----VRTTD-------LYQLL----LVIR-----RDDRLLLIQHPER-ERWGGMWGFP---ALSRTGPLLKS-------------RIDELQIA----SEAAEWGCD------V  297
c_000001933926_1          217  -----VARPT-------RHGVV----FWATR----GDGCVLLRRRPEH-GLLGGMMEFP---STEWRECPW---------------RIRSAV--------KAAPFA------A  276
c_000003136334_1          229  -----KRKPT-------RYATV----FWLLD----GRGNVLLRRREEK-GLLGGMMEVP---STDWLEENT---------------PTKENS--------MFAAAA------V  288
c_000005849454_9          234  -----KKWVD-------LEMVF----LVHR-----VGVRVGLQKRTS--GWSPGLYEPP---SAICESFATANE----------ASSAHESTAH----SLAREHR--------A  299
c_000001169194_1          234  -----VPARE-------ISFLL---VILQT-----EEGEVMLVQRPPE-GLLAGMWAFP---EQELAKPLD---------------CAATSRD---RAIELAMSLGA---EVV  301
c_000002786947_2          234  -----APARE-------IFFLL---VILQT-----EEGEVMLVRRPDK-GLLAGMWAFP---EQELHKPLD---------------CSTTSRD---RAIGLAMTLGA---EVV  301
c_000001279808_52         220  KKTTKRRKHR-------IEVGI----ACIW-----REGKYLVQARPKG-KSFEGSWEFP---GGKREKGES---------------FRGCVKR----EIEEEVGLN------V  287
c_000000358065_2          224  -----KKIIN-------KFYLA----TLYK-----HDDQVLLIKNDKF-KFLKNLLIFP---MKEISQS-----------------------------------------C  268
c_000003745941_1          218  -----KIKQE-------TIDRA----WVEQ-----GGKLLLHRANDKS-RRLSGILELP---TLDVMGY----------------------SAT----ALMK------------  268
c_000004187032_4          224  -----KKNKE-------KYFIL----KVYK-----KNQKYLLVKNTKF-NFLKNLAIFP---MEELSKP-----------------------------------------  267
c_000006067315_3          224  -----KFNKI-------KYFEA----NIYQ-----NQNKYLLIKNSKF-KFLNNLLIFP---MKEVEKN-----------------------------------------K  268
c_000005543774_1          186  -----IESKT-------YDIFC----YLRK-----NKKQIALTKNNDL-GFLKKFNLPE---IKATSKK-----------------------------------------N  230
c_000006211484_1          208  -----IQSKN-------YDIFC----YLQK-----NKKQIALTKNNDL-GFLKKFNLPK---IKETSEK-----------------------------------------N  252
c_000002718976_3          227  -----IKSKN-------YDIFC----YI-N-----TKKQIALTKNNQI-SFLKDFNIPE---IKE-TNNVSN-----------------------------------N  272
c_000002992548_1          226  -----KKEQK-------YNVYC----YLKK-----KKKEIALTKNKHL-SFLSNFNIPE---TKIETSNNKK-----------------------------------N  273
c_000000456751_1          192  -----IKEKK-------INIYC----YLNR-----YKKEIALTKNKNL-SFLNNFKMPE---IQMVVANNNLK-----------------------------------K  240
c_000002655634_1          223  -----TKEKK-------FNVYC----YLNK-----YKREIALAKNRNL-GFLYNFKMPE---IQMEVTNNNLK-----------------------------------R  271
c_000004232403_2          223  -----KKIPT-------RNLAA----ALIE-----YGDYIISKRPIN-GQLGGLWELP---NIELKNGKS---------------PEKSLKE----SINNHYGFT------I  285
6u7t.after.N146.pdb       226  -----TAVKQ-------VPLAV----AVLAD----DEGRVLIRKRDST-GLLANLWEFP---SCETDGADG---------------KEKLEQ----MVGEQ----------  283
Consensus_aa:                  ....................h.....hh.........phhl.pp....shh.shh.hP.....p.................................p...........

Conservation:                            6
sp_P17802_MUTY_ECOL       282  DNLTQLTAFRHTFSHFH---LDIVPMWLPV----SSFTG-------------------CMDEGNA---------LWYNLAQPPSVG--LAAPVERLLQQLR--TGAPV------  350
sp_P83847_MUTY_GEOS       295  ELTEPIVSFEHAFSHLV---WQLTVFPGRL----VHGG--------------------PVEEPY---------RLAPEDELKAYA--FPVSHQRVWREYK--EWASGVRRPD-  366
c_000003652391_2          287  DWRTLRGEVRHTFTHFH---LILRIRVAEL----PDDI----------------------TPAVG---------FLLSKHAFRPSD--LPTV----------------------  338
c_000000031207_6          287  ELIGHADRFQHAITYRK---LFIHPLVYRY----QGE----------------------IPAGV---------ETIPIDGGDR----LPALHRKSIAAAR--ELLLEVFR---  352
c_000000430975_4          294  DWKPHPGRVYHSFTHFD---LELKIMTINS----EVT----------------------PEGT---------FWCRKEEVKAQA--LPSLMHKIMTHMY--KA---------  354
c_000001007148_4          298  EFQEWMDEFQHGVTRFR---LHLQPIVAMW----RSGSW-------------------EGPVTA---------EWVGIDELATRP--LSVAGRRVATRVE--RILRHNGAQA-  370
c_000001092848_3          303  PWEPC-GTVTHTFTHFT---LALSVYRARV----ADMP----------------------APAGS---------WWCAPGLLAGEA--LPTVMKKVVEAAI--SGATKARRPIG  373
c_000001151029_1          251  DWKPQPGLVYHSFTHFD---LELKI-----------------------------------------------------------------------------------  272
c_000001797282_1          313  RWKRLPGLVRHGFTHFP---LELTVFVARV----ALATP-------------------APAGT---------RFTPRSALDDEP--LPGLMRKVLAHAF--DPKPEPEKKPR  385
c_000001803648_25         290  DWQDPGAEVRHSFTHFH---LRLSLRISVV-----GNA----------------------KPTAG---------HFIADADFDPET--LPTVMRKAYRIAR--TKFDGDANG--  358
c_000002078955_4          290  DWQVLNAEVRHTFTHFH---LRLSLRVATI-----GNK----------------------PPATG---------DFISKNAFDPET--LPTVMRKAYRLAC--TQFDDN-----  355
c_000002106160_4          287  TIKNKIGEIRHRYSHFG---ITLSGFHCRL----INGSK-------------------PEVNRPV---------RWIRLNEINQYA--FPKANHKLFNLMQ--E----------  351
c_000003607531_2          297  QWVKIAGTVNHAFTHFQ---LQLQVRVAVV----NDQT----------------------QISGG---------IWVDRDRLDEFA--LPNLMKKVWAFAS--KRIQPLPRLDC  368
c_000003872363_4          286  DWQRLDAQVRHTFTHFH---LRLSLRIVVA-----KNV----------------------IPTVG---------MFIPSDKFDPDD--LPTVMRKAHALAD--AAFDHG-----  351
c_000004820107_1          292  ESEQTNLSVKHSFTHFD---LIL-------------------------------------------------------------------------------------  311
c_000005774797_1          276  SEWKGIGQIRHIFTHID---L-----------------------------------------------------------------------------------------  293
c_000005867021_3          290  DWHVLGAEVRHTFTHFH---LRLMVHVATV-----NEN----------------------SPDVG---------KFIADDCFDPAD--LPTVMRKAHKLAL--GSFVGD-----  355
```

```
c_000004546210_2      287  LWVDSWEPFKHSFSHFD---LYIHPIQIRM----DGQFM-----------------NTAYKSKT---------QSFSLNQLDSLG--LSAPVKGLVNKLY-------------  351
c_000004551008_5      287  LWVDPWEPFKHSFSHYD---LYIHPIQIRM----DGQFM-----------------NTAHTPKT---------QSFSLNQLGSLG--LSAPVKGLVNQLY-------------  351
c_000002511148_4      282  LNIKKLSKVKHSFSHYN---LEAIPYLAKL----ESNK----------------------KYKNS---------VWVDYKNVESLG--LPAPIKKTINQLT--KP---------  344
c_000000577378_2      270  LNIKKLGKVKHSFSHYN---LEAIPYLATL----GSNK----------------------KYKNS---------VWVDYKNVESLG--LPAPVKKTINQIT--KP---------  332
c_000000598175_2      293  QESIALAEIKHTFSHFH---LYIQPLIINV----ETPIE----------------LGVMEDNES---------LWYNITTEFNGG--LAAPVQTLINITK--NIEKEIENGTN  369
c_000000754627_3      287  SDLNEVAPFTHVFSHFE---LTINPHVLNI----EKAPD------------------VVNDKQL---------VWYPLDQSIEVG--LAAPTKKLVKQIT--AIV--------  353
c_000000811118_1      274  PVEEDFSSFEHHLSHKK---LLISVKVLTI----KNKDR----------------LSIDRRIF---------NWVNISDSVNMG--IPKPVKDILKKYD-------------  338
c_000000899687_2      290  VSLNKVMTIKHTYTQFR---VTLTVFNCKL----QKKQI-----------------RPDGCEQW---------KWVSLSKLKKYP--FPAANVKIVKYLT--EK---------  356
c_000001176522_24     289  --GEALSAFRHTFTHFH---LDITPYLWHT----HSDD-------------------MPQQAG---------QWLPLAQALKMN--IPAAVKRVLQEVI--EYEQNR-----  354
c_000001345122_1      262  EQAVI----KHTFSHYH---LRMKPMM-------------------------------------------------------------------------------------  281
c_000001515736_6      279  RSWQDLPGFLHVLTHKD---LHLHPVLVPV----DQAYA-----------------ALVSEADES---------CWADAAAWAELG--LPAPIRKLLDAEL--A----------  345
c_000001682161_6      283  DSVEILKQLKHTFSHYH---LMMNPVMLQI----ERE----------------------AIEEV---------SWFELENLNGVG--LPTPIKKILRESF--A----------  343
c_000001923643_29     289  ADWTELPAFLHVLTHRD---LHLHPVRVPV----HGSGP-----------------DAAMADGQG---------CWADARAWADMG--LPAPVRKLLDAQF--AG---------  356
c_000002030994_5      282  SEADKLSKIKHTFSHFH---LHIQPLIINT----ETPLK----------------MGVMEHDDS---------LWYNVTTEFNGG--LAAPVQTLLNITK--DIVKGTQNDTN  358
c_000002038721_4      276  EEKESLPLIVHNLSHQK---LSIFTKKYFL----KNKDR-----------------LPIDRRGF---------VWVNKKESVKMG--IPKPVKTILENHL-------------  340
c_000002523527_15     316  ARAYLAGELTHVFTHFR---LLIRAIRVDL----DTVAL-------------------EGDAAS---------RWLSLDDLDMLG--TPAPVRKLLEDQA--RFGLF------  383
c_000002608528_2      282  DEAIKLSEIKHTFSHFH---LHIHPLIINM----ETPLK----------------IGVMEHDD-------------------------------------------------  321
c_000002961510_2      290  ISLNKVMTIKHTYTQFR---VTLNVFNCEL----QKKQI-----------------RPDGCEQW---------KWVSLSSLKKYP--FPAANVKIVKYLT--EK---------  356
c_000002979152_8      290  PSLQAYAEIEHEFSHFH---LSAKVLRIEL----DSDKR-----------------LGVMEETNA---------LWYNMHKDNPGG--FPAPIKKLLSRNL--L----------  356
c_000003248034_1      302  KHPKKLCKFRHSITFRL---YDVSIYTAIL----EGE----------------------LPEPF---------CWVPQRKLATIP--LTTVTRKALAVLS--PNSSDRNDEQN  372
c_000003294679_3      286  LKISRWKTLQHKFSHFD---LKIHLFKAEP----DIAIK-----------------NIANNDSA---------LWYKSDSQTRIG--LAAPVKKLLETQF--NIQ--------  353
c_000003333364_1      288  AASETIAVIRHGVTSFR---VTLTAFLLRF----GEPPEGW-------------PEPVLHDATEG---------VWVSFAELDRYT--FPAGHRKLIDQMS--RDMRYAWLV--  366
c_000003492667_5      288  SVRPHFYEELHEFERTE---LLLRFHRCQI----QAGEP-----------------SPQEEQEI---------KWVSPDDFGDID--FLKTNHKALDKLK--EMRV-------  356
c_000003492925_4      253  CASEYIGEVTHAITHFK---LICKVYVYDE------------------------------LDAEQN---------SYFTSSKIQKLA--ISKVDDKILKLYL--DTIQTY-----  316
c_000003766664_5      302  RIEGQCSSVRHVLTHRD---MDIAVFALSV----KGREPS---------------PALSTYTGG--------MWIHSPEALEDLG--VGRVTRKILEAAA--THEISGQER--  378
c_000003948186_72     284  DSCTELAPRRHTFSHFH---LLITPLHIKS----SQQTT-----------------HVMEDQQQ---------LWYNMRKPASVG--LAAPVKRLLRQTS--NQ---------  350
c_000004057120_4      282  QSVQHIEKQRHTFSHFH---LDYTAVLVKT----ENRIN-----------------NVMESNQS---------VWYKSDKINALG--LPAPIKKLLQNQN--RNEV-------  350
c_000005590109_5      282  KLIKQLDEQRHTFSHYH---LDYTGIMVKI----ENPIN-----------------IVMEANQS---------VWYKSTQLKFLG--LPTPIKKLLQNQY-------------  346
c_000006063368_2      282  PSLIQIKKVKHSFSHFN---LEAIPYLVKI----KEKK----------------------SIKNT---------VWVDSKNIESLG--IPSPVKKTIKEIR--VP---------  344
c_000000467631_2      261  TSKKQLDGQRHTFSHYH---LDYTPIEVKI----ENPIN-----------------NVMEANQS---------VWYKSEQINSLG--LATPIKRLLQNQ--------------  324
c_000000788235_1      283  PSATTESHFSHTFSHFR---LQAQTLLITQ----STPIK-----------------RVMEGTPS---------LWYNSNTTFMGG--FPAPIKTIINKLN--EDF--------  350
c_000001187176_11     283  AEPEALSPLRHTFSHFH---LDIQPLLVRV----DAPDR-----------------VTESGQ---------VWYNLREPSRLG--LAAPVKKLLQRAE--LNTQADERTSP  356
c_000001660697_64     297  SDVEAAGALTHTFTHFR---LHMHLLHADI----TKP-------------------ATLDDDW---------RWVPLAQLNSVG--LPAPVKLALETLV--QPSLL------  363
c_000003378864_2      301  EVGPLVTQVRHAYSHFKITLHAYHCRLLRG-----TPH-----------------PKVAVES---------RWVPLNGLDQFA--FPKANKRVLEVLEQPPAMPEVTP---  375
c_000003910742_3      284  ENTEKETKLSHTFSHFR---LHIHALLIRQ----TPPIK-----------------RVMEGTPS---------LWYNRSTKFTGG--FPAPIEKLFTRIL--T----------  349
c_000004116181_2      249  KTSEVLPTLRHTFSHFH---LQYTLLVRT----DNPLN-----------------FVMEVDQA---------VWYKAEQIPGLG--LPAPVKRLLDNLI--LNEDNND----  320
c_000006223903_1      284  LFESAWSVLRHSFTHFE---LAITPQPARA----PKRLT-----------------RRMQSTGL---------LWYKPGEPF-------------------------------  332
c_000001029068_2      257  QVRKKLPLINHSFTHYY---LTI-----------------------------------------------------------------------------------------  276
c_000001128125_11     291  ESLVEWNYFRHTFTHYH---LDIYPLVYRLD--EGEVSPVFTIG--------KSQNNDEKASAEM---------NWIELDRLNKGEVGVPAPVNKMLERLS--ETLDK------  374
c_000001595844_4      288  NIKNKLGPVTQLLTHQK---VIGQFFDITL----VGE----------------------KDLDI---------APISKVDLQVFP--FPKMMMMYINRIM--A----------  348
c_000001719155_2      292  HIDEKLMTVKHSYTRFR---VILHVFLCRV----LSGRV-----------------SPTHCEEW---------DWVKTEELDRYP--FPAANVKIIKSLR--KNQGRS-----  362
c_000002363038_6      293  HSLSEGEPFRHSFSHYH---LDIFPLIF-------QASTDYEIGQP---SNRVQETPAEYSVAHH---------QWFDINRLKSGEIGVPVPIRKMISSL--------------  370
c_000002561400_1      247  EGLIEWDDFRHTFSHYH---LDIYPLIFDE----RGAPKNTAKI--------ISEPGGVYDAEQG---------SWFDIQLLLDGGLAVPAPVNKLIGRLD--ELVE-------  327
c_000002566035_2      284  KQQKEYKPIKHAFTHFH---LTIRPIHIKI----RIYKD----------------KDYDTKVL---------KWIHPKTNPNLG--LPAPVLSMLIDLN--NTKRAVLNES-  358
c_000002598045_3      282  SNIKKLKKVKHNFSHFN---LEAIPYLVKV----KKGK----------------------KIKNT---------VWVNFKNVESLG--LPAPVKKTINQIT--MP---------  344
c_000004188575_41     280  TVSKQPAVFRHIFTHFK---LDIHPHLYHA----KQSAA----------------KDYNTGQA---------IWYKIEDALTLG--LPAPVKKTLTLTS--KG--------  345
c_000000240049_1      282  NSHNDLAAFRHTFSHFH---LDITPCEVDI----TQSAY-----------------AIAENDRF---------QWH--ADITQLA--LAAPISAIFRENN--SLK--------  347
c_000002139456_2      282  NSHNDLAAFRHTFSHFH---LEITPCEVDV----AQSAH-----------------AIAENDRF---------QWH--ADITQLA--LAAPISAIFRENN--SLK--------  347
c_000003535614_2      284  KKHHDLPVFRHTFSHFH---LDITPCEIEI----KSISN-----------------AIGDDSCY---------QWQ--VDISKLA--LAAPVSAIFNAK--------------  345
c_000002657784_4      281  KDLKTYPDINHSFSHLN---LNAKLFLLKL----NEDIN-----------------YNINKGKY---------VWVDKKGLKALG--IPTPVKNIFEVLK--LNDQTS-----  351
c_000003800129_15     263  ---NNNREIKHGLSHRE---LHLEFLVTSG-----------------------------TPKGG---------KWINKNEFNNLP--TPKPISDKLEQYA--KSFL-------  320
c_000003716781_1      273  LQREEVGELRHSITKHR---IRARVLRAVP----GRDE------------------------LPEGW---------AWFERSEIEALG--VTGMTSKALTLGR-----ARSR----  337
c_000001556689_3      294  RLGPKLGEIRHGMTDRR---ISCHVYEASL----SGSQTTMAGA------PTGLEKSDHNIPDS---------GWFTIEEVMTLA--LAASARKTLTTLL--DIDLTTAHNPG  380
c_000001834452_1      254  KLGEMMGSARHTITRHR---IQARVHQVEG------------------------------------W---------NGKGAVDPSSVP--LSSLGRKVLSIAG--VLGGLLLLSPD  317
c_000004255004_2      275  QQGDPVALIHHAITDMR---ITAPVYLLQC----KIPV------------------------KGPEW---------RWVSLANLHAYP--LSSLSTKALSAAR--QHLPTTRKLGG  346
c_000000605438_3      263  ENMTGLGEFKQHYTHFT---LDARVVLLKD------------------------------KKQE---------NYFTVEEIHDLA--LSGADRKALALLE--KYLQ-------  322
c_000000581237_15     263  KDMIKLGEFKQHYTHFT---LEASVVVLQN------------------------------KEQE---------RYFRINEIHALA--LSGADRKGLALIE--KHLA-------  322
c_000002529579_2      263  NNMTELGEFKQHYTHFT---LDGSVVLVTN------------------------------KKQK---------NYFTIDEIHDLA--LSGADRKALALLE--KYLQ-------  322
c_000006097838_1      268  VEPESLCIINHSITCYR---NRLEMFVIFR----TRKPL-----------------LAKLKG---------KWVPLAKLDEIP--LTTAHRKAANQFA--G----------  331
c_000000583727_6      294  AKFQPLCNINHSITRYR---NRLEMIVAEP----KRKPR-----------------LAKLKG---------EWLPLAKLAELP--LTAAHRKASDYFL--KTANSR-----  362
c_000005807640_2      295  AELEPLCTINHSITRYR---NRLEMFVIEP----ARKPR-----------------LAKLNG---------EWIPLAKLDELP--FTTSHHKAANQFC-------------  357
c_000000134878_1      270  KPETIIHQVDHHYSHFS---VTINAINCKY----NGGDI-----------------KLSGPTDF---------KWIHTSKLHSYA--F-----------------------  321
c_000005136725_3      285  KPEKLIHQVDHHYSHFS---VTINAINCYY----NGGKI-----------------KLSGPTDF---------RWVKTSRLHSYA--FPKASIKLFNAIK--SIE--------  352
c_000001684786_4      280  ETGEELKAIKHSITFRR---LTLRPRLLET----GSFTI-----------------RGKPSF---------RWALPEELGSVL-PVSSLCLKIADRLS-------------  343
c_000002971826_23     290  EEGAELEAIRHSITFRR---LTLRPRILGA----GSFST-----------------RAKPSF---------RWGRADELGNVL-PVSSLCLKIADRLS-------------  353
c_000002040695_2      291  ASWTLLGQIQHVFTHFT---LKLDLWYTIR------------------------------TVKTQG---------QWVLLNQMNDYP--LSSLTKKSLAFLV--KNLELEKEQAR  359
c_000003263657_32     303  KTDLVLPTVLHVFTHFK---ITLYGFGCTF----IKGAA-----------------KALTCHDA---------RWVSQTQLADFA--FSRAHNRLIDIMK--KEAAKGQVEMF  378
c_000005989041_2      297  QWVKITGTVTQPFTHFE---LQLQVRVAVV----ENKP------------------------KSSEG---------IWVDRNRLDDFA--LPNIMKKVLAFAL--RHNQHFSREFF  368
c_000002391082_2      298  EVGARLTSVDHAYSHFS---ITLHALEARY----LRGEP-----------------QAIGCADW---------KWIVPDELDDYA--LPRADRLIVEFLR--QDGRQLGLFDS  373
c_000003839553_2      305  EAVARLTSVDHAFSHFS---ITLHAIAARY------------------------------------------KEQE---------RYFRINEIHALA--LSGADRKGLALIE--KHLA-------  331
c_000003996707_2      295  EVGGLLAKVDHGFSHFS---ITLNLFEAMY----RGGEP-----------------EAVECDDW---------RWIDPAELDDLP--MPRADRRALEELS-EEEGQRVPVLMA  371
c_000002044706_1      271  ----LVVTSEHRFSHRH---WTTRVYRA------RVEDE---------------CSGPADASHWSEGL-----RWTCAGDLGKLA--VPEASRKTLRPCLRASRGSLPG----  353
c_000006007075_1      268  --KNLQKTLNFKMSNMN---MNIVIKIMKI----SSN-------------------------ISDS---------SWIKETKFNHYT--FPTFTKKIFNSLE--NYL--------  327
c_000002094036_4      245  SLLKSKNKLNLKMSNMN---MNICIDFVKI----KKK-----------------------PKNG---------LWIESSKLENCM--IPTFTKKIFASVK--NNL--------  306
c_000002762689_1      268  --LSSTDIVNFKMSNMN---MNIKIEHKNK----VGS----------------------LPTS---------HWIDPKKLKNYT--LPTFTKKIVRYLE--SNK--------  327
c_000005516980_2      268  --KNFNENLNFKMSNMN---MNIKIQYSKN----NIK----------------------LPSS---------YWVDQKKLNNYM--LPTFTKKVVKYLE--NNK--------  327
c_000003920004_1      293  KKFTKCKPILYSFSHFD-----------------------------------------------------------------------------------------------  309
c_000004008511_2      287  NKLKLKGFVDYSFSHFD---LETEVFFTKV----KKDY------------------------FANQ---------KWIKKNNIKISG--LPTVMKKIIEVAL-------------  346
c_000004481347_1      294  VKYKRIGKITYSFSHFD---LDIGIYKSKV----VKKK------------------------FKNF---------NWIKSNKIDTSG--MPTVMKEIVKKSL--KI---------  355
c_000003787733_3      289  NIGKKIGHVKHAYSHFR---ITLTLYQCKT----NRSVK----------------------NIYKEY---------KWIIPTEVNNYP--FPKSNHKLFK-----------  348
c_000000754657_2      289  NIKKKVGNVKHAYSHFK---ITLTLYRCQS----QGSVK----------------------SNNKEY---------RWITPAEIKYFP--FPKSNHKLFDVLN--SNGWNV-----  357
c_000000990943_3      289  NIEKEVGYVKHAYSHFK---ITQTLFQCQT----QESVK----------------------SINKEY---------RWITPIEVNNYP--FPKSNHKLFTILN--SDGWNV-----  357
c_000000447496_2      288  NIKNKVGYVKHAYSHFK---ITLTLYQCQT----QESVK----------------------IMNKEY---------RWITPSEIQYFP--FSKSNHKLFNLLN--SNGWNV-----  356
c_000001463500_11     276  QIQNKICTIKHSYSHFD---IRVSLYHCRI----NDINI----------------------KNMKNV---------CWIRPKDINNYA--FSKSNHKLFKILN--TNGWTI-----  344
c_000001286181_5      263  QNINELPSFKHSLSHKD---FHISPRIINI----PDDME----------------------TDDKKT---------IWIEKNKIAKLG--APKPVLDIVKKIL--NQNDKGLL---  333
c_000001293628_3      302  VRFSPLPEVRHTFTHLK---AVYRPMLVIG----TRCEG----------------SNPDSEH---------KWVLPDQVEDLP--LPVAQRRILDLAW--AALEGC-----  371
c_000001535696_8      292  VLRYNHKIIVHKLSHQQ---LHTHFWIVVI----SNELD------------------------------GSILLSEIENYP--VPVLIANFINTFK--NSYFWPNNSIR  359
c_000001614067_2      298  KWKILDQNLNHIFTHFK---LNCSLAIATI----NDENAL---------------VNDLSKSSY---------RFVQKTNINELA--LPSLIKKILNALK--NNEILDF----  371
c_000001765289_1      262  SIKKRIGAVDHAYSHFS---ITFHGYHCIE----NGEK------------------------------------------------------------------------  292
c_000001961666_1      274  NRILSKNVVRHEFSHFK---LELEVYYIET----RFP-----------------------QITHG---------EWLKTKQAGKQ---LPSVMKKVLDNI-------------  331
c_000002018097_4      284  DILNCFDKVKHSYSHYK---LLVTPFECEV----LNA----------------------DSRIG---------RWIRVGEIDKYA--FSKVNHKLFNAIG--IKNY-------  347
c_000002843512_36     254  CASKYIGEVTHAYTHFK---LIIQVYVYDE------------------------------LDAEKN---------SYFSSQEIQKLA--ISKVDDKILNLFL--NSIECD-----  317
c_000002930199_1      281  SNIKNKYTIEHNLSHQK---LKISFWDVDA----KNIN----------------------TQKGF---------EKITLDNISNYP--FAKPLEKYFQTQH--LIT--------  344
c_000003033795_2      277  SNPIFLGNITHKYSHFG---VNISFFISFP----KSIKS-----------------FNSSQEY---------RWIRIKNILNYP--LPKANHKMLDILK--KL----------  342
c_000003888107_2      278  KNPNYLGQIKHQYSHFK---VLISLFLADV----NDISS-----------------LKPNQNY---------KWTTRKQLDKFA--LPKANYKMLEILD--KLN--------  344
c_000004013286_6      307  QIGNKIGQIRVVTRFR---TTVDCYEAHF----SNIGKPNRSKNDISVNTLQKEQQEAGSLWHM---------QWATRKQIDSFP--MSVPARRVIRLLA--DFD--------  392
c_000004852258_1      265  KVGHKLGMINHTFTHFK---MNVLLYNCQA----KDQP----------------------LSKNTE---------KWVSFSDLDQYA--FSKANHKLFKLVE--LRND-------  330
c_000005254087_1      300  RTDRFLTSIKHAYTHFR---ITLHAFHATH----LRGEP-----------------QHIGVADH---------AWVTLADLDKYA-----------------------------  351
c_000005603677_1      290  VWALISGQEQQQLSHQK---IIARFLWPKQ----PLPQ----------------------ALSADW---------ELVDAQSWPNYP--LPRLIAQYWEKKQ--LAFF-------  355
c_000003283462_3      276  VPKKKIESIKHKYSHFS---ITLTAFHCKI----KNGAK-----------------VPSKQST---------RWITPGEIGDYP--FPRANHKLFNQLE--NFYSNG-----  345
c_000002687221_1      290  APVAFVKQIKHAYSHFS---ITVDAYSCTF----RGGRP-----------------RALGCADY---------RWIYPHETGKYA--FHSAGHKLFDKIL--ETAS-------  358
c_000003347358_19     285  CASEYIGKVTHAYTHFK---LICKVYVYYE------------------------NHPEEA---------DYFSEEEIQKIA--ISKVDEKIVNLYG--NFLKYD-----  348
c_000002701031_2      286  QILKISATIKHQLTHQT---IFARFWHIDT----KIQ-----------------------KVENS---------KYISKKELKNYP--ISRLLEKYLENIG--MI---------  347
c_000000582753_3      292  AIRKVHLKIKHVLSHQN---IYAKLIEVEV----LDVI----------------------RLENY---------IKVEKEDFESYP--IHRLMEKMMEQVG--K----------  353
c_000001713769_5      288  IILFNSSPIIHKLSHKT---IYAKFWILPV----NGSNS------------------------------------NSIKFSDVNKYP--VPRLIEKFLDKFN--YKHF-------  348
c_000004369364_1      288  KIKHKSAPSKQLLTHQK---IIGIF---------------------------------------------------------------------------------------  309
c_000006057486_30     292  ISKYNQKQIVHKLTHQH---LFTTFWIINT----NTCSK------------------------------------NTVNWQNLSNFA--LPTLIQNFVDDFK--K----------  349
c_000005494072_10     290  VSLFNPKPIIHKLSHQH---IYTSFWIVNY----NGSVS------------------------------------NTISIDELKKYP--VPVLIDNFIKDFE--M----------  347
c_000001742634_3      293  PYLFNEKTVIHKLSHQH---IYTKFWIVKL----DVNLE------------------------------------QTIPLNSIEKYP--VSTLIQNFIKEFN--F----------  350
c_000002826998_2      290  QIKKISLPFRQNLTHRK---IIATFFEIDF----MDKFS-----------------NKNNNY---------IKIERKNLSKFA--FPKIIDLYLKDNS--LNLAMK-----  358
c_000003159439_6      289  TSLYSDNIKAHKLSHQH---IFSKFWIVET----KQNIL------------------------------------NSILIEDLEKYP--VPKLIQNFIEKFD--NIKLNPATK--  354
c_000004887214_1      294  LQLLTPKVVVHKLSHQH---LYIKFWEIRV----PTFRS------------------------------------ATISWQELLKLP--FPIVIFKFIREFL--ETSMPNFDTPL  361
c_000005037037_2      288  KVKKRSKNFTQRLTHQK---LNAIFIEIDL----KTK-------------------INKQF---------INTDINNLSKFA--FP-----------------------  336
c_000001059964_1      280  QLGKEIHTLEHRYTSFR---TRVSFFTATV----HVLPE-------------------SSQKRV---------RFVKTSEVE-----------------------------  326
```

```
c_000002498472_1        289  EIVEKITNTVHHYTRYK---VTLHCFLLRL----KKGSDT---------------DPVLHAAQDF---------NWVPFKALQEYA--FPAGHRKLITYL--------------  355
c_000004615912_3        285  QTKK----YQQELTHQK---ISAYFWEIN-----LNVAP--------------------VNIYDF---------ILIDSKNLAKFA--FPKIIALYLENNS--LNL-------  345
c_000004766858_2        292  RIGPELFKMEHDYTRFH---ATLHCRLCTV----AEIPS--------------------VLPDNA---------QWLTLGEISALA--LPRVFQRLRKRLL--DEVSA------  359
c_000002747260_18       286  PWRDPGAEVRHTFTHFH---LRLALRVADL----PASA----------------------TPRTG---------QFIPHHAFRPGQ--LPTVMRKAYDLAS--ATFLGN-----  352
c_000005371561_38       287  NWRTVPGEVRHTFTHFH---LVLELRRADL----PEDC----------------------TTMRG---------QFLAPGGFRPSD--LPTAMRKAFDLAR--DG---------  349
c_000000169465_2        290  DETRTLEGFRHTFSHFH---LDITPVVAVV----NSAPS-----------------KRVAETAF---------RWFSLNEPIEVG--LAAPTTKIIQQLM--R----------  355
c_000003254110_11       290  EETKTLETFRHTFSHFH---LDITPVVAVV----NSPPT-----------------KRVADTAF---------RWFSLGEPIEVG--LAAPTTKIIKQLI--G----------  355
c_000004750284_20       286  EEIQQLDEFRHTFSHFH---LDITPLIAVV----NSTPQ-----------------KRVAENES---------RWFLLDEPIEVG--LAAPTTKIIKTLA--K----------  351
c_000000141782_15       292  TRPQAWPGLRHTFTHFQ---LDIQPVRLRL----TAAPA-----------------RLMDGDRH---------VWYNVESPDARG--LAAPVARLLQGLA--LEDTRA-----  362
c_000002717847_8        264  ---KFNAANKHALSHRH---IHFRFKQIAN----LPT------------------------GMKG---------EWFSKKDLANIA--LPKPISDKLLQDG-------------  319
c_000004054799_1        260  NQKYSKIKVNHPLSHLN---LDMKIKTFKI----DKKCD-----------------------------------------LDSN--L--------------------------  296
c_000006126673_1        302  KPYKQLPFIMHVFTHFS---LKLYLCKTSE----FICDN--------------------VPEEH---------KWVFISTLEDIG--FPSVFEKVKKII-------------  362
c_000000339186_3        248  -NGQPFASARHSITRYR---IQLNAHLIHA-----------------------------KRGKG---------MWCTSDEIRSLP--FSSAHLKLLGKLP--HRLDK------  308
c_000003029168_2        299  QKLTKLVTITYTVMHTR---TTLHVFEGML----DLETS-----------------TMQEASSR---------LWVRLDGLHKHP--FSSPQHEIIKTLD--KEAALQLAGLV  374
c_000001863436_1        316  TPRVELGEVVHHFSHIR---MTITAEHRVA----SSEQVRL--------------RSDDEVGQEA---------RWVSAVELVDLP--MSAAMRKIEGLYR--ELGSGAI----  391
c_000002830137_4        304  EPVRQLTTIQHGVTRFQ---ITLQCHLAQC----VSGRK---------------------RGPHL---------SWRMPHELEHLP--LSVTARKIGRLVA--KLAD-------  369
c_000004612302_3        298  EVEEWLQEFRHGVTRFR---LHLQPVVASW----RAGHW--------------------IGPVVA---------EWVEIDELTARP--LSVAARRVADGVQ--AGQRCGGASG-  370
c_000001933926_1        277  AWRELPGAVRHSFTHFH---LELQIMTVET----NEEVEG--------------------------------------------------------------------------  309
c_000003136334_1        289  DWQMLSGKVRHTFTHFH---LEL-------------------------------------------------------------------------------------------  308
c_000005849454_9        300  KLGAYQGQVRHTITHHR---IRAHVYAAEG-------------------------------Q---------APTHLEDPTQVP--LTGLARKVLDRIT--SA---------  354
c_000001169194_1        302  GVASALAEVQHRFTHLQ---ARYRPWVVPV----AKL----------------------LNGEGN---------VWMTPGEPVDFP--IPVAQHKVLDALA--ECGATLGSSES  373
c_000002786947_2        302  GHASALAEIQHRFTHLQ---ARYRPWVVPV----AKL----------------------LTGEGS---------MWMTPGEPVDFP--IPAAQHKVLDALA--ECRATPDSLES  373
c_000004278868_52       288  SVRSHFYEELHEFEHTE---LLLRFHRCQI----QAGEP-----------------QALENQEL---------KWVSPDEFDSVG--FLKTNHKALEKLK--SMRV-------  356
c_000000358065_2        269  SLSKPDHKLNLKMSNMN---MNISIDFSNN----KKK------------------------LKNG---------LWVEKTKLENHM--IPTFTKKIFASVK--HNL--------  330
c_000003745941_1        269  RGNTPIVTRKRGISNSQ---ITEPIWRLSD----NSEA---------------------DHAEM---------EWIDLEQIKKVT--LSGPHKRWINEIL--TDFNKQK----  336
c_000004187032_4        268  --KNFNENFNFKMSNMN---MNIKIQYLKN----LKK------------------------FPSS---------YWFDKRRLDNYM--LPTFTKKVVKYLE--KN---------  326
c_000006067315_3        269  FNTSLNKKINIKMSNMD---MKIILNKKNK----LIK------------------------IKNS---------YLVDKKNIKDLI--LPSFTKKIFNSVS--NYL--------  330
c_000005543774_1        231  KNWKFLCNYKNSISNKK---LNLNLYYKFS----SRI------------------------PSKY---------NWYSL-KKNKEF--IPTFTKKIFKQIT--HLY--------  291
c_000006211484_1        253  KNWRFLCNYKNSISNKK---LNINLYYKFT----SKI------------------------PSEY---------NWYSL-NKNNEF--IPSFTK--------------------  303
c_000002718976_3        273  QNWIFLKNYKNSISNLK---LNINLYYKFS----KKI------------------------PRSF---------NWYSL-KENKEF--IPSFTKRIFRQVS--TLF--------  333
c_000002992548_1        274  NGWIYLCNYKNNISNIK---MNINLFYKFI----KNK------------------------PKKY---------MWYPIDRRSNEF--IPSFTK--------------------  325
c_000000456751_1        241  RGWYYLCSYKNNISNIK---MNIDLFYKFT----KKI------------------------PKKF---------NWYSIDRSNQEF--IPSFTKKIFKKIA--KVYS-------  303
c_000002655634_1        272  REWNYLCNYKNNISNIK---MNINLFYKFT----KNK------------------------PKKF---------DWYSVDKSSQEF--MPSFTKKII-----------------  326
c_000004232403_2        286  EVGKNIGLIQHTFTHFK---MNITLFCCLL----KYDQ--------------------KIDSSA---------KWVRLFELDQYA--FSRGNHKLFQLYQ--KENV-------  351
6u7t.after.N146.pdb     284  ELTEPIVSFEHAFSHLV---WQLTVFPGRL----VHGG--------------------PVEEPY---------RLAPEDELKAYA--FPVSHQRVWREYK--EWAS-------  349
Consensus_aa:                ........hp+.ho+bp...l.hp.h.h..............................................@h..pph.phs..hs...p+hhp.h.............
```

S2 Table. Percent Identity Matrix.

	G. stearothermophilus MutY MutY	LCHF MutY 4	LCHF MutY 1	LCHF MutY 2	*E. coli* MutY	LCHF MutY 3
G. stearothermophilus MutY	100	37.5	31.2	31.2	38.0	33.3
LCHF MutY 4	37.5	100	34.4	33.4	33.3	33.7
LCHF MutY 1	31.2	34.4	100	65.5	37.0	34.8
LCHF MutY 2	31.2	33.4	65.5	100	39.9	34.8
E.coli MutY	38.0	33.3	37.0	39.9	100	48.3
LCHF MutY 3	33.3	33.7	34.8	34.8	48.3	100



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.

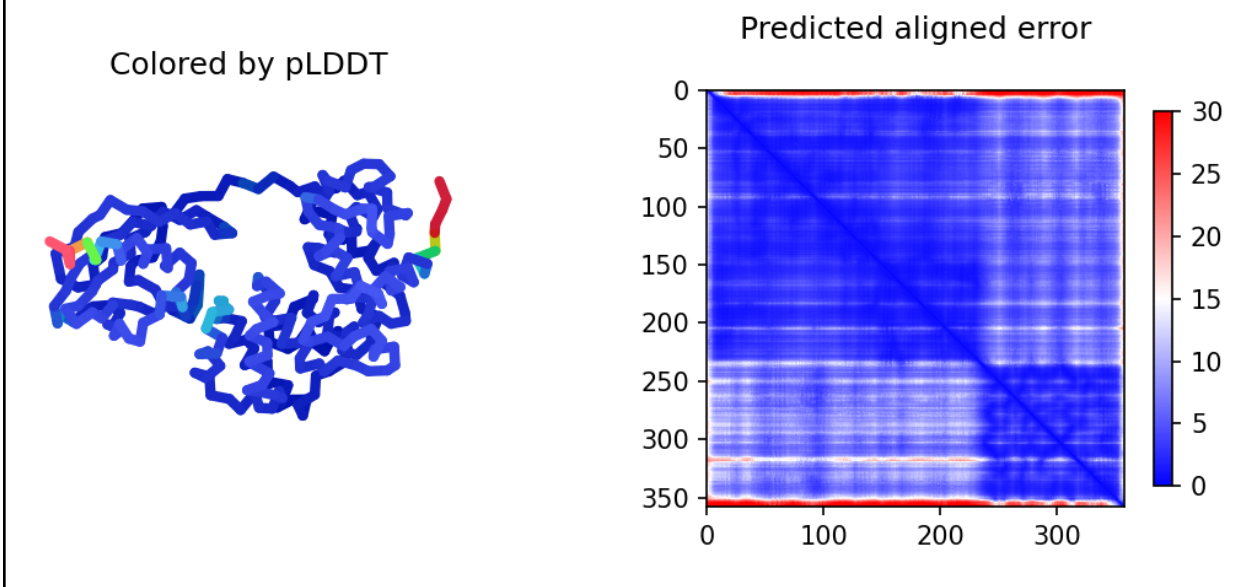
S4 Table. Metabolic gene identification

			Marinosulfonomonas[a]		*Rhodobacteraceae*	*Thiotrichaceae*	*Flavobacteriaceae*
			MAG 1	MAG 2			
Cytochrome Oxidases	UQCRFS1	K00411	X	X	X	X	
	coxA	K02274			X	X	
	ccoN	K00404	X	X	X		
	cydA	K00425		X			
	cyoB	K02298			X		
Aerobic CODH	coxS[b]	K03518					
	coxM[b]	K03519					
	coxL[b]	K03520					
Methanogenesis	mcrA[b]	K00399					
	mcrB[b]	K00401					
	mcrG[b]	K00402					
Methane Oxidation or Nitrification	pmoA[b]	K10944					
	pmoB[b]	K10945					
	pmoC[b]	K10946					
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	
	nrfA[b]	K03385					
	nirK	K00368			X		
	norB	K04561			X		
	norC	K02305			X		
	nosZ	K00376			X		
	MAG Completeness[c]		88.37	88.21	93.69	66.12	44.27
	MAG Contamination[c]		16.43	0.61	1.36	11.75	1.65

[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).

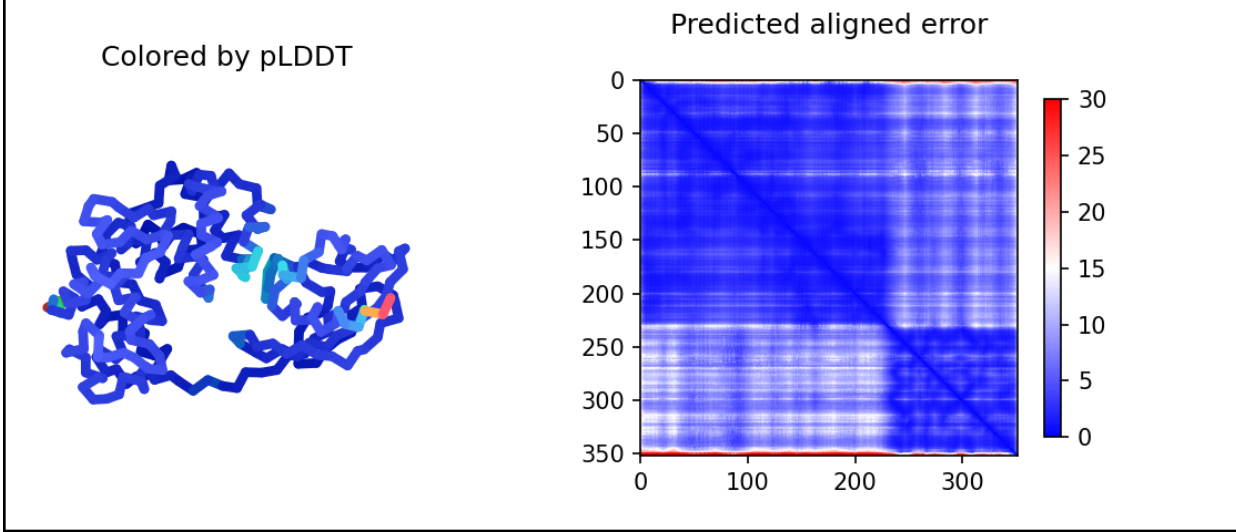
A KEGG ID analysis was used to identify the potential metabolic strategies of the MutY encoding organisms at the LCHF. The full metabolic KEGG ID search is shown above.

S5 Fig. AlphaFold Structure Prediction pLDDT Scores
pLDDT scores represent the confidence in the prediction calculated by AlphaFold2.
Marinosulfonomonas MutY Structure Prediction



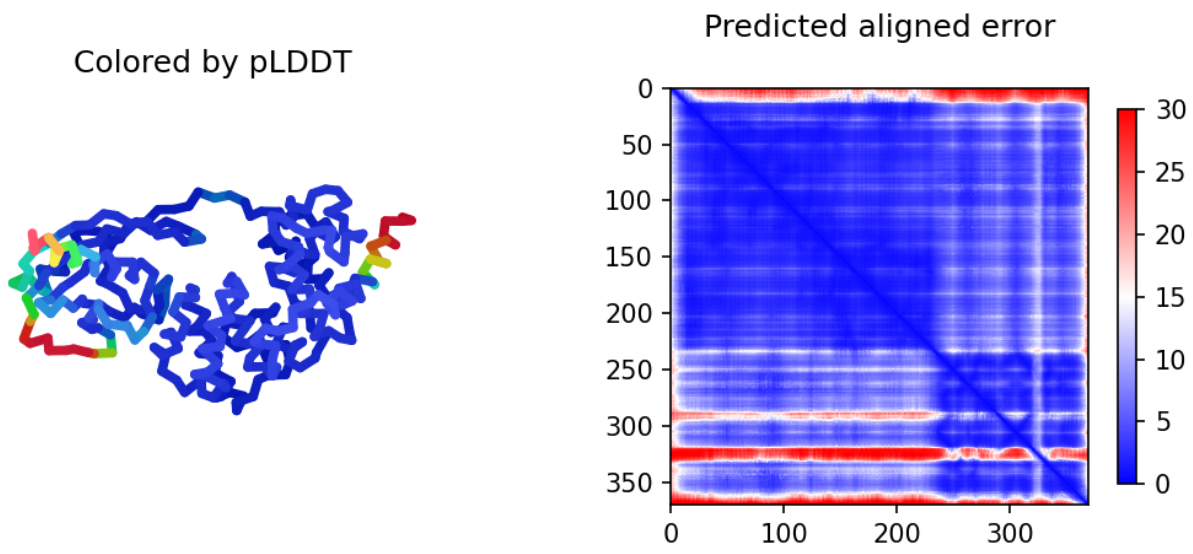
pLDDT score: 93.6
ptm score: 0.851

Rhodobacteraceae MutY Structure Prediction



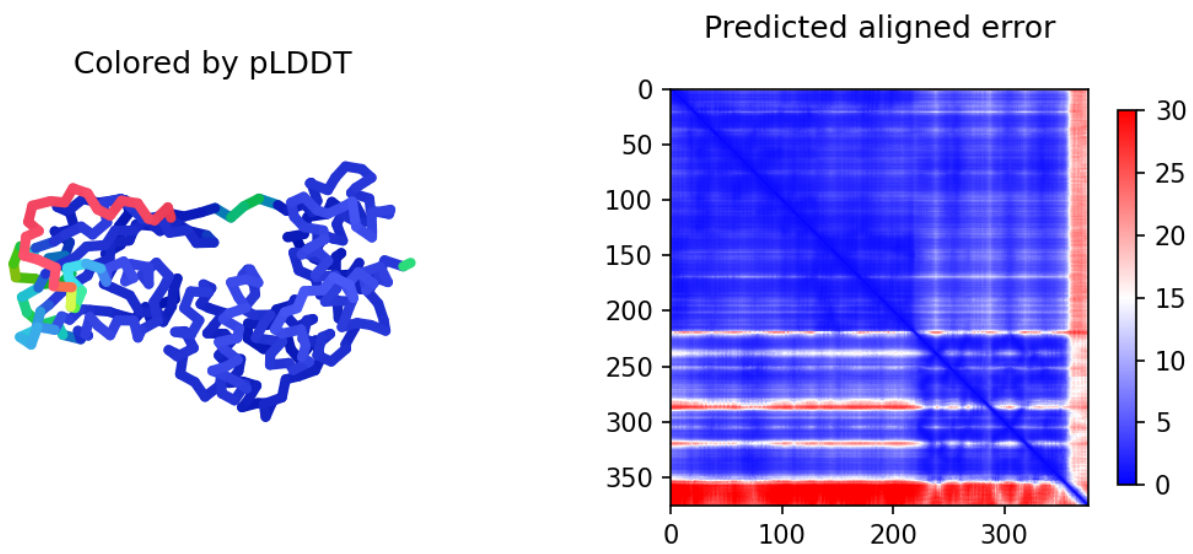
pLDDT score: 94.4
ptm score: 0.852

Thiotrichaceae MutY Structure Prediction



pLDDT score: 90.9
ptm score: 0.86

Flavobacteriaceae MutY Structure Prediction



pLDDT score: 91.3
ptm score: 0.862

S6 Table. Ligand binding affinity (kcal / mol)

Adenine	*Gs* MutY	*Ms* MutY	*Rb* MutY	*Tt* MutY	*Fb* MutY
Mode 1	-5.8	-5.1	-5.4	-5.0	-5.4
Mode 2	-5.5	-5.1	-4.7	-4.8	-5.2
Mode 3	-5.5	-4.5	-4.5	-4.7	-5.2
Mode 4	-5.2	-4.3	-4.5	-4.7	-5.0
Mode 5	-5.1	-4.2	-4.5	-4.6	-5.0
Mode 6	-5.1	-4.1	-4.5	-4.6	-5.0
Mode 7	-4.8	-4.1	-4.3	-4.5	-5.0
Mode 8	-4.7	-4.1	-4.2	-4.5	-4.8
Mode 9	-4.7	-4.1	-4.2	-4.4	-4.8

OG	*Gs* MutY	*Ms* MutY	*Rb* MutY	*Tt* MutY	*Fb* MutY
Mode 1	-5.7	-5.5	-5.3	-5.7	-5.3
Mode 2	-5.5	-5.4	-5.3	-5.5	-5.3
Mode 3	-5.4	-5.2	-5.1	-5.3	-5.1
Mode 4	-5.3	-5.2	-5.0	-5.2	-5.0
Mode 5	-5.2	-5.2	-5.0	-5.2	-5.0
Mode 6	-5.2	-4.9	-4.9	-5.2	-4.9
Mode 7	-5.0	-4.9	-4.9	-5.2	-4.9
Mode 8	-4.9	-4.8	-4.9	-5.1	-4.9
Mode 9	-4.9	-4.8	-4.9	-5.1	-4.9

*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.

S7 Table. Rifampicin resistance assay.

MutY Expressed	median (ci.low[a], ci.high[a])	Mutation Frequency[b] (per 10^8 cells)	Fold Change of EcMutY (ci.low[a], ci.high[a])	n
null	101 (81, 145)	20.26	8.78 (6.44, 13.3)	90
E. coli MutY	12 (9, 14)	6.70	1.00 (NA, NA)	79
Marinosulfonomonus MutY	7 (4, 10)	4.13	0.57 (0.27, 0.91)	30
Marinosulfonomonus MutY (catalysis-)	70 (51, 97)	30.77	6.09 (4.32, 9.30)	29
Marinosulfonomonus MutY (recognition-)	74 (59, 90)	11.53	6.39 (4.69, 8.80)	40
Rhodobacteraceae MutY	14 (8, 22)	3.48	1.22 (0.71, 2.00)	30
Rhodobacteraceae MutY (catalysis-)	61 (50, 80)	41.37	5.26 (4.00, 8.00)	38
Rhodobacteraceae MutY (recognition-)	70 (50, 82)	37.65	6.09 (4.12, 8.75)	30
Thiotrichaceae MutY	54 (42, 63)	14.77	4.70 (3.23, 6.44)	47
Thiotrichaceae MutY (catalysis-)	98 (50, 157)	28.89	8.48 (4.46, 14.28)	48
Thiotrichaceae MutY (recognition-)	72 (48, 183)	10.75	6.26 (3.77, 18.00)	20

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