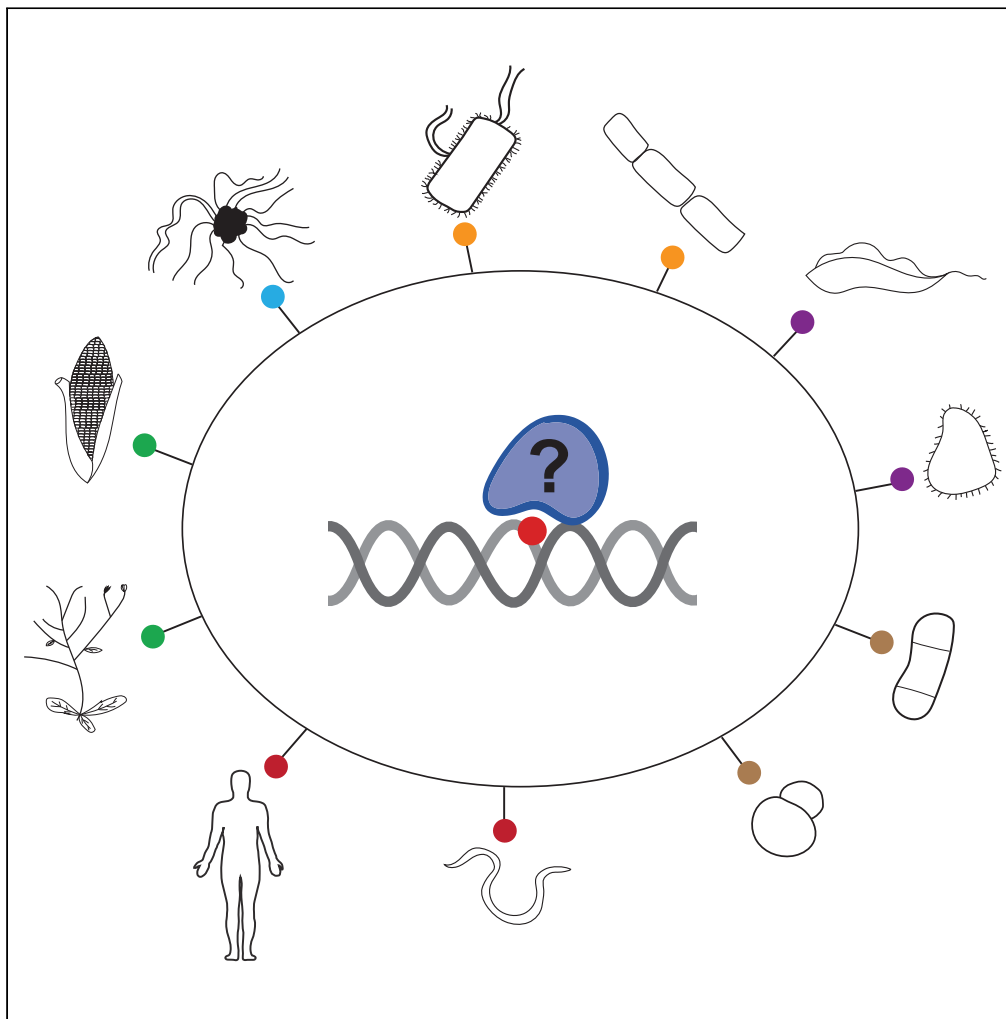


## Article

## DNA damage repair proteins across the Tree of Life



Emily Nischwitz,  
Vivien A.C.  
Schoonenberg,  
Albert Fradera-  
Sola, ..., Brian  
Luke, Falk Butter,  
Marion Scheibe

f.butter@imb.de (F.B.)  
m.scheibe@imb.de (M.S.)

**Highlights**

Phylointeractomics  
resource of 3 DNA  
damage lesions, across 11  
species

337 binding proteins  
identified with 99 known  
DNA damage repair  
factors

Linked 44 previously  
unconnected proteins to  
DNA damage repair

Resource for future study  
of crosstalk and  
evolutionary conservation

Nischwitz et al., iScience 26,  
106778  
June 16, 2023 © 2023 The  
Author(s).  
[https://doi.org/10.1016/  
j.isci.2023.106778](https://doi.org/10.1016/j.isci.2023.106778)

## Article

## DNA damage repair proteins across the Tree of Life

Emily Nischwitz,<sup>1,3</sup> Vivien A.C. Schoonenberg,<sup>1,3</sup> Albert Fradera-Sola,<sup>1</sup> Mario Dejung,<sup>1</sup> Olga Vydzhak,<sup>2</sup> Michal Levin,<sup>1</sup> Brian Luke,<sup>1,2</sup> Falk Butter,<sup>1,4,\*</sup> and Marion Scheibe<sup>1,\*</sup>

## SUMMARY

**Genome maintenance is orchestrated by a highly regulated DNA damage response with specific DNA repair pathways. Here, we investigate the phylogenetic diversity in the recognition and repair of three well-established DNA lesions, primarily repaired by base excision repair (BER) and ribonucleotide excision repair (RER): (1) 8-oxoguanine, (2) abasic site, and (3) incorporated ribonucleotide in DNA in 11 species: *Escherichia coli*, *Bacillus subtilis*, *Halobacterium salinarum*, *Trypanosoma brucei*, *Tetrahymena thermophila*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, *Homo sapiens*, *Arabidopsis thaliana*, and *Zea mays*. Using quantitative mass spectrometry, we identified 337 binding proteins across these species. Of these proteins, 99 were previously characterized to be involved in DNA repair. Through orthology, network, and domain analysis, we linked 44 previously unconnected proteins to DNA repair. Our study presents a resource for future study of the crosstalk and evolutionary conservation of DNA damage repair across all domains of life.**

## INTRODUCTION

The stability of the genome is constantly threatened by both exogenous and endogenous mutagens. These genotoxic stressors can damage the architecture of the DNA, causing single-stranded breaks, double-stranded breaks, or chemical modifications to individual bases. These alterations may prevent the successful storage of genetic information and its transmission from one generation to the next and may potentially affect cellular fitness. To maintain genome integrity, there is a carefully orchestrated DNA damage response that functions to identify and subsequently repair damaged DNA.<sup>1</sup> Base excision repair (BER) and ribonucleotide excision repair (RER) represent two pathways that are responsible for resolving some of the most frequently encountered DNA lesions.

BER is primarily responsible for removing nonhelix-distorting lesions.<sup>2</sup> Some of the most prevalent lesions removed via the BER pathway are alkylated or oxidized bases and uracil misincorporation. The most frequent oxidative base lesion is 7,8-dihydro-8-oxoguanine (8-oxoG/8-oxoGuanine), which has been reported to occur up to 1,500 times per mammalian cell per day.<sup>3</sup> There is strong conservation of the BER pathway in archaea, protozoa, fungi, metazoa, and plantae.<sup>4–8</sup> In higher eukaryotes, the repair process generally begins with damage recognition by a DNA glycosylase, which then removes the damaged base and creates an apurinic/apyrimidinic site (AP site/abasic site). Abasic sites can be formed not only as BER intermediates but also endogenously. It has been estimated that there are up to 10,000 abasic sites arising per day in a single mammalian cell.<sup>9</sup> When abasic sites are generated, a 5'-cleavage event is typically triggered by an AP endonuclease, resulting in a 3'-hydroxyl and 5'-deoxyribose phosphate. In single-nucleotide repair, the 5'-deoxyribose is removed primarily by DNA polymerase  $\beta$  and in some cases by DNA polymerase  $\gamma$ , and the resulting gap is then filled. If two or more nucleotides are repaired, the 3'-hydroxyl is used for strand displacement synthesis via either DNA polymerase  $\beta$  or  $\delta$  and  $\epsilon$ , usually in conjunction with PCNA (Proliferating Cell Nuclear Antigen).<sup>10</sup> The previously cleaved 5'-deoxyribose strand, often referred to as a 5'-flap, is removed by FEN1. In both instances, the nick is sealed with ligase I or III.<sup>7</sup>

Even more common than the generation of abasic sites is ribonucleotide misincorporation into double-stranded DNA during DNA replication. This occurs at a rate of one million sites per genome in mammalian cells, rendering it the most common endogenous DNA damage.<sup>11</sup> DNA polymerases have a highly conserved amino acid pocket that enforces sugar selectivity, referred to as a steric gate. While this steric gate helps polymerases prevent the entry of ribonucleotide triphosphates (rNTPs), there is still a large

<sup>1</sup>Institute of Molecular Biology (IMB), 55128 Mainz, Germany

<sup>2</sup>Institute of Developmental Biology and Neurobiology (IDN), Johannes-Gutenberg-University, 55128 Mainz, Germany

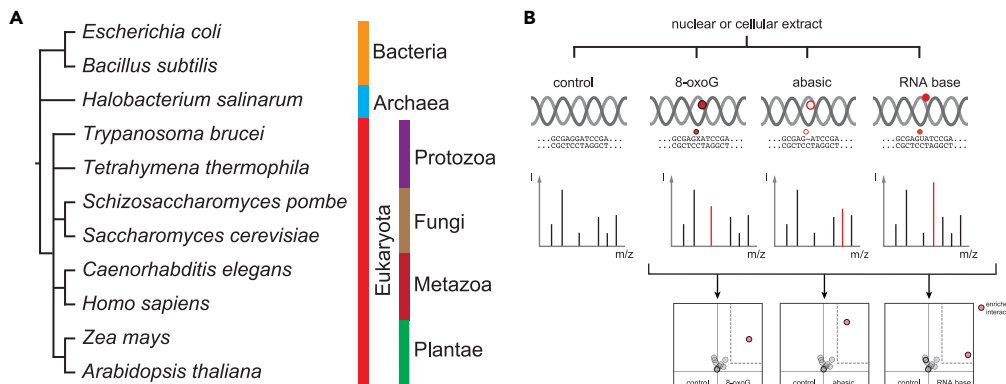
<sup>3</sup>These authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence: f.butter@imb.de (F.B.), m.scheibe@imb.de (M.S.)

<https://doi.org/10.1016/j.isci.2023.106778>





**Figure 1. Overview of screen for proteins interacting with DNA damage marks**

(A) Phylogenetic tree and overview of the eleven species included in this study.

(B) Experimental setup of the interactomics screen. Pull downs were performed for a control, and for an 8-oxoG, abasic, and RNA base lesion. Pull downs of the respective DNA damage lesion were compared to the common control to calculate enriched interaction partners passing a fold change threshold  $> 2$  with Welch t-test p values  $< 0.05$  (dashed gray line).

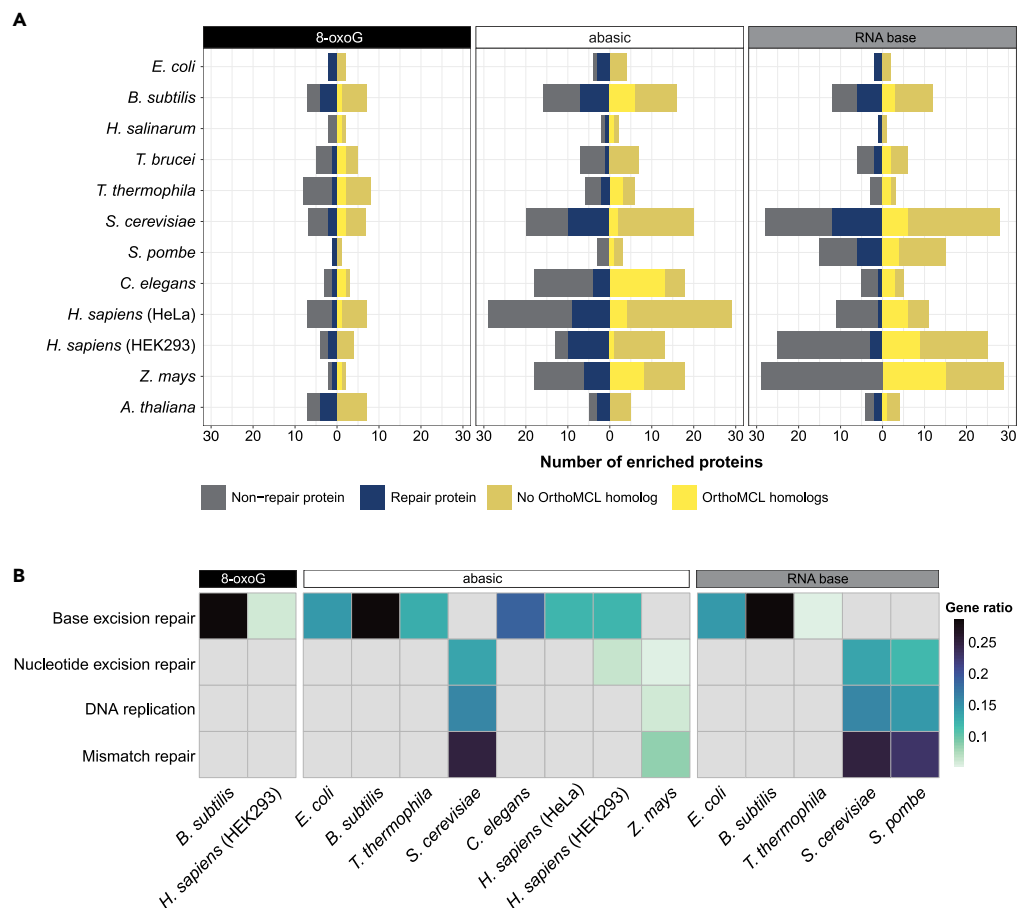
rate of ribonucleoside incorporation into DNA due to the imbalance of the nucleotide pools. For example, in *Saccharomyces cerevisiae*, there are 30- to 200-fold more ribonucleotides than nucleotides.<sup>12</sup> The *S. cerevisiae* replicative polymerases  $\alpha$ ,  $\delta$ , and  $\epsilon$  add approximately 1,900, 2,200, and 9,600 ribonucleotides per round of replication, respectively.<sup>13</sup> Across different organisms, there is a variable bias within the type of ribonucleotides incorporated into DNA. In this study, we selected rU, which in *S. cerevisiae* and *Schizosaccharomyces pombe* has comparable incorporation rates to rC and rA in nuclear genomes<sup>14</sup> but has thus far been studied less. When misincorporated ribonucleoside monophosphate (rNMP), also known as DNA-incorporated rNTPs, are integrated into DNA, they are most frequently repaired by RNase H2-mediated RER. RNase H2 recognizes the rNMP and incises at the 5'-side of the ribonucleoside, leaving a 3'-hydroxyl and 5'-phosphate. As in BER, the 3'-hydroxyl is used for strand displacement DNA synthesis via either DNA polymerase  $\delta$  supported by PCNA or by DNA polymerase  $\epsilon$ . The flap that is formed, beginning with the 5'-phosphate, is removed by FEN1 or EXO1, after which the repaired strand is ligated.<sup>15,16</sup>

Previously, we used a phylointeractomic screen to study the evolution of proteins binding telomeres across the vertebrate lineage.<sup>17</sup> Here, we revisit this concept, investigating the phylogenetic diversity in the recognition and repair of three well-established DNA lesions, primarily repaired by BER or RER: (1) 8-oxoguanine, (2) an abasic site, and (3) incorporated ribonucleotide in DNA. Previous literature has highlighted strong conservation among fundamental proteins in both of these pathways.<sup>18</sup> However, only by studying these pathways across the Tree of Life can the conservation and divergence of these different repair machinery be elucidated. Including organisms across all three domains of life, this study recapitulates previous findings and reveals new candidate proteins with the potential to be involved in DNA damage repair. We provide a large resource dataset that can be used to propel new discoveries within these specific DNA repair pathways and model organisms.

## RESULTS AND DISCUSSION

### Wide-scale identification of proteins interacting with DNA damage marks

In this study, we selected 11 species from a broad phylogenetic range encompassing all three domains of life: *Escherichia coli* and *Bacillus subtilis* (bacteria); *Halobacterium salinarum* (archaea); *Trypanosoma brucei* and *Tetrahymena thermophila* (eukaryota, protists); *S. pombe* and *S. cerevisiae* (eukaryota, fungi); *Caenorhabditis elegans* and *Homo sapiens* (eukaryota, metazoa); *Zea mays* and *Arabidopsis thaliana* (eukaryota, plantae) (Figure 1A). We used oligonucleotides that were 79 bases long with three different site-specific synthesized DNA alterations, to which a biotinylated counterstrand was annealed (Table S1). These double-stranded nucleic acid baits were immobilized on paramagnetic streptavidin beads and then incubated with protein lysates from the different species. Bound proteins were eluted from the beads and prepared for mass spectrometry measurements on a high-resolution orbitrap platform (Figure 1B). We quantified between 1,357 and 3,615 protein groups per species (Figure S1A). The replicates



**Figure 2. Interactors of the DNA damage lesions per species**

(A) Number of proteins enriched at each lesion in each species highlighted for Gene Ontology annotation “DNA repair” (GO:0006281) (blue) and presence of orthologs in OrthoMCL (yellow).

(B) KEGG term overrepresentation of enriched proteins at each lesion across species. Conditions with no enriched KEGG terms are not shown or presented in gray. Gene ratio refers to genes in the dataset (enriched proteins at lesion) over genes in the background (whole genome).

of every single experiment showed good technical reproducibility covering a similar range of protein intensities (Figure S1B). Each of the three DNA lesions, 8-oxoG, abasic, and RNA, was compared to a common nonmodified oligonucleotide with four replicates per condition to allow the calculation of an average enrichment value (fold change) and a p value for the reproducibility of the enrichment (Welch t-test) (Figure 1). Those proteins that had a log<sub>2</sub> fold change > 2 and a p value < 0.05 were considered enriched. Overall, we enriched 337 proteins across all lesions and species.

### Functional enrichment and network analysis reveal novel insights into the enriched interactors

We classified the 337 enriched proteins as either “DNA repair” or “non-DNA repair” using the Gene Ontology (GO) term GO:0006281 (Figure 2A). Of the 337 proteins, 99 were related to DNA repair, and 13 proteins were orthologs of DNA repair proteins (Figure 2A and Table 1, proteins with asterisks). Thus, our experimental conditions allowed for the identification of both known direct and indirect binders to the DNA damage lesions. Next, we used OrthoMCL to trace protein orthologies between species (Tables S2, S3, and S4).<sup>19</sup> The orthology group predictions are based on sequence similarity (reciprocal BLAST) and normalization of interspecies differences, followed by Markov clustering. In total, the OrthoMCL database contains 70,388 ortholog groups across more than 55 species.<sup>20</sup> Proteins detected in our DNA damage interactome screen across eleven species belonged to 10,329 of these groups. We identified 82 proteins that possessed no OrthoMCL orthology with the other 10 species included within

**Table 1. Overview of enriched interactors of each DNA damage lesion, per species (fold change > 2, Welch t-test p value < 0.05)**

Species	8-oxoG	abasic	RNA base
<i>E. coli</i>	mutY, phrB	fadJ, nfo, phrB, polA	nfo, polA
<i>B. subtilis</i>	exoA, mutY, nfo, ydaT, yhaZ, yisX, yxIJ	dinG <sup>a</sup> , disA, exoA, hupA, mutM, nfo, parC, parE, priA, topB <sup>a</sup> , ydaT, ydel, yfjM, yhaZ, yqxK, yxIJ	dinG <sup>a</sup> , exoA, mutM, nfo, topB <sup>a</sup> , ydcG, ydel, yfjM, yhaZ, yisX, yusI, yxIJ
<i>H. salinarum</i>	cydB, VNG_2525H	ogg, VNG_2498H	ogg
<i>T. brucei</i>	GLE2, Tb927.11.14995, Tb927.7.1290, Tb927.8.4240, Tb927.8.5510	DRBD9, GLE2, PPL2, Tb927.10.6550, Tb927.3.5150, Tb927.8.5510, TOP2	DRBD9, NST4, SET30, Tb927.2.6100, Tb927.6.1580, Tb927.8.5510
<i>T. thermophila</i>	PHR2 <sup>a</sup> , TTHERM_000530789, TTHERM_00145210, TTHERM_00147470, TTHERM_00361370, TTHERM_00463150, TTHERM_00614680, TTHERM_00852850	APN2 <sup>a</sup> , PARP4, PARP6, PCP1, PCP2, PHR2 <sup>a</sup>	PARP6, PCP1, TTHERM_00013250
<i>S. pombe</i>	myh1	sac11, SPAC3H8.08c, top2	alp5, hmo1, hpz1, kin1, mca1, mlo3, moc3, nop12, rfc1, rfc2, rfc3, rfc4, rfc5, SPAC3H8.08c, SPCC126.11c
<i>S. cerevisiae</i>	APN1, ASG1, MYO4, NUT1, PHR1, POL5, RNQ1	APN1, ASG1, CMR1, INO80, MAK5, MYO4, PDR1, PHR1, POL5, RFC1, RFC2, RFC3, RFC4, RFC5, RSC1, RSC58, RSC6, SNF2, SWI6, TOP2	APL4, APN1, ASG1, CMR1, HAP1, INO80, MBP1, MGM101, MYO4, OAF3, PDR1, POL5, RFC1, RFC2, RFC3, RFC4, RFC5, RSC1, RSC30, RSC58, RSC6, RSC9, SFH1, SNF2, STH1, SWI6, TOP2, YPL245W
<i>C. elegans</i>	col-143, exo-3, hmg-5	apn-1, col-119, col-140, col-143, dpy-17, exo-3, F07A5.2, F07H5.8, his-74, K07C5.3, obr-1, parp-2, perm-2, phat-1, phat-2, T01E8.8, Y14H12B.2, Y37D8A.19	C27D8.2, exo-3, F07A5.2, hmg-12, T01E8.8
<i>H. sapiens</i> (HeLa)	FANCI, FERMT2, KPNA6, MYL12A, NACC1, PPWD1, RTRAF	APTX, ATP5MG, BEND3, BLM, BOP1, COQ6, DNAJC13, EXOSC3, GATAD2A, HNRNPF, HNRNPH2, HPF1, ISG20L2, LIG3, MRTO4, MYL12A, NAP1L1, NIP7, NOP53, PARP1, POLB, PPIG, RIOX1, RPL21, RPLP1, RPS26, S100A8, UBE2N, XRCC1	AHCTF1, CENPV, CHD2, FXR1, KAT6A, MECP2, MPG, PCGF1, SAP130, ZMYND11, ZNF512B
<i>H. sapiens</i> (HEK293)	MAX, MUTYH, NTHL1, SEPTIN11	APTX, CMSS1, DDB1, DDB2, DNAJC13, LIG3, NOC3L, PARP2, PNKP, POLB, WRN, XPC, XRCC1	AHCTF1, APOBEC3C, BCOR, BCORL1, BRPF1, CENPV, CHD1, CHD2, CTCF, GLYR1, KAT6A, KRI1, KRR1, MPG, MSANTD7, NIP7, NOC3L, NSD2, NUP205, PCGF1, PITX2, RNF2, SUB1, TRIP12, ZNF512B

(Continued on next page)

**Table 1. Continued**

Species	8-oxoG	abasic	RNA base
<i>Z. mays</i>	B4FTT9 <sup>a</sup> , P06678	A0A1D6F6W7 <sup>a</sup> , A0A1D6JZF1 <sup>a</sup> , A0A1D6K922, A0A1D6LV91, A0A1D6NSE6, A0A1D6P5Y9, A0A804P6S3, B4FDA0, B4FER3 <sup>a</sup> , B4FJC2, B4FQT5, B4FRR3, B4FWP8, B4FX14 B6SNB5, B6U4F1, K7UTP1, K7VBU4 <sup>a</sup>	A0A1D6F4B6, A0A1D6GRJ8, A0A1D6HK01, A0A1D6HW59, A0A1D6LV91, A0A1D6LVY7, A0A1D6MYU1, A0A1D6N2N7, A0A1D6NSE6, A0A1D6QEP6, A0A804MH07, A0A804MT25, A0A804NRM4, A0A804R2N8, B4FDA0, B4FDW2, B4FRR3, B4FX14, B4G1M3, B4G1W8, B6SNB5, B6UA70, COP7N5, COP9C9, C4J4W6, C4J9R0, C4JC33, K7UTP1, Q6R9L4
<i>A. thaliana</i>	ARP, At1g09150, At4g32105, At5g16990, CRYD, PHR1, TRE1	At1g06260, At1g07080, CRYD, MOC1, PHR1	ARP, HON5, MOC1, TRE1

<sup>a</sup>Indicates orthology to known DNA damage repair factor, bold indicates previously known role in DNA damage repair, italics indicates no OrthoMCL orthology with the other 10 species included in the study.

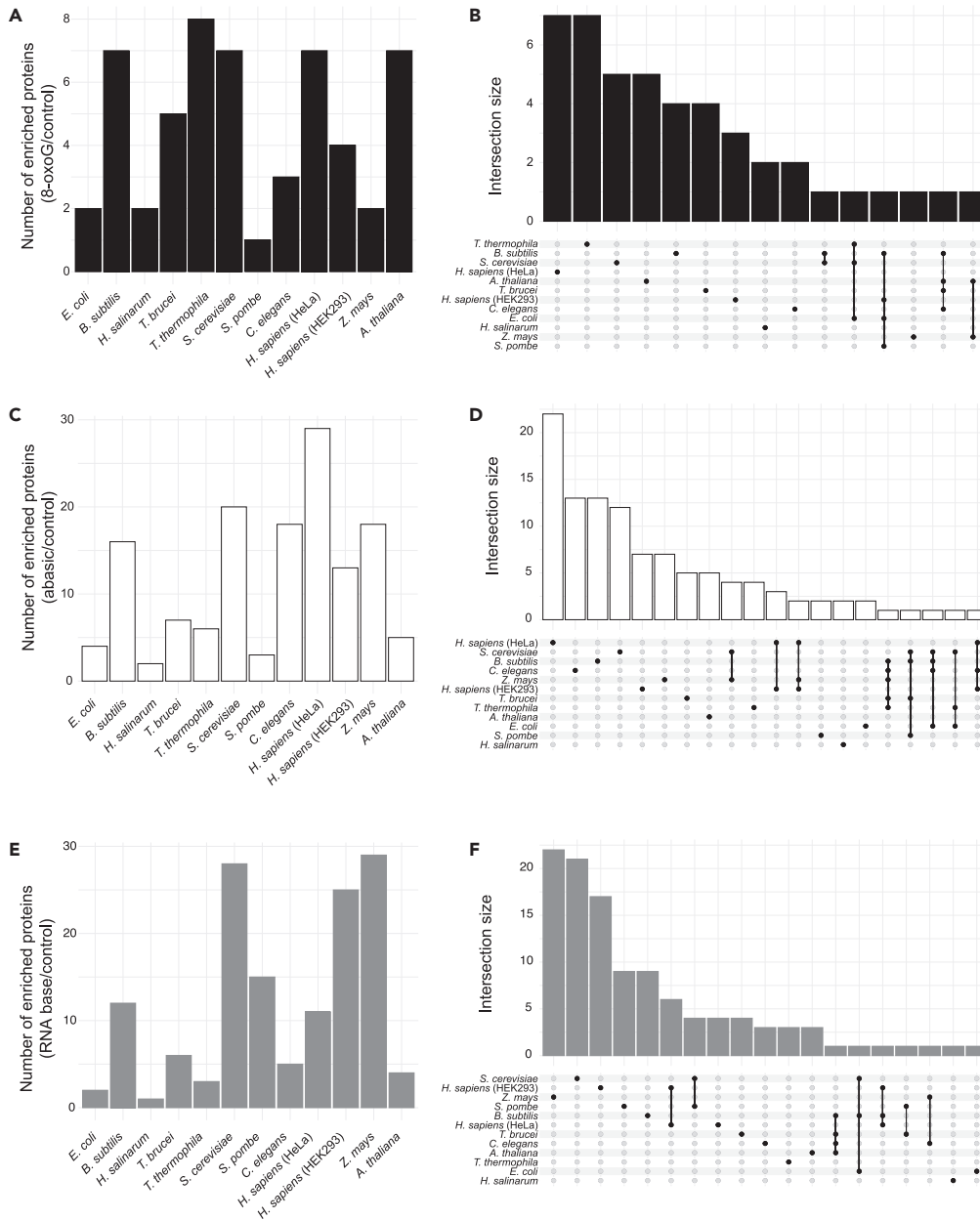
the study (Table 1, italicized protein names), four of which were repair proteins (Figure 2A). This suggests that in addition to finding conserved and previously established DNA repair factors, we also enriched species-specific DNA repair proteins.

To determine which functionalities were overrepresented, in addition to general “DNA repair”, within the interactors of 8-oxoG, abasic, and RNA lesions, we utilized both the Kyoto Encyclopedia for Genes and Genomes (KEGG) and GO (Table S5).<sup>21,22</sup> We found an overrepresentation of the KEGG term “base excision repair” for all lesions. There was additional enrichment of “nucleotide excision repair”, “mismatch repair”, and “DNA replication” (Figure 2B). Further interrogation of the enriched interactors of 8-oxoG showed enrichment of the GO biological processes “base-excision repair”, “base-excision repair ap site formation”, and “photoreactive repair” (Table S5, Figure S2). Within the interactors of the abasic lesion, there was enrichment of “DNA repair” annotated proteins in multiple species, and there were seven more terms belonging to the parent term of “DNA repair”. Four DNA repair-related GO terms (“UV damage excision repair”, “double-strand break repair”, “DNA repair”, and “base excision repair”) were overrepresented among the interactors of the RNA base lesion.

To investigate the context of our enriched proteins at each of the lesions, we created lesion- and species-specific networks using previously established interactions and proteins included in the STRING database (Tables S6, S7, and S9, Figure S3A).<sup>23</sup> We found a total of 339 interactions across our enriched proteins and species (Figure S3B). Of these enriched protein sets (3 lesions, 12 conditions, 36 total), ~61% had previously reported interactions among them. The largest number of known interactions (90) was found for the RNA lesion in *S. cerevisiae*. The 8-oxoG, abasic, and RNA-enriched proteins exhibited 7, 187, and 151 previously established interactions, respectively. This indicates the relative specificity of the 8-oxoG recognition and a more complex response resolving abasic and RNA lesions.

### Interactors of 8-oxoG, abasic, and RNA lesions across phylogenetic branches

To compare the overlap of enriched orthologs across the included species at the 8-oxoG lesion, abasic lesion, and RNA base, we again used orthology group predictions by OrthoMCL (Table S2), only counting proteins that surpassed our enrichment threshold (Figure 3, Tables S4 and S10). Within the interactors of the 8-oxoG lesion, we identified protein families that were conserved in up to four species (Figures 3A, 3B, and S4). The most conserved protein families were photolyases, MUTYH, and ExoIII-like and EndoIV-like AP endonucleases. Photolyases are critical repair proteins in bacteria, archaea, plantae, fungi, and animals. Despite their importance, they lost all DNA repair functionality in placental mammals.<sup>24</sup> The five enriched photolyases were grouped into two orthology groups (hsap\_CRY1/OG6\_100453 and atha\_PHR1/OG6\_104135). The divergence in these orthology groups indicates a specialization of the photolyases between species. It was unanticipated that photolyases would be enriched at 8-oxoG, as



**Figure 3. Interactors of the different lesions across phylogenetic branches**

- (A) Bar plot of the total number of enriched proteins at 8-oxoG across species.  
 (B) UpSet plot showing overlap of enriched proteins at the 8-oxoG lesion for the different species based on assigned orthology groups via OrthoMCL.  
 (C) Bar plot of the total number of enriched proteins at abasic lesions per species.  
 (D) UpSet plot showing overlap of enriched proteins at the abasic lesion for the different species based on assigned orthology groups via OrthoMCL.  
 (E) Bar plots of the total number of enriched proteins at the uracil RNA base per species.  
 (F) UpSet plot showing overlap of enriched proteins at the RNA base lesion for the different species based on assigned orthology groups via OrthoMCL.

typically these proteins recognize and resolve pyrimidine dimers. However, with the enrichment traversing five different species, there is a strong argument to suggest that a base conversion or lesion intermediate interacts with these photolyases and is resolved similarly across the Tree of Life. Other conserved interactors enriched at the 8-oxoG lesions were four members of the hsp\_MUTYH group (OG6\_102506). This

enrichment was specific to 8-oxoG in *B. subtilis*, *S. pombe*, and *H. sapiens*, whereas mutY in *E. coli* was also bound to the abasic lesion. Although this is a well-characterized BER glycosylase, it has thus far been shown primarily to bind 8-oxoG:A as opposed to the 8-oxoG:C used here. It is possible that the MUTYH orthologs generally bind to 8-oxoG due to their strong affinity, or they bind to a shared intermediate state of 8-oxoG:A and 8-oxoG:C.<sup>25</sup>

At the abasic lesion, we found a higher degree of overlapping proteins with seven instances of three or more orthologs enriched in two or more species (hsap\_DNAJC13, hspap\_TOP2B, scer\_PHR1, hspap\_LIG3, scer\_APN1, hspap\_APTX1, and hspap\_APEX1) (Figures 3C, 3D, and S5, Table S10). Two anticipated groups were the hspap\_APEX1 (ExoIII-like) and scer\_APN1 (EndoIV-like) AP endonucleases (OG6\_101139 and OG6\_104339, respectively), which are critical to the removal of abasic sites. Members of hspap\_LIG3 and hspap\_APTX1 are also critical to the BER pathway.<sup>2</sup> While LIG3 has been well studied in *H. sapiens*, the enriched ortholog in *C. elegans* has not been studied in the context of BER (K07C5.3, UniProt ID: Q19138). It is still unclear which ligase is involved in BER in *C. elegans*.<sup>26</sup> There were three homologs enriched in the hspap\_APTX1 group in HeLa and HEK cell lines and in *Z. mays*. APTX removes AMP from BER intermediates to form 3'-OH utilized by repair polymerases. A similar enrichment pattern was present in the hspap\_DNAJC13 group. DNAJC13 is a heat shock protein that is critical to the heat stress response and has been associated with Parkinson's disease.<sup>27,28</sup> DNAJC13 has not been studied in the context of BER.

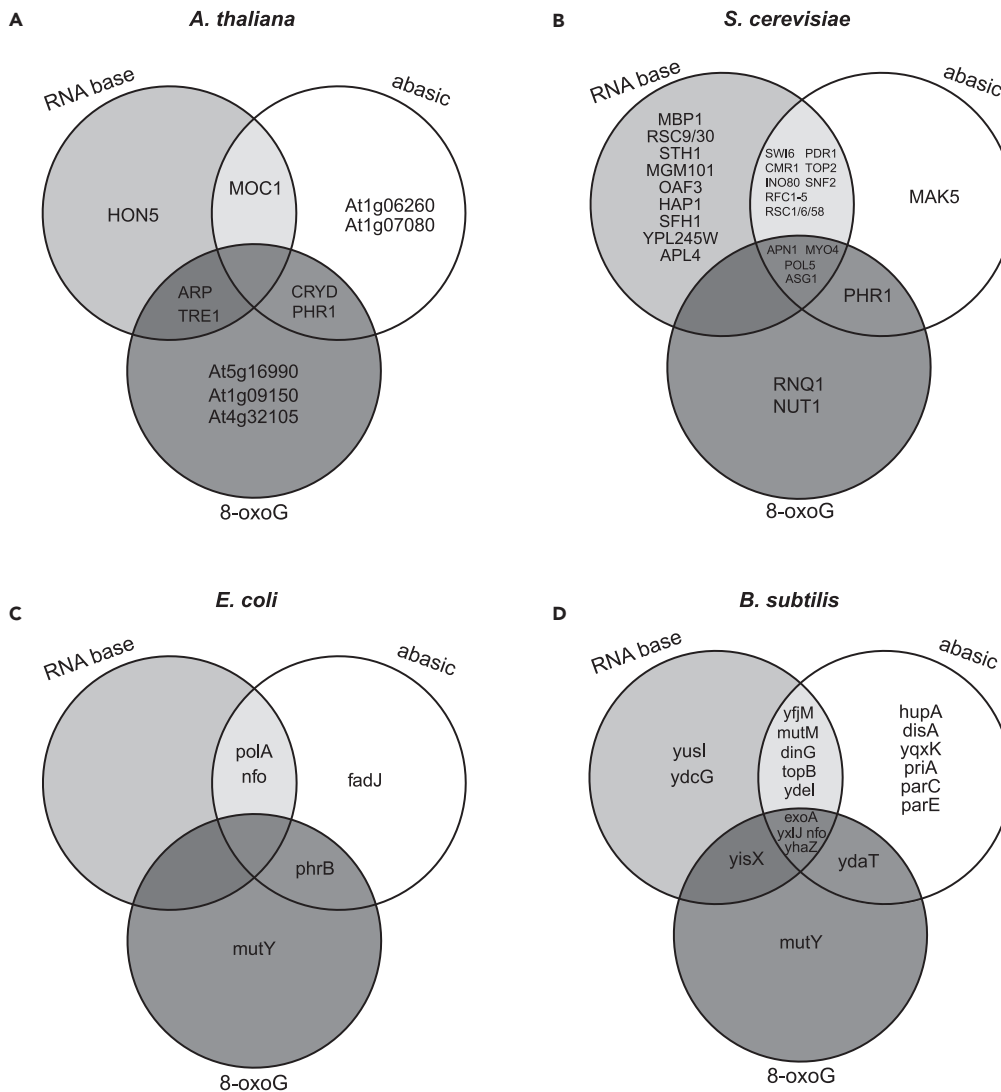
Among the enriched proteins interacting with rU across species, members of the RFC (Replication factor C) complex were enriched in both *S. cerevisiae* and *S. pombe* (Table S10). RFC is critical to the loading of PCNA, which is a well-established interactor of RNase H2, an initiator of RER. There was significant enrichment of the hspap\_APEX group in *B. subtilis*, *T. brucei*, *C. elegans*, and *A. thaliana*. Additionally, proteins of the scer\_APN group in *E. coli*, *B. subtilis*, and *S. cerevisiae* were enriched at rU. While the striking amount of enrichment of AP endonuclease was expected at the abasic and 8-oxoG lesions, this was unanticipated for the RNA lesion. There was also a noticeable enrichment of chromatin remodelers (Figures 3E, 3F, and S6). In both HeLa and HEK293 cells, PCGF1 and CHD2 were enriched. PCGF1 is part of the polycomb repressive complex 1, which is critical to epigenetic alterations repressing gene expression. In HEK293 cells, two interactors of the polycomb repressive complex were enriched, BCOR and BCORL1.<sup>29</sup> Additionally, within HEK293 cells, CHD1 and CTCF, which also mediate chromatin architecture in the presence of damage, were enriched.<sup>30,31</sup> In *S. cerevisiae*, we observed enrichment of the chromatin remodelers Ino80, Snf2, Swi6, and seven members of the remodels the structure of chromatin (RSC) family (Sfh1, Sth1, and Rsc1/6/9/30/58) (Table S4). All of the described chromatin remodelers have not yet been characterized in the misincorporated uracil from DNA but have been directly linked to the promotion of BER.<sup>32</sup>

### DNA damage interactors conserved across lesions

In this study, we observed potential DNA repair crosstalk through preferential binding of the same proteins at multiple lesions (Figure 4). We included two DNA damage lesions that are canonical substrates for BER, 8-oxoG, and abasic lesions, as well as a uracil ribonucleotide incorporated into DNA. As 8-oxoG is a common trigger for BER and abasic lesion is a common BER intermediate, we anticipated finding joint interactors between these two lesions. Of the 55 8-oxoG interactors, 19 overlapped with the abasic interactors (Table S11). Within this overlap, we unexpectedly found four instances of photolyases (Figures 4A–4C, Table S11). Additionally, in *B. subtilis*, ydaT was shared between the 8-oxoG and the abasic lesion (Figure 4D). This is an uncharacterized stress response protein that increases resistance to ethanol and low temperatures.<sup>33</sup>

There were 47 instances in which a protein was enriched both at the abasic site and rU. Such a large degree of overlap between the RNA base and abasic lesion was not initially expected. However, there has been evidence that abasic sites can occur within RNA and are primarily resolved by APE1 and MPG.<sup>34</sup> In HEK and HeLa cells, we enriched MPG and its ortholog yx1J in *B. subtilis*. Additionally, APE1 and APN1 orthologs were enriched in 6 of the 11 species. Thus, the removal of abasic sites from RNA may share mechanisms with uracil and abasic site removal when incorporated into DNA. Our data also suggest that in *S. cerevisiae*, the chromatin remodeling mechanisms that are needed to repair abasic sites are shared for the repair of rU (Ino80, Rsc1, Rsc6, Rsc58, Swi6, and Snf2) (Figure 4B), in line with chromatin state being a critical factor for the removal of both ribonucleotides and BER intermediates.<sup>16,32,35</sup> Beyond the overlaps of enriched proteins between two lesions, we also observed a notable overlap between all three lesions. In





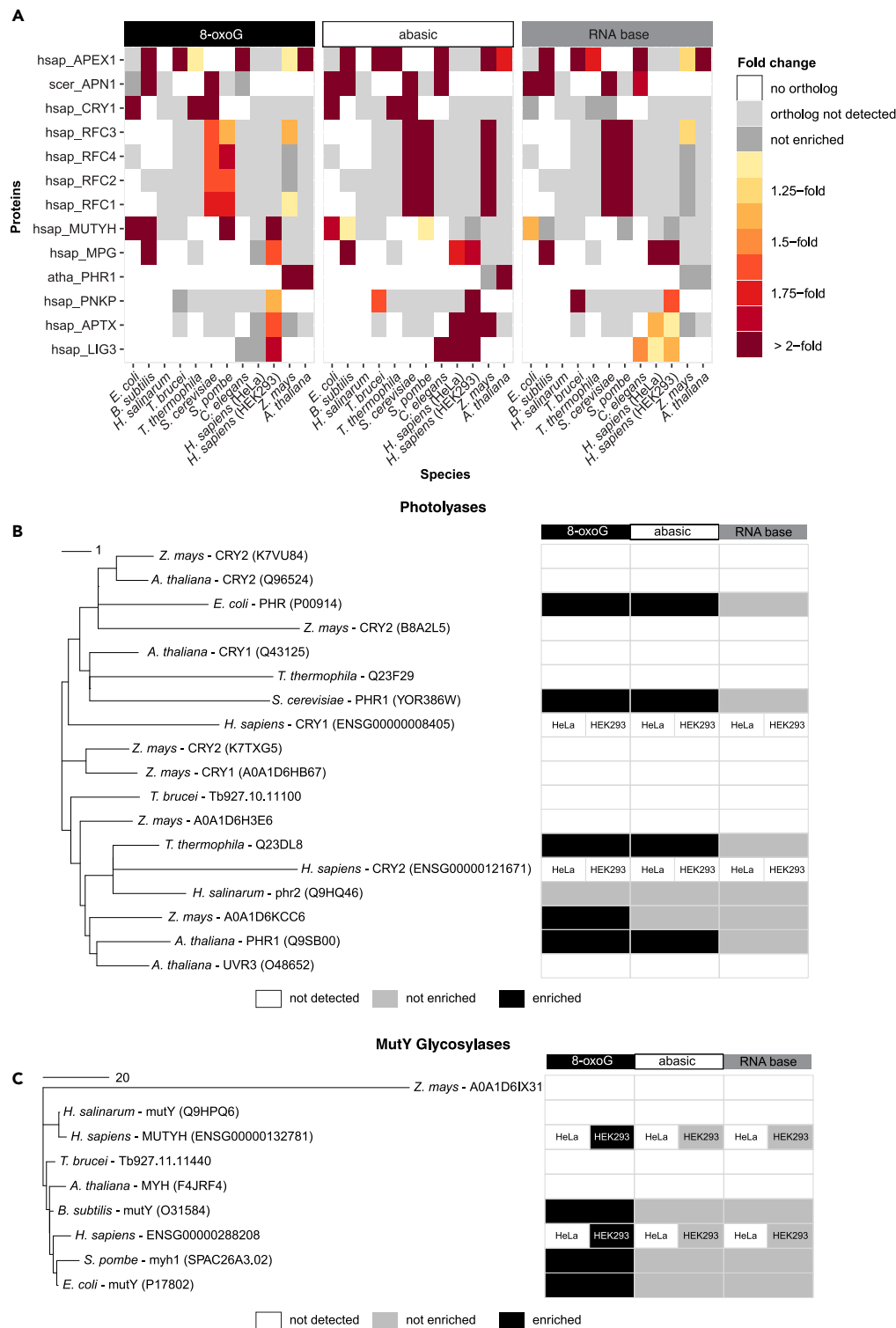
**Figure 4. Conserved interaction partners across the lesions**

Venn diagrams showing the overlapping enriched proteins at the RNA base, abasic site, and 8-oxoG lesions for (A) *A. thaliana*, (B) *S. cerevisiae*, (C) *E. coli*, and (D) *B. subtilis*. Overlap in other species is detailed in [Table S11](#).

*B. subtilis*, *T. brucei*, *S. cerevisiae*, and *C. elegans*, AP endonuclease orthologs are enriched at all three lesions. In *B. subtilis*, we observed two uncharacterized glycosylases, *yhaZ*, and *yxjJ*, at all three lesions ([Figure 4D](#)). Although ASG-1, POL5, and MYO4 are not characterized as DNA repair proteins, they were also found to interact with all three lesions in *S. cerevisiae* ([Figure 4B](#)). Taken together, our screen reiterates a broader profile for DNA repair factors in the repair of 8-oxoG, abasic, and RNA lesions and a potential crosstalk between the different repair pathways ([Figure 4](#), [Table S11](#)).

### Binding patterns by DNA repair factors are evolutionarily conserved across all domains of life

As the maintenance of genome stability is critical in each organism, many DNA damage factors are conserved in both sequence and functionality across species.<sup>18</sup> Across species and lesions, we enriched for classical BER-related proteins, including orthologs of the glycosylases MUTYH and MPG, deadenylase APTX, LIG3 and XRCC1, PCNA clamp loader RFC1-4, POLB, and the AP endonucleases APEX1 and Apn1 ([Figure 5A](#), [Table S4](#)). The APEX1/APE1 and Apn1 orthology groups represent the ExoIII-like AP exonucleases and EndoIV-like AP endonucleases, respectively. These groups of conserved AP endonucleases have been studied at length due to their evolutionary history.<sup>2,36,37</sup> Using a maximum likelihood phylogenetic



**Figure 5. Conservation of DNA repair orthologs across the Tree of Life**

(A) Heatmap representing enrichment levels of OrthoMCL orthology groups with GO annotation “DNA repair” (GO:0006281) with two or more enriched proteins across eleven species and 8-oxoG (black), abasic (white) and RNA base (gray) lesions. The color scale represents the fold change in comparison to control samples. Abbreviations: hsap, *Homo sapiens*; scer, *S. cerevisiae*; cele, *C. elegans*; atha, *A. thaliana*; spom, *S. pombe*.

**Figure 5. Continued**

(B) Maximum likelihood phylogenetic tree of the photolyase gene family including information on detection and enrichment (fold change > 2, Welch t-test p value < 0.05) for the different lesions. White boxes represent proteins that were not detected in the respective experiment. The scale bar in the plots indicates the number of amino acid substitutions per site.

(C) Maximum likelihood phylogenetic tree of the MUTY glycosylase gene family. Same as (B).

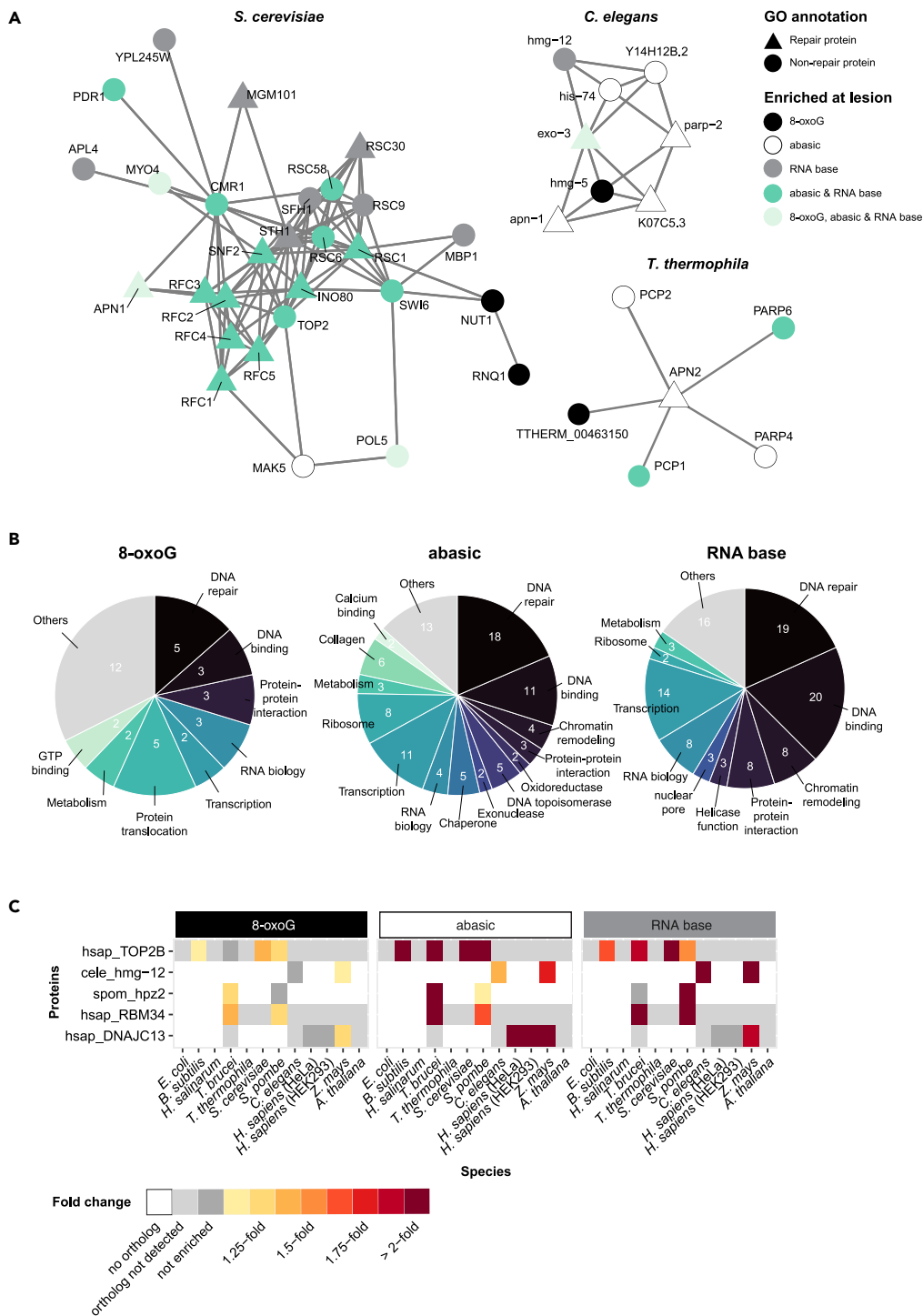
tree including all AP endonucleases across the 11 species, we demonstrate the potential enrichment differences between the two groups (Figure S7). For both groups of endonucleases, we found 2-fold or greater binding to 8-oxoG and abasic lesions in eight of the eleven species. Additionally, more unexpectedly spanning both groups was the enrichment of AP endonucleases at the RNA base in six of the eleven species. While AP endonucleases have been well characterized within BER, thus far, they have been shown to play a more minor role in RER.<sup>16</sup> It is possible that both types of AP endonucleases play a larger role than originally anticipated.

Two additional protein families that had highly conserved enrichment patterns were the photolyases (scer\_PHR1 and atha\_PHR1) and MUTYH-related glycosylases (hsap\_MUTYH). Despite both being DNA repair proteins, the binding of these proteins was unexpected in this particular context (Figures 5B and 5C). Photolyases are known to have specific repair activity for cyclobutane pyrimidine dimers and 6-4 pyrimidine-pyrimidone photoproducts caused by UV light.<sup>38</sup> However, the *S. cerevisiae* PHR1 orthologs in *E. coli*, *T. thermophila*, and *S. cerevisiae* were significantly enriched at both the 8-oxoG and abasic lesions (Figure 5B). Both orthologs in the atha\_PHR1 group were also significantly enriched at the 8-oxoG lesion. The enrichment at the abasic lesion in *A. thaliana* and for *Z. mays* was 1.9-fold, just below our threshold. For these orthology groups, the maximum likelihood phylogenetic tree showed a clear divergence of the plant photolyases, despite their similar *in vitro* binding characteristics. We did not observe the enrichment of any orthologs of PHR1 (atha\_PHR1 and scer\_PHR1) at the RNA lesion, which extended across all species regardless of evolutionary relation (Figure 5B).

MutY-related glycosylases are well characterized in the removal of 8-oxoG:A, but there are few studies showing their binding to 8-oxoG:C, which was used in this study. In an *in vitro* setting when the diffusion rate was measured, MUTYH orthologs would reside much longer at 8-oxoG:A but also have moderate stalling at 8-oxoG:C.<sup>39</sup> MUTYH orthologs were found to bind specifically to 8-oxoG in *E. coli*, *B. subtilis*, *S. pombe*, and *H. sapiens* (Figure 5C). There were no instances of detection of an MUTYH ortholog without enrichment at 8-oxoG, indicating highly specific binding that was independent of the evolutionary relation of the protein sequences. MUTYH has recently been suggested to facilitate the overall DNA damage response as a scaffolding protein.<sup>40</sup> While this function has been primarily explored within vertebrates, our findings indicate that its multiple functionalities might have emerged far earlier in evolution than originally estimated (Figure 5C).

**Identification of uncharacterized DNA repair proteins across multiple species**

In addition to the known DNA repair proteins, one-third of the enriched proteins were previously not associated with the “DNA repair” GO term (GO:0006281). We found enrichment of 35, 85, and 105 non-DNA repair classified proteins at the 8-oxoG lesion, abasic lesion, and RNA base, respectively. To investigate these proteins further, we created species-specific networks for all three DNA damage lesions using the STRING database (Figure S8 and Tables S6, S7, S8, and S9). Within these networks, we marked proteins categorized as repair (triangle) and non-DNA repair proteins (circle) and indicated at which lesion they were enriched. Here, we highlighted the *S. cerevisiae*, *C. elegans*, and *T. thermophila* networks (Figure 6A). Within *S. cerevisiae*, all enriched proteins contained in STRING interacted and formed one large network (Figure 6A). Included are five chromatin remodelers (RSC6, RSC9, RSC58, SFH1, and SWI6) that, although not characterized as DNA repair proteins, had a prominent number of interactions with both repair and non-DNA repair proteins. Within the RSC family, RSC1, RSC30, and STH1 have been classified as DNA repair proteins and are specifically linked to BER.<sup>32,41</sup> This suggests that the other RSC proteins likely play a role in chromatin remodeling surrounding DNA repair. Additionally, the non-DNA repair protein CMR1 had 11 interaction partners, five of which were “DNA repair” proteins. Notably, although not included in the “DNA repair” GO term, CMR1 has been shown to be needed to resolve genotoxic stress and has a preference for binding UV lesions *in vitro*.<sup>42,43</sup> For the *C. elegans* interactors, we identified three different subnetworks (Figures 6A and S8). Within one subnetwork, the “DNA repair” proteins *parp-2*,



**Figure 6. Network, domain, and phylogenetic analyses implicate novel proteins in DNA repair**

(A) Networks of enriched proteins across lesions for *S. cerevisiae*, *C. elegans*, and *T. thermophila*. Interactions as established in the STRING database.

(B) Classification of non-DNA repair proteins based on Pfam domain annotation. The total number of proteins classified at 8-oxoG was 29, at abasic 75, and at the RNA base 74.

(C) Heatmap representing enrichment levels of OrthoMCL orthology groups without GO annotation “DNA repair” (GO:0006281) with two or more enriched proteins across all eleven species and 8-oxoG (black), abasic (white) and RNA base (gray) lesions. The color scale represents the fold change in comparison to control samples. Abbreviations: hsap, *Homo sapiens*; cele, *C. elegans*; spom, *S. pombe*.

exo-3, and apn-1 interacted with 3, 4, and 1 non-DNA repair proteins, respectively. All three proteins were mutually linked to hmg-5. Hmg-5 was studied in a *C. elegans* Parkinson’s disease model, and together with nth-1, BER glycosylase, and other associated proteins reduced mitochondrial stress and oxidative damage.<sup>44</sup> Within the *T. thermophila* network, there were mutual interactions between APN2, identified as a DNA repair protein based on its orthology to the *S. cerevisiae* AP endonuclease, and four different PARP-related proteins, as well as THERM\_00463150, which has not been characterized. This indicates that APN2 might orchestrate the recruitment of PARP-related proteins or that PARP-related proteins are needed for APN2 to access DNA.

We evaluated the Pfam domains found among the enriched non-DNA repair proteins to elucidate more of their potential functionalities (Table S12).<sup>45</sup> The two most frequently identified domains were DNA-binding domains: (1) “protein of unknown function, DUF573” (corresponding to Interpro protein family “GLABROUS1 enhancer-binding protein family”), which is often part of proteins associated with plant stress response, and (2) “Fungal Zn(2)-Cys(6)” often involved in growth and metabolism.<sup>46,47</sup> We assigned each Pfam domain into one of 15 categories to summarize its primary function (Table S12). In all three lesions, the majority of domains were related to DNA repair and DNA binding (Figure 6B). Thus, despite the lack of categorization as DNA repair genes under the GO term “DNA repair”, there was a clear link to DNA repair functionality within these proteins. For example, we identified the “poly(ADP-ribose) polymerase” and “DNA-Ligase Zn-finger region” in four different proteins. These included hpz1 in *S. pombe* and Tb927.10.6550 in *T. brucei*, which both belong to the same orthology group. The other two proteins are PARP-related proteins in *T. thermophila*, PCP1 and PCP2. We also detected the “PARP-associated WGR domain” and a “PARP catalytic domain” in PARP4 and PARP6 in *T. thermophila*.

Furthermore, we examined the conservation of enrichment of non-DNA repair proteins across species to further support a role in DNA damage repair and recognition of lesions. We found at least five instances in which non-DNA repair genes were enriched in multiple species (Figure 6C). Intriguingly, some of these proteins were also identified within our domain analysis. For example, both enriched proteins in the spom\_hpz2 orthology group in *T. brucei* and *S. pombe* contained a PARP-related domain. Furthermore, there was specific enrichment of the *T. brucei* ortholog (Tb927.10.6550) at the abasic lesion and the *S. pombe* ortholog (hpz1) at the RNA base. Additionally, all three proteins enriched within the hsap\_DNAJC13 orthology group have Pfam ‘DnaJ domains’. These proteins preferentially bound to the abasic lesion in both HEK293 and HeLa cell lines as well as the two paralogs in *Z. mays* (UniProt: A0A1D6K922 and A0A1D6P5Y9). The conservation of enrichment across various species in both cases suggests a very likely role in DNA repair.

Through the use of network, domain, and phylogenetics analysis, we have identified proteins that, despite not being classified as DNA repair proteins, likely have a role in the DNA damage response.

## Conclusions

Performing a mass spectrometry-based phylointeractomics screen across 11 species, we compared the binding capabilities of three well-established DNA damage lesions, an 8-oxoG modification, abasic site, and ribonucleotide base incorporation. We enriched 337 proteins across all lesions and selected species (Table 1). Of these 337 proteins, 99 were related to DNA repair, which in a proteome-wide generic screen with thousands of possible proteins strongly indicates the specificity of the experiment. Supporting the specificity even further, DNA repair-related KEGG and GO terms were overrepresented in the enriched group of proteins. Through phylogenetic analysis, we established that the enrichment of particular DNA damage proteins extends through many species.

In addition to DNA repair genes, we identified two other intriguing groups of interactors in our screen. Namely, we detected an enrichment of 82 species-specific proteins as well as proteins that have not been implicated previously in DNA repair. This group of proteins presents an avenue to study potentially unique aspects of repair or damage response in their corresponding model organism. To elucidate

functionality and connection to DNA damage repair for originally non-DNA repair proteins, we utilized network, domain, and phylogenetics analysis. With this, we indicated an additional 44 proteins to potentially play a role in the DNA damage response.

Our study systematically evaluates *in vitro* binding partners in both BER lesions and an RNA lesion in eleven model species across the Tree of Life. We recapitulate previous findings and nominate putative unknown candidates to be involved in the resolution of these lesions. Through the use of network, domain, and phylogenetics analysis, we identified a subset of non-DNA repair classified proteins to likely be involved in DNA repair. Overall, this study opens avenues for further investigation of newly identified candidates to explore key factors in the crosstalk between BER and RER DNA damage pathways.

### Limitations of the study

In some cases, we do not identify or enrich all expected interaction partners at the included lesions, which can be caused by a variety of reasons. For instance, preparation from a large range of different tissues and cellular material can lead to variation in the pool of proteins available for measurement. The lack of *in vivo* conditions, such as pH, salt concentrations, temperature, post-translational modifications, and many other cellular conditions, affects DNA-protein interactions. As we did not perform cross-linking mass spectrometry, it is possible that some more transient interactions were not maintained. Furthermore, it is important to highlight the likely creation of repair intermediates in the *in vitro* pull-down assays. The ability to repair 8-oxoG, abasic sites, and uracil residues *in vitro* has been previously demonstrated with human cell extract.<sup>48,49</sup> However, we did find that unrepaired lesions existed in our experiment; for example, 11 out of the 24 canonical DNA repair-related proteins were uniquely enriched at the 8-oxoG lesion, suggesting that unrepaired lesions persisted.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Cultivation and extract preparation
  - DNA pull-down experiments
  - Mass spectrometry sample preparation
  - Mass spectrometry measurement
  - Mass spectrometry data analysis
  - Bioinformatics analysis and statistical analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106778>.

### ACKNOWLEDGMENTS

We are indebted to Franziska Roth und Jasmin Cartano for their technical support. We thank Varvara Verkhova for cultivation of *B. subtilis*, Sabrina Dietz for growing *H. salinarum* and *C. elegans*, and Markus Engstler for providing *T. brucei* cell extract. We thank Alejandro Ceron-Noriega for help with evolutionary analysis. Assistance by the IMB Media Lab and Proteomics Core Facility is gratefully acknowledged.

Funding: This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [Project-ID 393547839-SFB 1361]; Joachim Herz Stiftung Add-on fellowship to V.A.C.S.

## AUTHOR CONTRIBUTIONS

Conceptualization, F.B. and M.S.; Investigation, E.N., V.A.C.S., and M.S.; Formal analysis, E.N., V.A.C.S., A.F.-S., M.D., M.L., and F.B.; Visualization, E.N., V.A.C.S., A.F.-S., M.D., M.L., F.B., and M.S.; Writing—Original draft, E.N., V.A.C.S., F.B., and M.S.; Writing—Review & Editing, all authors contributed; Supervision: F.B. and M.S.; Project administration: F.B.; Funding acquisition: F.B.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 31, 2022

Revised: February 27, 2023

Accepted: April 25, 2023

Published: April 29, 2023

## REFERENCES

- Ciccia, A., and Elledge, S.J. (2010). The DNA Damage Response: making it safe to play with knives. *Mol. Cell* **40**, 179–204. <https://doi.org/10.1016/j.molcel.2010.09.019>.
- Beard, W.A., Horton, J.K., Prasad, R., and Wilson, S.H. (2019). Eukaryotic base excision repair: new approaches shine light on mechanism. *Annu. Rev. Biochem.* **88**, 137–162. <https://doi.org/10.1146/annurev-biochem-013118-111315>.
- Klungland, A., Rosewell, I., Hollenbach, S., Larsen, E., Daly, G., Epe, B., Seeberg, E., Lindahl, T., and Barnes, D.E. (1999). Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proc. Natl. Acad. Sci. USA* **96**, 13300–13305. <https://doi.org/10.1073/pnas.96.23.13300>.
- Marshall, C.J., and Santangelo, T.J. (2020). Archaeal DNA repair mechanisms. *Biomolecules* **10**, 1472. <https://doi.org/10.3390/biom10111472>.
- Genois, M.-M., Paquet, E.R., Laffitte, M.-C.N., Maity, R., Rodrigue, A., Ouellette, M., and Masson, J.-Y. (2014). DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. *Microbiol. Mol. Biol. Rev.* **78**, 40–73. <https://doi.org/10.1128/MMBR.00045-13>.
- Yao, S., Feng, Y., Zhang, Y., and Feng, J. (2021). DNA damage checkpoint and repair: from the budding yeast *Saccharomyces cerevisiae* to the pathogenic fungus *Candida albicans*. *Comput. Struct. Biotechnol. J.* **19**, 6343–6354. <https://doi.org/10.1016/j.csbj.2021.11.033>.
- Robertson, A.B., Klungland, A., Rognes, T., and Leiros, I. (2009). DNA repair in mammalian cells: base excision repair: the long and short of it. *Cell. Mol. Life Sci.* **66**, 981–993. <https://doi.org/10.1007/s00018-009-8736-z>.
- Córdoba-Cañero, D., Morales-Ruiz, T., Roldán-Arjona, T., and Ariza, R.R. (2009). Single-nucleotide and long-patch base excision repair of DNA damage in plants. *Plant J.* **60**, 716–728. <https://doi.org/10.1111/j.1365-3113X.2009.03994.x>.
- Lindahl, T., and Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610–3618. <https://doi.org/10.1021/bi00769a018>.
- Gredilla, R., Garm, C., and Stevnsner, T. (2012). Nuclear and mitochondrial DNA repair in selected eukaryotic aging model systems. *Oxid. Med. Cell. Longev.* **2012**, 282438. <https://doi.org/10.1155/2012/282438>.
- Reijns, M.A.M., Rabe, B., Rigby, R.E., Mill, P., Astell, K.R., Lettice, L.A., Boyle, S., Leitch, A., Keighren, M., Kilanowski, F., et al. (2012). Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149**, 1008–1022. <https://doi.org/10.1016/j.cell.2012.04.011>.
- Nick McElhinny, S.A., Watts, B.E., Kumar, D., Watt, D.L., Lundström, E.B., Burgers, P.M.J., Johansson, E., Chabes, A., and Kunkel, T.A. (2010). Abundant ribonucleotide incorporation into DNA by yeast replicative polymerases. *Proc. Natl. Acad. Sci. USA* **107**, 4949–4954. <https://doi.org/10.1073/pnas.0914857107>.
- Williams, J.S., Lujan, S.A., and Kunkel, T.A. (2016). Processing ribonucleotides incorporated during eukaryotic DNA replication. *Nat. Rev. Mol. Cell Biol.* **17**, 350–363. <https://doi.org/10.1038/nrm.2016.37>.
- Balachander, S., Gombolay, A.L., Yang, T., Xu, P., Newnam, G., Keskin, H., El-Sayed, W.M.M., Bryksin, A.V., Tao, S., Bowen, N.E., et al. (2020). Ribonucleotide incorporation in yeast genomic DNA shows preference for cytosine and guanosine preceded by deoxyadenosine. *Nat. Commun.* **11**, 2447. <https://doi.org/10.1038/s41467-020-16152-5>.
- Sassa, A., Yasui, M., and Honma, M. (2019). Current perspectives on mechanisms of ribonucleotide incorporation and processing in mammalian DNA. *Gene Environ.* **41**, 3. <https://doi.org/10.1186/s41021-019-0118-7>.
- Kellner, V., and Luke, B. (2020). Molecular and physiological consequences of faulty eukaryotic ribonucleotide excision repair. *EMBO J.* **39**, e102309. <https://doi.org/10.15252/embj.2019102309>.
- Kappei, D., Scheibe, M., Paszkowski-Rogacz, M., Bluhm, A., Gossmann, T.I., Dietz, S., Dejung, M., Herlyn, H., Buchholz, F., Mann, M., and Butter, F. (2017). Phylointeractomics reconstructs functional evolution of protein binding. *Nat. Commun.* **8**, 14334. <https://doi.org/10.1038/ncomms14334>.
- Kovalchuk, I. (2016). Chapter 38 - conserved and divergent features of DNA repair: future perspectives in genome instability research. In *Genome Stability*, I. Kovalchuk and O. Kovalchuk, eds. (Academic Press), pp. 651–666. <https://doi.org/10.1016/B978-0-12-803309-8.00038-0>.
- Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189. <https://doi.org/10.1101/gr.1224503>.
- Chen, F., Mackey, A.J., Stoeckert, C.J., and Roos, D.S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Res.* **34**, D363–D368. <https://doi.org/10.1093/nar/gkj123>.
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592. <https://doi.org/10.1093/nar/gkac963>.
- Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.* **49**, D325–D334. <https://doi.org/10.1093/nar/gkaa1113>.
- Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N.T., Legeay, M., Fang, T., Bork, P., et al. (2021). The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612. <https://doi.org/10.1093/nar/gkaa1074>.
- Mei, Q., and Dvornyk, V. (2015). Evolutionary history of the photolyase/cryptochrome superfamily in eukaryotes. *PLoS One* **10**,

- e0135940. <https://doi.org/10.1371/journal.pone.0135940>.
25. Yudkina, A.V., Shilkin, E.S., Endutkin, A.V., Makarova, A.V., and Zharkov, D.O. (2019). Reading and misreading 8-oxoguanine, a paradigmatic ambiguous nucleobase. *Crystals* 9, 269. <https://doi.org/10.3390/cryst9050269>.
26. Elsakrmy, N., Zhang-Akiyama, Q.-M., and Ramotar, D. (2020). The base excision repair pathway in the nematode *Caenorhabditis elegans*. *Front. Cell Dev. Biol.* 8, 598860. <https://doi.org/10.3389/fcell.2020.598860>.
27. Besemer, A.S., Maus, J., Ax, M.D.A., Stein, A., Vo, S., Freese, C., Nalbach, K., von Hilchen, C., Pfalzgraf, I.F., Koziollek-Drechsler, I., et al. (2021). Receptor-mediated endocytosis 8 (RME-8)/DNAJC13 is a novel positive modulator of autophagy and stabilizes cellular protein homeostasis. *Cell. Mol. Life Sci.* 78, 645–660. <https://doi.org/10.1007/s00018-020-03521-y>.
28. Gorenberg, E.L., and Chandra, S.S. (2017). The role of Co-chaperones in synaptic proteostasis and neurodegenerative disease. *Front. Neurosci.* 11, 248. <https://doi.org/10.3389/fnins.2017.00248>.
29. Astolfi, A., Fiore, M., Melchionda, F., Indio, V., Bertuccio, S.N., and Pession, A. (2019). BCOR involvement in cancer. *Epigenomics* 11, 835–855. <https://doi.org/10.2217/epi-2018-0195>.
30. Stadler, J., and Richly, H. (2017). Regulation of DNA repair mechanisms: how the chromatin environment regulates the DNA damage response. *Int. J. Mol. Sci.* 18, 1715. <https://doi.org/10.3390/ijms18081715>.
31. Luijsterburg, M.S., de Krijger, I., Wiegant, W.W., Shah, R.G., Smeenk, G., de Groot, A.J.L., Pines, A., Vertegaal, A.C.O., Jacobs, J.J.L., Shah, G.M., and van Attikum, H. (2016). PARP1 links CHD2-mediated chromatin expansion and H3.3 deposition to DNA repair by non-homologous end-joining. *Mol. Cell* 61, 547–562. <https://doi.org/10.1016/j.molcel.2016.01.019>.
32. Tanwar, V.S., Jose, C.C., and Cuddapah, S. (2019). Role of CTCF in DNA damage response. *Mutat. Res.* 780, 61–68. <https://doi.org/10.1016/j.mrrev.2018.02.002>.
33. Czaja, W., Mao, P., and Smerdon, M.J. (2014). Chromatin remodelling complex RSC promotes base excision repair in chromatin of *Saccharomyces cerevisiae*. *DNA Repair* 16, 35–43. <https://doi.org/10.1016/j.dnarep.2014.01.002>.
34. Pedreira, T., Elfmann, C., and Stülke, J. (2022). The current state of *Subti* Wiki, the database for the model organism *Bacillus subtilis*. *Nucleic Acids Res.* 50, D875–D882. <https://doi.org/10.1093/nar/gkab943>.
35. Liu, Y., Rodriguez, Y., Ross, R.L., Zhao, R., Watts, J.A., Grunseich, C., Bruzel, A., Li, D., Burdick, J.T., Prasad, R., et al. (2020). RNA abasic sites in yeast and human cells. *Proc. Natl. Acad. Sci. USA* 117, 20689–20695. <https://doi.org/10.1073/pnas.2011511117>.
36. Redrejo-Rodríguez, M., Vigouroux, A., Mursalimov, A., Grin, I., Alili, D., Koshenov, Z., Akishev, Z., Maksimenko, A., Bissenbaev, A.K., Matkarimov, B.T., et al. (2016). Structural comparison of AP endonucleases from the exonuclease III family reveals new amino acid residues in human AP endonuclease 1 that are involved in incision of damaged DNA. *Biochimie* 128–129, 20–33. <https://doi.org/10.1016/j.biochi.2016.06.011>.
37. Daley, J.M., Zakaria, C., and Ramotar, D. (2010). The endonuclease IV family of apurinic/aprimidinic endonucleases. *Mutat. Res.* 705, 217–227. <https://doi.org/10.1016/j.mrrev.2010.07.003>.
38. Kavakli, I.H., Baris, I., Tardu, M., Gül, Ş., Öner, H., Çal, S., Bulut, S., Yarparvar, D., Berkel, Ç., Ustaoglu, P., and Aydin, C. (2017). The photolyase/cryptochrome family of proteins as DNA repair enzymes and transcriptional repressors. *Photochem. Photobiol.* 93, 93–103. <https://doi.org/10.1111/php.12669>.
39. Nelson, S.R., Kathe, S.D., Hilzinger, T.S., Averill, A.M., Warshaw, D.M., Wallace, S.S., and Lee, A.J. (2019). Single molecule glycosylase studies with engineered 8-oxoguanine DNA damage sites show functional defects of a MUTYH polyposis variant. *Nucleic Acids Res.* 47, 3058–3071. <https://doi.org/10.1093/nar/gkz045>.
40. Raetz, A.G., and David, S.S. (2019). When you're strange: unusual features of the MUTYH glycosylase and implications in cancer. *DNA Repair* 80, 16–25. <https://doi.org/10.1016/j.dnarep.2019.05.005>.
41. Bohm, K.A., Hodges, A.J., Czaja, W., Selvam, K., Smerdon, M.J., Mao, P., and Wyrick, J.J. (2021). Distinct roles for RSC and SWI/SNF chromatin remodelers in genomic excision repair. *Genome Res.* 31, 1047–1059. <https://doi.org/10.1101/gr.274373.120>.
42. Gallina, I., Colding, C., Henriksen, P., Beli, P., Nakamura, K., Offman, J., Mathiasen, D.P., Silva, S., Hoffmann, E., Groth, A., et al. (2015). Cmr1/WDR76 defines a nuclear genotoxic stress body linking genome integrity and protein quality control. *Nat. Commun.* 6, 6533. <https://doi.org/10.1038/ncomms7533>.
43. Choi, D.-H., Kwon, S.-H., Kim, J.-H., and Bae, S.-H. (2012). Saccharomyces cerevisiae Cmr1 protein preferentially binds to UV-damaged DNA in vitro. *J. Microbiol.* 50, 112–118. <https://doi.org/10.1007/s12275-012-1597-4>.
44. SenGupta, T., Palikaras, K., Esbensen, Y.Q., Konstantinidis, G., Galindo, F.J.N., Achanta, K., Kassahun, H., Stavgiannoudaki, I., Bohr, V.A., Akbari, M., et al. (2021). Base excision repair causes age-dependent accumulation of single-stranded DNA breaks that contribute to Parkinson disease pathology. *Cell Rep.* 36, 109668. <https://doi.org/10.1016/j.celrep.2021.109668>.
45. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
46. Huang, J., Zhang, Q., He, Y., Liu, W., Xu, Y., Liu, K., Xian, F., Li, J., and Hu, J. (2021). Genome-wide identification, expansion mechanism and expression profiling analysis of GLABROUS1 enhancer-binding protein (GeBP) gene family in gramineae crops. *Int. J. Mol. Sci.* 22, 8758. <https://doi.org/10.3390/ijms22168758>.
47. Tianqiao, S., Xiong, Z., You, Z., Dong, L., Jiaoling, Y., Junjie, Y., Mina, Y., Huijuan, C., Mingli, Y., Xiayan, P., et al. (2021). Genome-wide identification of Zn2Cys6 class fungal-specific transcription factors (ZnFTFs) and functional analysis of UvZnFTF1 in *Ustilago violacea*. *Rice Sci.* 28, 567–578. <https://doi.org/10.1016/j.rsci.2021.03.001>.
48. Parsons, J.L., and Dianov, G.L. (2012). In vitro base excision repair using mammalian cell extracts. In *DNA Repair Protocols Methods in Molecular Biology*, L. Bjergbæk, ed. (Humana Press), pp. 245–262. [https://doi.org/10.1007/978-1-61779-998-3\\_17](https://doi.org/10.1007/978-1-61779-998-3_17).
49. Squillaro, T., Finicelli, M., Alessio, N., Del Gaudio, S., Di Bernardo, G., Melone, M.A.B., Peluso, G., and Galderisi, U. (2019). A rapid, safe, and quantitative in vitro assay for measurement of uracil-DNA glycosylase activity. *J. Mol. Med.* 97, 991–1001. <https://doi.org/10.1007/s00109-019-01788-8>.
50. de Albuquerque, B.F.M., Luteijn, M.J., Cordeiro Rodrigues, R.J., van Bergeijk, P., Waaijers, S., Kaaij, L.J.T., Klein, H., Boxem, M., and Ketting, R.F. (2014). PID-1 is a novel factor that operates during 21U-RNA biogenesis in *Caenorhabditis elegans*. *Genes Dev.* 28, 683–688. <https://doi.org/10.1101/gad.238220.114>.
51. Scherer, M., Levin, M., Butter, F., and Scheibe, M. (2020). Quantitative proteomics to identify nuclear RNA-binding proteins of Malat1. *Int. J. Mol. Sci.* 21, E1166. <https://doi.org/10.3390/ijms21031166>.
52. Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896–1906. <https://doi.org/10.1038/nprot.2007.261>.
53. Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazza, A., and Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinform. Oxf. Engl.* 21, 3439–3440. <https://doi.org/10.1093/bioinformatics/bti525>.
54. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* 4, 1184–1191. <https://doi.org/10.1038/nprot.2009.97>.
55. R Core Team (R Foundation for Statistical Computing, 2022). R: A Language and Environment for Statistical Computing.



56. Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* 28, 1947–1951. <https://doi.org/10.1002/pro.3715>.
57. Csárdi, G., and Nepusz, T. (InterJournal, complex systems. 2006). *The Igraph Software Package for Complex Network Research*.
58. Tyner, S., Briatte, F., and Hofmann, H. (2017). Network visualization with ggplot2. *Rom. Jahrb.* 9, 27–59. <https://doi.org/10.32614/RJ-2017-023>.
59. Sievers, F., and Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 27, 135–145. <https://doi.org/10.1002/pro.3290>.
60. Paradis, E., and Schliep, K. (2019). Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. <https://doi.org/10.1093/bioinformatics/bty633>.
61. Perez-Riverol, Y., Bai, J., Bandla, C., Garcia-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50, D543–D552. <https://doi.org/10.1093/nar/gkab1038>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Bacterial and virus strains</b>		
<i>E. coli</i> : DH5 $\alpha$	NEB	C2987H
<i>B. subtilis</i> : DSM10	DSMZ	DSM10
<b>Chemicals, peptides, and recombinant proteins</b>		
Dynabeads Streptavidin C1	Thermo Scientific	65002
Complete protease inhibitor cocktail tablets	Roche	04693116001
Protein assay dye reagent concentrate 5x	Biorad	#500-0006
NuPAGE LDS Sample Buffer (4x)	Life Technologies	NP0008
DL-DITHIOTHREITOL, $\geq 98\%$ (TLC), $\geq 99.0\%$	Sigma	D0632
4-10% NuPage NOVEX PAGE gel	Novex	NP0321BOX
Iodoacetamide $\geq 99\%$ (NMR)	Sigma	I6125
Acetonitrile	VWR	20048.320
Trypsin, MS approved, from porcine pancreas	Serva	37286.03
ReproSil-Pur 120 C18-AQ, 1.9 $\mu\text{m}$ 15 % C endc.	Dr. Maisch GmbH	r119.aq.0001
<b>Deposited data</b>		
Original code	This paper (Mario Dejung)	Github: <a href="https://github.com/mariodejung/DNADamage_phylointeractome">https://github.com/mariodejung/DNADamage_phylointeractome</a>
Original code	This paper (Albert Fradera Sola)	Github: <a href="https://github.com/AFraderaSola/DNADamage_Phylointeracome">https://github.com/AFraderaSola/DNADamage_Phylointeracome</a>
Mass spectrometry proteomics data	This paper	ProteomeXchange; PXD036040
<b>Experimental models: Cell lines</b>		
<i>H. sapiens</i> : HeLa	ECACC	HeLa S3
<i>H. sapiens</i> : HEK293	ECACC	HEK293
<b>Experimental models: Organisms/strains</b>		
<i>H. salinarum</i> : NRC-1	Vencio lab (University of Sao Paulo)	NRC-1
<i>S. cerevisiae</i> : BY4742 $\alpha$	Luke lab (JGU)	BY4742 $\alpha$
<i>S. pombe</i> : pp265	Baumann lab (JGU)	pp265
<i>T. thermophila</i> : SB210	Tetrahymena Stock Center	Stock ID: SD00703
<i>T. brucei</i> :	Engstler lab (University of Wuerzburg)	Lister 427
<i>C. elegans</i> : N2	Caenorhabditis Genetics Center	Strain Name: N2 Genotype: <i>C. elegans</i> wild isolate
<i>A. thaliana</i> : Columbia	Wachter lab (JGU)	Columbia
<i>Z. mays</i>	LIDL	N/A
<b>Oligonucleotides</b>		
Control without Lesion: AGAGTAAGGGCCT GCGGCGAGGATCCGACCACGATTCGCGC AGAAGGGGCCGAAATTCGCCGTGGACTC CCTCAGTAAT	Bio-synthesis	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
8-oxoG lesion: AGAGTAAGGGCCTGCGGC GAG(8-Oxo-dG) ATCCGACCACGATTCGCG CAGAAGGGGCCGAAATTCGCCGTGGACT CCCTCAGTAAT	Bio-synthesis	N/A
abasic lesion: AGAGTAAGGGCCTGCGGC AG(dSpacer) ATCCGACCACGATTCGCGCA GAAGGGGCCGAAATTCGCCGTGGACTCC CTCAGTAAT	Bio-synthesis	N/A
RNA lesion: AGAGTAAGGGCCTGCGGC AG(rU) ATCCGACCACGATTCGCGCAGAA GGGGCCGAAATTCGCCGTGGACTCCCT CAGTAAT	Bio-synthesis	N/A
Annealed strand (reverse control): (Biotin)ATTAAGGAGGGAGTCCAC GGCGAATTCGGCCCCCTTCTGCG CGAATCGTGGTCGGATCCTCGCC GCAGGCCCTTACTCT	Metabion	N/A
<b>Software and algorithms</b>		
MaxQuant	Cox and Mann <sup>1</sup>	1.6.5.0
R	The R core team	4.2.0

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Falk Butter ([f.butter@imb.de](mailto:f.butter@imb.de)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD036040.
- All original code has been deposited into the GitHub repository used for the proteomics and STRING database analysis, which is available at [https://github.com/mariodejung/DNADamage\\_phylointeractome](https://github.com/mariodejung/DNADamage_phylointeractome) and [https://github.com/AFraderaSola/DNADamage\\_Phylointeracome](https://github.com/AFraderaSola/DNADamage_Phylointeracome).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

All cultivation and growth conditions relevant for *B. subtilis* (DSM10), *E. coli* (DH5α), *H. salinarum* (NRC-1), *S. cerevisiae* (BY4742α), *S. pombe* (pp265), *T. thermophila* (SB210), *C. elegans* (N2), *T. brucei* (Lister 427), *H. sapiens* cell lines (HeLa and HEK293), *A. thaliana* and *Z. mays* are included within the 'method details' section.

## METHOD DETAILS

### Cultivation and extract preparation

**Bacteria:** *B. subtilis* (DSM10) and *E. coli* (DH5α) were grown at 37°C in LB medium (IMB media lab) and harvested at OD<sub>600</sub>=0.7. Cell pellets were resuspended in PBB buffer (150 mM NaCl, 50 mM Tris/HCl pH 8.0, 0.5% Igepal CA-630, 10 mM MgCl<sub>2</sub>, Pierce protease inhibitor EDTA free) and sonicated with a sonifier 450

(Branson) 3 times for 45 s (cycle=70%, output level 2) with 2-minute breaks. The lysate was centrifuged at 4°C for 15 min at 20,200 x g. The supernatant was supplemented with 10% (f.c.) glycerol (Roth), shock-frozen in liquid nitrogen and stored at –80°C.

**Archaea:** *H. salinarum* strain NRC-1 was cultivated in Complex Media (4.3 M NaCl, 81 mM MgSO<sub>4</sub> x 7 H<sub>2</sub>O, 27 mM KCl, 12 mM sodium citrate, 1% w/v oxoid peptone) at 37°C and in light for ~52h/2.5 days and harvested at OD<sub>600</sub>=0.5. The cells were pelleted at 3,500 x g for 30 min at 4°C and washed twice in Basic Salt Solution (4.3 M NaCl, 81 mM MgSO<sub>4</sub> x 7 H<sub>2</sub>O, 27 mM KCl, 12 mM sodium citrate) to remove the medium. After washing, the cells were resuspended in 10 ml of lysis buffer (2.1M NaCl, 50 mM Tris/HCl pH 7.5, 10 mM MgCl<sub>2</sub>) and sonicated on ice using a Branson 450 sonifier 6 times for 30 s (cycle=50%, output level 2) with 1 min breaks. The sonicated lysate was cleared by centrifugation at 3,500 x g for 30 min at 4°C and supplemented with 10% (f.c.) glycerol (Sigma) before shock-freezing in liquid nitrogen and stored at –80°C.

**Yeast:** *S. cerevisiae* (BY4742a) was grown in YP medium containing 20% glucose (IMB media lab) at 37°C until OD<sub>600</sub> = 0.5 and harvested by centrifugation at 20,200 x g. *S. pombe* (pp265) was cultivated in YES media at 32°C until OD<sub>600</sub> = 1.0 and harvested by centrifugation. For both species, cells were lysed using 0.5 mm zirconia glass beads (Roth) in lysis buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 0.01% Igepal CA-630, 1x PMSF) at 4°C with 3 cycles alternating between 30 s milling and 30 s cooling using a FastPrep-24 system (MP Biomedicals). The supernatant was transferred to a new tube, shock-frozen in liquid nitrogen and stored at –80°C.

**T. thermophila:** A mid-log SB210 culture of 3x10<sup>7</sup> cells was grown in 2% proteose peptone (BD Biosciences), 0.2% yeast extract (BD Biosciences), 12 μM ferric chloride, and 1x penicillin/streptomycin/fungizone (HyClone) at 30 °C at 100–120 rotations per minute. Cells were pelleted at 1,500 x g for 3 minutes and washed in 10 mM Tris-HCl pH 7.4. Cells were transferred to a 1.5 ml centrifuge tube and centrifuged at 1,500 x g for 2 min, and the supernatant was removed. Cells were resuspended in 1.2 ml lysis buffer (350 mM NaCl, 40 mM Hepes pH 7.5, 1% Triton X-100, 10% glycerol, freshly added 1 mM DTT, and 1x complete protease inhibitors [Roche]), and approximately 200 μl zirconia glass beads (Roth) were added and vortexed for 3 minutes at 4°C. The tube was centrifuged at ≥ 16,000 x g at 4°C for 5 min, and the supernatant was transferred to a new tube. The sample was centrifuged at ≥ 16,000 x g at 4°C for 15 minutes. The supernatant was transferred to a new tube, shock-frozen in liquid nitrogen and stored at –80°C.

**C. elegans:** Nuclear extraction was performed with N2 gravid adult worms as in.<sup>50</sup> Worms were synchronized and grown on egg plates until they reached the gravid adult stage. Then, worms were washed with M9 buffer 3 times, pelleted, and frozen into pellets in Extraction Buffer (40 mM NaCl, 20 mM MOPS pH 7.5, 90 mM KCl, 2 mM EDTA, 0.5 mM EGTA, 10% glycerol, 2 mM DTT, and 1x complete protease inhibitors, Roche). Pellets were ground into a fine powder with a mortar and pestle. The powder was transferred to a precooled glass douncer (Kimble), and the samples were ruptured with piston B over 30 strokes. The debris was cleared twice at 200 x g for 5 minutes at 4°C. The nuclear pellet was isolated by centrifuging at 2,000 x g for 5 minutes at 4°C. This pellet was washed in extraction buffer twice. The nuclear pellet was resuspended in 200 μL Buffer C+ (420 mM NaCl, 20 mM Hepes/KOH pH 7.9, 2 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 20% glycerol, and freshly added 0.1% Igepal CA-630, 0.5 mM DTT, 1x complete protease inhibitors [Roche]). The lysate was centrifuged at 4°C for 15 min at 20,200 x g. The supernatant was supplemented with 10% (f.c.) glycerol (Roth), shock-frozen in liquid nitrogen and stored at –80°C.

**Plants:** *Z. mays* and *A. thaliana* (Columbia) were ground, frozen in liquid nitrogen and transferred to a liquid nitrogen precooled 50 ml steel container for cryomilling with an MM400 (Retsch) at 30 Hz for 4 min. *Z. mays* powder was resuspended in 35 ml PBB buffer (150 mM NaCl, 50 mM Tris/HCl pH 8.0, 0.5% IGEPAL-CA630, 10 mM MgCl<sub>2</sub>, Pierce protease inhibitor EDTA free) and incubated on ice for 10 min. For *A. thaliana*, powder was resuspended in 30 ml Buffer A (10 mM Hepes KOH pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl), incubated on ice for 10 min, and subsequently dounced with 40 strokes in a glass douncer using pestle B (Kimble). After centrifugation at 3,640 x g at 4°C, the pellet was washed with 1x DPBS (Gibco), centrifuged again and incubated in 4-6 ml Buffer C+ (420 mM NaCl, 20 mM Hepes/KOH pH 7.9, 2 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 20% glycerol, and freshly added 0.1% Igepal CA-630, 0.5 mM DTT, 1x complete protease inhibitors [Roche]) for 1 hour at 4°C on a rotation wheel. Cell fragments were removed by centrifugation at 20,200 x g and 4°C for 60 min. The supernatant was shock-frozen in liquid nitrogen and stored at –80°C.

**Cultured cells:** HeLa and HEK293 cells were grown in DMEM (Gibco) with 10% FBS (Gibco) and PennStrep (Sigma) at 37°C with 75% relative humidity and 5% CO<sub>2</sub> in an incubator (Thermo). Cells were harvested, washed in 1x DPBS (Gibco), resuspended in buffer A (10 mM Hepes KOH pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl) and incubated on ice for 10 min. Cells were centrifuged at 500 x g for 5 min and resuspended in Buffer A+ (10 mM Hepes KOH pH 7.9, 1.5 mM MgCl<sub>2</sub>, 10 mM KCl, Roche protease inhibitor EDTA free, 0.1% Igepal CA-630, 0.5 mM DTT) and then dounced with 40 strokes in a glass douncer using pestle B (Kimble). Cells were centrifuged at 2,640 x g for 15 min, and the cell pellet was washed with 1x DPBS (Gibco) prior to incubation of the pellet in buffer C+ (420 mM NaCl, 20 mM HEPES/KOH pH 7.9, 2 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 20% glycerol, and freshly added 0.1% Igepal CA-630, 0.5 mM DTT, 1x complete protease inhibitors [Roche]) for 1 hour at 4°C on a rotation wheel. Cell fragments were removed by centrifugation at 20,200 x g and 4°C for 60 min. The supernatant was shock-frozen in liquid nitrogen and stored at -80°C.

### DNA pull-down experiments

Chemically synthesized oligonucleotides (Table S1) were ordered HPLC-purified from BioSynthesis (Lewisville) and Metabion (Planegg). For pull-downs, 1 nmol of single-stranded DNA lesion (or nondamaged control) oligonucleotide was annealed with 1 nmol of 5'-biotinylated counterstrand with annealing buffer (20 mM Tris-HCl pH 8.0, 10 mM MgCl<sub>2</sub>, 100 mM KCl) by first heating to 85°C for 5 min and slowly cooling to RT. The double-stranded oligonucleotides were immobilized on 250 µg streptavidin Dynabeads C1 (Thermo) and incubated with different amounts of protein extract ranging from 200-1,000 µg (200 µg: *C. elegans*, *Z. mays* and *A. thaliana*; 400 µg: HEK293 and HeLa; 500 µg: *H. salinarum*, *T. thermophila*; 800 µg: *S. cerevisiae* and 1,000 µg: *B. subtilis*, *E. coli*, *S. pombe* and *T. brucei*) in 1x PBB buffer (150 mM NaCl, 50 mM Tris-HCl pH 8.0, 0.5% Igepal CA-630, 5 mM MgCl<sub>2</sub> and 1x protease inhibitor cocktail [Roche]) rotating at 4°C for 90 min. Protein concentrations were determined using Protein Assay Dye Reagent (Bio-Rad). All samples were prepared in quadruplicate. After incubation, unbound proteins were removed by 3 washes with PBB buffer. The Dynabeads were ultimately resuspended in 25 µl 1x LDS (Thermo) containing 100 mM DTT (Sigma) and heated to 70°C for 10 min.

### Mass spectrometry sample preparation

LDS supernatant was loaded on a 4-10% NuPage NOVEX PAGE gel (Thermo) and run for 10 min at 180 V. Samples were processed as previously described.<sup>51</sup> In short, gel pieces were cut, destained with 50% EtOH/50 mM ammonium bicarbonate (ABC), dehydrated with acetonitrile (VWR), reduced with 10 mM DTT (Sigma), alkylated using iodoacetamide (Sigma) and subsequently again dehydrated with acetonitrile (VWR) and digested with 1 µg of MS-grade trypsin (Sigma) at 37°C overnight. The peptides were eluted from the gel pieces, loaded onto a StageTip<sup>52</sup> and stored at 4°C until measurement.

### Mass spectrometry measurement

Peptides were eluted from the StageTips using 80% acetonitrile/0.1% formic acid and concentrated prior to loading either on an uHPLC nLC-1000 system coupled to a Q Exactive Plus mass spectrometer (Thermo) or an uHPLC nLC-1200 system coupled to an Exploris 480 mass spectrometer (Thermo). The peptides were loaded on a 20 cm (Q Exactive Plus) or 50 cm (Exploris 480) column (75 µm inner diameter) in-house packed with Reprosil C18 (Dr. Maisch GmbH) and eluted with a 73- or 88-min optimized gradient increasing from 2% to 40% mixture of 80% acetonitrile/0.1% formic acid at a flow rate of 225 nl/min or 250 nl/min. The Q Exactive Plus was operated in positive ion mode with a data-dependent acquisition strategy of one MS full scan (scan range 300 - 1,650 m/z; 70,000 resolution; AGC target 3e6; max IT 20 ms) and up to ten MS/MS scans (17,500 resolution; AGC target 1e5, max IT 120 ms; isolation window 1.8 m/z) with peptide match preferred using HCD fragmentation. The Exploris 480 was operated in positive ion mode with a data-dependent acquisition strategy of one MS full scan (scan range 300 - 1,650 m/z; 60,000 resolution; normalized AGC target 300%; max IT 28 ms) and up to twenty MS/MS scans (15,000 resolution; AGC target 100%, max IT 40 ms; isolation window 1.4 m/z) with peptide match preferred using HCD fragmentation.

### Mass spectrometry data analysis

MaxQuant (Version 1.6.5.0) was used to search and quantify the raw mass spectrometry files for each species individually.<sup>53</sup> Individual protein databases used as search space for MaxQuant can be found in Table S3. Oxidation and acetylation were set as variable modifications, and carbamidomethylation was set as a fixed modification. Label-free quantification (LFQ) was used to calculate and normalize intensities without activating fast LFQ. The minimum ratio count used was 2. Match between runs was used to match

within each lesion (control, abasic, 8-oxoG, RNA base), with a match time window of 0.7 min, match ion mobility window of 0.05, alignment time window of 20 min, and alignment ion mobility of 1. Matching of unidentified features was deactivated. For protein quantification, we used a label minimum ratio count of 2 and unique + razor peptides for quantification.

### Bioinformatics analysis and statistical analysis

MaxQuant proteinGroup results files of all species were combined into a single file, with a column “species” indicating the individual species and cell type (Table S4). The complete dataset was filtered by removing reverse database binders, potential contaminants or proteins identified only on a modification site. Additionally, all protein groups with fewer than 2 peptides (1 unique) were filtered out. Missing LFQ values were treated as if they were below the detection limit of the mass spectrometer. Imputation was performed for each replicate of a condition individually from a beta distribution, within a range of the 0.2 and 2.5 percentile of measured intensities of the replicate. Only proteins that were present in  $\geq 2$  replicates of 4 per pull-down condition were used to calculate enrichment values ( $\log_2$  fold change, p value by Welch t-test) (Table S4). Gene information and annotations were downloaded<sup>54,55</sup> and used to assign detected proteins to orthology groups, as per OrthoMCL.<sup>20</sup> Labeling of specific orthology groups for Figure 6 was performed based on the following hierarchy of species: *hsap*, *scer*, *spom*, *cele*, *ecol*, *atha*, *bsub*, *halo*, *tbrt*, *tetr*, and *zmay*. In other words, if an orthology group contained a human gene, it would be referred to as this. If not, the *S. cerevisiae* gene was taken, and so forth according to the listed hierarchy. If multiple genes of one species were present in the orthology group, the first one from the list was selected. Heatmap clustering was performed on a numerical matrix, where 1 was an enriched protein ( $\log_2$ -fold change  $>2$ , p value  $<0.05$ ), 0 a detected protein (i.e., not enriched but measured), and  $-1$  a protein not detected within a species at all. To find similar clusters of proteins, we applied the complete linkage method (default setting) in *hclust* from the stats package in the R framework.<sup>56</sup> For functional enrichment analysis, terms were queried in the Gene Ontology (GO)<sup>22</sup> and the Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>57</sup> databases. Terms for a particular group of enriched proteins were tested for overrepresentation (adjusted p value [FDR]  $< 0.05$ ; Fisher’s exact test) against all terms found in the background (whole genome). The top three most overrepresented terms in each database were selected for graphical representation. To determine known and predicted interactions, enriched proteins were queried in the STRING database version 11.5.<sup>23</sup> Hits from text mining and co-occurrence interaction sources were excluded. Hits with a score  $>150$  in any of the remaining interaction sources (experiments, databases, coexpression, gene fusion and neighborhood) were included in the downstream analysis. Thus, protein-protein networks were generated with in-house scripts based on an R framework incorporating *igraph*,<sup>58</sup> with the Fruchterman-Reingold force-directed layout algorithm implementation, and *ggnetwork*.<sup>59</sup> Enriched proteins were illustrated as nodes, where color indicates their associated experimental lesion and their shape indicates whether they are known repair proteins or not. STRING known and predicted interactions were visualized as edges. All networks were drawn with the spoke model.

For phylogenetic tree construction, the amino acid sequences of all orthologs from the respective OrthoMCL groups were extracted from the species-specific protein sequence FASTA files (Table S3). For AP endonucleases, the OrthoMCL groups OG6\_101139 and OG6\_104339 were chosen to represent the group. OG6\_104135 and OG6\_100453 contain the Photolyase family, and OG6\_102506 contains the MutY Glycosylase family. The evolutionary history was inferred by using the maximum likelihood method and JTT matrix-based model.<sup>60</sup> The tree with the highest log likelihood is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the JTT model and then selecting the topology with superior log likelihood value. Evolutionary analyses were conducted in MEGA X.<sup>61</sup>

Pfam analysis for proteins with no previous DNA repair associations was conducted using Pfam domain annotations downloaded from OrthoMCL.<sup>20</sup> To enable broader categorizations, Pfam terms were classified into more general terms based on text mining of the Pfam term description (Table S12). These classifiers were used to detect the distribution of Pfam functions across the proteins that have not been previously annotated as DNA repair proteins.

### QUANTIFICATION AND STATISTICAL ANALYSIS

All quantification and statistical analysis details and associated citations can be found in the method details in the ‘mass spectrometry data analysis’ and ‘bioinformatics analysis and statistical analysis’ sections. In

short, the pull downs performed in the analysis were performed in quadruplicate, and the p value was determined by Welch's t-test with an enrichment threshold of  $\log_2$  fold change  $>2$  and p value  $<0.05$ . Utilizing both GO and KEGG databases, enriched proteins were tested for overrepresentation using Fisher's exact test, determining an adjusted p value (false discovery rate)  $<0.05$ . The STRING database was used to determine previously established interactors to proteins of interest. To create phylogenetic trees, the evolutionary history was inferred in MEGA X<sup>61</sup> by using the maximum likelihood method and JTT matrix-based model.<sup>60</sup> Pfam domain annotations were downloaded from OrthoMCL.<sup>20</sup>