



RESEARCH

Open Access

Evolutionary loss of 8-oxo-G repair components among eukaryotes

Kristina Jansson, Anders Blomberg, Per Sunnerhagen, Magnus Alm Rosenblad*

Abstract

Background: We have examined the phylogenetic pattern among eukaryotes of homologues of the *E. coli* 7,8-dihydro-8-oxoguanine (8-oxo-G) repair enzymes MutY, MutM, and MutT.

Results: These DNA repair enzymes are present in all large phylogenetic groups, with MutM homologues being the most universally conserved. All chordates and echinoderms were found to possess all three 8-oxo-G repair components. Likewise, the red and green algae examined have all three repair enzymes, while all land-living plants have MutY and MutM homologues, but lack MutT. However, for some phyla, e.g. protostomes, a more patchy distribution was found. Nematodes provide a striking example, where *Caenorhabditis* is the only identified example of an organism group having none of the three repair enzymes, while the genome of another nematode, *Trichinella spiralis*, instead encodes all three. The most complex distribution exists in fungi, where many different patterns of retention or loss of the three repair components are found. In addition, we found sequence insertions near or within the catalytic sites of MutY, MutM, and MutT to be present in some subgroups of Ascomycetes.

Conclusion: The 8-oxo-G repair enzymes are ancient in origin, and loss of individual 8-oxo-G repair components at several distinct points in evolution appears to be the most likely explanation for the phylogenetic pattern among eukaryotes.

Background

To maintain structural integrity of DNA, organisms have developed DNA repair mechanisms. These have evolved both in complexity and specificity to ensure genomic integrity against the constant threats from damaging agents of endogenous and exogenous origins. Damage to DNA bases resulting from alkylation, oxidation, deamination, and UV-induced crosslinking, is mainly repaired by the base excision repair (BER) pathway, which is highly conserved throughout evolution and ubiquitously present in bacteria, archaea, and eukaryotes [1]. BER is the major pathway for repair of oxidative base damage, transcription-coupled repair (TCR) and mismatch repair (MMR) being important backup pathways. Moreover, several of the DNA glycosylases that initiate BER of oxidative damage have overlapping specificities and serve as alternative pathways for various DNA lesions [2]. Oxidative damage in DNA, specifically the 8-oxo-G lesion, is removed or prevented by the 8-

oxo-G-specific BER enzymes MutY, MutM, and MutT [3]. MutT is an 8-oxo-dGTPase that prevents incorporation of 8-oxo-G into DNA [4]. MutM excises 8-oxo-G paired with C [5], while MutY is an adenine-DNA glycosylase that excises A paired with 8-oxo-G [6]. The MutY and MutM glycosylases are both members of the helix-hairpin-helix (HhH) superfamily. This gene subfamily is the most diverse of the DNA glycosylases, with differing substrate specificities [7]. The MutT homologue belongs to the group of nudix hydrolases and is not classified as a DNA glycosylase, despite being a component of the 8-oxo-G repair system [1].

The existence of an 8-oxo-G repair system in all main organism groups; archaea, bacteria, fungi, animals, and plants, underscores the importance of this system to defend against deleterious 8-oxo-G mutagenesis. Despite the widespread conservation and importance of this repair system in maintaining genomic stability, limited phylogenetic data is available about the highly diverse and adaptable HhH gene family of DNA repair enzymes among eukaryotes. The expanding number of entire genome sequences from a wide range of eukaryotic

* Correspondence: Magnus.Almrosenblad@cmb.gu.se
Department of Cell and Molecular Biology, Lundberg Laboratory, University of Gothenburg, P.O. Box 462, S-405 30 Göteborg, Sweden

groups therefore encouraged an analysis of the phylogenetic distribution of the 8-oxo-G repair genes. As a broad phylogenetic analysis of the HhH superfamily of BER DNA glycosylases among prokaryotes has already been presented [7], prokaryotes were omitted from this study. Here, we have identified 8-oxo-G repair genes from metazoans, fungi and plants, along with a sequence analysis of the identified proteins.

We find that their phylogenetic distribution among eukaryotes strongly argues for group-specific gene losses. Thus, we reveal several cases of unexpected gene distributions, despite the fact that our analysis includes organisms where DNA repair has been extensively characterised both biochemically and genetically: yeast, mammals, and higher plants.

Results

Phylogenetic distribution

The phylogenetic distribution and sequence analysis of the 8-oxo-G repair components included a large number of species across the domains of fungi, animals, and plants. In general, homologues of all three types of 8-oxo-G involved enzymes MutY, MutM, and MutT were found in most surveyed subgroups of animals and plants with a few exceptions (Figures 1, 2 and 3). All three 8-oxo-G repair enzymes are found throughout the entire group of chordates (8 species examined; Fig. 1). In red and green algae, the distribution pattern is likewise uniform, with all three repair homologues found in all species examined. Notably, all eight species of land-living plants examined lack a MutT homologue, but do possess MutY and MutM homologues (Fig. 2). However, MutY homologues are apparently absent in insects and annelids (Fig. 1). Notably, the *Caenorhabditis* nematodes harbour none of the three 8-oxo-G repair genes of interest. Interestingly, all three enzymes are found in the nematode *Trichinella spiralis* and MutM is found in *Brugia malayi*, showing that nematodes have evolved several different ways to deal with 8-oxo-G damage. Each of the two nematodes *C. elegans* and *C. briggsae* harbours one HhH superfamily homologue from the Nth BER family [7], which may functionally overlap and explain the lack of the 8-oxo-G repair system. Similarly, the molluscs display a very scattered distribution, with no single species harbouring all three 8-oxo-G repair proteins. However, considering molluscs as a group, all enzyme members of the 8-oxo-G repair system are represented (Fig. 1). Incomplete genome sequencing cannot be ruled out as a source of error, however.

The situation in fungi looks more diverse. Overall, all three 8-oxo-G repair homologues are found in basidiomycetous fungi (Fig. 3). However, among ascomycetes, the phylogenetic distribution pattern is clear but patchy. Of the three proteins, MutT is the least prevalent in

Ascomycota, and is retained only in the “Candida” group. All ascomycetous fungi were found to harbour the MutM homologue, with the exception of *Schizosaccharomyces*, although in this clade the single species *Sz. japonicus* seems to have the MutM protein as well. The MutY homologue is absent not only from *Saccharomyces cerevisiae* and closely related species, but also from other organisms of Saccharomycotina. Interestingly, the MutY component is found in Sordariomycetes and Dothideomycetes (subgroups of Pezizomycotina), but not in Eurotiomycetes, indicating it has been lost several times during evolution of the ascomycetes. MutY was present in all *Schizosaccharomyces* species, which was expected as *Sz. pombe* has a well-characterised MutY homologue.

Overall, the MutM homologue emerges as the most prevalent repair 8-oxo-G component among eukaryotes.

Sequence divergence between orthologues

To see if sequence divergence of the three repair proteins correlates with phylogeny, or would be dependent on the presence or absence of other 8-oxo-G repair components within any given organism, multiple sequence alignments were constructed (Figs 4 and 5). Among the 8-oxo-G enzymes, only the sequence of MutY is highly conserved between prokaryotes and eukaryotes [8]. MutY has a catalytic domain containing a signature helix-hairpin-helix element, followed by a Gly/Pro-rich loop and a catalytically essential aspartate residue, referred to as the HhH-GPD motif. MutY is distinct among the HhH-GPD superfamily in that it contains an additional carboxy-terminal domain that seems to be responsible for 8-oxo-G recognition [9]. The two MutY helical domains form a positively-charged groove with the adenine specific pocket at their interface. Like MutY, other base excision repair HhH superfamily glycosylases (MutM, EndoIII and AlkA) use similar two-domain molecular scaffolds, and their DNA-binding HhH-GPD motifs is the most conserved superfamily structural element [10]. In general, a strong conservation of the defined protein domains of MutY, MutM, and MutT was found. Identified residues, critical for DNA binding and substrate interaction [11-13], are extremely well conserved through all species examined. However, the group of fungi shows some unique sequence features. The “Saccharomyces” subgroup within Saccharomycotina is found to harbour an insert in MutM, approximately 15 amino acids long, about 10 residues downstream from the HhH-PVD structural domain (Fig. 4). Furthermore, the Eurotiomycetes subgroup of Pezizomycotina harbours another MutM insert, located immediately upstream of the HhH-PVD domain, and of the same size (about 10 residues; Fig. 4). Conversely, the Saccharomycotina group “Candida” and the

		Organism	MutY	MutM	MutT
Chordata	Vertebr	<i>Homo sapiens</i>	Grey	Grey	Grey
		<i>Mus musculus</i>	Grey	Grey	Grey
		<i>Gallus gallus</i>	Grey	Grey	Grey
		<i>Xenopus tropicalis</i>	Grey	Grey	Grey
		<i>Tetraodon nigroviridis</i>	Grey	Grey	Grey
		<i>Danio rerio</i>	Grey	Grey	Grey
	Ceph	<i>Branchiostoma floridae</i>	Grey	Grey	Grey
Uro	<i>Ciona intestinalis</i>	Grey	Grey	Grey	
Echinoderm		<i>Strongylocentrotus purpuratus</i>	Grey	Grey	Grey
Arthropoda	Cheli	<i>Ixodes scapularis</i>	Grey	Grey	White
	Crust	<i>Daphnia pulex</i>	Grey	Grey	Grey
		<i>Parhyale hawaiiensis</i>	Grey	Grey	Grey
	Insect	<i>Drosophila melanogaster</i>	White	Grey	White
		<i>Tribolium castaneum</i>	White	Grey	Grey
		<i>Apis mellifera</i>	White	Grey	Grey
		<i>Nasonia vitripennis</i>	White	Grey	Grey
	<i>Pediculus humanus corporis</i>	White	Grey	White	
Nematoda		<i>Caenorhabditis elegans</i>	White	White	White
		<i>Brugia malayi</i>	White	Grey	White
		<i>Trichinella spiralis</i>	White	Grey	Grey
Lophotro	Mollu	<i>Lottia gigantea</i>	White	Grey	White
		<i>Aplysia californica</i>	White	White	White
		<i>Crassostrea gigas</i>	White	White	Grey
	Anne	<i>Capitella sp.</i>	White	Grey	White
		<i>Helobdella robusta</i>	White	Grey	White
Platyhelm		<i>Schistosoma mansoni</i>	White	Grey	Grey
		<i>Schistosoma japonicum</i>	White	Grey	Grey
Cnidaria		<i>Nematostella vectensis</i>	White	Grey	Grey
		<i>Hydra magnipapillata</i>	White	Grey	Grey
Placozoa		<i>Trichoplax adhaerens</i>	White	Grey	Grey
Choanoflag		<i>Monosiga brevicollis</i>	White	Grey	Grey

Figure 1 Phylogenetic distribution of MutY, MutM, and MutT homologues among metazoans. The table shows the presence (grey) or absence (white) of 8-oxo-G repair gene homologues found in species included in the survey. Abbreviations: "Vertebr", Vertebrata; "Ceph", Cephalochordata; "Uro", Urochordata; "Cheli", Chelicerata; "Crust", Crustacea; "Lophotro", Lophotrochozoa; "Mollu", Mollusca; "Anne", Annelida; "Platyhelm", Platyhelminthes; "Choanoflag", Choanoflagellata.

		Organism	MutY	MutM	MutT
Streptophyta	Vascular	<i>Arabidopsis thaliana</i>			
		<i>Ricinus communis</i>			
		<i>Populus trichocarpa</i>			
		<i>Vitis vinifera</i>			
		<i>Zea mays</i>			
		<i>Oryza sativa</i>			
		<i>Selaginella moellendorffii</i>			
	Moss	<i>Physcomitrella patens</i>			
Chlorophyta		<i>Chlamydomonas reinhardtii</i>			
		<i>Chlorella</i>			
		<i>Ostreococcus tauri</i>			
		<i>Micromonas</i>			
Rhodophyta		<i>Cyanidioschyzon merolae</i>			

Figure 2 Phylogenetic distribution of MutY, MutM, and MutT homologues in plants. The table shows the presence (grey) or absence (white) of 8-oxo-G repair gene homologues found in species included in the survey. "Vascular", Tracheophyta.

Pezizomycotina group Sordariomycetes, that do not harbour any of the two MutM insertions, instead were found to have a small MutT sequence insertion and a longer MutY insertion, respectively (Fig. 3). The MutY insertion is located in the beginning of the specific "adenine recognition site", the substrate binding domain of the MutY protein (Fig. 5). While the longer, about 20 residues, MutY insertion is located in the important substrate recognition site, the shorter MutT insertion is located outside the highly conserved structural "nudix motif" in the catalytic site of the protein (not shown). Thus, all identified sequence insertions are found only among ascomycetous species, largely paralleling the phylogenetic division into subgroups. Interestingly, no single fungal species carries more than one out of the four identified sequence insertions (two in MutM, one in MutY, one in MutT). Construction of phylogenetic trees from entire protein sequences, in trying to clarify any mutually dependent evolutionary relationship between the 8-oxo-G repair components and these sequence insertions, did not provide any further information.

Discussion

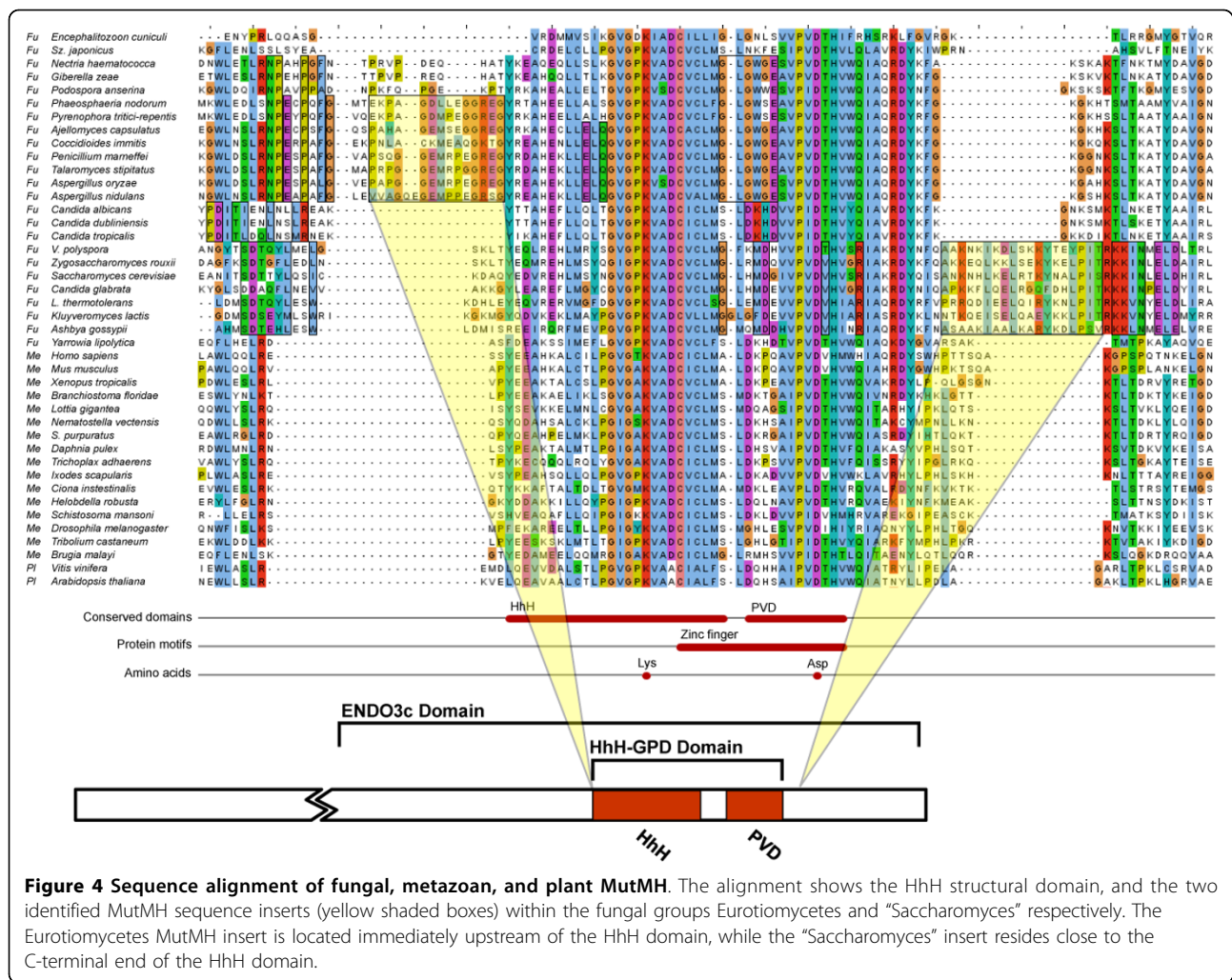
Evolutionary loss of individual 8-oxo-G repair proteins is the most likely event behind the observed distribution pattern of the MutM, MutY, and MutT proteins among eukaryotes. The existence of highly conserved full-length protein sequences throughout the lineage of evolution strongly argues for group-specific gene losses, and rules out convergent evolution of independently evolved catalytic domains of repair genes among diverse subgroups of eukaryotic species. However, it is not obvious why individual 8-oxo-G repair homologues have been lost or

retained in an organism group. In fungi, the loss of specific 8-oxo-G repair genes in distinct phylogenetic branches is very clear. Fungi within Pezizomycotina have lost the MutT homolog, while Saccharomycotina species seem to have lost the MutY homolog (Fig. 6). The Eurotiomycetes branch of Pezizomycotina moreover has lost the MutY protein, and harbours only the MutM homolog. The "Saccharomyces" and "Yarrowia" subgroups of Saccharomycotina also show loss of the MutT homolog, while it is still present in the "Candida" branch of Saccharomycotina (Fig. 6).

The "Saccharomyces", like the Eurotiomycetes subgroups, harbour only the MutM homolog. All subgroups of Ascomycota clearly harbour the MutM repair enzyme, except Schizosaccharomyces. Interestingly, in this clade the single species *Sz. japonicus* harbours a MutM homolog, with a sequence similar to other fungal homologues. By contrast, in *Sz. pombe*, MutT- and MutM-like proteins have not been identified either by sequence or experimental analysis [14]. Also, the nematode *C. elegans* deviates from the expected distribution pattern of 8-oxo-G repair enzymes, in that it harbours none of the 8-oxo-G repair components. However, *C. elegans* does hold the Nth HhH superfamily homologue, which may function in an alternative 8-oxo-G repair pathway [7]. It is noteworthy that *Caenorhabditis* thus only has one HhH superfamily member. By contrast, the nematode *Trichinella spiralis* harbours all three of the Mut proteins. This probably reflects the large evolutionary distance between nematode subgroups, with *T. spiralis* belonging to a basal lineage and evolutionary distant to *C. elegans*. Another noteworthy example is *Drosophila melanogaster*, which only has a MutM homolog,

		Organism	MutY	MutM	MutT	
A s c o m y c o t a	S a c c h a r o m y c o t i n a	Sac	<i>Saccharomyces cerevisiae</i>		* C-term	
		<i>Vanderwaltozyma polyspora</i>		* C-term		
		<i>Zygosaccharomyces rouxii</i>		* C-term		
		<i>Lachancea thermotolerans</i>		* C-term		
		<i>Candida glabrata</i>		* C-term		
		<i>Ashbya gossypii</i>		* C-term		
		<i>Kluyveromyces lactis</i>		* C-term		
		Can	<i>Pichia stipitis</i>			
		<i>Debaryomyces hansenii</i>			*	
		<i>Candida albicans</i>			*	
		<i>Candida tropicalis</i>			*	
		<i>Pichia guilliermondii</i>			*	
		<i>Candida dubliniensis</i>			*	
		<i>Candida lusitanae</i>			*	
	<i>Yarrowia lipolytica</i>					
	P e z i z o m y c o t i n a	S o r	<i>Podospira anserina</i>	*		
			<i>Chaetomium globosum</i>	*		
			<i>Gibberella zeae</i>	*		
			<i>Neurospora crassa</i>	*		
			<i>Verticillium albo-atrum</i>	*		
			<i>Nectria haematococca</i>	*		
		E u r	<i>Phaeosphaeria nodorum</i>			* N-term
			<i>Pyrenophora tritici-repentis</i>			* N-term
			<i>Ajellomyces capsulatus</i>			* N-term
			<i>Penicillium marneffei</i>			* N-term
	T a p h r i	<i>Talaromyces stipitatus</i>			* N-term	
		<i>Coccidioides immitis</i>			* N-term	
		<i>Aspergillus oryzae</i>			* N-term	
		<i>Aspergillus nidulans</i>			* N-term	
	B a s i d i o m y c e t e s	<i>Schizosaccharomyces japonicus</i>				
<i>Schizosaccharomyces cryophobus</i>						
<i>Schizosaccharomyces octosporus</i>						
<i>Schizosaccharomyces pombe</i>						
<i>Cryptococcus neoformans</i>						
Z y g o	<i>Malassezia globosa</i>					
	<i>Postia placenta</i>					
	<i>Ustilago maydis</i>					
	<i>Coprinopsis cinerea okayama</i>					
	<i>Rhizopus oryzae</i>					
M i c r o	<i>Batrachochytrium dendrobatidis</i>					
	<i>Encephalitozoon cuniculi</i>					
	<i>Enterocytozoon bieneusi</i>					
D i c t y	<i>Nosema ceranae</i>					
	<i>Dictyostelium discoideum</i>					

Figure 3 Phylogenetic distribution of MutY, MutM, and MutT homologues among fungi. The table shows the presence (grey) or absence (white) of 8-oxo-G repair gene homologues found in species included in the survey (fungi and slime mold). Repair enzyme homologues with any of the identified sequence inserts are marked with an asterisk, and where relevant their position indicated (C-term, N-term). "Sac", Saccharomyces; "Can", Candida; "Sor", Sordariomycetes; "Dot", Dothideomycetes; "Eur", Eurotiomycetes; "Taphri", Taphrinomycotina; "Basidiomyc", Basidiomycetes; "Zygo", Zygomycetes; "Chyt", Chytridiomycetes; "Micro", Microsporidia; "Dicty", Dictyostelia.

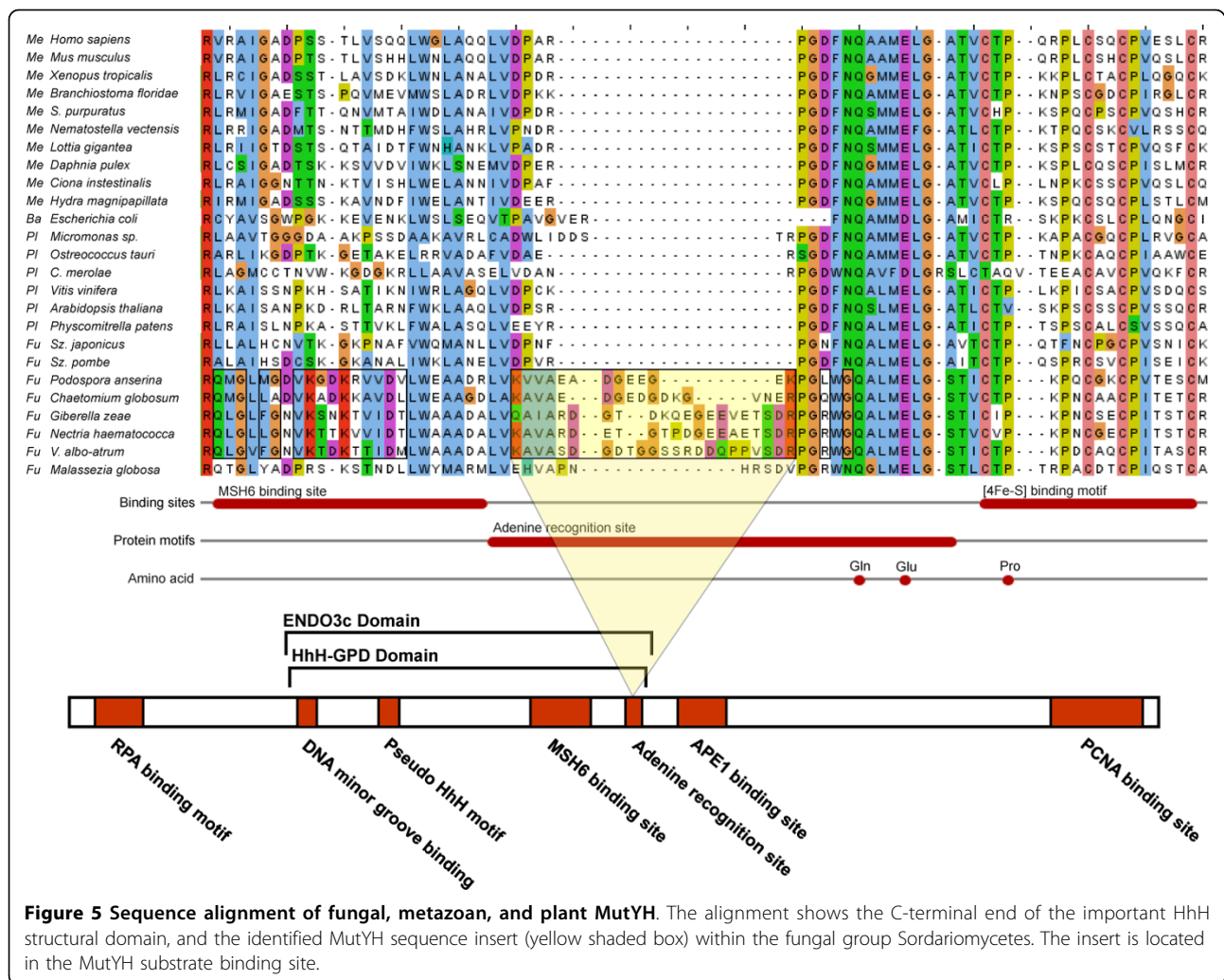


whereas most arthropods have two to three 8-oxo-G repair proteins. Therefore, the common picture of 8-oxo-G repair gene distribution, predicted from a typical "model organism", is not always the most representative view.

The widespread distribution of the MutM homologue in eukaryotic genomes, and the lack of either the MutY or the MutT homologue, or both, probably indicates 8-oxo-G in non-replicated DNA as the most abundant and important oxidative DNA damage to correct. The post-replicative adenine DNA glycosylase MutY mainly serves to excise adenines misincorporated opposite 8-oxo-G by replication, in cooperation with the MMR system [15]. This likely provides redundancy in post-replicative mismatch repair by separate pathways. All three 8-oxo-G repair components, however, are highly specific for their substrates [16], and possibly may have evolved from more "promiscuous" BER repair enzymes with catalytic activity toward alternative substrates. The situation of combining "promiscuous" broad substrate

enzymes with highly specific ones may provide an advantage in terms of specificity and redundancy within and between separate DNA repair pathways [16]. The organism thereby holds the capacity to deal with a larger variety of DNA damage in a new complex chemical environment.

The identified sequence insertions in MutY, MutM, and MutT, respectively, among subgroups of fungi, do not obviously correlate with the presence or absence of the other 8-oxo-G repair homologues. It is also not possible to predict the functional importance of these sequence insertions for specific enzymatic activities. The sequence diversity among HhH glycosylases, and to some extent observed among subgroups of fungi in this study of the 8-oxo-G repair system, may reflect the need of specific and highly adaptable systems to process complex patterns of DNA damage, caused by different environmental factors. Disparities in catalytic mechanisms and in DNA repair pathways, by which an organism processes DNA damage, probably is part of

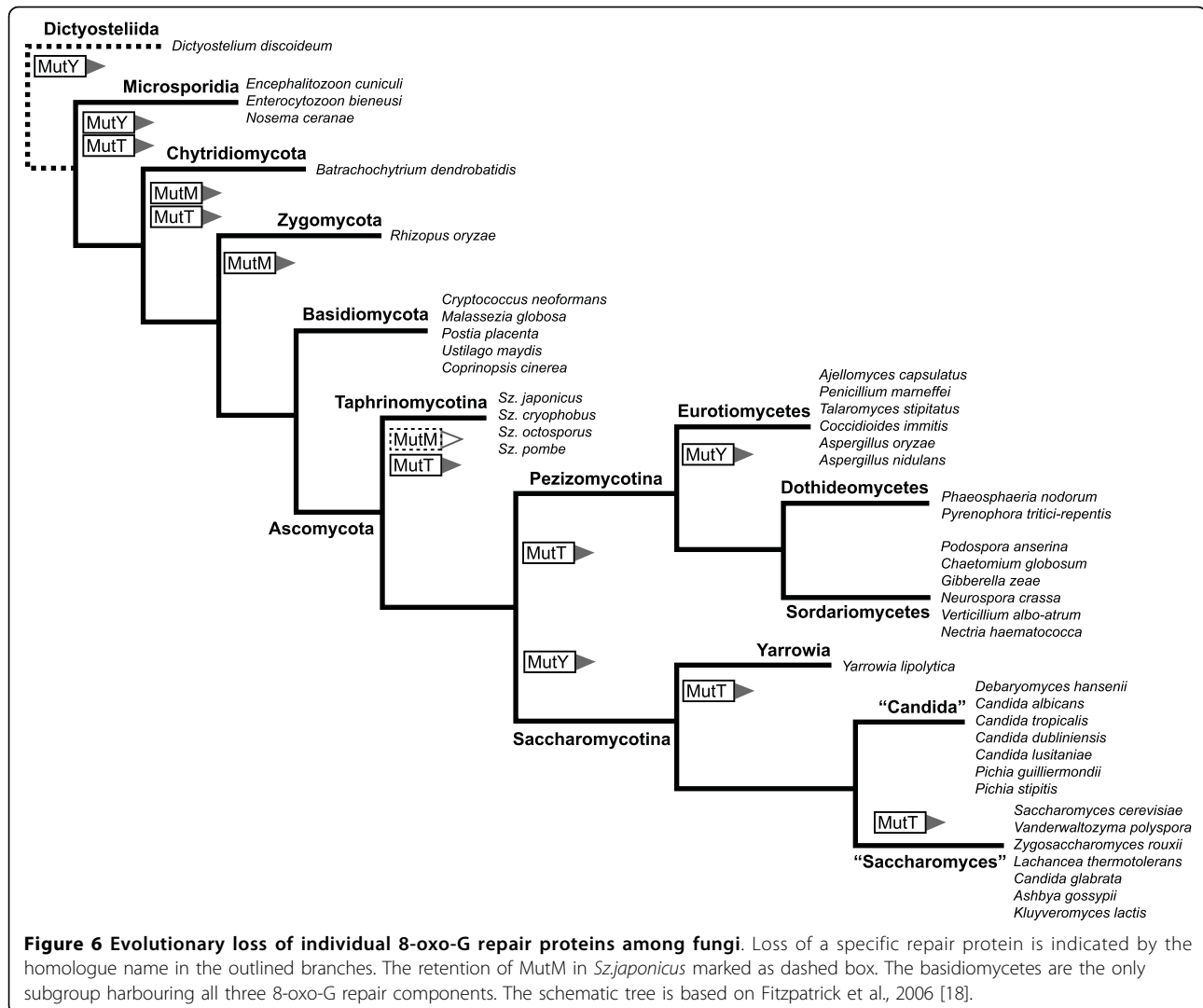


the explanation [7]. Even though there is an established functional redundancy between the MutY, MutM, and MutT proteins [17], and between separate repair pathways, in protecting against oxidative damage, more experimental data in substrate specificity and DNA repair pathway redundancy are clearly needed.

Methods

Eukaryotic organisms were included in the study if the entire genome sequence was available, with the aim of covering as wide a selection of organism groups as possible. Published sequences of MutY, MutM and MutT from *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, human, and *Arabidopsis thaliana* were used in BLASTP searches (E -values $< 1 \times 10^{-20}$) to retrieve candidate homologues in the respective groups of fungi, metazoans, and plants, plus the slime mold *Dictyostelium* and the choanoflagellate *Monosiga* (for summary of data, see Additional file 1, 2, 3 and 4). Evaluation of

candidates was based on the identification of domains in the NCBI Conserved Domains Database (CDD) that are specific for the different proteins: Endo3c (cd00056) and DNAGlycosylase_C (cd03431) for MutY, OGG_N (pfam07934, or in some cases cl06806) and Endo3c for MutM, and MTH1 (cd03427) for MutT. The domains were identified as part of the NCBI web-based BLAST interface which includes an RPS-BLAST search vs. the position-specific scoring matrices in CDD. While E -values were different for the four domains, they were always lower than 1×10^{-05} (cd03431). Only specific hits to domains were considered and best hits to similar domains (for instance Nth) were used as evidence to reject a candidate. For MutY and MutM, both domains had to be identified to score as positive. Found sequences were also subjected to reciprocal BLASTP searches, ensuring that they indeed were most similar to proteins of the respective family. Most searches were conducted using the non-redundant protein database at NCBI of December 2009.



Although the domain search made identifying proteins with low similarity to the initial query sequences easier, reliable candidates from the performed searches were also used as queries to retrieve more sequences from closely related groups, especially when a thorough evaluation was needed because the identified candidate had a different length.

As most protein sequences are based on gene predictions, many sequences had truncated ends due to the problems of identifying the exon-intron structure and thus the true ends of the gene. In those cases, the candidate protein sequence was extended by matching full sequences from closely related organisms to the genome using TBLASTN and adding the found segments to the protein sequence. For some organisms, also EST data were used to validate a predicted sequence. Long gaps or insertions within sequences that are not conserved in related species, and thus are indications of erroneous

gene predictions, were left uncorrected as long as they did not interfere with the identification of domains and could be aligned properly to the other sequences.

To get a better representation of species in the different groups and conduct searches against unpublished genomes and/or EST libraries, organism-specific databases were also used, many hosted by the Joint Genome Institute (JGI; Table 1). In the summary of results (Figs 1, 2 and 3), some species in groups with many representatives are left out (e.g. *Drosophila* and *Saccharomyces* species). In a few cases, species with only EST data available were added to show that a specific protein was indeed found within the group (e.g. the mollusc *Crassostrea gigas*).

Multiple alignments were constructed using clustalw or t_coffee and visualised in JalView where also annotated features, such as the MutY “adenine recognition site”, were included (Figs 1 and 2). The locations of

Table 1 Organism-specific databases utilised in this study.

Species	Organization	URL
<i>Apis mellifera</i>	BeeBase	http://genomes.arc.georgetown.edu/drupal/beebase/
<i>Batrachochytrium dendrobatidis</i>	Broad	http://www.broadinstitute.org/annotation/genome/batrachochytrium_dendrobatidis/
<i>Capitella teleta</i>	JGI	http://genome.jgi-psf.org/Capca1/Capca1.home.html
<i>Chlorella NC64A</i>	JGI	http://genome.jgi-psf.org/ChINC64A_1/ChINC64A_1.home.html
<i>Chlorella vulgaris</i>	JGI	http://genome.jgi-psf.org/Chlvu1/Chlvu1.home.html
<i>Ciona intestinalis</i>	JGI	http://genome.jgi-psf.org/Cioin2/Cioin2.home.html
<i>Cyanidioschyzon merolae</i>	Univ. of Tokyo	http://merolae.biol.s.u-tokyo.ac.jp/
<i>Daphnia pulex</i>	JGI	http://genome.jgi-psf.org/Dappu1/Dappu1.home.html
<i>Daphnia pulex</i>	wFleaBase	http://wfleabase.org/
<i>Dictyostelium discoideum</i>	dictyBase	http://www.dictybase.org/
<i>Helobdella robusta</i>	JGI	http://genome.jgi-psf.org/Helro1/Helro1.home.html
<i>Lottia gigantea</i>	JGI	http://genome.jgi-psf.org/Lotgi1/Lotgi1.home.html
<i>Monosiga brevicollis</i>	JGI	http://genome.jgi-psf.org/Monbr1/Monbr1.home.html
<i>Parhyale hawaiiensis</i>	JGI	http://genome.jgi-psf.org/parha/parha.home.html
<i>Rhizopus oryzae</i>	Broad	http://www.broadinstitute.org/annotation/genome/rhizopus_oryzae/
<i>Schizosaccharomyces octosporus</i>	Broad	http://www.broadinstitute.org/annotation/genome/schizosaccharomyces_group/
<i>Selaginella moellendorffii</i>	JGI	http://genome.jgi-psf.org/Selmo1/Selmo1.home.html
<i>Stagonospora (Phaeosphaeria) nodorum</i>	Broad	http://www.broadinstitute.org/annotation/genome/stagonospora_nodorum.2/
<i>Trichoplax adhaerens</i>	JGI	http://genome.jgi-psf.org/Triad1/Triad1.home.html
<i>Xenopus (Silurana) tropicalis</i>	JGI	http://genome.jgi-psf.org/Xentr4/Xentr4.home.html

found insertions were mapped to known 3D structures of the proteins from the Protein Data Bank, accessed via NCBI Structure database, using the program Cn3D. Structures used in the analysis were: MutY from *Geobacillus stearothermophilus*, PDB:1RRQ; MutM from *Homo sapiens*, PDB:2NOF; MutT from *E. coli*, PDB:3A6S.

Additional material

Additional file 1: Sequence Data. Excel file with information on all sequences included in study and data from searches. Sheets for Fungi, Metazoa and Plants plus overview for each group. Table contains accession numbers for MutY/M/T orthologous sequences as well as information on BLASTP hit values and domain hits. Bullets symbolize reliable hits for a species to a certain query and may be found in two columns for each Mut protein: "MutX" or "MutX additional refs" when additional query sequences from closely related species have been used. The initially used query sequences are listed in row 2. Comments, such as truncated domains, are listed in "Note" column.

Additional file 2: MutY alignment. For MutY and MutM most sequences from tables are included. For MutT only fungi and human sequences are included. All sequences are named with a code that identifies the organism group: 'Me' = metazoa; 'Fu' = fungi; 'Pl' = plants; 'Ba' = bacteria. Gaps in all sequences are the result of insertions in sequences removed from alignment. Annotations from literature are added. Secondary structure was taken from predictions using the "PHYRE automatic fold recognition server" (<http://www.sbg.bio.ic.ac.uk/phyre/>). Regions that are judged as different in a group of organisms are marked with boxes.

Additional file 3: MutM alignment. For MutY and MutM most sequences from tables are included. For MutT only fungi and human sequences are included. All sequences are named with a code that identifies the organism group: 'Me' = metazoa; 'Fu' = fungi; 'Pl' = plants; 'Ba' = bacteria. Gaps in all sequences are the result of insertions in sequences removed from alignment. Annotations from literature are added. Secondary structure was taken from predictions using the "PHYRE automatic fold recognition server" (<http://www.sbg.bio.ic.ac.uk/phyre/>). Regions that are judged as different in a group of organisms are marked with boxes.

Additional file 4: MutT alignment. For MutY and MutM most sequences from tables are included. For MutT only fungi and human sequences are included. All sequences are named with a code that identifies the organism group: 'Me' = metazoa; 'Fu' = fungi; 'Pl' = plants; 'Ba' = bacteria. Gaps in all sequences are the result of insertions in sequences removed from alignment. Annotations from literature are added. Secondary structure was taken from predictions using the "PHYRE automatic fold recognition server" (<http://www.sbg.bio.ic.ac.uk/phyre/>). Regions that are judged as different in a group of organisms are marked with boxes.

Acknowledgements

This work was supported by the Swedish Research Council (2007-5460) and the University of Gothenburg Science Faculty research platform in Ecotoxicogenomics.

Authors' contributions

PS conceived of the study and drafted the manuscript. MAR devised the design of the study and carried out some of the searches and gene predictions, as well as analyses of the results. KJ performed the majority of the searches and sequence alignments, put together the results and helped draft the manuscript. AB helped conceive the study and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 14 June 2010 Accepted: 1 September 2010

Published: 1 September 2010

References

1. Baute J, Depicker A: Base excision repair and its role in maintaining genome stability. *Crit Rev Biochem Mol Biol* 2008, **43**:239-276.

2. Slupphaug G, Kavli B, Krokan HE: **The interacting pathways for prevention and repair of oxidative DNA damage.** *Mutat Res* 2003, **531**:231-251.
3. Michaels ML, Tchou J, Grollman AP, Miller JH: **A repair system for 8-oxo-7,8-dihydrodeoxyguanine.** *Biochemistry* 1992, **31**:10964-10968.
4. Maki H, Sekiguchi M: **MutT protein specifically hydrolyses a potent mutagenic substrate for DNA synthesis.** *Nature* 1992, **355**:273-275.
5. Tchou J, Kasai H, Shibutani S, Chung MH, Laval J, Grollman AP, Nishimura S: **8-oxoguanine (8-hydroxyguanine) DNA glycosylase and its substrate specificity.** *Proc Natl Acad Sci USA* 1991, **88**:4690-4694.
6. 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 USA* 1992, **89**:7022-7025.
7. Denver DR, Swenson SL, Lynch M: **An evolutionary analysis of the helix-hairpin-helix superfamily of DNA repair glycosylases.** *Mol Biol Evol* 2003, **20**:1603-1611.
8. Slupska MM, Baikalov C, Luther WM, Chiang JH, Wei YF, Miller JH: **Cloning and sequencing a human homolog (hMYH) of the *Escherichia coli* mutY gene whose function is required for the repair of oxidative DNA damage.** *J Bacteriol* 1996, **178**:3885-3892.
9. 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, **38**:6374-6379.
10. Guan Y, Manuel RC, Arvai AS, Parikh SS, Mol CD, Miller JH, Lloyd S, Tainer JA: **MutY catalytic core, mutant and bound adenine structures define specificity for DNA repair enzyme superfamily.** *Nat Struct Biol* 1998, **5**:1058-1064.
11. Bessman MJ, Frick DN, O'Handley SF: **The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes.** *J Biol Chem* 1996, **271**:25059-25062.
12. Parker AR, Eshleman JR: **Human MutY: gene structure, protein functions and interactions, and role in carcinogenesis.** *Cell Mol Life Sci* 2003, **60**:2064-2083.
13. Shinmura K, Yokota J: **The OGG1 gene encodes a repair enzyme for oxidatively damaged DNA and is involved in human carcinogenesis.** *Antioxid Redox Signal* 2001, **3**:597-609.
14. Doi T, Yonekura S, Tano K, Yasuhira S, Yonei S, Zhang QM: **The *Schizosaccharomyces pombe* homolog (SpMYH) of the *Escherichia coli* MutY is required for removal of guanine from 8-oxoguanine/guanine mispairs to prevent G:C to C:G transversions.** *J Radiat Res (Tokyo)* 2005, **46**:205-214.
15. Gu Y, Parker A, Wilson TM, Bai H, Chang DY, Lu AL: **Human MutY homolog, a DNA glycosylase involved in base excision repair, physically and functionally interacts with mismatch repair proteins human MutS homolog 2/human MutS homolog 6.** *J Biol Chem* 2002, **277**:11135-11142.
16. O'Brien PJ: **Catalytic promiscuity and the divergent evolution of DNA repair enzymes.** *Chem Rev* 2006, **106**:720-752.
17. Tajiri T, Maki H, Sekiguchi M: **Functional cooperation of MutT, MutM and MutY proteins in preventing mutations caused by spontaneous oxidation of guanine nucleotide in *Escherichia coli*.** *Mutat Res* 1995, **336**:257-267.
18. Fitzpatrick DA, Logue ME, Stajich JE, Butler G: **A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis.** *BMC Evol Biol* 2006, **6**:99.

doi:10.1186/2041-9414-1-12

Cite this article as: Jansson *et al.*: Evolutionary loss of 8-oxo-G repair components among eukaryotes. *Genome Integrity* 2010 **1**:12.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

