



OPEN

## A minimal Fanconi Anemia complex in early diverging fungi

Drishtee Barua, Magdalena Płecha & Anna Muszewska

Fanconi Anemia (FA) pathway resolves DNA interstrand cross links (ICL). The FA pathway was initially recognized in vertebrates, but was later confirmed in other animals and speculated in fungi. FA proteins FANCM, FANCL and FANCI are present in *Saccharomyces cerevisiae* but, their mechanism of interaction to resolve ICL is still unclear. Unlike *Dikarya*, early diverging fungi (EDF) possess more traits shared with animals. We traced the evolutionary history of the FA pathway across *Opisthokonta*. We scanned complete proteomes for FA-related homologs to establish their taxonomic distribution and analyzed their phylogenetic trees. We checked transcription profiles of FA genes to test if they respond to environmental conditions and their genomic localizations for potential co-localization. We identified fungal homologs of the activation and ID complexes, 5 out of 8 core proteins, all of the endonucleases, and deubiquitination proteins. All fungi lack FANCC, FANCF and FANCG proteins responsible for post-replication repair and chromosome stability in animals. The observed taxonomic distribution can be attributed to a gradual degradation of the FA pathway from EDF to *Dikarya*. One of the key differences is that EDF have the ID complex recruiting endonucleases to the site of ICL. Moreover, 21 out of 32 identified FA genes are upregulated in response to different growth conditions. Several FA genes are co-localized in fungal genomes which also could facilitate co-expression. Our results indicate that a minimal FA pathway might still be functional in *Mucoromycota* with a gradual loss of components in *Dikarya* ancestors.

**Keywords** Interstrand crosslinks, Fanconi Anemia, Early diverging fungi, Ancestral trait, Fungal evolution, Opisthokonta

### Abbreviations

FA	Fanconi Anemia
ICL	Interstrand crosslinks
EDF	Early diverging fungi
DSB	Double-strand break
NER	Nucleotide excision repair
MMR	Mismatch repair
fToL	Fungal tree of life
MMS	Methyl methane sulfonate
HR	Homologous recombination
TLS	Translesion DNA synthesis
TPM	Transcripts per million
AM	Arbuscular mycorrhiza
UBZ	Ubiquitin binding Zinc finger
μM	Micromolar

Interstrand cross-links (ICL) are a lethal type of DNA damage where the two strands of DNA get cross-linked by covalent bonds, thereby halting the progress of DNA replication and transcription<sup>1</sup>. ICL may arise endogenously through nitrous acids and aldehydes, or exogenously through chemotherapeutic drugs such as Cisplatin and Mitomycin C<sup>2</sup>. Studies show that ICL are the major cytotoxic lesions produced by these drugs<sup>3</sup>. If left unrepaired, they stall DNA replication leading to mutations, chromosome breakage and cell death. ICL is repaired by several pathways: Nucleotide Excision Repair (NER), Homologous Recombination (HR) and Translesion DNA Synthesis (TLS) working in coordination<sup>4</sup>. The Fanconi Anemia (FA) pathway recognizes ICL and coordinates the ICL repair process<sup>5</sup>.

Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Pawińskiego 5A, 02-106 Warsaw, Poland.  
 email: musze@ibb.waw.pl

Fanconi Anemia is one of the rare genetic disorders characterized by a progressive bone marrow failure, congenital abnormalities along with cancer predisposition<sup>6,7</sup>. FA disease is linked to the failure of ICL repair<sup>8</sup>. Mutations in FANC genes are the most frequent cause of FA<sup>9</sup>. The FA pathway is activated by ICL during the S phase of the cell cycle in mammals<sup>10</sup> and possibly other eukaryotes<sup>4</sup>. Recent studies confirmed 22 genes to be FANC genes (Table 1). The remaining FA genes encode downstream effectors of the FA pathway and mutations in them do not cause classical FA symptoms<sup>6</sup>. It has been reported that the FA pathway is coupled with Double-strand Break (DSB) repair and NER<sup>11</sup>.

An activation complex of FANCM-FAAP24-MHF1-MHF2 recognizes ICL when a break in the replication process occurs. MHF1 (FAAP16) and MHF2 (FAAP10) form a (MHF1-MHF2) × 2 tetramer, FANCM forms a heterodimer with FAAP24 and together, they bind to MHF1/2 tetramer through a dual V-shaped structure<sup>12</sup>. This anchor complex recruits the FA core complex to sites of DNA damage.

The FA core complex, formed by eight FANC proteins (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, FANCM) and three FA associated proteins (FAAP20, FAAP24, FAAP100), provides the essential E3 ubiquitin ligase function for FANCD2 ubiquitination. FANCC regulates meiotic crossover along with FANCE and

Gene	Alias	Pfam/Interpro accession	FA subcomplex
MHF1	CENPS/ FAAP16	PF15630/IPR029003	Activation
MHF2	CENPX/ FAAP10	PF09415/IPR018552	
FANCA		PF03511/IPR003516	Core complex
FANCB		IPR033333	
FANCC		PF02106/IPR000686	
FANCE		PF11510/IPR039685	
FANCF		PF11107/IPR035428	
FANCG	XRCC9	IPR039684	
FANCI	BRIP1	PF06733/IPR010614	
FANCL		PF11793/IPR026848	
FANCM		PF16783/IPR031879	
FAAP20		PF15750/IPR031490	
FAAP24		PF17949/IPR026985	
FAAP100		PF15146/IPR029251	
FANCT	UBE2T	PF00179/IPR000608	Monoubiquitination
ATR		PF00454/IPR000403	
UHRF1		PF02182/IPR047406	
UHRF2		PF02182/IPR047468	
FANCD2		PF14631/IPR029448	ID complex
FANCI		PF14674-80/IPR026171	Endonuclease
SLX1		PF21202/IPR048749	
FANCP	SLX4	PF09494/IPR018574	
FAN1		PF08774/IPR014883	
EME1		PF21292/IPR033310	
MUS81		PF21136/IPR033309	
FANCQ	XPF	PF02732/IPR006167	
ERCC1		PF03834/IPR047260	
FANCS	BRCA1	PF00533/IPR011364	HR and DNA repair
FANCD1	BRCA2	PF09103-4/IPR015525	
FANCN	PALB2	PF16756/IPR042417	
FANCO	RAD51C	PF08423/IPR013632	
FANCR	RAD51	PF08423/IPR011941	
FANCU	XRCC2	PF08423/IPR030547	
FANCV	REV7	PF02301/IPR003511	
FANCW	RFWD3	PF13639/IPR037381	
REV1		PF16727/IPR031991	
REV3		IPR030559	
DPOLN		PF00476/IPR001098	
USP1	UBP1	PF00443/IPR033815	Deubiquitination
UAF1	WDR48	PF11816/IPR021772	

**Table 1.** Composition of Fanconi Anemia pathway genes, gene name aliases and protein domains in the protein products of FA genes (subcomplex nomenclature after Niraj and co-workers<sup>7</sup>).

FANCF<sup>13</sup>. The FANCA-FANCG-FAAP20 aids in the process of strand annealing and strand exchange<sup>14</sup>. FANCM, a multi-domain protein, binds to the core complex by the interaction of its MM1 domain with FANCF. Its MM2 domain on the other hand, interacts with the BTR complex and participates in homologous recombination<sup>15</sup>. The molecular functions for the remaining core proteins remain elusive, apart from binding to DNA.

The monoubiquitination of FANCD2 and FANCI, forming the ID complex, is the hallmark of the FA pathway. This ID complex facilitates recruitment of the downstream effector proteins to ICL<sup>8</sup>. The process involves attachment of a single ubiquitin molecule to a specific site in FANCD2 and FANCI<sup>16</sup> and is carried out by FA core complex together with UBE2T. The E3 ligase FANCL acts as a catalyst, with its C-terminal RING domain binding to UBE2T and a central double RWD domain that binds to FANCD2<sup>8</sup> to carry out the ubiquitination. The RWD domain is also said to stimulate the activity of UBE2T<sup>17</sup>. Along with UBE2T, ATR phosphorylates FANCD2 and FANCI via its effector kinase Chk1 to stabilize its association with DNA and FANCD2<sup>18</sup>. A recent study identified a novel ICL sensor protein—E3 ubiquitin ligase UHRF2, which along with its paralogue UHRF1, interacts with FANCD2 after its recruitment in the DNA and facilitates the retention of FANCD2 to the site of ICL<sup>19</sup>.

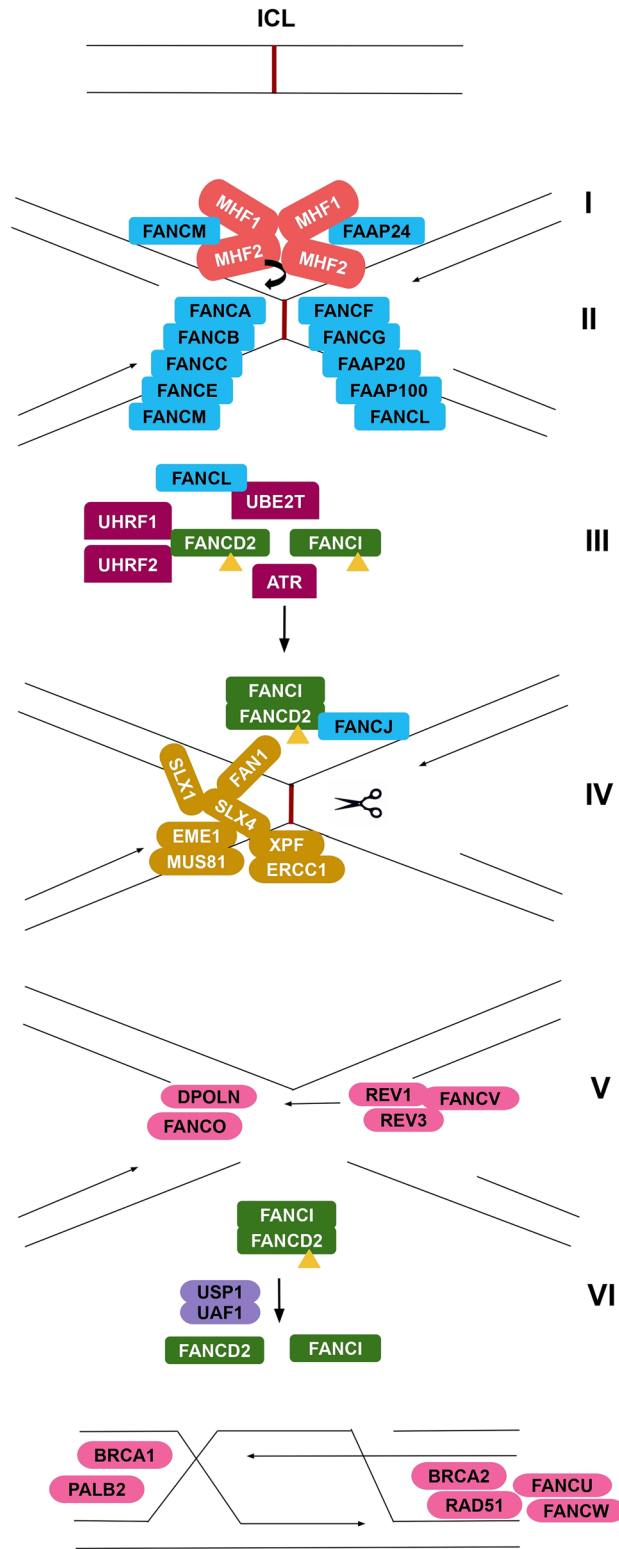
The downstream effector proteins comprises endonucleases and repair proteins that contribute to ICL repair. Coordination of nuclease activity begins with SLX4 (FANCP) that functions as a scaffold, modulator and a docking platform for SLX1, MUS81-EME1 and XPF-ERCC1 structure-specific nucleases<sup>5,20</sup>. The FAN1 nuclease contains a UBZ domain in its N-terminus that acts as a platform for binding to FANCD2. The monoubiquitinated ID complex recruits this endonuclease complex to the site of ICL that carries out incisions to unhook the crosslinked bases<sup>16</sup>. Studies identified that XPF-ERCC1 is the most important for resistance to ICL<sup>5</sup>. This process of unhooking is necessary for initiation of TLS. The FANCI helicase binds to FANCD2 and regulates chromatin localization<sup>21</sup>.

The TLS polymerase consisting of REV1, REV3-REV7 or Pol ζ complex and DPOLN (DNA polymerase  $\nu$ ) is subsequently recruited to carry out lesion bypass. While REV1 inserts the first nucleotide opposite a damaged base, REV3 performs the extension step. The polymerase REV3 is particularly specialized in extending distorted base pairs (such as mismatches due to inaccurate base insertion by another TLS polymerase). REV7, a newly identified FA gene, acts as an adaptor between REV1 and REV3<sup>22</sup>, and together they perform lesion bypass and are considered the main players of post-replication repair<sup>23</sup>. DPOLN (DNA polymerase  $\nu$ ), a Y-family polymerase protein, is able to carry out insertions as well extensions in a diverse set of minor and major groove ICL with no stalling of replication forks<sup>24</sup>. In the case of TLS with A-family polymerase proteins (such as pol $\eta$ ), a monoubiquitinated PCNA is required for recruiting pol $\eta$  to the site of break and ensure accurate replicative bypass TLS activation<sup>25</sup>. TLS involving DPOLN is controlled by the FA core complex and is independent of the PCNA ubiquitination event in mammals<sup>26</sup>. Proteins FANCO, FANCU and RAD51 are components of the RAD51 paralogue complex BCDX2<sup>27</sup>. The BCDX2 complex contains RAD51B, RAD51C, RAD51D, and XRCC2; it binds to single-stranded DNA, nicks in duplex DNA and to single-stranded regions in duplex DNA<sup>28</sup>. FANCO forms RAD51 foci in response to the damage<sup>29</sup>. It facilitates phosphorylation of the checkpoint kinase CHEK2 and thereby transduction of the damage signal, leading to cell cycle arrest and HR activation<sup>26</sup>. XRCC2, classified as an FA gene (FANCU), stabilizes RAD51<sup>27</sup>. The FA components, BRCA1, BRCA2 and PALB2 take part in the subsequent steps of DNA repair involving HR<sup>5</sup>. BRCA1 functions upstream of BRCA2, and recruits PALB2<sup>30</sup>. BRCA2 helps load RAD51 to the DNA by interaction of its BRC4 repeat with the RecA domain of RAD51<sup>31</sup>. PALB2 binds directly to both BRCA1 and BRCA2, and together they remove the CMG helicase from stalled replication forks<sup>7</sup>.

The end of the FA pathway is marked with the deubiquitination of FANCD2/ FANCI heterodimer by the USP1-UAF1 complex<sup>32</sup>. It is carried out by the interaction of deubiquitylating enzymes USP1 and UAF1 with FANCD2 in the chromatin. The ID complex is deubiquitylated by USP1 whose activity is enhanced by the interaction with UAF1<sup>33</sup>.

ICL repair by FA pathway is best studied in humans<sup>2</sup> and mice<sup>35</sup> (Fig. 1). However, studies show that FA core binding proteins are present in all classes of sponges, pointing to the ancient origin of the FA pathway in animal evolution<sup>36</sup>. The nematode *Caenorhabditis elegans* has a functional ortholog of the DEAD-box helicase FANCI, termed as DOG-1 (Deletions of G-rich DNA) that functions alongside FANCD2, FANCM, FANCO, FANCI proteins and maintains the stability of its G-rich DNA<sup>37</sup>. The function of FA proteins was also recently studied in *Drosophila melanogaster*, where the monoubiquitination of FANCD2 was found to be linked to a mitosis-specific DNA DSB response<sup>38</sup>. Proteins MHF1 and MHF2, were previously identified as anti-crossover factors during meiosis in *Arabidopsis thaliana*<sup>13</sup>. Singh and colleagues identified FANCC as another anti-crossover gene and showed that FANCC, FANCE and FANCF subcomplex was conserved from vertebrates to plants and it regulates meiotic recombination<sup>13</sup>. There have also been attempts to study the FA pathway in non-animal organisms. For instance, studies in *Saccharomyces cerevisiae* report the presence of putative homologs of FA activation proteins MHF1 and MHF2 (CENPS and CENPX respectively) as well as FA core binding proteins FANCM, FANCI and FANCP (MPH1, CHL1 and SLX4 respectively)<sup>4,39</sup>. However, yeast cells deficient in aforementioned protein-coding genes displayed no significant sensitivity to ICL. Instead, they require a combination of different DNA repair systems: NER, HR and post-replication repair to mitigate ICL<sup>4</sup>.

In light of these studies, there is substantial evidence of the FA pathway beyond mammals, with some missing proteins. The discoveries in amoebozoia, non-mammalian animals and fungi led us to ask questions about the existence of FA pathway in ancestral Opisthokonts and early diverging fungal (EDF) lineages. Here, we conduct a genomic survey of FA proteins across the fungal tree of life. We propose a hypothetical model of FA pathway for distinct evolutionary lineages and recapitulate the evolutionary history of FA complexes across taxa.



**Homologous recombination**

**Figure 1.** Schematic representation of the Fanconi Anemia pathway in humans, derived from<sup>34</sup>. [I—Activation of FA pathway; II—Binding of FA core complex to the ICL site; III—Monoubiquitination of FANCD2 & FANCI to form the ID complex; IV—Formation of endonuclease complex and subsequent ICL unhooking; V—Translesion synthesis or lesion bypass; VI—Deubiquitination of ID complex and closure of FA pathway.]

## Results

### Distribution of proteins

Genes encoding proteins from the FA pathway are conserved in evolution because of their involvement in cellular homeostasis. To recover the distribution of individual components of FA pathway, we mapped 40 reference proteins on a collection of 183 fungal and five early diverging *Opisthokonta* proteomes (protein sets derived from whole genome sequencing projects) (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). Out of the aforementioned 40 FA selected reference sequences, 32 have homologs in fungal proteomes (Fig. 2, Supplementary Results).

### Transcriptomics of FA proteins

#### *FA proteins expressed in pure culture transcriptomic studies of EDF*

Since most of the identified homologs of FA components are annotated as hypothetical unknown proteins without experimental characterization, we looked for their gene expression levels in the available EDF whole transcriptomes. The number of transcripts found serves as an intermediate evidence that the predicted FA homologs in fungi originate from active genes in normal conditions. The analyzed transcriptomes corroborate the expression of 27 out of 32 identified FA homologs in representatives of *Mucorales* (*M. lusitanicus*, *J. flammicorona*), *Endogonales* (*Endogone* sp.), *Umbelopsidales* (*U. isabellina*) and *Mortierellomycotina* (*L. transversale*). For instance, *M. lusitanicus* expressed four out of eight core binding genes (FANCA, FANCE, FANCL, FANCM) and one ID complex gene FANCD2 predicted for *Mucoromycotina*. The activation genes MHF1 and MHF2 and the monoubiquitination gene UBE2T were also expressed, along with endonucleases SLX1, MUS81, FAN1 and repair genes REV1, REV3, FANCV, BRCA2 and RAD51. The *Mortierellomycotina* *L. transversale* expressed copies of MHF2, UBE2T, ATR, FANCD2 and FANCI, FANCL, FANCM, FANCO, SLX1, MUS81, FAN1, REV1, REV3, FANCV, DPOLN, FANCW and BRCA1 genes. The transcriptomes of *Endogone* sp. expressed UHRF1, MUS81, ERCC1 and RAD51 and *J. flammicorona* expressed FANCI, ATR, DPOLN and FANCW, while *U. isabellina* expressed FANCD2, UBP1, REV1 and two copies of UBE2T. (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1).

#### *FA proteins expressed in condition-specific transcriptomic studies of EDF*

In order to test the functionality of the FA candidate genes, if they are expressed and regulated depending on environmental conditions, we analyzed expression profiles of genes corresponding to the proteins with available

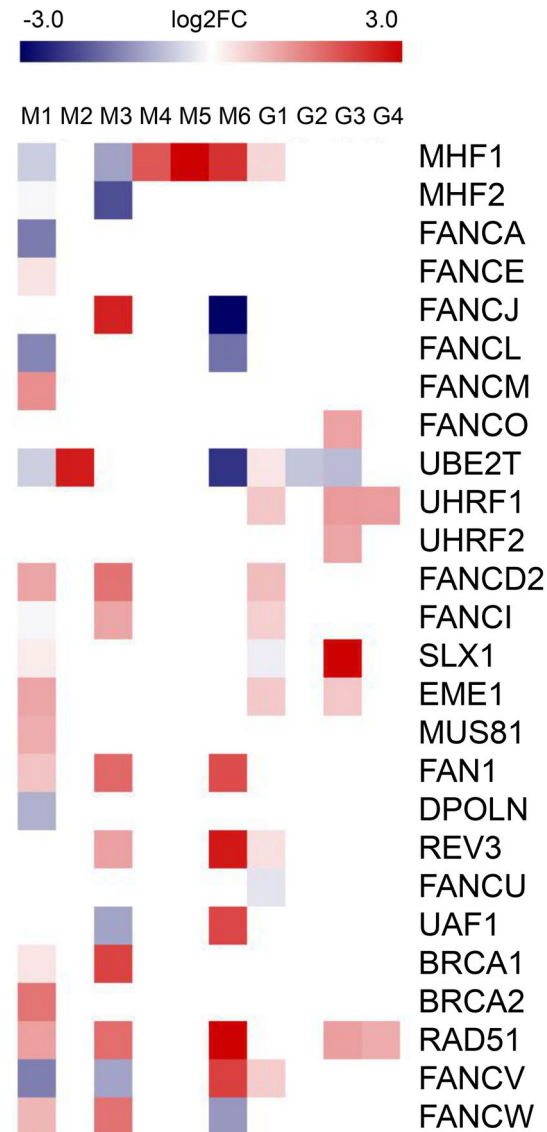


**Figure 2.** Distribution of 40 FA and FA associated components among selected eukaryotes.



transcriptomic data obtained from various environmental conditions. A total of 26 out of 32 genes encoding homologs of FA pathway components were expressed under multiple conditions. This result suggests that the genes are actively regulated and may be involved in stress responses.

Our analysis of *M. lusitanicus* (M1) transcripts expressed under anaerobic vs aerobic growth conditions<sup>40</sup> identified 19 out of 26 predicted FA genes. Differential expression analysis showed upregulation in the FA core binding and ID complex genes: FANCD2, FANCE and FANCM, along with endonucleases SLX1, EME1, MUS81, FAN1 and repair genes BRCA1, BRCA2, RAD51 and FANCV. The activation complex genes MHF1 and MHF2, ubiquitylating gene UBE2T, 3 FA core genes FANCA, FANCI and FANCL along with DPOLN and FANCV repair genes were downregulated (Fig. 3). Transcriptomics of *R. delemar* (M3) during exposure to murine macrophages<sup>41</sup> showed upregulation of FANCD2, FANCI, FANCI, FAN1, REV3, BRCA1, RAD51 and



**Figure 3.** Differential expression profiling of 26 out of 32 FA pathway genes expressed in different environmental conditions in members of *Mucoromycotina* and *Glomeromycotina*. The empty cells signify the absence of gene expression from the organism in the given environmental condition. M1: *Mucor lusitanicus* growth in anaerobic vs aerobic conditions; M2: *Rhizopus microsporus* growth in presence vs absence of murine macrophage; M3: *Rhizopus delemar* growth in presence vs absence of murine macrophage; M4: *Rhizopus delemar* host-pathogen interaction (6 h) using human airway epithelial cells (A549); M5: *Rhizopus delemar* host-pathogen interaction (16 h) using human airway epithelial cells (A549); M6: *Rhizopus delemar* host-pathogen interaction (18 h) using mouse bone marrow-derived macrophages (BM); G1: *Gigaspora rosea* response to plant signals in the switch from asymbiotic to presymbiotic growth; G2: *Rhizophagus irregularis* growth in *Lotus japonicus* roots exposed to different concentrations of phosphate (20  $\mu$ M, 100  $\mu$ M, 300  $\mu$ M, 500  $\mu$ M); G3: *Rhizophagus irregularis* association with *Medicago truncatula* and treatment with strigolactone for 24 h; G4: *Rhizophagus irregularis* association with *Medicago truncatula* and treatment with strigolactone for 1 week.

FANCW and downregulation of MHF1 and MHF2, UAF1 and FANCV (Fig. 3). On the other hand, a similar study conducted on *R. microsporus* (M2) displayed upregulation of only UBE2T. Transcriptomic studies on *R. delemar* were also found centering human host–pathogen interaction against mucormycosis. In one of the studies, A549 airway epithelial cells were subjected to in vitro infection with *R. delemar* for 6 h (M4) and 16 h (M5)<sup>42</sup>. The MHF1 gene was found to be upregulated at both timepoints. In another study, mouse bone-marrow derived macrophages (BMDMs) were infected with *R. delemar* for 1 h, 4 h and 18 h<sup>43</sup>, and we found the expression of FA genes from the samples infected for 18 h (M6). The genes MHF1, FAN1, REV3, RAD51, UAF1 and FANCV were upregulated while FANCI, FANCL, UBE2T and FANCW were downregulated (Fig. 3).

The study involving *G. rosea* treated with G24 (G1), a synthetic analog of strigolactone, addressed its gene regulation in response to plant signals in the switch from asymbiotic growth to presymbiotic growth<sup>44</sup>. Our data analysis displayed upregulation of MHF1, UBE2T, UHRF1, FANCD2, FANCI, EME1, REV3 and FANCV respectively with multiple copies of UBE2T. SLX1 and FANCU were downregulated (Fig. 3).

Two different transcriptomic studies were obtained for *R. irregularis*. One study concerned the transcriptomic profiling of AM roots exposed to varying concentrations of phosphate (20  $\mu$ M, 100  $\mu$ M, 300  $\mu$ M, 500  $\mu$ M) (G2)<sup>45</sup>, where multiple copies of UBE2T were downregulated. Another study focused on the differential expression of *R. irregularis* during AM symbiosis by harvesting strigolactone-treated spores a day (G3) and a week (G4) after inoculation respectively<sup>46</sup>. Our analysis revealed that spores harvested a day post inoculation had upregulation of FANCO, UHRF1, UHRF2, SLX1, EME1 and RAD51, and downregulation of UBE2T, while on the other hand, UHRF1 and RAD51 were upregulated in the spores harvested a week post inoculation (Fig. 3).

We also checked the expression profiling of genes involved in the NER, MMR and DSB repair pathways in order to find their correlation with the down- or upregulation of FA genes. All of the NER, MMR and DSB genes were upregulated in *R. delemar*, *M. lusitanicus* and *R. microsporus*, regardless of the environmental conditions in contrast to the results found for FA genes. However, in the case of *G. rosea* and *R. irregularis*, genes involved in DSB repair were particularly downregulated, while those involved in NER and MMR were upregulated.

#### FA genes not expressed in transcriptomic analysis

Generally, at least one gene from every FA pathway subcomplex was detected in the transcriptomes of pure culture and treated fungal species. The transcriptomes obtained from axenic cultures of *Mucoro-* and *Glomeromycotina* members did not detect the genes encoding endonucleases: EME1, SLX4 and XPF (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). While on the other hand, transcriptomes obtained from organisms subjected to condition-specific expression did not detect genes encoding ATR, XPF, ERCC1 and SLX4, REV1, and UBP1 (not shown in Fig. 3).

#### Genomic co-occurrence of FA genes

Fungal genes involved in one metabolic process are often clustered in fungi<sup>47</sup>. We found that several genes involved in FA pathway are located on the same contig. Usually, only two or three FA genes were colocalized, however, we found up to ten genes co-occurring on a single contig in selected *Ascomycota* (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). Colocalization with another FA gene within a distance of 250 kb was observed for 18% of *Dikarya* FA genes and for 19% of EDF genes encoding FA proteins. The distance of 250 kb was chosen as a threshold based on the occurrence of biologically important interactions (such as contacts between enhancers and promoters) in this range<sup>48</sup>. A hypergeometric distribution test performed on 56 genome datasets showed that the observed co-occurrences of FA genes are not expected at random (with the highest *p* value = 0.009) (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). Co-occurrence of proteins in distances greater than 250 kb were classified into long-range colocalization patterns<sup>49</sup>. FA genes were not localized on the same contig with any other FA gene in a minority of taxa (7/36 *Dikarya* and 30/116 EDF) (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). Not all FA genes are equally likely to co-occur with others; we did not observe any proximity to other FA genes for FANCA, FANCE, UHRF1, UHRF2, BRCA1, REV1 and UAF1 within 250 kb distance range (Fig. 4A). FA genes in *Chytrids* and *Blastocladiomycota* co-occurred in groups of four, the distances among each gene ranging between 200 and 300 kb (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1).

In *Dikarya*, we mainly observed long-range colocalization between FA genes. However, genes UHRF1, UHRF2, DPOLN, REV1 and UAF1 did not co-occur even in large distances for all fungal groups (Fig. 4B).

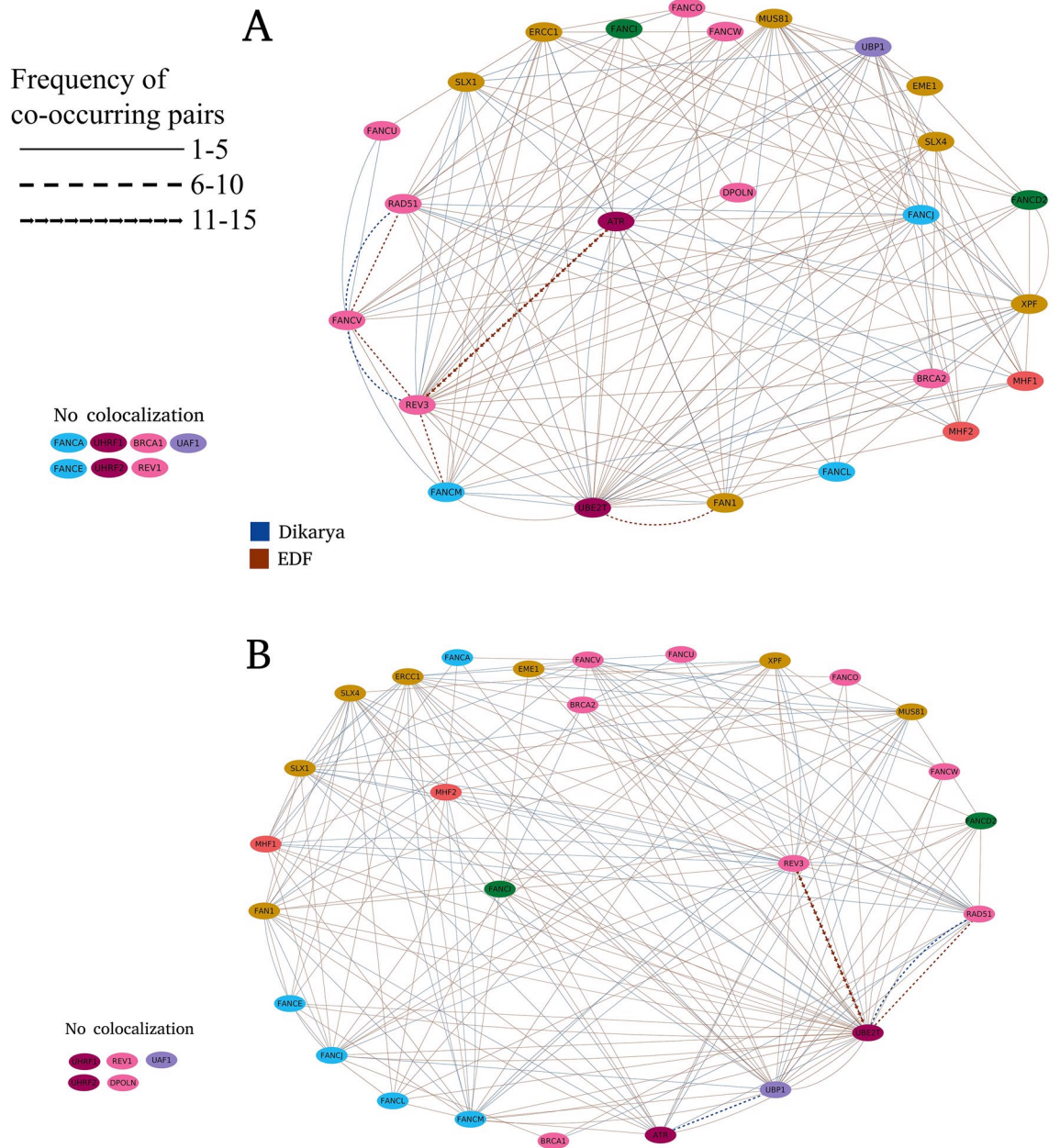
Genes coding proteins which directly interact, for instance FANCD2 and FANCI, were not located together in any of the analyzed genomes. However, these genes tend to co-occur with a ubiquitination, an endonuclease or a deubiquitination gene across all EDF.

Endonucleases are the products of genes most often co-occurring with other FA components, particularly the genes for REV3 and UBE2T or ATR are located close to each other (Fig. 4).

Four *Chytridiomycota* representatives have two-domain FA proteins: an N-terminal FANCM domain and a C-terminal XPF endonuclease (TPX77543.1, *Chytriomycetes confervae*; TPX61370.1, *Powellomyces hirtus*; ORY52110.1, *Rhizoclostridium globosum*; KNC96281.1, *Spizellomyces punctatus* DAOM BR117) (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1).

#### Differences in domain architecture

For 30 out of 32 predicted FA proteins, fungal sequences clustered together with the animal reference sequences. However, SLX4 and FANCI sequences formed separate clusters. Further analysis of these proteins revealed that SLX4 and FANCI homologs display a few differences in their lengths and domain composition. This points to a possible change in protein function or specificity in fungi.

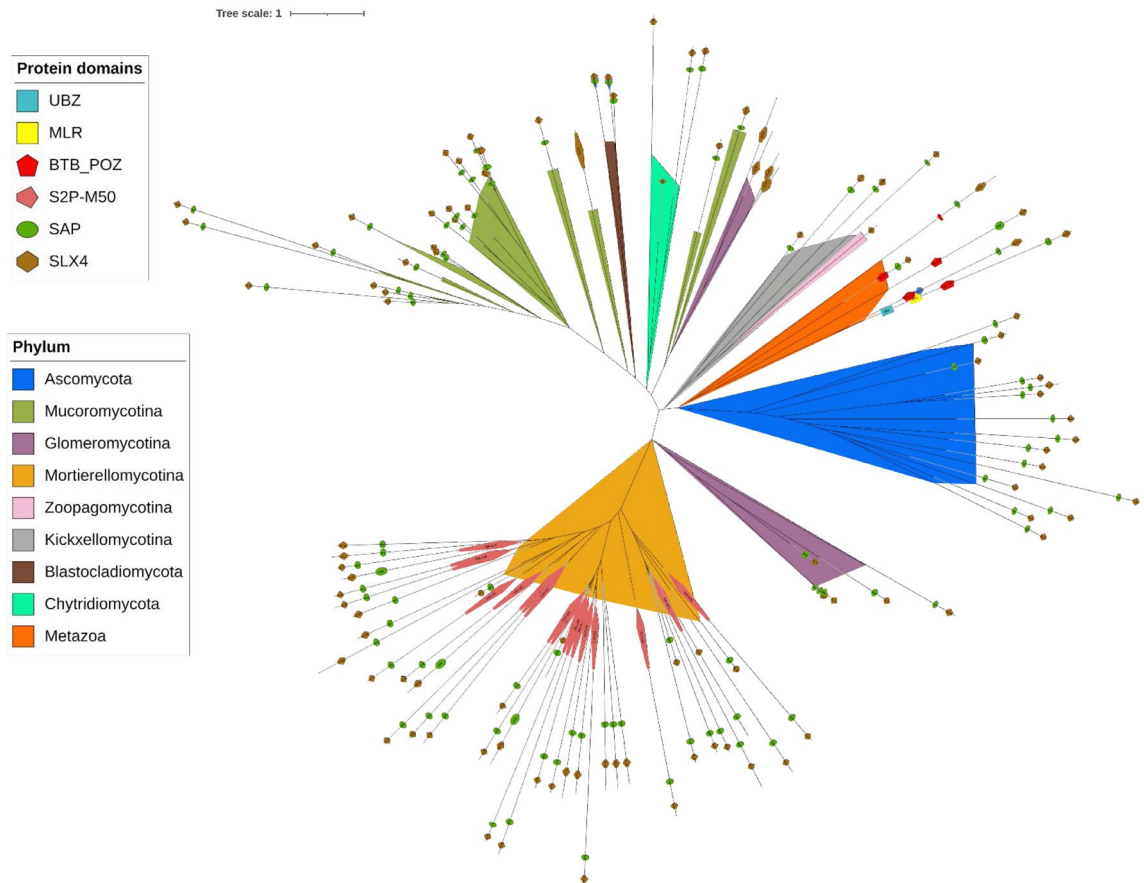


**Figure 4.** Co-occurrence of FA genes within distances of (A) 250 kb and (B) greater than 250 kb observed in *Dikarya* and EDF assemblies.

We found that SLX4 harbors an N-terminal SAP domain and a C-terminal SLX4 domain across diverse fungal lineages (Fig. 5). On the other hand, the human SLX4 is a 1834 amino acid long protein which, apart from the two domains, contains three additional domains in its N-terminus: UBZ, MLR and BTB/POZ. The BTB/POZ domain is present in SLX4 homologs of animals including sponges, but absent from all fungal homologs. The UBZ and MLR domains are also absent in fungi. Moreover, the fungal homologs are significantly shorter than their animal homologs with lengths ranging from 151 aa in *Allomyces macrogynus* (KNE55791.1, *Blastocladiomycota*) up to 423 aa in *Rhizophagus irregularis* (POG76120.1, *Glomeromycotina*). Mortierellomycotina SLX4 homologs are as long as the ones in animals (for instance, SLX4 of *Haplosporangium bisporale* KAF8951561.1, is 1993 aa long) and they also contain an additional N-terminal S2P-M50 domain.

Differences were also observed in the domain architecture of FANCI. The FANCI is a 5'-3' DNA binding helicase with conserved N-terminal DEAH/DEAD box helicase and C-terminal RAD3 helicase domains. Multiple sequence alignment of FANCI homologues in EDF points towards a deletion of the DEAD box domain. FANCI homologs in *Opidiomycota* and *Microsporidia* are significantly truncated with a solo DEAD/DEAH box domain without the C-terminal RAD3 helicase. These deviations from the core FANCI domain architecture might impair its function.





**Figure 5.** An unrooted Maximum Likelihood phylogenetic tree of SLX4 scaffold protein across selected eukaryotic lineages, drawn in iTOL<sup>50</sup>.

### Phylogenetics of Fanconi Anemia proteins

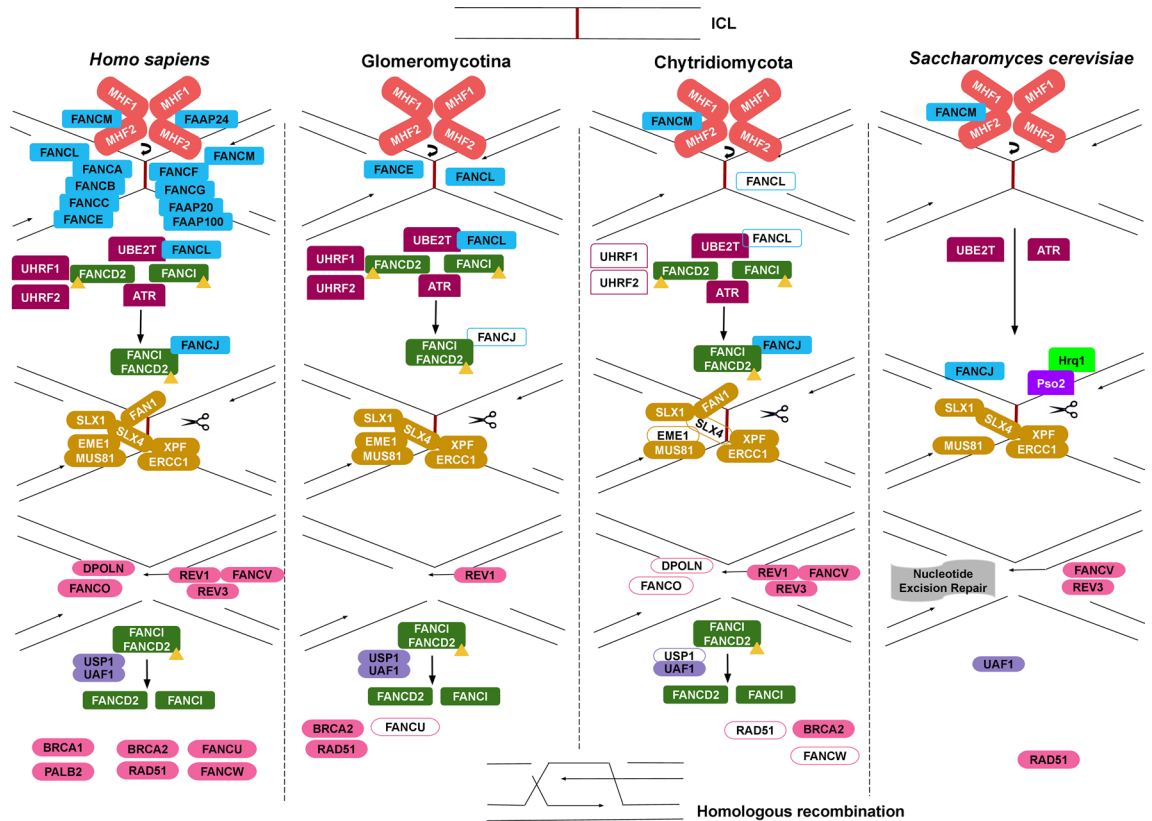
Phylogenetic trees for 32 FA proteins are in agreement with the species tree, pertaining to the vertical inheritance of housekeeping genes in eukaryotes (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Materials SM1, Supplementary Dataset SD1). Most FA proteins occur as a single copy per proteome. A conserved domain architecture was observed in all of the analyzed Opisthokonts, with exceptions in FANCI helicase and SLX4 scaffold protein. In addition to the differences in protein length as well as domain architecture, gene duplications were observed in members of *Blastocladiomycota*, *Entomophthoromycotina* and *Mucoromycotina* respectively. For instance, in *Allomyces macrogynus*, 2 copies were observed for proteins FANCL, FANCM, FANCO, ATR, UHRF1, UHRF2, SLX4, FAN1, whereas proteins FANCI, and ERCC1 had four copies each. A series of duplications were observed in 6 FA proteins in *Entomophthora muscae*, namely FANCD2, FANCL, UBE2T, SLX1, ERCC1 and FAN1. The MUS81 endonuclease was found in 10 copies in *J. flammicorona*. Duplications were also observed in EDF species containing the ID proteins (FANCD2-FANCI), FANCI helicase and endonucleases MUS81, SLX1 and FAN1.

Taking into account all the homologs of FA components found in EDF, a model of minimal FA machinery could be proposed (depicted as an example in *Glomeromycotina* and *Chytridiomycota*) (Fig. 6). The presence of FANCM-MHF1-MHF2 activation complex could signal the existing core proteins FANCL and FANCI to bind to the ICL site, bringing together the ubiquitination proteins. This process would be followed by monoubiquitination of FANCD2 and FANCI forming the ID complex, which along with FANCI, would enable the build-up of endonucleases around the site of ICL. The subsequent unhooking of ICL would be followed by TLS and deubiquitination of the ID complex with USP1-UAF1 complex. The final step would be complemented with the HR proteins along with proteins from NER and DSB repair pathways. The absence of ID proteins in *Dikarya* might enable Pso2 nuclease and Hrq1 helicase to associate with FANCI helicase and endonuclease components of FA in order to complete the unhooking of ICL.

### Discussion

In this study we show that the FA pathway is fragmentarily present in basal Opisthokonts like *S. rosetta*, *C. owczarzakii*, *M. brevicollis* and *S. arctica*, as well as in many fungal lineages, especially EDF. Recently, diverse traits were reported to exist in EDF long after their separation from the common ancestor with animals<sup>52</sup>. Now, the FA pathway extends the list of known ancient pathways retained in fungi.

The work of Tao and coworkers revealed that MHF 1/2 functioning requires a stable association with FANCM<sup>12</sup>. FANCM homolog in *Archaea* (Hef protein) resolves replication forks and possesses both helicase and



**Figure 6.** Speculative FA pathway model for *Glomeromycotina* and *Chytridiomycota* compared to *Saccharomyces cerevisiae*<sup>51</sup> and *Homo sapiens*<sup>7</sup>. The filled shapes indicate presence of the protein in all members of the group; the shapes outlined indicate presence of the protein in 50% of the members of that group.

endonuclease activities<sup>53</sup>. The FANCM ortholog of *Schizosaccharomyces pombe* (Fml1), also has FA-independent roles in DNA damage response. FANCM-MHF worked in parallel with EME1-MUS81 to process meiotic joint DNA molecules limiting crossovers<sup>54</sup>. On the other hand, the FANCM ortholog in *A. thaliana* was found to have no direct role in ICL DNA repair, but necessary in HR pathways in somatic cells<sup>55</sup>. FANCM also promotes RAD51-dependent gene conversion at stalled replication forks<sup>56</sup>. The absence of FANCM in *Glomeromycotina* hence, puts an open question on their mechanism of FA activation and subsequent DNA repair process. The endonuclease complex is the most conserved across all lineages; this is consistent with their involvement in universal repair pathways beyond FA. FA-specific proteins forming the core binding complex have a patchy taxonomic distribution with the whole set limited to mammals. The currently known ensemble of core binding proteins likely originated in the ancestor of vertebrates. The work of Alpi et al. (2008) demonstrated that FANCL is the crucial E3 ligase subunit and that UBE2T, FANCL and FANCI were sufficient for robust FANCD2 monoubiquitination<sup>17</sup>. They also speculated that flies, worms and *Dictyostelium* may possess a simplified FA monoubiquitination pathway with just three components: UBE2T, FANCL, and FANCI. The conservation of FANCD2, FANCI, UBE2T and FANCL in EDF (Fig. 2), along with their expression during stress conditions (Fig. 3) provides evidence of a simplified, functioning core complex.

Ubiquitination of both FANCD2 and FANCI forms a dual ubiquitin-locking mechanism needed for the ID complex function<sup>57</sup>. The ID complex is retained in EDF but lost in *Dikarya* (Fig. 2). Its absence in *Dikarya* might point to the lack of sensitivity to ICL in yeasts. It is also possible that other DNA repair systems are involved in ICL repair in this group of organisms.

A significant part of the FA pathway is the process of unhooking, covered mostly by the SLX1-SLX4 complex. Studies of SLX4 in humans have pointed out that loss of the UBZ domain (which recruits SLX4 to the site of ICL) in SLX4 would abolish FA-independent interactions, increasing the number of chromosomal aberrations<sup>58</sup>. The absence of the UBZ domain in members of the fungal kingdom leaves a gap in knowledge of the docking sites for SLX4 (Fig. 5). Interestingly, SLX4 genes were expressed neither in normal nor in stress conditions.

The transcriptomic analysis carried out in this study showed that SLX1 is highly upregulated in *Rhizophagus irregularis* associated with *Medicago truncatula* treated with strigolactone for 24 h.

This result suggests that SLX1 would partner with another inactive nuclease for substrate binding instead of SLX4 or that SLX4 would function differently in EDF than in mammals. The upregulation of FA core and repair genes during exposure of *R. delemar* and *R. microsporus* spores to phagocytosis (murine macrophages) leads to active catabolic pathways, nucleic acid binding, transcription and regulation via polymerase, indicating the involvement of genome maintenance pathways<sup>41</sup>. The association of AM fungi with plant roots triggers plant host defense mechanisms. We recognized downregulation of FA genes in *R. irregularis* in response to its

colonization to *Lotus japonicus* roots at different phosphate concentrations of 20, 100, 300, and 500  $\mu\text{M}$  (Fig. 3). It indicates that higher phosphate concentrations silence the FA gene expression, coupled with decrease in AM fungal colonization in *Glomeromycotina*.

Yeasts lack most of the FA pathway components and deal with the ICL threat by a replication and recombination-independent mechanism. In such a case, endonucleolytic unhooking depends on the Pso2 nuclease facilitated by Hrq1 helicase<sup>51</sup>, which is followed by translesion synthesis. This pathway may be responsible for the removal of ICL in FA pathway deficient cells<sup>59</sup>. We found Pso2 homologs in all classes of fungi, except *Microsporidia*. This opens up the possibility of an alternative for FA pathway in most of the fungal taxa. The presence of Pso2 is the basis of our model for ICL repair in *Saccharomycotina* and could, in general, be extended to *Dikarya* (Fig. 6).

FANCI consists of N-terminal DEAH/DEAD box helicase and C-terminal Rad3 helicase domains. DEAD box helicase domain is missing from some of the FANCI homologues in EDF. This might be a valid domain truncation or gene calling error missing the first exon in some of the taxa. Studies report that the DEAH/DEAD box domain binds to helix extensions, helps in target recognition and unwinding<sup>60</sup>. Lack of DEAD box helicase may reduce the catalytic activity of FANCI in some of the EDF. However, it is also possible that Rad3 helicase domain alone can perform the necessary function.

The repair of ICL DNA damage undergoes a final step involving the vertebrate—specific DNA polymerase  $\nu$  (DPOLN). DPOLN depletion results in low HR efficiency and an increased sensitivity to ICL causing agents<sup>23</sup>. With the occurrence of DPOLN in *Chytridiomycota*, we can speculate that this protein is of ancestral origin. A study by Huang and Cook points out that the rate of HR mediated DNA repair is not the same across fungal species<sup>61</sup>. It is likely that the uneven distribution of HR proteins across the fungal tree explains their observation.

The duplications observed in the phylogenetic trees of FA proteins are likely attributed to whole genome duplications or polyploidy in a given taxon (like in the case of *A. macrogynus*), a bigger genome size (240 Mb genome size in the case of *J. flammicorona*) and occurrence of genomes full of repetitive elements and potential functional diploidy (in case of *E. muscae*)<sup>62</sup>.

We observed previously unreported FA components to be conserved beyond fungi and mammals. *D. discoideum* is remarkably resistant to DNA damage<sup>63</sup> and was previously reported to have a minimal FA pathway consisting of FANCD2, FANCI, FANCL, FANCM, FANCI and UBE2T, components which possibly evolved in the last eukaryotic common ancestor<sup>64</sup>. We additionally found other FA members: MHF1, FANCO, ATR, FAN1, MUS81, XPF, ERCC1, REV1, REV3, USP1 and UAF1 (Fig. 2) in the social amoeba proteome (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1), which supports the hypothesis of the early evolutionary origins of FA pathway in general. Another experimental study on *D. discoideum* proved that excision repair nuclease XPF was necessary to repair ICL<sup>64</sup>. Also, early animals including sponges possess a relatively complete FA pathway (Fig. 2).

Interestingly, in *Arthropods*, we observed a rather depleted FA protein repertoire. The presence of FA core binding proteins: FANCI, FANCL, FANCM, with conserved functions of FANCD2 and FANCI in *D. melanogaster* points to a reduction of FA pathway in arthropods<sup>38</sup>. It was also found that in flies, each nuclease can act individually, without the need to form a complex during ICL repair<sup>65</sup>. Based on our results in nematodes, we may speculate there is a functional equivalent of the FA pathway for ICL repair due to the conservation of activation proteins, ID complex proteins, along with FANCI, FANCM and FANCO, all of which are essential for pathway progression and ICL repair.

The co-occurrence of FA pathway genes in fungal genomes suggests their involvement in a common network. The genomic proximity in fungi is often linked with co-expression of genes which are needed by the cell to work in one process (Fig. 4). This non-random organization of FA genes is another confirmation of a possible functionally active FA repair system for ICL in EDF.

The localization of ID complex genes with ubiquitins, endonucleases and DNA repair genes stands true with the chronology of events (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). This is particularly visible in *Mucorales* and *Mortierellales*.

Despite the divergence of *Dikarya* around 650 million years ago<sup>66</sup>, the genomic proximity of FA components is preserved, particularly in *Ascomycota* (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). The co-localization of FA genes across the fToL implies an evolutionary pressure exerted on these genes to cluster together and maintain a functional pathway, perhaps for genome maintenance.

The existence of two FA pathway components fused in a single protein in most members of *Chytridiomycetes* (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1), supports the existence of a gene fusion event around *Chytrid* evolution. However, the proteins are atypically long (1800–1900 aa) and there is only one transcript in EST database supporting such a gene model (XM\_016756588.1: *Spizellomyces punctatus* DAOM BR117) with the presence of both protein domains in a single protein. This does not rule out the possibility of gene fusion since transcriptomic data for *Chytridiomycota* is scarce. Regardless of the robustness of gene fusion, there is genetic proximity between the FANCM and XPF endonuclease in this taxon.

The lack of genomic proximity of FA core genes in *Glomeromycotina* can be attributed to high activity and proliferation of transposons in their genomes leading to genome reshuffling. The high interspecific diversity in the genomes of *Glomeromycotina* affects all known protein domains<sup>67</sup>.

Observed differences in FA conservation among *Opisthokonta* were summarized in the proposed model of FA repair pathway in *Glomero-* and *Chytridiomycota* groups (Fig. 6). We speculate the presence of a minimal FA pathway in early diverging fungi that promotes ICL repair. The expression of genes encoding proteins from NER, DSB and MMR pathways co-occurs with the expression of FA components, opening a possibility of coordination of these pathways to maintain genome stability in EDF. The presence of FANCI and FANCL in EDF, proteins that play a crucial role in mediating the monoubiquitination and formation of the ID complex, could be sufficient to kick-start an active FA pathway. The absence of the above proteins coupled with the absence of

an ID complex in *Dikarya* could be the answer to insensitivity of ICL reported in yeasts, despite the discovery of FA homologs<sup>4</sup>. Numerous animal models for FA pathway have been developed and yet, no model has come up regarding the dynamics of FA pathway proteins in the fungal kingdom. Identification of putative homologs of FA proteins in EDF groups could pave the way for understanding the biology of ICL repair in the fungal tree of life.

Taken together, our study points to the ancient origin of the FA pathway and its conservation beyond mammals. At the same time, the pathway was shaped by massive gene loss in model animals and *Dikarya*. We hypothesize the existence of a minimal form of Fanconi Anemia pathway in the early diverging fungal lineages.

## Methods

FA pathway reference protein list was built based on keywords “fanconi”, “Fanconi Anemia”, “FANC” searched in the following databases: UniProt<sup>68</sup>, Reactome version 83<sup>69</sup>, Interpro<sup>70</sup> and OrthoDB<sup>71</sup>. This resulted in a list of 40 FA proteins. The dataset of FA proteins comprises sequences from five model organisms: *Homo sapiens* (GCF\_000001405.40), *Mus musculus* (GCF\_000001635.27), *Caenorhabditis elegans* (GCF\_000002985.6), *Dictyostelium discoideum* (GCA\_000004695.1) and *Saccharomyces cerevisiae* acquired from UniProt database (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). These reference FA proteins were used as BLASTp queries against 183 fungal proteomes representing the diversity of sequenced fungi (e-value  $\leq 1e-5$ ) (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). For each protein, clustering in CLANS<sup>72</sup> was performed to group the sets of fungal homologous sequences from potentially non-specific hits ( $p$  value of  $1e-10$ ,  $1e-20$  and  $1e-30$  and attraction exponent value of 2). Sequence groups were inspected for protein domain conservation and domain architecture using NCBI’s conserved domain database<sup>73</sup> and PfamScan assessed by pfam\_scan.pl<sup>74</sup>.

In order to verify the ancestral origin of the FA pathway, we also searched for the protein sets of five basal Opisthokonts [*Capsaspora owczarzewski* (GCF\_000151315.2)<sup>75</sup>, *Sphaeroforma arctica* (GCF\_001186125.1)<sup>76</sup>, *Fonticula alba* (GCF\_000388065.1)<sup>76</sup>, *Salpingoeca rosetta* (GCF\_000188695.1)<sup>77</sup>, *Monosiga brevicollis* (GCF\_000002865.3)<sup>78</sup>]. Multiple sequence alignments were performed using MAFFT (v7.407) local alignment method<sup>79</sup> with a maximum number of iterative refinements set to 100. Phylogenetic tree construction with maximum likelihood method was carried out using IQ-TREE (v1.6.9) with automated model selection, tree search and SH-aLRT test and ultrafast bootstrap<sup>80</sup> (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Dataset SD1). A phylogenomic species tree was built with OrthoFinder<sup>81</sup> (with mmseqs) for a set of 39 fungal taxa, 5 basal Opisthokonts and 5 Metazoa with *Dictyostelium discoideum* (GCA\_000004695.1) and *Arabidopsis thaliana* (GCF\_000001735.4) as outgroups. The trees were visualized and represented using the iTOL online tool<sup>50</sup> (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Materials SM1, Fig. 5).

The expression of FA proteins was checked in publicly available EDF transcriptomes (<https://doi.org/10.5281/zenodo.10911400>, Supplementary Table S1). The limited dataset contained transcriptomes obtained from pure cultures of *Mucor lusitanicus* MU402, *Jimgerdemannia flammicorona* AD002, *Umbelopsis isabellina* M6-22, *Endogone sp.* FLAS-F59071, *Lobosporangium transversale* NRRL 3116 along with transcriptomes of EDF obtained from diverse cultural environments: *Mucor lusitanicus* MS12, *Rhizopus microsporus* FP469, *Rhizopus delemar* 99-880, *Gigaspora rosea* DAOM 194757 and *Rhizophagus irregularis* DAOM 197198.

The data were downloaded from the ENA server<sup>82</sup> as RNA-Seq fastq files. Their quality was checked using FASTQC (v0.11.8)<sup>83</sup>. The adapters were trimmed with fastp (v0.19.6) using default parameters<sup>84</sup>. The Hisat2 tool (v2.1.0)<sup>85</sup> was used to align the fastq reads with the reference genome of the respective organism downloaded from Ensembl Fungi database (release 53, downloaded between May–June 2022). The SAM files obtained from the alignment were compressed into binary file format (BAM) using samtools (v1.10)<sup>86</sup>, and the aligned reads were counted using featureCounts (v1.6.3)<sup>87</sup>. The Transcript Per Million reads (TPM) values were calculated from the aligned reads using StringTie (v2.1.3b)<sup>88</sup>.

The differential expression analysis was carried out using the DESeq2 R package<sup>89</sup> for protein-coding gene expression from condition-specific transcriptomic datasets. The *Padj* value was set to  $\leq 0.05$ , and RNA-Seq reads were mapped on the FA protein-coding gene sequences. The log2fold change criteria [*downregulation* < 0 > *upregulation*] was used to determine the gene expression profiles. In addition to this, using the same approach, we also looked at the expression profiling of genes involved in DNA repair pathways of NER, DSB and mismatch repair (MMR).

To determine genomic localization of FA protein coding genes, contig names and coordinates were retrieved from NCBI using edirect tools. The genes occurring in a single contig were grouped together and their corresponding distances were calculated. Genomic distances up to 250 kb between locus pairs were grouped in colocalization pattern<sup>48</sup>, while distances greater than 250 kb were grouped under long-range colocalization patterns<sup>49</sup>. We tested the possibility of randomly finding two FA genes in a window of 250 kb by taking into account: average gene length, gene distance and gene number in a given genome. We applied a hypergeometric distribution test using values derived from GFF files. We were able to download 56/183 GFF files. The hypergeometric probabilities were calculated for each of the 56 GFF files and the standard P-value ( $P < 0.01$ ) was chosen as the filtering criterion. The frequency of co-localization was computed and the results were visualized using Cytoscape v3.10.1<sup>90</sup>.

## Data availability

All metadata processed in this study are deposited in zenodo: <https://doi.org/10.5281/zenodo.10911400>. All protein identifiers, genomic assemblies, transcriptomic datasets, hypergeometric test values are listed in Supplementary Table S1. The figures of phylogenetic trees for FA proteins are depicted in Supplementary Materials SM1 and available as a newick file format in Supplementary Dataset SD1. Supplementary Results provide a detailed commentary on the taxonomic distribution depicted in Fig. 2.



Received: 2 January 2024; Accepted: 22 April 2024

Published online: 30 April 2024

## References

- Krishnan, V., Tay, L. S. & Ito, Y. The fanconi anemia pathway of DNA repair and human cancer. in *Advances in DNA Repair* (InTech, 2015).
- Kee, Y. & D'Andrea, A. D. Expanded roles of the Fanconi anemia pathway in preserving genomic stability. *Genes Dev.* **24**, 1680–1694 (2010).
- Cheng, S.-Y. *et al.* Cytotoxicity, crosslinking and biological activity of three mitomycins. *Bioorg. Chem.* **123**, 105744 (2022).
- McHugh, P. J., Ward, T. A. & Chovanec, M. A prototypical Fanconi anemia pathway in lower eukaryotes?. *Cell Cycle* **11**, 3739–3744 (2012).
- Kottemann, M. C. & Smogorzewska, A. Fanconi anaemia and the repair of Watson and Crick DNA crosslinks. *Nature* **493**, 356–363 (2013).
- Del Valle, J. *et al.* Exploring the role of mutations in Fanconi anemia genes in hereditary cancer patients. *Cancers* **12**, 829 (2020).
- Niraj, J., Färkkilä, A. & D'Andrea, A. D. The Fanconi anemia pathway in cancer. *Annu. Rev. Cancer Biol.* **3**, 457–478 (2019).
- Lopez-Martinez, D., Liang, C.-C. & Cohn, M. A. Cellular response to DNA interstrand crosslinks: The Fanconi anemia pathway. *Cell. Mol. Life Sci.* **73**, 3097–3114 (2016).
- Peake, J. D. & Noguchi, E. Fanconi anemia: Current insights regarding epidemiology, cancer, and DNA repair. *Hum. Genet.* **141**, 1811–1836 (2022).
- Longerich, S., Li, J., Xiong, Y., Sung, P. & Kupfer, G. M. Stress and DNA repair biology of the Fanconi anemia pathway. *Blood* **124**, 2812–2819 (2014).
- Dehé, P.-M. & Gaillard, P.-H.L. Control of structure-specific endonucleases to maintain genome stability. *Nat. Rev. Mol. Cell Biol.* **18**, 315–330 (2017).
- Tao, Y. *et al.* The structure of the FANCM–MHF complex reveals physical features for functional assembly. *Nat. Commun.* **3**, 1–12 (2012).
- Singh, D. K. *et al.* The FANCC–FANCE–FANCF complex is evolutionarily conserved and regulates meiotic recombination. *Nucleic Acids Res.* **51**, 2516–2528 (2023).
- Benitez, A. *et al.* FANCA promotes DNA double-strand break repair by catalyzing single-strand annealing and strand exchange. *Mol. Cell* **71**, 621–628.e4 (2018).
- Panday, A. *et al.* FANCM regulates repair pathway choice at stalled replication forks. *Mol. Cell* **81**, 2428–2444.e6 (2021).
- Walden, H. & Deans, A. J. The Fanconi anemia DNA repair pathway: Structural and functional insights into a complex disorder. *Annu. Rev. Biophys.* **43**, 257–278 (2014).
- Alpi, A. F., Pace, P. E., Babu, M. M. & Patel, K. J. Mechanistic insight into site-restricted monoubiquitination of FANCD2 by Ube2t, FANCL, and FANCI. *Mol. Cell* **32**, 767–777 (2008).
- Tan, W., van Twest, S., Murphy, V. J. & Deans, A. J. ATR-mediated FANCI phosphorylation regulates both ubiquitination and deubiquitination of FANCD2. *Front Cell Dev Biol* **8**, 2 (2020).
- Motnenko, A. *et al.* Identification of UHRF2 as a novel DNA interstrand crosslink sensor protein. *PLoS Genet.* **14**, e1007643 (2018).
- Payliss, B. J. *et al.* Phosphorylation of the DNA repair scaffold SLX4 drives folding of the SAP domain and activation of the MUS81-EME1 endonuclease. *Cell Rep.* **41**, 111537 (2022).
- Chen, X. *et al.* The Fanconi anemia proteins FANCD2 and FANCI interact and regulate each other's chromatin localization. *J. Biol. Chem.* **289**, 25774–25782 (2014).
- Bluteau, D. *et al.* Biallelic inactivation of REV7 is associated with Fanconi anemia. *J. Clin. Invest.* **126**, 3580–3584 (2016).
- Sharma, S. & Canman, C. E. REV1 and DNA polymerase zeta in DNA interstrand crosslink repair. *Environ. Mol. Mutagen.* **53**, 725–740 (2012).
- Roy, U. & Schärer, O. D. Involvement of translesion synthesis DNA polymerases in DNA interstrand crosslink repair. *DNA Repair* **44**, 33–41 (2016).
- Huang, T. T. *et al.* Regulation of monoubiquitinated PCNA by DUB autocleavage. *Nat. Cell Biol.* **8**, 339–347 (2006).
- Moldovan, G.-L. *et al.* DNA polymerase POLN participates in cross-link repair and homologous recombination. *Mol. Cell. Biol.* **30**, 1088–1096 (2010).
- Shamseldin, H. E., Elfaki, M. & Alkuraya, F. S. Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. *J. Med. Genet.* **49**, 184–186 (2012).
- Masson, J. Y. *et al.* Identification and purification of two distinct complexes containing the five RAD51 paralogs. *Genes Dev.* **15**, 3296–3307 (2001).
- Rothkamm, K. *et al.* DNA damage foci: Meaning and significance. *Environ. Mol. Mutagen.* **56**, 491–504 (2015).
- Foo, T. K. & Xia, B. BRCA1-dependent and independent recruitment of PALB2-BRCA2-RAD51 in the DNA damage response and cancer. *Cancer Res.* **82**, 3191–3197 (2022).
- Pellegrini, L. *et al.* Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. *Nature* **420**, 287–293 (2002).
- Rodríguez, A. & D'Andrea, A. Fanconi anemia pathway. *Curr. Biol.* **27**, R986–R988 (2017).
- Garner, E. & Smogorzewska, A. Ubiquitylation and the Fanconi anemia pathway. *FEBS Lett.* **585**, 2853–2860 (2011).
- Ceccaldi, R., Sarangi, P. & D'Andrea, A. D. The Fanconi anaemia pathway: New players and new functions. *Nat. Rev. Mol. Cell Biol.* **17**, 337–349 (2016).
- Bona, N. & Crossan, G. P. Fanconi anemia DNA crosslink repair factors protect against LINE-1 retrotransposition during mouse development. *Nat. Struct. Mol. Biol.* <https://doi.org/10.1038/s41594-023-01067-8> (2023).
- Riesgo, A., Farrar, N., Windsor, P. J., Giribet, G. & Leys, S. P. The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Mol. Biol. Evol.* **31**, 1102–1120 (2014).
- Jones, M. & Rose, A. A DOG's view of Fanconi anemia: Insights from *C. elegans*. *Anemia* **2012**, 323721 (2012).
- Clay, D. E., Jezuit, E. A., Montague, R. A. & Fox, D. T. Conserved function of *Drosophila* Fancd2 monoubiquitination in response to double-strand DNA breaks. *G3* **12**, jkac129 (2022).
- Daee, D. L. *et al.* Rad5-dependent DNA repair functions of the *Saccharomyces cerevisiae* FANCM protein homolog Mph1. *J. Biol. Chem.* **287**, 26563–26575 (2012).
- Homa, M. *et al.* Differential gene expression of *Mucor lusitanicus* under aerobic and anaerobic conditions. *J. Fungi (Basel)* **8**, 404 (2022).
- Sephton-Clark, P. *et al.* Bacterial endosymbionts influence fungal transcriptional profiles with implications for host response in the human fungal pathogens *Rhizopus microsporus* and *Rhizopus delemar*. *bioRxiv* <https://doi.org/10.1101/580746> (2020).
- Chibucos, M. C. *et al.* An integrated genomic and transcriptomic survey of mucormycosis-causing fungi. *Nat. Commun.* **7**, 12218 (2016).
- Andrianaki, A. M. *et al.* Iron restriction inside macrophages regulates pulmonary host defense against *Rhizopus* species. *Nat. Commun.* **9**, 3333 (2018).
- Tang, N. *et al.* A survey of the gene repertoire of *Gigaspora rosea* unravels conserved features among glomeromycota for obligate biotrophy. *Front. Microbiol.* **7**, 233 (2016).

45. Sugimura, Y. & Saito, K. Transcriptional profiling of arbuscular mycorrhizal roots exposed to high levels of phosphate reveals the repression of cell cycle-related genes and secreted protein genes in *Rhizophagus irregularis*. *Mycorrhiza* **27**, 139–146 (2017).
46. Tsuzuki, S., Handa, Y., Takeda, N. & Kawaguchi, M. Strigolactone-induced putative secreted protein 1 is required for the establishment of symbiosis by the arbuscular mycorrhizal fungus *Rhizophagus irregularis*. *Mol. Plant. Microbe Interact.* **29**, 277–286 (2016).
47. Wisecaver, J. H., Slot, J. C. & Rokas, A. Correction: The evolution of fungal metabolic pathways. *PLoS Genet.* **11**, e1005449 (2015).
48. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
49. Torres, D. E. *et al.* Three-dimensional chromatin organization promotes genome evolution in a fungal plant pathogen. *bioRxiv* <https://doi.org/10.1101/2023.04.04.535574> (2023).
50. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
51. Rogers, C. M. *et al.* The yeast Hrq1 helicase stimulates Pso2 translesion nuclease activity and thereby promotes DNA interstrand crosslink repair. *J. Biol. Chem.* **295**, 8945–8957 (2020).
52. Merényi, Z. *et al.* Genomes of fungi and relatives reveal delayed loss of ancestral gene families and evolution of key fungal traits. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-023-02095-9> (2023).
53. Meetei, A. R. *et al.* A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat. Genet.* **37**, 958–963 (2005).
54. Lorenz, A. *et al.* The fission yeast FANCM ortholog directs non-crossover recombination during meiosis. *Science* **336**, 1585–1588 (2012).
55. Knoll, A. *et al.* The Fanconi anemia ortholog FANCM ensures ordered homologous recombination in both somatic and meiotic cells in *Arabidopsis*. *Plant Cell* **24**, 1448–1464 (2012).
56. Sun, W. *et al.* The FANCM ortholog Fml1 promotes recombination at stalled replication forks and limits crossing over during DNA double-strand break repair. *Mol. Cell* **32**, 118–128 (2008).
57. Smogorzewska, A. *et al.* Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289–301 (2007).
58. Yamamoto, K. N. *et al.* Involvement of SLX4 in interstrand cross-link repair is regulated by the Fanconi anemia pathway. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6492–6496 (2011).
59. Moldovan, G.-L. & D'Andrea, A. D. How the Fanconi anemia pathway guards the genome. *Annu. Rev. Genet.* **43**, 223–249 (2009).
60. Talwar, T. *et al.* The DEAD-box protein DDX43 (HAGE) is a dual RNA-DNA helicase and has a K-homology domain required for full nucleic acid unwinding activity. *J. Biol. Chem.* **292**, 10429–10443 (2017).
61. Huang, J. & Cook, D. E. The contribution of DNA repair pathways to genome editing and evolution in filamentous pathogens. *FEMS Microbiol. Rev.* **46**, fuac035 (2022).
62. Elya, C. & De Fine Licht, H. H. The genus *Entomophthora*: Bringing the insect destroyers into the twenty-first century. *IMA Fungus* **12**, 34 (2021).
63. Pears, C. J., Brustel, J. & Lakin, N. D. Dictyostelium discoideum as a model to assess genome stability through DNA repair. *Front. Cell Dev. Biol.* **9**, 752175 (2021).
64. Zhang, X.-Y. *et al.* Xpf and not the Fanconi anaemia proteins or Rev3 accounts for the extreme resistance to cisplatin in *Dictyostelium discoideum*. *PLoS Genet.* **5**, e1000645 (2009).
65. Chennuri, P., Cox, L. S. & Saunders, R. D. C. EXD2 and WRN exonucleases are required for interstrand crosslink repair in *Drosophila*. *bioRxiv* <https://doi.org/10.1101/284307> (2018).
66. Berbee, M. L. *et al.* Genomic and fossil windows into the secret lives of the most ancient fungi. *Nat. Rev. Microbiol.* **18**, 717–730 (2020).
67. Morin, E. *et al.* Comparative genomics of *Rhizophagus irregularis*, *R. cerebriforme*, *R. diaphanus* and *Gigaspora rosea* highlights specific genetic features in Glomeromycotina. *New Phytol.* **222**, 1584–1598 (2019).
68. UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
69. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **50**, D687–D692 (2022).
70. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).
71. Zdobnov, E. M. *et al.* OrthoDB in 2020: Evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **49**, D389–D393 (2021).
72. Frickey, T. & Lupas, A. CLANS: A Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* **20**, 3702–3704 (2004).
73. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).
74. Bateman, A. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **32**, D138–D141 (2004).
75. Suga, H. *et al.* The Capsaspora genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325 (2013).
76. Ruiz-Trillo, I. *et al.* The origins of multicellularity: A multi-taxon genome initiative. *Trends Genet.* **23**, 113–118 (2007).
77. Fairclough, S. R. *et al.* Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* **14**, R15 (2013).
78. King, N. *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **451**, 783–788 (2008).
79. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
80. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
81. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
82. Leinonen, R. *et al.* The European nucleotide archive. *Nucleic Acids Res.* **39**, D28–31 (2011).
83. Andrews, S. *et al.* FastQC: A quality control tool for high throughput sequence data. *Preprint at* (2010).
84. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
85. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
86. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
87. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
88. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
89. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
90. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

## Acknowledgements

This work was supported by the National Science Centre grant #2021/41/B/NZ2/02426 to A.M.

### Author contributions

D.B and A.M. designed the study; D.B. and A.M. prepared the dataset and performed the analyses; D.B., M.P. and A.M. interpreted the data and wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-60318-w>.

**Correspondence** and requests for materials should be addressed to A.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024