BioTechnologia vol. 97(4) • 2016
Journal of Biotechnology, Computational Biology and Bionanotechnology SUPPLEMENTARY MATERIALS

CC BY NC ND

A novel approach for identifying DNA repair pathways proteins using an evolutionary approach: *Plasmodium falciparum* case study

KAJA MILANOWSKA[1, 2, 3]*, JUSTYNA WOJTCZAK[1]

[1] Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Poznań, Poland

[2] Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warszawa, Poland

[3] European Center for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland

Abstract

Analysis of evolution-related proteins is one of the most effective methods for assigning functions to newly discovered proteins. We present a novel approach for identifying homologues that can be involved in DNA repair mechanisms in organisms without proteome annotation. This approach is based on the profiles built with sequences and secondary structures of proteins with known and well-described function. For this analysis, the profiles of DNA repair proteins from three model organisms (*Homo sapiens*, *Saccharomyces cerevisiae*, and *Escherichia coli*) were prepared and deposited in the DNA repair pathways database – REPAIRtoire (http://repairtoire.genesilico.pl) (Milanowska et al., 2011). The methodology formulated was used to analyze all proteins from *Plasmodium falciparum* and annotate those with *unreviewed* status that were recognized as taking part in DNA repair. The analysis identified novel proteins that have neither been previously described in the InParanoid database (O'Brien et al., 2005) as homologous or orthologous for the chosen model organisms, nor been identified by the HHsearch program (Soding, 2005). This approach is included as a search module on the REPAIRtoire website (http://repairtoire.genesilico.pl/homologs) and is available for usage.

Key words: DNA repair pathways, homologs detection, orthologs, pipeline, *Plasmodium falciparum*

* Corresponding author: Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, ul. Umultowska 89, 61-614 Poznań, Poland; e-mail: kaja@amu.edu.pl

Table 1. Summary of proteins taking part in MMR pathway in *Plasmodium falciparum* (based on KEGG database: www.kegg.jp/kegg-bin/show_pathway?org_name=pfa&mapno=03410). First column (1) is a shortcut used to name the protein, second column (2) shows the name of the organism, third column (3) bears the UniProt ID of the protein, fourth column (4) describes function of the protein. Columns (5) and (6) shows weather the protein was found as a homolog for given organism using sequence approach and structural approach

MMR protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch result (5)	Algorithm result (6)
SSB	*Plasmodium falciparum*	Q8I415	single-strand binding protein, putative		
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
PMS2	*Plasmodium falciparum*	Q8IBJ3	mismatch repair protein pms1 homologue, putative		
	Homo sapiens	**P54278**	mismatch repair endonuclease PMS2	×	
	Saccharomyces cerevisiae	**P14242**	DNA mismatch repair protein PMS1		×
	Escherichia coli	–			
MLH1	*Plasmodium falciparum*	Q8IIJ0	DNA mismatch repair protein MLH1, putative		
	Homo sapiens	**P40692**	DNA mismatch repair protein Mlh1	×	×
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
MSH6	*Plasmodium falciparum*	Q8I447	DNA repair protein, putative		
	Homo sapiens	**P52701**	DNA mismatch repair protein Msh6	×	
	Saccharomyces cerevisiae	**Q03834**	DNA mismatch repair protein MSH6		×
	Escherichia coli	P23909	DNA mismatch repair protein MutS		
PMS2	*Plasmodium falciparum*	Q8IBJ3	mismatch repair protein pms1 homologue, putative		
	Homo sapiens	**P54278**	mismatch repair endonuclease PMS2	×	
	Saccharomyces cerevisiae	**P14242**	DNA mismatch repair protein PMS1		×
	Escherichia coli	–			
MSH2	*Plasmodium falciparum*	8ILI9	DNA mismatch repair protein Msh2p, putative		
	Homo sapiens	P43246	DNA mismatch repair protein Msh2		
	Saccharomyces cerevisiae	**P25847**	DNA mismatch repair protein MSH2	×	×
	Escherichia coli				
RFC	*Plasmodium falciparum*	Q8IIQ1	replication factor C subunit 5, putative		
	Homo sapiens	P40938	replication factor C subunit 3		
	Saccharomyces cerevisiae	P38251	replication factor C subunit 5		
	Escherichia coli				
RFC	*Plasmodium falciparum*	Q8IKK4	replication factor C3		
	Homo sapiens	P40937	replication factor C subunit 5		
	Saccharomyces cerevisiae	P38629	replication factor C subunit 3		
	Escherichia coli	–			

MMR protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch result (5)	Algorithm result (6)
RFC	*Plasmodium falciparum*	O96260	replication factor C, subunit 2		
	Homo sapiens	P35249	replication factor C subunit 4		
	Saccharomyces cerevisiae	P40348	replication factor C subunit 2		
	Escherichia coli	–			
RFC	*Plasmodium falciparum*	O96271	replication factor C subunit 1, putative		
	Homo sapiens	**P35251**	replication factor C subunit 1	×	×
	Saccharomyces cerevisiae	P38630	replication factor C subunit 1		
	Escherichia coli	–			
RFC	*Plasmodium falciparum*	Q8I512	replication factor C subunit 4		
	Homo sapiens	P35250	replication factor C subunit 2		
	Saccharomyces cerevisiae	P40339	replication factor C subunit 4		
	Escherichia coli	–			
PCNA	*Plasmodium falciparum*	P61074	proliferating cell nuclear antigen		
	Homo sapiens	P12004	proliferating cell nuclear antigen		
	Saccharomyces cerevisiae	**P15873**	proliferating cell nuclear antigen		×
	Escherichia coli	–			
Exo1	*Plasmodium falciparum*	**Q8IBK1**			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RPA	*Plasmodium falciparum*	C6S3I6			
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RPA	*Plasmodium falciparum*	**Q9U0J0**		×	×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
Pol-d	*Plasmodium falciparum*	Q7KQL4	DNA polymerase		
	Homo sapiens	**P28340**	DNA polymerase delta catalytic subunit		×
	Saccharomyces cerevisiae	**P15436**	DNA polymerase delta catalytic subunit	×	
	Escherichia coli	P21189	DNA polymerase II		
Pol-e	*Plasmodium falciparum*	O77321	DNA polymerase epsilon subunit B, putative		
	Homo sapiens	P49005	DNA polymerase delta subunit 2		
	Saccharomyces cerevisiae	**P46957**	DNA polymerase delta small subunit	×	×
	Escherichia coli				
Lig1	*Plasmodium falciparum*	Q8IES4	DNA ligase		
	Homo sapiens	**P18858**	DNA ligase 1	×	×
	Saccharomyces cerevisiae	P04819	DNA ligase 1		
	Escherichia coli	–			

Table 2. Summary of proteins taking part in HRR pathway in *Plasmodium falciparum* (based on KEGG database: www.kegg.jp/kegg-bin/show_pathway?org_name=pfa&mapno=03410). First column (1) is a shortcut used to name the protein, second column (2) shows the anme of the organism, third column (3) bears the UniProt ID of the protein, fourth column (4) describes function of the protein. Columns (5) and (6) shows weather the protein was found as a homolog for given organism using sequence approach and structural approach

HHR protein (1)	Organism (2)	UniProt ID and InParanoid reference (3)	Function (4)	HHSearch result (5)	Algorith result (6)
SSB	*Plasmodium falciparum*	Q8I415			
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
DPO1	*Plasmodium falciparum*	Q8ILY1		×	×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
DPO1	*Plasmodium falciparum*	C6KT89			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
BLM	*Plasmodium falciparum*	Q8I2W7	DNA helicase, putative		
	Homo sapiens	P46063	ATP-dependent DNA helicase Q1		
	Saccharomyces cerevisiae	–			
	Escherichia coli	**P15043**	ATP-dependent DNA helicase RecQ	×	×
TOP3	*Plasmodium falciparum*	Q8IDM7	DNA topoisomerase III, putative		
	Homo sapiens	Q13472	DNA topoisomerase 3-alpha		
	Saccharomyces cerevisiae	P13099	DNA topoisomerase 3		
	Escherichia coli	**P06612**	DNA topoisomerase 1	×	×
RAD50	*Plasmodium falciparum*	C6KSQ6	probable DNA repair protein RAD50		
	Homo sapiens	–			
	Saccharomyces cerevisiae	P12753	DNA repair protein RAD50		
	Escherichia coli	–			
RPA3	*Plasmodium falciparum*	C6S3I6			
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RPA1		**Q9U0J0**		×	×
		–			
		–			
		–			
RAD51	*Plasmodium falciparum*	Q8IIS8	Rad51 homolog		
	Homo sapiens	Q06609	DNA repair protein RAD51 homolog 1		
	Saccharomyces cerevisiae	**P25454**	DNA repair protein RAD51	×	×
	Escherichia coli	P0A7G6	protein RecA		

HHR protein (1)	Organism (2)	UniProt ID and InParanoid reference (3)	Function (4)	HHSearch result (5)	Algorith result (6)
DSS1	*Plasmodium falciparum*	Q8IBN3			
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RAD54	*Plasmodium falciparum*	Q8IAN4	DNA repair protein rad54, putative		
	Homo sapiens	**Q92698**	DNA repair and recombination protein RAD54-like	×	×
	Saccharomyces cerevisiae	P32863	DNA repair and recombination protein RAD54		
	Escherichia coli	–			
POL-d, catalytic subunit	*Plasmodium falciparum*	Q7KQL4	DNA polymerase		
	Homo sapiens	**P28340**	DNA polymerase delta catalytic subunit		×
	Saccharomyces cerevisiae	**P15436**	DNA polymerase delta catalytic subunit	×	
	Escherichia coli	P21189	DNA polymerase II		
POL-d subunit 2	*Plasmodium falciparum*	O77321	DNA polymerase epsilon subunit B, putative		
	Homo sapiens	P49005	DNA polymerase delta subunit 2		
	Saccharomyces cerevisiae	**P46957**	DNA polymerase delta small subunit	×	×
	Escherichia coli	–			

Table 3. Summary of proteins taking part in BER pathway in *Plasmodium falciparum* (based on KEGG database: www.kegg.jp/kegg-bin/show_pathway?org_name=pfa&mapno=03410). First column (1) is a shortcut used to name the protein, second column (2) shows the anme of the organism, third column (3) bears the UniProt ID of the protein, fourth column (4) describes function of the protein. Columns (5) and (6) shows weather the protein was found as a homolog for given organism using sequence approach and structural approach

BER protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch result (5)	Algorithm result (6)
APE1	*Plasmodium falciparum*	Q8IE02			
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
DPO1	*Plasmodium falciparum*	**Q8ILY1**		×	×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
DPO1	*Plasmodium falciparum*	**C6KT89**			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			

BER protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch result (5)	Algorithm result (6)
Lig1	*Plasmodium falciparum*	Q8IES4	DNA ligase		
	Homo sapiens	**P18858**	DNA ligase 1	×	×
	Saccharomyces cerevisiae	P04819	DNA ligase 1		
	Escherichia coli				
PCNA	*Plasmodium falciparum*	P61074	proliferating cell nuclear antigen		
	Homo sapiens	P12004	proliferating cell nuclear antigen		
	Saccharomyces cerevisiae	**P15873**	proliferating cell nuclear antigen		×
	Escherichia coli				
PCNA2	*Plasmodium falciparum*	Q7KQJ9			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
Pol-d	*Plasmodium falciparum*	Q7KQL4	DNA polymerase		
	Homo sapiens	**P28340**	DNA polymerase delta catalytic subunit		×
	Saccharomyces cerevisiae	**P15436**	DNA polymerase delta catalytic subunit	×	
	Escherichia coli	P21189	DNA polymerase II		
Pol-e	*Plasmodium falciparum*	O77321	DNA polymerase epsilon subunit B, putative		
	Homo sapiens	P49005	DNA polymerase delta subunit 2		
	Saccharomyces cerevisiae	**P46957**	DNA polymerase delta small subunit	×	×
	Escherichia coli				
Pol-e, subunit A	*Plasmodium falciparum*	C6KTD8	DNA polymerase epsilon, catalytic subunit a, putative		
	Homo sapiens	**Q07864**	DNA polymerase epsilon catalytic subunit A		×
	Saccharomyces cerevisiae	**P21951**	DNA polymerase epsilon catalytic subunit A	×	
	Escherichia coli	–			
Pol-e, subunit B	*Plasmodium falciparum*	Q8I579	DNA polymerase epsilon subunit B, putative		
	Homo sapiens	P56282	DNA polymerase epsilon subunit 2		
	Saccharomyces cerevisiae	**P24482**	DNA polymerase epsilon subunit B		×
	Escherichia coli	–			
Fen1	*Plasmodium falciparum*	Q7K734	flap endonuclease 1		
	Homo sapiens	**P39748**	flap endonuclease 1	×	×
	Saccharomyces cerevisiae	P26793	flap endonuclease 1		
	Escherichia coli	–			

Table 4. Summary of proteins taking part in NER pathway in *Plasmodium falciparum* (based on KEGG database: www.kegg.jp/kegg-bin/show_pathway?org_name=pfa&mapno=03410). First column (1) is a shortcut used to name the protein, second column (2) shows the anme of the organism, third column (3) bears the UniProt ID of the protein, fourth column (4) describes function of the protein. Columns (5) and (6) shows weather the protein was found as a homolog for given organism using sequence approach and structural approach

NER protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch recult (5)	Algorithm result (6)
DPO1	*Plasmodium falciparum*	Q8ILY1		×	×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
DPO1	*Plasmodium falciparum*	C6KT89			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RBX1	*Plasmodium falciparum*	O77367	ubiquitin-protein ligase, putative		
	Homo sapiens	P62877	E3 ubiquitin-protein ligase RBX1		
	Saccharomyces cerevisiae	Q08273	RING-box protein HRT1		
	Escherichia coli	–			
HR23B	*Plasmodium falciparum*	Q8IJS8	DNA repair protein RAD23, putative		×
	Homo sapiens	P54725	UV excision repair protein RAD23 homolog A		
	Saccharomyces cerevisiae	P32628	UV excision repair protein RAD23		
	Escherichia coli	–			
CDK7	*Plasmodium falciparum*	Q8IJQ1			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
MNAT1	*Plasmodium falciparum*	Q8I3Y3	CDK-activating kinase assembly factor, putative		×
	Homo sapiens	P51948	CDK-activating kinase assembly factor MAT1		
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
CCNH	*Plasmodium falciparum*	Q8IKK0	cyclin homologue		×
	Homo sapiens	P51946	cyclin-H		
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
XPB	*Plasmodium falciparum*	Q8IJ31	DNA repair helicase rad25, putative		
	Homo sapiens	P19447	TFIIH basal transcription factor complex helicase XPB subunit	×	×
	Saccharomyces cerevisiae	Q00578	DNA repair helicase RAD25		
	Escherichia coli	–			

NER protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch recult (5)	Algorithm result (6)
XPD	*Plasmodium falciparum*	Q8I2H7	DNA excision-repair helicase, putative		
	Homo sapiens	P18074	TFIIH basal transcription factor complex helicase XPD subunit		
	Saccharomyces cerevisiae	P06839	DNA repair helicase RAD3	×	×
	Escherichia coli	–			
TTDA	*Plasmodium falciparum*	Q8IL51			
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
TFIIH2	*Plasmodium falciparum*	Q8IEG6	TFIIH basal transcription factor subunit		
	Homo sapiens	Q13888	general transcription factor IIH subunit 2		
	Saccharomyces cerevisiae	Q04673	suppressor of stem-loop protein 1	×	
	Escherichia coli	–			
TFIIH3	*Plasmodium falciparum*	Q8IDG5	uncharacterized protein		
	Homo sapiens	Q13889	general transcription factor IIH subunit 3		
	Saccharomyces cerevisiae	Q12004	RNA polymerase II transcription factor B subunit 4		×
	Escherichia coli	–			
TFIIH4	*Plasmodium falciparum*	Q8I4Y8	transcription factor Tfb2, putative		
	Homo sapiens	Q92759	general transcription factor IIH subunit 4		
	Saccharomyces cerevisiae	Q02939	RNA polymerase II transcription factor B subunit 2		×
	Escherichia coli	–			
XPG	*Plasmodium falciparum*	O96154	DNA repair endonuclease, putative		
	Homo sapiens	P28715	DNA repair protein complementing XP-G cells		×
	Saccharomyces cerevisiae	P07276	DNA repair protein RAD2		
	Escherichia coli	–			
XPA	*Plasmodium falciparum*	Q8IBY6			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RPA	*Plasmodium falciparum*	C6S3I6			
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RPA	*Plasmodium falciparum*	Q9U0J0		×	×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			

NER protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch recult (5)	Algorithm result (6)
XPF	*Plasmodium falciparum*	Q8ID22	DNA repair endonuclease, putative		
	Homo sapiens	Q92889	DNA repair endonuclease XPF		
	Saccharomyces cerevisiae	P06777	DNA repair protein RAD1		
	Escherichia coli	–			
ERCC1	*Plasmodium falciparum*	O96136	ERCC1 nucleotide excision repair protein, putative		
	Homo sapiens	P07992	DNA excision repair protein ERCC-1		
	Saccharomyces cerevisiae	P06838	DNA repair protein RAD10		
	Escherichia coli	–			
Pol-d	*Plasmodium falciparum*	Q7KQL4	DNA polymerase		
	Homo sapiens	P28340	DNA polymerase delta catalytic subunit		×
	Saccharomyces cerevisiae	P15436	DNA polymerase delta catalytic subunit	×	
	Escherichia coli	P21189	DNA polymerase II		
Pol-e	*Plasmodium falciparum*	O77321	DNA polymerase epsilon subunit B, putative		
	Homo sapiens	P49005	DNA polymerase delta subunit 2		
	Saccharomyces cerevisiae	P46957	DNA polymerase delta small subunit	×	×
	Escherichia coli	–			
Pol-e, subunit A	*Plasmodium falciparum*	C6KTD8	DNA polymerase epsilon, catalytic subunit a, putative		
	Homo sapiens	Q07864	DNA polymerase epsilon catalytic subunit A		×
	Saccharomyces cerevisiae	P21951	DNA polymerase epsilon catalytic subunit A	×	
	Escherichia coli	–			
Pol-e, subunit B	*Plasmodium falciparum*	Q8I579	DNA polymerase epsilon subunit B, putative		
	Homo sapiens	P56282	DNA polymerase epsilon subunit 2		
	Saccharomyces cerevisiae	P24482	DNA polymerase epsilon subunit B		×
	Escherichia coli	–			
PCNA	*Plasmodium falciparum*	P61074	proliferating cell nuclear antigen		
	Homo sapiens	P12004	proliferating cell nuclear antigen		
	Saccharomyces cerevisiae	P15873	proliferating cell nuclear antigen		×
	Escherichia coli				
PCNA2	*Plasmodium falciparum*	Q7KQJ9			×
	Homo sapiens	–			
	Saccharomyces cerevisiae	–			
	Escherichia coli	–			
RFC	*Plasmodium falciparum*	Q8IIQ1	replication factor C subunit 5, putative		
	Homo sapiens	P40938	replication factor C subunit 3		
	Saccharomyces cerevisiae	P38251	replication factor C subunit 5		
	Escherichia coli	–			

NER protein (1)	Organism (2)	UniProtID and InParanoid reference (3)	Function (4)	HHSearch recult (5)	Algorithm result (6)
RFC	*Plasmodium falciparum*	Q8IKK4	replication factor C3		
	Homo sapiens	P40937	replication factor C subunit 5		
	Saccharomyces cerevisiae	P38629	replication factor C subunit 3		
	Escherichia coli	–			
RFC	*Plasmodium falciparum*	O96260	replication factor C, subunit 2		
	Homo sapiens	P35249	replication factor C subunit 4		
	Saccharomyces cerevisiae	P40348	replication factor C subunit 2		
	Escherichia coli	–			
RFC	*Plasmodium falciparum*	O96271	replication factor C subunit 1, putative		
	Homo sapiens	P35251	replication factor C subunit 1	×	×
	Saccharomyces cerevisiae	P38630	replication factor C subunit 1		
	Escherichia coli	–			
RFC	*Plasmodium falciparum*	Q8I512	replication factor C subunit 4		
	Homo sapiens	P35250	replication factor C subunit 2		
	Saccharomyces cerevisiae	P40339	replication factor C subunit 4		
	Escherichia coli	–			
Lig1	*Plasmodium falciparum*	Q8IES4	DNA ligase		
	Homo sapiens	P18858	DNA ligase 1	×	×
	Saccharomyces cerevisiae	P04819	DNA ligase 1		
	Escherichia coli	–			

Table 5. Results obtained with querying REPAIRtoire database with all proteins from *Plasmodium falciparum* proteome

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
C0H4L8	DNA mismatch repair protein	yes	complete proteome reference proteome	P25847	4.6e-79	P25847	1e-100
C0H4V8	DNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	P32863	2e-53	Q03468	1.5e-69
C6KSP4	DNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			Q92698	3.1e-62
C6KSR0	Ubiquitin conjugating enzyme E2	yes	complete proteome reference proteome			P52490	9.1e-51

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
C6KSY9	Endonuclease III homologue	yes	complete proteome endonuclease hydrolase lyase nuclease reference proteome			P78549	2.1e-53
C6KT89	DNA polymerase 1	yes	complete proteome nucleotidyl transferase reference proteome transferase			P00582	3.8e-90
C6KTD3	Cullin-like protein	yes	complete proteome reference proteome			Q13619	2.9e-96
C6KTD8	DNA polymerase epsilon, catalytic subunit A	yes	complete proteome nucleotidyl transferase reference proteome transferase	P21951	6e-208	Q07864	0.0
C6KTE4	DEAD/DEAH box ATP-dependent RNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome RNA-binding	O94762	5.6e-52	P35187	3.6e-52
C6S3D8	Topoisomerase	yes	complete proteome isomerase reference proteome			P23179	1.3e-67
C6S3H8	RNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P46063	6e-52
O77321	DNA polymerase epsilon subunit B	yes	complete proteome reference proteome	P46957	8.2e-64	P46957	1e-115
O77380	CPSF (Cleavage and polyadenylation specific factor), subunit A	yes	complete proteome reference proteome			Q16531	1.9e-94
O77397	Ubiquitin conjugating enzyme	yes	complete proteome ligase reference proteome Ubl conjugation pathway			P52490	5.8e-52
O96136	ERCC1 nucleotide excision repair protein	yes	complete proteome reference proteome			P07992	5.1e-59
O96139	5′-3′ exonuclease, N-terminal resolvase-like domain	yes	complete proteome exonuclease hydrolase nuclease reference proteome			P00582	6.2e-65
O96154	DNA repair endonuclease	yes	complete proteome endonuclease hydrolase nuclease nucleus reference proteome			P28715	4.2e-74

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
O96239	DEAD/DEAH box helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	Q92698	2.2e-65	Q92698	5.9e-74
O96264	DEAD/DEAH box helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	9.8e-54		
O96271	Replication factor C subunit 1	yes	complete proteome reference proteome	P35251	3e-83	P35251	2e-102
O97285	ATP-dependent RNA Helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	1.8e-52	P46063	5.7e-52
O97290	ATP-dependent RNA Helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P35187	9.3e-52
Q8I0W7	Snrnp protein	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	1.7e-54		
Q8I0W8	Deoxyribodipyrimidine photolyase (Photoreactivating enzyme, DNA photolyase)	yes	complete proteome lyase reference proteome			P05066	6.1e-92
Q8I1X9	ATP-dependent RNA Helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P46063	1e-53
Q8I2H7	DNA excision-repair helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding nucleus reference proteome	P06839	9e-131	P06839	3e-125
Q8I2U3	Ubiquitin conjugating enzyme	yes	complete proteome ligase reference proteome Ubl conjugation pathway			P52490	8.7e-51

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8I2W7	DNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	P15043	6.2e-64	P15043	6.7e-61
Q8I2Y2	N-glycosylase/DNA lyase	yes	complete proteome glycosidase hydrolase lyase reference proteome			O15527	9e-62
Q8I301	Ubiquitin conjugating enzyme	yes	3D-structure complete proteome ligase reference proteome Ubl conjugation pathway			P52490	4.3e-51
Q8I347	Uncharacterized protein	yes	complete proteome reference proteome			Q9UBZ9	4.2e-89
Q8I3A1	Replication factor A-related protein	yes	complete proteome reference proteome			P27694	8.3e-84
Q8I3J4	Ubiquitin conjugating enzyme	yes	3D-structure complete proteome ligase reference proteome Ubl conjugation pathway			P52490	2.3e-53
Q8I3V5	Phosphatidylinositol 3-kinase	yes	complete proteome kinase reference proteome transferase			P38111	2.5e-54
Q8I3W6	Helicase, belonging to UvrD family	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P03018	1.4e-74
Q8I3Y3	CDK-activating kinase assembly factor	yes	complete proteome kinase metal-binding reference proteome transferase zinc zinc-finger			Q03290	7.6e-55
Q8I416	ATP-dependent RNA Helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P35187	1.6e-51
Q8I4R9	DEAD/DEAH box helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	9.9e-52	P35187	5.3e-53
Q8I4S6	DNA repair protein rhp16	yes	complete proteome reference proteome	P32863	1.8e-54	P32863	1.8e-61

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8I4W3	Rac-beta serine/threonine protein kinase	yes	ATP-binding complete proteome kinase nucleotide-binding reference proteome serine/threonine-protein kinase transferase	O14757	5.6e-56		
Q8I511	DEAD/DEAH box helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P35187	1.4e-52
Q8I534	Calcium/calmodulin-dependent protein kinase 2	yes	ATP-binding complete proteome kinase nucleotide-binding reference proteome transferase			O96017	2e-52
Q8I574	Splicing factor 3b, subunit 3, 130kD	yes	complete proteome reference proteome	Q16531	2e-111	Q16531	9e-145
Q8I5E7	ATP-dependent RNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	3.5e-59		
Q8I5R8	RNA polymerase subunit 8c	yes	complete proteome reference proteome			P52434	1.9e-65
Q8I607	Ubiquitin conjugating enzyme E2	yes	complete proteome ligase reference proteome Ubl conjugation pathway			P52490	2e-52
Q8IAN4	DNA repair protein rad54	yes	complete proteome reference proteome	Q92698	4e-96	Q92698	6.3e-96
Q8IAU1	RNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	2.8e-55		
Q8IAU5	Cullin-like protein	yes	complete proteome reference proteome	Q13619	4.6e-71	Q13619	2e-124
Q8IAW2	Ubiquitin conjugating enzyme	yes	complete proteome ligase reference proteome Ubl conjugation pathway			P52490	1e-51
Q8IBJ3	Mismatch repair protein pms1 homologue	yes	complete proteome reference proteome	P54278	6.1e-71	P14242	4.4e-98

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8IBK1	Exonuclease I	yes	complete proteome exonuclease hydrolase nuclease reference proteome			P39875	2.6e-65
Q8IBN8	DEAD box helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome RNA-binding			P35187	8.5e-51
Q8IBY6	Nucleotide excision repair protein	yes	complete proteome reference proteome			P23025	5.9e-69
Q8ID22	DNA repair endonuclease	yes	complete proteome endonuclease hydrolase nuclease reference proteome			P06777	1e-117
Q8ID59	DNA-directed RNA polymerase 2	yes	complete proteome DNA-directed RNA polymerase nucleotidyl transferase reference proteome transcription transferase	P19388	2.9e-62	P19388	2.7e-83
Q8ID74	Polynucleotide kinase	yes	complete proteome kinase reference proteome transferase			Q96T60	3e-65
Q8IDD0	DNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			Q92698	1.1e-66
Q8IDM7	DNA topoisomerase III	yes	complete proteome isomerase reference proteome	P06612	4e-69	P06612	6e-128
Q8IDQ5	DNA helicase	yes	ATP-binding complete proteome helicase hydrolase metal-binding nucleotide-binding reference proteome zinc zinc-finger	P32863	4.2e-52	Q9Y620	1.1e-68
Q8IDV5	Calcium-dependent protein kinase	yes	ATP-binding calcium complete proteome kinase nucleotide-binding reference proteome transferase	O96017	1.1e-58		

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8IE01	ATP-dependent RNA Helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P46063	3.3e-52
Q8IE72	DEAD box helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			P06839	6.2e-94
Q8IFN9	Eukaryotic initiation factor	yes	ATP-binding complete proteome helicase hydrolase initiation factor nucleotide-binding protein biosynthesis reference proteome	O94762	7e-57		
Q8IFP1	U5 small nuclear ribonucleoprotein-specific protein	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome ribonucleoprotein viral nucleoprotein virion			O75417	8.9e-62
Q8II68	A/G-specific adenine glycosylase	yes	complete proteome reference proteome			Q9UIF7	1.5e-72
Q8II71	26S proteasome regulatory complex subunit	yes	complete proteome proteasome reference proteome	Q13098	1e-51	Q13098	6.8e-72
Q8IIJ0	DNA mismatch repair protein MLH1	yes	complete proteome reference proteome	P40692	7e-104	P40692	6e-151
Q8IJ31	DNA repair helicase rad25	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	P19447	4e-111	P19447	2e-127
Q8IJN8	Ribonucleotide reductase small subunit	yes	complete proteome reference proteome			P49723	8.5e-83
Q8IJQ1	Cdk7	yes	ATP-binding complete proteome kinase nucleotide-binding reference proteome transferase			P50613	1.2e-53
Q8IJS8	DNA repair protein RAD23	yes	complete proteome reference proteome			P54727	1.2e-77

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8IJW1	Endonuclease	yes	complete proteome endonuclease hydrolase nuclease reference proteome			P39748	5.8e-57
Q8IKG6	DNA-3-methyladenine glycosylase	yes	complete proteome reference proteome			P29372	2.7e-63
Q8IKK3	DNA polymerase alpha subunit	yes	complete proteome reference proteome			P24482	3.6e-56
Q8IKT6	Serine/threonine protein kinase	yes	ATP-binding complete proteome kinase nucleotide-binding reference proteome serine/threonine-protein kinase transferase	O14757	1.7e-51		
Q8IKX5	Serine/threonine protein kinase	yes	ATP-binding complete proteome kinase nucleotide-binding reference proteome serine/threonine-protein kinase transferase			O14757	1.6e-51
Q8IKY1	DNA replication related protein	yes	complete proteome reference proteome			Q96NY9	2.7e-56
Q8IL13	Helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	2.4e-60	O94762	6.9e-51
Q8IL78	DEAD/DEAH box helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome			O75417	8.5e-58
Q8ILG5	ATP-dependent DNA helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	P15043	2.7e-74	P15043	4.3e-80
Q8ILI9	DNA mismatch repair protein Msh2p	yes	complete proteome reference proteome	P25847	1.2e-83	P25847	1e-102

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8ILK9	DNA-directed DNA polymerase	yes	ATP-binding complete proteome DNA-directed DNA polymerase helicase hydrolase nucleotide-binding nucleotidyl transferase reference proteome transferase			O75417	1.8e-81
Q8ILL6	Calcium/calmodulin-dependent protein kinase	yes	ATP-binding complete proteome kinase nucleotide-binding reference proteome transferase	O96017	7.3e-59	O96017	2.3e-51
Q8IM12	DNA-repair helicase	yes	ATP-binding complete proteome helicase hydrolase nucleotide-binding nucleus reference proteome	P06839	1.3e-82	P06839	5.2e-96
Q9U0J0	Replication factor a protein	yes	complete proteome reference proteome	P27694	4.1e-56	P27694	8.2e-80
C0H4G2	Serine/threonine-protein kinase Nek-2	no	ATP-binding complete proteome kinase nucleotide-binding reference proteome transferase	O14757	2.7e-51	O14757	4.2e-51
C0H4N8	Protein kinase	no	ATP-binding complete proteome kinase nucleotide-binding reference proteome transferase			O14757	6e-53
C0H4W3		no	ATP-binding chromatin regulator coiled coil complete proteome DNA-binding helicase hydrolase merozoite nucleotide-binding nucleus reference proteome transcription transcription regulation	Q92698	7.5e-63	Q03468	6.5e-63

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
C6KSQ1	Serine/threonine protein kinase, Pfnek-5	no	ATP-binding complete proteome kinase nucleotide-binding reference proteome serine/threonine-protein kinase transferase			O14757	5.2e-51
C6KT82	Smarca-related protein	no	complete proteome metal-binding reference proteome zinc zinc-finger	Q92698	2.9e-57	Q92698	1.7e-66
P61074		no	complete proteome DNA replication DNA-binding nucleus reference proteome			P15873	1.8e-53
P61075		no	ATP-binding complete proteome cytoplasm kinase nucleotide-binding phosphoprotein reference proteome serine/threonine-protein kinase transferase	P50613	5.5e-62	P50613	1.5e-53
P62344		no	ATP-binding calcium complete proteome kinase lipoprotein membrane metal-binding myristate nucleotide-binding phosphoprotein reference proteome repeat serine/threonine-protein kinase transferase	O96017	4e-57	O96017	7.2e-51
Q7K6A0	cAMP-dependent protein kinase catalytic subunit	no	ATP-binding complete proteome kinase nucleotide-binding reference proteome serine/threonine-protein kinase transferase			O14757	4.1e-52

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q7K734		no	complete proteome DNA damage DNA repair DNA replication endonuclease exonuclease hydrolase magnesium metal-binding mitochondrion nuclease nucleus phosphoprotein reference proteome	P39748	7.2e-75	P39748	2.7e-75
Q7KQJ9		no	complete proteome DNA replication DNA-binding nucleus reference proteome			P15873	4.9e-51
Q7KQL4		no	complete proteome DNA replication DNA-binding DNA-directed DNA polymerase nucleotidyl transferase reference proteome transferase	P15436	2e-206	P28340	2e-213
Q8I1Q9	Cdc2-related protein kinase 1	no	complete proteome kinase reference proteome transferase	P50613	6.2e-51		
Q8I3I8	RNA helicase-1	no	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	4.5e-60		
Q8I447		no	ATP-binding complete proteome DNA damage DNA repair DNA-binding nucleotide-binding reference proteome	P52701	3e-123	Q03834	2e-150
Q8I4Y8	Uncharacterized protein	no	complete proteome reference proteome			Q02939	1e-126
Q8I579	Uncharacterized protein	no	complete proteome reference proteome			P24482	2.7e-98
Q8I5N7	Conserved Plasmodium protein	no	complete proteome reference proteome			P23179	1.1e-65
Q8IB05	Meiotic recombination protein dmc1-like protein	no	ATP-binding complete proteome nucleotide-binding reference proteome	P25454	1.3e-67	P25454	1e-58

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8IBS5		no	3D-structure ATP-binding calcium complete proteome developmental protein differentiation kinase lipoprotein metal-binding myristate nucleotide-binding phosphoprotein reference proteome repeat serine/threonine-protein kinase transferase	O96017	2.8e-56	O96017	2.3e-51
Q8ICR0		no	ATP-binding calcium complete proteome kinase lipoprotein metal-binding myristate nucleotide-binding reference proteome repeat serine/threonine-protein kinase transferase	O96017	5.2e-63		
Q8IDG5	Uncharacterized protein	no	complete proteome reference proteome			Q12004	6.9e-72
Q8IDW1	Protein kinase 6	no	ATP-binding complete proteome kinase nucleotide-binding reference proteome transferase	P50613	1.1e-51		
Q8IEG4	Serine/threonine protein kinase	no	ATP-binding complete proteome kinase nucleotide-binding reference proteome serine/threonine-protein kinase transferase	O14757	1e-53	O96017	3.6e-51
Q8IEG6	TFIIH basal transcription factor subunit	no	complete proteome metal-binding reference proteome zinc zinc-finger	Q04673	5.8e-70	Q13888	2e-108

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q8IES4		no	ATP-binding complete proteome DNA damage DNA recombination DNA repair DNA replication ligase nucleotide-binding reference proteome	P18858	9e-115	P18858	3e-138
Q8IIS8	Rad51 homolog	no	ATP-binding complete proteome nucleotide-binding reference proteome	P25454	1.2e-72	P25454	6.2e-62
Q8IIW0	PfSNF2L	no	complete proteome reference proteome	Q92698	5e-81	P40352	8.5e-80
Q8IJ38		no	complete proteome DNA replication DNA-binding DNA-directed DNA polymerase nucleotidyl transferase reference proteome transferase	P14284	7.3e-66	P14284	4e-167
Q8IJ90	Uncharacterized protein	no	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	6.6e-52		
Q8IJG6	Uncharacterized protein	no	complete proteome nucleus reference proteome	Q92698	3.1e-70	Q92698	1.7e-70
Q8IKF0	Helicase 45	no	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	5e-60		
Q8IKK0	Cyclin homologue	no	complete proteome cyclin reference proteome			P37366	5.9e-57
Q8ILU6		no	complete proteome DNA damage DNA repair glycosidase hydrolase mitochondrion nucleus reference proteome	P12295	1e-69	P13051	7.7e-87
Q8ILY1	Pfprex	no	complete proteome reference proteome	P00582	5.1e-63	P00582	1.2e-94
Q8IM38	Ribonucleotide reductase small subunit	no	complete proteome reference proteome	P49723	2.1e-72	P49723	1.1e-84

Uniprot	Full name	Putative	Function	Sequence only	E-value	Sequence, structure	E-value
Q9NJU9		no	3D-structure ATP-binding calcium complete proteome developmental protein differentiation kinase metal-binding nucleotide-binding reference proteome repeat serine/threonine-protein kinase transferase	O96017	7.9e-57		
Q9TY94	DEAD box helicase, UAP56	no	ATP-binding complete proteome helicase hydrolase nucleotide-binding reference proteome	O94762	1.9e-53		
Q9U0H1		no	complete proteome DNA replication DNA-binding DNA-directed DNA polymerase nucleotidyl transferase reference proteome transferase	P15436	6.2e-79	P15436	7e-149

Table 6. Other libraries and frameworks used in the project

Name	Description	Role in the project
csb	set of Python libraries for bioinformatics analyses	HHsearch results parsing.
NumPy	Python library for numeric calculations	analysis of the data obtained from hhsearch.py
SciPy	set of Python libraries for mathematic and scientific calculations	methods for histogram calculations
matplotlib	Python library for plotting	methods for histogram plotting
BeautifulSoup	Python library for extracting website data	retrieval of data through UniProt API
jinja2	Python library similar to Django template system used to run calculations outside Django and import results afterwards into main application	generation of results table for REPAIRtoire website
jQuery DataTables.js	JavaScript library based on JQuery library used to add features to online tables	sorting, searching, listing and styles for results on REPAIRtoire website