

## Supporting Information

### **A 3'-5' exonuclease activity embedded in the helicase core domain of *Candida albicans* Pif1 helicase**

Xiao-Bin WEI<sup>1§</sup>, Bo ZHANG<sup>1§</sup>, Nicolas BAZEILLE<sup>2§</sup>, Ying YU<sup>1</sup>, Na-Nv LIU<sup>1</sup>, Brigitte RENE<sup>2</sup>, Olivier MAUFFRET<sup>2</sup>, Xu-Guang XI<sup>1,2, \*</sup>

<sup>1</sup>College of Life Sciences, Northwest A&F University, Yangling, Shaanxi 712100, China


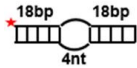
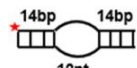


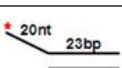


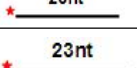
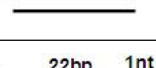
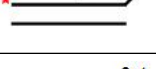
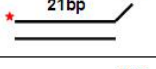
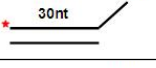

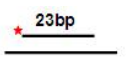
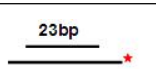
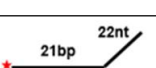

<sup>2</sup>Laboratoire de Biologie et Pharmacologie Appliquée, Ecole Normale Supérieure de Cachan, Université Paris-Saclay, Centre National de la Recherche Scientifique, 61 Avenue du Président Wilson, 94235 Cachan, France

\*Corresponding authors: email: [xxi01@ens-cachan.fr](mailto:xxi01@ens-cachan.fr)

Tel: 33 01 4740 7754; Fax: 33 01 4740 7754;

§These authors contributed equally to this work.

**Table S1.** substrates used for activities assays

Name	Length (nt)	Structure	Sequences (5'-3')
S <sup>20</sup>	20		*TTGATTTGATGTACGGATGT
Bb <sup>4</sup>	40/40		*GCAGTAGGCCTAGCATGAGTTAGGACTGACAGCTGCATGG CCATGCAGCTGTCAGTCCATTGTCATGCTAGGCCTACTGC
Bb <sup>12</sup>	40/40		*GCAGTAGGCCTAGCATGAGTTAGGACTGACAGCTGCATGG CCATGCAGCTGTCACAGGATTGAGTAGCTAGGCCTACTGC
S <sup>12</sup> -5F	12		*GTACGGATGTCT
S <sup>12</sup> -3F	12		GTACGGATGTCT*
S <sup>20</sup> D <sup>23</sup> -5F	43/23		*ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC GACATCCTTACACTTTATTTTTG
S <sup>20</sup> D <sup>23</sup> -3F	43/23		ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC* GACATCCTTACACTTTATTTTTG
A	23		*GTGTAAGGATGTCACGATCATTG
B	23/23		*GTGTAAGGATGTCACGATCATTG CAATGATCGTGACATCCTTACAC
C	23/22		*GTGTAAGGATGTCACGATCATTG AATGATCGTGACATCCTTACAC
D	23/21		*GTGTAAGGATGTCACGATCATTG ATGATCGTGACATCCTTACAC
E	43/30		*ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC TTTATTTTTGCTACAAAATAAATGTAACAT
F	43/25		*ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC TTTTGCTACAAAATAAATGTAACAT
G	23/43		*GTGTAAGGATGTCACGATCATTG TTTTTTTTTTCAATGATCGTGACATCCTTACACTTTTTTTTTT
H	23/43		CACTTTATTTTTGCTACAAAATA *ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC
I	43/21		*ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC GCTACAAAATAAATGTAACAT
J	43/43		*ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC CTGTAGGAATGTGAAATAAAAAGCTACAAAATAAATGTAACAT
K	43/43		*ATGTTACATTTATTTTGTAGCAAAAATAAAAGTGTAAGGATGTC CTGTAGGAATGTGAAATAAAAAGCTACAAAATAAATGTAACAT*

\* , Fluorescein

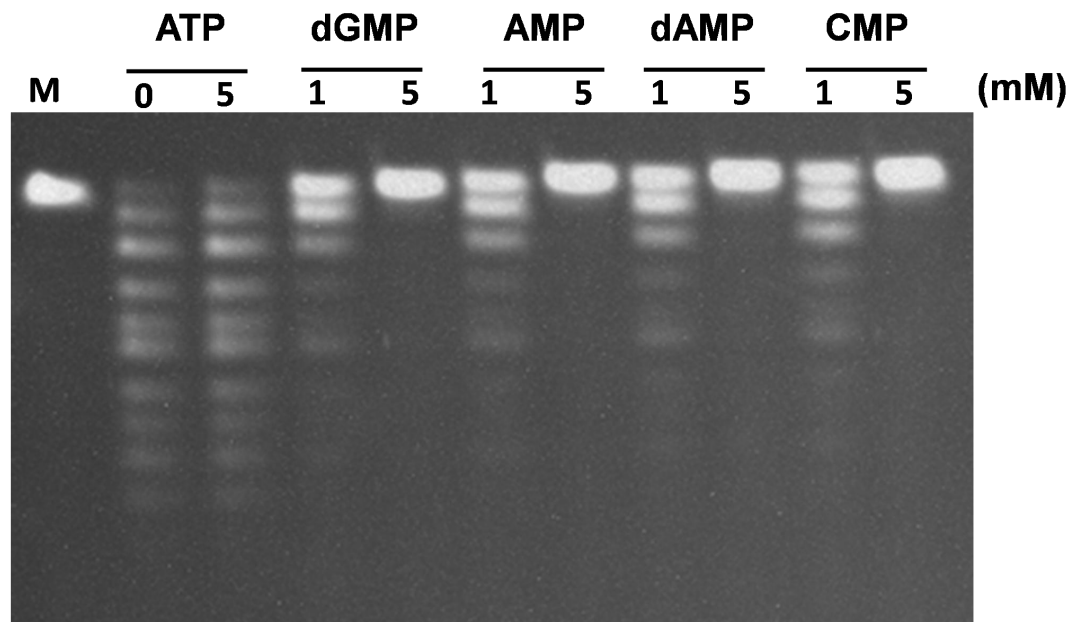
\* , hexachlorofluorescein

**Table S2.** Primers used for mutagenesis

Name	Sequence (5'-3')
CaPif1 <sup>D582V</sup> -F	GACGACACCGCGCGT <u>GT</u> ATTCTGCCGTCTGTCT
CaPif1 <sup>D582V</sup> -R	AGACAGACGGCAGAACTCACGCGCGGTGTCGTC
CaPif1 <sup>D794V</sup> -F	CAGTGGACCACCGAAG <u>TT</u> GAAGACGGTACCGTT
CaPif1 <sup>D794V</sup> -R	AACGGTACCGTCTTCAACTTCGGTGGTCCACTG
CaPif1 <sup>E473A</sup> -F	GTTCTGATCATCGAC <u>GCG</u> ATCTCTATGGTTGAC
CaPif1 <sup>E473A</sup> -R	GTCAACCATAGAGATCGCGTCGATGATCAGAAC

**Table S3.** Proteins accession numbers for sequence alignment

<b>Orgnism</b>	<b>Strain</b>	<b>Accession numbers</b>
<i>Naumovozya dairenensis</i>	CBS421	XP_003670036.1
<i>Naumovozya castellii</i>	CBS4309	XP_003673236.1
<i>Saccharomyces cerevisiae</i>	S288c	NP_013650.1
<i>Kazachstania africana</i>	CBS2517	XP_003958958.1
<i>Candida glabrata</i>	CBS138	XP_449255.1
<i>Zygosaccharomyces rouxii</i>	CBS732	XP_002495556.1
<i>Torulaspora delbrueckii</i>	CBS1146	XP_003680891.1
<i>Vanderwaltozyma polyspora</i>	DSM70294	XP_001643959.1
<i>Tetrapisispora phaffii</i>	CBS4417	XP_003684156.1
<i>Tetrapisispora blattae</i>	CBS6284	XP_004181037.1
<i>Ashbya gossypii</i>	ATCC10895	NP_984987.1
<i>Eremothecium cymbalariae</i>	DBVPG#7215	XP_003644227.1
<i>Kluyveromyces lactis</i>	NRRL Y-1140	XP_451763.1
<i>Lachancea thermotolerans</i>	CBS6340	XP_002553410.1
<i>Millerozyma farinosa</i>	CBS7064	XP_004197165.1
<i>Debaryomyces hansenii</i>	CBS767	XP_459513.2
<i>Meyerozyma guilliermondii</i>	ATCC6260	XP_001484923.1
<i>Lodderomyces elongisporus</i>	NRRL YB-4239	XP_001523552.1
<i>Candida orthopsilosis</i>	Co90-125	XP_003870993.1
<i>Candida tropicalis</i>	MYA-3404	XP_002548760.1
<i>Candida albicans</i>	SC5314	XP_712376.1
<i>Candida dubliniensis</i>	CD36	XP_002422246.1
<i>Komagataella pastoris</i>	GS115	XP_002491346.1
<i>Emiliana huxleyi virus</i>	86	YP_293894.1
<i>Homo sapiens</i>		NP_079325.2



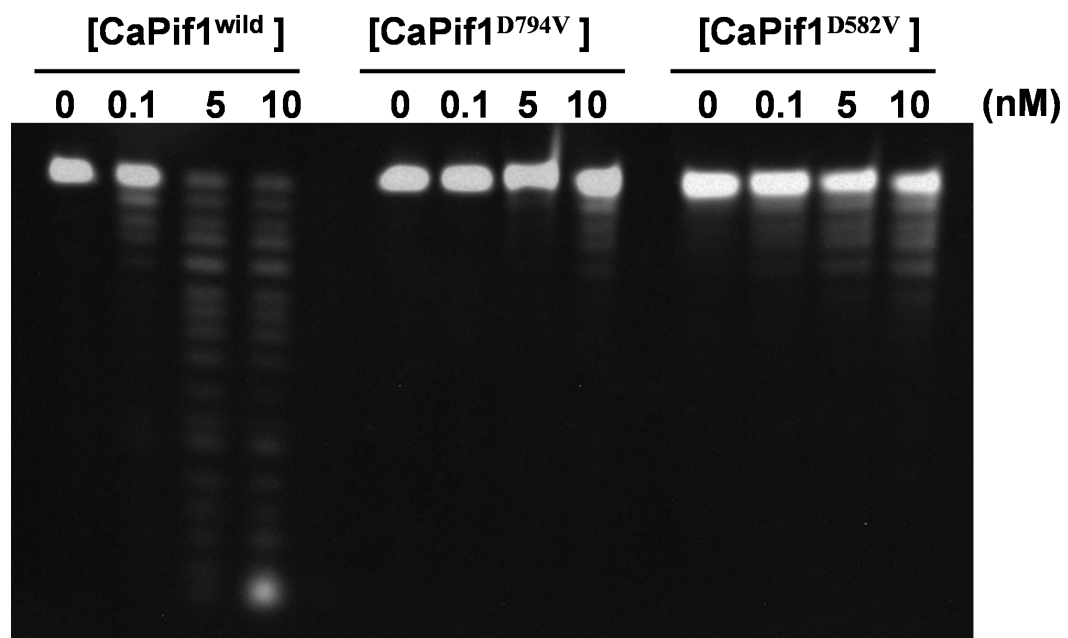
**Figure S1.** CaPif1 exonuclease activity was inhibited by nucleotide monophosphates. The experiments were performed under the standard exonuclease assay conditions with substrate S12-5F.

Strains	Conserved E/D	Conserved E/D
<i>C. albicans</i> SC5314	572 DGNVPDDTARDFCRLSRPLKC 592	785 VVEPEQWTTEDDEDGT-VLVSR 804
<i>C. dubliniensis</i> CD36	573 DGNVPDDTVRDFCRLSRPLKC 593	786 VVEPEQWTTEDDEGK-VLVSR 805
<i>C. tropicalis</i> MYA-3404	512 DGMINKESIVRFQQLERPLKC 532	685 IVDVESWSTEDVETESILAKR 705
<i>L. elongisporus</i> NRRL YB-4239	715 DGRISDRTVREFQRLSRPLSC 735	927 LVEPEEWTVEDEDGH-VLVSR 946
<i>C. orthopsilosis</i> Co90-125	622 DGRISESTIQEFRRLSRPLEC 642	821 LVEPEQWTIEDEDGR-ALVSR 840
<i>M. farinosa</i> CBS7064	428 SGKVTSDTIAEFNKLTRPLET 448	654 LVEPEEWTVEDEHEN-VLVRR 673
<i>D. hansenii</i> CBS767	480 NGKVNNETIAEFKRLSRKLDN 500	707 LVEPEQWTIEDENEN-VLARR 726
<i>M. guilliermondii</i> ATCC6260	368 SGQIGPDTEMEFKKLARKLEC 388	596 LVEPETWTVEDEHQQ-VLASR 615
<i>N. dairenensis</i> CBS421	473 LGKIDENTEREFKKLTRPLA- 492	706 LVEPENWAIEDENEK-PIVSR 725
<i>N. castellii</i> CBS4309	475 LGKIDDETEREFRKLSRSLP- 494	709 LVEPECWEIEDENSK-PIVSR 728
<i>Z. rouxii</i> CBS732	441 LGQIDEDTEMEFKRLNRPLP- 460	678 LVEPEDWAIEDENEK-PLVSR 697
<i>T. delbrueckii</i> CBS1146	526 LGQIDAEATEREFKKLNRPLP- 545	764 LVEPESWAIEDDNEK-SLVSR 783
<i>K. africana</i> CBS2517	418 LGQIDEDTEREFKKLNRPLP- 437	654 LVEPEDWAIEDEHEK-PIVSR 673
<i>S. cerevisiae</i> S288c	433 LGNIDDETEREFKKLNRPLP- 452	672 LVEPEDWAIEDENEK-PLVSR 691
<i>V. polyspora</i> DSM70294	553 LGEIDIEETEDEFKCLSRELE- 572	791 LVEPEDWSIEDENEK-PIVSR 810
<i>C. glabrata</i> CBS138	455 LGKIDMETEMEFKKLNRPLP- 474	690 LVEPEDWAIEDENEK-PLVSR 709
<i>T. phaffii</i> CBS4417	547 VGNVSDTEMEFKKLNRPLP- 566	787 LVEPEDWAIEDEDEK-PIVSR 806
<i>T. blattae</i> CBS6284	442 LGKLDFKMEIEFKKLQRPLP- 461	678 LVEPEDWSVEDENEK-PIVSR 697
<i>A. gossypii</i> ATCC10895	393 LGKISEETEAEFRKLSRPLV- 412	634 LVREEAWAIEDENEK-PLVTR 653
<i>E. cymbalariae</i> DBVPG#7215	474 LGKISPETELEFRKLSRPLE- 493	715 LVREEAWAVEDENEK-PLVTR 734
<i>K. lactis</i> NRRL Y-1140	454 MGVISDRTLKEFAELSRPLP- 473	696 LVRPETWAMEDEKDV-PLVSR 715
<i>L. thermotolerans</i> CBS6340	488 MGKISLETELEFKKLNRPLP- 507	723 LVQPEDWAVEDELQK-PIVSR 742
<i>K. pastoris</i> GS115	427 NGRISPRVEAEFQKLNRPLPV 447	655 LVEPEQFSINDIDEK-PLASR 674
<i>E. huxleyi</i> virus 86	196 IGKCATETENALAKCHVSVKP 216	327 VMQMERAEETEOKFMN-SKCTR 346
<i>H. sapiens</i>	396 LGRCSDVETRQLQATASHKVG 416	523 VIHADRWTVQATGGQ--LLSR 542

**Figure S2.** Sequence alignment of selected fungal species using BLAST. The accession numbers of the proteins sequences were summarized in the Supplementary Table S3.

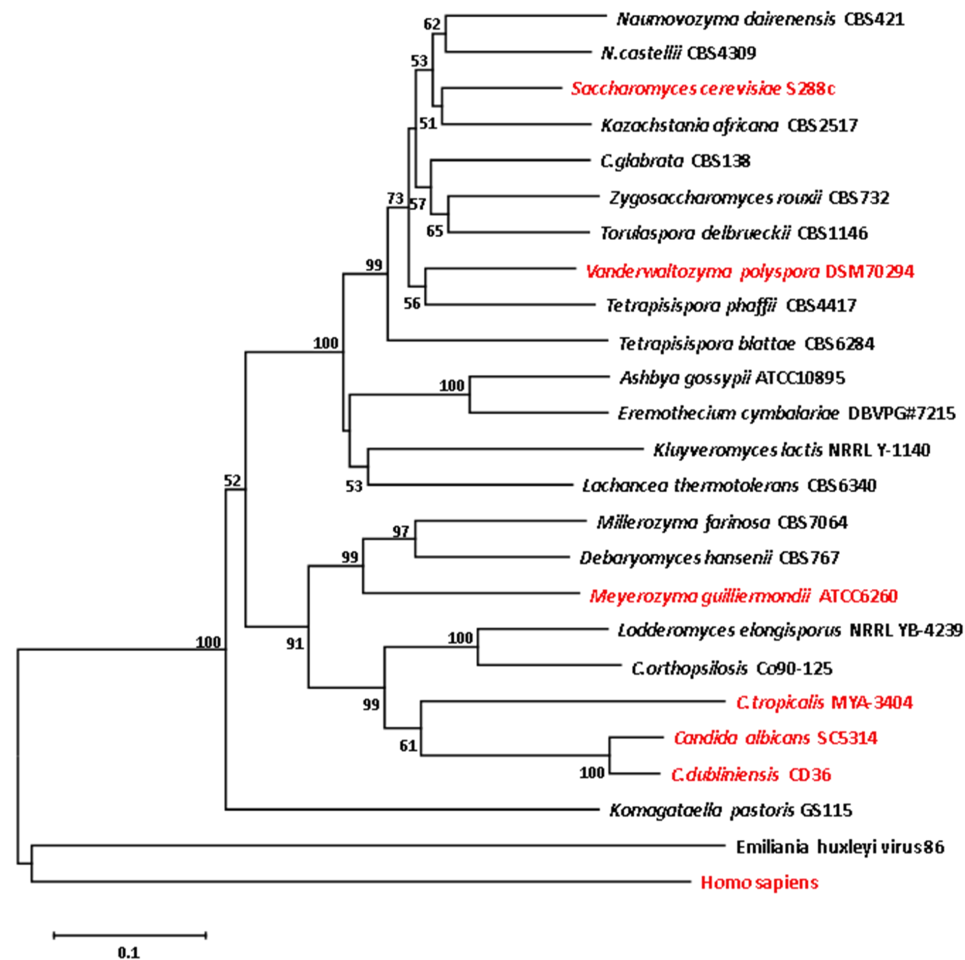
<i>C.albicans</i>	<i>pif1</i>	563	FIDMLNLRDGNVDDTARDFCRLSRPLKC-PEG	595
<i>H.sapiens</i>	<i>WRN-exo</i>	124	LKMLLENKAVKKAGVGIEGDQWKLLRDFDIKLN	157
<i>E.coli</i>	<i>KF-exo</i>	405	LKPLLEDEKALKVGNLKYDRGILAN-YGIELRG	437
<i>A.Thaliana</i>	<i>1VK0</i>	100	LYRFFASKFVTFVGVQIEEDLD-LLRENHG-LVI	131
<i>C.albicans</i>	<i>pif1</i>	776	DGITFRTVVVE-PEQWTTEDEDGTVLVSRIOFPL	808
<i>H.sapiens</i>	<i>WRN-exo</i>	197	CSNWSKFPLTEDQKLYAATDAYAGFIIYRNLELL	230
<i>E.coli</i>	<i>KF-exo</i>	483	QLTFNQIAL-EEAGRYAAEDADVTLQLHLKMWPD	515
<i>A.Thaliana</i>	<i>1VK0</i>	173	IEAKWEKAGPEEQLEAAAIEGWLIIVNVWDQLSDE	206

**Figure S3.** Sequence alignment of CaPif1 with WRN-exo, Kf-exo and *A. thaliana* homolog was carried out using PHI-BLAST. The numbers between the sequences indicate the amino-acid position relative to the amino terminus of each protein and the conserved aspartic/glutamic residues were colored in red.

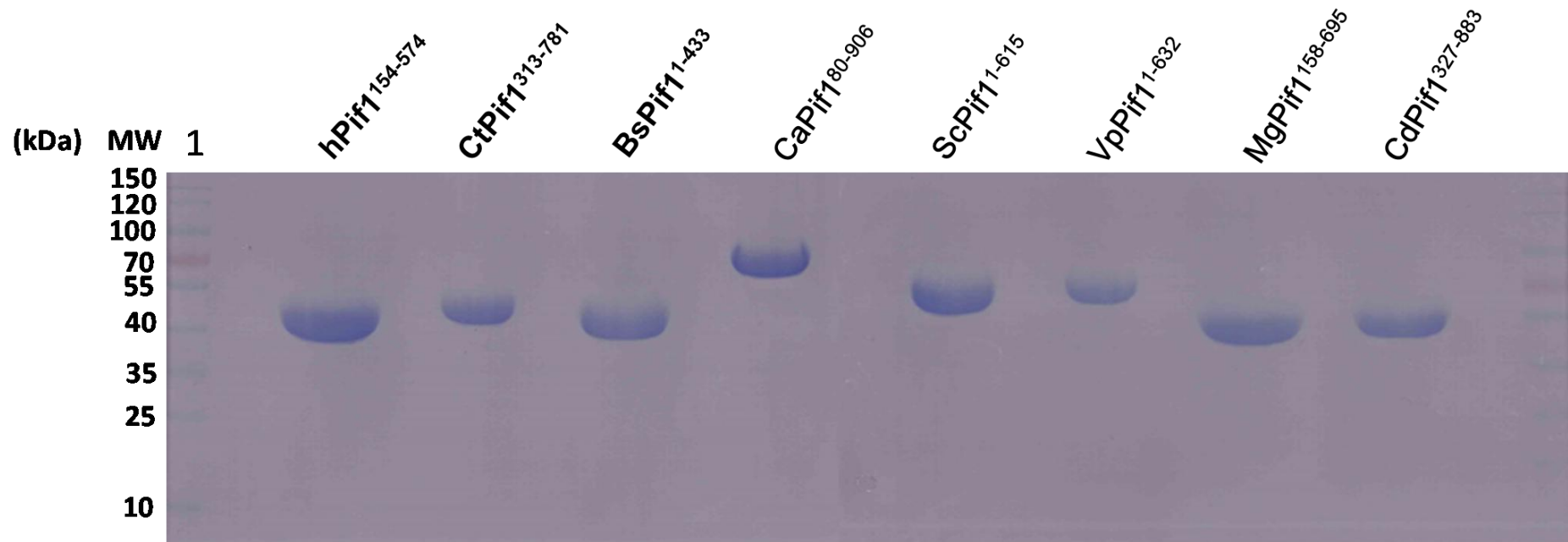


**Figure S4.** Gel electrophoresis measurements of exonuclease activity of wild-type CaPif1 , mutants CaPif1<sup>D582V</sup> and CaPif1<sup>D794V</sup> purified proteins. The reactions were carried out at 37°C for 5min in buffer A with 400nM 5'-fluorescein labeled 20-nt ssDNA (substrate S<sup>20</sup>) at the indicated protein concentrations.

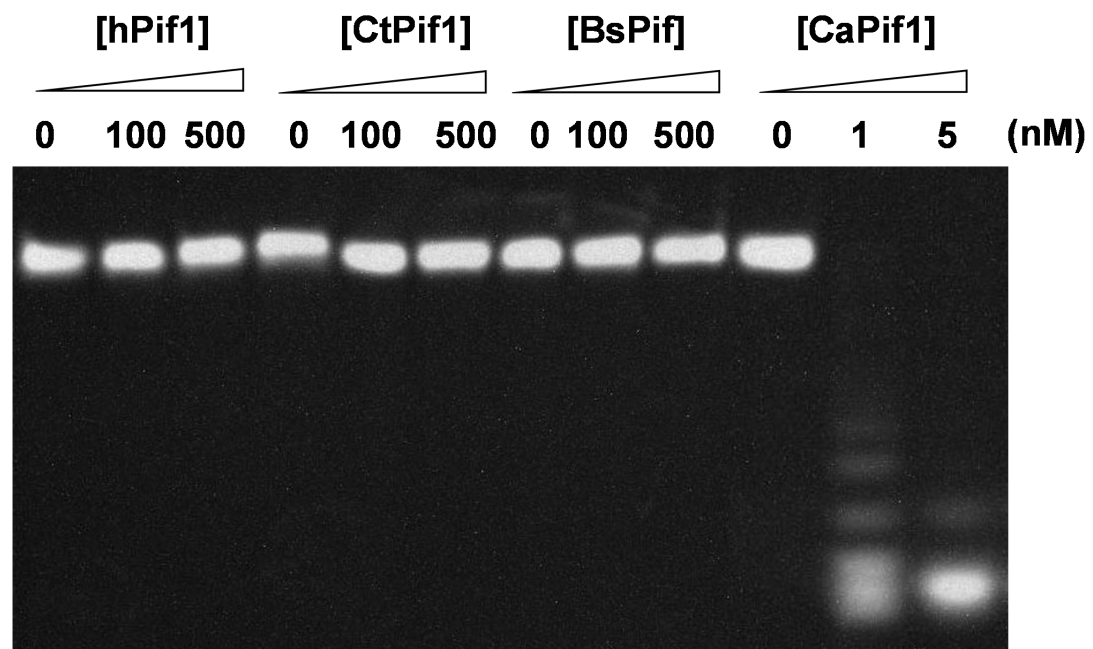




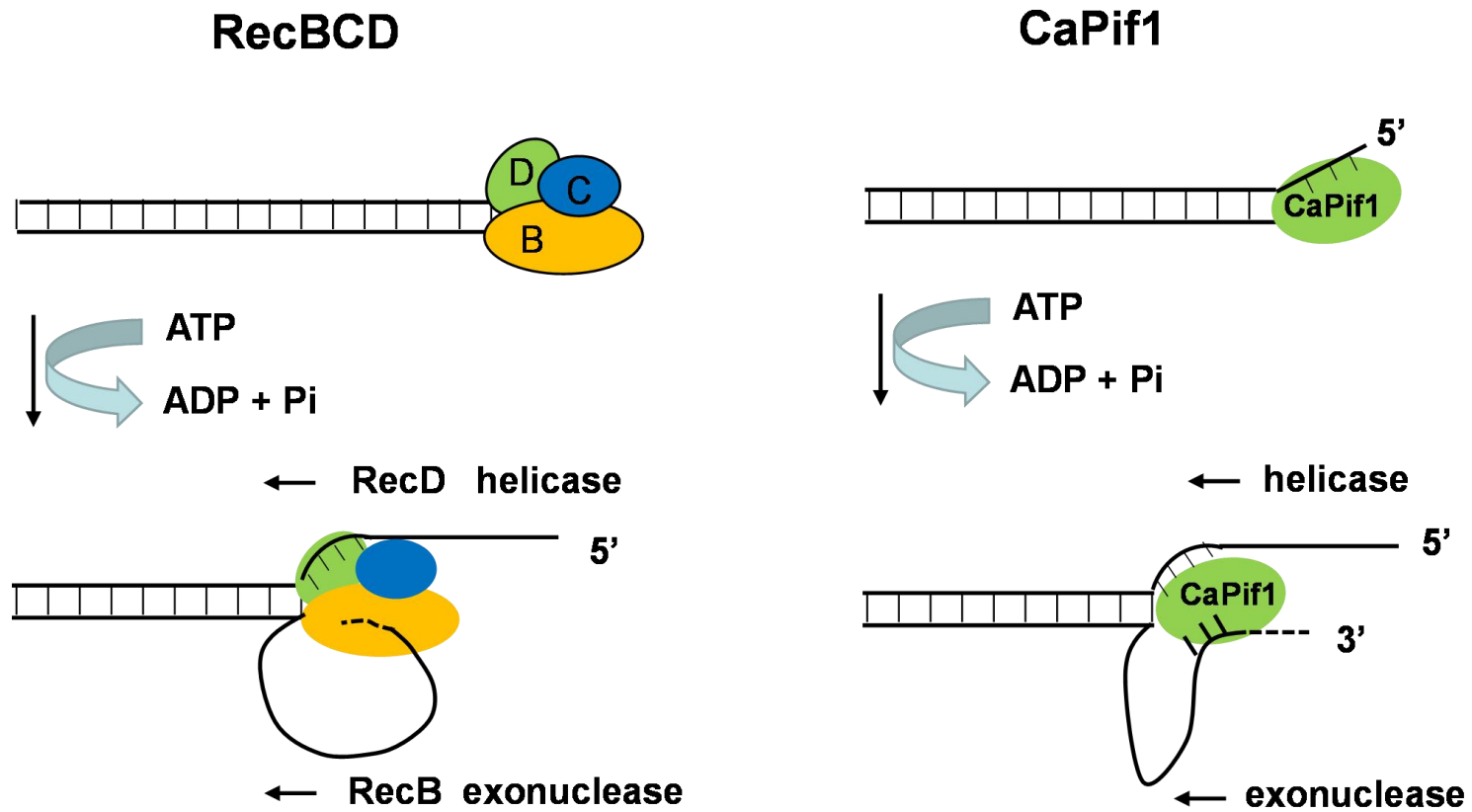
**Figure S5.** Selected Pif1 helicases from the constructed phylogenetic tree were used for confirming the relationship between the presence of the conserved aspartic/glutamic residues and exonuclease activities. The protein sequences accession numbers are summarized in Table S3.



**Figure S6.** Different kinds of pif1 protein purity. SDS-PAGE analysis of the purified 8 kinds of Pif1 protein. The gel (10% acrylamide) was stained using Coomassie brilliant blue R250. Lane 1 show protein molecular maker and all the protein (20 $\mu$ g), respectively.



**Figure S7.** hPif1, BsPif1 and CtPif1 do not display any appreciable exonuclease activities when high concentration of protein was added. The experiments were performed under the standard exonuclease assay conditions with substrate S12-5F.



**Figure S8.** In accordance to the properties of RecBCD helicase, we predict the model that coupling of helicase and nuclease activity in CaPif1.