

# Genome-wide survey and phylogenetic analysis on subunit sequences of eukaryotic DNA polymerase delta

H.F. Song<sup>1</sup>, H.Q. Chen<sup>1</sup>, Y. Wang<sup>2</sup> and Y.J. Zhou<sup>1</sup>

<sup>1</sup>Institute of Life Sciences, Jiangsu University, Zhenjiang, China

<sup>2</sup>School of Food and Biological Engineering, Jiangsu University, Zhenjiang, China

Corresponding author: Y.J. Zhou

E-mail: yajingzhou@ujs.edu.cn

Genet. Mol. Res. 13 (4): 9558-9568 (2014)

Received October 22, 2013

Accepted May 10, 2014

Published November 12, 2014

DOI <http://dx.doi.org/10.4238/2014.November.12.4>

**ABSTRACT.** DNA polymerase  $\delta$  is not only the major replicative enzyme in eukaryotic chromosomal DNA synthesis but is also the primary polymerase for most DNA repair pathways. However, the subunit composition of polymerase  $\delta$  varies in different organisms. While polymerase  $\delta$  in many eukaryotic species has all 4 subunits (POLD1, 2, 3, and 4), many other organisms do not possess POLD4. Whether POLD4 is indispensable and why these differences exist are unknown. In the present study, we identified the POLD4 protein sequences of 218 eukaryotic species and determined the POLD1, 2, and 3 protein sequences of 55 species representing various taxonomic groups. No insect and nematode species examined possessed POLD4. Approximately 80% of protozoan species did not contain POLD4. Nearly 50% of fungal species did not contain POLD4. Other animal and plant species are expected to contain POLD4. Phylogenetic analyses of POLD1, 2, 3, and 4 sequences revealed that most animal and plant species inherited DNA polymerase  $\delta$  from protozoa, whereas some other animal and plant species may have inherited polymerase  $\delta$  directly from

fungi. Because a large number of protozoan and fungal species do not possess POLD4, current insect and nematode species lacking POLD4 may have evolved from ancestor protozoan species lacking POLD4; thus, other protozoan and animal species lacking POLD4 may share a similar evolutionary history. Future studies should examine the origin and indispensability of POLD4 in various organisms.

**Key words:** BLAST search; DNA polymerase delta; Eukaryotic species subunits; Phylogenetic analysis

## INTRODUCTION

A significant characteristic of all living organisms is the ability to reproduce and transmit genomic information to their offspring. Three major multi-subunit DNA polymerases, including polymerase  $\alpha$  (pol  $\alpha$ ), polymerase  $\delta$  (pol  $\delta$ ), and polymerase  $\epsilon$  (pol  $\epsilon$ ), are responsible for eukaryotic chromosomal DNA replication. Pol  $\delta$  is thought to play a central role in eukaryotic DNA synthesis of both the leading and the lagging strands as well as in DNA re-synthesis (gap-filling) of various DNA repair processes (Waga and Stillman, 1998; Bell and Dutta, 2002; Hubscher et al., 2002; Garg and Burgers, 2005).

The subunit composition of pol  $\delta$  varies among different eukaryotic species. For instance, pol  $\delta$  in the budding yeast *Saccharomyces cerevisiae* is composed of 3 subunits, Pol3p, Pol31p/Hys2, and Pol32p (Burgers and Gerik, 1998; Gerik et al., 1998), while that of the fission yeast *Schizosaccharomyces pombe* is a 4-subunit complex consisting of Pol3, Cdc1, Cdc27, and Cdm1 (Reynolds et al., 1998; Zuo et al., 2000; Bermudez et al., 2002). Cdm1 is the smallest subunit of fission yeast pol  $\delta$  and has no apparent homolog in budding yeast. Similar to *S. pombe*, mammalian pol  $\delta$  consists of 4 subunits, including p125, p50, p68/p66, and p12 (Byrnes et al., 1976; Hughes et al., 1999; Mo, 2000; Liu et al., 2000). The catalytic subunit p125, harboring both 5'-3'-DNA polymerase and 3'-5'-exonuclease activities, is homologous to fission yeast Pol3 and budding yeast Pol3p. The small subunit p50 is a homolog of Cdc1 and Pol31p/Hys2, whereas the 3rd subunit p68/p66, containing a proliferating cell nuclear antigen-binding motif, is viewed as a homolog of Cdc27 and Pol32p. The smallest subunit, p12, is a homolog of Cdm1. Thus, DNA polymerase  $\delta$  consists of multiple subunits, 4 of which are referred to as the catalytic subunit (POLD1), small subunit (POLD2), regulatory subunit (POLD3), and 4th subunit (POLD4).

POLD4 appears to play a minor role in pol  $\delta$  function because Cdm1 of *S. pombe* pol  $\delta$  is non-essential and is absent in *S. cerevisiae*. However, recent studies indicate that the 4th subunit of mammalian pol  $\delta$ , p12, serves a structural function, which not only helps to maintain the stability of pol  $\delta$  but also interacts with the core dimer to alter the conformation and/or structure of the catalytic site of pol  $\delta$  (Li et al., 2006; Meng et al., 2010; Zhou et al., 2012). It also functions in cell proliferation and maintenance of genomic stability of human cells (Huang et al., 2010a,b). Depletion of the p12 subunit through its degradation, which is regulated by the ataxia telangiectasia and Rad3-related/chk1 complex and dependent on ubiquitination, can result in a conversion of pol  $\delta$  from a heterotetramer to a trimer with altered enzymatic properties (Zhang et al., 2007; Meng et al., 2009). Moreover, *in vitro* base excision repair (BER) assays revealed that p12 plays an important role in uracil-initiated BER by modulating the rates of single-nucleotide BER and long patch BER during the repair process (Zhou et al., 2011).



To identify organisms with and without POLD4, we conducted BLASTp and/or tBLASTn searches against sequences of all organisms listed in GenBank's Map View database (<http://www.ncbi.nlm.nih.gov/mapview/>). For both BLAST searches, human POLD4 (GenBank accession No. NP\_066996.3) was used as the query sequence with default settings. For each organism, a BLASTp search was conducted first against protein databases (including RefSeq protein, Non-RefSeq protein, Build protein, and *Ab initio*) of that organism to determine whether a potential POLD4 protein sequence existed in that organism. If BLASTp failed to return the POLD4 sequence, a tBLASTn search was conducted against various nucleotide databases (including genomic, expressed sequence tag, and whole genome short-gun sequence databases) of that organism. When tBLASTn hit sequences were separated into 2 or more coding regions, intron splice sites were analyzed using the NetGene2 online tool (<http://www.cbs.dtu.dk/services/NetGene2/>).

### BLAST searches and identification of POLD1, POLD2, and POLD3 sequences

Identification of the POLD1, 2, and 3 sequences followed the same strategy used for identifying the POLD4 sequence. Similarly, POLD1, 2, and 3 sequences are highly conserved across various organisms. POLD1 sequences show many highly conserved motifs as shown in the multiple alignments of POLD1 sequences ([Figure S1](#)). We selected its conserved motif from sites 626-643 (DPDVIxGYxIxxxxxYL, where x indicates any amino acid) as the standard for determining whether an unknown sequence was a POLD1 sequence. POLD2 sequences contain a highly conserved motif from sites 1402-1413 (PTAPDTLxxYPF), which was used as the standard for identifying POLD2 sequences ([Figure S1](#)). POLD3 sequences are not as highly conserved as POLD1, 2, and 4 sequences. Rather, they show less highly conserved motifs from sites 1461-1470 (LxxIxxxVxD), 1489-1496 (IVTYKxLS), 1517-1529 (VxxNxAKxMLxxY), and 1563-1571 (LxxxYLVxG) ([Figure S1](#)). An unknown sequence with 2 or more of these conserved motifs was regarded as a POLD3 sequence.

To identify the POLD1, 2, and 3 sequences, human POLD1, 2, and 3 (GenBank accession Nos. NP\_002682.2, NP\_001120690.1, and NP\_006582.1, respectively) were used as queries to conduct BLASTp searches against protein databases (RefSeq protein and *Ab initio* protein) of the 55 species chosen to represent different taxonomic groups (Table 1) with default settings.

### Multiple sequence alignment

Multiple sequence alignment was performed using ClustalW (version 5.0) implemented in MEGA 5 (Tamura et al., 2011) with default settings. Multiple aligned sequences were highlighted in GeneDoc Multiple Sequence Alignment Editor and Shading Utility (Version 2.6.02) (Nicholas et al., 1997) and copied as a rich text file for further annotation.

### Phylogenetic analysis

Phylogenetic analysis was performed using MEGA 5 (Tamura et al., 2011). Three different algorithms were employed for tree reconstruction: neighbor-joining (NJ), maximum parsimony, and maximum likelihood (ML). The NJ and ML trees were constructed using the *p*-distance and Jones-Taylor-Thornton (Jones et al., 1992) models, respectively. Bootstrap values were obtained using 1000 replicates. All other parameters used were default values.

**Table 1.** Protein accession numbers of POLD1, 2, 3, and 4 in 55 eukaryotic species.

No.	Species	POLD1	POLD2	POLD3	POLD4
1	<i>Homo sapiens</i>	NP_002682.2	NP_001120690.1	NP_006582.1	NP_066996.3
2	<i>Taeniopygia guttata</i>	XP_002191467.2	XP_002196136.2	XP_002187642.2	EF191831.1†
3	<i>Gallus gallus</i>	XP_426179.4	XP_003643965.2	NP_001006284.1	Not found
4	<i>Meleagris gallopavo</i>	XP_003204329.1	hmm.65064.p*	hmm.139424.p*	Not found
5	<i>Anolis carolinensis</i>	XP_003222764.1	XP_003224261.1	XP_003227524.1	XP_003215017.1
6	<i>Xenopus (Silurana) tropicalis</i>	NP_001039204.1	NP_001005794.1	NP_001016556.1	Not found
7	<i>Danio rerio</i>	NP_001034899.1	NP_998468.1	NP_001036234.1	NP_001243129.1
8	<i>Tetraodon nigroviridis</i>	CAG02717.1	CAF90299.1	CAF97746.1	Not found
9	<i>Salmo salar</i>	NP_001167152.1	NP_001136183.1	NP_001133358.1	NP_001140097.1
10	<i>Aedes aegypti</i>	XP_001648163.1	XP_001658442.1	XP_001657202.1	Not found
11	<i>Apis mellifera</i>	XP_623795.1	XP_001121190.2	XP_003251482.1	Not found
12	<i>Bombus impatiens</i>	XP_003484918.1	XP_003488499.1	XP_003491626.1	Not found
13	<i>Drosophila melanogaster</i>	NP_524099.2	NP_647659.1	NP_477397.1	Not found
14	<i>Ixodes scapularis</i>	XP_002399458.1	XP_002410434.1	XP_002402639.1	Not found
15	<i>Pediculus humanus corporis</i>	XP_002429857.1	XP_002432055.1	XP_002431427.1	Not found
16	<i>Brugia malayi</i>	XP_001892851.1	XP_001896497.1	XP_001898030.1	Not found
17	<i>Caenorhabditis elegans</i>	NP_506017.1	NP_502177.1	NP_506088.1	Not found
18	<i>Branchiostoma floridae</i>	XP_002608243.1	XP_002607199.1	XP_002593439.1	XP_002612025.1
19	<i>Strongylocentrotus purpuratus</i>	XP_003725042.1	XP_782368.2	XP_791644.1	XP_003726058.1
20	<i>Hydra magnipapillata</i>	XP_004209898.1	XP_002155623.1	XP_002156160.2	Not found
21	<i>Amphimedon queenslandica</i>	XP_003388637.1	XP_003391229.1	XP_003387538.1	XP_003384125.1
22	<i>Ciona intestinalis</i>	XP_002125320.1	XP_002126992.1	XP_002125821.1	XP_002126252.1
23	<i>Saccoglossus kowalevskii</i>	hmm.49154.p	XP_002738996.1	XP_002730626.1	FF435957.1†
24	<i>Nematostella vectensis</i>	XP_001641357.1	XP_001630066.1	XP_001624310.1	XP_001633359.1
25	<i>Plasmodium falciparum</i> 3D7	XP_001347450.1	XP_001351155.1	XP_002808715.1	Not found
26	<i>Leishmania infantum</i> IPCM5	XP_001468284.1	XP_001466176.1	XP_001468694.1	Not found
27	<i>Cryptosporidium parvum</i> Iowa II	XP_627772.1	XP_627069.1	XP_626108.1	XP_627350.1
28	<i>Dictyostelium discoideum</i> AX4	XP_638283.1	XP_640553.1	XP_001732994.1	Not found
29	<i>Capsaspora owczarzaki</i> ATCC 30864	EFW39690.2	EFW46408.1	EFW46520.2	EFW43759.1
30	<i>Monosiga brevicollis</i> MX1	XP_001748370.1	XP_001750194.1	XP_001743902.1	XP_001743111.1
31	<i>Paramecium tetraurelia</i> strain d4-2	XP_001443230.1	XP_001426415.1	XP_001452275.1	XP_001447625.1
32	<i>Perkinsus marinus</i> ATCC 50983	XP_002773286.1	XP_002783195.1	XP_002786522.1	XP_002782838.1
33	<i>Salpingoeca</i> sp ATCC 50818	EGD74190.1	EGD78703.1	EGD82577.1	EGD77988.1
34	<i>Guillardia theta</i> CCMP2712	EKX51650.1	EKX48257.1	EKX36550.1	EKX54558.1
35	<i>Oryza sativa</i> Japonica Group	NP_001067405.1	NP_001048841.1	NP_001042334.2	NP_001063696.1
36	<i>Sorghum bicolor</i>	XP_002450450.1	XP_002465940.1	XP_002454978.1	XP_002462682.1
37	<i>Arabidopsis thaliana</i>	NP_201201.2	NP_973664.1	NP_565185.1	NP_563854.1
38	<i>Populus trichocarpa</i>	EEE94038.1	EEE92828.1	EEF11454.1	XP_002319772.1
39	<i>Chlamydomonas reinhardtii</i>	XP_001689909.1	XP_001694397.1	hmm510274*	XP_001701601.1
40	<i>Physcomitrella patens</i> subsp <i>patens</i>	XP_001754656.1	XP_001772952.1	XP_001786191.1	XP_001781255.1
41	<i>Chlorella variabilis</i>	EFN58392.1	EFN55049.1	EFN51119.1	EFN53288.1
42	<i>Ostreococcus lucimarinus</i> CCE9901	XP_001416112.1	XP_001420728.1	XP_001416472.1	XP_001418707.1
43	<i>Micromonas</i> sp RCC299	XP_002502470.1	XP_002505212.1	XP_002507982.1	XP_002500135.1
44	<i>Selaginella moellendorffii</i>	XP_002983728.1	XP_002971713.1	XP_002980528.1	XP_002984250.1
45	<i>Nectria haematococca</i> mpVI 77-13-4	XP_003050401.1	XP_003050394.1	XP_003044639.1	XP_003052063.1
46	<i>Podospora anserina</i> S mat+	XP_001911857.1	XP_001911351.1	XP_001910383.1	XP_001909378.1
47	<i>Saccharomyces cerevisiae</i> S288c	NP_010181.2	NP_012539.1	NP_012577.1	Not found
48	<i>Yarrowia lipolytica</i>	XP_499875.1	XP_502431.1	XP_501056.1	XP_501078.1
49	<i>Schizosaccharomyces pombe</i> 972h-	NP_596124.1	NP_594406.1	NP_595419.1	NP_595948.1
50	<i>Cryptococcus gattii</i> WM276	XP_003196775.1	XP_003192401.1	XP_003193166.1	XP_003197140.1
51	<i>Ustilago maydis</i> 521	XP_757605.1	XP_761123.1	XP_759913.1	XP_756674.1
52	<i>Laccaria bicolor</i> S238N-H82	XP_001883646.1	XP_001878224.1	XP_001874578.1	XP_001875069.1
53	<i>Encephalitozoon intestinalis</i> ATCC 50506	XP_003073533.1	XP_003073562.1	XP_003073220.1	Not found
54	<i>Nosema ceranae</i> BRL01	XP_002995919.1	XP_002995059.1	XP_002995129.1	Not found
55	<i>Batrachochytrium dendrobatidis</i> JAM81	EGF80322.1	EGF78336.1	EGF79296.1	EGF81212.1

Of the selected species, No. 1 is the representative of mammal, Nos. 2-9 of other vertebrates, Nos. 10-15 of insect, Nos. 16 and 17 of nematode, Nos. 18-24 of other invertebrates, Nos. 25-34 of protozoa, Nos. 35 and 36 of monocots, Nos. 37 and 38 of eudicots, Nos. 39-44 of algae and mosses, Nos. 45 and 46 of Pezizomycotina from Ascomycota of fungi, Nos. 47 and 48 of Saccharomycotina from Ascomycota of fungi, No. 49 of Schizosaccharomycetes from Ascomycota of fungi, Nos. 50-52 of Basidiomycota of fungi, Nos. 53 and 54 of Microsporidia of fungi, No. 55 of Chytridiomycota of fungi. Those labeled “\*” and “†” are from “*Ab initio* protein” and “mRNA” sequences, respectively.

## RESULTS AND DISCUSSION

### Distribution of POLD4 in eukaryotes

Through exhaustive BLAST searches against all available protein and nucleotide sequence databases in GenBank (<http://www.ncbi.nlm.nih.gov/mapview/>), we obtained a general view of the distribution of POLD4 in eukaryotic species. Table 2 and [Table S1](#) summarize the distribution of POLD4 in 218 eukaryotic species, each of which was chosen to represent a unique genus of the surveyed eukaryotic species. The results showed that no insect or nematode species contains POLD4, and a large number of protozoan and fungal species do not contain POLD4. The distribution of POLD4 in other eukaryotic species also varies widely.

**Table 2.** Summary of POLD4 distribution in various eukaryotes.

Taxon	Organisms with POLD4	Organisms without POLD4
Mammals	<i>Homo sapiens</i> and 18 other mammalian species	<i>Nomascus leucogenys</i> and <i>Ornithorhynchus anatinus</i>
Other vertebrates	<i>Anolis carolinensis</i> , <i>Taeniopygia guttata</i> , <i>Danio rerio</i> and 5 other bony fish species	<i>Meleagris gallopavo</i> , <i>Gallus gallus</i> , <i>Tetraodon nigroviridis</i> , <i>Xenopus laevis</i> , <i>Xenopus (Silurana) tropicalis</i>
Insects	Not found	<i>Drosophila melanogaster</i> and 16 other insect species
Nematodes	Not found	<i>Caenorhabditis elegans</i> and 6 other nematode species
Other invertebrates	<i>Strongylocentrotus purpuratus</i> and 6 other invertebrate species	<i>Hydra magnipapillata</i>
Protozoa	<i>Cryptosporidium</i> and 8 other species	<i>Babesia</i> and 33 other species
Flowering plants	<i>Oryza sativa</i> and 8 other flowering plant species	Not found
Algae and mosses	<i>Chlamydomonas reinhardtii</i> and 6 other algae or moss species	<i>Hemiselmis andersenii</i>
Fungi - Ascomycota	<i>Aspergillus</i> and 36 other species of Pezizomycotina	<i>Aciculosporium</i> and 10 other species of Pezizomycotina
Pezizomycotina		<i>Saccharomyces cerevisiae</i> and 17 other genera of Saccharomycotina
Fungi - Ascomycota	<i>Yarrowia lipolytica</i>	Not found
Saccharomycotina		
Fungi - Ascomycota	<i>Schizosaccharomyces pombe</i>	Not found
Schizosaccharomycetes		
Fungi - Basidiomycota	<i>Cryptococcus</i> and 9 other species of Basidiomycota	<i>Melampsora</i> and 5 other species of Basidiomycota
Fungi - Microsporidia	Not found	<i>Encephalitozoon</i> and 6 other species of
Microsporidia		
Fungi - Chytridiomycota	<i>Batrachochytrium dendrobatidis</i>	Not found

This table is a summary based on the distribution of POLD4 sequence in 218 eukaryotic species in our survey. Please refer to Table S1 for sequence accession number of POLD4 and species names of individual organisms.

First, most mammalian species contain POLD4. Among the 21 mammalian species surveyed, 19 possessed POLD4. Only *Nomascus leucogenys* (gibbon) and *Ornithorhynchus anatinus* (platypus) did not have POLD4 (Table 2 and [Table S1](#)). Because the other 5 primate species examined possessed POLD4, we considered the absence of POLD4 in the gibbon to be the result of incompleteness of their sequence databases. Since the platypus is the most primitive mammal, the absence of POLD4 likely also resulted from an incomplete sequence database seeing as POLD4 is present in other mammals. Therefore, we conclude that all mammals possess POLD4.

Second, among the 13 other vertebrates evaluated (i.e., non-mammalian vertebrates), 8 contained POLD4 and 5 did not (Table 2 and [Table S1](#)). The 8 vertebrates with POLD4 included 6 species of bony fishes, 1 species of lizard, and 1 species of bird, while the 5 vertebrates without POLD4 included 2 species of birds, 2 species of frogs and toads, and 1 species of bony fish. The absence of POLD4 in the bird and bony fish species was likely because of in-

completeness in their sequence databases. Whether specific bird species and frog species do not contain POLD4 requires verification upon the release of new sequence data.

Third, none of the 17 insect and 7 nematode species in our survey contained POLD4 (Table 2 and [Table S1](#)). Because we examined all available protein and nucleotide sequence databases and the surveyed organisms included model organisms such as *Drosophila melanogaster* and *Caenorhabditis elegans*, we concluded that no insect and nematode species contain POLD4.

Fourth, among the 8 other invertebrates (i.e., non-insect non-nematode invertebrates), 7 possessed POLD4. Only *Hydra magnipapillata* did not have POLD4 (Table 2 and [Table S1](#)). Therefore, most non-insect non-nematode invertebrates possess POLD4.

Fifth, among the 43 protozoan species examined in our survey, 9 contained POLD4 and 34 did not (Table 2 and [Table S1](#)), suggesting that most protozoan species (approximately 80%) do not have POLD4. The 9 protozoan species included 3 species of apicomplexans, 3 species of 'eukaryotes' (a name used in GenBank to classify a group of protozoa that belongs to various orders and/or families), 2 species of choanoflagellates, and 1 species of ciliates. The 34 protozoan species included 7 species of eukaryotes, 6 species of apicomplexans, 4 species of oomycetes, 2 species of flatworms, 2 species of cellular slime molds, 2 species of ciliates, 2 species of kinetoplastids, 2 species of diatoms, 1 species of forams, 1 species of pelagophytes, 1 species of cercozoans, 1 species of cryptomonads, 1 species of brown algae, 1 species of diplomonads, and 1 species of placozoans. Protozoan species of apicomplexans, eukaryotes, and ciliates were found in both groups with and without POLD4.

Sixth, POLD4 was found in all 9 species of flowering plants and 7 species of algae and mosses studied. Only *Hemiselmis andersenii* did not contain POLD4 (Table 2 and [Table S1](#)). Thus, all plant species possess POLD4.

Finally, the distribution of POLD4 in fungal species varied widely in different phyla or subphyla. In the phylum Ascomycota, 37 species of subphylum Pezizomycotina possessed POLD4, while 11 did not (i.e., most had POLD4), 1 species of subphylum Saccharomycotina had POLD4 and 18 did not (i.e., most did not have POLD4), and 1 species of subphylum Schizosaccharomycetes had POLD4 (no species of other genera were available for our survey). In the phylum Basidiomycota, 10 species had POLD4, while 6 did not. In the phylum Microsporidia, none of the 7 species examined had POLD4. In the phylum Chytridiomycota, 1 species had POLD4 (no species of other genera were available for our survey). Thus, a total of 50 species of fungi possess POLD4, while 42 fungus species do not.

In summary: 1) no insect and nematode species possessed POLD4; 2) mammals, non-mammalian vertebrates, non-insect non-nematode invertebrates, flowering plants, algae, and mosses contain POLD4; 3) approximately 80% of protozoan species do not have POLD4; and 4) the number of fungal species with POLD4 is slightly higher than those without POLD4 (50 vs 42).

### **Distribution of POLD1, POLD2, and POLD3 sequences**

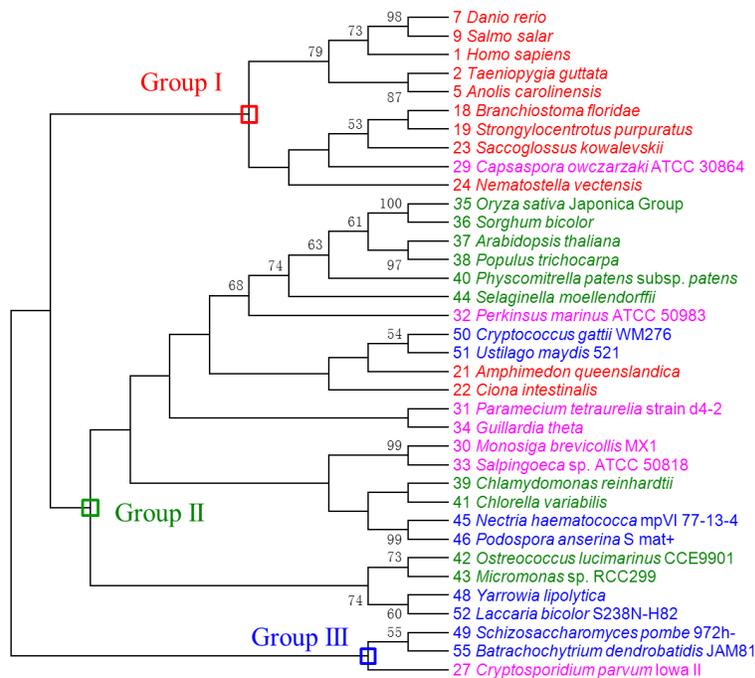
We determined the distribution of POLD4 in 218 eukaryotic species (Table 2 and [Table S1](#)). To facilitate phylogenetic analyses, we also searched for POLD1, 2, and 3 sequences in 55 species from different taxonomic groups (Table 1). Unsurprisingly, all 55 species were found to possess POLD1, 2, and 3 sequences.

### **Phylogenetic analyses of POLD1, 2, 3, and 4 sequences**

The survey results showed that POLD1, 2, and 3 are present in all eukaryotic species,

whereas POLD4 is only present in particular organisms. To understand this difference, we conducted phylogenetic analyses of subunit sequences of DNA polymerase  $\delta$ .

First, among the 55 species analyzed from different taxonomic groups, 36 species were found to possess POLD4. The multiple aligned conserved regions of the POLD4 sequences are shown in Figure 1. The phylogenetic tree constructed using these conserved amino acids (Figure 2) shows that the 36 species could be divided into 3 groups based on their phylogenetic relationships. Group I includes 9 animal and 1 protozoan species, showing a relatively simple POLD4 origin in the animal species (i.e., from protozoa). Group II consists of 10 plant, 2 animal, 5 protozoan, and 6 fungal species, indicating that plant POLD4 sequences share their most recent common ancestor with fungi or protozoa; POLD4 sequences of some lower animal species have a close relationship with both fungal and protozoan POLD4. Group III includes 2 fungal and 1 protozoan species, indicating a common ancestor of POLD4 for both plant and animal species. The phylogenetic relationship shown in Figure 2 suggests that most animal and plant POLD4 sequences originated in protozoa, and a few animal species (e.g., Nos. 21 and 22) and plant species (e.g., No. 39 and Nos. 41 to 43) may have inherited POLD4 directly from fungi.

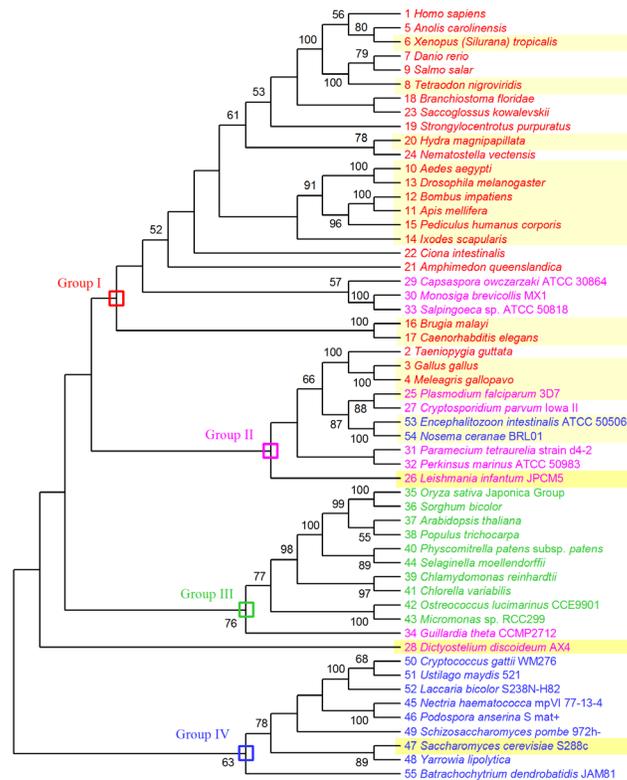


**Figure 2.** Neighbor-joining phylogenetic tree of POLD4 protein sequences from 36 eukaryotic species. The tree was constructed using conserved regions of multiple aligned POLD4 protein sequences as shown in Figure 1. For simplicity, only the bootstrap values  $\geq 50$  are shown and branch lengths are not proportional to distances between sequences. The color of species names and numbers correspond to the groups identified in the legend to Figure 1.

Our studies revealed the origin of plant and animal POLD4 sequences; however, the question as to why some animal species do not contain POLD4 remained unanswered. To answer this question, we conducted a second phylogenetic analysis using the POLD1, 2, and

3 sequences from 55 representative species listed in Table 1. In 2001, Herniou et al. reported that generation of phylogenies based on the combined sequences of all shared genes is the most effective approach for resolving relationships among species. Therefore, this approach was adopted for this study. Briefly, the conserved regions of POLD1, 2, and 3 sequences were manually spliced together and were then used for phylogenetic analysis ([Figure S1](#)).

The phylogenetic relationship between the POLD1, 2, and 3 sequences among the 55 species is shown in Figure 3, with species lacking POLD4 highlighted in yellow.



**Figure 3.** Neighbor-joining phylogenetic tree of POLD1 and 2 protein sequences from 55 eukaryotic species. The tree was constructed using manually spliced conserved regions of multiple aligned POLD1 and 2 protein sequences as shown in [Figure S1](#). For simplicity, only the bootstrap values  $\geq 50$  are shown and branch lengths are not proportional to distances between sequences. The color of species names and numbers correspond to the groups identified in the legend to Figure 1.

Figure 3 shows that the phylogenetic analysis divided POLD1, 2, and 3 in eukaryotic species into 4 distinct groups. Group I includes 21 animal and 3 protozoan species, among which 11 animal species do not have POLD4. In this group, the 6 insect species (Nos. 10 to 15) and the 2 nematode species (Nos. 16 and 17) lacking POLD4 form a separate monophyletic clade. Group II contains 3 animal, 5 protozoan, and 2 fungal species. Here, the 3 bird species (Nos. 2 to 4) have a closer evolutionary relationship with protozoa and fungi than with other animal species. In this group, 2 bird species (Nos. 3 and 4), 1 protozoan species (No. 25), and 2 fungal species (Nos. 53 and 54) do not have POLD4, suggesting that specific bird species inher-

ited polymerase  $\delta$  from ancestor species originally lacking POLD4. Group III includes 10 plant and 1 protozoan species. Here, all plant species possess POLD4, suggesting that their common ancestor also had POLD4. Group IV contains 9 fungal species, 1 of which lacks POLD4. In addition, 1 protozoan species (No. 28), which also lacks POLD4, was not included in any of the 4 groups. It is shown as an outgroup of other protozoan, animal, and plant species. The 2 fungal species *Encephalitozoon intestinalis* ATCC 50506 (No. 53) and *Nosema ceranae* BRL01 (No. 54) were previously classified as protozoa (Sprague and Vavra, 1977) but were recently reclassified as fungi (Baldauf et al., 2000). Our analysis of the phylogeny of POLD1, 2, and 3 sequences demonstrates that these organisms have a closer relationship with protozoa.

Our phylogenetic analyses led to the following conclusions: 1) fungi possess the most primitive form of DNA polymerase  $\delta$  in all eukaryotic species, 2) most animal and plant species inherited DNA polymerase  $\delta$  from protozoa, and 3) some animal and plant species may have inherited polymerase  $\delta$  directly from fungi. Furthermore, approximately 80% of protozoan species and nearly 50% of fungal species do not possess POLD4. Therefore, current insect and nematode species lacking POLD4 may have evolved from ancestor protozoan species originally lacking POLD4, and other protozoan (e.g., Nos. 25, 26, and 28) and animal (e.g., No. 20 and possibly Nos. 6 and 8) species lacking POLD4 may have a similar evolutionary history.

## CONCLUSIONS

Variation in the subunit composition of eukaryotic polymerase  $\delta$  suggested that POLD4 may be indispensable in some cases and can be used for evolutionary analyses. In this study, we searched protein and nucleotide sequence databases of 218 eukaryotic species and found that: 1) no insect or nematode species possessed POLD4, 2) approximately 80% of protozoan species did not have POLD4, 3) nearly 50% of fungal species did not have POLD4, and 4) mammals, non-mammalian vertebrates, non-insect non-nematode invertebrates, flowering plants, algae, and mosses contain POLD4. Phylogenetic analyses of POLD1, 2, 3, and 4 sequences revealed that most animal and plant species inherited DNA polymerase  $\delta$  from protozoa, while some other animal and plant species may have inherited polymerase  $\delta$  directly from fungi. Because a considerable number of protozoan and fungal species did not possess POLD4, current insect and nematode species lacking POLD4 may have evolved from ancestor protozoan species lacking POLD4, while other protozoan and animal species lacking POLD4 may have a similar evolutionary history. This is the first study to survey the distribution of POLD4 among various eukaryotic species, to conduct phylogenetic analyses on the 4 subunit sequences of polymerase  $\delta$ , and to examine why some organisms do not contain POLD4. Our data are useful for further studies on the origin and indispensability of POLD4 in various organisms.

## Conflicts of interest

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31370790, #31100118) and the Startup Scientific Research Fund from Jiangsu University for Advanced Professionals (#09JDG002).

## [Supplementary material](#)

## REFERENCES

- Baldauf SL, Roger AJ, Wenk-Siefert I and Doolittle WF (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290: 972-977.
- Bell SP and Dutta A (2002). DNA replication in eukaryotic cells. *Annu. Rev. Biochem.* 71: 333-374.
- Bermudez VP, MacNeill SA, Tappin I and Hurwitz J (2002). The influence of the Cdc27 subunit on the properties of the *Schizosaccharomyces pombe* DNA polymerase delta. *J. Biol. Chem.* 277: 36853-36862.
- Burgers PM and Gerik KJ (1998). Structure and processivity of two forms of *Saccharomyces cerevisiae* DNA polymerase delta. *J. Biol. Chem.* 273: 19756-19762.
- Byrnes JJ, Downey KM, Black VL and So AG (1976). A new mammalian DNA polymerase with 3' to 5' exonuclease activity: DNA polymerase delta. *Biochemistry* 15: 2817-2823.
- Garg P and Burgers PM (2005). DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol.* 40: 115-128.
- Gerik KJ, Li X, Pautz A and Burgers PM (1998). Characterization of the two small subunits of *Saccharomyces cerevisiae* DNA polymerase delta. *J. Biol. Chem.* 273: 19747-19755.
- Herniou EA, Luque T, Chen X, Vlak JM, et al. (2001). Use of whole genome sequence data to infer *baculovirus* phylogeny. *J. Virol.* 75: 8117-8126.
- Huang QM, Akashi T, Masuda Y, Kamiya K, et al. (2010a). Roles of POLD4, smallest subunit of DNA polymerase delta, in nuclear structures and genomic stability of human cells. *Biochem. Biophys. Res. Commun.* 391: 542-546.
- Huang QM, Tomida S, Masuda Y, Arima C, et al. (2010b). Regulation of DNA polymerase POLD4 influences genomic instability in lung cancer. *Cancer Res.* 70: 8407-8416.
- Hubscher U, Maga G and Spadari S (2002). Eukaryotic DNA polymerases. *Annu. Rev. Biochem.* 71: 133-163.
- Hughes P, Tratner I, Ducoux M, Piard K, et al. (1999). Isolation and identification of the third subunit of mammalian DNA polymerase delta by PCNA-affinity chromatography of mouse FM3A cell extracts. *Nucleic Acids Res.* 27: 2108-2114.
- Jones DT, Taylor WR and Thornton JM (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8: 275-282.
- Li H, Xie B, Zhou Y, Rahmeh A, et al. (2006). Functional roles of p12, the fourth subunit of human DNA polymerase delta. *J. Biol. Chem.* 281: 14748-14755.
- Liu L, Mo J, Rodriguez-Belmonte EM and Lee MY (2000). Identification of a fourth subunit of mammalian DNA polymerase delta. *J. Biol. Chem.* 275: 18739-18744.
- Meng X, Zhou Y, Zhang S, Lee EY, et al. (2009). DNA damage alters DNA polymerase delta to a form that exhibits increased discrimination against modified template bases and mismatched primers. *Nucleic Acids Res.* 37: 647-657.
- Meng X, Zhou Y, Lee EY, Lee MY, et al. (2010). The p12 subunit of human polymerase delta modulates the rate and fidelity of DNA synthesis. *Biochemistry* 49: 3545-3554.
- Mo J, Liu L, Leon A, Mazloum N, et al. (2000). Evidence that DNA polymerase delta isolated by immunoaffinity chromatography exhibits high-molecular weight characteristics and is associated with the KIAA0039 protein and RPA. *Biochemistry* 39: 7245-7254.
- Nicholas KB, Nicholas HB and Deerfield DW (1997). GeneDoc: analysis and visualization of genetic variation. *Embnet News* 4: 14.
- Reynolds N, Watt A, Fantes PA and MacNeill SA (1998). Cdm1, the smallest subunit of DNA polymerase d in the fission yeast *Schizosaccharomyces pombe*, is non-essential for growth and division. *Curr. Genet.* 34: 250-258.
- Sprague V and Vavra J (1977). Systematics of the Microsporidia. In: Comparative Pathobiology. Vol. 2 (Bulla LA and Cheng TC, eds.). Plenum Press, New York.
- Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
- Waga S and Stillman B (1998). The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* 67: 721-751.
- Zhang S, Zhou Y, Trusa S, Meng X, et al. (2007). A novel DNA damage response: rapid degradation of the p12 subunit of dna polymerase delta. *J. Biol. Chem.* 282: 15330-15340.
- Zhou Y, Chen H, Li X, Wang Y, et al. (2011). Production of recombinant human DNA polymerase delta in a *Bombyx mori* bioreactor. *PLoS One* 6: e22224.
- Zhou Y, Meng X, Zhang S, Lee EY, et al. (2012). Characterization of human DNA polymerase delta and its subassemblies reconstituted by expression in the MultiBac system. *PLoS One* 7: e39156.