NIH Public Access
Author Manuscript
DNA Repair (Amst). Author manuscript; available in PMC 2015 October 01.

Published in final edited form as:
DNA Repair (Amst). 2014 October ; 22: 77–88. doi:10.1016/j.dnarep.2014.07.003.

Phylogenetic analysis and evolutionary origins of DNA polymerase X-family members

Rachelle J. Bienstock, **William A. Beard**, and **Samuel H. Wilson**[*]
Laboratory of Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, 111 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States

Abstract

Mammalian DNA polymerase (pol) β is the founding member of a large group of DNA polymerases now termed the X-family. DNA polymerase β has been kinetically, structurally, and biologically well characterized and can serve as a phylogenetic reference. Accordingly, we have performed a phylogenetic analysis to understand the relationship between pol β and other members of the X-family of DNA polymerases. The bacterial X-family DNA polymerases, *Saccharomyces cerevisiae* pol IV, and four mammalian X-family polymerases appear to be directly related. These enzymes originated from an ancient common ancestor characterized in two *Bacillus* species. Understanding distinct functions for each of the X-family polymerases, evolving from a common bacterial ancestor is of significant interest in light of the specialized roles of these enzymes in DNA metabolism.

Keywords

DNA polymerase; Evolution; Function; Genomics; Phylogenetic; Structure; X-family

1. Introduction

Phylogenetic analysis is a computational method for quantifying evolutionary changes and relationships between protein sequences from different species over time. This type of analysis contributes to our understanding the functional development of a specific enzyme and the relationship between enzymes within cellular pathways, as well as the origin of enzymatic pathways within a cell. Additionally, evolutionary pathway analysis can provide insight into the origin of a particular disease pathway, as well as the relationship between genes in a model organism, as compared to humans.

DNA polymerases catalyze DNA synthesis during repair, replication, and recombination of DNA, as well as specialized DNA synthesis functions during viral replication and antibody gene maturation. Human cells have at least 16 distinct DNA polymerases, now characterized as members of different groups or “families” (designated A, B, X and Y) based on primary

© 2014 Published by Elsevier B.V.
[*]Corresponding author: Tel.: +1 911 541 4701; fax: +1 919 541 2843. wilson5@niehs.nih.gov (S.H. Wilson).
Conflict of interest
The authors declare that there are no conflicts of interest.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

protein sequences and putative structural motifs [1,2]. DNA polymerases are usually multi-domain proteins that include an accessory domain in addition to the polymerase domain [3]. The accessory domain (e.g., proofreading exonuclease) can complement the biological function of the polymerase. The nucleotidyl transferase super-family structural fold consists of a catalytic subdomain with a β-sheet and 2 α-helices [4]. The catalytic subdomain includes three carboxylate side chains that coordinate two divalent metal cations (usually Mg^{2+}) within the polymerase catalytic site.

The human DNA polymerase X-family, or family-X, is comprised of DNA polymerase (pol) β, pol λ, polμ and terminal deoxynucleotidyl transferase (Tdt). Only vertebrates possess members of all four of these family-X DNA polymerases, with plants, fungi and simpler organisms having only one or two family members; in some cases X-family members are not present at all (e.g., *Caenorhabditis elegans* and *Drosophila melanogaster*) [5,6]. Tdt has only been identified in vertebrates. Family-X DNA polymerases also include the African swine fever virus pol X, *Saccharomyces cerevisiae* pol IV and a large and emerging series of recently identified pol X-family members in bacterial systems [7,8].

Human pol β has been kinetically, structurally, and biologically characterized [9], and in the current work serves as a reference for comparison with the bacterial X-family polymerases. Bioinformatic and phylogenetic analyses were used to establish an evolutionary, structural and functional relationship between human pol β and bacterial DNA polymerase X-family members. These *in silico* studies may facilitate use of bacterial systems as models in understanding DNA transactions in more complex organisms with pol X-family members and/or provide insights into the role of the bacterial enzymes in their native environment.

Aravind and Koonin [10] characterized a broad group of nucleotidyl transfer enzymes that included DNA polymerase X-family members as well as members from related families: archaeal and bacterial CCA-adding enzymes/polyA polymerases, protein nucleotidyltransferases, antibiotic nucleotidyltransferases, and proteobacterial adenylyl cyclases. All of these enzymes transfer a nucleotide to an acceptor hydroxyl group, and their common active site suggested an evolutionary relationship.

Analyses of phylogenetic relationships of X-family members have been reported more recently. Uchiyama et al. [6] suggested that all X-family members evolved from a single pol λ-like gene involved in non-homologous end-joining (NHEJ) and that the other X-family member polymerases arose due to gene duplication of this pol λ-like gene. Similarly, Kodera et al. [11] concluded that since the most basal phylum (e.g., *Echinodermata* in metazoans) contained three X-family DNA polymerase genes (i.e., βλ, and Tdt/μ-like), it is likely that the eukaryotic pol X-family diverged from a single pol λ-like coelenterate phylum gene. In contrast, the computational analyses presented here suggest that all X-family members evolved from a polymerase nucleotidyl transfer catalytic core protein present in ancient bacterial organisms, and gene duplication and alterations occurred over time providing increasing complexity and organelle differentiation within species. This work was aimed at achieving an understanding of the functional and chronological development of different X-family members, especially in relation to pol β.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

DNA polymerase X-family members are present and conserved throughout many of the oldest and most varied forms of life. We applied several established methods for sequence alignment followed by phylogenetic analysis to assess the hypothesis that the various X-family polymerases evolved from a DNA polymerase X present in ancient bacterial species. The analysis was more extensive than that in previously published studies. For example, one such study [12] included only 27 X-family polymerase sequences. With recent advances in genomic DNA sequencing, the present study represents a sampling of more than 100 diverse species' sequences. Additionally, as crystal structures for many of the polymerase X-family members have been solved, including bacterial representatives such as from *Deinococcus radiodurans* [13], *Thermus thermophilus HB8* [14], and the African swine fever virus (ASFV) pol X [15,16], many structure-function relationships important in phylogenetic considerations of DNA polymerase X-family members could be evaluated.

2. Materials and methods

Several established algorithms were utilized for sequence alignment and analysis of the assembled DNA polymerase X-family phylogenetic tree. For creating the phylogenetic tree, we used the following: Phylome DB v. 3.0 [17], PhyML v 3.0 [18,19], ETE [20], iTOL [21,22], Phylemon 2.0 [23], Archaeopteryx [24], and Tree-Graph2 [25]. MUSCLE v. 3.7 was used as the sequence alignment algorithm [26]. JalView was used for creating visual images of sequence alignments produced by MUSCLE [27].

Phylogenetic analysis was initiated from an alignment of 111 identified X-family DNA polymerase sequences. The resulting phylogenetic tree was developed by beginning with the defined "Phylome" deposited in the PhylomeDB v3.0 [17] for human pol β. PhylomeDB (http://orthology.phylomedb.org), is a database of complete collections of gene phylogenies (phylomes) including a number of model species. Phylome uses orthology prediction on a genomic scale to determine the evolutionary relationship between genes from multiple independent phylogenetic trees. The Phylome DB contains 717 species and provides redundant orthology and paralogy predictions. To the original 63 sequences present in this defined phylogenetic pol β tree in the Phylome DB v. 3.0, were added an additional 48 sequences; 23 identified bacterial DNA X-family polymerases, 5 trypanosomatid sequences, and 20 eukaryote (animal) sequences belonging to the X-family. These sequences were then realigned with the original group defining the "Phylome" using the MUSCLE v. 3.7 sequence alignment algorithm [26] and a new phylogeny tree was generated using PhyML v. 3.0 [18,19]. All the identified pol β sequences analyzed in Asagoshi et al. [5], including bacterial DNA polymerase X sequences, and all the bacterial sequences analyzed by Banos et al. [28] were included in the alignment. Many of these sequences are annotated by GenBank as "hypothetical proteins," "phosphotransferase domain containing proteins," and "DNA polymerase X or X-family DNA polymerase" or "predicted to be or similar to DNA polymerase β."

Eggnog 3.0 [29] was also searched to identify 376 protein members of the COG1796 DNA polymerase IV X-family group or "cog" when searched with the human pol β sequence in 261 species. However, most of these orthologs are based on sequence similarity and polymerase enzymatic function has not been confirmed within these species. KOG2534

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

includes 198 enzymes in the DNA polymerase X-family category from 86 different species. The DNA polymerase X-family members used in the sequence alignments and subsequent phylogenetic analysis are tabulated in Table 1.

3. Results

3.1. Phylogenetic relationships

The comprehensive phylogenetic tree developed from the DNA polymerase X-family sequences included in this study is illustrated in Fig. 1. The detailed sequence alignment used to develop this tree is prohibitively large (Supplementary FASTA file). Nevertheless, the quality and consensus of a portion of the alignment is illustrated in Fig. 2. This figure illustrates the sequence conservation exhibited over the human pol β sequence (residues 1–335). Key residues of pol β involved in substrate binding, catalysis, conformational changes, and its deoxyribose phosphate (dRP) lyase activity are tabulated in Table 2.

From the standpoint of a broad overview, the bacterial X-family polymerases are the evolutionary ancestors of all family-X polymerases (including the eukaryotic cellular enzymes designated pol β, pol λ pol μ TdT, and pol IV) (Figs. 1 and 3). The ancient bacterial X-family polymerases evolved into an ancestral pol IV from *Kluyveromyces lactis* and later *S. cerevisiae* pol IV. After this point, two distinct branches arose: (1) the Tdt and pol μ branch (two gram-negative rod shaped bacteria also have polymerases clustered with this group, as well as several fungi), and the (2) the pol β and λ branch. The pol β branch includes a sub-branch of slime mold and trypanosomatid pol β; and the pol λ sub-branch includes the plant *Arabidopsis thaliana* pol X and some fungal polymerases.

As illustrated in Figs. 1 and 3, the evolutionary origin of the DNA polymerase X-family members begins with the most ancient pol X from the *Bacillus* bacterial species, *B. subtilis and B. amyloliquefaciens* and then *B. pumilus.* These are gram-positive bacteria with protective endospores permitting the organisms to tolerate extreme environmental conditions. The oldest ancestral *Bacillus* polymerases X are followed by the *Listeria* and *Staphylococcus* pol X that lead to the *Desulfotomaculum reducens* polymerase and then to the *Thermoplasma* and *Thiobacillus*, followed by the thermophilic species, *T. thermophilius* and *T. aquaticus* pol X; these are followed by the *D. radiodurans*, *Mycobacterium tuberculosis*, and *Nematostella vectensis* pol X species. Most bacterial pol X-family members are ancestors and do not cluster specifically with either pol β, pol λ, pol μ, or pol IV. There are exceptions, however: *Vibrio anguillarum*, and *Aeromonas hydrophila. V. anguillarum* pol X is pol μ/Tdt-like and sequesters with these sequences. These species are members of the bacterial phylum Proteobacteria Class: Gamma Proteobacteria Order: Vibrionales Family: Vibrionaceae. This gram negative curved rod-shaped bacterium with one polar flagellum is an important pathogen of cultured salmonid fish and causes the disease known as vibriosis or red pest of eels. *A. hydrophila* pol X is also a pol μ/Tdt like-DNA polymerase. It differs in class from the Vibrionales and belongs to the bacteria Proteobacteria: Gammaproteobacteria, Aeromonadales, Aeromonadaceae. *Aeromonas* bacterium is found in freshwater environments and in brackish water; it is a gram-negative rod that has polar flagella and is a facultative anaerobe.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

The pol IV containing fungi species branch evolved from *N. vectensis* that then broke off into the fungal species group, then pol μ and Tdt followed by a split from the pol λ and pol β groups. In the pol β branch, the trypanosomids form a distinct and earlier pol β sub-group, as do the slime mold, *Salpingoeca rosetta* sp., *Tribolium castaneum*, and *Strongylocentrotus purpuratus* species. The pol λ branch also has an older off-branch comprised of the *Arabidopsis* and fungi *Neurospora crassa, Gibberella zeae, Yarrowia lipolytica*, and *Cryptococcus neoformans*.

3.2. Relationships

3.2.1. X-family polymerases in fungi—These enzymes are divided into two distinctive classes (Fig. 1). One group of fungal species has a pol IV-like DNA polymerase, whereas others have two family-X polymerases, one that clusters with and is similar to mammalian pol λ and another that clusters with mammalian pol μ and Tdt. Plants are represented in this tree by *Arabidopsis* that only has pol λ.

3.2.2. Functional relationships to S. cerevisiae pol IV—DNA polymerase IV has 5′-dRP lyase activity, an essential activity for single-nucleotide base excision repair (BER), low processivity for DNA synthesis on open template DNA (i.e., not gapped) and can fill short DNA gaps. DNA polymerase IV does not have a proofreading activity and is highly inaccurate. It has the greatest homology with mammalian pol λ, has a BRCT domain, and both pol λ and pol IV contain a large number of positively charged residues in their dRP lyase domain [30]. *Schizosaccharomyces pombe* pol IV has dRP lyase activity, and properties that are similar to pol β, pol λ and pol μ. DNA polymerases β and λ share 32% sequence identify and pol λ can substitute for pol β during *in vitro* and *in vivo* BER [31]. DNA polymerase μ and Tdt share 41% sequence identity. *S. pombe* pol IV shares the highest sequence similarity to pol μ [31].

3.2.3. X-family polymerases in trypanosomatids—Trypanosomatids have more than one pol β and appear to have evolved a separate and distinct packing and presentation scheme for their pol β-like enzymes compared with other species. These enzymes within some of these parasites are unusual in that they are not nuclear, but instead are mitochondrial. Trypanosomatids have unusual mitochondrial DNA packed in a special structure called kinetoplast DNA (kDNA). The pol β gene has been cloned and characterized from several parasites, including *Trypanosoma cruzi, Trypanosoma brucei, Crithidia fasciculata and Leishmania*. The *Leishmania major* pol β, while localized to the nucleus is also involved in kinetoplast mitochondrial DNA replication, BER and translesion DNA synthesis [32]. In the case of *T. brucei*, two pol β genes have been identified within the mitochondria [33]. In *C. fasciculata* and the *Tryponosoma*, pol β with dRP lyase activity has been demonstrated in mitochondria [11,34,35]. *Leishmania infantum* pol β uses manganese as its divalent cation, prefers gapped DNA substrates with a 5′-phosphate in the gap, and has dRP lyase activity, all features characteristic of mammalian pol β. *Leishmania* is one of the most primitive eukaryotes, and *C. fasciculata* and *Trypanosoma* are related [36]. DNA polymerase β in *C. fasciculata* mitochondria is error prone *in vitro* and can only be found when replicating kDNA and is, therefore not thought to be involved in BER. Critical residues for the 5′-dRP lyase activity are conserved between *Crithidia* and mammalian pol β

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

(Table 3): Lys35, Lys60 and Lys68, important for single-stranded DNA binding; Lys35, important for 5′-phosphate group recognition; and Lys72 and Tyr39, dRP lyase catalytic residues [37,38]. The *C. fasciculata* pol β-like Schiff base intermediate can be trapped experimentally with $NaBH_4$ reduction, although the enzyme's dRP lyase activity is lower than the mammalian enzyme.

3.2.4. DNA polymerase β-like enzymes in lower organisms—DNA polymerase β genes from many lower organisms have been cloned and sequenced and the products of these genes are believed to function within BER. These enzymes are highly conserved, and in many species, the pol β proteins are similar in size retaining significant sequence similarity and protein domain structure. Functionally significant amino acid residues involved in catalysis are conserved within species and preserved throughout evolution (Table 2 and Supplementary Table 1). For example, *Oncorhynchus masou* (cherry salmon), *Xiphophorus maculates* and *Danio rerio* (zebrafish) pol β have 337 residues, only a few residues more than human, rodent (335 residues) and frog (334 residues). Zebrafish share 79, 80, and 79% sequence similarity with rat, human, and frog pol β, respectively. Jellyfish (*Aurelia sp. 1* As) pol β has also been characterized (335 residues) and found to share 55% sequence identity with the vertebrate pol β and have similar behavior in DNA repair assays as the mammalian enzymes [11]. As an example, the jellyfish enzyme is sensitive to inhibition by dideoxythymidine triphosphate, like mammalian pol β, and demonstrates similar response to several other mammalian pol β inhibitors [11]. However, the X-family polymerases do not share sequence similarity with, and are not related evolutionarily to, *Escherichia coli* pol I (A-family), pol II (B-family), or pol III (C-family) [39]. Surprisingly, there are no pol X-family members in the genomes of *Drosophila* or *C. elegans* [5,6]. Yet as illustrated in Fig. 1, pol β has been found in other ecdysozoans, such as *T. castaneum* (red flour beetle) and sea anemone (*N. vectensis*) [11]. To date, no homologs of pol β or Tdt have been identified in fungi, plants and lower fungi, such as slime mold. These organisms have a pol λ-like enzyme only.

3.3. Domains and enzymatic activities

Although DNA polymerase X-family members only share partial sequence homology, they share a conserved domain organization. Fig. 4 illustrates the domain structure of different X-family members. The polymerase catalytic domain of all X-family members have the common subdomain architecture with DNA binding, catalytic, and dNTP binding subdomains (D-, C-, and N-subdomains, respectively) [40]. Some DNA polymerase X-family members are small enzymes. DNA polymerase β is the smallest mammalian DNA polymerase, while the ASFV pol X is only 20 kDa (174 residues) [41]. Mammalian pol β and pol λ have been shown to have 5′-dRP lyase and polymerase activities required in BER. DNA polymerase λ, pol μ and Tdt also have an amino-terminal BRCT domain, a protein-protein interaction domain involved in NHEJ and possibly other forms of double strand break repair [8]. Importantly, pol μ and Tdt do not have 5′-dRP lyase activity [42]. The X-family members with the lyase activity possess a conserved catalytic lysine nucleophile (pol β, Lys72; pol λ, Lys312) that forms a Schiff base intermediate in the lyase β-elimination reaction mechanism. These enzymes also have a conserved tyrosine residue (pol β, Tyr39; pol λ, Tyr279) suggested to be involved in the lyase catalytic mechanism [39]. This tyrosine

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

residue corresponds to a phenylalanine in family-X members lacking the dRP lyase activity (i.e., pol μ and Tdt). The residues necessary for efficient lyase activity are conserved in the polymerase X-family members of many lower species, e.g., in the 8 kDa domain of the parasitic protist *C. fasciculata* pol X (Tyr38 and Lys71) and yeast pol IV (Tyr214 and Lys248).

B. subtilis pol X has been shown to have DNA synthesis activity, manganese dependent 3′ to 5′ exonuclease, and apurinic/apyrimidinic (AP) endonuclease activities [28]. *B. subtilis* pol X shares 24% sequence identity with human pol β and requires a template and divalent metal for polymerase activity. DNA polymerase μ and Tdt also have been shown to have template-independent DNA synthesis activity. These template-independent enzymes have a conserved histidine residue (His329, pol μ; His342, Tdt) near the active site that may stabilize the primer terminus in the absence of a complementary templating base (i.e., single-stranded primer terminus). Previously reported sequence alignments identified a conserved Helix-hairpin-Helix motif within bacterial and archeal pol X family members; these Helix-hairpin-Helix motifs were predicted to indicate that these enzymes possessed dRP lyase activity based on the presence of homologous conserved lysine and tyrosine residues observed with enzymes that have this activity, although this activity has not been demonstrated experimentally [28].

Bacillus pol X has 570 residues; residues 1–317 define the polymerase domain, and residues 335–570 define a histidinol phosphatase (PHP) domain unique to bacterial polymerase X-family members (Fig. 4). The enzyme has been shown to possess intrinsic AP endonuclease activity [43]. Other than the polymerase domain, the functions of the bacterial pol X domains are less well established than those of the mammalian polymerases and are still under active investigation. To date, the bacterial pol X enzymes appear to lack a BRCT domain [44]. Several catalytically important residues for DNA synthesis activity have been identified in *T. thermophilus* pol X [45]. In addition, a PHP domain with intrinsic 3′–5′ exonuclease activity has been identified in this organism. Recently, it has been reported that the *T. thermophilus* PHP domain has 3′ phosphatase and AP endonuclease activities [45].

3.4. Structural comparisons

The ASFV pol X is the smallest (174 residues) nucleotidyl transfer polymerase identified to date (Figs. 4 and 5). ASFV causes hemorrhagic fever in pigs, and pol X repairs viral DNA damaged as the host fights viral infection. The ASFV pol X has a unique structure involving the polymerase C- and N-subdomains, but without the D-subdomain [15,16]. Despite the lack of a DNA binding subdomain, the enzyme binds DNA with high affinity. The C-subdomain of ASFV pol X is considerably more electropositive than in pol β, even though ASFV pol X shares 55% sequence homology with the carboxyl-terminal half of pol β [41]. The ASFV active site metal coordinating ligands (Asp49, Asp51, and Asp100) are structurally super-imposable with the catalytic aspartate triad of human pol β (Asp190, Asp192, and Asp256). Interestingly, ASFV pol X has a disulfide bond between Cys81 and Cys86 in one solution structure [15], but is reduced in a solution structure determined in a different laboratory [16]. The resultant structures indicate that the N-subdomains are in alternate positions reminiscent of those observed when mammalian pol β binds a nucleoside

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

triphosphate [41]. The significance for this redox activity is not known and has not been observed with the mammalian polymerase. As noted above, ASFV pol X binds DNA intermediates of base excision repair with high affinity, and the enzyme does not have an amino-terminal lyase domain that would facilitate single-nucleotide BER. The enzyme does not interact with the BER proteins AP endonuclease or XRCC1 and does not have a proofreading exonuclease. This polymerase is an example of a minimal functional version of the evolutionary conserved pol β-type core polymerase and is anomalous to other X-family member structures [46].

Crystal structures of X-family bacterial polymerases from two organisms have been solved (Fig. 5): *D. radiodurans*, a gram-positive bacterium, and *T. thermophilus HB8*, a gram-negative eubacterium. The *Thermus* pol X is atypical in that it binds Mg^{2+}-dNTP prior to binding DNA [45]. *Thermus* pol X has DNA/RNA polymerase activity and 3′–5′ proofreading exonuclease activity, and structures have been solved of the binary complex with dGTP, template/primer DNA, and with gapped DNA and ddGTP (ternary substrate complex). A pol β-like active site, with 2 metals and aspartate catalytic triad (Asp198, Asp200, and Asp243) is conserved in this bacterial polymerase. Lys263 is essential for Mg^{2+} binding to dNTP and this lysine residue is conserved in bacterial and archeal pol X enzymes and is also found in pol μ and Tdt. In pol β, however, the corresponding residue is an aspartate. *Thermus* pol X has Mg^{2+} and Mn^{2+}-dependent DNA synthesis activities and Mn^{2+}-dependent exonuclease activity [44]. Site-directed mutational studies determined that His344, His374, His468 and Asp529 in the PHP domain were essential residues for the proofreading exonuclease activity, but not polymerase activity [44]. These residues are highly conserved among all bacterial pol X structures [47].

D. radiodurans is highly resistant to ionizing radiation damage. Deletion of the *D. radiodurans* pol X decreases the rate of DNA double-strand break repair. The polymerase has strong proofreading exonuclease activity stimulated by Mn^{2+}. Its structure displays an extended fold in contrast to the compact structure displayed for the polymerase domain of the mammalian enzymes [13]. Similar to *Thermus* pol X, it has a carboxyl-terminal PHP domain with a tri-nuclear zinc-binding site. However, the isolated PHP domain does not appear to be functional [13]. The carboxyl-terminal PHP domain isolated from *T. thermophilus* was shown to have proofreading activity, as well as 3′-phosphatase and AP endonuclease activities that can be utilized during base excision repair [45]. The 3′-phosphatase activity has not been reported in other pol X enzymes. Only prokaryotes have been shown to contain a PHP domain within their X-family DNA polymerase. In higher organisms, a separate and distinct enzyme performs this catalytic function. The PHP domain in *Bacillus* has AP endonuclease and proofreading exonuclease activities [47].

A significant difference between the *D. radiodurans* enzyme and higher organism pol X-family members is the replacement of the three conserved active site aspartate residues comprising the catalytic triad that binds divalent metals with two glutatmate residues (Fig. 6A). This is the only X-family polymerase, among all of the sequences examined in this study, that has an ‘AAE’ motif for residues corresponding to the ‘DXD’ motif (residues 190–192) in human pol β and conserved in almost all pol X-family members (Supplementary Table 1). The Asp256 residue in human pol β is also a glutatmate in the *D.*

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

radiodurans enzyme. In addition to binding the catalytic metal, this residue also participates in deprotonation of the primer terminus [48]. The PHP domain is similar to the YcdX *E. coli* zinc-binding domain, and zinc is found in the *D. radiodurans* pol X crystal structure. A 5′-dRP lyase and BER activities have been demonstrated in the *Deinococcus* pol X [49]. This pol X does not have the conserved lysine found in both pol β and λ (Lys72 and Lys312, respectively) that acts as the Schiff base nucleophile for lyase activity. This lysine is Glu72 in *Deinococcus* pol X. Lys33 of *D. radiodurans* that corresponds to Lys35 of pol β and Arg275 of pol λ is the only conserved basic residue in this family-X polymerase. In *D. radiodurans* pol X, Lys64 and Lys67 may be involved in DNA binding and Tyr37 may be involved in 5′-phosphate recognition necessary for DNA gap binding. Tyr37 correspond to Tyr39 in pol β and Tyr279 in pol λ that bind a 5′-phosphate in gapped DNA. The carboxyl-terminal PHP domain, present in many bacterial polymerase X-family members, is not present in mammalian pol X members but has been reported in the structure of *E. coli* pol IIIa subunit (family-C replicative DNA polymerase) [50]. The *D. radiodurans* pol X PHP domain superimposes with the Pol IIIa PHP domain, with a RMSD of 2.8 Å (136 Cα), and these PHP domains share only 15% sequence identity [13].

As is illustrated in Fig. 5 where the structures of various X-family polymerases are compared, the domain orientation in *D. radiodurans* pol X is different from that of pol β. As noted above, the polymerase nucleotidyl transfer reaction involving two metals bound to three aspartate resides must be different, as these have been replaced with two glutamate residues in *D. radiodurans* pol X (Fig. 6A). Accordingly, different metal binding sites must exist in this situation.

3.5. Evolutionary conservation of functionally significant residues

Since mammalian pol β has been extensively characterized functionally and structurally [9], it is useful to compare the conservation of functionally significant residues of pol β with those of other family members (Table 2 and Supplementary Table 1).

3.5.1. Active site metal-coordinating aspartates—As noted above, *D. radiodurans* (PDB ID 2W9M) is unique in that it does not include the active site "DXD" motif (pol β, Asp190 and Asp192). There are only two other X-family polymerases that we have examined that do not have this signature motif conserved, *M. tuberculosis* and *N. vectensis* (sea anemone). Fig. 6A illustrates a structural alignment (superposition) of the polymerase domain from human pol β (PDB ID 2FMS) and *D. radiodurans* pol X (PDB ID 2W9M). The root-mean squared deviations of the superposition of the C- and N-subdomains domains is 1.2 Å (74 Cα). The DXD motif is otherwise conserved in all other bacterial pol X and mammalian pol X sequences examined in this study. Asp256 is highly conserved with the exception of four pol X bacterial species; *D. radiodurans, Sphingobacterium spiritivorum, M. tuberculosis, and A. hydrophila*. Asparagine replaces this aspartate in one fungal pol λ-like enzyme in *N. crassa.*

3.5.2. Deoxynucleoside triphosphate binding residues—Arg183 of pol β interacts directly with the triphosphate tail of the incoming nucleotide (Fig. 6B). It is the most conserved residue throughout all X-family members (98%, Supplementary Table 1). Ser180

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

of pol β also interacts with the triphosphate tail and is conserved in bacterial pol X sequences but is glycine in fungi family-X polymerases, Tdt and pol μ. The backbone of pol β Gly189 interacts with the γ-phosphate of the incoming nucleotide and is a histidine in pol μ and Tdt sequences and a lysine in some bacterial pol X species.

The pol β dNTP binding pocket is formed by Tyr271, Phe272, Asp276, and Asn279 (Fig. 6B). The pol β YFTGSD motif (residues 271–276) is GWTGSK/R in pol μ and Tdt sequences and (Y/H)FTGSK in the bacterial pol X sequences. DNA polymerase β Asn279 is conserved in pol λ (Asn514), but not in pol μ (Gln432) or Tdt (Glu457), but is conserved in the majority of bacterial X-family polymerases. Tyr271 and Phe272 significantly alter their positions and interactions when a nucleotide binds to the binary DNA complex and have been suggested to play important roles in subdomain motions during catalytic cycling [9,51]. Asn279 forms a minor groove hydrogen bond with the base and Asp276 contributes van der Waals interactions with the base of the incoming nucleotide. Alanine substitution for Asn279 weakens nucleotide binding [52,53], but replacing the charged side chain of Asp276 with a hydrophobic side chain increases binding affinity [54]. Tyr271 has also been implicated in substrate specificity during ribo- and deoxyribosugar discrimination [55]. The YFTGSD sequence is highly conserved in all pol β sequences. It is TFTGSK in the trypanosome pol β sequences, YFTGSA in the majority of pol λ sequences, and GWTGSR/K/Q in many pol μ and Tdt sequences. This motif is not conserved in many of fungal species (e.g., *S. pombe, Ashbya gossypii, Candida glabrata, N. crassa, G. zeae and K. lactis*) and is also not conserved in many of the bacterial pol X sequences. It is HFTGSK in the majority of pol X bacterial sequences, however it diverges significantly in *S. spiritivorum, D. radiodurans, M. tuberculosis, N. vectensis, A. hydrophila* subsp. *hydrophila* ATCC 7966 and *V. anguillarum* 775.

3.5.3. DNA (template) binding residues—Arg283 of pol β plays a critical role in proper positioning of the templating base (Fig. 6C). Alanine substitution for this residue strongly reduces base substitution [52,56] and frameshift fidelity [57]. This is believed to be due to its interactions with the template strand as well as its role in structural transitions during substrate binding. Mutation of this residue destabilizes the closed catalytically active conformation of pol β [52,58]. The loss in fidelity is specifically due to the inability to insert the correct nucleotide [59]. Lys280 forms van der Waals contact with the base of the templating nucleotide. Interestingly, amino acid substitution at this position is strongly dependent on the identity of the templating base as well as the specific side chain at this position. These observations indicate that the interactions provided by substrate binding residues are specific for each base pair [40]. Lys280 aligns with arginine or lysine in most pol β, pol λ, Tdt, and slime mold sequences. It is serine in *Monodelphis domestica* pol μ (Supplementary Table 1) and valine in all the trypanosome pol β species. It is not conserved in the bacterial pol X, but is generally isoleucine in the prokaryotic enzymes (Supplementary Table 1). DNA polymerase β Arg283 is highly conserved in bacterial pol X enzymes, but is not in *S. spiritivorum, D. radiodurans V. anguillarum, or M. tuberculosis* (93% overall conservation, Supplementary Table 1).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

3.5.4. Lyase domain—The dRP lyase nucleophile Lys72 of pol β that forms a Schiff-base intermediate is highly conserved in all pol β and pol λ species, but is not conserved in Tdt and pol μ (Supplementary Table 1). This lysine appears to be conserved in some bacterial pol X sequences but is valine in several others. Glu26 and Ser30 are not conserved in the bacterial *Staphylococcus* and *D. radiodurans* X-family polymerases and Ser30 is not widely conserved even in some mammalian pol β sequences (e.g., *Canis lupus familiaris, Bos taurus*).

4. Conclusions

Phylogenetic analysis has shown that mammalian X-family DNA polymerases and yeast pol IV share a common origin with bacterial X-family polymerases. These polymerases share a common bacterial ancestor characterized in the *Bacillus* species designated the evolutionary grandfather of all family-X polymerases. The initial *Bacillus* X-family polymerases evolutionarily differentiated into branches; one of which became pol μ and Tdt, and a second that includes pol β and pol λ. A third branch includes a plant pol X and fungi pol IV sub-branch. The fungi enzymes are divided into two separate classes one which has a pol IV-like polymerase and another that has two X-family polymerases. Trypanosomatids are unusual and have more than one pol β and share fewer of the highly conserved residues of the X-family. DNA polymerase β is not nuclear and is mitochondrial in some of these trypanosomatid species. Over the millennium of evolution from *Bacillus* to the present, pol β has retained size, sequence, domain and residue conservation. Bacterial pol X species do have a unique PHP domain not present in other X-family species. The polymerase active site residues that bind the incoming nucleotide are highly conserved with the exception of *D. radiodurans, M. tuberculosis* and *N. vectensis*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by Research Project Numbers Z01-ES050158 and Z01-ES050159 in the Intramural Research Program of the National Institutes of Health, National Institute of Environmental Health Sciences and was in association with the National Institutes of Health Grant 1U19CA105010.

Abbreviations

AP	apurinic/apyrimidinic
ASFV	African swine fever virus
BER	base excision repair
dNTP	deoxynucleoside triphosphate
dRP	deoxyribose phosphate
kDNA	kinetoplast DNA
NHEJ	non-homologous end-joining

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

PHP	polymerase and histidinol phosphatase
Tdt	terminal deoxynucleotidyl transferase

References

1. Braithwaite DK, Ito J. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucl Acid Res. 1993; 21:787–802.
2. Bebenek K, Kunkel TA. Functions of DNA polymerases. Adv Protein Chem. 2004; 69:137–165. [PubMed: 15588842]
3. Wu S, Beard WA, Pedersen LG, Wilson SH. Structural comparison of DNA polymerase architecture suggests a nucleotide gateway to the polymerase active site. Chem Rev. 2014; 114:2759–2774. [PubMed: 24359247]
4. Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J. Crystal structure of rat DNA polymerase β: evidence for a common polymerase mechanism. Science. 1994; 264:1930–1935. [PubMed: 7516581]
5. Asagoshi K, Lehmann W, Braithwaite EK, Santana-Santos L, Prasad R, Freedman JH, Van Houten B, Wilson SH. Single-nucleotide base excision repair DNA polymerase activity in *C. elegans* in the absence of DNA polymerase β. Nucl Acids Res. 2012; 40:670–681. [PubMed: 21917855]
6. Uchiyama Y, Takeuchi R, Kodera H, Sakaguchi K. Distribution and roles of X-family DNA polymerases in eukaryotes. Biochimie. 2009; 91:165–170. [PubMed: 18706967]
7. Lecointe F, Shevelev IV, Bailone A, Sommer S, Hübscher U. Involvement of an X family DNA polymerase in double-stranded break repair in the radioresistant organism *Deinococcus radiodurans*. Mol Microbiol. 2004; 53:1721–1730. [PubMed: 15341650]
8. Ramadan K, Shevelev I, Hübscher U. The DNA-polymerase-X family: controllers of DNA quality? Nat Rev Mol Cell Biol. 2004; 5:1038–10343. [PubMed: 15573140]
9. Beard WA, Wilson SH. Structure and mechanism of DNA polymerase β. Chem Rev. 2006; 106:361–382. [PubMed: 16464010]
10. Aravind L, Koonin EV. DNA polymerase β-like nucleotidyltransferase superfamily: identification of three new families, classification and evolutionary history. Nucl Acids Res. 1999; 27:1609–1618. [PubMed: 10075991]
11. Kodera H, Takeuchi R, Uchiyama Y, Takakusagi Y, Iwabata K, Miwa H, Hanzawa N, Sugawara F, Sakaguchi K. Characterization of marine X-family DNA polymerases and comparative analysis of base excision repair proteins. Biochem Biophys Res Comm. 2011; 415:193–199. [PubMed: 22033415]
12. Filée J, Forterre P, Sen-Lin T, Laurent J. Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. J Mol Evol. 2002; 54:763–773. [PubMed: 12029358]
13. Leulliot N, Cladière L, Lecointe F, Durand D, Hübscher U, van Tilbeurgh H. The family X DNA polymerase from *Deinococcus radiodurans* adopts a non-standard extended conformation. J Biol Chem. 2009; 284:11992–11999. [PubMed: 19251692]
14. Lin Y, Chen S, Si S, Xie Y. Expression, purification, crystallization and preliminary X-ray crystallographic analysis of the hyperthermophilic nucleotidyltransferase TTHA1015 from *Thermus thermophilus* HB8. Acta Crystallogr F. 2011; 67:782–784.
15. Showalter AK, Byeon IJL, Su MI, Tsai MD. Solution structure of a viral DNA polymerase X and evidence for a mutagenic function. Nat Struct Biol. 2001; 8:942–946. [PubMed: 11685239]
16. Maciejewski MW, Shin R, Pan B, Marintchev A, Denninger A, Mullen MA, Chen K, Gryk MR, Mullen GP. Solution structure of a viral DNA repair polymerase. Nat Struct Biol. 2001; 8:936–941. [PubMed: 11685238]
17. Huerta-Cepas J, Capella-Gutierrez S, Pryszcz LP, Denisov I, Kormes D, Marcet-Houben M, Gabaldón T. PhylomeDB v3.0: an expanding repository of genome-wide collections of trees, alignments and phylogeny-based orthology and paralogy predictions. Nucl Acids Res. 2011; 39:D556–D560. [PubMed: 21075798]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

18. Guindon S, Dufayard J-Fo, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59:307–321. [PubMed: 20525638]
19. Guindon S, Gascuel O. A Simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52:696–704. [PubMed: 14530136]
20. Huerta-Cepas J, Dopazo J, Gabaldon T. ETE: a python environment for tree exploration. BMC Bioinform. 2010; 11:24.
21. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007; 23:127–128. [PubMed: 17050570]
22. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. Nucl Acids Res. 2011; 39:W475–W478. [PubMed: 21470960]
23. Sánchez R, Serra F, Tárraga J, Medina I, Carbonell J, Pulido L, de María A, Capella-Gutíerrez S, Huerta-Cepas J, Gabaldón T, Dopazo J, Dopazo H. Phylemon 2.0: a suite of web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing. Nucl Acids Res. 2011; 39:W470–W474. [PubMed: 21646336]
24. Han M, Zmasek C. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinform. 2009; 10:356.
25. Stöver B, Müller K. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinform. 2010; 11:7.
26. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]
27. Waterhouse AM, Procter JB, Martin DMA, Clamp Ml, Barton GJ. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009; 25:1189–1191. [PubMed: 19151095]
28. Baños B, Lázaro JM, Villar L, Salas M, de Vega M. Characterization of a *Bacillus subtilis* 64-kDa DNA polymerase X potentially Involved in DNA repair. J Mol Biol. 2008; 384:1019–1028. [PubMed: 18938175]
29. Powell S, Szklarczyk D, Trachana K, Roth A, Kuhn M, Muller J, Arnold R, Rattei T, Letunic I, Doerks T, Jensen LJ, von Mering C, Bork P. eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. Nucl Acids Res. 2012; 40:D284–D289. [PubMed: 22096231]
30. Bebenek K, Garcia-Diaz M, Patishall SR, Kunkel TA. Biochemical properties of *Saccharomyces cerevisiae* DNA polymerase IV. J Biol Chem. 2005; 280:20051–20058. [PubMed: 15778218]
31. González-Barrera S, Sánchez A, Ruiz JF, Juáez R, Picher AJ, Terrados G, Andrade P, Blanco L. Characterization of SpPol4, a unique X-family DNA polymerase in *Schizosaccharomyces pombe*. Nucl Acids Res. 2005; 33:4762–4774. [PubMed: 16120966]
32. Venegas JA, Åslund L, Solari A. Cloning and characterization of a DNA polymerase β gene from *Trypanosoma cruzi*. Parasitol Int. 2009; 58:187–192. [PubMed: 19567232]
33. Bruhn DF, Sammartino MP, Klingbeil MM. Three mitochondrial DNA polymerases are essential for kinetoplast DNA replication and survival of bloodstream form *Trypanosoma brucei*. Eukaryot Cell. 2011; 10:734–743. [PubMed: 21531873]
34. Saxowsky TT, Matsumoto Y, Englund PT. The mitochondrial DNA polymerase β from *Crithidia fasciculata* has 5′-deoxyribose phosphate (dRP) lyase activity but is deficient in the release of dRP. J Biol Chem. 2002; 277:37201–37206. [PubMed: 12151410]
35. Lopes, DdO; Schamber-Reis, BLF.; Regisda-Silva, CG.; Rajão, MA.; DaRocha, WD.; Macedo, AM.; Franco, GR.; Nardelli, SC.; Schenkman, S.; Hoffmann, J-Sb; Cazaux, C.; Pena, SDJ.; Teixeira, SMR.; Machado, CR. Biochemical studies with DNA polymerase β and DNA polymerase β-PAK of *Trypanosoma cruzi* suggest the involvement of these proteins in mitochondrial DNA maintenance. DNA Repair (Amst). 2008; 7:1882–1892. [PubMed: 18761429]
36. Alonso A, Terrados G, Picher AJ, Giraldo R, Blanco L, Larraga V. An intrinsic 5′-deoxyribose-5-phosphate lyase activity in DNA polymerase beta from *Leishmania infantum* supports a role in DNA repair. DNA Repair (Amst). 2006; 5:89–101. [PubMed: 16174567]
37. Prasad R, Beard WA, Chyan JY, Maciejewski MW, Mullen GP, Wilson SH. Functional analysis of the amino-terminal 8-kDa domain of DNA polymerase β as revealed by site-directed mutagenesis:

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

DNA binding and 5′-deoxyribose phosphate lyase activities. J Biol Chem. 1998; 273:11121–11126. [PubMed: 9556598]

38. Prasad R, Batra VK, Yang XP, Krahn JM, Pedersen LC, Beard WA, Wilson SH. Structural insight into the DNA polymertase β deoxyribose phosphate lyase mechanism. DNA Repair (Amst). 2005; 4:1347–1357. [PubMed: 16172026]
39. Garcia-Diaz M, Bebenek K, Gao G, Pedersen LC, London RE, Kunkel TA. Structure-function studies of DNA polymerase lambda. DNA Repair (Amst). 2005; 4:1358–1367. [PubMed: 16213194]
40. Beard WA, Shock DD, Yang XP, DeLauder SF, Wilson SH. Loss of DNA polymerase β stacking interactions with templating purines, but not pyrimidines, alters catalytic efficiency and fidelity. J Biol Chem. 2002; 277:8235–8242. [PubMed: 11756435]
41. Beard WA, Wilson SH. DNA polymerases lose their grip. Nat Struct Biol. 2001; 8:915–917. [PubMed: 11685231]
42. Moon AF, Garcia-Diaz M, Batra VK, Beard WA, Bebenek K, Kunkel TA, Wilson SH, Pedersen LC. The X family portrait: structural insights into biological functions of X family polymerases. DNA Repair (Amst). 2007; 6:1709–1725. [PubMed: 17631059]
43. Baños B, Villar L, Salas M, de Vega M. Intrinsic apurinic/apyrimidinic (AP) endonuclease activity enables *Bacillus subtilis* DNA polymerase X to recognize, incise, and further repair abasic sites. Proc Natl Acad Sci U S A. 2010; 107:19219–19224. [PubMed: 20974932]
44. Nakane S, Nakagawa N, Kuramitsu S, Masui R. Characterization of DNA polymerase X from *Thermus thermophilus* HB8 reveals the POLXc and PHP domains are both required for 3′–5′ exonuclease activity. Nucl Acids Res. 2009; 37:2037–2052. [PubMed: 19211662]
45. Nakane S, Ishikawa H, Nakagawa N, Kuramitsu S, Masui R. The structural basis of the kinetic mechanism of a gap-filling X-family DNA polymerase that binds Mg^{2+}-dNTP before binding to DNA. J Mol Biol. 2012; 417:179–196. [PubMed: 22306405]
46. Oliveros M, Yánez RJ, Salas ML, Salas J, Vinuela E, Blanco L. Characterization of an African swine fever virus 20-kda DNA polymerase involved in DNA repair. J Biol Chem. 1997; 272:30899–30910. [PubMed: 9388236]
47. Baños B, Laro JM, Villar L, Salas M, de Vega M. Editing of misaligned 3′-termini by an intrinsic 3′–5′ exonuclease activity residing in the PHP domain of a family X DNA polymerase. Nucl Acids Res. 2008; 36:5736–5749. [PubMed: 18776221]
48. Batra VK, Perera L, Lin P, Shock DD, Beard WA, Pedersen LC, Pedersen LG, Wilson SH. Amino acid substitution in the active site of DNA polymerase β explains the energy barrier of the nucleotidyl transfer reaction. J Am Chem Soc. 2013; 135:8078–8088. [PubMed: 23647366]
49. Khairnar NP, Misra HS. DNA polymerase X from *Deinococcus radiodurans* implicated in bacterial tolerance to DNA damage is characterized as a short patch base excision repair polymerase. Microbiology. 2009; 155:3005–3014. [PubMed: 19542005]
50. Bailey S, Wing RA, Steitz TA. The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases. Cell. 2006; 126:893–904. [PubMed: 16959569]
51. Kirby TW, DeRose EF, Beard WA, Wilson SH, London RE. A thymine isostere in the templating position disrupts assembly of the closed DNA polymerase β ternary complex. Biochemistry. 2005; 44:15230–15237. [PubMed: 16285726]
52. Beard WA, Osheroff WP, Prasad R, Sawaya MR, Jaju M, Wood TG, Kraut J, Kunkel TA, Wilson SH. Enzyme-DNA interactions required for efficient nucleotide incorporation and discrimination in human DNA polymerase β. J Biol Chem. 1996; 271:12141–12144. [PubMed: 8647805]
53. Kraynov VS, Werneburg BG, Zhong XJ, Lee H, Ahn JW, Tsai MD. DNA polymerase β: analysis of the contributions of tyrosine-271 and asparagine-279 to substrate specificity and fidelity of DNA replication by pre-steady-state kinetics. Biochem J. 1997; 323:103–111. [PubMed: 9173867]
54. Vande Berg BJ, Beard WA, Wilson SH. DNA structure and aspartate 276 influence nucleotide binding to human DNA polymerase β: implication for the identity of the rate-limiting conformational change. J Biol Chem. 2001; 276:3408–3416. [PubMed: 11024043]
55. Cavanaugh NA, Beard WA, Batra VK, Perera L, Pedersen LG, Wilson SH. Molecular insights into DNA polymerase deterrents for ribonucleotide insertion. J Biol Chem. 2011; 286:31650–31660. [PubMed: 21733843]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

56. Osheroff WP, Beard WA, Wilson SH, Kunkel TA. Base substitution specificity of DNA polymerase β depends on interactions in the DNA minor groove. J Biol Chem. 1999; 274:20749–20752. [PubMed: 10409611]
57. Osheroff WP, Beard WA, Yin S, Wilson SH, Kunkel TA. Minor groove interactions at the DNA polymerase β active site modulate single-base deletion error rates. J Biol Chem. 2000; 275:28033–28038. [PubMed: 10851238]
58. Freudenthal BD, Beard WA, Wilson SH. Structures of dNTP intermediate states during DNA polymerase active site sssembly. Structure. 2012; 20:1829–1837. [PubMed: 22959623]
59. Beard WA, Shock DD, Vande Berg BJ, Wilson SH. Efficiency of correct nucleotide insertion governs DNA polymerase fidelity. J Biol Chem. 2002; 277:47393–47398. [PubMed: 12370169]
60. Batra VK, Beard WA, Shock DD, Krahn JM, Pedersen LC, Wilson SH. Magnesium induced assembly of a complete DNA polymerase catalytic complex. Structure. 2006; 14:757–766. [PubMed: 16615916]
61. Delarue M, Boulé JB, Lescar J, Expert-Bezancon N, Jourdan N, Sukumar N, Rougeon F, Papanicolaou C. Crystal structures of a template-independent DNA polymerase: murine terminal deoxynucleotidyltransferase. EMBO J. 2002; 21:427–439. [PubMed: 11823435]
62. Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA, Pedersen LC. Structural insight into the substrate specificity of DNA polymerase μ. Nat Struct Mol Biol. 2007; 14:45–53. [PubMed: 17159995]
63. Garcia-Diaz M, Bebenek K, Krahn JM, Pedersen LC, Kunkel TA. Structural analysis of strand misalignment during DNA synthesis by a human DNA polymerase. Cell. 2006; 124:331–342. [PubMed: 16439207]
64. Date T, Yamamoto S, Tanihara K, Nishimoto Y, Matsukage A. Aspartic acid residues at positions 190 and 192 of rat polymerase β are involved in primer binding. Biochemistry. 1991; 30:5286–5292. [PubMed: 2036395]
65. Menge KL, Hostomsky Z, Nodes BR, Hudson GO, Rahmati S, Moomaw EW, Almassy RJ, Hostomska Z. Structure-function analysis of the mammalian DNA polymerase β active site: role of aspartic acid 256, arginine 254, and arginine 258 in nucleotidyl transfer. Biochemistry. 1995; 34:15934–15942. [PubMed: 8519750]
66. Date T, Yamamoto S, Tanihara K, Nishimoto Y, Liu N, Matsukage A. Site-directed mutagenesis of recombinant rat DNA polymerase β: involvement of arginine-183 in primer recognition. Biochemistry. 1990; 29:5027–5034. [PubMed: 2198936]
67. Li SX, Vaccaro JA, Sweasy JB. Involvement of phenylalanine 272 of DNA polymerase beta in discriminating between correct and incorrect deoxynucleoside triphosphates. Biochemistry. 1999; 38:4800–4808. [PubMed: 10200168]
68. Lavrik OI, Prasad R, Beard WA, Safronov IV, Dobrikov MI, Srivastava DK, Shishkin GV, Wood TG, Wilson SH. dNTP binding to HIV-1 reverse transcriptase and mammalian DNA polymerase β as revealed by affinity labeling with a photoreactive dNTP analog. J Biol Chem. 1996; 271:21891–21897. [PubMed: 8702991]
69. Liu J, Tsai MD. DNA Polymerase β: pre-steady-state kinetic analyses of dATPαS stereoselectivity and alteration of the stereoselectivity by various metal ions and by site-directed mutagenesis. Biochemistry. 2001; 40:9014–9022. [PubMed: 11467964]
70. Miller H, Prasad R, Wilson SH, Johnson F, Grollman AP. 8-OxodGTP incorporation by DNA polymerase β is modified by active-site residue Asn279. Biochemistry. 2000; 39:1029–1033. [PubMed: 10653647]
71. Ahn J, Werneburg BG, Tsai MD. DNA polymerase β: structure-fidelity relationship from pre-steady-state kinetic analyses of all possible correct and incorrect base pairs for wild type and R283A mutant. Biochemistry. 1997; 36:1100–1107. [PubMed: 9033400]
72. Bakhtina M, Roettger MP, Tsai MD. Contribution of the reverse rate of the conformational step to polymerase β fidelity. Biochemistry. 2009; 48:3197–3208. [PubMed: 19231836]
73. Starcevic D, Dalal S, Sweasy J. Hinge residue Ile260 of DNA polymerase β is important for enzyme activity and fidelity. Biochemistry. 2005; 44:3775–3784. [PubMed: 15751954]

74. Roettger MP, Bakhtina M, Tsai MD. Mismatched and matched dNTP incorporation by DNA polymerase β proceed via analogous kinetic pathways. Biochemistry. 2008; 47:9718–9727. [PubMed: 18717589]

75. Opresko PL, Sweasy JB, Eckert KA. The mutator form of polymerase β with amino acid substitution at tyrosine 265 in the hinge region displays an increase in both base substitution and frame shift errors. Biochemistry. 1998; 37:2111–2119. [PubMed: 9485358]

76. Matsumoto Y, Kim K. Excision of deoxyribose phosphate residues by DNA polymerase β during DNA repair. Science. 1995; 269:699–702. [PubMed: 7624801]

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dnarep.2014.07.003.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

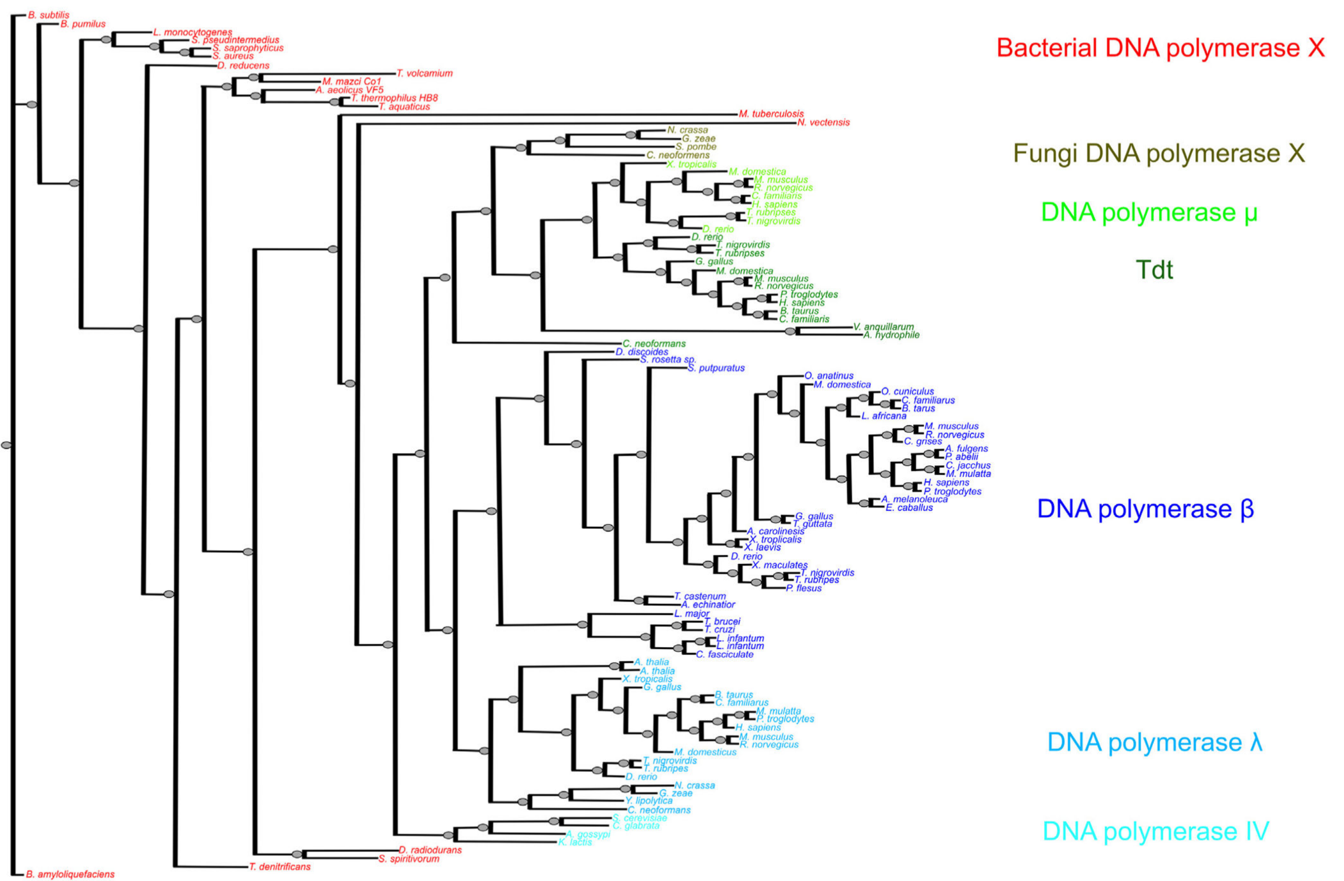

Fig. 1.

Phylogenetic tree for X-family DNA polymerase sequences. The tree was generated using PhyML v. 3.0 [18,19]. The branches are colored according to general subclasses.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

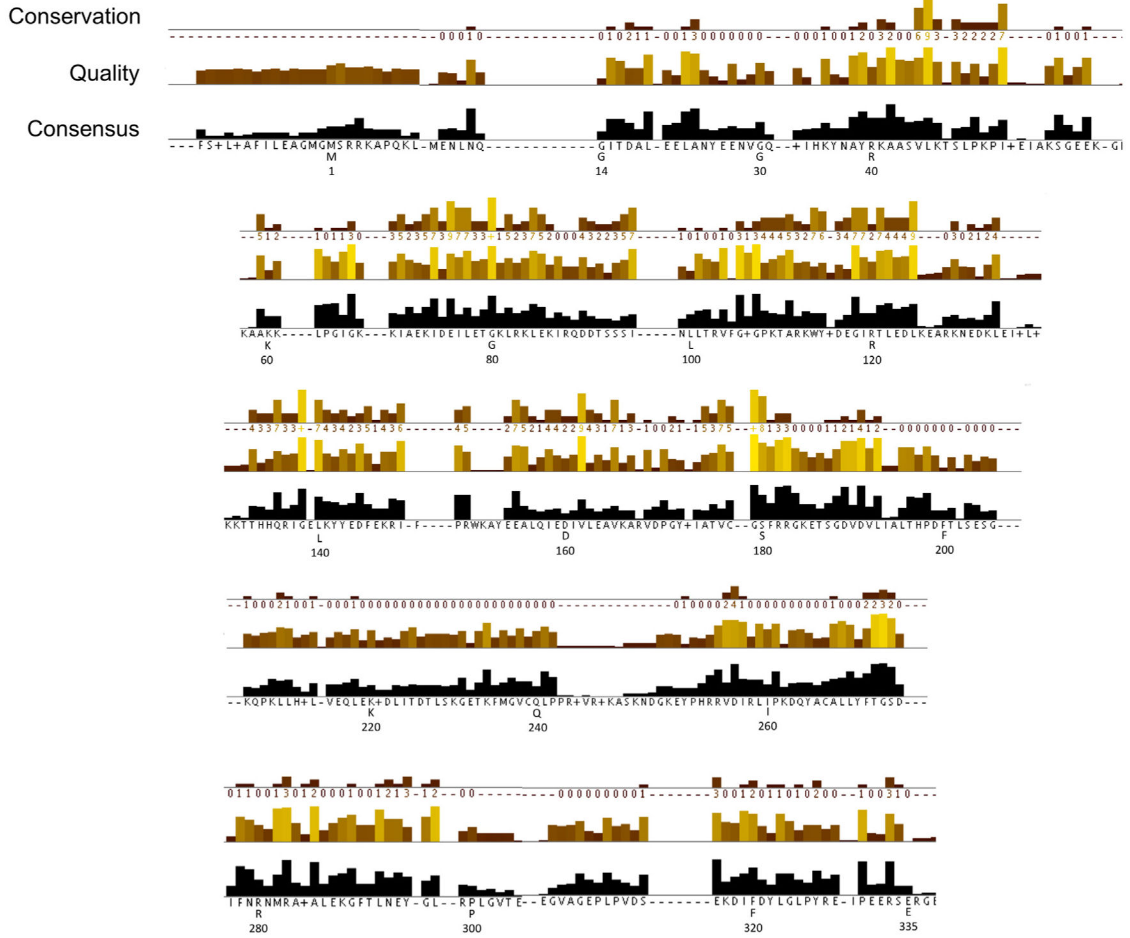

Fig. 2.

Amino acid sequence conservation and analysis corresponding to human pol β. The numbering under the sequence corresponds to that for human pol β, while the letter indicates the consensus sequence at that position. The JalView alignment quality displayed below the conservation columns is a measure of the likelihood of observing mutations at that position in the alignment [27]. Specifically, the quality score is calculated for each column in the alignment by summing, for all mutations, the ratio of the two BLOSUM 62 scores for a mutation pair and each residue's conserved BLOSUM62 score (which is higher). This value is normalized for each column, and then plotted on a scale from 0 to 1. Conservation is measured as a numerical index reflecting the conservation of chemical properties in the alignment; identical residues score highest with the next most conserved group containing substitutions to amino acids within the same chemical class. Conservation is visualized on the alignment as a histogram giving the score for each column. Conserved columns are indicated by '*' (score of 11 with default amino acid property grouping), and columns with mutations with conserved chemical properties are marked with a '+' (score of 10, indicating all properties are conserved).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

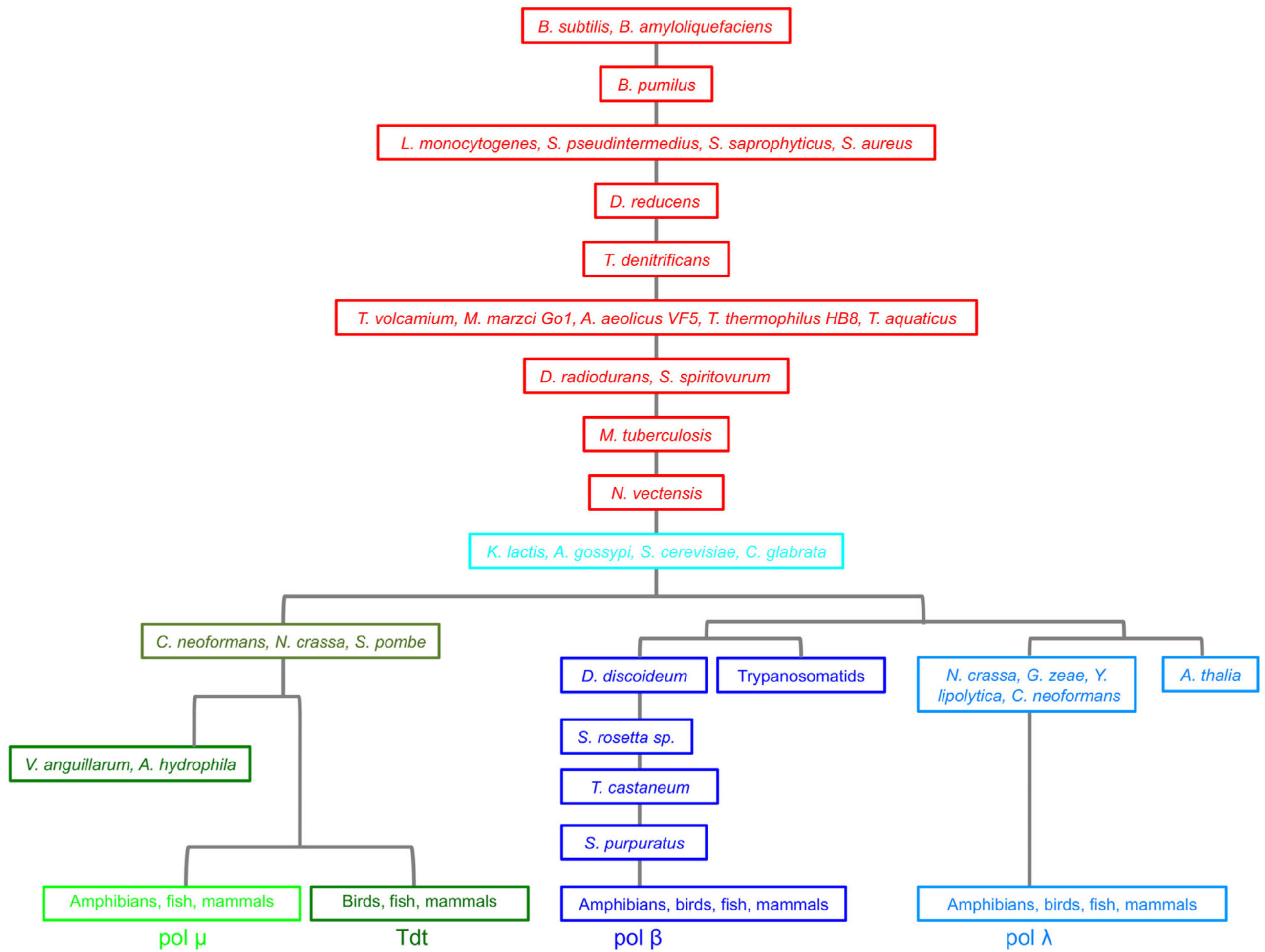

Fig. 3.

Schematic family tree illustrating the origin and evolution of the X-family DNA polymerases. The most ancient family members are from *Bacillus* bacterial species.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

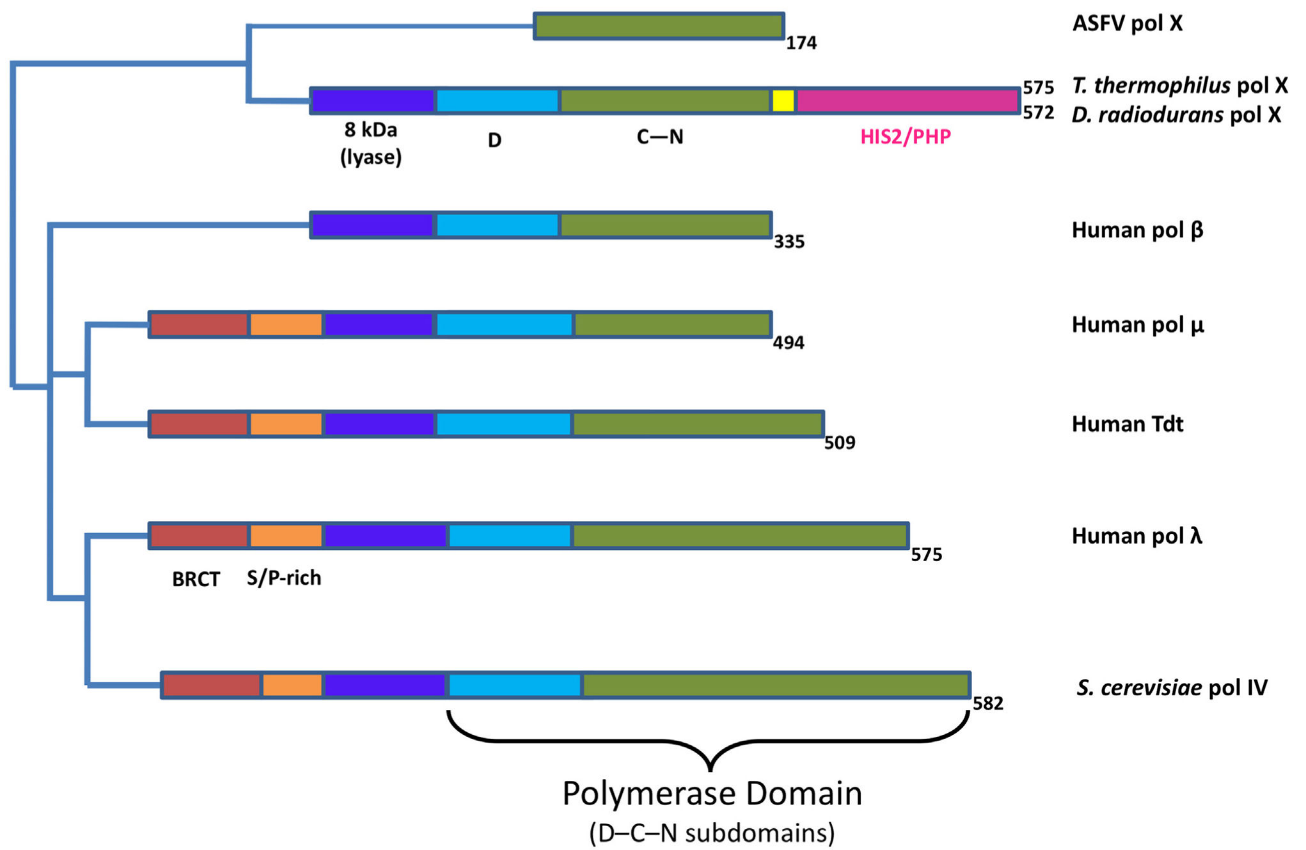

Fig. 4.

Domain organization of selected X-family DNA polymerases. Mammalian DNA polymerase β is composed of two domains that complement its biological role in base excision repair: lyase (purple) and polymerase (light blue and green). The polymerase domain includes three subdomains: catalytic (C, green) and nucleotide binding (N, green) and DNA binding (D, light blue). DNA polymerase μ, Tdt, pol λ and Pol IV have an amino-terminal BRCT domain (light red). The two bacterial enzymes include a carboxyl-terminal PHP/HIS2 domain (pink) connected to the polymerase domain with a linker (yellow). Pol λ includes a serine-proline-rich region (S/P-rich, orange).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

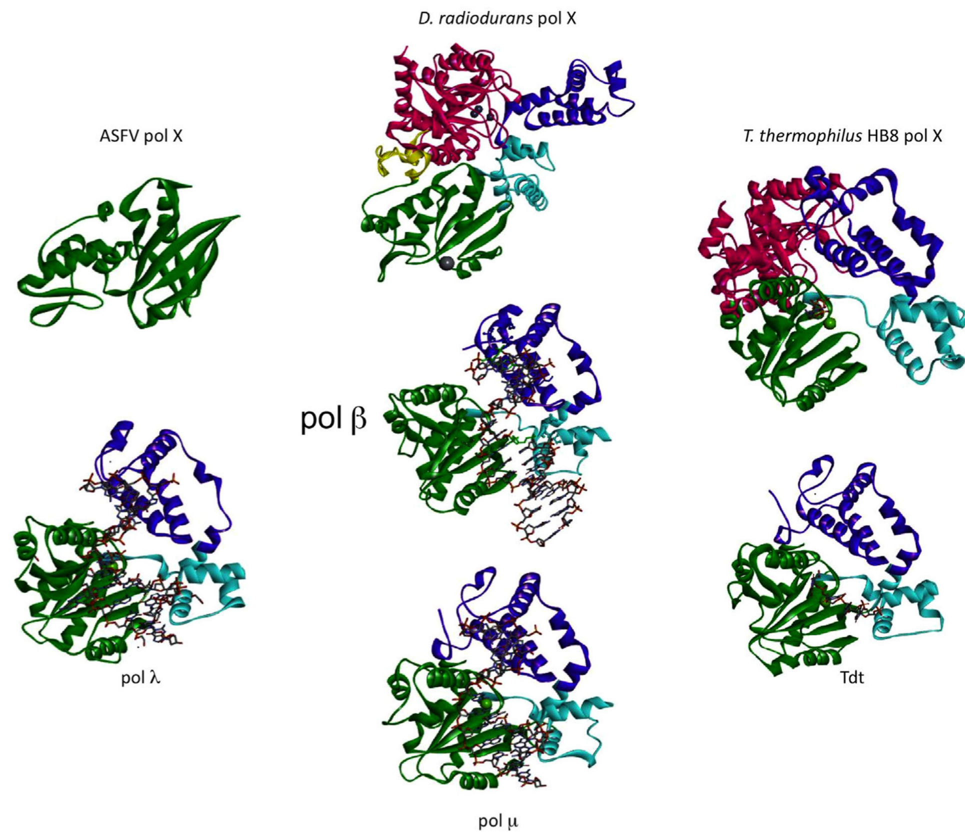

Fig. 5.

Representative crystal structures of X-family DNA polymerases. The domains/subdomains are colored according to the scheme used in Fig. 4 and are arranged to show similar orientations. The PDB IDs are: human pol β, 2FMS [60]; ASFV pol X, 1JAJ [16]; *D. radiodurans* pol X, 2W9M [13]; *T. thermophilus* HB8 pol X, 3AU2 [45]; truncated mouse Tdt, 1KDH [61]; truncated mouse μ, 2IHM [62]; truncated human pol λ, 2BCR [63]. When present, DNA is shown in a stick representation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

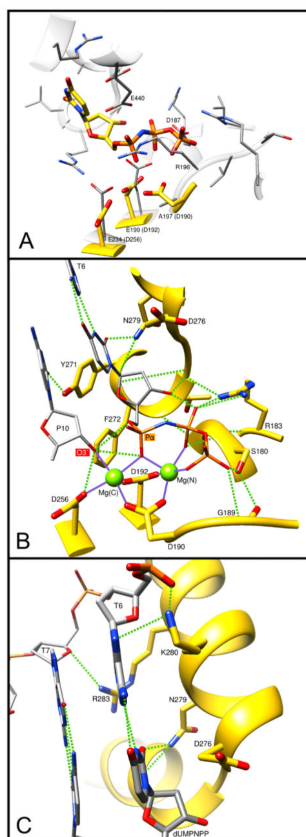

Fig. 6.

Molecular interactions of key DNA polymerase β side-chains. The ternary substrate (DNA/dUMPNPP) complex of human pol β (PDB ID 2FMS) was used to illustrate these interactions. (A) Overlay of the pol β active site (yellow carbons) with that of *D. radiodurans* pol X (gray carbons, PDB ID 2W9M). Key *D. radiodurans* residues are indicated and the equivalent pol β residues are given in parentheses. Arg196 (R196) of *D. radiodurans* appears to occlude the triphosphate binding pocket. In addition, pol β Asp190 (D190) is not conserved; *D. radiodurans* Ala197 (A197). (B) Key protein interactions with the incoming nucleotide (dUMPNPP). Hydrogen bonds are indicated with green dashed lines and metal coordinations are shown as solid purple lines. The catalytic and nucleotide binding metals are shown as green spheres: Mg(C) and Mg(N), respectively. Protein and substrate carbons are yellow and gray, respectively. The primer terminus (P10) O3′ and the α-phosphate of the incoming nucleotide are indicated. T6 identifies the templating base. The function and conservation of these residues are tabulated in Table 2 and Supplementary Table 1. (C) Key protein interactions with the bases of the nascent base pair. Hydrogen bonds are indicated with green dashed lines. T7 identifies the templating base immediately upstream of the coding templating base T6. The base of the incoming nucleotide (dUMPNPP) is indicated. The function and conservation of these residues are tabulated in Table 2 and Supplementary Table 1.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

DNA polymerase family X members used for sequence alignments and subsequent phylogenetic analyses.

	Prokaryotic DNA polymerase family X sequences
1.	*Aeromonas hydrophila subsp. hydrophila* ATCC 7966 (gb|ABK36337.1|)
2.	*Aquifex aeolicus VF5* (ref|NP 213981.1|)
3.	*Aurelia sp. 1 sensu* (UniProtKB: F7J5W8 (F7J5W8 9CNID))
4.	*Bacillus amyloliquefaciens* (UniProtKB: F4EI46 (F4EI46 BACAM))
5.	*Bacillus pumilus SAFR-032* (gb|ABV63179.1|)
6.	*Bacillus subtilis subsp. subtilis str. 168* (emb|CAB14819.1|)
7.	*Deinococcus radiodurans R1* (ref|NP 294190.1|)
8.	*Desulfotomaculum reducens MI-1* (ref|YP 001112987.1|)
9.	*Listeria monocytogenes serotype 4b str. F2365* (ref|YP 013839.1|)
10.	*Methanosarcina mazei Go1* (ref|NP 633918.1)
11.	*Mycobacterium tuberculosis* (ref|NP 218373.1)
12.	*Salpingoeca rosetta sp.* ATCC 50818 (gb|EGD82858.1|)
13.	*Sphingobacterium spiritivorum* ATCC 33300 (ref|ZP 03968066.1|)
14.	*Staphylococcus aureus subsp. aureus JH9* (ref|YP 001246578.1|)
15.	*Staphylococcus pseudintermedius str. HKU10-03* (UniProtKB: E8SFK2 STAPH)
16.	*Staphylococcus saprophyticus subsp. saprophyticus* ATCC 15305 (ref|YP 301742.1|)
17.	*Thermoplasma volcanium GSS1* (ref|NP 111375.1|)
18.	*Thermus aquaticus* (tr|P77987|P77987 THEAQ)
19.	*Thermus thermophilus HB8* (ref|YP 144416.1|)
20.	*Thiobacillus denitrificans* ATCC 25259 (gb|AAZ97399.1|)
21.	*Vibrio anguillarum 775* (gb|AEH34679.1|)
	Eukaryotic DNA polymerase family X sequences
1.	*Acromyrmex echinatior* (ant; gb|EGI68014.1|)
2.	*Ailuropoda melanoleuca* (giant panda; ref|XP 002918406.1|)
3.	*Ailurus fulgens* (red panda; gb|ADT64454.1|)
4.	*Anolis carolinensis* (lizard; ref|XP 003228273.1|)
5.	*Arabidopsis thaliana* (mouse ear cress flowering plant; λ-NCBI GenPept: ADM33939.1 gi:304440990)
6.	*Ashbya gossypii* (filamentous fungus; UniProtKB: Q757Q1 ASHGO)
7.	*Bos taurus* (cow; β-ref|NP 001029936.1|, λ-ref|NP 001179488.1|, μ-gb|AAX46342.1|, Tdt-UniProtKB: P06526 TDT BOVIN)
8.	*Callithrix jacchus* (marmoset; β-ref|XP 002757050.1|)
9.	*Candida glabrata* (haploid yeast; UniProtKB: Q6FKY2 CANGA)
10.	*Canis lupus familiaris* (domestic dog; β-NCBI GenPept: AAV66968, λ-ref|XP 861980.1|, Tdt-ref|XP 005637632.1|, μ-RefSeq status WITHDRAWN)
11.	*Cricetulus griseus* (guinea pig; gb|EGV95023.1|)
12.	*Crithidia fasciculate* (parasitic protist; gb|AAA68599.2|)
13.	*Cryptococcus neoformans var. neoformans B-3501A* (encapsulated yeast; ref|XP 777918.1|)
14.	*Cryptococcus neoformans var. neoformans JEC21* (encapsulated yeast; ref|XP 568043.1|)
15.	*Cryptococcus neoformans var. grubii H99* (fungal pathogen causing fatal meningitis; λ-ref|XP 773693.1|)
16.	*Danio rerio* (zebrafish; β-ref|NP 001003879|, λ-ref|NP 998408.1|, Tdt-gb|AAI63775.1|)
17.	*Dictyostelium discoideum* (slime mold; UniProtKB: Q1ZXF2 DICDI)

NIH-PA Author Manuscript

18. *Equus caballus* (horse; β-ref|XP 001489108.2|)
19. *Fusarium graminearum PH-1* (fungal plant pathogen; ref|XP 382072.1|)
20. *Gallus gallus domesticus* (chicken; β-ref|XP 001143904|, λ-ref|XP 001232209.2|, Tdt-UniProtKB: P36195 TDT CHICK)
21. *Homo sapiens* (human; β-ref|NP 002681.1|, λ-UniProtKB: Q9UGP5 DPOLL HUMAN, μ-UniProtKB: Q9NP87 DPOLM HUMAN, Tdt-UniProtKB: P04053 TDT HUMAN)
22. *Kluyveromyces lactis* (yeast; UniProtKB: Q6CX59 KLULA)
23. *Leishmania infantum* (parasite; gb|AAF00495.1|)
24. *Leishmania major* (parasite; UniProtKB: Q4QI80 LEIMA)
25. *Lepeophtheirus salmonis* (sea louse; tr|C1BUF7|C1BUF7 9MAXI)
26. *Loxodonta africana* (elephant; ref|XP 003412553.1|)
27. *Macaca mulatta* (rhesus monkey; β-ref|XP 001097548.1|, λ-ref|NP 001253835.1|)
28. *Monodelphis domestica* (gray short-tailed opossum; β-ref|XP 001373104.1|, λ-ref|XP 001369819.1|, μ-UniProtKB: F6SV89 MONDO, Tdt-UniProtKB:O02789 TDT MONDO)
29. *Mus musculus* (mouse; β-ref|NP 035260.1|, λ-UniProtKB: Q9QXE2 DPOLL MOUSE, μ-UniProtKB: Q9JIW4 DPOLM MOUSE, Tdt-UniProtKB: P09838 TDT MOUSE)
30. *Nematostella vectensis* (starlet sea anemone; β-ref|XP 001618045.1|)
31. *Neurospora crassa OR74A* (bread mold; β-ref|XP 961407.1|, λ-ref|XP 963912.2|)
32. *Ornithorhynchus anatinus* (platypus; ref|XP 001509890.2|)
33. *Oryctolagus cuniculus* (rabbit; ref|XP 002720811.1|)
34. *Pan troglodytes* (chimpanzee; β-ref|XP 001143904|, λ-ref|NP 001267179.1|, Tdt-ref|XP 521569.1|)
35. *Platichthys flesus* (European flounder; emb|CAC28866.1|)
36. *Pongo abelii* (orangutan; ref|XP 002819090.1|)
37. *Rattus norvegicus* (rat; β-ref|NP 058837.2|, λ-UniProtKB: Q5RKI3 DPOLL RAT, μ-ref|NP 001011912.1|, Tdt-gb|EDL94195.1|)
38. *Saccharomyces cerevisiae* (budding yeast; UniProtKB: P25615 DPO4 YEAST)
39. *Schizosaccharomyces pombe 972h* (fission yeast; ref|NP 592977.1|)
40. *Strongylocentrotus purpuratus* (purple sea urchin; ref|XP 787665.2|)
41. *Taeniopygia guttata* (zebra finch; ref|XP 002186553.1|)
42. *Takifugu rubripes* (Fugu rubripes puffer fish; β-ref|XP 003974975.1|, λ ref|XP 003972614.1|, μ-ref|XP 003965471.1|, Tdt-ref|NP 001027915.1|)
43. *Tetraodon nigrovirdis* (green spotted puffer fish; β-ref|XP 003974975.1|, λ-emb|CAG03351.1|, μ-emb|CAG05362.1|, Tdt-emb|CAG10152.1|)
44. *Tribolium castaneum* (red flour beetle; dbj|BAK40156.1|)
45. *Trypanosoma brucei* (protozoan parasite; gb|AAQ56191.1|)
46. *Trypanosoma cruzi* (protozoan parasite; gb|ACQ66108.1|)
47. *Xenopus laevis* (frog; sp|O57383|DPOLB XENLA)
48. *Xenopus tropicalis* (western clawed frog; β-ref|NP 001006894|, λ-ref|NP 001093716.1|, μ-UniProtKB: Q5FVA7 XENTR)
49. *Xiphophorus maculates* (southern platyfish; gb|AAU11318.1|)
50. *Yarrowia lipolytica* (non-conventional yeast; λ-ref|XP 502740.1|)

Gene identifiers are indicated in parenthesizes: ref – RefSeq (NCBI Reference Sequence Database); UniProtKB – The UniProt Knowledgebase (UniProtKB), UniProtKB/Swiss-Prot and UniProtKB/TrEMBL; NCBI GenPept, gb, gi – GenBank; emb, EMBL-EBI – European Nucleotide archive; dbj – DNA Databank of Japan; sp – Swiss-Prot; tr – TrEMBL.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

DNA polymerase β residues involved in ligand binding/catalysis.

Residue	Interaction	Reference[a]
Catalytic site		
Asp190	Coordinates two divalent metals	[64]
Asp192	Coordinates two divalent metals	[64]
Asp256	Coordinates catalytic metal	[48,65]
dNTP binding		
Ser180	Coordinates Pγ	[53]
Arg183	Coordinates Pβ	[53,66]
Ser188	H-bonds with Ser180	[53]
Gly189	Coordinates Pγ	
Tyr271	Sugar discrimination	[52,53,55]
Phe272	Sugar discrimination; subdomain motions	[67]
Asp276	Base van der Waals interactions	[54,68,69]
Asn279	H-bond with the minor groove edge of the base	[52,53,70]
Templating (coding) nucleotide		
Lys280	Base van der Waals interactions	[40,53]}
Arg283	Interactions with the minor groove edge of the base	[52,71]
Subdomain motions (open–closed)		
Arg258	Alters salt bridge	[72]
Ile260	Modulates α-helix M rotation	[72–74]
Tyr265	Modulates α-helix M rotation	[75]
Glu295	Altered H-bonding	[53]
Lyase activity (dRP binding pocket)		
Glu26	Facilitate sugar deprotonation (C2′)	
Ser30	Facilitate sugar deprotonation (C2′)	
His34	Active site pocket	[37]
Lys35	5′-PO_4 binding; sugar ring opening	[37,76]
Tyr39	Stabilize deprotonated Lys72	[76]
Lys41	Active site pocket	
Lys60	Active site pocket	[37]
Lys68	5′-PO_4 binding	[37,76]
Glu71	Facilitate sugar deprotonation (C2′)	[37]
Lys72	5′-PO_4 binding; Nucleophile involved in C1′ attack	[37,76]
Lys84	Alternative nucleophile	[37,76]

[a] DNA polymerase β site-directed mutagenesis studies.