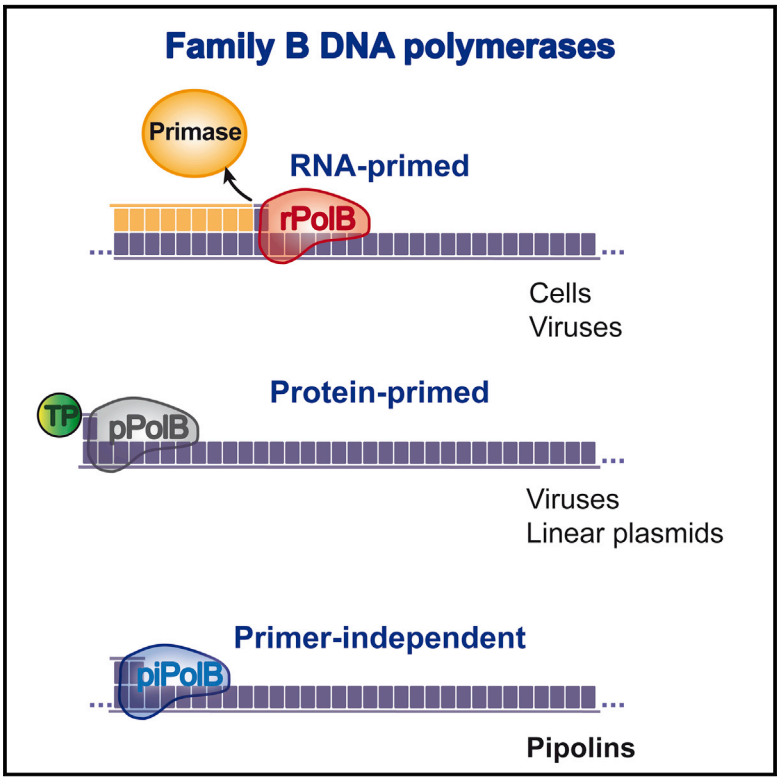
Article

Cell Reports

Primer-Independent DNA Synthesis by a Family B DNA Polymerase from Self-Replicating Mobile Genetic Elements

Graphical Abstract



Highlights

- Primer-independent PolBs (piPolBs) display templated *de novo* DNA synthesis capacity
- piPolBs denote a third major group of family B DNA polymerases
- piPolBs are the hallmark of *pipolins*, self-replicating mobile genetic elements
- *Pipolins* are widespread among diverse bacterial phyla and mitochondria

Authors

Modesto Redrejo-Rodríguez, Carlos D. Ordóñez, Mónica Berjón-Otero, ..., Patrick Forterre, Margarita Salas, Mart Krupovic

Correspondence

modesto.redrejo@csic.es (M.R.-R.), msalas@cbm.csic.es (M.S.), krupovic@pasteur.fr (M.K.)

In Brief

Redrejo-Rodríguez et al. report and characterize a DNA polymerase group (piPolB) from the B family that can perform primer-independent DNA replication. PiPolBs are encoded by Pipolins, diverse self-replicating genetic elements that are widespread among bacterial phyla and in mitochondria.

Redrejo-Rodríguez et al., 2017, Cell Reports *21*, 1574–1587
November 7, 2017
https://doi.org/10.1016/j.celrep.2017.10.039

OPEN ACCESS
CellPress

Cell Reports
Article

Primer-Independent DNA Synthesis by a Family B DNA Polymerase from Self-Replicating Mobile Genetic Elements

Modesto Redrejo-Rodríguez,[1,5,*] Carlos D. Ordóñez,[1] Mónica Berjón-Otero,[1,3] Juan Moreno-González,[1] Cristian Aparicio-Maldonado,[1,4] Patrick Forterre,[2] Margarita Salas,[1,*] and Mart Krupovic[2,*]

[1]Centro de Biología Molecular "Severo Ochoa," Consejo Superior de Investigaciones Científicas and Universidad Autónoma de Madrid, Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain

[2]Institut Pasteur, Unité Biologie Moléculaire du Gène chez les Extrêmophiles, Paris, France

[3]Present address: Max Plank Institute for Medical Research, Heidelberg, Germany

[4]Present address: Department of Bionanoscience, Delft University of Technology, Delft, the Netherlands

[5]Lead Contact

*Correspondence: modesto.redrejo@csic.es (M.R.-R.), msalas@cbm.csic.es (M.S.), krupovic@pasteur.fr (M.K.)

https://doi.org/10.1016/j.celrep.2017.10.039

SUMMARY

Family B DNA polymerases (PolBs) play a central role during replication of viral and cellular chromosomes. Here, we report the discovery of a third major group of PolBs, which we denote primer-independent PolB (piPolB), that might be a link between the previously known protein-primed and RNA/DNA-primed PolBs. PiPolBs are encoded by highly diverse mobile genetic elements, *pipolins*, integrated in the genomes of diverse bacteria and also present as circular plasmids in mitochondria. Biochemical characterization showed that piPolB displays efficient DNA polymerization activity that can use undamaged and damaged templates and is endowed with proofreading and strand displacement capacities. Remarkably, the protein is also capable of template-dependent *de novo* DNA synthesis, i.e., DNA-priming activity, thereby breaking the long-standing dogma that replicative DNA polymerases require a pre-existing primer for DNA synthesis. We suggest that piPolBs are involved in self-replication of pipolins and may also contribute to bacterial DNA damage tolerance.

INTRODUCTION

DNA polymerases (DNAPs) are key enzymes essential for genome replication, recombination, and repair across all cellular life forms and their viruses. Family B DNAPs (PolBs) are involved in genome replication in Eukarya and Archaea, but also in viruses from the three domains of life (Koonin, 2006). For the initiation of DNA synthesis, all PolBs characterized thus far depend on the presence of an external primer, a hydroxyl group presented either by a nucleic acid (RNA or DNA) or the so-called terminal protein (TP). Thus, based on the primer requirement for the initiation of genome replication as well as phylogenetic clustering, these enzymes can be broadly divided into two major groups (Filée et al., 2002): RNA primed (rPolBs) and protein primed (pPolBs). The evolutionary relationship between the two groups is unknown, and it is thus unclear whether the ancestral PolB would have employed a protein or an RNA oligonucleotide as a primer. The rPolB group contains mainly replicases devoted to accurate and efficient copying of large cellular and viral genomes. By contrast, pPolBs are exclusive to selfish mobile genetic elements (MGEs) and viruses with moderately sized linear genomes (<50 kb) (Kazlauskas and Venclovas, 2011; Krupovic and Koonin, 2015). The signature of pPolBs is the presence of specific subdomains, named TPR1 and TPR2, which were originally described in Φ29 DNAP. TPR1 is required for the DNAP interaction with the TP, whereas TPR2 endows pPolB with the processivity and strand displacement capacity (Dufour et al., 2000; Rodríguez et al., 2005; Salas et al., 2016), the two properties ensuring the superiority of Φ29DNAP in various molecular biology applications, such as multiple-displacement single-molecule DNA amplification (Hutchison et al., 2005; Sidore et al., 2016).

Protein priming has been described for a number of viruses (Berjón-Otero et al., 2016; Hoeben and Uil, 2013; Salas, 1991) as well as for linear eukaryotic plasmids (Klassen and Meinhardt, 2007). More recently, pPolB-encoding genes were also identified in two superfamilies of MGE integrated into various cellular genomes (Krupovic and Koonin, 2016). The first superfamily comprises eukaryotic virus-like transposable elements, called Polintons (also known as Mavericks), which, besides pPolB, encode retrovirus-like integrases and a set of proteins predicted to be involved in the formation of viral particles (Krupovic et al., 2014a). The second supergroup, denoted casposons, is present in a wide range of archaea and some bacteria (Krupovic et al., 2014b). For integration into the cellular genome, casposons employ endonucleases homologous to Cas1, a signature enzyme of the CRISPR-Cas systems (Béguin et al., 2016; Krupovic et al., 2017). It has been postulated that pPolBs participate in the replication of the casposon and polinton genomes (Kapitonov and Jurka, 2006; Krupovic et al., 2014b); accordingly, these MGEs are referred to as self-synthesizing or self-replicating elements.

Here we uncover a highly diverse superfamily of self-replicating MGEs, dubbed *pipolins*, which are present in three major

bacterial phyla, as well as in mitochondria, and encode divergent PolB carrying TPR1 and TPR2 subdomains. Biochemical characterization of a representative enzyme encoded by a pipolin from *Escherichia coli* showed that the protein displays a versatile and efficient DNA replication capacity. Strikingly, the protein is also capable of an intrinsic *de novo* DNA synthesis, i.e., DNA-priming activity, not previously described in members of the PolB family. This group of DNAPs, which we denote primer-independent PolB (piPolB), should be sufficient to initiate and carry out an entire replication cycle of the circular pipolin DNA *in vivo*. Moreover, enhanced survival of *E. coli* cells expressing piPolB upon replication blockage by DNA-damaging agents suggests an additional role of piPolB in bacterial DNA damage tolerance.

RESULTS

A Third Major Group of PolBs

Position-specific iterative (PSI)-BLAST searches against the RefSeq bacterial genome database at NCBI seeded with the sequence of experimentally characterized pPolB from bacteriophage Bam35 (NP_943751) retrieved numerous hits to highly divergent PolBs (~16%–20% sequence identity to the pPolB of Bam35) encoded within chromosomes and several plasmids from widely diverse bacteria, such as Firmicutes, Actinobacteria, and Proteobacteria. Nevertheless, analysis of multiple sequence alignments of these divergent DNAPs showed that all of them contained the TPR1 and TPR2 subdomains, a hallmark of pPolBs (Figure 1A), and the active site residues of the exonuclease and DNAP domains of PolBs (Braithwaite and Ito, 1993) were conserved, albeit with notable variations within the KxY and PolC motifs (Figure 1B, see also below). The PolC motif (YxDTDS) is almost universally conserved in PolBs (Braithwaite and Ito, 1993) and contains two catalytic aspartic acid residues required for protein activity (Bernad et al., 1990; Copeland and Wang, 1993). We noticed that, in all members of the divergent PolB group, the first of the two aspartates within the PolC motif was substituted for a threonine residue (SxTTDG, see Figure 1B). Notably, some archaeal pPolBs also showed variation within this motif (Bath et al., 2006; Krupovic et al., 2014b; Peng et al., 2007), but none of these proteins has been experimentally characterized.

Additional searches seeded with representative sequences of the divergent PolB group from proteobacteria, such as *Escherichia coli* (KDU42669) or *Rhodobacterales* bacterium Y4I (WP_008555115), yielded significant hits to several homologs encoded by pCRY1-like circular mitochondrial plasmids. Notably, the latter plasmids were distinct from the extensively studied linear mitochondrial plasmids, which encode pPolBs (see below). Sequence analysis of the mitochondrial proteins confirmed their close similarity to the divergent group of bacterial PolBs (Figures 1A and 1B).

Maximum likelihood phylogenetic analysis of representative sequences from all known clades of PolBs revealed that the divergent pPolBs formed a distinct, well-supported clade, which we denote piPolB (see below), separated from all other pPolBs (Figure 1C), suggesting that it diverged early in the evolution of PolBs. Thus, piPolB represents the third major group of PolBs, besides rPolBs and pPolBs. Within the piPolB clade, there are two major groups, which are roughly congruent with the bacterial taxonomy (Figure 1C; Figure S1). The first one predominantly includes sequences from Actinobacteria and several orders of Firmicutes, namely, Bacillales, Lactobacillales, and Clostridiales. The second group contains sequences from different classes of Proteobacteria. Notably, the latter group also includes piPolBs from circular mitochondrial plasmids, which cluster with sequences from alphaproteobacteria (Figure 1C; Figure S1).

piPolBs Are Encoded within Self-Replicating MGEs

Genomic context analysis provided compelling evidence that the majority of piPolBs are encoded within MGEs integrated into bacterial chromosomes. Unlike casposons and polintons, the vast majority of piPolB-carrying MGEs encoded integrases of the tyrosine recombinase superfamily (Y-integrases). Some of the elements carried additional copies of Y-integrases or integrases/invertases of the serine recombinase superfamily (Figure 1D; Figure S2). Nevertheless, several bacterial and all mitochondrial piPolB homologs were encoded by extrachromosomal, rather than integrated, plasmids, and, accordingly, they lacked the integrase genes, suggesting that integration into the chromosome is optional for these MGEs. Hence, we refer to all these bacterial and mitochondrial elements as piPolB-encoding MGEs (*pipolins*).

The Y-integrases typically catalyze recombination between homologous sites present on the cellular genome and the circular double-stranded DNA (dsDNA) molecules of the MGEs. Thorough analysis of the piPolB-encompassing genomic regions allowed us to define the precise integration sites for many pipolins from diverse bacterial taxa (Table S1; Figure S2), with tRNA genes being the most common integration site. Comparative genomic analysis of pipolins showed that they formed groups that were generally consistent with the phylogeny of the piPolBs (Figure 1D; Figure S2). Besides piPolB and integrases, pipolins often encoded excisionases, components of type I and type II restriction modification systems, and various plasmid mobilization proteins (Figure S2; Table S2). In addition, the less conserved genes found in pipolins encoded different DNA-binding proteins with ribbon-helix-helix, zinc-finger, or helix-turn-helix motifs, as well as histone-like H-NS chromatin proteins, various nucleases, and toxin-antitoxin systems. None of the elements encoded virus-specific proteins. By contrast, the pangenome of pipolins consisted of various genes typical of plasmids (Table S2). Consistent with this assertion, four of the bacterial and five mitochondrial piPolBs were encoded by circular plasmids (Figure S1). Notably, the mitochondrial plasmids carry no other genes than those encoding piPolB (Gobbi et al., 1997; Li and Nargang, 1993; Schulte and Lambowitz, 1991), suggesting that, following the introduction of a mitochondrial ancestor into a proto-eukaryotic host, the MGE underwent reductive evolution.

Although piPolB is the only DNA replication-associated protein conserved in all pipolins, some elements encode putative helicases of superfamilies 1 and 2, 3′-5′ exonucleases, uracil-DNA glycosylases, ribonucleases H, and an Orc1/Cdc6-like AAA+ ATPase. Unlike pPolB-encoding plasmids and viruses, which, as a rule, have linear genomes, pipolins represent circular dsDNA molecules and, thus, the protein-priming mechanism is unlikely to be applicable. The overwhelming majority (94%) of dsDNA viruses encoding RNA-primed DNAPs also encode their

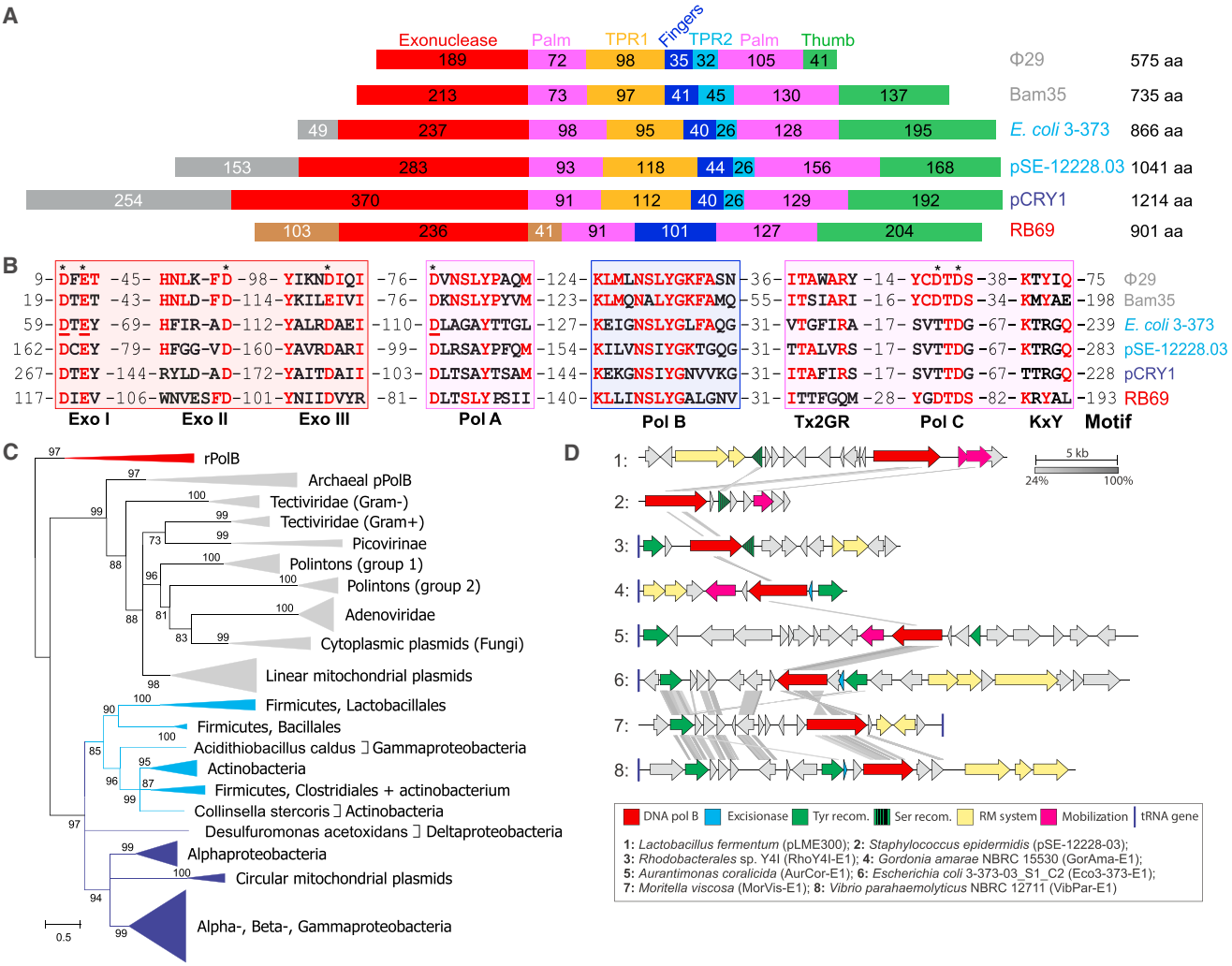


Figure 1. A Major Group of PolBs

(A) Schematic representation of representative PolBs. Conserved domains are represented as colored boxes and their length is indicated. Non-conserved N-terminal domains of RB69 and piPolBs are colored in tawny and gray, respectively.

(B) Detail of conserved PolB motifs. Conserved residues D59/E61 and D368 in *E. coli* 373 piPolB that were mutated to generate exonuclease- and polymerase-deficient variants, respectively, are underlined.

(C) Maximum likelihood phylogeny of piPolBs. The tree is rooted with diverse cellular and viral RNA-primed PolBs (red) and other protein-primed PolBs are also shown (gray). The two groups of piPolBs are colored with light and dark shades of blue (see text for details). The scale bar represents the number of substitutions per site. Branches with support values below 70% were collapsed. The tree in which piPolB clades are expanded is shown in Figure S1.

(D) Genomic organization of representative pipolins (see Figure S2 and Table S2 for details).

own primases (Kazlauskas et al., 2016). By contrast, none of the pipolins possesses genes for recognizable primases, raising questions regarding the priming mechanism.

Collectively, results of the phylogenetic and comparative genomic analyses underscore the uniqueness of piPolBs and pipolins, which may be considered as the third major superfamily of self-replicating MGEs, next to polintons and casposons.

Pipolin DNAP Is a Proficient Replicase

To verify whether piPolBs were indeed active DNAPs, we chose a representative enzyme from *E. coli* 3-373-03-S1_C2 pipolin, and we purified its recombinant form. We first analyzed the synthetic and degradative activities of this protein in a primer extension assay (Figure 2A, lanes 4–9). As expected, only degradation products could be detected in the absence of dinucleotide triphosphates (dNTPs). However, the addition of dNTPs resulted in a switch from exonucleolysis to polymerization activity, indicating that both activities were coordinated. Protein variants with deficient polymerization (D368A; Figure 2A, lanes 10 and 11; Figure S3A) or exonuclease (D59A/E61A; Figure S3A) activities confirmed that 5′-3′ synthetic and 3′-5′ degradative capacities were intrinsic to the recombinant purified piPolB. The presence of proficient DNA polymerization activity in piPolB confirmed that only the second carboxylate moiety in the PolC motif is

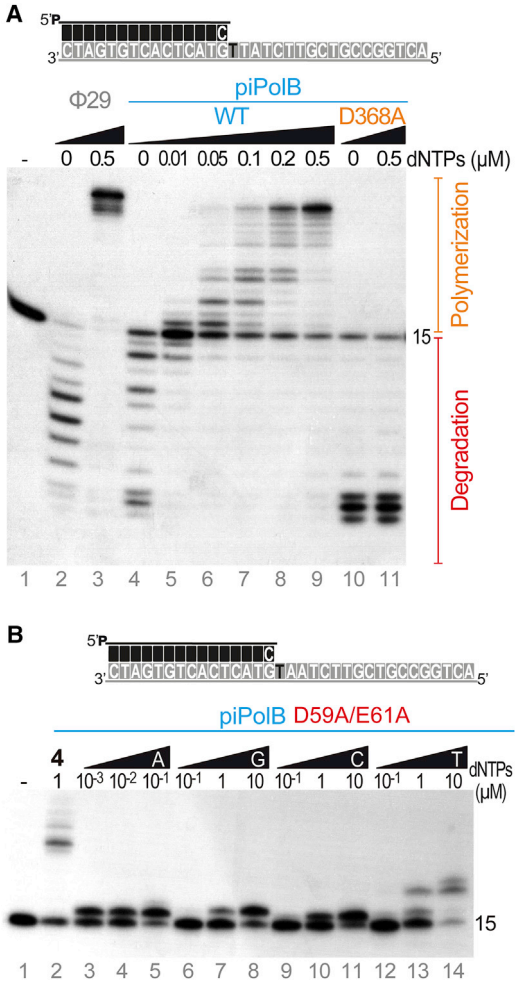


Figure 2. Recombinant piPolB from *E. coli* 3-373-03_S1_C2 Pipolin Is an Active and Faithful DNAP with Intrinsic Proofreading Activity

(A) Primer extension assays with an oligonucleotide template/primer duplex substrate as depicted above the gel. Reactions were incubated for 10 min at 30°C in the presence of either wild-type or D368A polymerase-deficient variant of piPolB and the indicated amount of dNTPs and triggered with 10 mM $MgCl_2$.

(B) Nucleotide insertion preference by the D59A/E61A exonuclease-deficient piPolB variant in the presence of increasing amounts of each dNTP as indicated.

required for metal coordination, in agreement with the previous suggestions (Brautigam and Steitz, 1998; Wang et al., 1997).

We next analyzed the insertion preference for Watson-Crick base pairs using the piPolB exonuclease-deficient variant D59A/E61A. As shown in Figure 2B, insertion of the correct nucleotide could be detected at approximately 1,000-fold lower dNTP concentration compared with the incorrect dNTP. These results confirm that piPolB of pipolins is an efficient and faithful DNAP.

PiPolB Is Endowed with Intrinsic Translesion Synthesis across DNA Containing Non-bulky Nucleotide Analogs

Abasic (AP) sites constitute the most common DNA lesion that may arise from spontaneous depurination, but they also occur as intermediates in base excision repair. A prevailing model is that high-fidelity replicative DNAPs are unable to replicate through such lesions, leading to stalled replication and subsequent triggering of DNA damage tolerance mechanisms, involving specialized DNAPs that can bypass the DNA damage by translesion synthesis (TLS) (Broyde et al., 2008; Vaisman and Woodgate, 2017). However, recent works reported examples of TLS by cellular and viral replicases from families A, B, or C during processive genome replication (Berjón-Otero et al., 2015; Nevin et al., 2017; Sun et al., 2015). As shown in Figure 3A, piPolB was also able to insert the first nucleotide and extend the primer beyond a tetrahydrofuran (THF) moiety, a stable analog of an abasic site (lines 4–9), whereas Φ29DNAP only gave rise to negligible replication, as expected (Berjón-Otero et al., 2015) (lines 2 and 3). The bypass capacity often depends on the sequence context and is counteracted by the proofreading activity (Berjón-Otero et al., 2015; Choi et al., 2010; Tanguy Le Gac et al., 2004; Zhu et al., 2008). However, piPolB TLS capacity did not seem to be affected by the template sequence context (Figure S3A). A minor band of partial product at the lesion site (16-mer) could be detected, suggesting that elongation of the primer beyond the abasic site was a limiting step in the TLS by piPolB, despite the fact that replication of both undamaged and damaged oligonucleotide templates could be processive (Figure S3B).

We then analyzed the incorporation preference opposite to the THF site. Using the exonuclease-deficient variant D59A/E61A, we found that piPolB preferentially inserted purines over pyrimidines (Figure 3B, lanes 3–6), in the preference order A > G > T > C, in agreement with the so-called "A rule" previously described for many DNAPs (Strauss, 2002). The TLS by DNAPs may occur via a misalignment mechanism, resulting in a 1- or 2-nt deletion and, accordingly, a shorter DNA product (Laverty et al., 2017; Yang, 2014). Thus, we monitored step-by-step polymerization in a primer extension assay in the presence of different dNTP combinations (Figure 3B, lanes 7–11). In particular, we provided deoxyadenosine triphosphate (dATP) in combination with another single dNTP (deoxythymidine triphosphate [dTTP], deoxycytidine triphosphate [dCTP], and deoxyguanosine triphosphate [dGTP], lanes 7–9, respectively). Whereas insertion opposite to the THF was detected in all cases, only the combination dATP and dTTP (AT, lane 7) allowed primer extension beyond the abasic site, giving rise to a product that corresponded with the 19-mer marker (Figure 3B, lane M), indicating accurate replication (see substrate scheme above the gel). Consistently, the presence of dATP, dTTP, and dGTP (ATG, lane 10) allowed the copy of the template up to the 22-mer product length, and only when the four dNTPs were provided could the full-length replication product be detected (lane 11). Taken together, these results indicate that TLS capacity of piPolB preferably inserts an A opposite to the abasic sites and subsequently elongates the primer processively without introducing frameshift mutations.

We next analyzed the abasic site bypass with different metal cofactors and replication-blocking DNA damages. As shown in Figure 3C, abasic site TLS in the presence of manganese ions was more efficient at lower dNTP concentrations than with magnesium (lanes 19–22 versus 8–11), in agreement with previous

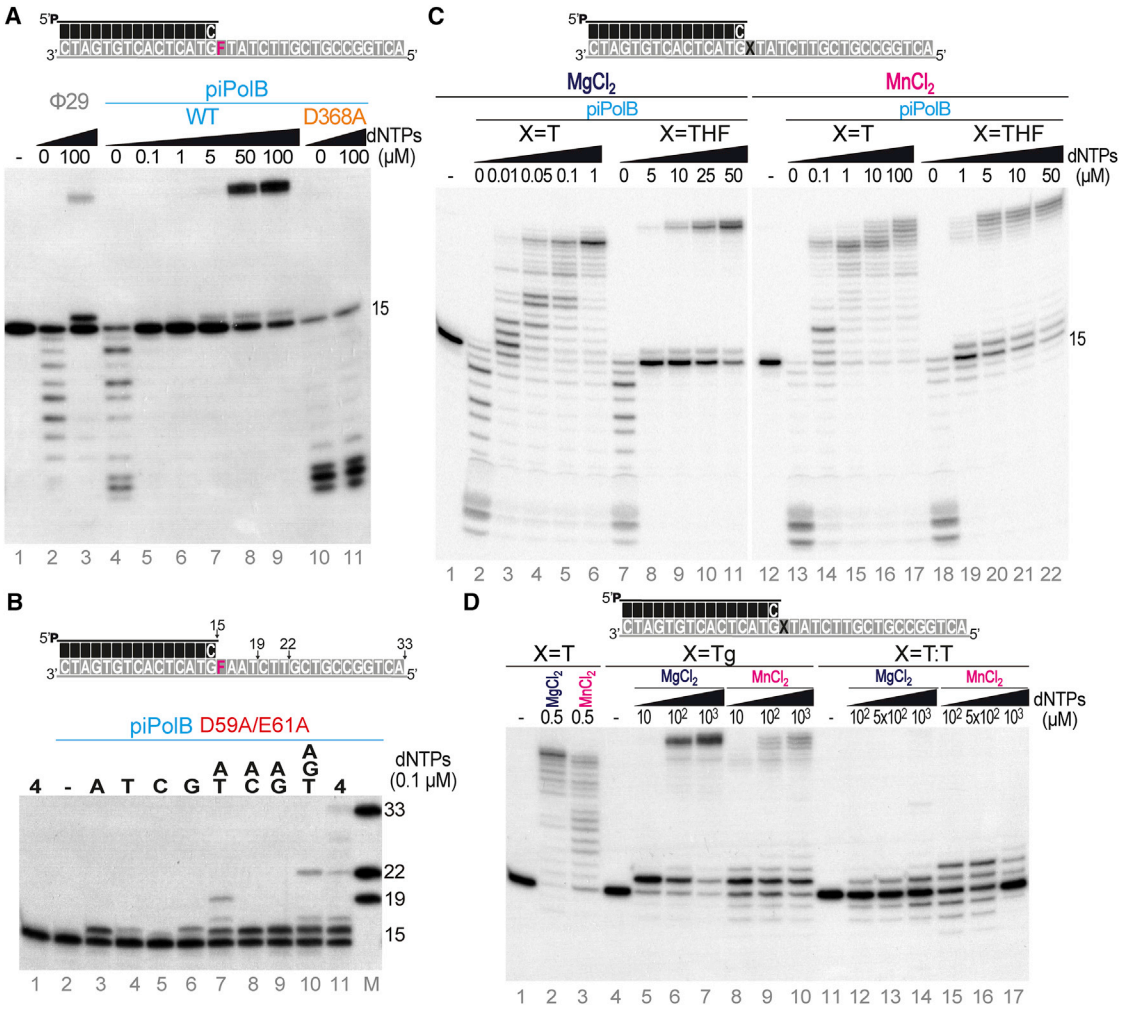


Figure 3. Characterization of piPolB TLS Capacity

(A) Primer extension experiment opposite abasic site-containing template. "F" stands for THF abasic site analog.

(B) Step-by-step monitoring of piPolB replication of THF-containing template by the sequential addition of dNTPs (0.1 μM).

(C) Effect of divalent metal cofactors on piPolB polymerization capacity on undamaged and damaged templates. Reactions were triggered either with 1 mM $MnCl_2$ or 10 mM $MgCl_2$, as indicated.

(D) TLS capacity of piPolB on alternative DNA-damaged templates.

reports on other PolBs (Tanguy Le Gac et al., 2004; Villani et al., 2002). We noted that replication of undamaged template required a higher dNTP concentration in the presence of manganese ions (lanes 14–17) when compared with the magnesium-triggered reactions (lanes 3–6). We also explored the template specificity of piPolB TLS capacity with substrates containing thymine-glycol (Tg)-oxidized base and cyclobutane thymine dimers (T:T). Our results indicated that piPolB was able to bypass Tg in the presence of magnesium ions (Figure 3D, lanes 5–7). However, primer extension beyond the damage was less efficient, since the 16-mer pause was stronger than in the case of the THF-containing template (lanes 9–11 in Figure 3C versus lanes 5–7 in Figure 3D). In line with the impairment in processive primer extension beyond the damage, manganese ions apparently did not stimulate the TLS. On the contrary, Tg bypass was reduced in the presence of this metal cofactor (lanes 9 and 10). In the case of T:T, insertion of only 1 or 2 nt opposite to the damage could be detected (lanes 12–17). In conclusion, piPolB has an efficient TLS capacity that allows it to bypass abasic sites and oxidative base modifications, but it is unable to overcome bulkier modifications such as T:T, likely because this damage induces major structural changes in the DNA helix that strongly obstruct DNA replication.

Primer-Independent DNA Replication

Due to its high processivity, coupled with strand displacement capacity, Φ29DNAP was able to synthesize very large single-stranded DNA (ssDNA) fragments in singly primed M13 DNA rolling circle replication assays (Salas and de Vega, 2016) (Figure 4A, lane 1). By contrast, piPolB gave rise to a smeared signal of replication products spanning 0.5–10 kb, with an apparent peak at ~3 kb (Figure 4A, lanes 3–6), which indicates that piPolB

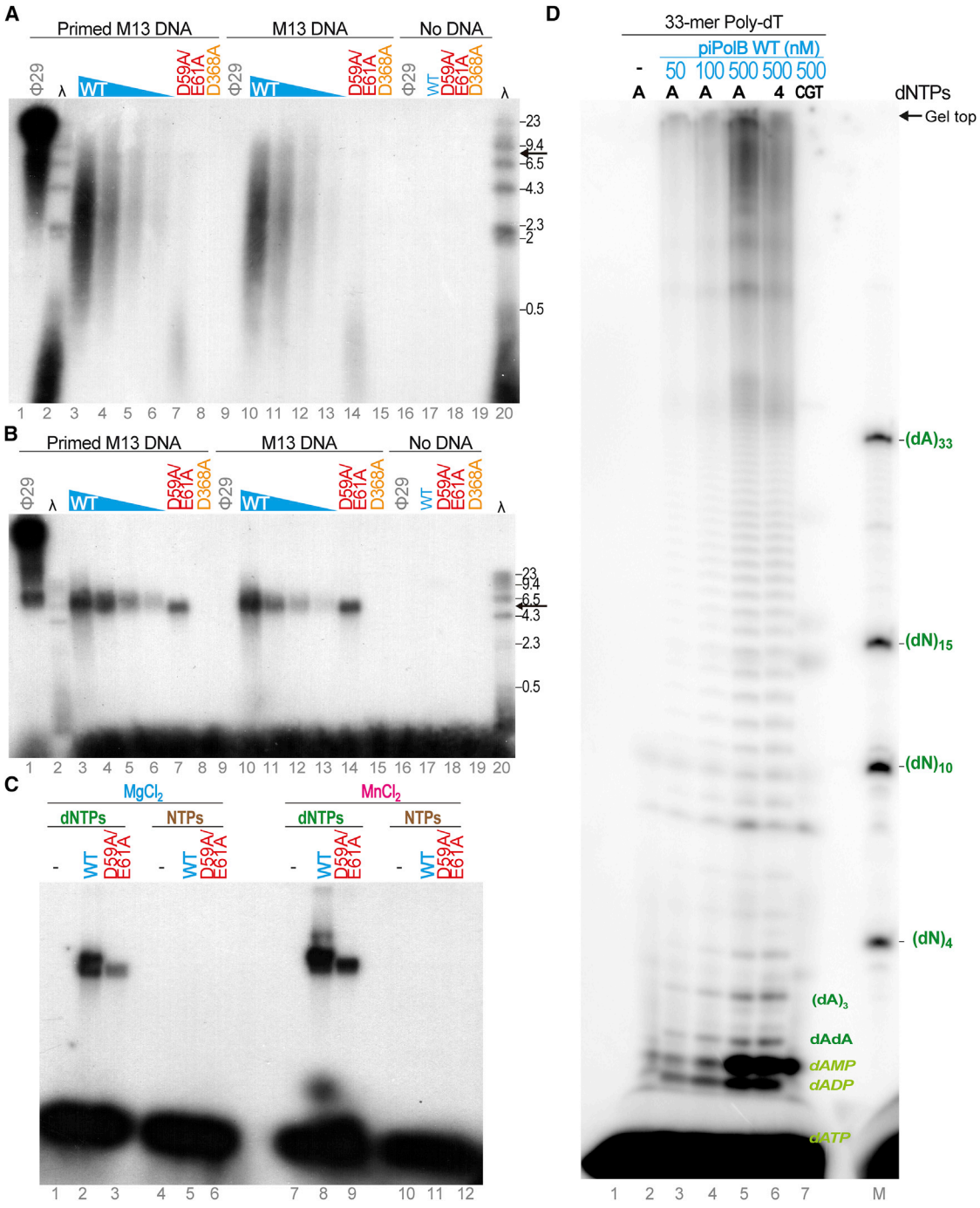


Figure 4. Primer-Independent DNA Synthesis by piPolB

(A and B) Alkaline (A) and non-denaturing TAE (B) agarose electrophoresis of primed (lanes 1 and 3–8) or not-primed (lanes 9–15) M13 ssDNA replication products. Wild-type and D368A and D59A/E61A variants of piPolB were assayed at 500 nM or, when indicated, decreasing concentrations of wild-type piPolB (500, 250, 100, and 50 nM, lanes 3–6 and 10–13). See the Experimental Procedures for details.

(C) Non-denaturing TAE agarose electrophoresis of M13 replication products in the presence of either 100 μM dNTPs (lanes 1–3 and 7–9) or NTPs (lanes 4–6 and 10–12), as indicated. Replication assays were carried out with wild-type or D59A/E61A piPolB variants and triggered with either 10 mM $MgCl_2$ (lanes 1–6) or 1 mM $MnCl_2$ (lanes 7–12).

(D) Primer synthesis and replication of homopolymeric poly-dT DNA template (1 μM) by piPolB. Reactions were triggered with 1 mM $MnCl_2$ and resolved in high-resolution 8 M urea-20% PAGE.

was not as processive as Φ29DNAP. However, the maximal product length obtained with piPolB remained similar even at 20-fold lower enzyme concentration, suggesting that it is a processive DNA replicase. A considerable portion of replication products was larger than the M13 DNA, suggesting that piPolB DNA replication is coupled with strand displacement. The latter activity was subsequently confirmed using an oligonucleotide template/primer substrate with a 5-nt gap (Figure S4).

Strikingly, a very similar replication pattern was detected regardless of whether the M13 was primed or not (Figure 4A, lanes 10–13). By contrast, as expected, Φ29DNAP was unable to synthesize any product in the absence of a primer (lane 9). When the same samples were loaded on a non-denaturing agarose gel (Figure 4B), the replication product appeared as a single band that corresponded to the expected M13 unit length, suggesting that the ssDNA products detected in alkaline-denaturing electrophoresis gel are M13 replication products. Consistently, no product could be detected either with the D368A variant deficient for polymerization activity (lanes 8 and 15) or in the absence of input DNA template (lanes 16–19). Notably, the fragments detected with the D59A/E61A mutant were slightly smaller (by <0.5 kb) than with the wild-type enzyme (Figure 4A, lanes 7 and 14), presumably because exonuclease deficiency gives rise to the accumulation of replication mistakes that may result in the impairment of strand displacement or processivity (Soengas et al., 1992). These results further confirmed that M13 DNA replication, with or without the added primer, was intrinsic to piPolB. *De novo* DNA synthesis on non-primed M13 DNA could be detected using both, magnesium or manganese ions, as cofactors (Figure 4C, lanes 2 and 3 and 8 and 9), albeit with a somewhat higher intensity of total replication product with manganese ions. However, replication was not detected when deoxyribonucleotides were substituted with ribonucleotides (lanes 5 and 6 and 11 and 12), as expected for a PolB that contains the conserved tyrosine steric gate (Bonnín et al., 1999; Brown and Suo, 2011) (Pol B motif in Figure 1B). Smaller DNA fragments were detected with the wild-type polymerase that might be products of the exonucleolytic degradation (lane 8).

To investigate a possible sequence requirement for *de novo* initiation of DNA replication, we performed assays using a single-stranded homopolymeric poly-dT 33-mer as a template. As shown in Figure 4D, DNA replication in the presence of the complementary dATP gave rise to large DNA products and a laddered pattern, indicating that replication started *de novo*, with the synthesis of short primers. This laddered pattern could correspond to either a distributive replication or alternative initiation positions throughout the template. Replication products obtained using the exonuclease-deficient piPolB were overall shorter, suggesting a processivity impairment, as found in the case of M13 replication (Figure S5, lanes 8–10 versus 11–13). Using this short, homopolymeric substrate, DNA primer synthesis was negligible with magnesium ions (Figure S5, lanes 2–7 versus 8–13), underlining the higher efficiency of manganese as a cofactor for DNA priming. Interestingly, when all dNTPs were added, generation of large DNA products was somewhat reduced (Figure 4D, lane 6), and, if dATP was reduced to the labeled nucleotide (16 nM compared with 100 μM of the non-labeled, lane 7), replication products were negligible, which suggested that formation of correct Watson-Crick base pairs was required for replication initiation.

Collectively, these results indicate that piPolB from *E. coli* 3-373-03_S1_C2 pipolin is able to initiate and perform DNA replication of circular and linear templates in the absence of pre-existing primers or additional protein factors. Furthermore, replication of homopolymeric DNA substrates suggests that, contrary to canonical DNA primases (Frick and Richardson, 2001; García-Gómez et al., 2013), piPolB DNA-priming capacity does not rely on a specific template sequence.

De Novo Synthesis of DNA Primers

To further confirm that piPolB is able to synthesize DNA *de novo*, we performed M13 ssDNA replication using γ^{32}P-ATP as a labeled nucleotide. Thus, only newly synthesized DNA fragments would be detected. As shown in Figure 5A, small DNA fragments (up to 4–5 nt in length) were generated in a distributive manner by wild-type and exonuclease-deficient piPolBs, but not by the D368A variant. Again, this reaction was considerably more efficient in the presence of manganese ions than with magnesium ions (lanes 1–8 versus 9–16). Furthermore, the products were only detected in the presence of dNTPs, but not with NTPs (not shown). Instead of the large DNA fragments detected in the assays described above (Figure 4A), we observed only di- and trinucleotide primers, which may be abortive initiation products resulting from the incorporation of a ribonucleotide (rather than dNTP) as a terminal 5′ nucleotide.

The use of high-resolution PAGE allowed us to identify alternative di- and trinucleotide primers with similar intensity, suggesting that DNA synthesis initiation by piPolB does not require a specific template sequence. In line with this, when each dNTP was provided separately (Figure 5B), the reaction was clearly stimulated by dGTP, in the presence of either magnesium or manganese ions (lanes 4 and 14) and, to a lesser extent, by dCTP and dTTP, either alone or in combination with other deoxyribonucleotides. The fact that A-dG dinucleotide was the most efficiently synthetized initiation product is in agreement with the observation that pyrimidines are the preferential template substrates for the priming reaction by most DNA primases (Frick and Richardson, 2001). In line with these results, single-nucleotide changes in the poly-dT homopolymer substrate did not substantially change the efficiency of *de novo* DNA synthesis (Figure S6), although short di- and trinucleotides could be detected when one or two Cs were included in the template sequence, even at the 5′ end of the template molecule (lanes 7 and 12). Taken together, these results demonstrate that piPolB was able to initiate de novo DNA primer synthesis without a strong requirement for specific template sequence.

An Invariable Lysine Plays a Role in TLS and Primer Synthesis Activities

PolBs contain a conserved KxY motif within a β strand in the palm domain involved in stabilization of the primer terminus (Berman et al., 2007; Blasco et al., 1995). We hypothesized that structural adaptations of this motif or nearby residues would be required for stable binding of a nucleoside triphosphate at the 5′ side of the nascent primer to allow dinucleotide formation.

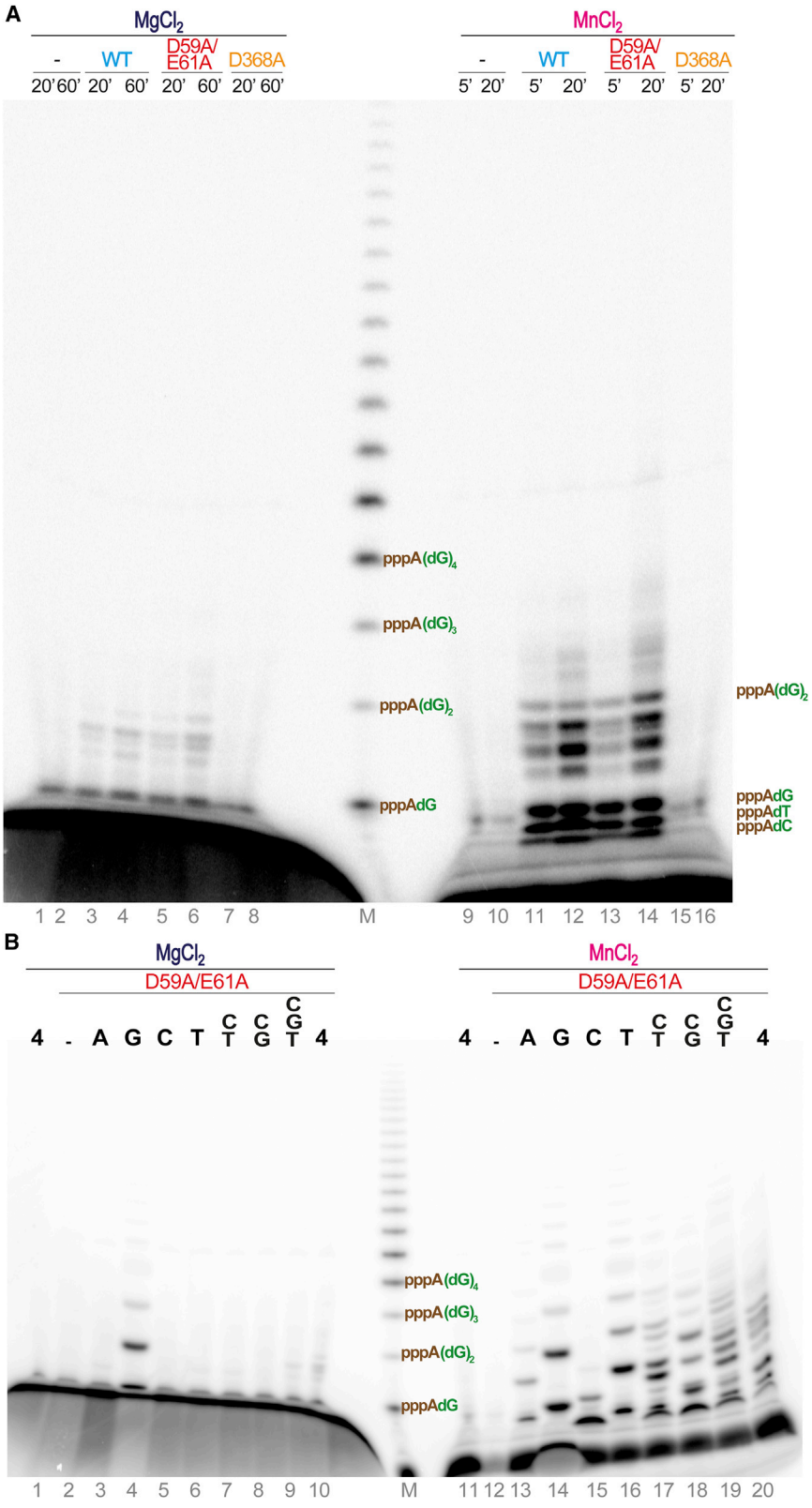


Figure 5. *De Novo* DNA Synthesis by piPolB

(A) Primer synthesis by piPolB. M13 DNA was incubated with either dNTPs or NTPs and wild-type (WT) piPolB or the polymerase- (D368A) or exonuclease- (D59A/E61A) deficient variants. Detected products are labeled with [γ^{32}P]ATP (1 μCi) that only could be incorporated in the 5′ position of the newly synthetized primers. A [γ^{32}P]ATP-(dGMP)n ladder was used a size marker (lane M).

(B) Insertion preference for the first steps of DNA primer synthesis by exonuclease-deficient piPolB. The assay was performed as in (A) but with each dNTP provided independently or in the indicated combinations. Reactions were triggered with either 10 mM $MgCl_2$ or 1 mM $MnCl_2$ and resolved in high-resolution 8 M urea-20% PAGE.

Figure 6. The piPolBs Invariant K613 Residue Plays a Role in TLS and *De Novo* Primer Synthesis

(A) Details of a sequence alignment of piPolBs showing a predicted palm β sheet containing the conserved KxY motif. For reference, RB69, Bam35, and Φ29 DNAPs were included. Significantly different sequences are highlighted in magenta, and the predicted secondary structures of these sequence are in blue and red for β sheet and α helix, respectively. The PolB KxY and the piPolB counterpart KTRG motifs are boxed in green, whereas the KH motif is boxed in orange. The mutated K613 and H614 residues in the *E. coli* 3-373-03_S1_C2 pipolin piPolB (GI 693097161) are underlined.

(B) Primer extension assays of wild-type, K613A, and H614A His-tagged piPolBs. Assays were carried out for 10 min at 30°C in the presence of 1 nM primer/template duplex, 10 μM dNTPs, and the indicated concentration of piPolB variants.

(C) Primer synthesis by wild-type, K613A, and H614A His-tagged piPolBs.

(D) Comparison of primer extension capacity of wild-type and K613A His-tagged piPolBs opposite to undamaged (X = T) or damaged (X = THF) templates. Reactions were triggered with 1 mM $MnCl_2$ and resolved in 8 M urea-20% PAGE.

Indeed, analysis of the multiple sequence alignment showed that the piPolBs lack the canonical KxY motif (Figure 1B) and instead contain an alternative conserved sequence KTRG (Figure 6A). An additional KH pattern within an N-terminal extension of the same β strand is also highly conserved in piPolB homologs, defining an extended KH-X_8-KTRG motif. We generated alanine variants of K613, H614, K623, and R625 residues. In agreement with a putative role in primer terminus stabilization, K623A and R625A variants had impaired primer extension capacity (Figure S7A) and primer synthesis beyond the dinucleotide formation (Figure S7B). On the other hand, K613A and H614A proteins had normal primer extension capacity under the tested conditions (Figure 6B). However, whereas H614A was able to synthetize new primers with a similar pattern as the wild-type piPolB (Figure 6C, lanes 6 and 7), K613A priming capacity was strongly reduced (lanes 4 and 5), suggesting a specific role of this residue during the *de novo* DNA synthesis.

Furthermore, the TLS capacity of K613A protein was also strongly impaired compared to the wild-type piPolB (Figure 6D, lanes 3 and 4 versus 7 and 8). Thus, although DNA primer synthesis and primer extension opposite to the undamaged and damaged substrates appeared to rely on the same conserved catalytic residues, as shown for the D368A variant (see above), we were able to partially uncouple these activities. This result further confirmed the unique intrinsic TLS and DNA primase capacities of piPolB, and also it unveils the role of the extended primer stabilization motif of the piPolB group that would be required for these activities.

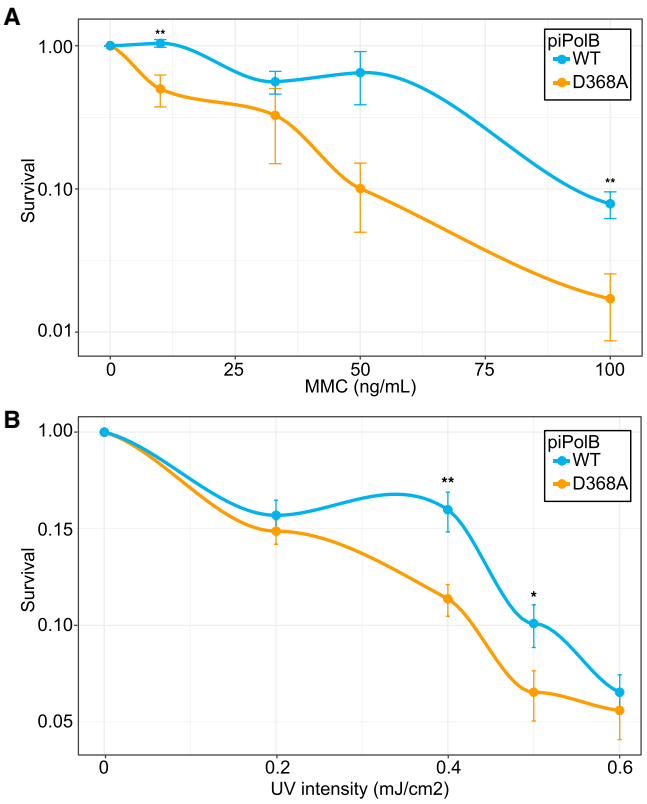


Figure 7. Survival of *E. coli* (DE3) Cells Expressing Wild-Type of Inactive D368A piPolB Variants upon DNA Damage Challenges

(A and B) The graphs show relative survival (mean and SE of four independent experiments) of cells overexpressing wild-type of DNA polymerization-deficient piPolB variants after genotoxic challenge with MMC (A) or UV irradiation (B). The p values are indicated as *p $\leq$ 0.1, **p $\leq$ 0.05, and ***p $\leq$ 0.01. See the Supplemental Experimental Procedures for details.

Biological Role of piPolB in *De Novo* DNA Synthesis

Considering the DNA-priming capacity of piPolB, which is unprecedented in PolB family enzymes, we decided to investigate its biological role *in vivo*. To this end, we challenged the piPolB-expressing bacteria with Mitomycin C (MMC) and UV irradiation, the two DNA-damaging agents known to block DNA replication by introducing bulky base modifications and interstrand cross-links. Since piPolB was unable to replicate a T:T-containing template (Figure 3D), it is unlikely that its TLS capacity may allow bypass of DNA damage induced by MMC treatment or UV irradiation. However, given that replication blockage on the leading strand can be circumvented by re-priming events downstream of the UV-generated lesions (Heller and Marians, 2006), we hypothesized that the *de novo* DNA synthesis by piPolB might contribute to relieving the genotoxic stress generated by DNA-damaging agents. Our results suggested that this is indeed the case, since expression of the wild-type piPolB in *E. coli* Bl21(DE3) cultures significantly enhanced cell survival upon both MMC treatment and UV irradiation, as compared with bacteria expressing D368A inactive piPolB variant (Figure 7). These results indicate a possible role of piPolB in DNA damage tolerance or repair in the context of *E. coli* cells.

DISCUSSION

Here we report the discovery and biochemical characterization of a previously overlooked major group of replicative PolBs, which we named piPolB due to their unique capacity to perform primer-independent, templated DNA synthesis. Within the global PolB phylogeny, piPolB forms a distinct, ancient clade on par with the two previously described groups, rPolB and pPolB. The piPolB-encoding genes are found in MGEs, dubbed pipolins, most of which are integrated into genomes of bacteria from phyla Firmicutes, Actinobacteria, and Proteobacteria, but also replicating as circular plasmids in mitochondria. The distribution of pipolins is rather patchy, which is typical of integrated MGEs (Forterre, 2012; Makarova et al., 2014). To a large extent, pipolins seem to have co-evolved with their hosts, because piPolB-based phylogeny is congruent with the general bacterial taxonomy. Notably, phylogenetic analysis showed that piPolBs from mitochondrial plasmids cluster with alphaproteobacterial homologs (Figure S1). Given that in all likelihood mitochondria have evolved from an alphaproteobacterial ancestor at the onset of eukaryogenesis (Gray, 2012), it is tempting to speculate that piPolBs were introduced into eukaryotes along with the proto-mitochondrial alphaproteobacterial endosymbiont. According to conservative estimates based on the microfossil record, eukaryotes emerged ~2 billion years ago (Dyall et al., 2004; López-García and Moreira, 2015). Thus, the piPolB clade should be at least as old if not older, especially if the emergence of pipolins predated the divergence of the major bacterial phyla.

The piPolBs share the conserved active site as well as the TPR1 and TPR2 subdomains with pPolBs (Figure 1). Consistently, we showed that piPolB displays efficient DNA polymerization and strand displacement activities. Furthermore, piPolB also showed intrinsic TLS capacity across non-bulky base damages (Figure 3). Strikingly, unlike all other PolBs, piPolB does not require an externally provided primer for DNA replication. Conversely, we found that piPolB is able to initiate DNA synthesis *de novo*, a capacity so far exclusive to DNA primases. In the case of Φ29DNAP, the TPR1 motif makes contacts with the template strand and plays a key role in the interaction with the TP during the early steps of protein-primed replication (Dufour et al., 2000; Kamtekar et al., 2006). Given that piPolBs do not interact with a TP, the function of TPR1 region may be limited to the interaction with the DNA or certain cellular cofactors, which would modulate the piPolB activity *in vivo*.

The use of manganese as divalent cofactor instead of magnesium increased TLS across abasic sites (Figure 3C) as well as *de novo* DNA synthesis (Figures 4C and 5; Figure S5). Although the roles of divalent metal ions in DNA synthesis have been controversial, a number of recent findings suggest a physiological role of manganese ions as a cofactor in DNA damage tolerance and repair pathways (Andrade et al., 2009; Cannavo and Cejka, 2014; Kent et al., 2016), as in the case of the human PrimPol (García-Gómez et al., 2013). PrimPol domain-containing DNA primases have been shown to possess multiple enzymatic activities *in vitro*, including primer-dependent and primer-independent DNAP activity, nucleotidyl-transferase, TLS, and even reverse-transcriptase activities (Gill et al., 2014; Guilliam et al., 2015; Iyer et al., 2005; Lipps et al., 2003; Martínez-Jiménez et al.,

2015). Although in certain virus- or plasmid-encoded proteins the PrimPol domain is fused to various helicases and can synthesize large DNA products (Zhu et al., 2017), these enzymes lack the exonuclease domain, and their DNA polymerization on longer templates appears to be mainly distributive. Thus, it is generally considered that the role of PrimPol proteins *in vivo* is largely restricted to the synthesis of short primers, which are extended by the cellular replicative DNAPs (Beck et al., 2010; Gill et al., 2014). By contrast, piPolBs are full-fledged replicative DNAPs endowed with the proofreading and strand displacement capacities. Thus, our results challenge a long-standing dogma in the field, which states that replicative DNAPs are unable to synthesize DNA *de novo*, without a pre-existing primer providing a hydroxyl moiety to anchor the incoming nucleotide (Kornberg and Baker, 1992; Kuchta and Stengel, 2010).

Our finding that one enzyme acts both as a primase and a DNAP poses evolutionary and mechanistic questions. In particular, why do cellular organisms require two enzymes to fulfill *de novo* DNA synthesis? The inability of cellular replicases to perform as primases might be dictated by the challenges associated with the binding of the priming 5′ nucleotide, because the active site pocket of a primase has to accommodate the triphosphate moiety, which is substantially more negatively charged compared to the pre-existing primer typically faced by a DNAP. The maintenance of the unstable, short oligonucleotides on the templates is also a considerable challenge that arguably cannot be tackled by processive replicases (Kuchta and Stengel, 2010). We have shown that piPolBs have a unique KTRG motif, alternative to the conserved KxY motif of PolBs, which interacts with the primer terminus (Blasco et al., 1995). Moreover, an invariant lysine nearby the KTRG motif plays a key role both in TLS and *de novo* primer synthesis (Figure 6). Given the positive charge of this and nearby residues in the piPolB group, it is likely that the extended KH-X_8-KTRG motif may induce a highly stable primer terminus-binding mechanism that may favor the binding of the incoming nucleotide and the subsequent stabilization of the ternary complex, which would result in enhanced polymerization capacity. These results establish a structural liaison between TLS and priming capacities of piPolB that should be further explored in the future by structural and biochemical approaches.

As mentioned above, all DNA primases lack proofreading capacity. This seems advantageous for the efficient synthesis of short-lived Okazaki fragments. Conversely, the 3′-5′ exonuclease-proofreading activity, which is necessary for faithful DNA replication, could hinder the primase capacity. Thus, piPolB synthetic and degradative activities must be highly coordinated to allow efficient primer synthesis and faithful DNA replication. Furthermore, the piPolB exonuclease activity is also compatible with TLS of non-bulky base damages, which, as reported previously for pPolB of bacteriophage Bam35 (Berjón-Otero et al., 2015), does not require template strand misalignment but tolerates damage-containing mismatches during processive DNA synthesis. The TLS capacity might be of particular importance for mitochondrial pipolins, which, similar to mitochondrial genomes, are likely to be frequently exposed to reactive oxygen species regularly produced during normal mitochondrial respiration (Valentine et al., 1998). Previous studies have shown that replication of pCRY1-like pipolins from fungal mitochondria (Gobbi et al., 1997; Li and Nargang, 1993) can be initiated from multiple origins rather than from a fixed origin (Baidyaroy et al., 2012). However, this observation remained unexplained. In light of our current results, such a replication pattern is consistent with the possibility that pCRY1-like pipolins are replicated by their cognate piPolBs in a primer-independent manner. Analogously, the circular episomal form of bacterial pipolins could be replicated by piPolBs from multiple origins.

Replication across bulkier DNA lesions that could not be bypassed by piPolB might benefit from possible downstream re-priming. Replication re-start in UV-exposed *E. coli* chromosome was suggested in the late 60s (Rupp and Howard-Flanders, 1968), and an origin-independent leading strand re-initiation has been demonstrated experimentally (Heller and Marians, 2006). Accordingly, we have shown that expression of the wild-type piPolB promotes survival of *E. coli* cells exposed to replication-blocking DNA-damaging agents (Figure 7). Hence, we hypothesize that piPolB might have evolved to maintain pipolins' DNA by providing faithful and processive *de novo* DNA replication as well as tolerance to DNA damage, which may also increase the fitness of the host bacteria. The latter mechanism resembles the recently proposed roles of human PrimPol in DNA damage tolerance and bypass (Guilliam and Doherty, 2017; Martínez-Jiménez et al., 2015; Mourón et al., 2013), which in the case of piPolB would be provided by an MGE. According to this hypothesis, pipolins may act as bacterial symbionts contributing to maintenance of the host genome upon genotoxic stress.

Importantly, piPolB holds a great promise for developing novel biotechnological applications. For instance, *in vitro* activities of piPolB, namely, strand displacement and faithful, processive DNA polymerization, can be harnessed for efficient primer-independent whole-genome amplification, whereas the TLS can be useful for amplification of damaged or ancient DNA templates. Given that piPolBs do not display strong sequence requirement for replication initiation, replication origins may be selected in a random manner, a property useful for whole-genome amplification. In this framework it is worth mentioning that a recently developed method, dubbed TruePrime, using a combination of an AEP primase, TthPrimPol, and the Φ29DNAP, has been proposed for whole-genome amplification from single cells (Picher et al., 2016). Similarly, piPolB could become a single-enzyme solution to achieve the same goal in single-cell genomic applications. Further structural and functional characterization of piPolBs and the pipolins that encode them will help to understand the details of this unique replication mechanism and to harness the potential of these enzymes for versatile biotechnological applications.

EXPERIMENTAL PROCEDURES

Primer Extension Assays

Assays were performed in 20 μL final volume containing 50 mM Tris-HCl (pH 7.5), 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/mL BSA, 0.05% (v/v) Tween 20, and, unless otherwise stated, 1 nM of the indicated 5′-labeled primer/template duplex, 10 nM DNAP, and the indicated dNTP concentration. Reactions were triggered by the addition of either 10 mM $MgCl_2$ or 1 mM $MnCl_2$, as indicated, and, after incubation for the indicated times at 30°C, the reactions were stopped by adding 10 μL formamide loading buffer (98% formamide, 20 mM EDTA,

0.5% [w/v] bromophenol blue, and 0.5% [w/v] xylene cyanol). Samples were analyzed by 8 M urea-20% PAGE (20 × 30 × 0.5 mm) in 1× Tris-borate-EDTA (TBE) buffer. Gel bands were detected either by autoradiography or phosphorimages (Typhoon FLA 7000) and processed with ImageJ software.

Replication of ssDNA

The reaction mixture contained, in a final volume of 25 μL, 50 mM Tris-HCl (pH 7.5), 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/mL BSA, 0.05% (w/v) Tween 20, 20 mM ammonium sulfate, 100 μM dNTPs, 0.5 μCi [α-^{32}P]dATP, 3.2 nM primed or non-primed M13mp18 ssDNA, and the indicated concentrations of each DNAP. Reactions were triggered by the addition of either 10 mM $MgCl_2$ or 1 mM $MnCl_2$ and incubated for 20 min at 30°C. Reactions were then quenched by adding 5 μL 250 mM EDTA and 5% (w/v) SDS, and they were directly loaded in Tris-Acetate-EDTA buffer (40 mM tris, 20 mM acetic acid, and 1 mM EDTA, pH 8.0) (TAE)1x non-denaturing agarose electrophoresis. For alkaline agarose electrophoresis, an aliquot (15 μL) was subjected to gel filtration through Sephadex G-15 spin columns containing 0.1% (w/v) SDS. Lambda DNA ladder used as a size marker was labeled by filling in with Klenow fragment (New England Biolabs) in the presence of [α-^{32}P]dATP (Sambrook and Russell, 2001).

Replication of homopolymeric single-stranded oligonucleotides is described in the Supplemental Experimental Procedures.

Statistical Methods

Data analysis and representation was performed using R and R-Studio (Studio, Boston, MA; http://www.rstudio.com), using packages Dplyr, Stats, and Ggplot2, available from CRAN (the comprehensive R archive network; https://cran.rstudio.com). Based on Shapiro-Wilk normality tests, results were analyzed by either paired t test or Dependent 2-group Wilcoxon signed rank test.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and four tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2017.10.039.

AUTHOR CONTRIBUTIONS

Conceptualization, M.R.-R.; Investigation, M.R.-R., C.D.O., M.B.-O., J.M.-G., C.A.-M., and M.K.; Writing – Original Draft, M.R.-R., M.S., and M.K.; Writing – Review & Editing, M.R.-R., P.F., M.S., and M.K.; Funding Acquisition, M.R.-R., P.F., and M.S.; Resources, M.S.

ACKNOWLEDGMENTS

We thank Laurentino Villar for protein purifications, Dr. Miguel de Vega for valuable suggestions and critical reading of the manuscript, and Professor Luis Blanco for helpful discussions and the [γ-^{32}P]ATP-(dGMP)n ladder. M.S.'s lab is funded by the Spanish Ministry of Economy and Competitiveness (BFU2014-52656P). M.R.-R. was supported by a ComFuturo Grant (NewPols4Biotech) from Fundación General CSIC. C.D.O. was holder of a "Plan de Empleo Juvenil" contract from Madrid Regional Government (funded by YEI program from European Social Fund, EC). An institutional grant from Fundación Ramón Areces to the Centro de Biología Molecular Severo Ochoa is also acknowledged. P.F. was funded by the European Research Council under the European Union's Seventh Framework Program (FP/2007-2013)/Project EVOMOBIL - ERC Grant Agreement 340440.

Received: March 10, 2017
Revised: September 19, 2017
Accepted: October 11, 2017
Published: November 7, 2017

REFERENCES

Andrade, P., Martín, M.J., Juárez, R., López de Saro, F., and Blanco, L. (2009). Limited terminal transferase in human DNA polymerase mu defines the required balance between accuracy and efficiency in NHEJ. Proc. Natl. Acad. Sci. USA *106*, 16203–16208.

Baidyaroy, D., Hausner, G., and Bertrand, H. (2012). In vivo conformation and replication intermediates of circular mitochondrial plasmids in Neurospora and Cryphonectria parasitica. Fungal Biol. *116*, 919–931.

Bath, C., Cukalac, T., Porter, K., and Dyall-Smith, M.L. (2006). His1 and His2 are distantly related, spindle-shaped haloviruses belonging to the novel virus group, Salterprovirus. Virology *350*, 228–239.

Beck, K., Vannini, A., Cramer, P., and Lipps, G. (2010). The archaeo-eukaryotic primase of plasmid pRN1 requires a helix bundle domain for faithful primer synthesis. Nucleic Acids Res. *38*, 6707–6718.

Béguin, P., Charpin, N., Koonin, E.V., Forterre, P., and Krupovic, M. (2016). Casposon integration shows strong target site preference and recapitulates protospacer integration by CRISPR-Cas systems. Nucleic Acids Res. *44*, 10367–10376.

Berjón-Otero, M., Villar, L., de Vega, M., Salas, M., and Redrejo-Rodríguez, M. (2015). DNA polymerase from temperate phage Bam35 is endowed with processive polymerization and abasic sites translesion synthesis capacity. Proc. Natl. Acad. Sci. USA *112*, E3476–E3484.

Berjón-Otero, M., Villar, L., Salas, M., and Redrejo-Rodríguez, M. (2016). Disclosing early steps of protein-primed genome replication of the Gram-positive tectivirus Bam35. Nucleic Acids Res. *44*, 9733–9744.

Berman, A.J., Kamtekar, S., Goodman, J.L., Lázaro, J.M., de Vega, M., Blanco, L., Salas, M., and Steitz, T.A. (2007). Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. EMBO J. *26*, 3494–3505.

Bernad, A., Lázaro, J.M., Salas, M., and Blanco, L. (1990). The highly conserved amino acid sequence motif Tyr-Gly-Asp-Thr-Asp-Ser in alpha-like DNA polymerases is required by phage phi 29 DNA polymerase for protein-primed initiation and polymerization. Proc. Natl. Acad. Sci. USA *87*, 4610–4614.

Blasco, M.A., Méndez, J., Lázaro, J.M., Blanco, L., and Salas, M. (1995). Primer terminus stabilization at the phi 29 DNA polymerase active site. Mutational analysis of conserved motif KXY. J. Biol. Chem. *270*, 2735–2740.

Bonnín, A., Lázaro, J.M., Blanco, L., and Salas, M. (1999). A single tyrosine prevents insertion of ribonucleotides in the eukaryotic-type phi29 DNA polymerase. J. Mol. Biol. *290*, 241–251.

Braithwaite, D.K., and Ito, J. (1993). Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. *21*, 787–802.

Brautigam, C.A., and Steitz, T.A. (1998). Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. Curr. Opin. Struct. Biol. *8*, 54–63.

Brown, J.A., and Suo, Z. (2011). Unlocking the sugar "steric gate" of DNA polymerases. Biochemistry *50*, 1135–1142.

Broyde, S., Wang, L., Rechkoblit, O., Geacintov, N.E., and Patel, D.J. (2008). Lesion processing: high-fidelity versus lesion-bypass DNA polymerases. Trends Biochem. Sci. *33*, 209–219.

Cannavo, E., and Cejka, P. (2014). Sae2 promotes dsDNA endonuclease activity within Mre11-Rad50-Xrs2 to resect DNA breaks. Nature *514*, 122–125.

Choi, J.Y., Lim, S., Kim, E.J., Jo, A., and Guengerich, F.P. (2010). Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and REV1. J. Mol. Biol. *404*, 34–44.

Copeland, W.C., and Wang, T.S. (1993). Mutational analysis of the human DNA polymerase alpha. The most conserved region in alpha-like DNA polymerases is involved in metal-specific catalysis. J. Biol. Chem. *268*, 11028–11040.

Dufour, E., Méndez, J., Lázaro, J.M., de Vega, M., Blanco, L., and Salas, M. (2000). An aspartic acid residue in TPR-1, a specific region of protein-priming DNA polymerases, is required for the functional interaction with primer terminal protein. J. Mol. Biol. *304*, 289–300.

Dyall, S.D., Brown, M.T., and Johnson, P.J. (2004). Ancient invasions: from endosymbionts to organelles. Science *304*, 253–257.

Filée, J., Forterre, P., Sen-Lin, T., and Laurent, J. (2002). Evolution of DNA polymerase families: evidences for multiple gene exchange between cellular and viral proteins. J. Mol. Evol. *54*, 763–773.

Forterre, P. (2012). Darwin's goldmine is still open: variation and selection run the world. Front. Cell. Infect. Microbiol. *2*, 106.

Frick, D.N., and Richardson, C.C. (2001). DNA primases. Annu. Rev. Biochem. *70*, 39–80.

García-Gómez, S., Reyes, A., Martínez-Jiménez, M.I., Chocrón, E.S., Mourón, S., Terrados, G., Powell, C., Salido, E., Méndez, J., Holt, I.J., and Blanco, L. (2013). PrimPol, an archaic primase/polymerase operating in human cells. Mol. Cell *52*, 541–553.

Gill, S., Krupovic, M., Desnoues, N., Béguin, P., Sezonov, G., and Forterre, P. (2014). A highly divergent archaeo-eukaryotic primase from the Thermococcus nautilus plasmid, pTN2. Nucleic Acids Res. *42*, 3707–3719.

Gobbi, E., Carpanelli, A., Firrao, G., and Locci, R. (1997). The Cryphonectria parasitica plasmid pUG1 contains a large ORF with motifs characteristic of family B DNA polymerases. Nucleic Acids Res. *25*, 3275–3280.

Gray, M.W. (2012). Mitochondrial evolution. Cold Spring Harb. Perspect. Biol. *4*, a011403.

Guilliam, T.A., and Doherty, A.J. (2017). PrimPol-Prime Time to Reprime. Genes (Basel) *8*, E20.

Guilliam, T.A., Keen, B.A., Brissett, N.C., and Doherty, A.J. (2015). Primase-polymerases are a functionally diverse superfamily of replication and repair enzymes. Nucleic Acids Res. *43*, 6651–6664.

Heller, R.C., and Marians, K.J. (2006). Replication fork reactivation downstream of a blocked nascent leading strand. Nature *439*, 557–562.

Hoeben, R.C., and Uil, T.G. (2013). Adenovirus DNA replication. Cold Spring Harb. Perspect. Biol. *5*, a013003.

Hutchison, C.A., 3rd, Smith, H.O., Pfannkoch, C., and Venter, J.C. (2005). Cell-free cloning using phi29 DNA polymerase. Proc. Natl. Acad. Sci. USA *102*, 17332–17336.

Iyer, L.M., Koonin, E.V., Leipe, D.D., and Aravind, L. (2005). Origin and evolution of the archaeo-eukaryotic primase superfamily and related palm-domain proteins: structural insights and new members. Nucleic Acids Res. *33*, 3875–3896.

Kamtekar, S., Berman, A.J., Wang, J., Lázaro, J.M., de Vega, M., Blanco, L., Salas, M., and Steitz, T.A. (2006). The phi29 DNA polymerase:protein-primer structure suggests a model for the initiation to elongation transition. EMBO J. *25*, 1335–1343.

Kapitonov, V.V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. Proc. Natl. Acad. Sci. USA *103*, 4540–4545.

Kazlauskas, D., and Venclovas, C. (2011). Computational analysis of DNA replicases in double-stranded DNA viruses: relationship with the genome size. Nucleic Acids Res. *39*, 8291–8305.

Kazlauskas, D., Krupovic, M., and Venclovas, Č. (2016). The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. Nucleic Acids Res. *44*, 4551–4564.

Kent, T., Mateos-Gomez, P.A., Sfeir, A., and Pomerantz, R.T. (2016). Polymerase θ is a robust terminal transferase that oscillates between three different mechanisms during end-joining. eLife *5*, e13740.

Klassen, R., and Meinhardt, F. (2007). Linear protein-primed replicating plasmids in eukaryotic microbes. In Microbial Linear Plasmids (Springer), pp. 187–226.

Koonin, E.V. (2006). Temporal order of evolution of DNA replication systems inferred by comparison of cellular and viral DNA polymerases. Biol. Direct *1*, 39.

Kornberg, A., and Baker, T.A. (1992). DNA Replication (W.H. Freeman).

Krupovic, M., and Koonin, E.V. (2015). Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. Nat. Rev. Microbiol. *13*, 105–115.

Krupovic, M., and Koonin, E.V. (2016). Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems. Curr. Opin. Microbiol. *31*, 25–33.

Krupovic, M., Bamford, D.H., and Koonin, E.V. (2014a). Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. Biol. Direct *9*, 6.

Krupovic, M., Makarova, K.S., Forterre, P., Prangishvili, D., and Koonin, E.V. (2014b). Casposons: a new superfamily of self-synthesizing DNA transposons at the origin of prokaryotic CRISPR-Cas immunity. BMC Biol. *12*, 36.

Krupovic, M., Béguin, P., and Koonin, E.V. (2017). Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. Curr. Opin. Microbiol. *38*, 36–43.

Kuchta, R.D., and Stengel, G. (2010). Mechanism and evolution of DNA primases. Biochim. Biophys. Acta *1804*, 1180–1189.

Laverty, D.J., Averill, A.M., Doublié, S., and Greenberg, M.M. (2017). The A-Rule and Deletion Formation During Abasic and Oxidized Abasic Site Bypass by DNA Polymerase θ. ACS Chem. Biol. *12*, 1584–1592.

Li, Q., and Nargang, F.E. (1993). Two Neurospora mitochondrial plasmids encode DNA polymerases containing motifs characteristic of family B DNA polymerases but lack the sequence Asp-Thr-Asp. Proc. Natl. Acad. Sci. USA *90*, 4299–4303.

Lipps, G., Röther, S., Hart, C., and Krauss, G. (2003). A novel type of replicative enzyme harbouring ATPase, primase and DNA polymerase activity. EMBO J. *22*, 2516–2525.

López-García, P., and Moreira, D. (2015). Open Questions on the Origin of Eukaryotes. Trends Ecol. Evol. *30*, 697–708.

Makarova, K.S., Wolf, Y.I., Forterre, P., Prangishvili, D., Krupovic, M., and Koonin, E.V. (2014). Dark matter in archaeal genomes: a rich source of novel mobile elements, defense systems and secretory complexes. Extremophiles *18*, 877–893.

Martínez-Jiménez, M.I., García-Gómez, S., Bebenek, K., Sastre-Moreno, G., Calvo, P.A., Díaz-Talavera, A., Kunkel, T.A., and Blanco, L. (2015). Alternative solutions and new scenarios for translesion DNA synthesis by human PrimPol. DNA Repair (Amst.) *29*, 127–138.

Mourón, S., Rodriguez-Acebes, S., Martínez-Jiménez, M.I., García-Gómez, S., Chocrón, S., Blanco, L., and Méndez, J. (2013). Repriming of DNA synthesis at stalled replication forks by human PrimPol. Nat. Struct. Mol. Biol. *20*, 1383–1389.

Nevin, P., Gabbai, C.C., and Marians, K.J. (2017). Replisome-mediated translesion synthesis by a cellular replicase. J. Biol. Chem. *292*, 13833–13842.

Peng, X., Basta, T., Häring, M., Garrett, R.A., and Prangishvili, D. (2007). Genome of the Acidianus bottle-shaped virus and insights into the replication and packaging mechanisms. Virology *364*, 237–243.

Picher, A.J., Budeus, B., Wafzig, O., Krüger, C., García-Gómez, S., Martínez-Jiménez, M.I., Díaz-Talavera, A., Weber, D., Blanco, L., and Schneider, A. (2016). TruePrime is a novel method for whole-genome amplification from single cells based on TthPrimPol. Nat. Commun. *7*, 13296.

Rodríguez, I., Lázaro, J.M., Blanco, L., Kamtekar, S., Berman, A.J., Wang, J., Steitz, T.A., Salas, M., and de Vega, M. (2005). A specific subdomain in phi29 DNA polymerase confers both processivity and strand-displacement capacity. Proc. Natl. Acad. Sci. USA *102*, 6407–6412.

Rupp, W.D., and Howard-Flanders, P. (1968). Discontinuities in the DNA synthesized in an excision-defective strain of Escherichia coli following ultraviolet irradiation. J. Mol. Biol. *31*, 291–304.

Salas, M. (1991). Protein-priming of DNA replication. Annu. Rev. Biochem. *60*, 39–71.

Salas, M., and de Vega, M. (2016). Protein-Primed Replication of Bacteriophage Φ29 DNA. In The Enzymes, S.K. Laurie and O. Marcos Túlio, eds. (Academic Press), pp. 137–167.

Salas, M., Holguera, I., Redrejo-Rodríguez, M., and de Vega, M. (2016). DNA-binding proteins essential for protein-primed bacteriophage Φ29 DNA replication. Front. Mol. Biosci. *3*, 37.

Sambrook, J., and Russell, D. (2001). Molecular cloning: a laboratory manual, Fourth Edition (New York: Cold Spring Harbor Laboratory Press).

Schulte, U., and Lambowitz, A.M. (1991). The LaBelle mitochondrial plasmid of Neurospora intermedia encodes a novel DNA polymerase that may be derived from a reverse transcriptase. Mol. Cell. Biol. *11*, 1696–1706.

Sidore, A.M., Lan, F., Lim, S.W., and Abate, A.R. (2016). Enhanced sequencing coverage with digital droplet multiple displacement amplification. Nucleic Acids Res. *44*, e66.

Soengas, M.S., Esteban, J.A., Lázaro, J.M., Bernad, A., Blasco, M.A., Salas, M., and Blanco, L. (1992). Site-directed mutagenesis at the Exo III motif of phi 29 DNA polymerase; overlapping structural domains for the 3′-5′ exonuclease and strand-displacement activities. EMBO J. *11*, 4227–4237.

Strauss, B.S. (2002). The "A" rule revisited: polymerases as determinants of mutational specificity. DNA Repair (Amst.) *1*, 125–135.

Sun, B., Pandey, M., Inman, J.T., Yang, Y., Kashlev, M., Patel, S.S., and Wang, M.D. (2015). T7 replisome directly overcomes DNA damage. Nat. Commun. *6*, 10260.

Tanguy Le Gac, N., Delagoutte, E., Germain, M., and Villani, G. (2004). Inactivation of the 3′-5′ exonuclease of the replicative T4 DNA polymerase allows translesion DNA synthesis at an abasic site. J. Mol. Biol. *336*, 1023–1034.

Vaisman, A., and Woodgate, R. (2017). Translesion DNA polymerases in eukaryotes: what makes them tick? Crit. Rev. Biochem. Mol. Biol. *52*, 274–303.

Valentine, J.S., Wertz, D.L., Lyons, T.J., Liou, L.L., Goto, J.J., and Gralla, E.B. (1998). The dark side of dioxygen biochemistry. Curr. Opin. Chem. Biol. *2*, 253–262.

Villani, G., Tanguy Le Gac, N., Wasungu, L., Burnouf, D., Fuchs, R.P., and Boehmer, P.E. (2002). Effect of manganese on in vitro replication of damaged DNA catalyzed by the herpes simplex virus type-1 DNA polymerase. Nucleic Acids Res. *30*, 3323–3332.

Wang, J., Sattar, A.K., Wang, C.C., Karam, J.D., Konigsberg, W.H., and Steitz, T.A. (1997). Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. Cell *89*, 1087–1099.

Yang, W. (2014). An overview of Y-Family DNA polymerases and a case study of human DNA polymerase η. Biochemistry *53*, 2793–2803.

Zhu, Y., Song, L., Stroud, J., and Parris, D.S. (2008). Mechanisms by which herpes simplex virus DNA polymerase limits translesion synthesis through abasic sites. DNA Repair (Amst.) *7*, 95–107.

Zhu, B., Wang, L., Mitsunobu, H., Lu, X., Hernandez, A.J., Yoshida-Takashima, Y., Nunoura, T., Tabor, S., and Richardson, C.C. (2017). Deep-sea vent phage DNA polymerase specifically initiates DNA synthesis in the absence of primers. Proc. Natl. Acad. Sci. USA *114*, E2310–E2318.

Cell Reports, Volume *21*

Supplemental Information

Primer-Independent DNA Synthesis by a Family B DNA Polymerase from Self-Replicating Mobile Genetic Elements

Modesto Redrejo-Rodríguez, Carlos D. Ordóñez, Mónica Berjón-Otero, Juan Moreno-González, Cristian Aparicio-Maldonado, Patrick Forterre, Margarita Salas, and Mart Krupovic

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Bioinformatic analyses

Phylogenetic analysis. The non-redundant database of protein sequences at the NCBI was searched using the PSI-BLAST (Altschul et al., 1997). For phylogenetic analyses protein sequences were aligned with Promals3D (Pei and Grishin, 2014). Poorly aligned (low information content) positions were removed using the Gappyout function of Trimal (Capella-Gutierrez et al., 2009). The dataset of viral, plasmid and polintons pDNAP sequences was collected previously (Krupovic and Koonin, 2015). Maximum likelihood phylogenetic tree was constructed using the PhyML program (Guindon et al., 2010) the latest version of which (http://www.atgc-montpellier.fr/phyml-sms/) includes automatic selection of the best-fit substitution model for a given alignment. The best model identified by PhyML was LG +G6 +I +F (LG, Le-Gascuel matrix; G6, Gamma shape parameter: fixed, number of categories: 6; I, proportion of invariable sites: fixed; F, equilibrium frequencies: empirical).

Identification and annotation of integrated MGE. The pipolins' were identified by thorough analysis of genomic neighborhoods of the piPolB-encoding genes. The precise boarders of integration were defined based on the presence of direct repeats corresponding to attachment sites. The repeats were searched for using Unipro UGENE (Okonechnikov et al., 2012). Pipolin genes were annotated based on the PSI-BLAST searches (Altschul et al., 1997) against the non-redundant protein database at NCBI and HHpred searches (Soding et al., 2005). Pipolins were compared to each other and visualized using EasyFig (Sullivan et al., 2011).

Protein expression and purification

Primer-independent DNA polymerase (piPolB) from *E. coli* 3-373-03_S1_C2 Pipolin (NCBI GI:693097161) was obtained from GeneScript into NdeI-XhoI sites of pET23a (see Table S4). A stop codon was included to obtain the untagged recombinant protein. Polymerase (D368A) and exonuclease (D59A/E61A) deficient proteins, as well as wild type, K614A, H615A, K624A and R626A his-tagged variants, were obtained by site directed mutagenesis (Table S4).

All piPolB variants were expressed in Bl21(DE3) *E. coli* cells, using ZYM-5052 autoinduction medium (Studier, 2005) in the presence of 100 mg/L ampicillin. Cultures were grown for 20 h at 28 ºC. For purification of untagged piPolB variants, cells were disrupted by grinding with alumina and suspended in buffer A (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 7 mM β-mercaptoethanol, 5% (v/v) glycerol) containing 1 M NaCl. Alumina and cell debris were removed by centrifugation, and absorbance at 260 nm was adjusted to 120 units/ml prior to DNA precipitation with 0.3% (w/v) polyethyleneimine. After centrifugation at 20,000 x g for 20 min, ammonium sulfate was added to the supernatant to 69% saturation and centrifuged at 20,000 x g for 30 min. The piPolB (wild type and mutants) containing pellet was resuspended in buffer A and applied to serial Q Sepharose® fast flow (GE Healthcare) and phosphocellulose (P11, Whatman) columns, at an ionic strength about 0.2 M NaCl. After extensive wash with increasing concentrations of NaCl in buffer A, purified DNA polymerase was eluted with 0.35 M NaCl and applied to Heparin-Sepharose® CL-6B column (GE Healthcare), where, after washing with 0.35, 0.4 and 0.45 M NaCl, they were eluted at 1 M NaCl in buffer A.

Histidine-tagged variants were purified by standard method (Spriestersbach et al., 2015). Briefly, cells were resuspended in buffer C (50 mM phosphate buffer, pH 8, 7 mM β-mercaptoethanol, 5% (v/v) glycerol, 1 M NaCl, 5 mM imidazole) and incubated for 30 min at room temperature with 1 mg/mL lysozyme (Sigma) and 1 unit of benzonase (Sigma), prior to cell disruption by sonication. After centrifugation at 20,000 x g for 30 min, the soluble fraction was applied to a Ni-NTA column (Qiagen). After extensive wash with 5, 10, 25 and 50 mM imidazole, the protein was eluted with 200 mM imidazole and subsequently applied to Heparin-Sepharose® CL-6B column (GE Healthcare), where, after washing with 0.35, 0.4 and 0.45 M NaCl, it was eluted at 1 M NaCl in buffer A.

In all cases, pooled fractions containing pure piPolB variants were dyalized overnight against 500 volumes of buffer B (50 mM Tris-HCl, pH 7.5, 1 mM EDTA, 7 mM β-mercaptoethanol, 0.25 M NaCl and 50% (v/v) glycerol) and kept at -20ºC, or at -70ºC for long storage. Final purity of the proteins was estimated to be >90 % by SDS–PAGE followed by Coomassie blue staining.

DNA substrates

Oligonucleotides (Table S3) were purchased from Sigma or IDT in PAGE purification grade. To form a primer/template substrate as indicated in the top of each figure, the P15 oligonucleotide (Table S3) was 5´-labeled with [γ-^{32}P]ATP using T4 Polynucleotide Kinase and hybridized to 1.2-fold molar excess of

complementary unlabeled template oligonucleotides (T33GTA, T33GTT or T33GFA, Table S3) in the presence of 50 mM NaCl and 50 mM Tris-HCl, pH 7.5.
Genomic M13mp18 single-stranded circular DNA (laboratory stock) was diluted up to 50 ng/μL in a buffer containing 0.2 M NaCl and 60 mM Tris-HCl, pH 7.5 with or without M13 UP primer (Table S3), heated for 5 min at 65 ºC and cooled slowly overnight to allow the annealing of the primer. Primed and non-primed M13 substrates were stored at -20 ºC in small aliquots to minimize DNA nicking due to repetitive cycles of freeze-thaw.

Replication of homopolymeric single stranded oligonucleotides
Assays were made in 20 μL final volume containing 50 mM Tris-HCl, pH 7.5, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/ml BSA, 0.05% (v/v) Tween 20, 1 μM oligonucleotide template, 100 μM dNTPs, 500 nM of the indicated piPolB variant and 0.5 μCi [α-^{32}P]dATP. After incubation for the indicated times at 30 ºC, the reactions were stopped by adding 10 μL of formamide loading buffer (98% formamide, 20 mM EDTA, 0.5% (w/v) bromophenol blue, and 0.5% (w/v) xylene cyanol). Samples were analyzed by 8 M urea-20% polyacrylamide gel electrophoresis (20x30x0.5 mm) in 1X TBE buffer. Gel bands were detected either by autoradiography or phosphorimages (Typhoon FLA 7000) and processed with ImageJ software.

De novo primer synthesis detection
To detect de novo primer synthesis we used [γ-^{32}P]ATP as the labeled nucleotide. M13 ssDNA (3.2 nM) was used as template. The reaction mixture contained, in a final volume of 25 μL, 50 mM Tris-HCl, pH 7.5, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/ml BSA, 0.05% (v/v) Tween 20, 100 μM dNTPs, 0.5 μCi [γ-^{32}P]ATP, the indicated template and DNA polymerase (500 nM). Reactions were triggered by addition of either 10 mM $MgCl_2$ or 1 mM $MnCl_2$ and incubated for indicated times at 30 ºC. Then, the reactions were stopped by adding 10 μL of formamide loading buffer (98% formamide, 20 mM EDTA, 0.5% (w/v) bromophenol blue, and 0.5% (w/v) xylene cyanol). Samples were analyzed by 8 M urea-20% polyacrylamide gel electrophoresis (20x30x0.5 mm) in 1X TBE buffer. When indicated, high resolution gels (40 cm long) were used. Gel bands were detected by phosphorimages (Typhoon FLA 7000) and processed with ImageJ software. The [γ^{32}P]ATP-(dGMP)n DNA ladder used as size marker, generated by human PrimPol (García-Gómez et al., 2013) was a gift from Dr. Luis Blanco (CBMSO, Madrid).

Survival of piPolB expressing bacteria upon DNA damaging agents challenges
Starter cultures of *E. coli* Bl21(DE3) harboring pET23a::piPolB or pET23a::piPolB(D368A) plasmids were inoculates in LB media in the presence of ampicillin (150 μg/mL) and glucose (40 mM) and grown overnight shaking at 37 ºC. Saturated cultures were diluted (1:100) in fresh LB media with ampicillin and grown 1-2 h at 28 ºC until DO600nm=0.4. Recombinant protein expression was then induced by 0.5 mM IPTG during one hour prior to genotoxic challenge. At this point an aliquot was withdrawn to verify recombinant protein expression by SDS-PAGE (not shown).
For MMC treatment, the indicated drug concentration was added directly into the cultures that were grown for an additional hour and then serial-diluted in fresh LB and plate onto LB-agar plates (without antibiotic). In the case of UV-exposure, the induced cultures were serial-diluted in sterile PBS and plated onto LB-agar plates prior to irradiation with the indicated UV-light intensity in a Spectrolinker™ XL-1000 (Spectronics Corporation).

Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Φ29DNAP	M.Salas laboratory stock	Lazaro et al. 1995
Experimental Models: Cell Lines		
E. coli Bl21(DE3)	Stratagene	C6000-03
Recombinant DNA		
pET23a(+)::piPolB	This paper (GeneScript)	See Table S4
pET23a(+)::piPolB derivatives	This paper	See Table S4
Sequence-Based Reagents		
M13mp18 ssDNA	M.Salas laboratory stock	Sambrok and Russel, 2001
Oligonucleotide substrates for DNA replication assays	Sigma/IDT	See Table S3
Software and Algorithms		
PROMALS3D	University of Texas Southwestern Medical Center	Pei et al.NAR 2008 and http://prodata.swmed.edu/promals3d/promals3d.php
Illustrator	Adobe	adobe.com/illustrator

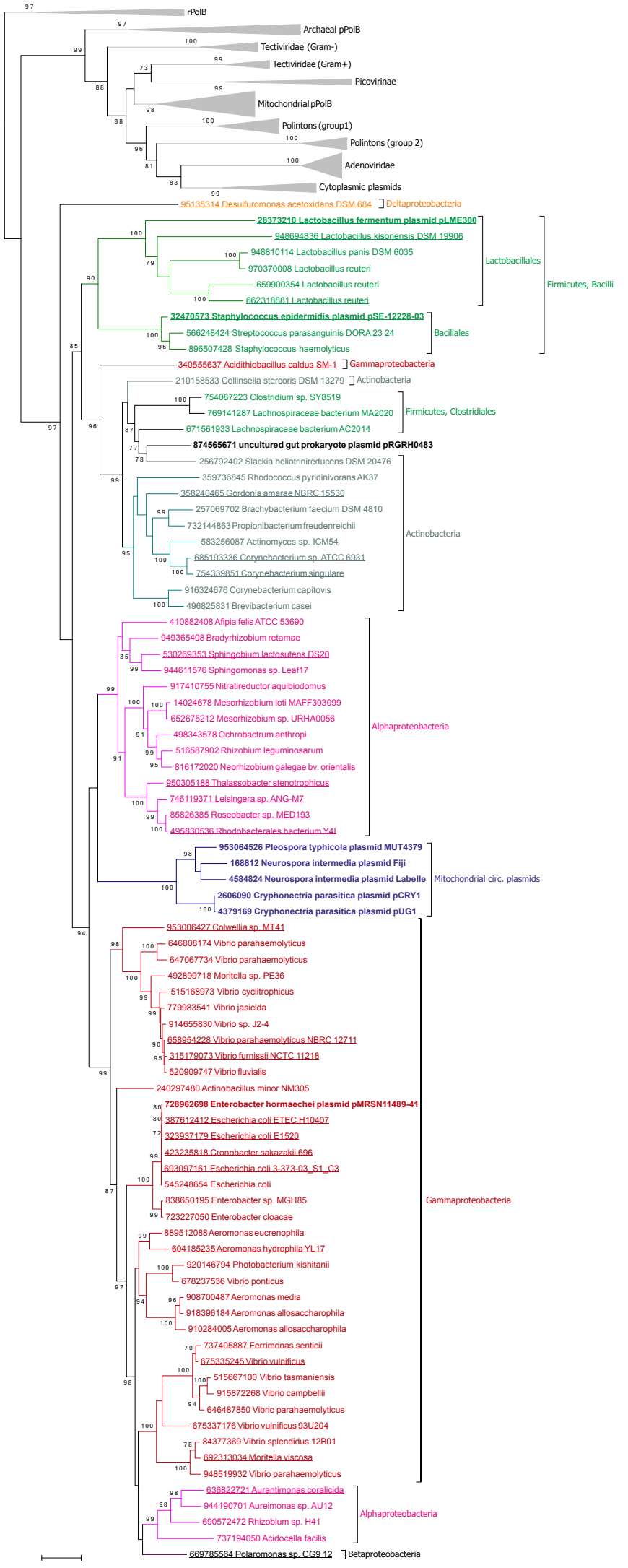

rPolB
Archaeal pPolB
Tectiviridae (Gram-)
Tectiviridae (Gram+)
Picovirinae
Mitochondrial pPolB
Polintons (group1)
Polintons (group 2)
Adenoviridae
Cytoplasmic plasmids
95135314 Desulfuromonas acetoxidans DSM 684
Deltaproteobacteria
28373210 Lactobacillus fermentum plasmid pLME300
948694836 Lactobacillus kisonensis DSM 19906
948810114 Lactobacillus panis DSM 6035
970370008 Lactobacillus reuteri
659900354 Lactobacillus reuteri
662318881 Lactobacillus reuteri
Lactobacillales
32470573 Staphylococcus epidermidis plasmid pSE-12228-03
566248424 Streptococcus parasanguinis DORA 23 24
896507428 Staphylococcus haemolyticus
Bacillales
Firmicutes, Bacilli
340555637 Acidithiobacillus caldus SM-1
Gammaproteobacteria
210158533 Collinsella stercoris DSM 13279
Actinobacteria
754087223 Clostridium sp. SY8519
769141287 Lachnospiraceae bacterium MA2020
671561933 Lachnospiraceae bacterium AC2014
Firmicutes, Clostridiales
874565671 uncultured gut prokaryote plasmid pRGRH0483
256792402 Slackia heliotrinireducens DSM 20476
359736845 Rhodococcus pyridinivorans AK37
358240465 Gordonia amarae NBRC 15530
257069702 Brachybacterium faecium DSM 4810
732144863 Propionibacterium freudenreichii
583256087 Actinomyces sp. ICM54
685193336 Corynebacterium sp. ATCC 6931
754339851 Corynebacterium singulare
916324676 Corynebacterium capitovis
496825831 Brevibacterium casei
Actinobacteria
410882408 Afipia felis ATCC 53690
949365408 Bradyrhizobium retamae
530269353 Sphingobium lactosutens DS20
944611576 Sphingomonas sp. Leaf17
917410755 Nitratireductor aquibiodomus
14024678 Mesorhizobium loti MAFF303099
652675212 Mesorhizobium sp. URHA0056
498343578 Ochrobactrum anthropi
516587902 Rhizobium leguminosarum
816172020 Neorhizobium galegae bv. orientalis
950305188 Thalassobacter stenotrophicus
746119371 Leisingera sp. ANG-M7
85826385 Roseobacter sp. MED193
495830536 Rhodobacterales bacterium Y4I
Alphaproteobacteria
953064526 Pleospora typhicola plasmid MUT4379
168812 Neurospora intermedia plasmid Fiji
4584824 Neurospora intermedia plasmid Labelle
2606090 Cryphonectria parasitica plasmid pCRY1
4379169 Cryphonectria parasitica plasmid pUG1
Mitochondrial circ. plasmids
953006427 Colwellia sp. MT41
646808174 Vibrio parahaemolyticus
647067734 Vibrio parahaemolyticus
492899718 Moritella sp. PE36
515168973 Vibrio cyclitrophicus
779983541 Vibrio jasicida
914655830 Vibrio sp. J2-4
658954228 Vibrio parahaemolyticus NBRC 12711
315179073 Vibrio furnissii NCTC 11218
520909747 Vibrio fluvialis
240297480 Actinobacillus minor NM305
728962698 Enterobacter hormaechei plasmid pMRSN11489-41
387612412 Escherichia coli ETEC H10407
323937179 Escherichia coli E1520
423235818 Cronobacter sakazakii 696
693097161 Escherichia coli 3-373-03_S1_C3
545248654 Escherichia coli
838650195 Enterobacter sp. MGH85
723227050 Enterobacter cloacae
889512088 Aeromonas eucrenophila
604185235 Aeromonas hydrophila YL17
920146794 Photobacterium kishitanii
678237536 Vibrio ponticus
908700487 Aeromonas media
918396184 Aeromonas allosaccharophila
910284005 Aeromonas allosaccharophila
737405887 Ferrimonas senticii
675335245 Vibrio vulnificus
515667100 Vibrio tasmaniensis
915872268 Vibrio campbellii
646487850 Vibrio parahaemolyticus
675337176 Vibrio vulnificus 93U204
84377369 Vibrio splendidus 12B01
692313034 Moritella viscosa
948519932 Vibrio parahaemolyticus
Gammaproteobacteria
636822721 Aurantimonas coralicida
944190701 Aureimonas sp. AU12
690572472 Rhizobium sp. H41
737194050 Acidocella facilis
Alphaproteobacteria
669785564 Polaromonas sp. CG9 12
Betaproteobacteria

Figure S1

Figure S1. Maximum likelihood phylogeny of pipolin-related PolBs. Related to Figure 1.
Clades that are only distantly related to the pipolin-encoded proteins were collapsed. The tree is rooted with cellular and viral RNA-primed PolBs. Branches are colored according to the classification of the corresponding taxa. Pipolins for which precise integration sites were identified (Table S1) are underlined. Numbers at the branch points represent the Bayesian-like transformation of aLRT (aBayes) local support values. The scale bar represents the number of substitutions per site.

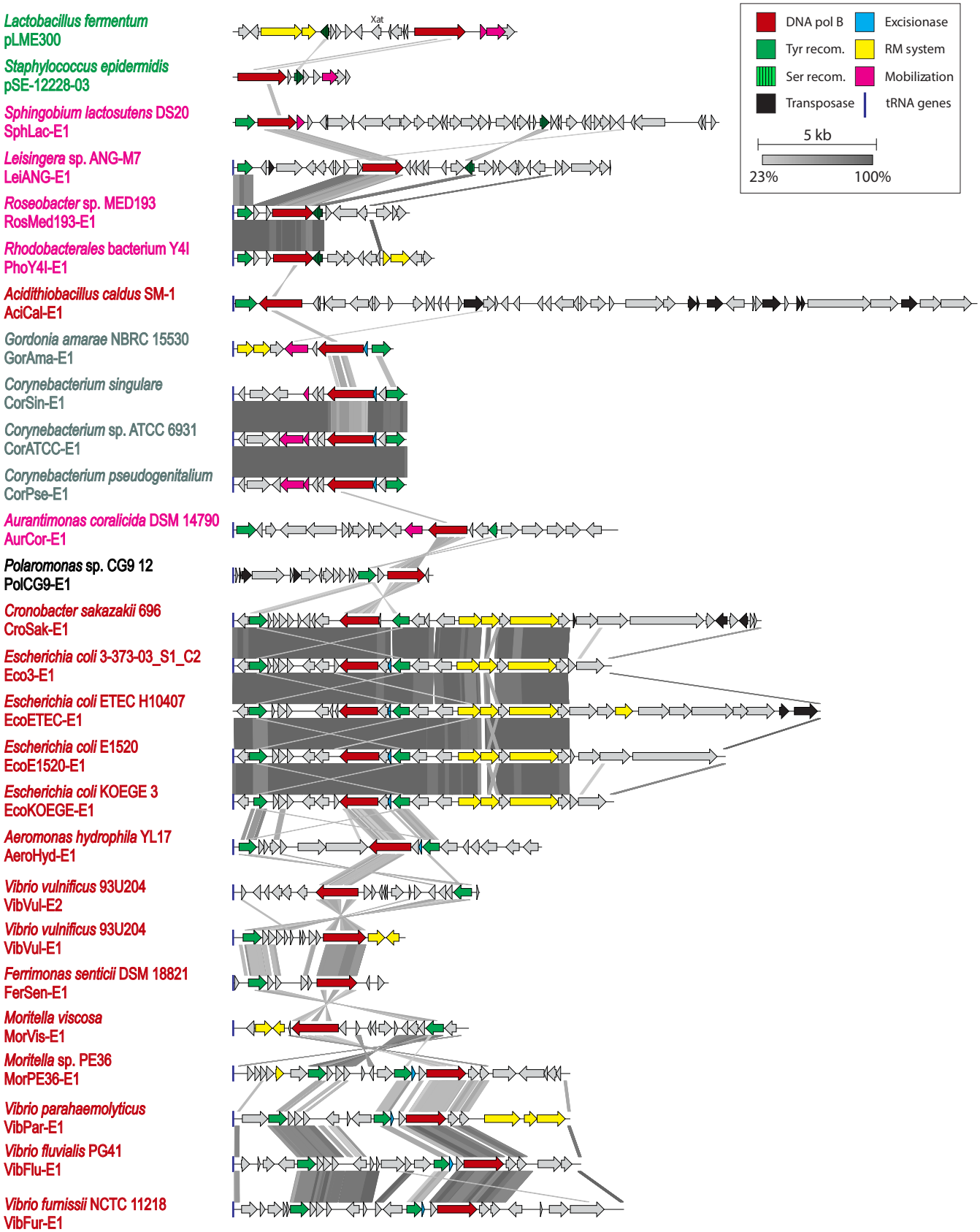


Figure S2. Comparative genomic analysis of bacterial Pipolins. Related to Figure 1.
Predicted protein-coding genes are indicated with arrows, indicating the direction of transcription. 'Xat' stands for xenobiotic acyltransferase. The color key for the designation of the common genes and the greyscale for homology degree are shown in the top right area of the figure. Names of pipolin carrying bacteria are colored as in Figure S1.

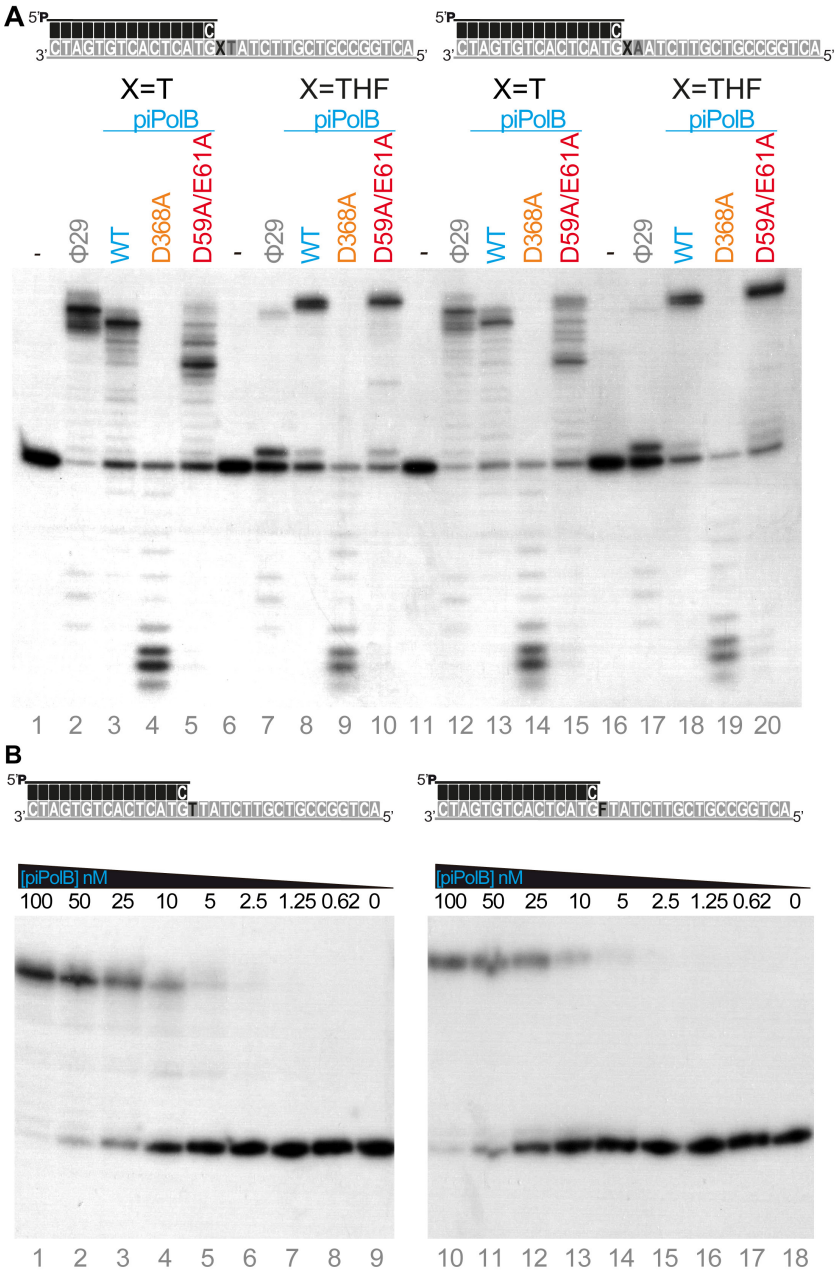


Figure S3. Effect of sequence context and enzyme concentration on TLS capacity by piPolB. Related to Figures 2 and 3.

(A) Denaturing PAGE analysis of primer extension products by Φ29 DNA polymerase (Φ29) and piPolB. As indicated, wild type, D368A (polymerase deficient) and D59A/E61A (exonuclease deficient) piPolB variants were tested using two different sequence contexts (lanes 1-10 vs. 11-20), in the absence (lanes 1-5 and 11-15) or presence (lanes 6-10 and 16-20) of a THF abasic site analog. Schematic representations of each template/primer substrate are depicted above. (B) Processive replication of primer/template substrates by decreasing concentrations of wild type piPolB. Reactions were performed in the presence of either 1 μM (lanes 1-9) or 100 μM (lanes 10-18) dNTPs for undamaged or damaged templates, respectively.

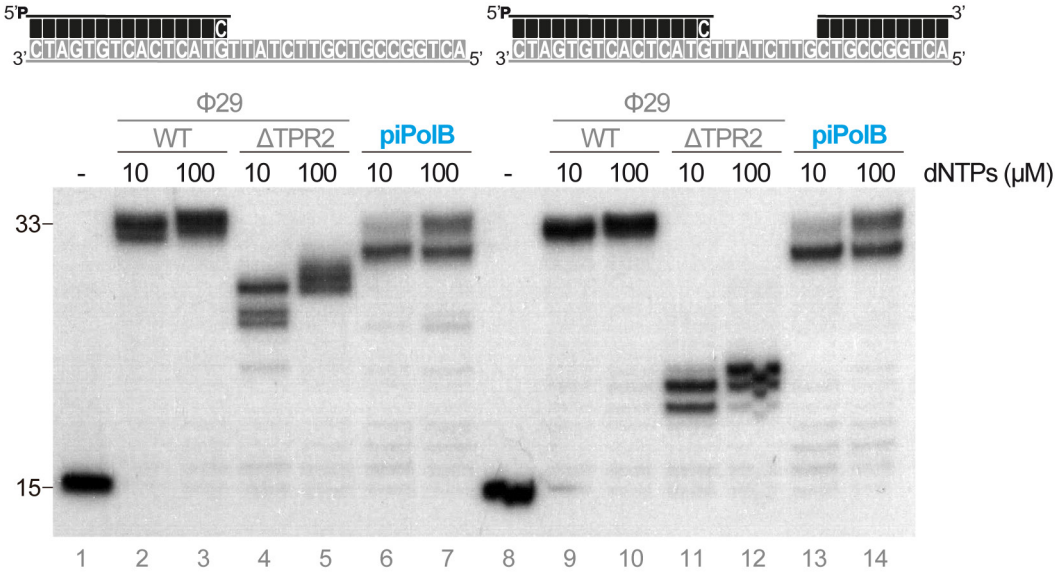


Figure S4. Strand displacement capacity of piPolB. Related to Figure 4.
Denaturing PAGE analysis of replication of a recessed primer/template duplex (lanes 1-7) or a 5 nucleotides gap (lanes 8-14) obtained with a downstream oligonucleotide (P20-33, Table S3) that must be displaced to resume replication by piPolB (lanes 6-7 and 13-14). Wild type and ΔTPR2 variants of Φ29 DNA polymerase were used as positive and negative controls, respectively.

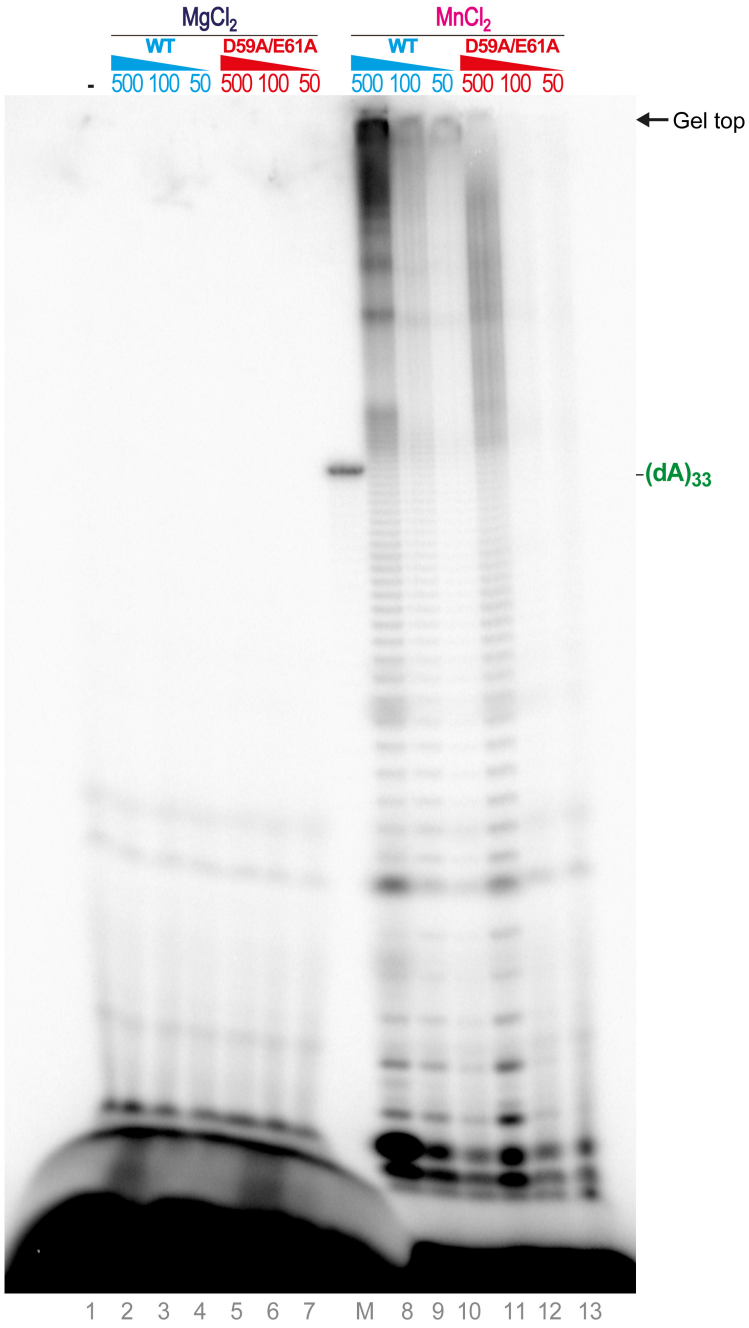


Figure S5. Effect of divalent cations on *de novo* replication of homopolymeric ssDNA substrate by piPolB. Related to Figure 4.

Primer synthesis and replication of homopolymeric poly-dT DNA template (1 μM) by wild type (WT) or exonuclease deficient (D59A/E61A) piPolB variants (500 nM). Reactions were triggered either with 10 mM $MgCl_2$ or 1 mM $MnCl_2$ and resolved in high resolution 8 M urea-20% PAGE. The P4, P10, P15 and 33A primers (see STAR Methods and Table S3) labeled with a 5'-^{32}P were loaded as size markers (lane M), and the size of the shorter products detected are indicated on the left and right, respectively.

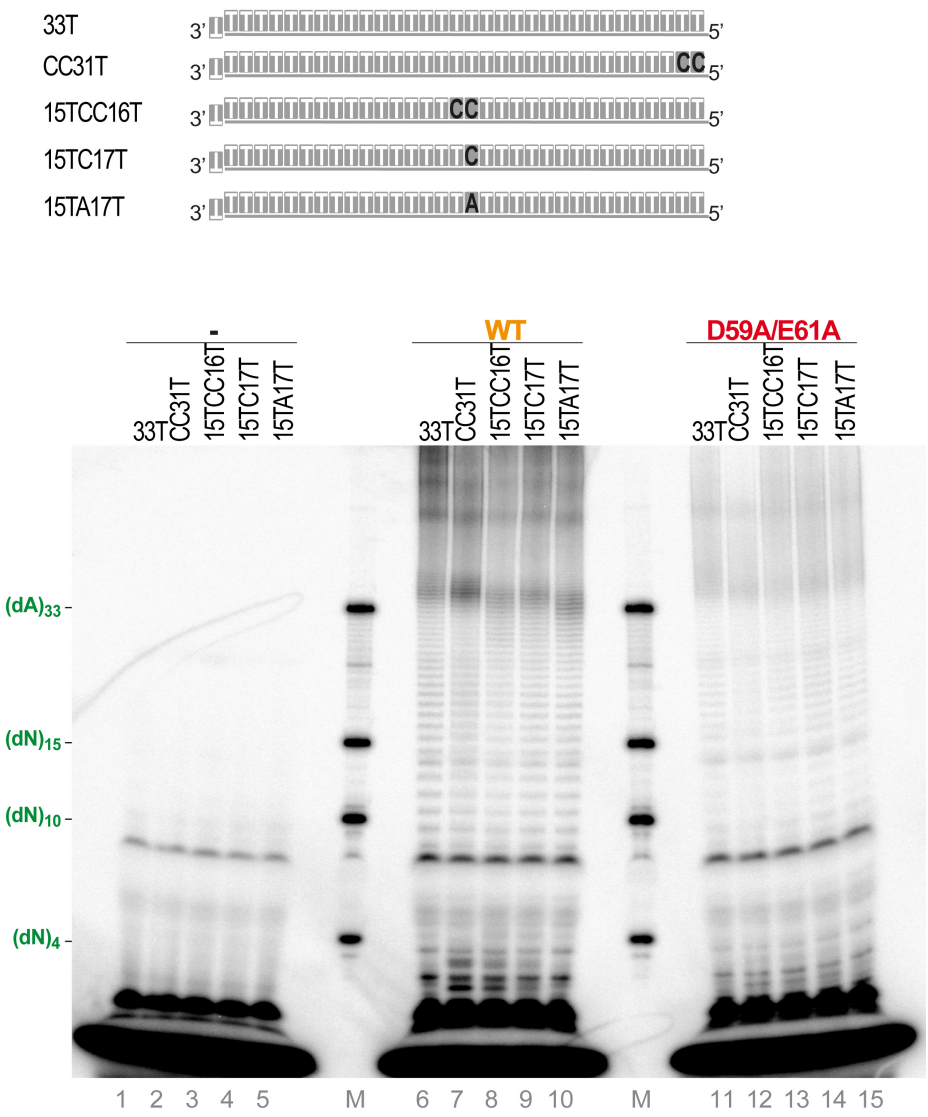


Figure S6. Small effect of single modifications of template sequence on *de novo* replication of homopolymeric ssDNA substrate by piPolB. Related to Figure 5.
Alternative ssDNA templates are depicted above the gel.

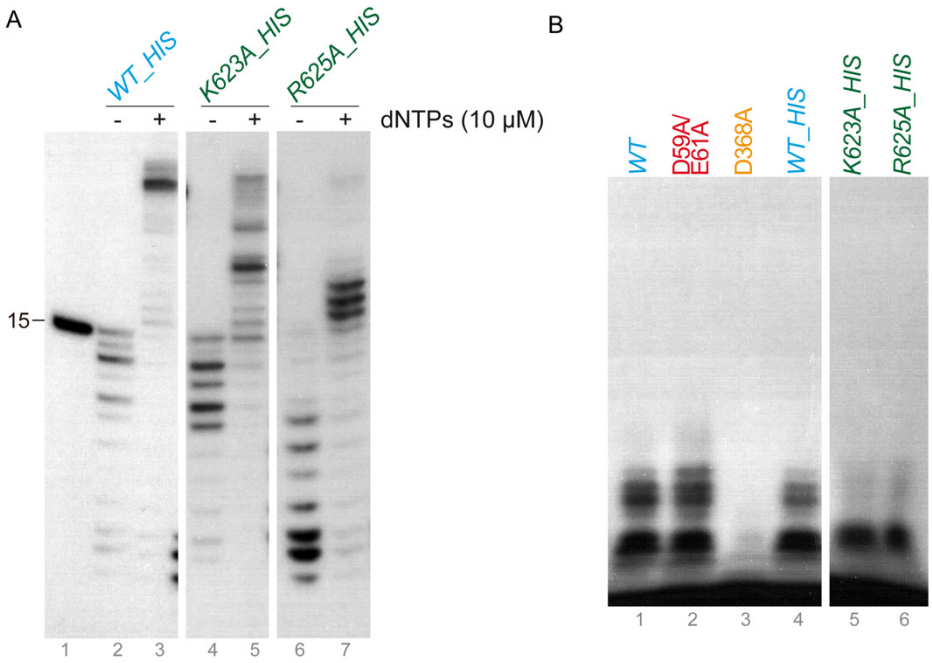


Figure S7. Polymerase and primase capacities of piPolB variants in the analog residues of PolB KxY motif. Related to Figure 6.

Primer extension (A) and primer synthesis (B) capacity of wild type, K624A and R626A His-tagged piPolBs. Assays were performed as in Figure 5. For reference, non-tagged wild type and exonuclease and polymerase deficient piPolB variants were also included in panel B.

Supplementary Table S1. Major characteristics of bacterial integrated pipolins. Related to Figure 1.

Name	Organism	Accession number	Coordinates	Size, bp	att, bp	Target
AciCal-E1	*Acidithiobacillus caldus* SM-1	CP002573	742597..793219	50,623	22	tRNA-Ser
ActICM54-E1	*Actinomyces* sp. ICM54	JDFI01000069	13259..25930	12,672	42	tRNA-Arg
AeroHyd-E1	*Aeromonas hydrophila* YL17	CP007518	1176890..1197918	21,029	57	tRNA-Leu
AurCor-E1	*Aurantimonas coralicida* DSM 14790	NZ_ATXK01000001	336290..362498	26,209	45	tRNA-Met
ColMT41	*Colwellia* sp. MT41	CP013145	3871337..3888198	16,862	70	tRNA-His
CorPse-E1	*Corynebacterium pseudogenitalium* ATCC 33035	GL542874	17142..28962	11,821	44	tRNA-Met
CorSin-E1	*Corynebacterium singular*	CP010827	813240..825107	11,868	77	tRNA-Met
CorATCC-E1	*Corynebacterium* sp. ATCC_6931	CP008913	649037..660848	11,812	35	tRNA-Met
CroSak-E1	*Cronobacter sakazakii* 696	CALF01000071	2319337..2355309	35,973	115	tRNA-Leu
DesAce-E1	*Desulfuromonas acetoxidans* DSM 684	AAEW02000002	57804..90620	32,817	12	tRNA-Glu*
Eco3-373-E1	*Escherichia coli* 3-373-03_S1_C2	NZ_JNMI01000006	64262..90063	25,802	113	tRNA-Leu
EcoE1520-E1	*Escherichia coli* E1520	AEHT01000035	186352..219882	33,531	116	tRNA-Leu
EcoETEC-E1	*Escherichia coli* ETEC H10407	NC_017633	2091087..2131115	40,029	126	tRNA-Leu#
EcoKOEGE-E1	*Escherichia coli* KOEGE_3	NZ_KE701180	343066..369002	25,934	122	tRNA-Leu
FerSen-E1	*Ferrimonas senticii* DSM 18821	NZ_AUGM01000004	223370..233973	10,604	118	tRNA-Trp
GorAma-E1	*Gordonia amarae* NBRC 15530	BAED01000024	21677..32602	10,926	36	tRNA-Lys
LacKis-E1	*Lactobacillus kisonensis* DSM 19906	AZEB01000002	15062..25687	10,626	53	intergenic
LacReu-E1	*Lactobacillus reuteri* TMW1.112	JOKX02000001	220481..239506	19,026	89	Intragenic (KEQ20280)
LeiANG-E1	*Leisingera* sp. ANG-M7	NZ_JWLI01000013	31462..57169	25,708	48	tRNA-Phe
MorPE36-E1	*Moritella* sp. PE36	NZ_ABCQ01000001	278975..301937	22,963	327	tRNA-Leu
MorVis-E1	*Moritella viscosa*	LN554852	4574306..4590347	16,042	187	tRNA-Leu*
PolCG9-E1	*Polaromonas* sp. CG9_12	CCJP01000005	2655959..2669601	13,643	179	tRNA-Phe
RhoY4I-E1	Rhodobacterales bacterium Y4I	NZ_DS995281	71969..85705	13,737	47	tRNA-Phe
RosMED193-E1	*Roseobacter* sp. MED193	AANB01000002	378959..391005	12,047	47	tRNA-Phe
SphLac-E1	*Sphingobium lactosutens* DS20	ATDP01000106	11226..44366	33,141	51	tRNA-Gly#
ThaSte-E1	*Thalassobacter stenotrophicus*	CYRX01000031	20244..36763	16,520	51	tRNA-Asn
VibFlu-E1	*Vibrio fluvialis* PG41	NZ_ASXS01000007	262170..285863	23,694	61	tRNA-Leu
VibFur-E1	*Vibrio furnissii* NCTC 11218	CP002377	779039..805644	26,606	95	tRNA-Leu
VibPar-E1	*Vibrio parahaemolyticus* NBRC 12711	NZ_BBQD01000016	8844..31803	22,960	57	tRNA-Leu
VibVul-E1	*Vibrio vulnificus* 93U204	CP009261	705355..717099	11,745	53	tRNA-Trp
VibVul-E2	*Vibrio vulnificus* 93U204	CP009261	2876933..2893717	16,785	174	tRNA-Leu*

* – integration into 5' end of the tRNA gene; # – tandem integration of two distinct elements.

Table S3. Sequences of oligonucleotides used in this work. Related to Experimental Procedures..

Name*	***Sequence (5'-3')***
P4	GATC
P10	GACTGCTTAC
P15	GATCACAGTGAGTAC
T33GTA	ACTGGCCGTCGTTCTATTGTACTCACTGTGATC
T33GTT	ACTGGCCGTCGTTCTAATGTACTCACTGTGATC
T33GFT	ACTGGCCGTCGTTCTAT**F**GTACTCACTGTGATC
P20-33	GAACGACGGCCAGT
33A	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
33T*	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT/invT/
CC31T*	CCTTTTTTTTTTTTTTTTTTTTTTTTTTTT/invT/
15TCC16T*	TTTTTTTTTTTTTCCTTTTTTTTTTTTTTT/invT/
15TC17T*	TTTTTTTTTTTTTCTTTTTTTTTTTTTTTT/invT/
15TA17T*	TTTTTTTTTTTTTATTTTTTTTTTTTTTTT/invT/
M13 UP	GTAAAACGACGGCCAGT

***F** represents the THF abasic site analog. */invT/ stands for a last dTMP nucleotide linked by an inverted 3'-3' bond.

Table S4. Gene Sequence Information and mutagenesis primers. Related to Experimental Procedures.

Codon-optimized sequence of B-family DNA polymerase from E. coli 3-373-03_S1_C2 pipolin (piPolB) cloned into NdeI-XhoI sites (underlined) of pET23a(+) plasmid	
CATATGAGCAATAACCTGCAAGACATCCTGGCTGCCGCTTCTGGCTACCAAAGTGTCACC TCGGAACCGGCCCTGAATCGCAAACGCCCGAAAACCCTGGATGACTATCCGGTTATTCCG CCGGCGAGCAAGAAAGTGAGCGTGATTAGCTCTGATCTGACCCTGCATATCGGTTTTGAC ACGGAATACGTGTTCAACCCGGAAACCCGCCAGAATGACATCCTGTCGTATCAAAGCTAC GTGGTTCTGCCGGATAACACGGGCATTTCCAATATTATCTATCCGCCGGACTCACAGAAA AAATCTCGTCTGAGTTTCAAAGATTTCCTGTGCCAAACCATTACGCCGCTGCTGGAAACC GGTGTTATCACGAAATGGCCGGGCATTATCAACATTTACGCCCACTTTATTCGCGCGGAC ATCGCCTCGTTTGCAAACTTCTGGAGCGATTACAAAATCCTGCTGAAAGGCATCCGTGGT ACCGTTAGTTCCTTTAAAAACCGCTACGGTATCGATTTCGACGAACAGCAAGAACGTCGC GTCAAAACCGAACAGATTATGTTTGATAAACGTACGTCTCCGCCGCGCTGCAGTAATGTG GCCTTCATTGATACCCTGCTGATCACGCCGGGCGGTATGGGTCTGGCAGAATGTGGCGAA CTGCTGGGTCTGCCGAAACTGACCATTCCGGCTCCGTATAGCATCACGAACATGCGCGAA TACCTGCTGGGTGACCGTGCAGGTTTTGAAGCGTATGCGCTGCGTGATGCTGAAATCGCG GTTCGCTACGCTCTGCAGGTCCGTAATTTCTGCGCGCGCGAACTGATGATTGATCGTGTGC CGGCAACCATTGGTGCCATGGCCGTTTCTCGTTTCACCAAAACGCTGAAAGAAAACAACA TGAGTCCGGAAGTGTGCCTGGGCACCCATATCAAAACGCGTGAACTGTGGCTGACCGAA AAACAGGCCTTTCGCACGATTAAAAACCCGGCATCCGTTCCGTCACGTGAACTGTTTGAA ACCTTCCCGATTAACTGCTATCATGGCGGTCGCAATGAATGTTTCATGATGGGTGTGACC CCGTCAGATCACTGGTATGATTACGACCTGGCAGGCGCTTATACCACGGGTCTGCTGGAT ATTCTGACCCCGGACTACGGCAACATCCGTCTGAGCAAAAATCCGGATGACTATTGCGGC CATGTGATGGGTTTTGCGCTGGTTACCTTTCGCTTCCCGGAATCCGTCCCGTATCCGTCAC TGCCGGTGCGTACGGATCAGTACGGTCTGTTTTTCCCGCTGAGCGGTGAAAGCTGGGCCA CCGCCCCGGAAATTGAACTGGCCCTGTCCCTGGGTGCAGAAATGACGATCCATAACGGCA TTATCGTGCCGTGGATTTGTGATACCAGCCCGCACAATAGTGAATCCACGTCAGTTTTTCT GCCGTTTGTGCAGCAAGTTCGCGAAAACCGTAATCGCCATATCAAAGGTTCCCTGGAAGA AAAATTCTGGAAAGAAATCGGCAACTCACTGTATGGTAAACTGGCTCAGGGCCTGCGTGC CAAAACCGCATTCGATACCGCGCGTGGCCTGAATCGCAGCCTGCCGCCGTCATCGGTCAC CCAACCGTTTTTCGCGGCCCACGTGACGGGTTTTATTCGCGCTGTCGTGGGCGAACTGAT GAATGCGCTGCCGTCTGATAGCTCTGTTGTCAGTGTGACCACGGACGGCTTTCTGACCAA CTGTCCGCTGGATAAAATCAATATGTCGGGTCCGCTGAGTTCCCGCTTCCAGAGCCTGTG CGATATTGTTGACCCGGGTTCATCGATGCTGACCTGTAAACATGAAGTCTCTCAACTGAT CGCCATGAAAACCCGTGGTCAGCTGACGTATAAAGCAATTCAAGGCAAACCGGTGGTTC ATGCACGCGCTGGTGTCAAACCGCCGGCCGACATTCCGCGTAGTGATTATAACGACTACA TGGTGGATCTGTACCTGAATCGTCTGCCGGGTCAGACCCTGTCGCGTAGCACCCTGATCT CGACGCGCGAAATGTGGCTGTCTGAAAGTGATCTGGTTAGCCGTGAACAGGACATTCGCC TGAACCTGGAATTTGATTTCAAACGTCAACCGGTGCGCCCGGCGATGAACGAAGGCCATC TGCTGATGTTTTCTCGTCCGTGGGATAATATGGAAGAAGCCCTGCAGCAACGTAGTCTGT TCGATGACTGGCGCCAGACCCACACGCTGAAAACCCTGGCCGATTGGGATGACTGGTGC GACTTTCTGTATTGTCGCACGGTTTTCTCTGATATGAAACTGAAAGTCGGCTCTAAACGTA GTGATGACATCCTGGTTCGTCTGTTTCTGCGCGCACTGACCCAGTGCCAATGGGGTCTGAT GCTGAAAGATAAAAAATCCTACTCATGTAAAGAAGTGGCGGAATGGCTGACCTCGGAAG GCTATAGCGTTACCGTCACGGATGTCAAAAATGCTGTGCGTGCGAAAATTCCGCAGATGA AATTTAGCTCTGTGACCCCGCGTATGAAATCCCTGATGGACATTATCGCCCGTAAATATC CGACCTTTTGCCTGCCGGTTTAACTCGAG	
Oligonucleotide pairs used for site-directed mutagenesis	
piPolB variant	*Sequence (5'-3')*
D368A	CCCCGTCAGATCACTGGTATGATTACGCCCTGGCAGGCGCTTATACCACG
	GCTGGTATAAGCGCCTGCCAGGGCGTAATCATACCAGTGATCTGACGGG G
D59A/E61A	CCCTGCATATCGGTTTTGCCACGGCATACGTGTTCAACCCGGAAACCC
	GGGTTTCCGGGTTGAACACGTATGCCGTGGCAAAACCGATATGCAGGG
WTHis	CCTTTTGCCTGCCGGTTTTACTCGAGCACCACCACCACCACCAC
	GTGGTGGTGGTGGTGGTGCTCGAGTAAAACCGGCAGGCAAAAGG
K614A	GGGTTCATCGATGCTGACCTGTGCACATGAAGTCTCTCAACTGATCGC
	GCGATCAGTTGAGAGACTTCATGTGCACAGGTCAGCATCGATGAACCC
H615A	GGGTTCATCGATGCTGACCTGTAAAGCTGAAGTCTCTCAACTGATCGC

	GCGATCAGTTGAGAGACTTCAGCTTTACAGGTCAGCATCGATGAACCC
K624A	GTCTCTCAACTGATCGCCATGGCAACCCGTGGTCAGCTGACG
	CGTCAGCTGACCACGGGTTGCCATGGCGATCAGTTGAGAGAC
R626A	GTCTCTCAACTGATCGCCATGAAAACCGCTGGTCAGCTGACGTATAAAG C
	GCTTTATACGTCAGCTGACCAGCGGTTTTCATGGCGATCAGTTGAGAGAC

Supplementary References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972-1973.

García-Gómez, S., Reyes, A., Martínez-Jiménez, M.I., Chocron, E.S., Mourón, S., Terrados, G., Powell, C., Salido, E., Méndez, J., Holt, I.J.*, et al.* (2013). PrimPol, an archaic primase/polymerase operating in human cells. Mol Cell *52*, 541-553.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol *59*, 307-321.

Krupovic, M., and Koonin, E.V. (2015). Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. Nat Rev Microbiol *13*, 105-115.

Okonechnikov, K., Golosova, O., Fursov, M., and team, U. (2012). Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics *28*, 1166-1167.

Pei, J., and Grishin, N.V. (2014). PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol Biol *1079*, 263-271.

Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res *33*, W244-248.

Spriestersbach, A., Kubicek, J., Schafer, F., Block, H., and Maertens, B. (2015). Purification of His-Tagged Proteins. Methods Enzymol *559*, 1-15.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr Purif *41*, 207-234.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011). Easyfig: a genome comparison visualizer. Bioinformatics *27*, 1009-1010.