Supplementary Information for

Quantifying XNA replication fidelity using nanopore sequencing

Authors: Nicholas A. Kaplan, Jayson R. Sumabat, Jane V. McKelvey, Jeantine E. Lunshof, Jorge A. Marchand

Materials and Methods

Supplementary Tables

Supplementary Figures

Materials and Methods

Commercial materials. Agarose (0710-500G; electrophoresis grade) and AMPure XP purification beads (NC9959336) were purchased from Thermo Fisher Scientific (Waltham, MA). Tris base (10708976001), 5 M betaine solution (B0300-1VL), 6 N hydrochloric acid (1430071000), GelGreen (SCT124), premixed TBE buffer (10x, 11666703001), potassium chloride (P3911-500G), potassium acetate (P1190-100G), bovine serum albumin (A7030-10G), dimethyl sulfoxide (DMSO; D8418-250ML), magnesium chloride hexahydrate (63068-250G), TritonTM X-100 (X100-500ML), and sodium chloride (S3014-5KG) were purchased from Sigma-Aldrich (St. Louis, MO). Restriction enzymes, Exo III (M0206L), thermolabile Exo I (M0568L), Exo I (M0293L), Exo VIII (truncated; M0545S), NEBNext Ultra™ II End Repair/dA-Tailing Module (E7546S), NEBNext Ultra™ II Ligation Module (E7595S), Salt T4 DNA Ligase (M0467S), 50 bp DNA ladder (N3236S), and 2′-Deoxynucleoside triphosphate solution mix (dNTP; N = A, T, G, C; N0446S) and 2′-Deoxynucleoside triphosphate solution set (dATP, dTTP, dGTP, dCTP; N0446S) were purchased from New England Biolabs (Ipswich, MA). Gene ruler 1 kb plus DNA ladder (SM1331) was purchased from Invitrogen (Carlsbad, CA). Thermostable polymerases Q5 (M0491L), Phusion (M0530S), DeepVent (M0258S), DeepVent (exo-) (M0259S), and OneTaq (M0480S) were purchased from New England Biolabs (Ipswich, MA). Thermostable polymerases Taq (50-103-7174), Titanium Taq (NC9806143), AccuPrime *Pfx* (12344024), *Tth* (NC1894531), SD (501685638), *Pfu* (501036614), KlenTaq1 (NC9707348), and Takara Taq (50443948) were purchased from Thermo Fisher Scientific (Waltham, MA). Deoxyoligonucleotides (oligos) sequences are included in **Supplementary Tables 2, 3, & 5**. Oligos containing only standard bases, in addition to oligos **B1-5** (/iisodG/), oligos **S1-5** (/iMe-isodC/) were purchased from Integrated DNA Technologies (Coralville, IA). Oligos **P1**, **Z1**, and **PB1** were purchased from Firebird Biomolecular Sciences LLC (Alachua, FL), and oligos **D1-4** were ordered from Xenolis Pte. Ltd (Singapore). Upon receipt, all oligos were resuspended at a stock concentration of 100 μM in elution buffer (10 mM Tris-HCl, pH = 8.2) and stored at either 4 °C for immediate usage or -20 °C for long-term storage. Xenonucleoside triphosphates dPTP, dZTP, dBTP (dPTP-201, dZTP-101, dBTP-301P) were purchased from FireBird Biomolecular Sciences LLC (Alachua, FL). Xenonucleoside triphosphate dS_nTP (C-001) was purchased from TriLink BioTechnologies (San Diego, CA). Xenonucleoside triphosphates dDsTP and dDiol-dPxTP (XTP-0100, XTP-0101) were purchased from Xenolis Pte. Ltd (Singapore). DNA purification kits (ZD4034, ZD7011) were purchased from Zymo Research (Irvine, CA). MinION sequencing device (MIN-101B), Flongle Adapter, Flongle Flow Cells (FLO-FLG114), Ligation Sequencing Kit (SQK-LSK114), and Flongle Sequencing Expansion kit (EXP-FSE002) were purchased from Oxford Nanopore Technologies (ONT; Oxford, United Kingdom). Unless otherwise specified, other commodity chemicals used in this work were purchased from major domestic distributors (Sigma-Aldrich, St. Louis, MO; Thermo Fisher Scientific, Waltham, MA).

Preparation of DNA for model training. DNA and XNA-containing oligos for training comparative basecalling models were designed to differ only at the XNA position. For a given replication error mode, variations in the XNA position were created to include the expected mutation. For example, to train a model that could capture **B→A** transitions, the training set included both 5'-C**B**G and 5'-C**A**G sequences. Synthetic DNA sequences used for training various comparative basecalling models are shown in **Supplementary Table 2**. All training reads were generated using Flongle flow cells from Oxford Nanopore Technologies (R10.4.1). While the

general training approach was the same, there were slight variations in training for each ubp XNA pair.

To generate training sequences for **B** and **S (B:** iso-guanosine; **S**: 5-Me-isocytosine**)**, complementary 90mer oligos containing the XNA base (**B1-B5, S1-S5**) were combined in an equal molar ratio and annealed by heating to 95 ℃ and slowly cooling to 20 ℃ (0.1 ℃ /s) in 100 mM NaCl using a thermocycler. The annealed oligos were purified using the Zymo DNA Clean & Concentrator kit, then prepared for nanopore sequencing as described in the "**Nanopore sequencing sample preparation and data acquisition**" section. Canonical DNA training data was generated by PCR amplifying oligos containing a standard base in the XNA position (**BA1-5, SA1-5, BN2, SN2**) using a high-fidelity polymerase (Phusion Polymerase in HF Buffer or Q5 polymerase in Q5 reaction buffer) using 200 μM dNTP concentration. Forward and reverse primers used in amplification were ordered containing 24mer ONT barcodes (**Supplementary Table 3 & 4**), allowing multiple canonical training sequences to be multiplexed using a single flow cell. Before training, canonical reads were demultiplexed using CutAdapt (v: 4.9).

To generate training sequences for **P** and **Z** (**P**: (2-amino-8-(1′-β-d-2′-deoxyribofuranosyl); **Z**: (6-amino-5-nitro-3-(1′-β-d-2′-deoxyribofuranosyl)-2(1H)pyridine)), 71mer oligos containing **P** and **Z** were used (**P1, Z1**). Due to concerns of poor post-sequencing alignment with short sequences, training sequences were prepared by TA ligation of annealed oligos to barcoded hairpins (**Supplementary Fig. S8**), as described in the "**Preparation of synthetic DNA controls**" section (hairpin sequences: **Supplementary Table 5**). After the hairpin products were digested with *ScaI-HF* to generate blunt ends, the barcoded dsDNA containing **P≡Z** was prepared for sequencing according to standard methods. Canonical DNA sequences containing the **G≡C** pair substituting the **P≡Z** position were prepared by PCR amplification of identical 71mer oligos containing a **G** at the **P** position (**PG1**), and a C at the Z position (**ZG1**).

To generate training data for the 6-letter template containing both **P** and **B** (**PB1**), primer **R_14** (**Supplementary Table 3**) was annealed to **PB1** and extended with Q5 polymerase at 72 ℃ for 15 minutes using Q5 Reaction Buffer (1X). dNTP stocks were added at the following final concentrations: dATP (0.4 mM), dGTP (0.4 mM), dCTP (0.4 mM), dTTP (0.4 mM), dSTP (0.4 mM), and dZTP (0.2 mM). An agarose gel for this extension is shown in **Supplementary Fig. S29A**.

Since **Px** was not commercially available as a product for phosphoramidite DNA synthesis, strands containing **Diol-Px** used for training were generated by performing a primer extension from a 60-mer oligo containing **Ds** (**D1-D4**). Primer extension reactions were performed following recommendations from Xenolis. Primer **Ext_D1-2_30** (**Supplementary Table 3**) was annealed to templates **D1** and **D2**, while primer **Ext_D3-4_25** was annealed to templates **D3** and **D4**. Primer extensions were performed using AccuPrime *Pfx* polymerase at 68 ℃ for 60 minutes. Deoxynucleoside triphosphate stocks were added at the following final concentrations: dATP (0.4 mM), dGTP (0.4 mM), dCTP (0.4 mM), dTTP (0.4 mM), and dDiol-PxTP (0.05 mM). An agarose gel for primer extension of templates **D1** and **D2** is shown in **Supplementary Fig. S29B**.

Nanopore sequencing sample preparation and data acquisition. Nanopore sequencing samples were prepared using the Ligation Sequencing Kit V14 from Oxford Nanopore Technologies. 500 ng total of DNA were first end-prepped using the NEBNext Ultra II End Repair/dA-Tailing module, followed by an AMPure XP cleanup using a 1.8-3X bead ratio, depending on the product length. Beads were washed twice using freshly prepared 70% ethanol. Clean samples were then ligated with ligation adapters (LA) in ligation buffer (LNB) using either

Quick T4 DNA Ligase (NEB) or Salt T4 DNA Ligase (NEB), as recommended by ONT. Ligations were purified using AMPure XP and washed twice with the ONT short fragment buffer (SFB). Samples were eluted using ONT elution buffer (EB). Flongle flow cells were first flushed with flow cell flush (FCF) containing flow cell tether (FCT). DNA prep quality was assessed by A_{260} before loading flow cell with library beads (LIB), sequencing buffer (SB) and full DNA sample. Data acquisition for nanopore sequencing was performed using MinKNOW (v:23.11.7, ONT).

Model training, basecalling, and classification. Comparative XNA basecalling models were trained on reads generated from prepared sequences (**Supplementary Table 2**). Raw nanopore sequencing data (.pod5 file format) were first basecalled using the Dorado basecaller (ONT, v: 0.8.0) with a high accuracy model (hac; dna_r10.4.1_e8.2_400bps_hac@v5.0.0) with a q-score filtering threshold = 5 and move table output. Output sequences were aligned to the canonical DNA-containing reference using minimap2 (v: 2.17-r941). The XNA position was specified using a BED file, which was used to extract signal training chunks using Remora (ONT, v: 2.1.3) `dataset prepare` command. Separate chunk files were generated for modified and canonical data, then merged into a balanced training set. All models were trained using `remora model train` with the following parameters: kmer context = (4,4); chunk context = (50 50), reference anchor = True unless otherwise noted. The default remora PyTorch model (LSTM with reference) was used for training. Training was allowed to proceed until convergence, using a default minimum of 5 epochs. 20% of the training reads were held for internal validation. Training set sizes and performance statistics for each model are shown in **Supplementary Table 6**.

Model testing and benchmarking. Models were tested using raw data (.pod5 files) withheld from model training and validation. Individual .pod5 files for ubp XNA or canonical DNA sequences were basecalled first using Dorado, then classified using a comparative basecalling model as input to Remora. Results from per-read modifications files for each aligned sequence were analyzed to calculate true positive (TP), true negative (TN), false positive (FP), and false negative (FN) for 10,000 basecalled reads from each dataset. In certain cases, this threshold was reduced if the testing dataset size was smaller than 10,000 reads. Equations used for reported benchmarking calculations are shown below. Benchmarking results are shown in **Supplementary Table 7.**

$$\text{Recall (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity (TNR)} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Validation of XNA fraction quantification *in silico* using mixed synthetic datasets. To validate model performance on mixed datasets, raw reads (.pod5) with a known ground truth were mixed in predefined ratios, and XNA classification models were used to estimate XNA fraction. A graphical overview of this process is shown in **Supplementary Fig. S6A**. Two 90-bp synthetic DNA constructs, identical except at the XNA position (**BS: B1,S1** or **AT: BA1, ST1**), were sequenced separately. Raw sequencing reads (.pod5 format) were then randomly sampled and merged using the `pod5` package in defined XNA/DNA ratios, from 0 to 1 in steps of 0.1. Each of these merged pod5 files (200,000 total reads) was basecalled, aligned, and classified using comparative basecalling models.

Purification, library preparation, and nanopore sequencing of PCR samples. After thermocycling, completed PCR reactions were held at 4 ℃ then stored at -20 ℃ until purification. To assess reaction completion and verify product length, 5 µL of each individual replicate was analyzed by gel electrophoresis (1X TBE buffer, 2% agarose, 1X GelGreen stain) before purification and compared against a 50 bp ladder (NEB). After gel verification, reactions were pooled (15 µL each) and purified using the Zymo DNA Clean & Concentrator kit. Pooled samples were eluted in 45 µL of MilliQ water. DNA concentration of each pooled product was measured by A_{260} using a SpectraMax-QuickDrop™ Spectrophotometer. For nanopore sequencing, PCR products were pooled equivalently by mass with appropriate synthetic controls and prepared following the details in the "**Nanopore sequencing sample preparation and data acquisition**" section. Pooled reaction products were sequenced using Flongle flow cells (Oxford Nanopore Technologies) with R10.4.1 flow cell chemistry. DNA preparation quality was assessed by A_{260} before loading the flow cell with library beads (LIB), sequencing buffer (SB), and the full DNA sample. All reads remaining after filtering, alignment, and demultiplexing were used for analysis.

Filtering, demultiplexing, and analysis pipeline for PCR datasets. Nanopore reads were demultiplexed by matching both forward and reverse 24mer barcode pairs to unique samples. The 'Natural Barcodes' (NB) set from ONT was used, **Supplementary Table 4**. For each barcode pair, specified forward-reverse and reverse-forward barcode combinations were generated and searched using CutAdapt with a minimum overlap requirement of 14 and error tolerance of 20%. For the expected 150 nt product, read lengths were filtered by specifying a minimum length of 120 nt and a maximum length of 200 nt. Length filters were adjusted for shorter expected product lengths. Reads matching either orientation were combined and deduplicated to form a single FASTQ file per barcode pair. Next, unique read IDs were extracted from each FASTQ file and annotated with the corresponding barcode pair. Reads that appeared in multiple barcode files were removed from all files. The resulting read ID list was merged with the per-read modification predictions from Remora (output from `remora validate`), linking classification results to barcode assignments. The number of reads predicted as modified (class_pred = 1) or unmodified (class_pred = 0) for each barcode pair were written to a summary file for data analysis.

Theory and assumptions for quantifying replication fidelity.

Case 1: Systems with only XNA-to-DNA mutations. We first define an average polymerase fidelity of a 6-letter ubp XNA system (A, T, G, C, X, Y) with one error mode: XNA to standard DNA mutations. Though the related theory was used originally by Sismour & Benner in 2005, we extend certain definitions for compatibility with our measurements.[37] The total number of dsDNA molecules (N_{tot}) in a theoretical PCR at some cycle (n), is related to the initial number of molecules

($N_{tot,i}$) according to **Eqn**. **1**, where $N_{tot,i}$ describes all possible strand pairings. In the theoretical framework, PCR cycles proceed with 100% amplification efficiency, which can be experimentally enforced by performing reactions with defined primer-to-template ratios.

$$N_{tot} = N_{tot,i} \cdot 2^n \qquad \textbf{[Eqn. 1]}$$

After 'n' number of PCR cycles, the total number of dsDNA molecules containing a ubp XNA pair (N_{XNA}) can be related to the initial number of XNA-containing molecules ($N_{XNA,i}$) and the average polymerase fidelity per cycle $\varphi_{\langle X,Y\rangle}$ for a particular ubp XNA system. $\varphi_{\langle X,Y\rangle}$ is assumed to be cycle invariant. Since PCR amplifies both strands simultaneously, the replication fidelities of individual bases become coupled. Therefore, this average fidelity, $\varphi_{\langle X,Y\rangle}$, is the only observable fidelity from PCR measurements.

$$N_{XNA} = N_{XNA,i} \cdot (1 + \varphi_{\langle X,Y\rangle})^n \qquad \textbf{[Eqn. 2]}$$

Using sequencing, we measure the fraction of DNA that contains, or does not contain, ubp XNA bases. By dividing **Eqn. 2** by **Eqn. 1** and assuming we start with the ubp XNA system ($N_{tot,i} = N_{XNA,i}$), we can define the XNA fraction (f_{XNA}) as the ratio ($N_{XNA}/\ N_{tot}$). XNA fraction (f_{XNA}) at some cycle (n) can then be related back to the average polymerase fidelity, $\varphi_{\langle X,Y\rangle}$, according to **Eqn. 3**.

$$f_{XNA} = \frac{N_{XNA}}{N_{tot}} = \left(\frac{1+\varphi_{\langle X,Y\rangle}}{2}\right)^n \qquad \textbf{[Eqn. 3]}$$

Since measurements produced in our sequencing assays are strand-specific, we need to relate f_{XNA} to strand-specific ubp XNA fractions f_X and f_Y. Fraction of arbitrary XNA '**X**' (f_X) in each DNA strand can be related to the total number of molecules containing **X** (N_X) and the total number of molecules containing a standard DNA base at that same position ($N_{N(+)}$, for (+) strand) by **Eqn. 4A**.

$$f_X = \frac{N_X}{N_X + N_{N(+)}} \qquad \textbf{[Eqn. 4A]}$$

Similarly, Fraction of the complementary XNA '**Y**' (f_Y) is related to the total number of molecules containing **Y** (N_Y) and the total number standard DNA base complements at the same position, ($N_{N(-)}$), by **Eqn. 4B**.

$$f_Y = \frac{N_Y}{N_Y + N_{X(-)}} \qquad \textbf{[Eqn. 4B]}$$

To relate f_{XNA} to the observed XNA fractions (f_X and f_Y), we first rewrite XNA fraction in terms of number of molecules of each species in **Eqn. 5**.

$$f_{XNA} = \frac{N_{XNA}}{N_{tot}} = \frac{N_X + N_Y}{N_X + N_{N(+)} + N_Y + N_{N(-)}} \qquad \textbf{[Eqn. 5]}$$

To relate the observed read counts to the underlying molecular abundances, we explicitly account for sequencing bias in our data. Let β represent sequencing bias for each species. The number of observed molecules (N^{obs}) is related to the number of true molecules (N) according to the **Eqns. 6A-6D** below.

$$N_X^{obs} = N_X \cdot \beta_X \qquad \textbf{[Eqn. 6A]}$$
$$N_Y^{obs} = N_Y \cdot \beta_Y \qquad \textbf{[Eqn. 6B]}$$
$$N_{N(+)}^{obs} = N_{N(+)} \cdot \beta_{N(+)} \qquad \textbf{[Eqn. 6C]}$$
$$N_{N(-)}^{obs} = N_{N(-)} \cdot \beta_{N(-)} \qquad \textbf{[Eqn. 6D]}$$

Although we detect systematic sequencing bias between sense (+) and antisense strands (-), bias within a given strand is minimal. We therefore make the following assumptions **Eqns. 7A** and **7B**.

$$\beta_X \approx \beta_{N(+)} = \beta_{(+)} \qquad \textbf{[Eqn. 7A]}$$
$$\beta_Y \approx \beta_{N(-)} = \beta_{(-)} \qquad \textbf{[Eqn. 7B]}$$

If the underlying system has equal numbers of sense and antisense strands (N sense = N antisense), then $N_X + N_{N(+)} = N_Y + N_{N(-)}$. We can relate the sequencing bias to observed number of reads for each species, as shown in **Eqn. 7C**.

$$\frac{\beta_{(+)}}{\beta_{(-)}} = \frac{N_X^{obs} + N_{N(+)}^{obs}}{N_Y^{obs} + N_{(-)}^{obs}} \qquad \textbf{[Eqn. 7C]}$$

Substituting **Eqns. 6A-6D**, and **7C** into **Eqn. 5** (algebra not shown) yields **Eqn. 8**, which reveals f_{XNA} is independent of these bias terms and simply the arithmetic mean of the measured XNA fractions (f_X and f_Y).

$$f_{XNA} = f_{\langle X,Y \rangle} = \frac{f_X + f_Y}{2} \qquad \textbf{[Eqn. 8]}$$

Finally, using **Eqn. 9a**, $\varphi_{\langle X,Y \rangle}$ can be calculated directly from observed fractions, f_X and f_Y. We refer to this calculation as a *single endpoint measurement* of fidelity.

$$\varphi_{\langle X,Y \rangle} = 2 * 2^{\log(f_{\langle X,Y \rangle})/n} - 1 \qquad \textbf{[Eqn. 9a]}$$

Alternatively, this function can be log-linearized, shown in this work as the negative $\log_2$ of the average XNA fraction, ($-\log_2(f_{\langle X,Y \rangle}(n))$) and the slope can be used to extract the $\varphi_{\langle X,Y \rangle}$ parameter (**Eqns. 9b-9d**).

$$log_2(f_{\langle X,Y \rangle}) = -n\log_2\left(\frac{1+\varphi_{\langle X,Y \rangle}}{2}\right) \qquad \textbf{[Eqn. 9b]}$$

$$-m = \log_2\left(\frac{1+\varphi_{\langle X,Y \rangle}}{2}\right) \qquad \textbf{[Eqn. 9c]}$$

$$(2 * 2^{-m}) - 1 = \varphi_{\langle X,Y \rangle} \qquad \textbf{[Eqn. 9d]}$$

Rather than using the average fidelity parameter, various sources in literature have instead parameterized ubp XNA fidelity using an observed decay constant, $\delta_{\langle X,Y\rangle}$, modeled from the exponential decay of ubp XNA pairs as a function of theoretical cycle (n). This decay rate, $\delta_{\langle X,Y\rangle}$, can be related to average fidelity by the average error rate, $\varepsilon_{\langle X,Y\rangle}$, through **Eqns. 10-12**.

$$f_{\langle X,Y\rangle} = \left(1 - \delta_{\langle X,Y\rangle}\right)^n \qquad \textbf{[Eqn. 10]}$$

$$\delta_{\langle X,Y\rangle} = 1 - 2^{\log_2(f_{\langle X,Y\rangle})/n} \qquad \textbf{[Eqn. 11]}$$

$$1 - \varphi_{\langle X,Y\rangle} = \varepsilon_{\langle X,Y\rangle} = 2 \cdot \delta_{\langle X,Y\rangle} \qquad \textbf{[Eqn. 12]}$$

Modeling PCR using decoupled individual nucleotide replication fidelities

As shown in **Eqn. 13**, individual nucleotide replication fidelities, φ_X and φ_Y, are related to the average fidelity, $\varphi_{\langle X,Y\rangle}$, by their geometric mean.

$$\varphi_{\langle X,Y\rangle} = \sqrt{\varphi_X \cdot \varphi_Y} \qquad \textbf{[Eqn. 13]}$$

The relationship shown in **Eqn. 13** arises from the dominant eigenvalue of the coupled replication ODE system (**Eqn. 14**), which captures the joint retention probability of the ubp during PCR.

$$\frac{d}{dn}\begin{bmatrix} N_X \\ N_Y \\ N_{N(+)} \\ N_{N(-)} \end{bmatrix} = \begin{bmatrix} 0 & \varphi_Y & 0 & 0 \\ \varphi_X & 0 & 0 & 0 \\ 0 & 1-\varphi_Y & 0 & 1 \\ 1-\varphi_X & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} N_X \\ N_Y \\ N_{N(+)} \\ N_{N(-)} \end{bmatrix} \qquad \textbf{[Eqn. 14]}$$

A similar coupled replication ODE system, shown in **Eqn. 15**, allows for the tracking of dsDNA species counts (N) as a function of theoretical cycle. Here, the dsDNA species include correctly paired duplexes containing the arbitrary XNA base pair **X** and **Y** (*X,Y*), duplexes consisting only of standard bases (*N,N*), mismatched duplexes between **X** and a standard base (*X,N*), and mismatched duplexes between **Y** and a standard base (*N,Y*). Plots generated by modeling this ODE as discrete difference equations are shown **in Supplementary Figure S10**.

$$\frac{d}{dn}\begin{bmatrix} N_{X,Y} \\ N_{X,N} \\ N_{N,Y} \\ N_{N,N} \end{bmatrix} = \begin{bmatrix} \varphi_X + \varphi_Y & \varphi_X & \varphi_Y & 0 \\ 1-\varphi_X & 1-\varphi_X & 0 & 0 \\ 1-\varphi_Y & 0 & 1-\varphi_Y & 0 \\ 0 & 1 & 1 & 2 \end{bmatrix}\begin{bmatrix} N_{X,Y} \\ N_{X,N} \\ N_{N,Y} \\ N_{N,N} \end{bmatrix} \qquad \textbf{[Eqn. 15]}$$

An 8-letter coupled replication ODE system, shown in **Eqn. 16**, tracks various ssDNA species counts containing ubp XNAs at two distinct positions (sense strand: **P** and **B**; antisense strand: **Z** and **S**. Here, the ssDNA species include molecules that contain both ubp XNAs **P** and **B** (*PB*) or **S** and **Z** (*SZ*), molecules that have lost one ubp XNA but retained the other (*PN*, *NB*, *SN*, *NZ*), and molecules that have lost both ubp XNAs (*NN*(+), *NN*(-)). Individual fidelities on template **P**, **B**, **S**, and **Z** are φ_P, φ_B, φ_S, and φ_Z, respectively. Error rate is represented by ε_X, equal to $(1 - \varphi_X)$,

where $X = P$, B, S, or Z. Retention of both XNAs within a strand requires simultaneous correct replication at each position. Therefore, joint retention fidelities are multiplicative rather than additive ($\varphi_{PB} = \varphi_P \varphi_B$). This ODE system was used to model fraction species in **Fig. 6E**, with individual per-cycle fidelities set to $\varphi_P = \varphi_Z = 0.956$ and $\varphi_B = \varphi_S = 0.930$.

$$\frac{d}{dn}\begin{bmatrix} N_{PB} \\ N_{SZ} \\ N_{PN} \\ N_{NB} \\ N_{SN} \\ N_{NZ} \\ N_{NN(+)} \\ N_{NN(-)} \end{bmatrix} = \begin{bmatrix} 0 & \varphi_S\varphi_Z & 0 & 0 & 0 & 0 & 0 & 0 \\ \varphi_P\varphi_B & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_S\varphi_Z & 0 & 0 & 0 & \varphi_Z & 0 & 0 \\ 0 & \varphi_S\varepsilon_Z & 0 & 0 & \varphi_S & 0 & 0 & 0 \\ \varepsilon_P\varphi_B & 0 & 0 & \varphi_B & 0 & 0 & 0 & 0 \\ \varphi_P\varepsilon_B & 0 & \varphi_P & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_S\varepsilon_Z & 0 & 0 & \varepsilon_S & \varepsilon_Z & 1 & 0 \\ \varepsilon_P\varepsilon_B & 0 & \varepsilon_P & \varepsilon_B & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} N_{PB} \\ N_{SZ} \\ N_{PN} \\ N_{NB} \\ N_{SN} \\ N_{NZ} \\ N_{NN(+)} \\ N_{NN(-)} \end{bmatrix} \quad \textbf{[Eqn. 16]}$$

Case 2: Systems with both XNA-to-DNA mutations and XNA inversions. For the **Ds**:**Px** system, we derive a separate 6-letter model (A, T, G, C, X, Y) that accounts for two distinct error modes: (i) XNA-to-DNA transversions (*i.e.,* **Ds**:**Px**→**N**:**N**) and (ii) XNA strand inversions (*i.e.,* **Ds**:**Px**→**Px**:**Ds**). Since these error modes are independent, and analogous to **Eqn. 9**, the XNA fraction can be used to isolate the average error due XNA-to-DNA transversions, $\varepsilon_{\langle X,Y\rangle}^{trnv}$, using **Eqn. 17a-c**. Similar to **Eqn. 9b-9c**, we obtain $\varepsilon_{\langle X,Y\rangle}^{trnv}$ from the slope of a fit to **Eqn. 17d**.

$$f_{XNA(+)} = \frac{N_{X(+)} + N_{Y(+)}}{N_{X(+)} + N_{Y(+)} + N_{N(+)}} \textbf{[Eqn. 17a]}$$

$$f_{XNA(-)} = \frac{N_{X(-)} + N_{Y(-)}}{N_{X(-)} + N_{Y(-)} + N_{N(-)}} \textbf{[Eqn. 17b]}$$

$$\frac{f_{XNA(+)} + f_{XNA(-)}}{2} = \left(\frac{1 + (1 - \varepsilon_{\langle X,Y\rangle}^{trnv})}{2}\right)^n \textbf{[Eqn. 17c]}$$

$$log_2\left(\frac{f_{XNA(+)} + f_{XNA(-)}}{2}\right) = -n\log_2\left(\frac{1 + \left(1 - \varepsilon_{\langle X,Y\rangle}^{trnv}\right)}{2}\right) \textbf{[Eqn. 17d]}$$

To model XNA strand inversion errors, we use a 4x4 state transition model (**Fig. 6B**) to track the evolution of sense (+) and antisense (-) strands independently (**Eqn. 15**).

$$\mathbf{S}(n+1) = \mathbf{A} \times \mathbf{S}(n) \textbf{[Eqn. 18]}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \varepsilon_{xx}^{+} & \mu_{yx}^{+} \\ 0 & 1 & \mu_{xy}^{+} & \varepsilon_{yy}^{+} \\ \varepsilon_{xx}^{-} & \mu_{yx}^{-} & 1 & 0 \\ \mu_{xy}^{-} & \varepsilon_{yy}^{-} & 0 & 1 \end{bmatrix}$$

Here, μ_{yx} and μ_{xy} are the proper ubp pairing probabilities, while ε_{xx} and ε_{yy} are inversion probabilities. The ubp XNA-to-DNA transversion rates are assumed to be approximately equal and strand-invariant for both ubp XNAs ($\varepsilon_X^{trnv} = \varepsilon_Y^{trnv} = \varepsilon_{\langle X,Y\rangle}^{trnv}$). This allows us to use the measured average transversion error from **Eqn. 17** to isolate and fit the inversion error terms (**Eqn. 19a-d**).

$$\mu_{xy}^{+} = 1 - \varepsilon_{xx}^{+} - \varepsilon_{\langle X,Y\rangle}^{trnv} \text{ [Eqn. 19a]}$$
$$\mu_{xy}^{-} = 1 - \varepsilon_{xx}^{-} - \varepsilon_{\langle X,Y\rangle}^{trnv} \text{ [Eqn. 19b]}$$
$$\mu_{yx}^{+} = 1 - \varepsilon_{yy}^{+} - \varepsilon_{\langle X,Y\rangle}^{trnv} \text{ [Eqn. 19c]}$$
$$\mu_{yx}^{-} = 1 - \varepsilon_{yy}^{-} - \varepsilon_{\langle X,Y\rangle}^{trnv} \text{ [Eqn. 19d]}$$

These inversion error terms are computed by minimizing the mean squared error between observed and predicted fraction ubp XNA ratios across measured theoretical cycles. Inversion probabilities were fit to measured data either by using unconstrained error terms (**Fig. 6F**) or allowing only **Ds**:**Ds**-type mismatches (i.e., ε_{YY}^{+} and $\varepsilon_{YY}^{-} = 0$, **Supplementary Figure S39**).

Experimental design considerations for quantifying replication fidelity. Measurements of $\varphi_{\langle X,Y\rangle}$ from sequencing data require us to specify the number of theoretical cycles for each condition. Theoretical cycles are equivalent to number of DNA doubling events and correspond to a regular thermocycle with 100% amplification efficiency. While the number of cycles can be programmed, ubp XNA PCR amplification efficiency is generally unknown and likely < 100%. One commonly employed strategy is to estimate actual theoretical cycles by quantifying end-point DNA concentration to obtain an amplification-fold. Alternatively, the number of theoretical cycles can be specified by setting the initial the ratio of primer-to-template molecules in the reaction.[37] If all primers are consumed, the maximum number of theoretical cycles is simply set by **Eqn. 20**, where the C_{primer} and $C_{template}$ are molar concentrations and *n* is the number of theoretical cycles.

$$2^n = \frac{C_{primer}}{C_{template}} \quad \text{[Eqn. 20]}$$

If amplification efficiency (η) is < 1, theoretical cycles is less than thermal cycles by **Eqn. 21**.

$$(1 + \eta)^n = \frac{C_{primer}}{C_{template}} \quad \text{[Eqn. 21]}$$

By setting the number of programmed thermal cycles to be much greater than the expected number of theoretical cycles, PCR reactions are expected to proceed to completion and undergo an *n* number of doublings. Validity of this assumption was verified experimentally using gel electrophoresis (**Supplementary Fig. S12**).

To account for reduced efficiency of ubp XNA replication, the number of thermal cycles was programmed to 40 for **BS** amplification, which sets our tolerance of amplification efficiency as low as $\eta_{min,BS} \approx 0.39$. Gel electrophoresis of **BS** PCR products showed minimal side products or leftover primers (**Supplementary Fig. S9**). For amplification of **PZ** and **DsPx**, 30 thermal cycles

were used to improve reaction completion while reducing the formation of undesired amplification products, corresponding to minimum amplification efficiencies of $\eta_{min,PZ} \approx 0.55$ and $\eta_{min,DsPx} \approx 0.33$.

In PCR reactions, primer concentrations were set 0.2 μM to prevent polymerase-limiting conditions during exponential amplification, which could result in linear amplification. Apart from the **Ds**:**Px** system, for which 3-4 minute extension times during PCR are recommended, extension times were minimized to discourage non-ideal behavior, including primer degradation by polymerases with 3`-5` exonuclease activity. The minimum extension time used was 10s (**Supplementary Fig. S12**). Nucleoside triphosphates were generally added in excess (200 μM) to avoid amplification in nucleotide-limiting regimes.

Preparation of synthetic DNA controls. Synthetic positive controls (100% XNA) for complementary oligos **B1**-**5**,**S1**-**5** and **P1**,**Z1** were prepared by TA ligation of annealed oligos to barcoded hairpins (**Supplementary Fig. S8**). 5`-phosphorylated hairpins (**Supplementary Table 5**) were ordered from IDT with a 3`-T overhang protected from exonuclease degradation by a phosphorothioate bond. These hairpins contain a 24-mer barcode (**Supplementary Table 4**) in the annealed region and a *ScaI* restriction site near the hairpins to generate blunt ends before sequencing. Annealed oligos were 5`-phosphorylated and 3`-A-tailed using the NEBNext Ultra II End Repair/dA-tailing Module. End-prepped DNA was immediately ligated with T-tailed barcode hairpins using the NEBNext Ultra II Ligation Module. Ligation products were purified using AMPure XP beads at a 1.8X bead-to-sample ratio. To generate blunt-end products for sequencing, the purified ligation products were digested with *ScaI-HF* at 37 °C for 3 hours. After an additional AMPure cleanup, blunt-end products were ready for sequencing prep.

For 8-letter positive control DNA, a primer extension reaction was performed by annealing primer **R_14** (**Supplementary Table 3**) to the 6-letter template **PB1** (**Supplementary Table 2**) and extending with Q5 polymerase at 72 °C for 15 minutes in Q5 Reaction Buffer. dNTP stocks were added at the following final concentrations: dATP (0.2 mM), dGTP (0.2 mM), dCTP (0.2 mM), dTTP (0.2 mM), dSTP (0.6 mM), and dZTP (0.2 mM). An agarose gel for this extension is shown in **Supplementary Fig. S29A**.

For **DsPx** positive control DNA, primer **Ext_D1-2_30** (**Supplementary Table 3**) was annealed to templates **D1** and **D2**, while primer **Ext_D3-4_25** was annealed to templates **D3** and **D4**. A primer extension was performed using AccuPrime *Pfx* polymerase at 68 °C for 60 minutes. Deoxynucleoside triphosphate stocks were added at the following final concentrations: dATP (0.4 mM), dGTP (0.4 mM), dCTP (0.4 mM), dTTP (0.4 mM), and dDiol-PxTP (0.05 mM). An agarose gel for primer extension of templates **D1** and **D2** is shown in **Supplementary Fig. S29B**.

PCR amplification of the BS base pair for fidelity quantification. Complementary oligos containing **B** or **S** (**B1**-**B5** and **S1**-**S5**, **Supplementary Table 2**) were combined in a 1:1 molar ratio. A 10-fold serial dilution was prepared 10 μM to 10 pM to serve as templates for PCR reactions. For standard PCR with known template concentration, 1 μL of template was transferred to 25 μL PCR reactions to achieve the desired molar primer-to-template ratio. Reactions contained 0.2 μM of each forward and reverse barcoded PCR primer (**Supplementary Table 3 & 4**). For multi-step PCR, the unpurified PCR product mix from a completed reaction was diluted 1000X into the subsequent PCR reaction as DNA template. Polymerases and buffer were prepared as a

1X mixture. Thermopol Buffer was used for Taq polymerase, and Q5 reaction buffer was used for Q5 polymerase. Nucleoside triphosphates dATP, dGTP, dTTP, dCTP, dBTP, and dSTP were added to the reaction mix at a final concentration of 200 μM unless otherwise specified. PCR thermal cycler conditions were as follows: denature 2 minutes at 92 ℃ followed by 40 cycles of denature 92 ℃ for 15 s, anneal at 58 ℃ for 15 s, and extend at 72 ℃ for 10s with a final extension at 72 ℃ for 1 min. Samples were held at 4 ℃ upon PCR program completion and moved to -20 ℃ for storage.

Reproducing PCR amplification of BS from literature. PCR reactions in Johnson et al. (2004)[39] and Karalkar et al. (2016)[36] used the same buffer consisting of 10 mM Tris-HCl, 40 mM potassium acetate, 2 mM magnesium chloride, and 0.1 mg/mL bovine serum albumin at pH 9.1. A 5X concentrated stock of this buffer was prepared and brought to pH 9.1 using 1M HCl. To reproduce data from Johnson et al. (2004), reactions contained the custom buffer described above (1X) with Titanium *Taq* (1U), dNTPs (25 μM), dBTP (25 μM), dSTP (50 μM), 0.2 μM F/R primers, and equimolar amounts of complementary DNA templates **B5** and **S5** (200 pM, 20 pM, and 2 pM, 0.2 pM, **Supplementary Table 2**). To reproduce data from Karalkar et al. 2016, reactions contained the custom buffer described above (1X) with Titanium *Taq* (1U), dNTPs (100 μM), dBTP (100 μM), dSTP (100 μM), 0.02 μM F/R primers, and equimolar amounts of templates **B2** and **S2** (2 nM, 200 pM, 20 pM, and 2 pM) For each set of experiments, thermal cycling conditions used were identical to those reported by the authors.

PCR amplification of BS for pH, divalent cation, and supplemental additive screening. A 10X concentrated stock of Thermopol buffer without $MgSO_4$ was prepared consisting of 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 200 mM Tris base, and 0.1% Triton X-100 (v/v). For the pH screen, buffers were brought to varying pH values (pH 7.75, 8.00, 8.25, 8.50, 8.80, 9.00, & 9.25) using 1M HCl. 25 μL PCR reactions were prepared using these buffers at varying pH values (1X) and included Taq polymerase (1U/25 μL reaction), dATP (200 μM), dGTP (200 μM), dCTP (200 μM), dTTP (200 μM), dBTP (200 μM), dSTP (200 μM), and $MgSO_4$ (2 mM). Template **B1** & **S1** (**Supplementary Table 2**) were annealed and added at 4 pg/μL (11.4 theoretical cycles). F/R primers were added to each reaction at 0.2 μM.

For the divalent cation screen, PCR reactions were supplemented with $CoCl_2$ or $MnSO_4$ at final concentrations of either 0.2 mM or 0.6 mM, with and without additional $MgSO_4$ (0 mM or 2 mM). PCR reaction mixes contained Thermopol buffer (1X, pH 8.80), prepared as above without added $MgSO_4$, Taq polymerase (1U/25 μL reaction), dATP (200 μM), dGTP (200 μM), dCTP (200 μM), dTTP (200 μM), dBTP (200 μM), dSTP (200 μM), and annealed template **B1** & **S1** at 4 pg/μL (11.4 theoretical cycles). F/R primers were added to each reaction at 0.2 μM.

For PCR reaction additive screening, PCR reactions were supplemented with either DMSO (2-10%) or betaine (100-500 mM). PCR reaction mixes contained Thermopol buffer (1X, pH 8.80), prepared as above without added $MgSO_4$, Taq polymerase (1U/25 μL reaction), dATP (200 μM), dGTP (200 μM), dCTP (200 μM), dTTP (200 μM), dBTP (200 μM), dSTP (200 μM), and annealed template **B1** & **S1** at 4 pg/μL (11.4 theoretical cycles). F/R primers were added to each reaction at 0.2 μM.

All PCR experiments described above were completed with an internal reference condition using Taq polymerase and standard Thermopol buffer (pH = 8.80, 2 mM $MgSO_4$, no supplement). PCR thermal cycler conditions were as follows: denature 2 minutes at 92 ℃ followed by 40 cycles of denature 92 ℃ for 15 s, anneal at 58 ℃ for 15 s, and extend at 72 ℃ for 10s with a final extension at 72 ℃ for 1 min. Samples were held at 4 ℃ upon PCR program completion and then moved to -20 ℃ for storage.

PCR amplification of BS for polymerase screening. 25 µL PCR reactions were prepared containing 1X polymerase buffer, polymerase (1 U/25µL reaction), dNTPs (200 µM each dATP, dGTP, dCTP, dTTP, dBTP, and dSTP). PCR conditions for each polymerase were first screened using buffers recommended by the supplier, including the recommended supplementation of Mg^{2+}. For improved amplification performance with select polymerases, Mg^{2+} concentration was increased. Buffers and Mg^{2+} concentrations for each polymerase screened are listed in **Supplementary Table 9A**. PCR was performed at two template concentrations to verify the expected loss as a function of theoretical cycles. Fidelity was calculated using endpoint XNA fraction measurements and averaged across the two template concentrations. For Phusion, Q5, Deep Vent, *Tth*, *Pfu*, Deep Vent (exo-), and SD polymerases, template **B1** & **S1** (**Supplementary Table 2**) were annealed and used in screening reactions at 40 pg/µL and 0.4 pg/uL, corresponding to 8.1 and 14.8 theoretical cycles, respectively. For KlenTaq and AccuPrime *Pfx*, template **B1** & **S1** were added to screening reaction at equimolar concentrations (400 pM and 4 pM), corresponding to 9.0 and 15.6 theoretical cycles. Each experimental batch was completed with a reference PCR condition performed using Taq polymerase in Thermopol buffer. Across experimental batches, measured average replication fidelities for **BS** with Taq polymerase were between 86.9-88.8 ± 1.0% (standard deviation) per theoretical cycle. PCR thermal cycler conditions were as follows: denature 2 minutes at 92 ℃ followed by 40 cycles of denature 92 ℃ for 15 s, anneal at 58 ℃ for 15 s, and extend at 72 ℃ for 10s with a final extension at 72 ℃ for 1 min. Samples were held at 4 ℃ upon PCR program completion and then moved to -20 ℃ for storage.

PCR amplification of BS for nucleotide optimization. Reaction mixes were prepared containing equimolar amounts of **B1**, **S1** or **B2, S2** DNA templates (**Supplementary Table 2**). Nucleoside triphosphates were added at final concentrations from 0.05 mM to 0.3 mM to achieve desired nucleotide ratios, as listed in **Supplementary Table 9B**. 25 µL PCR reactions for Taq contained NEB Thermopol Buffer (1X) and Taq polymerase (1U/reaction) while Q5 reactions contained Q5 Reaction Buffer (1X) and Q5 polymerase (1U/reaction). Primers were added to a final concentration of 0.2 µM. For screening experiments where a single endpoint fidelity was measured, DNA template was added to a final concentration of 4 pM, corresponding to n = 15.6 theoretical cycles. A standard internal reference condition (200 µM dNTP, dSTP, dBTP) was included to account for batch effects in experiments. Reference condition average fidelity values using Taq polymerase with template **B1**, **S1** ranged from 87.2-88.1% ± 0.4% (standard deviation). Experimental batch variation of the reference condition is not included in the reported error measurements. PCR thermal cycler conditions were as follows: denature 2 minutes at 92 ℃, followed by 40 cycles of denature 92 ℃ for 15 s, anneal at 58 ℃ for 15 s, and extend at 72 ℃ for 10s, with a final extension at 72 ℃ for 1 min. Samples were held at 4 ℃ upon PCR program completion and then moved to -20 ℃ for storage.

PCR amplification of the PZ base pair for fidelity quantification. 10X PCR buffers were prepared containing 100 mM Tris-HCl, 500 mM KCl, and 15 mM $MgCl_2$. The pH was adjusted with 1M HCl to pH 8.80 and 8.00. PCR reactions contained prepared buffer (1X), Takara Taq polymerase (1.25U/reaction), and deoxynucleoside triphosphates at the following final concentrations: dATP (100 μM), dGTP (100 μM), dCTP (200 μM), dTTP (100 μM), dZTP (100 μM), and dPTP (200 μM). For PCR reactions utilizing optimized dPTP concentrations, dNTPs were added to the reaction mix at the following final concentrations: dATP (100 μM), dCTP (200 μM), dGTP (100 μM), dTTP (100 μM), dZTP (100 μM), and dPTP (500 μM). Two complementary oligos containing **P** and **Z** (**P1** & **Z1**, **Supplementary Table 2**) were combined in an equimolar molar ratio and diluted to achieve final template concentrations of 400 pM, 40 pM, 4 pM, and 0.4 pM. PCR reactions contained 0.2 μM of each barcoded PCR primer (**Supplementary Table 3 & 4**). Thermal cycler conditions were 2 minutes at 92 °C followed by 30 cycles of 90 °C for 15s, 58 °C for 15s, and 72 °C for 60s with a final extension of 10 minutes at 72 °C. Samples were held at 4 °C upon PCR program completion and then moved to -20 °C for storage.

8-Letter PCR. A DNA oligo containing both **P** and **B** (**PB1**, **Supplementary Table 2**) was designed to have the ±10 bases flanking **B** and **P** to be identical to the flanking regions in sequences **B2** and **P1**. For 8-letter PCR with Taq polymerase, reaction mixes were prepared using NEB Thermopol Buffer (1X) and Taq polymerase (1U/reaction). For 8-letter PCR with Takara Taq polymerase, reactions contained 10 mM Tris-HCl (pH 8.80), 50 mM KCl, and 1.5 mM $MgCl_2$, and Takara Taq polymerase (1.25U/reaction). For 8-letter PCR reactions, template **PB1** was used the following final concentrations: 4 nM, 400 pM, 40 pM, and 4 pM. To investigate the effect of template on fidelity quantification, template **P1**, rather than **PB1**, was added at the same final concentrations (**Supplementary Figure S34**). All PCR reactions contained nucleotides at the following final concentrations: dATP (200 μM), dCTP (200 μM), dGTP (100 μM), dTTP (100 μM), dBTP (200 μM), dSTP (200 μM), dZTP (200 μM), and dPTP (200 μM). F/R barcoded PCR primers were added at 0.2 μM (**Supplementary Table 3 & 4**). Thermal cycling conditions for all experiments described above were 2 minutes at 92 °C followed by 30 cycles of 90 °C for 15s, 58 °C for 15s, and 72 °C for 60s with a final extension of 10 minutes at 72 °C. Samples were held at 4 °C upon PCR program completion and then moved to -20 °C for storage.

PCR amplification of the DsPx base pair for fidelity quantification. Ds-containing oligos (**D1** & **D2**, **Supplementary Table 2**) were added to 25 μL PCR reactions at the following template concentrations: 4 nM, 1 nM, 400 pM, 100 pM, and 40 pM. F/R barcoded PCR primers were added at 0.2 μM (**Supplementary Table 3 & 4**). AccuPrime *Pfx* Polymerase and AccuPrime *Pfx* Reaction Mix were used at final concentrations of 1U/25 μL and 1X, respectively. The AccuPrime *Pfx* Reaction Mix contains 0.3 mM dNTPs and 1 mM $MgSO_4$. Additional dNTPs and $MgSO_4$ were supplemented to bring the final concentrations to 0.4 mM and 1.5 mM, respectively. Nucleoside triphosphates dDsTP and dDiol-PxTP were added at a final concentration of 50 μM. Thermal cycling conditions were 2 minutes at 94 °C followed by 30 cycles of 94 °C for 15s, 58 °C for 15s, and 68 °C for 3 minutes. There was a final 10 min extension at 68 °C, and the samples were held at 4 °C upon PCR program completion then moved to -20 °C for storage.

Parameterizing Ds:Px strand inversion from sequencing data. Strand-specific **Ds**:**Px** (**Px** = **Diol-Px**) basecalls were determined using three independent binary classification models, applied in series. Each model was trained to distinguish between a specific pair of bases at the

target position: **Ds** vs. **N**, **Px** vs. **N**, and **Px** vs. **Ds**. The three binary model outputs were then combined to define a unique classification pattern for each read (e.g., 0-0-1 or 1-1-0), representing the joint outcome across all three pairwise comparisons. Final basecalls were assigned using a predefined rule set that maps each classification pattern to a replication outcome using the truth table shown in **Supplementary Table 10**. Strand inversion classification performance on a ground-truth test set withheld from training is shown in **Supplementary Fig. S37**.

Normalization of XNA fraction and fraction species. Calculated XNA fractions were normalized using XNA fraction values obtained from internal standards (**Eqn. 22**), where $f_{X,norm}$ is the normalized XNA fraction, $f_{X,raw}$ is the measured XNA fraction, $f_{X,NC}$ is the XNA fraction measured for the negative control (0% XNA), and $f_{X,PC}$ is the XNA fraction measured for the positive control (100% XNA).

$$f_{X,norm} = \frac{f_{X,raw} - f_{X,NC}}{f_{X,PC} - f_{X,NC}} \qquad \textbf{[Eqn. 22]}$$

For 8-letter PCR experiments, the positive control internal standard was generated by primer extension from a 6-letter template (**PB1**, **Supplementary Table 2**), while the negative control internal standard was generated by PCR amplification of template **PB-GA1**. The fraction species f_{XY} (*XY* = *PB* or *SZ*) and f_{NN} (*NN* = *GA* or *TC*) were normalized by **Eqn. 22** using these two DNA internal standards. The remaining fraction species ($1 - f_{XY} - f_{NN}$) was assigned to f_{XN} (*XN* = *PA* or *TZ*) and f_{NY} (*NY* = *GB* or *SC*), preserving their measured relative proportions, *R* ($R = f_{XN,raw}/f_{NY,raw}$). Experimental results with and without normalization for the 8-letter PCR experiments are shown in **Supplementary Fig. S31**.

For **DsPx** strand inversion calculations, fraction species were normalized using **Eqn. 22**, with $f_{X,PC}$ and $f_{X,NC}$ for each species parameterized by multiclassification performance on ground truth datasets shown in **Supplementary Figure S37**. Raw fraction species results DsPx PCR prior to normalization are shown in **Supplementary Fig. S38**.

Statistics and error propagation. All PCR reactions were performed in experimental triplicates (*n*=3) as described in the relevant methods sub-sections. For each condition, the standard deviation was calculated for each measured XNA fraction. The average XNA fraction ($f_{\langle X,Y \rangle}$) is calculated by averaging measurements of f_X and f_Y for each replicate, effectively treating $f_{\langle X,Y \rangle}$ as the measured value for downstream calculations. Therefore, differences between the measurements of f_X and f_Y are not propagated to the reported error of fidelity calculations.

To determine the statistical significance between two groups of measurements, a two-tailed z-test was performed using **Eqn. 23**, where $\bar{x}_1$ and $\bar{x}_2$ are the means of group 1 and 2, respectively. s_1 and s_2 are the standard deviations of group 1 and 2, respectively.

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 + s_2^2}} \qquad \textbf{[Eqn. 23]}$$

Standard error of the mean (SEM) of the average XNA fraction ($\sigma_{f_{\langle X,Y\rangle}}$) is calculated according to **Eqn. 24**, where n is the number of replicates.

$$\sigma_{f_{\langle X,Y\rangle}} = \frac{s}{\sqrt{n}} \quad \textbf{[Eqn. 24]}$$

Standard error of the mean (SEM, σ) for endpoint fidelity calculations was propagated using **Eqn. 25**. This equation was derived by taking the derivative $\left(\frac{d\varphi_{\langle X,Y\rangle}}{df_{\langle X,Y\rangle}}\right)$ of **Eqn. 9a**. SEM of average XNA fraction is represented by $\sigma_{f_{\langle X,Y\rangle}}$, which is calculated according to **Eqn. 24**. SEM of fidelity is represented by $\sigma_{\varphi_{\langle X,Y\rangle}}$.

$$\sigma_{\varphi_{\langle X,Y\rangle}} = \left| 2 \cdot \frac{2^{(\log_2 f_{\langle X,Y\rangle})/n}}{n * f_{XNA}} \right| \cdot \sigma_{f_{\langle X,Y\rangle}} \quad \textbf{[Eqn. 25]}$$

Similarly, the standard error of the slope (σ_m) was propagated to the standard error of the mean of fidelity ($\sigma_{\varphi_{\langle X,Y\rangle}}$) using **Eqn. 26**. This equation was derived by taking the derivative $\left(\frac{d\varphi_{\langle X,Y\rangle}}{dm}\right)$ of fidelity ($\varphi_{\langle X,Y\rangle}$) with respect to slope (m) in **Eqn. 9d**.

$$\sigma_{\varphi_{\langle X,Y\rangle}} = | \, 2 \cdot \ln 2 \cdot 2^{-m} \, | \cdot \sigma_m \quad \textbf{[Eqn. 26]}$$

Slope, y-intercept, coefficient of determination (R^2), and slope standard error (Slope SE) were obtained using Weighted Least Squares (WLS) Regression implemented in `statsmodels.WLS` library (Python) with weights (w) defined as $w = 1/\text{SEM}^2$, where SEM is the standard error of the mean of experimental replicates used for linear regression. Average fidelity was calculated from the slope using **Eqn. 9d**, and slope SE was propagated to fidelity using **Eqn. 26**. The 95% confidence interval (CI) of the fidelity was obtained using a two-tailed t-distribution with $N - 2$ degrees of freedom, where N is the number of data points used for linear regression.

Supplementary Tables

Supplementary Table 1. Summary of XNA replication fidelity quantification literature. Summarized experimental details of previously reported replication fidelities from literature. 'Base Pair' column indicates the ubp XNA species which may form the same base pair but be structurally distinct than those used in this work. 'Pol' column indicates polymerase originally used. 'Assay' column provides a brief description of the original analytical method used to determine XNA fractions. 'Calc' column describes the type of calculation used to determine fidelity from measured data (other = from slope but with a linear rather than log-linear approximation). 'Recalculated Fidelity' columns show the value of the average fidelities (i.e., $\varphi_{\langle X,Y \rangle}$values) we recalculated using reported XNA fraction measurements. Recalculations were done using slope method (**Eqn. 9d**) unless noted by an asterisk. Experimental reproduction of measurements using conditions from originally published work are reported in 'Reproduced Fidelity' column. 'N/A' denotes details that were not available in original publication.

Base Pair	Pol	Context	Assay	Calc.	Reported Fidelity	Recalculated Fidelity	Reproduced Fidelity	Ref
isoG:isoC (BS)	Titanium *Taq*	CBG	Gel, acid cleavage	Other	96%	96.8%	96%	[39]
isoG:isoC (BS)	Titanium *Taq*	GBT	Gel, acid cleavage	Slope	93%	93%	N/A	[37]
B:Sn (BS)	Titanium *Taq*	GBT	Gel, acid cleavage	Endpoint	86%	66.5%	92.0%	[36]
P:Zn (PZ)	Jumpstart *Taq*	CPG	Gel, restriction digest	Endpoint	99.8%	N/A	N/A	[24]
P:Zn (PZ)	Takara Taq HS	Multiple	Enzyme-assisted sequencing	Endpoint	99.34%	*98.7%	97.9%	[32]
P:Zn (PZ)	Takara Taq HS	GPG	Gel, restriction digest	N/A	N/A	92.5%	N/A	[19]
Ds: NH2-hx-Px (DsPx)	AccuPrime *Pfx*	GDsA	Dideoxy dye term. sequencing	Endpoint	99.22%	*98.4%	N/A	[15]

'*' Fidelity was recalculated using reported endpoint values rather since multiple data points at varying theoretical cycles were not reported.

Supplementary Table 2. DNA template sequences used for model training and PCR. All sequences are shown in the 5′ to 3′ direction. Ubp XNA bases at the variable position are bolded in red and standard bases in black.

Template	Base	Oligonucleotide Sequence	Length
B1	B	GGTCTGGTGCCACTGGTCTGCTACGAGTGATGGTATTGGACTAGG**B**CTCT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
BA1	A	GGTCTGGTGCCACTGGTCTGCTACGAGTGATGGTATTGGACTAGG**A**CTCT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
S1	S^n	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAAGAG**S**CCTAG TCCAATACCATCACTCGTAGCAGACCAGTGGCACCAGACC	90
ST1	T	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAAGAG**T**CCTAG TCCAATACCATCACTCGTAGCAGACCAGTGGCACCAGACC	90
B2	B	GGTCTGGTGCCACTGGTCTGCTATGAGTGATCGTATTGGACTACG**B**TCGT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
BA2	A	GGTCTGGTGCCACTGGTCTGCTATGAGTGATCGTATTGGACTACG**A**TCGT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
BN2	N	GGTCTGGTGCCACTGGTCTGCTATGAGTGATCGTATTGGACTACG**N**TCGT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
S2	S^n	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAACGA**S**CGTAG TCCAATACGATCACTCATAGCAGACCAGTGGCACCAGACC	90
ST2	T	GTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAACGA**T**CGTAGT CCAATACGATCACTCATAGCAGACCAGTGGCACCAGACC	90
SN2	N	GTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAACGA**N**CGTAGT CCAATACGATCACTCATAGCAGACCAGTGGCACCAGACC	90
B3	B	GGTCTGGTGCCACTGGTCTGCTATGAGTGATCGTATTGGACTACG**B**ACGT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
BA3	A	GGTCTGGTGCCACTGGTCTGCTATGAGTGATCGTATTGGACTACG**A**ACGT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
S3	S^n	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAACGT**S**CGTAG TCCAATACGATCACTCATAGCAGACCAGTGGCACCAGACC	90
ST3	T	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAACGT**T**CGTAG TCCAATACGATCACTCATAGCAGACCAGTGGCACCAGACC	90
B4	B	GGTCTGGTGCCACTGGTCTGCTATGAGTGATCGTATTGGACTACG**B**GCGT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
BA4	A	GGTCTGGTGCCACTGGTCTGCTATGAGTGATCGTATTGGACTACG**A**GCGT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
S4	S^n	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAACGC**S**CGTAG TCCAATACGATCACTCATAGCAGACCAGTGGCACCAGACC	90
ST4	T	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAACGC**T**CGTAG TCCAATACGATCACTCATAGCAGACCAGTGGCACCAGACC	90
B5	B	GGTCTGGTGCCACTGGTCTGCTACGAGTGATGGTATTGGACTAGC**B**GTCT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
BA5	A	GGTCTGGTGCCACTGGTCTGCTACGAGTGATGGTATTGGACTAGC**A**GTCT TACTGTTGTAACGCACCCGATCACGATAGGCAACCACACC	90
S5	S^n	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAAGAC**S**GCTAG TCCAATACCATCACTCGTAGCAGACCAGTGGCACCAGACC	90
ST5	T	GGTGTGGTTGCCTATCGTGATCGGGTGCGTTACAACAGTAAGAC**T**GCTAG TCCAATACCATCACTCGTAGCAGACCAGTGGCACCAGACC	90
P1	P	GGTCTGGTGCCACTGGTAACTGGGACAGCTGAAGT**P**CAGTCAGCCAGGGA AACACGATAGGCAACCACACC	71
PG1	G	GGTCTGGTGCCACTGGTAACTGGGACAGCTGAAGT**G**CAGTCAGCCAGGGA AACACGATAGGCAACCACACC	71
PN1	N	GGTCTGGTGCCACTGGTAACTGGGACAGCTGAAGT**N**CAGTCAGCCAGGGA AACACGATAGGCAACCACACC	71
Z1	Z	GGTGTGGTTGCCTATCGTGTTTCCCTGGCTGACTG**Z**ACTTCAGCTGTCCC AGTTACCAGTGGCACCAGACC	71

ZC1	C	GGTGTGGTTGCCTATCGTGTTTCCCTGGCTGACTG**C**ACTTCAGCTGTCCC AGTTACCAGTGGCACCAGACC	71
ZN1	N	GGTGTGGTTGCCTATCGTGTTTCCCTGGCTGACTG**N**ACTTCAGCTGTCCC AGTTACCAGTGGCACCAGACC	71
D1	Ds	GTCTGGTGCCACTGGTCTGCTGTATTGGACTATG**D**AACTTACTGTTGTAA CGCACCCGA	60
DN1	N	GTCTGGTGCCACTGGTCTGCTGTATTGGACTATG**N**AACTTACTGTTGTAA CGCACCCGA	60
D2	Ds	GTCTGGTGCCACTGGTCTGCTGTATTGGACTATC**D**TACTTACTGTTGTAA CGCACCCGA	60
DN2	N	GTCTGGTGCCACTGGTCTGCTGTATTGGACTATC**N**TACTTACTGTTGTAA CGCACCCGA	60
D3	Ds	TCGGGTGCGTTACAACAGTAAGTT**D**CATAGTCCAATACAGCAGACCAGTG GCACCAGACC	60
D4	Ds	TCGGGTGCGTTACAACAGTAAATA**D**GATAGTCCAATACAGCAGACCAGTG GCACCAGACC	60
PB1	P, B	GGTCTGGTGCCACTGGTCAGCTGAAGT**P**CAGTCAGCCATTGGACTACG**B**T CGTTACTGTCACGATAGGCAACCACACC	78
PB-GA1	G, A	GGTCTGGTGCCACTGGTCAGCTGAAGT**G**CAGTCAGCCATTGGACTACG**A**T CGTTACTGTCACGATAGGCAACCACACC	78
PB-NN1	N, N	GGTCTGGTGCCACTGGTCAGCTGAAGT**N**CAGTCAGCCATTGGACTACG**N**T CGTTACTGTCACGATAGGCAACCACACC	78

Supplementary Table 3. Primer sequences. Primers used in this work are tabulated including primer name, barcode, and sequence. Red sequence font corresponds to barcode sequences listed in **Supplementary Table 4**. All sequences are shown in the 5′ to 3′ direction.

Primer	Barcode	Sequence
F_01	NB01	CGATTCCACAAAGACACCGACAACTTTCTTGGTCTGGTGCCACTGGT
F_02	NB02	CGATTCACAGACGACTACAAACGGAATCGAGGTCTGGTGCCACTGGT
F_04	NB04	CGATTCTAGGGAAACACGATAGAATCCGAAGGTCTGGTGCCACTGGT
F_05	NB05	CGATTCAAGGTTACACAAACCCTGGACAAGGGTCTGGTGCCACTGGT
F_06	NB06	CGATTCGACTACTTTCTGCCTTTGCGAGAAGGTCTGGTGCCACTGGT
F_12	NB12	CGGTTCTCCGATTCTGCTTCTTTCTACCTGGGTCTGGTGCCACTGGT
F_13	NB13	CGGTTCAGAACGACTTCCATACTCGTGTGAGGTCTGGTGCCACTGGT
F_14	NB14	CGGTTCAACGAGTCTCTTGGGACCCATAGAGGTCTGGTGCCACTGGT
F_15	NB15	CGGTTCAGGTCTACCTCGCTAACACCACTGGGTCTGGTGCCACTGGT
F_16	NB16	CGGTTCCGTCAACTGACAGTGGTTCGTACTGGTCTGGTGCCACTGGT
R_07	NB07	CGATTCAAGGATTCATTCCCACGGTAACACGGTGTGGTTGCCTATCGTG
R_08	NB08	CGATTCACGTAACTTGGTTTGTTCCCTGAAGGTGTGGTTGCCTATCGTG
R_09	NB09	CGATTCAACCAAGACTCGCTGTGCCTAGTTGGTGTGGTTGCCTATCGTG
R_10	NB10	CGATTCGAGAGGACAAAGGTTTCAACGCTTGGTGTGGTTGCCTATCGTG
R_11	NB11	CGATTCTCCATTCCCTCCGATAGATGAAACGGTGTGGTTGCCTATCGTG
R_17	NB17	GTCGAAACCCTCCAGGAAAGTACCTCTGATGGTGTGGTTGCCTATCGTG
R_18	NB18	GTCGAACCAAACCCAACAACCTAGATAGGCGGTGTGGTTGCCTATCGTG
R_19	NB19	GTCGAAGTTCCTCGTGCAGTGTCAAGAGATGGTGTGGTTGCCTATCGTG
R_20	NB20	GTCGAATTGCGTCCTGTTACGAGAACTCATGGTGTGGTTGCCTATCGTG
R_21	NB21	GTCGAAGAGCCTCTCATTGTCCGTTCTCTAGGTGTGGTTGCCTATCGTG
R_22	NB22	GTCGAAACCACTGCCATGTATCAAAGTACGGGTGTGGTTGCCTATCGTG
R_23	NB23	GTCGAACTTACTACCCAGTGAACCTCCTCGGGTGTGGTTGCCTATCGTG
R_24	NB24	GTCGAAGCATAGTTCTGCATGATGGGTTAGGGTGTGGTTGCCTATCGTG
R_31	NB31	GTCGAATGCGTACAGCAATCAGTTACATTGTCGGGTGCGTTACAACAGTAAG
R_32	NB32	GTCGAACCAGTAGAAGTCCGACAACGTCATTCGGGTGCGTTACAACAGTAAG
R_33	NB33	GTCGAACAGACTTGGTACGGTTGGGTAACTTCGGGTGCGTTACAACAGTAAG
Ext_D1-2	None	GGTGTGGTTGCCTATCGTGAATCGTACTCTTGGAAATCTCACTCGGGTGCGTTACAACAG
Ext_D1-2_30	NB30	AGACTTGAATTCAGTGAGGATCTACTTCGACCCAAATCTCACTCGGGTGCGTTACAACAG
Ext_D3-4_25	NB25	AGACTTGAATCATTGCGTTGCATACCCAACTTACAATCTCACGGTCTGGTGCCACTGGTCTGCTG
lacI_DS_R	None	GGTAGGCAGGGTCTTCCAGCTCAGAGAGTCAATTCAGGGTGG
lacI_US_F	None	ACTTACATTAATTGCGTTGCG
F_60bpt_12	NB12	GAGCTGGAAGACCCTGCCTACCCGCTGCGAGTAAGTTCCGATTCTGCTTCTTTCTACCTGGGTCTGGTGCCACTGGT

Supplementary Table 4. Nanopore barcode sequences. Barcode sequences used for multiplexing experiments. All sequences are shown in the 5′ to 3′ direction.

Barcode	Forward sequence	Reverse sequence
NB01	CACAAAGACACCGACAACTTTCTT	AAGAAAGTTGTCGGTGTCTTTGTG
NB02	ACAGACGACTACAAACGGAATCGA	TCGATTCCGTTTGTAGTCGTCTGT
NB03	CCTGGTAACTGGGACACAAGACTC	GAGTCTTGTGTCCCAGTTACCAGG
NB04	TAGGGAAACACGATAGAATCCGAA	TTCGGATTCTATCGTGTTTCCCTA
NB05	AAGGTTACACAAACCCTGGACAAG	CTTGTCCAGGGTTTGTGTAACCTT
NB06	GACTACTTTCTGCCTTTGCGAGAA	TTCTCGCAAAGGCAGAAAGTAGTC
NB07	AAGGATTCATTCCCACGGTAACAC	GTGTTACCGTGGGAATGAATCCTT
NB08	ACGTAACTTGGTTTGTTCCCTGAA	TTCAGGGAACAAACCAAGTTACGT
NB09	AACCAAGACTCGCTGTGCCTAGTT	AACTAGGCACAGCGAGTCTTGGTT
NB10	GAGAGGACAAAGGTTTCAACGCTT	AAGCGTTGAAACCTTTGTCCTCTC
NB11	TCCATTCCCTCCGATAGATGAAAC	GTTTCATCTATCGGAGGGAATGGA
NB12	TCCGATTCTGCTTCTTTCTACCTG	CAGGTAGAAAGAAGCAGAATCGGA
NB13	AGAACGACTTCCATACTCGTGTGA	TCACACGAGTATGGAAGTCGTTCT
NB14	AACGAGTCTCTTGGGACCCATAGA	TCTATGGGTCCCAAGAGACTCGTT
NB15	AGGTCTACCTCGCTAACACCACTG	CAGTGGTGTTAGCGAGGTAGACCT
NB16	CGTCAACTGACAGTGGTTCGTACT	AGTACGAACCACTGTCAGTTGACG
NB17	ACCCTCCAGGAAAGTACCTCTGAT	ATCAGAGGTACTTTCCTGGAGGGT
NB18	CCAAACCCAACAACCTAGATAGGC	GCCTATCTAGGTTGTTGGGTTTGG
NB19	GTTCCTCGTGCAGTGTCAAGAGAT	ATCTCTTGACACTGCACGAGGAAC
NB20	TTGCGTCCTGTTACGAGAACTCAT	ATGAGTTCTCGTAACAGGACGCAA
NB21	GAGCCTCTCATTGTCCGTTCTCTA	TAGAGAACGGACAATGAGAGGCTC
NB22	ACCACTGCCATGTATCAAAGTACG	CGTACTTTGATACATGGCAGTGGT
NB23	CTTACTACCCAGTGAACCTCCTCG	CGAGGAGGTTCACTGGGTAGTAAG
NB24	GCATAGTTCTGCATGATGGGTTAG	CTAACCCATCATGCAGAACTATGC
NB25	GTAAGTTGGGTATGCAACGCAATG	CATTGCGTTGCATACCCAACTTAC
NB26	CATACAGCGACTACGCATTCTCAT	ATGAGAATGCGTAGTCGCTGTATG
NB27	CGACGGTTAGATTCACCTCTTACA	TGTAAGAGGTGAATCTAACCGTCG
NB28	TGAAACCTAAGAAGGCACCGTATC	GATACGGTGCCTTCTTAGGTTTCA
NB29	CTAGACACCTTGGGTTGACAGACC	GGTCTGTCAACCCAAGGTGTCTAG
NB30	TCAGTGAGGATCTACTTCGACCCA	TGGGTCGAAGTAGATCCTCACTGA
NB31	TGCGTACAGCAATCAGTTACATTG	CAATGTAACTGATTGCTGTACGCA
NB32	CCAGTAGAAGTCCGACAACGTCAT	ATGACGTTGTCGGACTTCTACTGG
NB33	CAGACTTGGTACGGTTGGGTAACT	AGTTACCCAACCGTACCAAGTCTG
NB34	GGACGAAGAACTCAAGTCAAAGGC	GCCTTTGACTTGAGTTCTTCGTCC
NB35	CTACTTACGAAGCTGAGGGACTGC	GCAGTCCCTCAGCTTCGTAAGTAG
NB36	ATGTCCCAGTTAGAGGAGGAAACA	TGTTTCCTCCTCTAACTGGGACAT
NB37	GCTTGCGATTGATGCTTAGTATCA	TGATACTAAGCATCAATCGCAAGC
NB38	ACCACAGGAGGACGATACAGAGAA	TTCTCTGTATCGTCCTCCTGTGGT
NB39	CCACAGTGTCAACTAGAGCCTCTC	GAGAGGCTCTAGTTGACACTGTGG
NB40	TAGTTTGGATGACCAAGGATAGCC	GGCTATCCTTGGTCATCCAAACTA
NB41	GGAGTTCGTCCAGAGAAGTACACG	CGTGTACTTCTCTGGACGAACTCC
NB42	CTACGTGTAAGGCATACCTGCCAG	CTGGCAGGTATGCCTTACACGTAG
NB43	CTTTCGTTGTTGACTCGACGGTAG	CTACCGTCGAGTCAACAACGAAAG
NB44	AGTAGAAAGGGTTCCTTCCCACTC	GAGTGGGAAGGAACCCTTTCTACT
NB45	GATCCAACAGAGATGCCTTCAGTG	CACTGAAGGCATCTCTGTTGGATC
NB46	GCTGTGTTCCACTTCATTCTCCTG	CAGGAGAATGAAGTGGAACACAGC
NB47	GTGCAACTTTCCCACAGGTAGTTC	GAACTACCTGTGGGAAAGTTGCAC
NB48	CATCTGGAACGTGGTACACCTGTA	TACAGGTGTACCACGTTCCAGATG

Supplementary Table 5. Synthetic hairpin sequences used for TA ligation. Names and sequences of hairpins used to barcode synthetic DNA sequences without PCR amplification. All sequences are shown in the 5′ to 3′ direction.

Hairpin name	Sequence
NB25_OuterHP_T-Tail	/5Phos/GGTTCTCATTGCGTTGCATACCCAACTTACAGTACTGATCCGTTTCTCGGATCAGTACTGTAAGTTGGGTATG CAACGCAATGAGAACC*T
NB26_OuterHP_T-Tail	/5Phos/GGTTCTATGAGAATGCGTAGTCGCTGTATGAGTACTGATCCGTTTCTCGGATCAGTACTCATACAGCGACTAC GCATTCTCATAGAACC*T
NB27_OuterHP_T-Tail	/5Phos/GGTTCTTGTAAGAGGTGAATCTAACCGTCGAGTACTGATCCGTTTCTCGGATCAGTACTCGACGGTTAGATTC ACCTCTTACAAGAACC*T
NB28_OuterHP_T-Tail	/5Phos/GGTTCTGATACGGTGCCTTCTTAGGTTTCAAGTACTGATCCGTTTCTCGGATCAGTACTTGAAACCTAAGAAG GCACCGTATCAGAACC*T

/5Phos/ = 5′-PO_4
*N = phosphorothioate bond

Supplementary Table 6. Training metrics for comparative basecalling models. Training metrics and model names of comparative basecalling models used in this work. 'Base' column indicates the ubp XNA base for classification, while the 'Conf' column indicates the base the model was trained to distinguish. 'Context' column contains the sequence of the 5` and 3` surrounding nucleotides. The 'Reads' column contains the number of reads of each dataset (ubp XNA vs conf.) used for training. All datasets used for training were balanced 1:1. The remaining columns contain training and validation output metrics. 'Epoch' column indicates how many epochs were used for training. Accuracy and loss at of the best model after convergence for both training and internal validation are shown in respective columns. Asterisk '*' denotes model trained with a '75 75' chunk context, and double asterisk '**' denotes model trained with a '125 125' chunk context, rather than the default '50 50' context. Red italicized font corresponds to training datasets in which the XNA was incorporated via primer extension and therefore may contain misincorporation of standard DNA bases.

Model	Base	Conf.	Context	Training Set Size	Epoch	Trn. Acc.	Trn. Loss	Val. Acc.	Val. Loss
BA1	B	A	AGGBCTC	35583	23	0.979	0.071	0.971	0.118
ST1	S	T	GAGSCCT	46836	14	0.999	0.006	0.998	0.013
BA2	B	A	ACGBTCG	50904	21	0.999	0.004	0.994	0.026
ST2	S	T	CGASCGT	35472	11	0.997	0.016	0.993	0.029
BN2	B	N	ACGBTCG	93091	16	0.996	0.013	0.985	0.052
SN2	S	N	CGASCGT	60568	14	0.996	0.012	0.992	0.029
BA3	B	A	ACGBACG	26151	19	0.998	0.006	0.991	0.035
ST3	S	T	CGTSCGT	42850	16	0.993	0.024	0.988	0.024
BA4	B	A	ACGBGCG	29337	15	0.991	0.028	0.977	0.070
ST4	S	T	CGCSCGT	50202	19	0.997	0.011	0.993	0.037
BA5	B	A	AGCBGTC	167201	16	0.997	0.010	0.986	0.051
ST5	S	T	GACSGCT	103229	12	0.995	0.019	0.993	0.027
**PG1	P	G	AGTPCAG	88846	20	0.994	0.018	0.986	0.050
ZC1	Z	C	CTGZACT	47018	15	0.998	0.007	0.992	0.032
ZN1	Z	N	CTGZACT	82751	20	0.998	0.007	0.992	0.034
8L-BN2	B	N	ACGBTCG	126523	19	0.993	0.020	0.983	0.058
8L-SN2	*S*	N	CGASCGT	151512	16	0.992	0.030	0.987	0.049
*8L-PG1	P	G	AGTPCAG	113521	19	0.998	0.006	0.990	0.040
8L-ZC1	*Z*	C	CTGZACT	45454	26	0.998	0.004	0.987	0.055
DA1	Ds	A	ATGDsAAC	30248	17	0.993	0.026	0.992	0.033
XT1	*Px*	T	GTTPxCAT	8632	17	0.997	0.006	0.997	0.016
DN1	Ds	N	ATGDsAAC	96202	20	0.997	0.009	0.993	0.027
XN1	*Px*	N	GTTPxCAT	86708	10	0.996	0.015	0.995	0.017
I-XN1	*Px*	N	ATGDsAAC	282792	21	0.997	0.009	0.992	0.031
I-DN1	Ds	N	GTTPxCAT	265894	16	0.997	0.008	0.992	0.028
XD1	*Px*	Ds	ATGDsAAC	128084	11	0.997	0.008	0.994	0.019
DX1	Ds	*Px*	GTTPxCAT	114967	24	0.999	0.002	0.992	0.036
DN2	Ds	N	ATCDsTAC	18152	9	0.995	0.022	0.992	0.026
XN2	*Px*	N	GTAPxGAT	15698	13	0.997	0.012	0.993	0.030
I-XN2	*Px*	N	ATCDsTAC	57151	19	0.998	0.010	0.993	0.030
I-DN2	Ds	N	GTAPxGAT	15179	15	0.996	0.013	0.991	0.031
XD2	*Px*	Ds	ATCDsTAC	22642	26	0.999	0.001	0.991	0.040
DX2	Ds	*Px*	GTAPxGAT	12817	26	0.991	0.034	0.957	0.206

Supplementary Table 7. Benchmarking metrics for comparative basecalling models. 'Base' column indicates the ubp XNA base for classification, while the 'Conf' column indicates base the model was trained to distinguish. 'Context' column contains the sequence of the 5` and 3` surrounding nucleotides. 'XNA' and 'DNA'. 'Reads' column contains the number of reads used for each training label. Remaining columns contain validation outputs metrics: TP = True Positives; FN = False Negatives; FP = False Positives; TN = True Negatives. Recall, specificity (Spec.), precision (Prec.), and F1-score (F1) are defined in "**Model Testing and Benchmarking**" section. Asterisk '*' denotes model trained with a '75 75' chunk context, and double asterisk '**' denotes model trained with a '125 125' chunk context, rather than the default '50 50' context. Red italicized font corresponds to testing datasets in which the XNA was incorporated via primer extension and therefore may contain misincorporation of standard DNA introduced during strand synthesis.

Model	Base	Conf.	Context	Reads	TP	FN	FP	TN	Recall	Spec.	Prec.	F1
BA1	B	A	AGGBCTC	10000	9454	546	66	9934	0.945	0.993	0.993	0.969
ST1	S	T	GAGSCCT	10000	9962	38	30	9970	0.996	0.997	0.997	0.997
BA2	B	A	ACGBTCG	10000	9926	74	78	9922	0.993	0.992	0.992	0.992
ST2	S	T	CGASCGT	10000	9897	103	35	9965	0.990	0.997	0.996	0.993
BN2	B	N	ACGBTCG	10000	9854	146	168	9832	0.985	0.983	0.983	0.984
SN2	S	N	CGASCGT	5000	4952	48	29	4971	0.990	0.994	0.994	0.992
BA3	B	A	ACGBACG	10000	9891	109	109	9891	0.989	0.989	0.989	0.989
ST3	S	T	CGTSCGT	10000	9840	160	160	9840	0.984	0.984	0.984	0.984
BA4	B	A	ACGBGCG	10000	9758	242	390	9610	0.976	0.961	0.962	0.969
ST4	S	T	CGCSCGT	10000	9885	115	74	9926	0.989	0.993	0.993	0.991
BA5	B	A	AGCBGTC	10000	9837	163	152	9848	0.984	0.985	0.985	0.984
ST5	S	T	GACSGCT	10000	9898	102	50	9950	0.990	0.995	0.995	0.992
**PG1	P	G	AGTPCAG	10000	9813	187	184	9816	0.981	0.982	0.982	0.981
ZC1	Z	C	CTGZACT	10000	9914	86	101	9899	0.991	0.990	0.990	0.991
ZN1	Z	N	CTGZACT	10000	9876	124	106	9894	0.988	0.989	0.989	0.988
8L-BN2	B	N	ACGBTCG	7500	7289	211	59	7441	0.972	0.992	0.992	0.982
8L-SN2	*S*	N	CGASCGT	7500	7369	131	63	7437	0.983	0.992	0.992	0.987
*8L-PG1	P	G	AGTPCAG	7500	7171	329	152	7348	0.956	0.980	0.979	0.968
8L-ZC1	*Z*	C	CTGZACT	7500	7127	373	85	7415	0.950	0.989	0.988	0.969
DA1	Ds	A	ATGDsAAC	10000	9950	50	41	9959	0.995	0.996	0.996	0.995
XT1	*Px*	T	GTTPxCAT	2500	2485	15	9	2491	0.994	0.996	0.996	0.995
DN1	Ds	N	ATGDsAAC	10000	9923	77	41	9959	0.992	0.996	0.996	0.994
XN1	*Px*	N	GTTPxCAT	10000	9928	72	47	9953	0.993	0.995	0.995	0.994
I-XN1	*Px*	N	ATGDsAAC	10000	9939	61	97	9903	0.994	0.990	0.990	0.992
I-DN1	Ds	N	GTTPxCAT	10000	9927	73	121	9879	0.993	0.988	0.988	0.990
XD1	*Px*	Ds	ATGDsAAC	10000	9915	85	82	9918	0.992	0.992	0.992	0.992
DX1	Ds	*Px*	GTTPxCAT	10000	9906	94	68	9932	0.991	0.993	0.993	0.992
DN2	Ds	N	ATCDsTAC	7500	7421	79	27	7473	0.989	0.996	0.996	0.993
XN2	*Px*	N	GTAPxGAT	800	792	8	5	795	0.990	0.994	0.994	0.992
I-XN2	*Px*	N	ATCDsTAC	7500	7424	76	34	7466	0.990	0.995	0.995	0.993
I-DN2	Ds	N	GTAPxGAT	800	790	10	5	795	0.988	0.994	0.994	0.991
XD2	*Px*	Ds	ATCDsTAC	7500	7397	103	74	7426	0.986	0.990	0.990	0.988
DX2	Ds	*Px*	GTAPxGAT	800	768	32	40	760	0.960	0.950	0.950	0.955

Supplementary Table 8. Weighted least squares (WLS) regression statistics for slope-based fidelity quantification. Slope, y-intercept, coefficient of determination (RSQ), and slope standard error (Slope SE) were obtained using Weighted Least Squares (WLS) Regression implemented in `statsmodels.WLS` (Python) with weights (w) defined as $w = 1/\mathrm{SEM}^2$, where SEM is the standard error of the mean for each group of experimental replicates (n=3) used for linear regression. Fidelity was calculated from the slope using **Eqn. 9d**. Slope SE was propagated using **Eqn. 26** to obtain the 95% confidence interval (CI) of the fidelity.

Ubp	Figure	Label	Temp.	Pol.	Slope	Intercept	RSQ	Slope SE	Fid. (φ)	±95% CI
BS	3C	⟨B,S⟩	B1,S1	Taq	0.0817	0.0649	0.999	0.0013	0.8899	0.0041
BS	4C	5`-GBC	B1,S1	Taq	0.0809	0.0622	0.996	0.0016	0.8909	0.0091
BS	4C	5`-GBC	B1,S1	Q5	0.0313	-0.0664	0.999	0.0006	0.9571	0.0037
BS	4C	5`-GBT	B2,S2	Taq	0.0498	0.1335	0.998	0.0016	0.9322	0.0092
BS	4C	5`-GBT	B2,S2	Q5	0.0337	0.0192	0.999	0.0006	0.9538	0.0033
BS	4C	5`-GBA	B3,S3	Taq	0.0656	0.1463	0.998	0.0022	0.9111	0.0124
BS	4C	5`-GBA	B3,S3	Q5	0.0266	0.0567	0.992	0.0017	0.9635	0.0098
BS	4C	5`-GBG	B4,S4	Taq	0.0646	0.1837	0.985	0.0057	0.9125	0.0324
BS	4C	5`-GBG	B4,S4	Q5	0.0418	0.1220	0.996	0.0019	0.9428	0.0109
BS	4C	5`-CBG	B5,S5	Taq	0.0635	0.2452	0.997	0.0026	0.9138	0.0149
BS	4C	5`-CBG	B5,S5	Q5	0.0407	0.0500	0.999	0.0009	0.9443	0.0050
BS	S25D	-	B5,S5	Ti. Taq	0.0301	0.0553	0.998	0.0009	0.9587	0.0053
BS	S26E	-	B2,S2	Ti. Taq	0.0587	0.2018	0.998	0.0020	0.9203	0.0117
BS	S27B	-	B2,S2	Q5	0.0126	-0.0202	0.998	0.0004	0.9826	0.0024
BS	S27E	-	B2,S2	Q5	0.0060	0.0009	0.979	0.0008	0.9917	0.0152
PZ	5C	pH 8.0	P1,Z1	Takara	0.0188	-0.0348	0.997	0.0009	0.9741	0.0040
PZ	5C	pH 8.8	P1,Z1	Takara	0.0369	-0.0150	0.999	0.0007	0.9494	0.0042
PZ	5C	dPTP opt	P1,Z1	Takara	0.0151	-0.0474	0.998	0.0007	0.9791	0.0030
BS,PZ	5F	⟨B,S|P,Z⟩	PB1	Taq	0.0832	0.0732	0.998	0.0029	0.8879	0.0162
BS,PZ	S30B	⟨B,S⟩	PB1	Taq	0.0516	0.1055	0.999	0.0016	0.9297	0.0068
BS,PZ	S30B	⟨P,Z⟩	PB1	Taq	0.0323	0.0790	0.996	0.0014	0.9558	0.0083
BS,PZ	S32C	⟨B,S|P,Z⟩	PB1	Takara	0.0708	0.0952	0.993	0.0043	0.9042	0.0243
BS,PZ	S32E	⟨B,S⟩	PB1	Takara	0.0529	0.1577	0.987	0.0056	0.9280	0.0243
BS,PZ	S32E	⟨P,Z⟩	PB1	Takara	0.0237	0.0062	0.998	0.0008	0.9674	0.0047
BS	S33B	-	B2,S2	Taq	0.0363	0.1235	0.995	0.0019	0.9503	0.0111
PZ	S33D	-	P1,Z1	Taq	0.0265	0.0908	0.999	0.0006	0.9637	0.0036
PZ	S34B	-	P1	Taq	0.0340	0.1560	0.960	0.0066	0.9534	0.0284
DsPx	6E	*Pu*-Ds-*Pu*	D1	AP *Pfx*	0.0192	-0.0057	0.947	0.0026	0.9736	0.0115
DsPx	6E	*Py*-Ds-*Py*	D2	AP *Pfx*	0.0052	0.0033	0.950	0.0007	0.9928	0.0030

Supplementary Table 9. Screening for optimizing BS replication fidelity.

9A. Polymerase screening details. Polymerase, supplier, and supplied buffers used and divalent cation additives ([Mg2+]) used for each polymerase in the BS polymerase screen.

Polymerase	Supplier	Buffer	[Mg2+] (mM)
Q5	NEB	Q5 Reaction Buffer (5 X)	2.0
Phusion	NEB	Phusion HF Buffer (5 X)	1.5
DeepVent	NEB	Thermopol Buffer (10 X)	10.0
DeepVent (exo-)	NEB	Thermopol Buffer (10 X)	2.0
OneTaq	NEB	OneTaq Standard Buffer (5 X)	1.8
Pfu	G-Biosciences	Pfu Buffer (10 X)	2.0
Tth	Boca Scientific	Tth PCR buffer (10 X)	5.5
SD	Boca Scientific	SD Polymerase Buffer (10 X)	3.0
AccuPrime *Pfx*	Invitrogen	AccuPrime *Pfx* Buffer Mix	1.0
Taq	G-Biosciences	Thermopol Buffer (10 X)	2.0
KlenTaq1	AB Bioscience	KlenTaq1 Buffer (10 X)	3.5

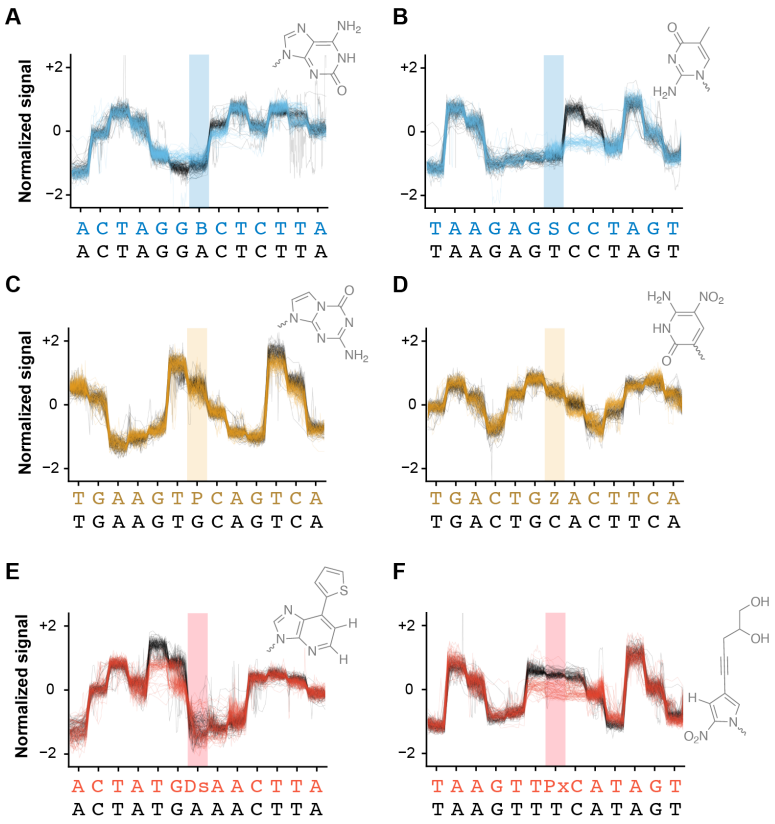
9B. Nucleotide screening ratios. 'Label' column indicates the varied nucleotide ratio indicated in the 'Ratio' column. Individual final concentrations of nucleotide (μM) used are listed in respective columns (dBTP, dSTP, dATP, dTTP, dGTP, dCTP).

		Nucleotide Concentration (μM)					
Label	Ratio	dBTP	dSTP	dATP	dTTP	dGTP	dCTP
B/A	0.5	100	200	200	200	200	200
B/A	1	200	200	200	200	200	200
B/A	2	200	200	100	200	200	200
B/A	4	200	200	50	200	200	200
B/A	6	300	200	50	200	200	200
S/T	0.5	200	100	200	200	200	200
S/T	1	200	200	200	200	200	200
S/T	2	200	200	200	100	200	200
S/T	4	200	200	200	50	200	200
S/T	6	200	300	200	50	200	200
S/T	8	200	400	200	50	200	200
S/T	10	200	500	200	50	200	200
B/A and S/T	0.5	100	100	200	200	200	200
B/A and S/T	1	200	200	200	200	200	200
B/A and S/T	2	200	200	100	100	200	200
B/A and S/T	4	200	200	50	50	200	200
B/A and S/T	6	300	300	50	50	200	200
B/A and S/T	8	400	400	50	50	200	200
B/A and S/T	10	500	500	50	50	200	200

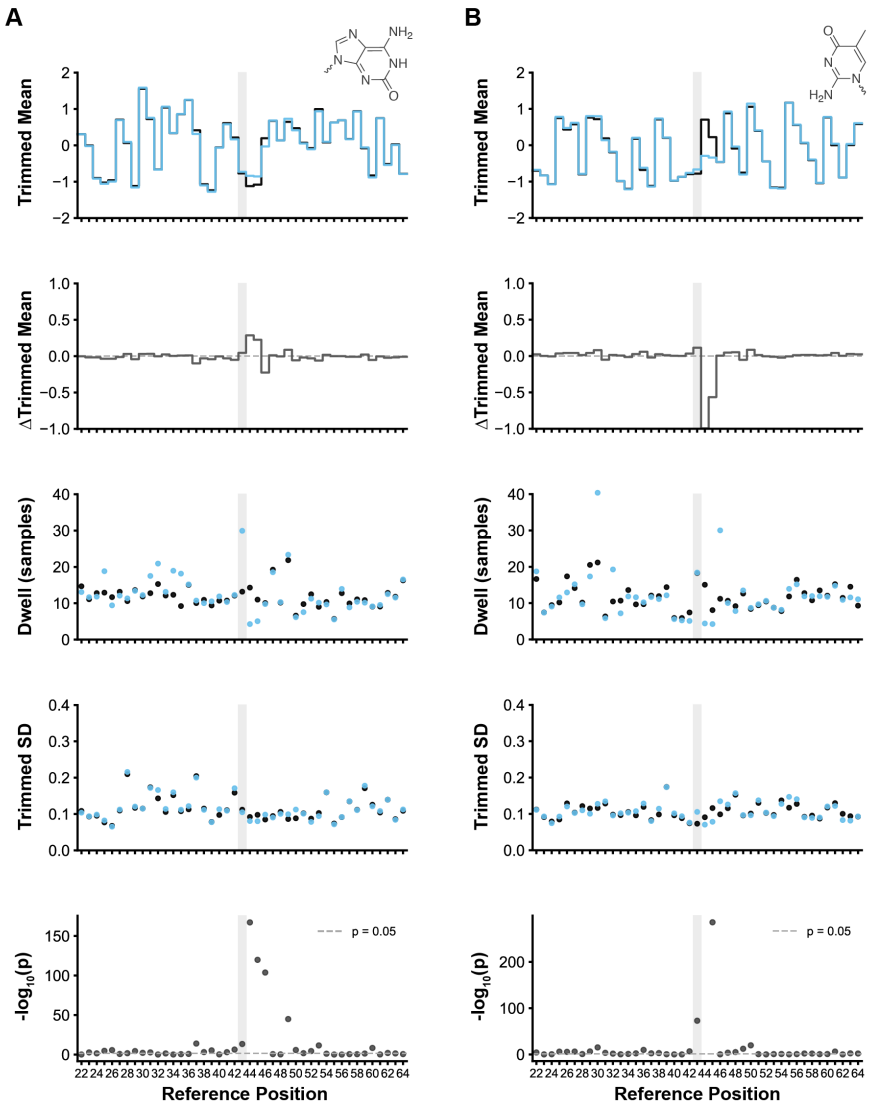
Supplementary Table 10. Strand inversion multiclassification mapping. Multiple strand-specific models were trained on sense and antisense strand data to identify inversion events and distinguish them from XNA retention or DNA transversion ('trnv'). The strand-specific model names (**Supplementary Table 6 & 7**) are listed in column headers. Each model was applied independently to the corresponding strand at the XNA position to determine comparative calls (**Ds** vs **N**, **Px** vs **N**, and **Ds** vs **Px**).

(+) Models				(-) Models					
DN1/2	**I-XN1/2**	**XD1/2**	**Basecall**	**XN1/2**	**I-DN1/2**	**DX1/2**	**Basecall**	**Output**	**Outcome**
N	N	Ds	N	N	N	Px	N	000	DNA trnv
N	N	Px	N	N	N	Ds	N	001	DNA trnv
N	Px	Ds	N	N	Ds	Px	N	010	DNA trnv
N	Px	Px	Px	N	Ds	Ds	Ds	011	Inverted
Ds	N	Ds	Ds	Px	N	Px	Px	100	Retained
Ds	N	Px	N	Px	N	Ds	N	101	DNA trnv
Ds	Px	Ds	Ds	Px	Ds	Px	Px	110	Retained
Ds	Px	Px	Px	Px	Ds	Ds	Ds	111	Inverted

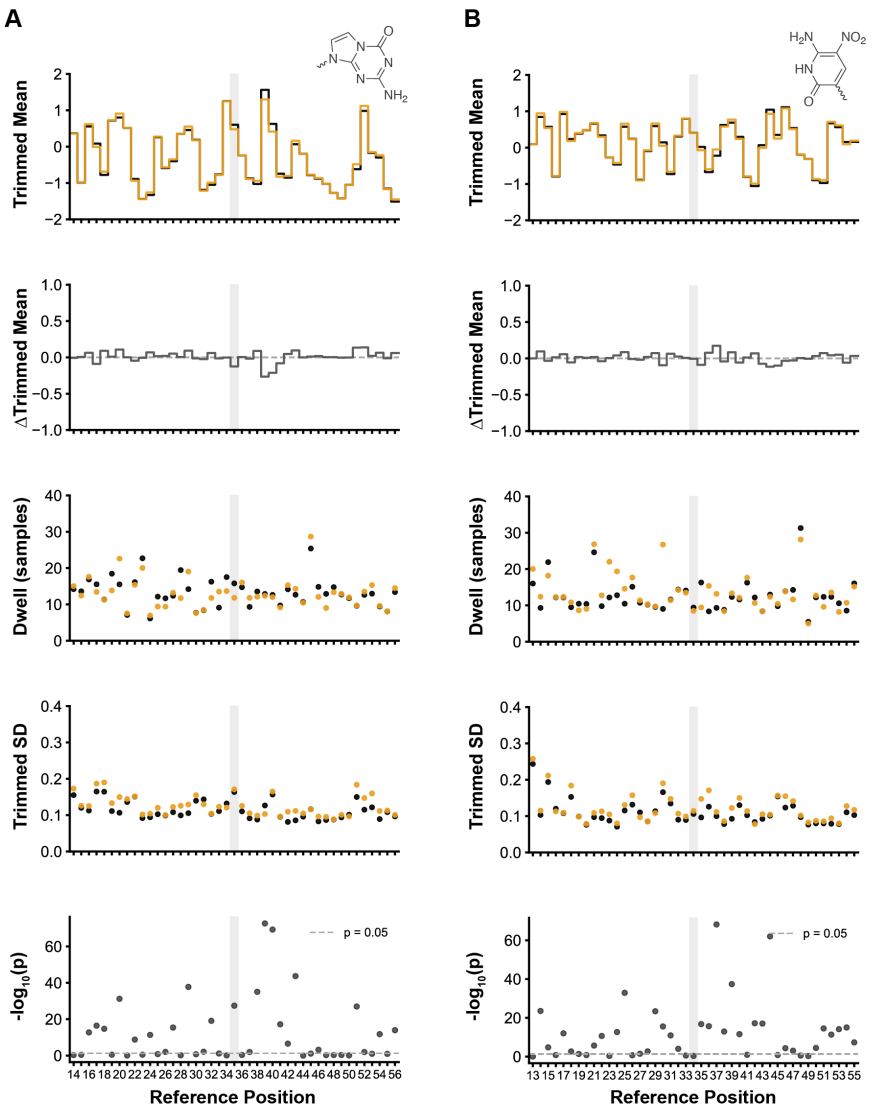
Supplementary Figures



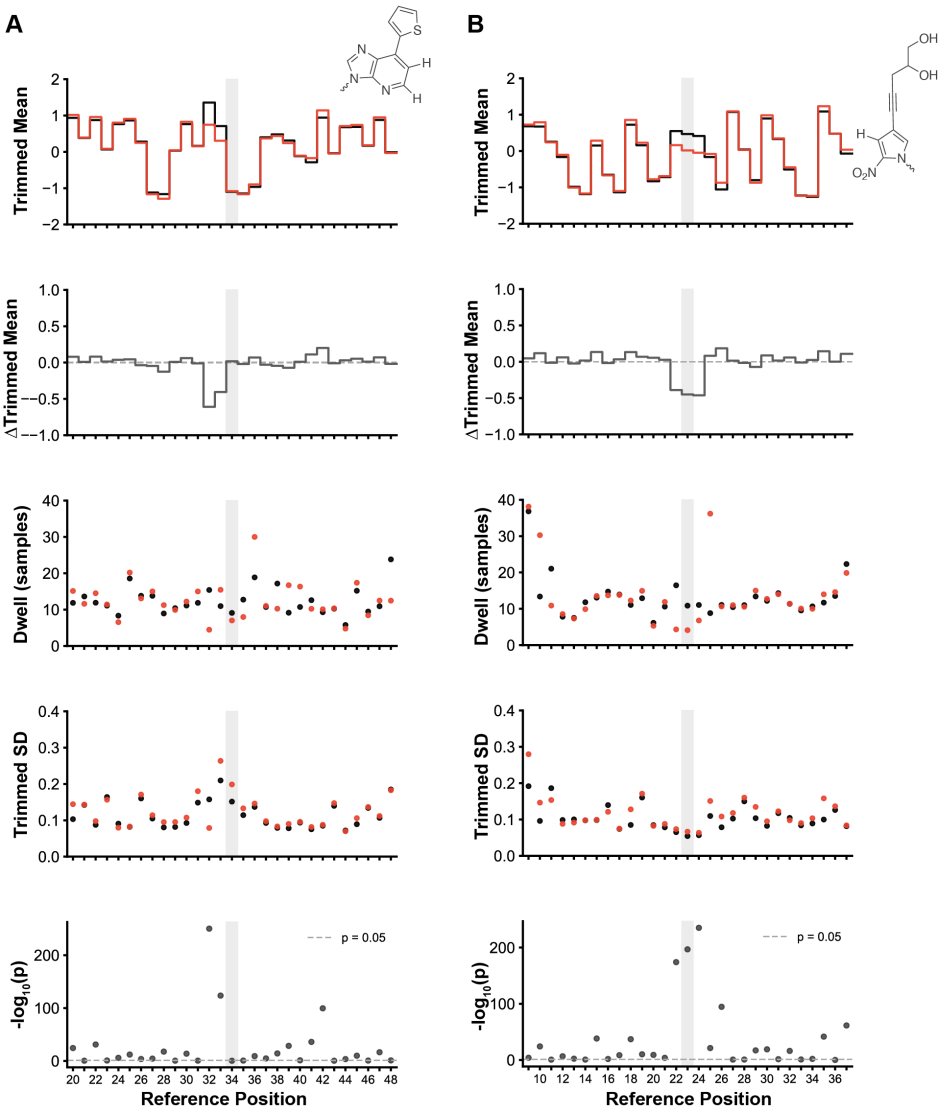
Supplementary Figure S1. Sample ionic current signal traces for ubp XNAs in single sequence contexts. Per-read ionic current signal traces are shown for reads containing a single ubp XNA, or a substituted canonical DNA base, within an otherwise identical sequence context. Raw nanopore signal was extracted from POD5 files and mapped to reference coordinates using aligned BAM files. Signal-to-sequence mapping was refined using Remora with default 9mer k-mer model and rough rescaling without iterative refinement. For each sequence, 75 reads were randomly sampled and plotted on a segmented-time scale, aligned to the corresponding reference base positions. The modified base position is highlighted, and six flanking bases are shown symmetrically on either side. All sequences are written in the 5` to 3` direction. **(A) B** versus **A**, in the context corresponding to model BA1. **(B) S** versus **T**, in the context corresponding to model ST1. **(C) P** versus **G**, in the context corresponding to PG1**. (D) Z** versus **C**, in the context corresponding to ZC1. **(E) Ds** versus **N** (**N** = **A**, **T**, **G**, or **C**), in the context corresponding to DN2. **(F) Diol-Px** versus **N**, in the context corresponding to XN2.



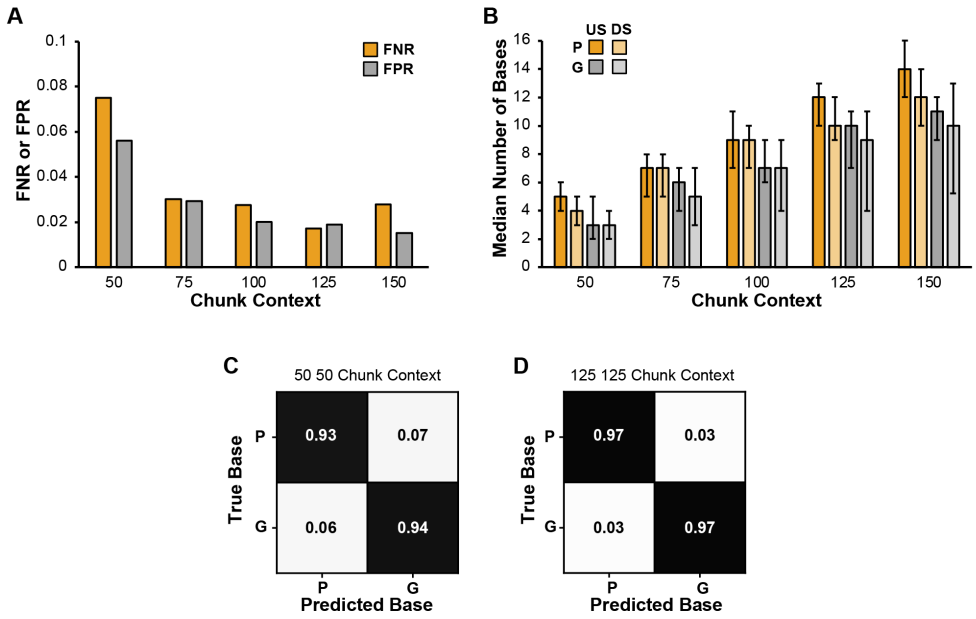
Supplementary Figure S2. Comparison of extracted signal metrics for the ubp XNA B/S vs DNA in single-sequence contexts. Two dsDNA sequences differing only in a single position, **(A) B/A** and **(B) S/T**, were basecalled corresponding to sequences used to train BA1 and ST1 models, respectively. Signal metrics for 500 randomly sampled reads after Remora signal refinement were extracted. Metrics are calculated from normalized current values. Plots show comparison between DNA base (black) and XNA base (blue) for signal trimmed mean ionic current (Trimmed Mean), difference between DNA and XNA trimmed mean ionic current (ΔTrimmed Mean), dwell time in terms of number of average sample points collected (Dwell), standard deviation of normalized signal within the signal region (Trimmed SD), and statistical significance of a Welch's t-test performed on DNA vs XNA for trimmed mean current values at the given position ($-\log_{10}(p)$). Gray highlight denotes the location of the ubp XNA within the original reference sequence.



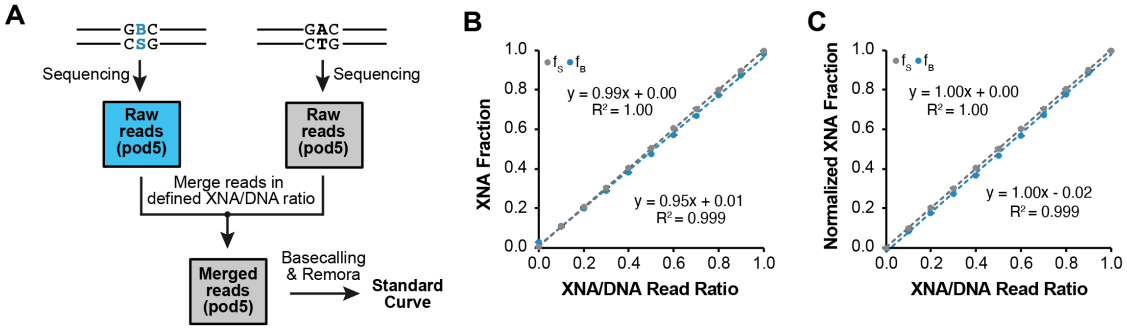
Supplementary Figure S3. Comparison of extracted signal metrics for the ubp XNA P/Z vs DNA in single-sequence contexts. Two dsDNA sequences differing only in a single position, **(A) P/G** and **(B) Z/C**, were basecalled corresponding to sequences used to train PG1 and ZC1 models, respectively. Signal metrics for 500 randomly sampled reads after Remora signal refinement were extracted. Metrics calculated from normalized current values. Plots show comparison between DNA base (black) and XNA base (orange) for signal trimmed mean ionic current (Trimmed Mean), difference between DNA and XNA trimmed mean ionic current (ΔTrimmed Mean), dwell time in terms of number of average sample points collected (Dwell), standard deviation of normalized signal within the signal region (Trimmed SD), and statistical significance of a Welch's t-test performed on DNA vs XNA for trimmed mean current values at the given position ($-\log_{10}(p)$). Gray highlight denotes the location of the ubp XNA within the original reference sequence.



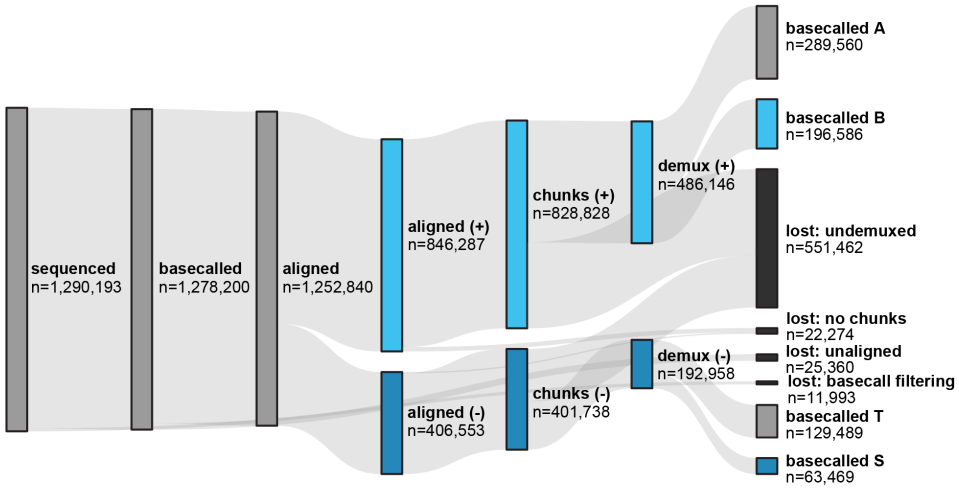
Supplementary Figure S4. Comparison of extracted signal metrics for the ubp XNA Ds/Diol-Px vs DNA in single-sequence contexts. Two dsDNA sequences differing only in a single position, **(A) Ds/A** and **(B) Diol-Px/T**, were basecalled, corresponding to sequences used to train models DA1 and XT1, respectively. Signal metrics for 500 randomly sampled reads after Remora signal refinement were extracted. Metrics calculated from normalized current values. Plots show comparison between DNA base (black) and XNA base (red) for signal trimmed mean ionic current (Trimmed Mean), difference between DNA and XNA trimmed mean ionic current (ΔTrimmed Mean), dwell time in terms of number of average sample points collected (Dwell), standard deviation of normalized signal within the signal region (Trimmed SD), and statistical significance of a Welch's t-test performed on DNA vs XNA for trimmed mean current values at the given position ($-\log_{10}(p)$). Gray highlight denotes the location of the ubp XNA within the original reference sequence.



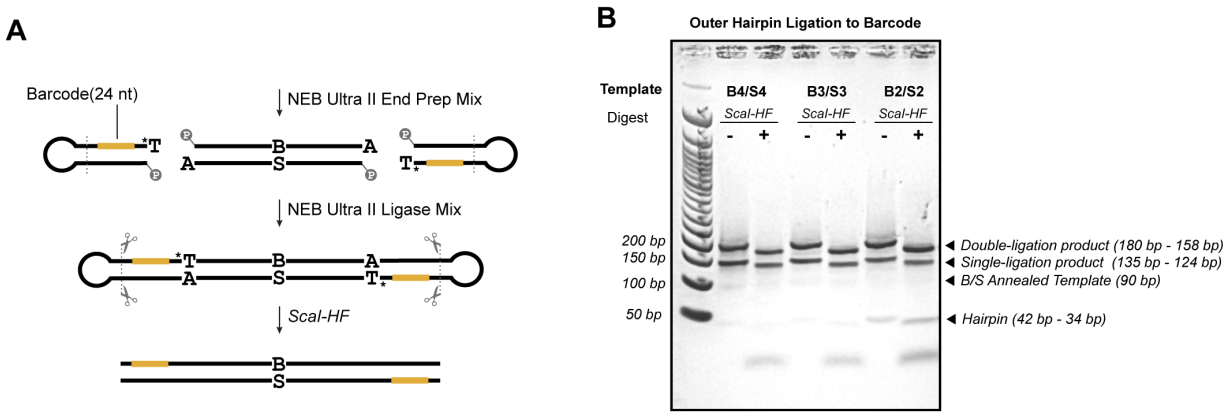
Supplementary Figure S5. Effects of chunk context on model performance for P vs G (PG1) model training. **(A)** Model performance, displayed as the false negative rate (FNR) and false positive rate (FPR) on the testing dataset for varying symmetric chunk contexts used in model training and testing. **(B)** Median number of bases upstream (US) and downstream (DS) of the XNA focus position (Either P for template **P1** or G for template **PG1**), for which the specified chunk context includes the current signal. Error bars represent the interquartile range. **(C)** Confusion matrix showing model performance for PG1 models using the chunk context '50 50' **(D)** Confusion matrix showing model performance for PG1 models using the chunk context '125 125'.



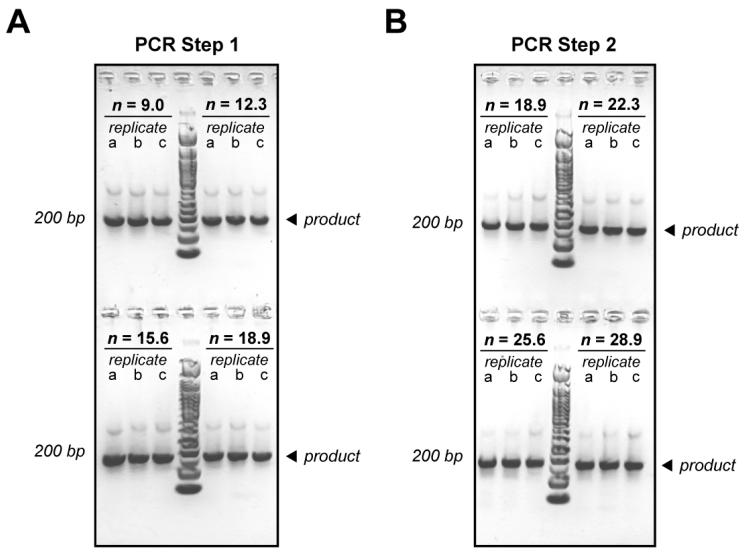
Supplementary Figure S6. Validation of XNA fraction quantification *in silico* using mixed synthetic datasets. **(A)** Schematic overview of the *in silico* testing quantifiable range. Raw nanopore reads (POD5) were generated from sequencing two synthetic constructs (annealed templates **B1**:**S1** and annealed templates **BA1**:**ST2**) using two separate flow cells. Reads from each POD5 file were randomly sampled and merged at predefined ratios to generate datasets with known ground truth XNA fractions. **(B)** Estimated XNA fraction as a function of ground truth XNA fraction for **B** and **S**, obtained by applying XNA classification models to each mixed dataset. Each point represents a single mixed dataset, with ground truth fractions spanning 0 to 1 in increments of 0.1. **(C)** XNA fractions normalized to the XNA fraction estimates for the ground truth 0% XNA and 100% XNA datasets.



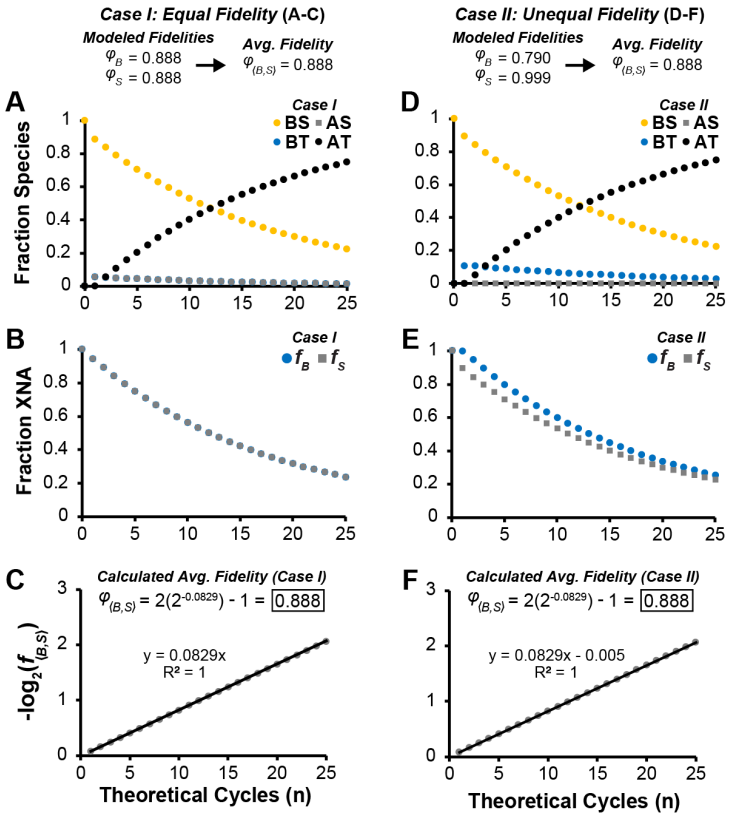
Supplementary Figure S7. Sample Sankey diagram for BS PCR basecalling. Sankey diagram constructed for the sequencing run used to generate XNA fraction data presented in **Fig. 3B**. Number of reads (n) is listed per step of the data processing pipeline including sequencing (Oxford Nanopore Technologies, Flongle flow cell), basecalling (Dorado), alignment (Minimap2), chunk generation (Remora), and demultiplexing (CutAdapt).



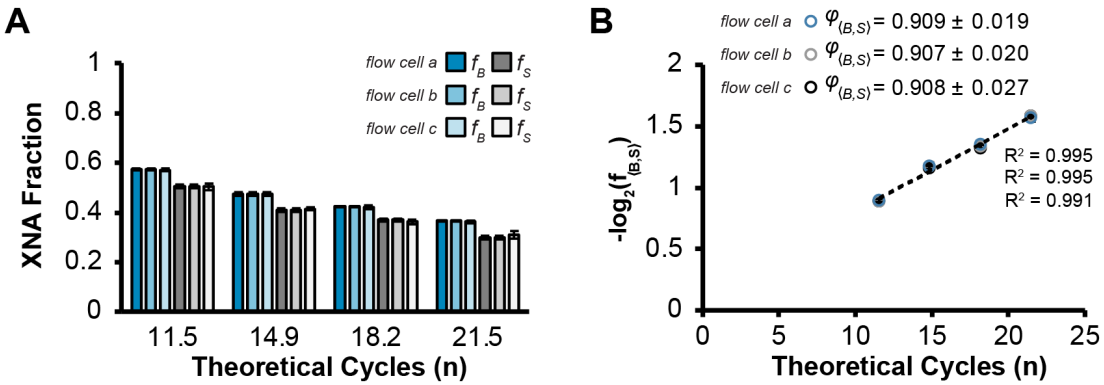
Supplementary Figure S8. Barcoding XNA-containing DNA using TA ligation of synthetic constructs with outer hairpins. (A) Schematic overview for the generation of barcoded, lengthened synthetic sequences for model training and testing without PCR. 5'-phosphorylated hairpin-forming DNA oligos were ordered from IDT with a single 3'-**T** overhang protected from cleavage by a phosphorothioate bond. Synthetic DNA oligos containing ubp XNAs were annealed, end-prepped, and ligated to annealed outer hairpins. Purified ligation products were then digested using *ScaI*-HF to generate blunt-end DNA suitable for nanopore sequencing preparation. **(B)** Agarose gel showing ligation products for three dsDNA **BS** templates with '+' and without '-' *ScaI*-HF digestion. Expected lengths of various ligation/digestion products are labeled to the right of the gel. Ladder shown is the 50 bp ladder (NEB).



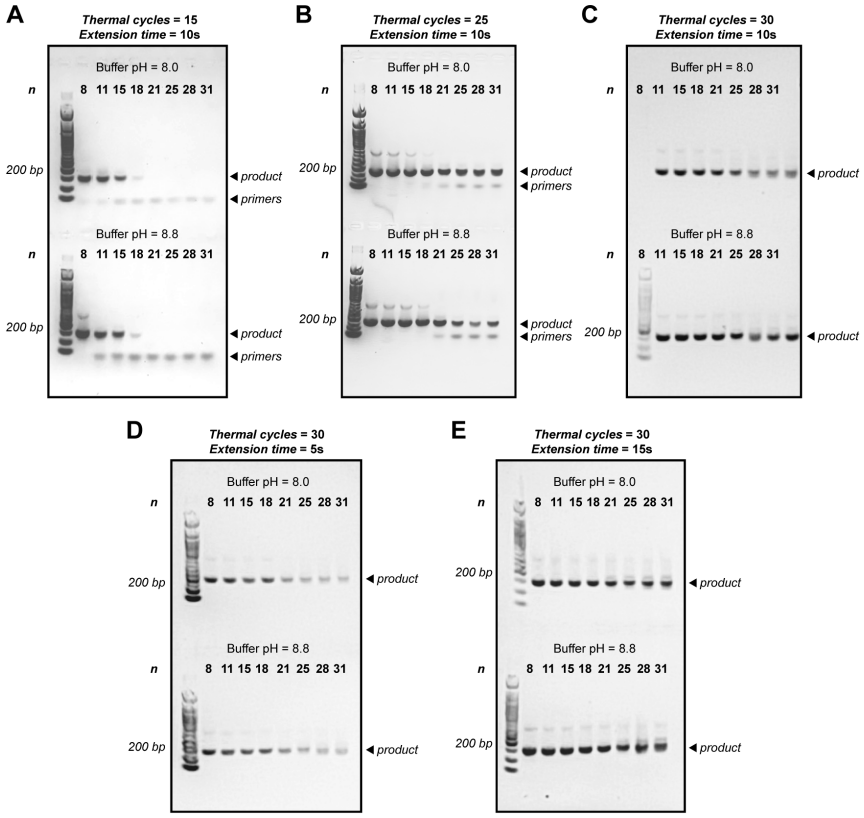
Supplementary Figure S9. Two-step PCR amplification of annealed DNA template B1 and S1. Agarose gel corresponding to the two-step PCR data presented in **Fig. 3B**. The number of theoretical cycles (*n*) corresponding to each replicate group is labeled on the gel. Each reaction was performed in triplicate, as labeled by replicates a, b, and c. 50 bp DNA Ladder (NEB) is included for reference. The intended product is labeled to the right of each gel. **(A)** PCR step 1, using diluted DNA template at known concentrations. **(B)** PCR step 2, using a 1000 X diluted products of PCR step 1 as template.



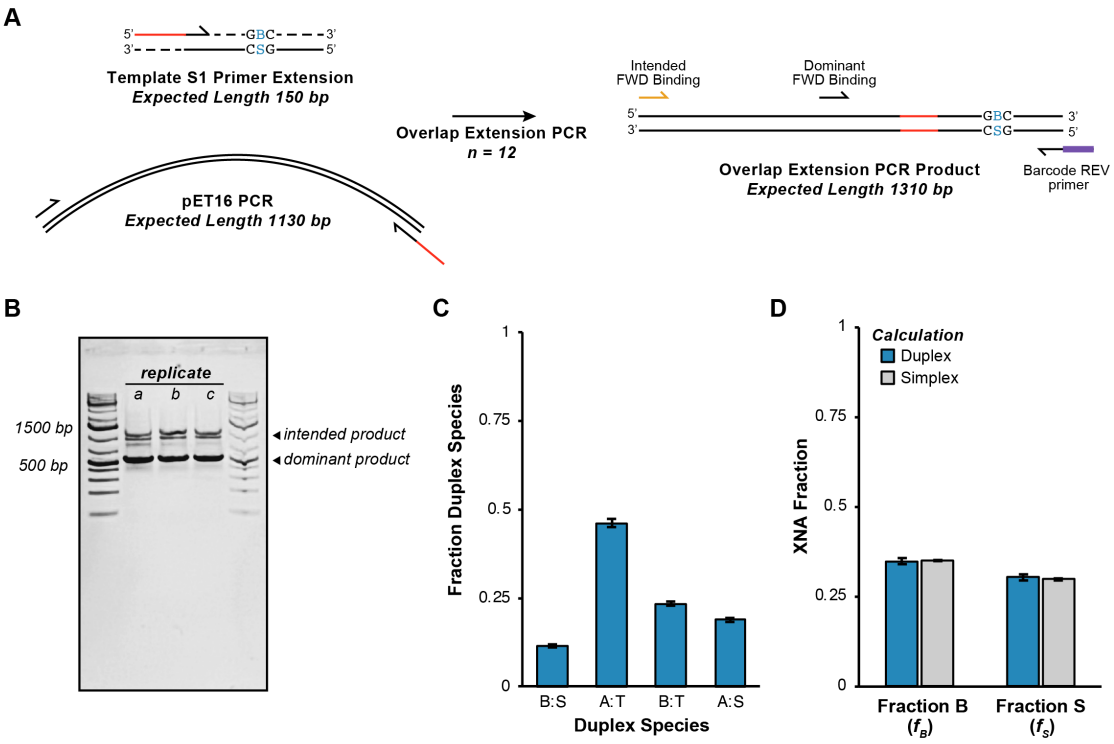
Supplementary Figure S10. Modeling PCR of the BS pair using decoupled replication fidelities φ_B and φ_S. The number of dsDNA species versus the number of theoretical cycles was modeled using the ODE system described by **Eqn. 15** for two distinct cases. For Case I (**A-C**), individual replication fidelities were set to be equal ($\varphi_B = \varphi_S = 0.888$), resulting in an average fidelity ($\varphi_{\langle B,S\rangle}$) of $\varphi_{\langle B,S\rangle} = 0.888$. For Case II (D-F), φ_B and φ_S were set to be assymetric ($\varphi_B = 0.790,\ \varphi_S = 0.999$), resulting in an identical observed value of $\varphi_{\langle B,S\rangle}$ to Case I ($\varphi_{\langle B,S\rangle} = \sqrt{\varphi_B \cdot \varphi_S} = 0.888$). **(A)** and **(D)** show the modeled fraction of dsDNA species as a function of theoretical cycle (n) for Case I and Case II, respectively. **(B)** and **(E)** show the modeled XNA fraction for B (f_B) and S (f_S) as a function of theoretical cycle (n) for Case I and Case II, respectively. **(C)** and **(F)** show log-transformed, average modeled XNA fraction ($-\log_2 f_{\langle B,S\rangle}$) as a function of theoretical cycle (n) for Case I and Case II, respectively. Both cases have the same slope, corresponding to the expected average fidelity $\varphi_{\langle B,S\rangle} = 0.888$.



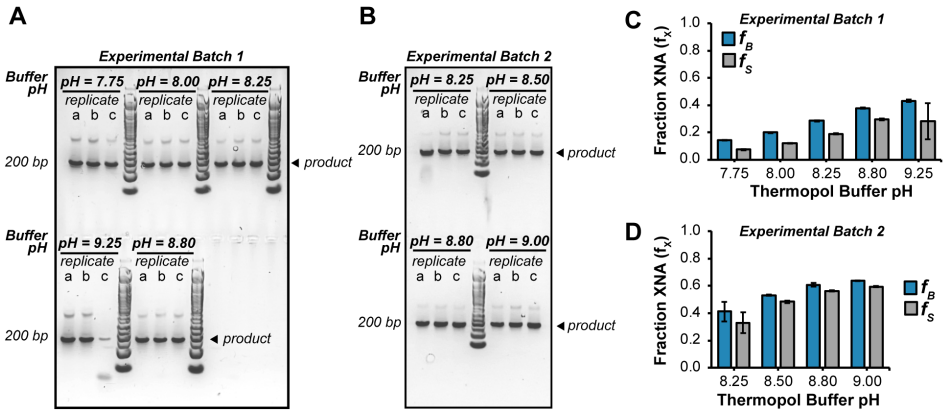
Supplementary Figure S11. Reproducibility of fidelity measurements across different ONT Flongle flow cells. (A) The average fraction **B** (f_B) and fraction **S** (f_S) for each value of theoretical cycles (n) as collected from sequencing identical, purified PCR samples on three separate Flongle flow cells. Error bars represent the standard deviation of triplicate reactions (n = 3). **(B)** Log-plots constructed for the data in panel A, with calculated fidelity ($\varphi_{\langle B,S\rangle}$) $\pm$ 95% CI displayed on the plot for each of the three flow cells. Error bars represent the standard deviation of triplicate reactions (n =3).



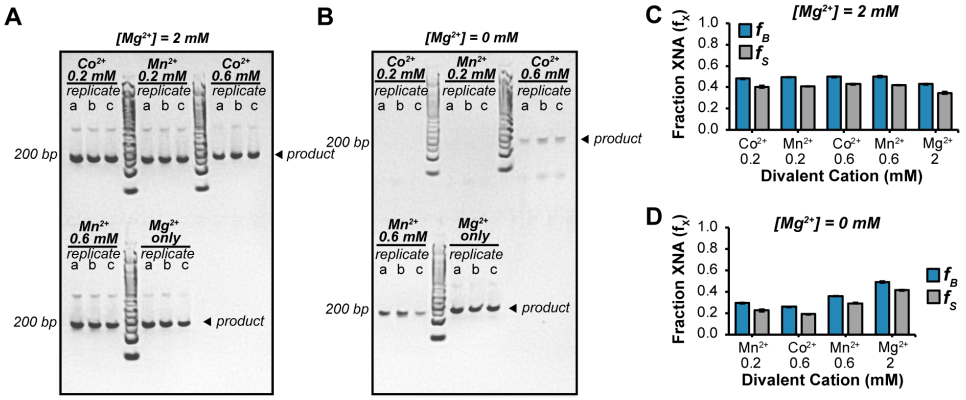
Supplementary Figure S12. Agarose gels showing PCR reaction progress for various cycle numbers and extension times. Agarose gel electrophoresis of PCR reaction completion across varying primer-to-template ratios in reactions subject to 15, 25, or 30 programmed thermal cycles. Sets of eight PCR reactions were prepared using 10-fold serial dilutions of dsDNA template (**BA1**,**ST1**; **Supplementary Table 2**), resulting in DNA template concentrations from 40 pg/µL (n = 8) to 4 ag/µL (n = 31) using Taq polymerase in Thermopol buffer at pH 8.0 (top) or 8.8 (bottom). Theoretical cycle numbers (n) corresponding to each primer/template ratio are labeled above each lane. The 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled. Product bands and visible primers are labeled to the right of the gels. **(A)** 15 thermal (real) cycles with a 10s extension time (72 °C). **(B)** 25 thermal cycles with a 10s extension time. **(C)** 30 thermal cycles with a 10s extension time. **(D)** 30 thermal cycles with a 5s extension time. **(E)** 30 thermal cycles with a 15s extension time.



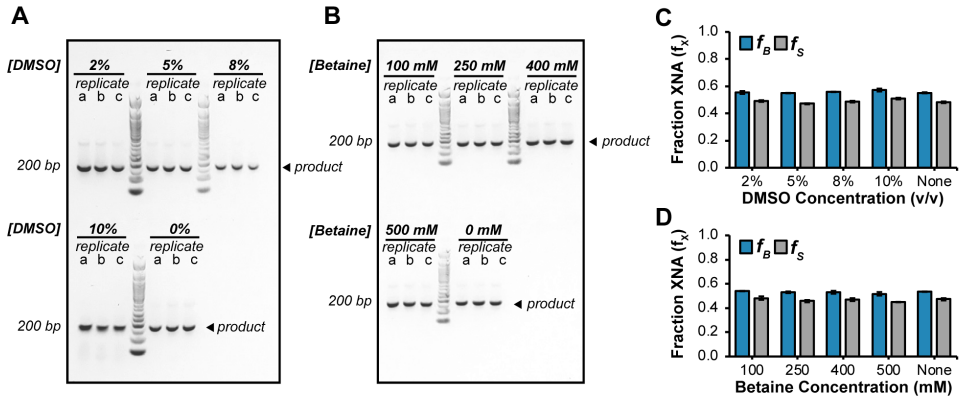
Supplementary Figure S13. Overlap extension PCR of template S1 and duplex sequencing analysis. **(A)** Schematic overview for the generation of barcoded, lengthened (>500 bp) PCR products using overlap extension PCR (OEP). **(B)** Agarose gel showing OEP products for triplicate reactions, labeled a-c. GeneRuler 1kb plus DNA ladder is included for reference, with 1500 bp and 500 bp bands labeled. Intended product and dominant product lengths are labeled on the gel. **(C)** The fraction of total duplex species is shown for the four expected species. **(D)** Comparison of XNA fraction between duplex (blue) and simplex (gray) calculation methods. Using the duplex calculation method, Fraction **B** was calculated as the sum of fraction B:S species and B:T species divided by the total number of duplex species. Similarly, Fraction **S** was calculated as the sum of fraction **B**:**S** species and **A**:**S** species divided by the total number of duplex species.



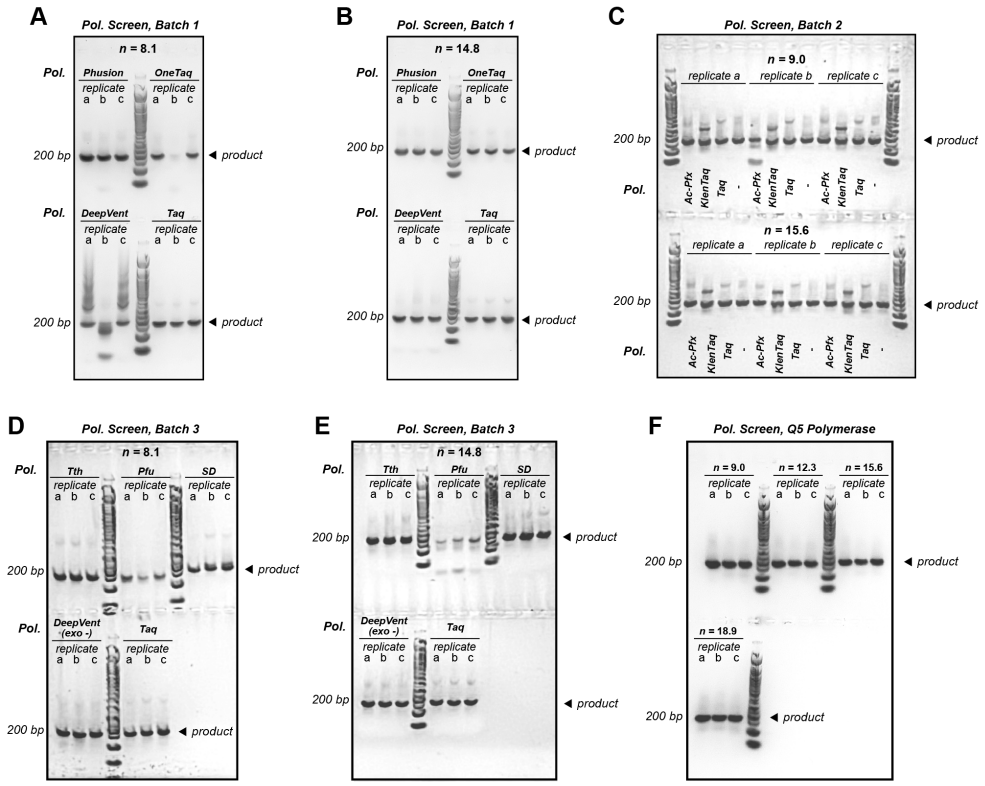
Supplementary Figure S14. Source agarose gels and raw XNA fraction data for Thermopol buffer pH screening. Thermopol Buffer pH screening with Taq Polymerase and template **B1** & **S1** was completed in two experimental batches. Agarose source gels, including all experimental replicates (labeled a-c; n=3), are shown for **(A)** experimental batch 1 and **(B)** experimental batch 2. The 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled. The average fraction **B** (f_B) and fraction **S** (f_S) remaining after 11.4 theoretical cycles for various buffer pH values are shown for **(C)** experimental batch 1 and **(D)** experimental batch 2. Each batch includes an internal reference condition (pH 8.80) used for comparison in **Fig. 4A**. Error bars represent the standard deviation of experimental replicates (n =3).



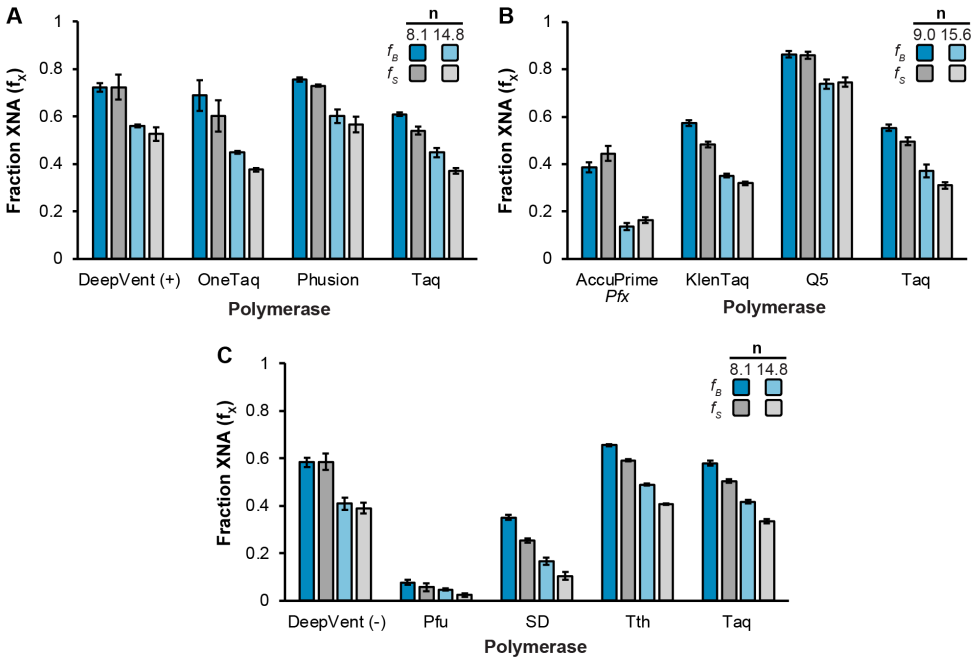
Supplementary Figure S15. Source agarose gels and raw XNA fraction data for divalent cation supplementation screening. Divalent cation (M^{2+}) screening with Taq Polymerase and template B1/S1 was completed with (2 mM) and without (0 mM) magnesium (Mg^{2+}) added to the reaction. M^{2+} conditions tested include Co^{2+} (0.2, 0.6 mM) and Mn^{2+} (0.2, 0.6 mM) supplementation. Agarose source gels, including all experimental replicates (n=3), are shown for **(A)** samples with 2 mM Mg^{2+} supplementation and **(B)** samples without Mg^{2+} supplementation. It was observed that PCR reactions in panel **B** were incomplete at 0.2 mM Co^{2+}, 0.2 mM Mn^{2+}, and 0.6 mM Co^{2+}. A 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled. The average fraction B (f_B) and fraction S (f_S) remaining after 11.4 theoretical cycles for various M^{2+} conditions are shown for **(C)** samples with 2 mM Mg^{2+} supplementation and **(D)** samples without Mg^{2+} supplementation. Each batch includes an internal reference condition (2 mM Mg^{2+}; no other M^{2+} supplement) used for comparison in **Fig. 4A**. Error bars represent the standard deviation of experimental replicates (n =3).



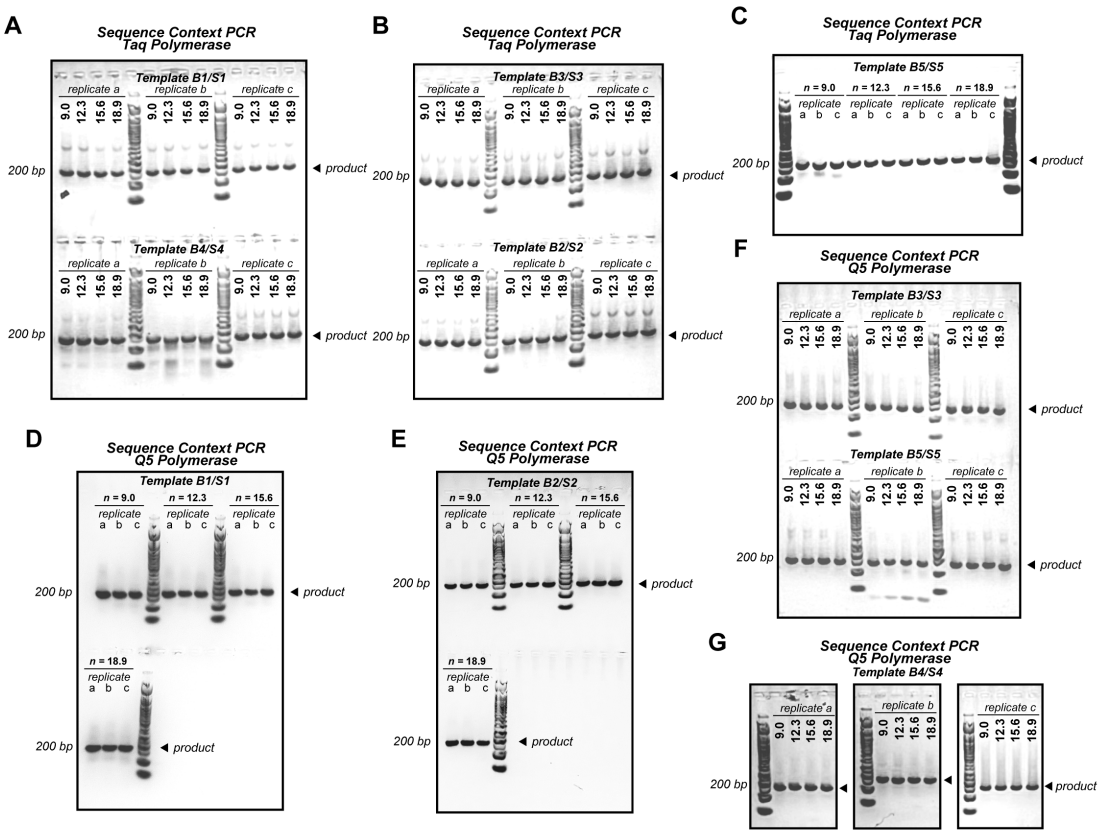
Supplementary Figure S16. Source agarose gels and raw XNA fraction data for DMSO and betaine supplementation screening. PCR supplement screening with Taq Polymerase and template **B1/S1** was completed by varying concentrations of betaine (100-500 mM) or DMSO (2-10%) in the PCR reaction. Agarose source gels, including all experimental replicates (n=3), are shown for **(A)** DMSO supplementation and **(B)** betaine supplementation. The 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled. The average fraction **B** (f_B) and fraction **S** (f_S) remaining after 11.4 theoretical cycles for various supplementation conditions are shown for **(C)** DMSO and **(D)** betaine. Each batch includes an internal reference condition (0% DMSO or 0 mM betaine) used for comparison in **Fig. 4A**. Error bars represent the standard deviation of experimental replicates (n =3).



Supplementary Figure S17. Source agarose gels for BS polymerase screening. PCR polymerase (pol.) screening using template **B1, S1** was completed in several batches, with each batch run at multiple theoretical cycle (n) values. Agarose source gels, including all experimental replicates (n=3), are shown. The 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled. The batch 1 polymerase screen, including Phusion HF, OneTaq, DeepVent, and Taq (reference condition) polymerases, was completed at **(A)** $n = 8.1$ and **(B)** $n = 14.8$ theoretical cycles. The batch 2 polymerase screen, including AccuPrime *Pfx* (Ac-*Pfx*), KlenTaq1 (KlenTaq), and Taq (reference condition) polymerases, was completed at **(C)** $n = 9.0$ and $n = 15.6$ theoretical cycles. The batch 3 polymerase screen, including *Tth*, *Pfu*, SD, DeepVent (exo-) and Taq (reference condition) polymerases, was completed at **(D)** $n = 8.1$ and **(E)** $n = 14.8$ theoretical cycles. **(F)** Q5 polymerase was screened separately, at four theoretical cycle values ($n = 9.0$-18.9).

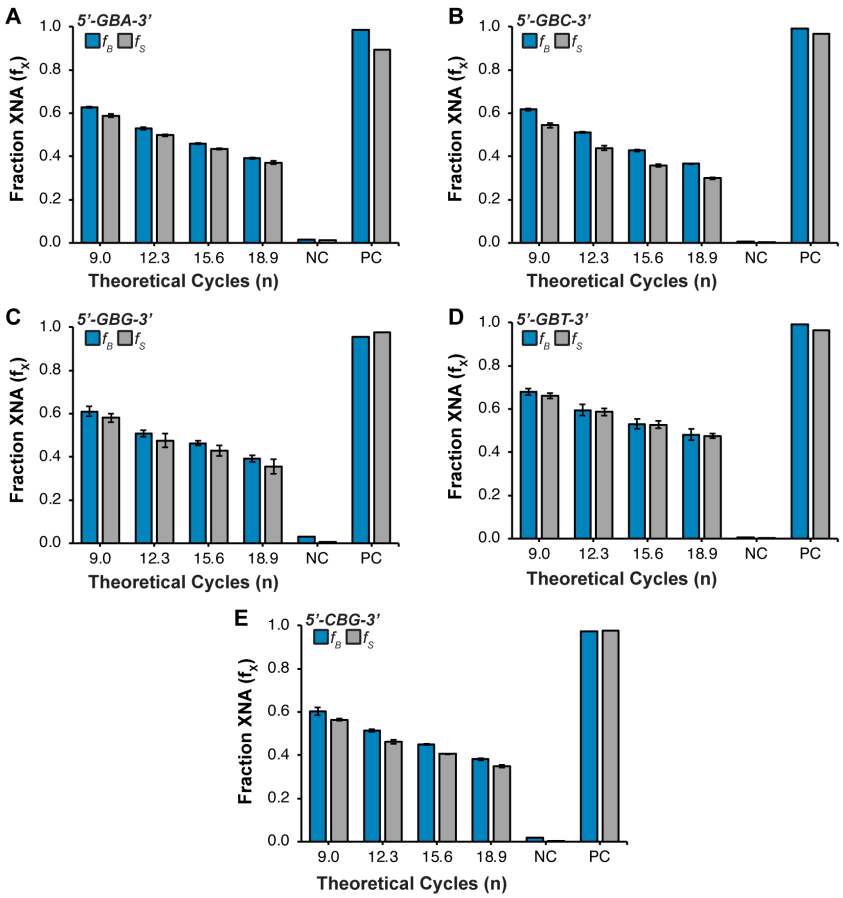


Supplementary Figure S18. Raw XNA fraction data for PCR polymerase screening experiments. The average fraction **B** (f_B) and fraction **S** (f_S) remaining after n theoretical cycles (shown in legend) for various polymerases alongside the Taq polymerase reference used for comparison in **Fig. 4A**. Raw XNA fractions shown here were used to calculate fidelity ($\varphi_{\langle B,S\rangle}$) values presented in Fig. 4B. **(A)** Polymerase screen XNA fractions for DeepVent (exo+), OneTaq, Phusion, and Taq polymerases. **(B)** Polymerase screen XNA fractions for AccuPrime *Pfx*, KlenTaq1 (KlenTaq), Q5, and Taq polymerases. **(C)** Polymerase screen XNA fractions DeepVent (exo-), *Pfu*, SD, *Tth*, and Taq polymerases. Error bars represent the standard deviation of experimental replicates (n = 3).

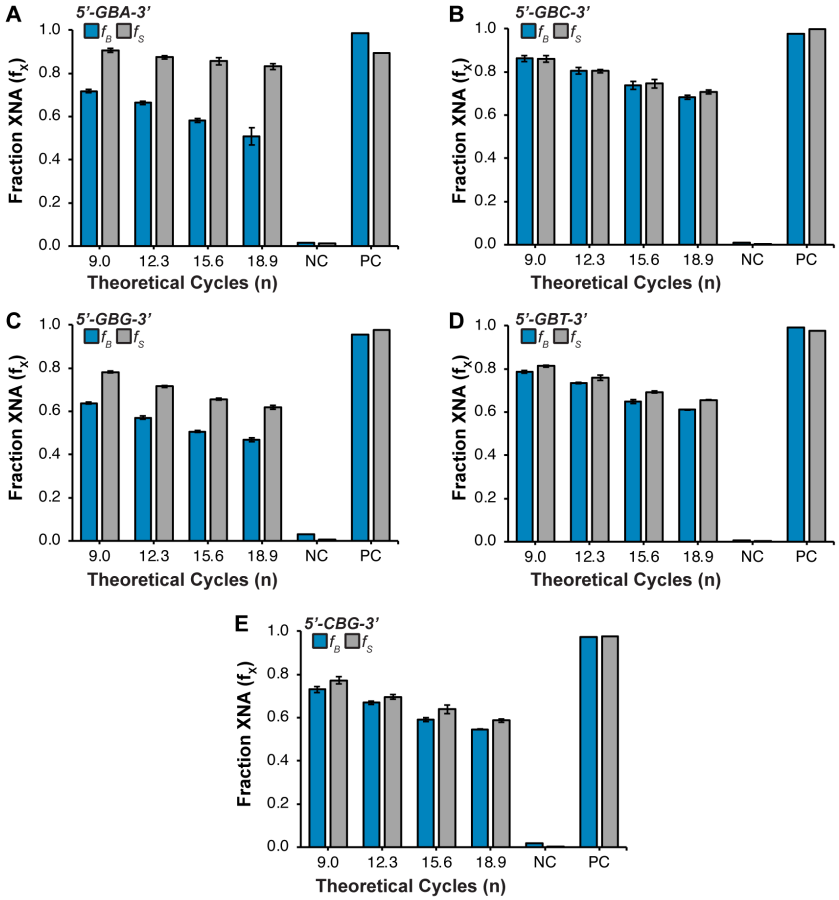


Supplementary Figure S19. Source agarose gels for PCR fidelity quantification of BS sequence contexts using Taq polymerase and Q5 polymerase. Source agarose gels for sequence context screening at four theoretical cycle values (n = 9.0, 12.3, 15.6, and 18.9), including experimental triplicates (a-c) are shown. The 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled. PCR amplification of **(A)** template **B1,S1** (top) and **B4,S4** (bottom) using Taq polymerase, **(B)** template **B3,S3** (top) and **B2,S2** (bottom) using Taq polymerase. **(C)** template **B5,S5** using Taq polymerase, **(D)** template **B1,S1** using Q5 polymerase, **(E)** template **B2,S2** using Q5 polymerase, **(F)** template **B3,S3** (top) and **B5,S5** (bottom) using Q5 polymerase, and **(G)** template **B4,S4** using Q5 polymerase.

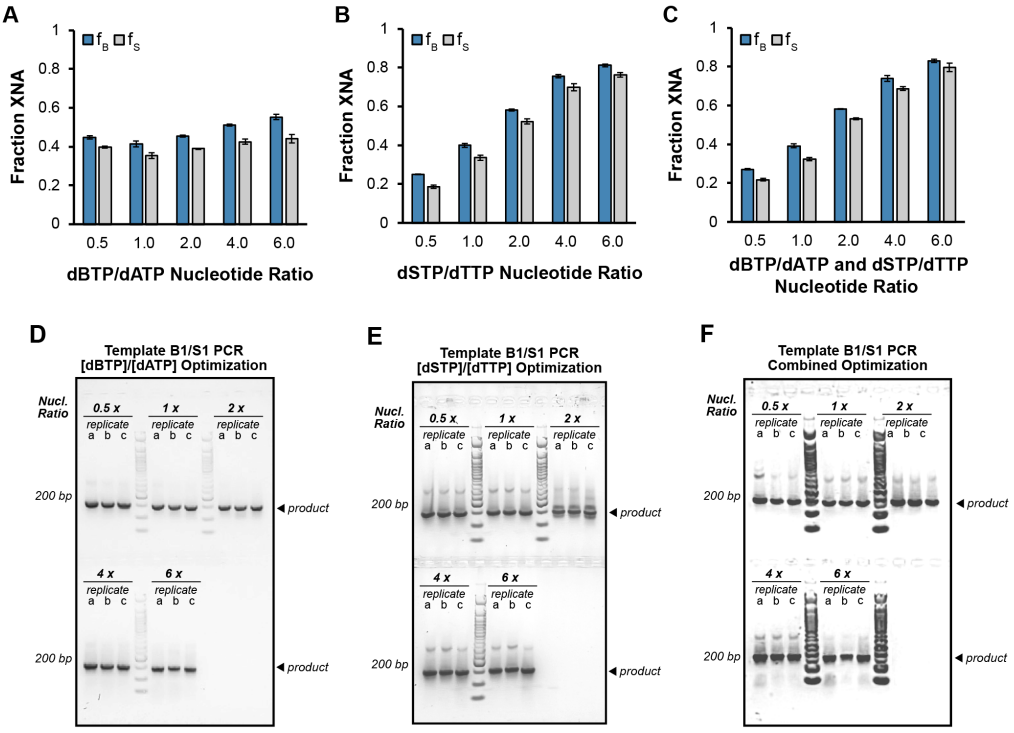
.



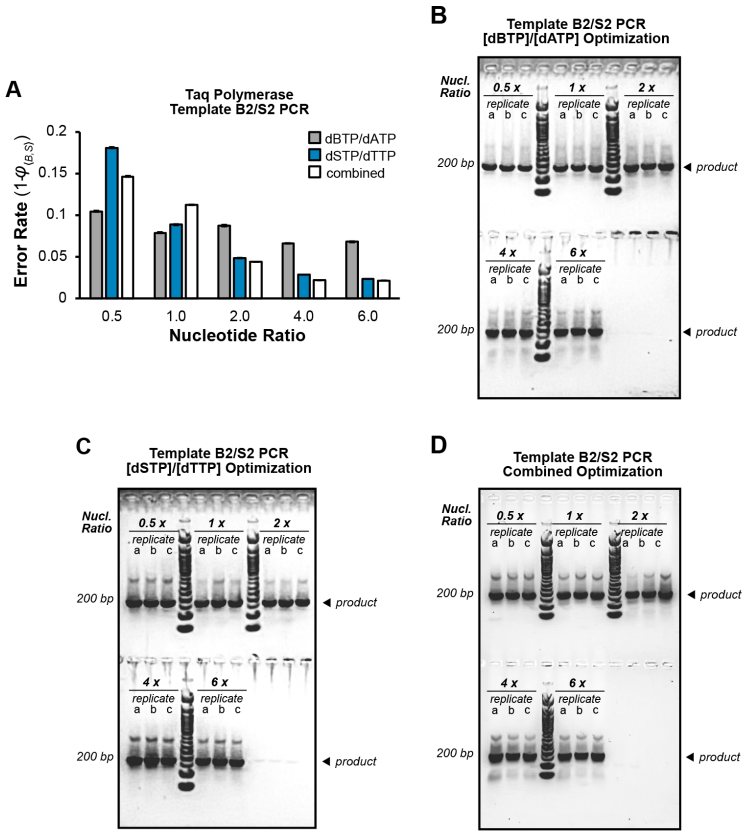
Supplementary Figure S20. Raw XNA fraction data for sequence context PCR experiments with Taq polymerase. The average fraction **B** (f_B) and fraction **S** (f_S) remaining after *n* theoretical cycles for PCR amplification of various BS templates by Taq polymerase. The XNA fractions for barcoded DNA controls are included for 0% XNA (NC) and 100% XNA (PC). Raw XNA fractions shown here were transformed onto a log scale, and linear regression was used to calculate fidelity ($\varphi_{\langle B,S\rangle}$) values presented in **Fig. 4C**. Linear regression statistics are shown in **Supplementary Table 8**. Each panel corresponds to a template sequence context (**Supplementary Table 2**), including **(A)** 5`-GBA-3`; B3,S3**(B)** 5`-G**B**C-3`; **B1,S1**, **(C)** 5`-G**B**G-3`; **B3,S3**, **(D)** 5`-G**B**T-3`; **B2,S2**, and **(E)** 5`-C**B**G-3`; **B5,S5**. Error bars represent the standard deviation of experimental replicates (n = 3).



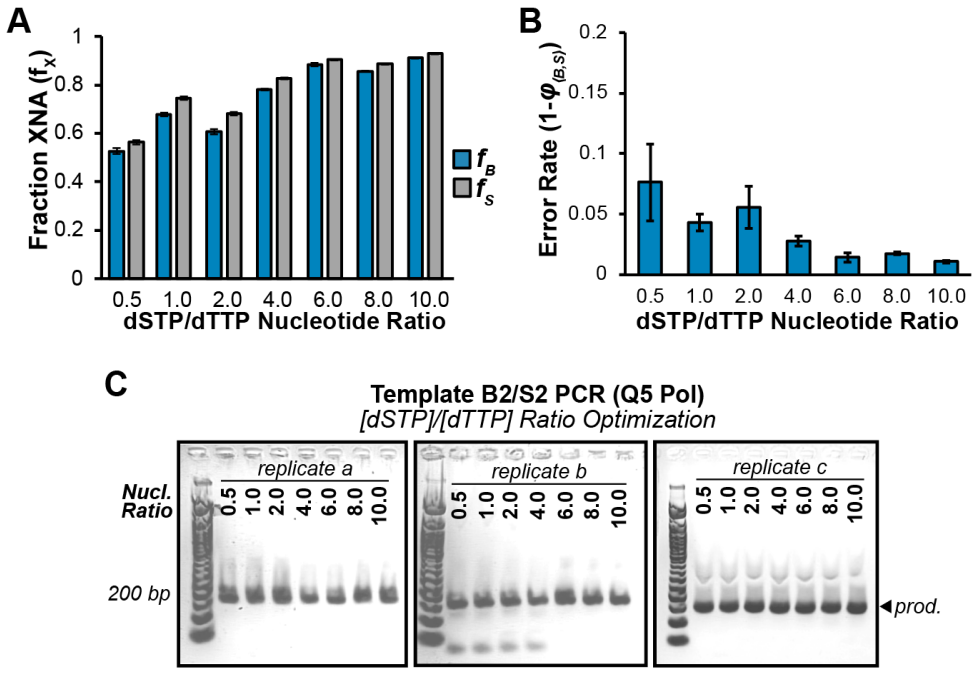
Supplementary Figure S21. Raw XNA fraction data for sequence context PCR experiments with Q5 polymerase. The average fraction **B** (f_B) and fraction **S** (f_S) remaining after n theoretical cycles for PCR amplification of various BS templates by Q5 polymerase. Raw XNA fractions shown here were transformed onto a log scale, and linear regression was used to calculate fidelity ($\varphi_{\langle B,S\rangle}$) values presented in **Fig. 4C**. Linear regression statistics are shown in **Supplementary Table 8**. Each panel corresponds to a template sequence context, including **(A)** 5`-G**B**A-3`; **B3,S3**, **(B)** 5`-G**B**C-3`; **B1,S1**, **(C)** 5`-G**B**G-3`; **B3,S3**, **(D)** 5`-G**B**T-3`; **B2,S2**, and **(E)** 5`-C**B**G-3`; **B5,S5**. Error bars represent the standard deviation of experimental replicates (n = 3).



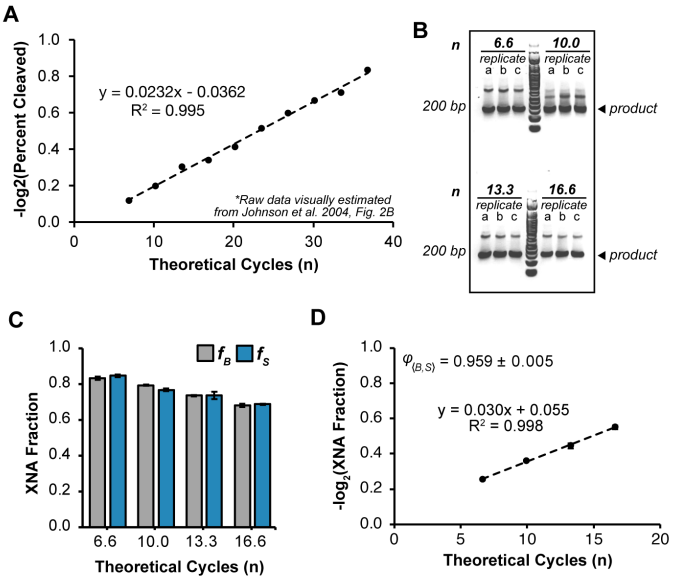
Supplementary Figure S22. Raw XNA fraction data and agarose source gels for nucleotide optimization screening with Taq polymerase. Template **B1,S1** was amplified by Taq polymerase using 15.61 theoretical cycles. Nucleotide concentrations used to specify ratios are shown in **Supplementary Table 9B**. **(A)** The average fraction **B** (f_B) and fraction **S** (f_S) for various dBTP/dATP nucleotide ratios. **(B)** f_B and f_S for various dSTP/dTTP nucleotide ratios. **(C)** f_B and f_S for various combined dBTP/dATP and dSTP/dTTP nucleotide ratios. Error bars represent standard deviation (n = 3). Source gels are shown for **(D)** dBTP/dATP, **(E)** dSTP/dTTP, and **(F)** combined dBTP/dATP and dSTP/dTTP nucleotide (nucl.) ratio screening. The 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled.



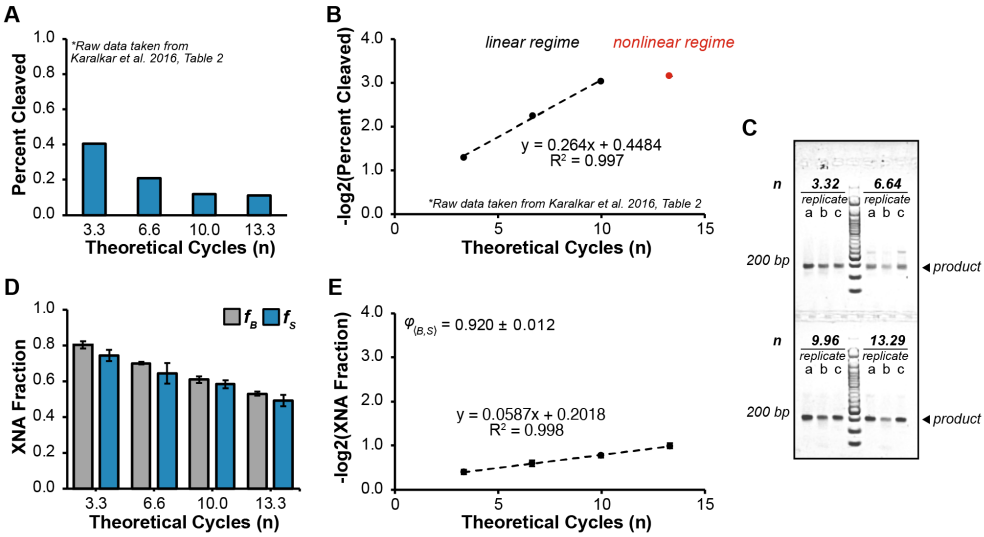
Supplementary Figure S23. Nucleotide optimization for template B2,S2 using Taq polymerase. **(A)** Polymerase error rate for varying nucleotide ratios after amplification of template **B2,S2** by Taq polymerase. Nucleotide ratios include dBTP/dATP, dSTP/dTTP, and both dBTP/dATP and dSTP/dTTP (combined). Nucleotide concentrations corresponding to each ratio are shown in **Supplementary Table 9B**. Error bars represent the propagated standard deviation of experimental replicates (n = 3). **(B)** Source agarose gel for dBTP/dATP ratio screening after 15.61 theoretical cycles. **(C)** Source agarose gel for dSTP/dTTP ratio screening after 15.61 theoretical cycles. **(D)** Source agarose gel for combined dBTP/dATP and dSTP/dTTP ratio screening after 15.61 theoretical cycles. For all gels shown the 50 bp DNA ladder (NEB) was included for reference.



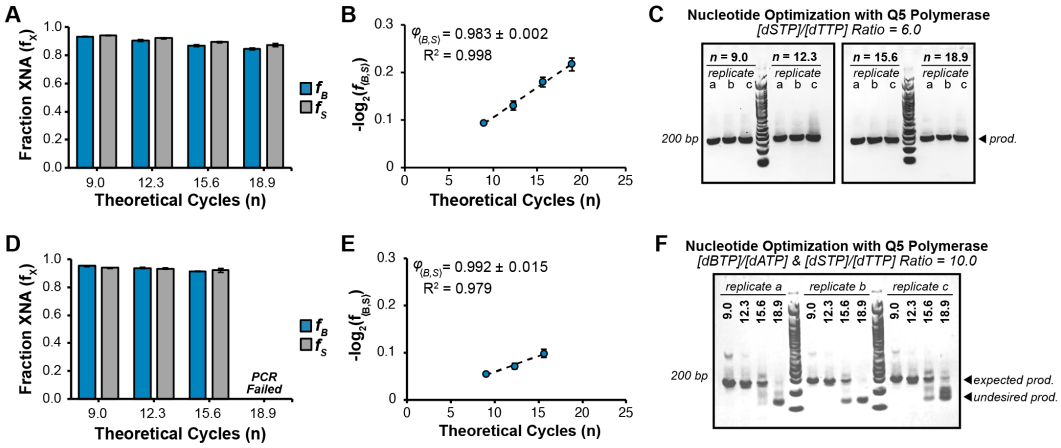
Supplementary Figure S24. Nucleotide optimization for template B2,S2 using Q5 polymerase. (A) The average fraction **B** (f_B) and fraction **S** (f_S) remaining after 15.61 theoretical cycles for various dSTP/dTTP nucleotide ratios. Error bars represent standard deviation (n=3). **(B)** Polymerase error rate for varying dSTP/dTTP nucleotide ratio. Nucleotide concentrations corresponding to each ratio are shown in **Supplementary Table 9B**. Error bars represent the propagated standard deviation of experimental replicates (n = 3). **(C)** Source agarose gel for dSTP/dTTP nucleotide (Nucl.) ratio screening. The 50 bp DNA ladder (NEB) was included for reference, with the 200 bp band labeled.



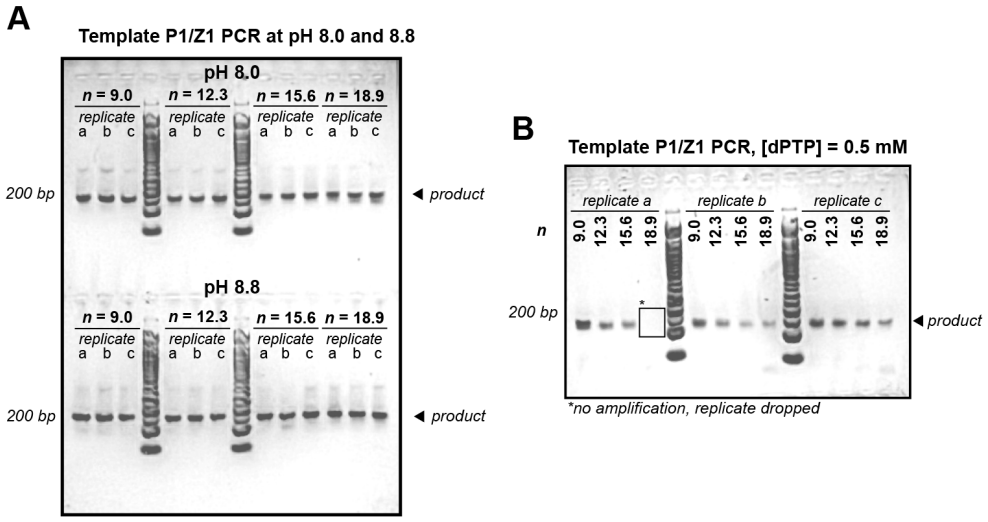
Supplementary Figure S25. Analysis and reproduction of experiments reported in Johnson et al. 2004. **(A)** Percent cleaved values were visually estimated from Johnson et al. 2004 (reference figure from publication: Figure 2B) and re-plotted on a $\log_2$ scale versus theoretical cycles for linear regression. Using the slope of the log-plot, the average fidelity was estimated $\varphi_{\langle B,S\rangle}$ = 96.8%, which is slightly higher than the average fidelity metric originally reported by the authors ($\varphi_{\langle B,S\rangle}$ = 96%). We reproduced the PCR experiment described by Johnson et al. 2004, (using template **B5,S5**), with four different theoretical cycle values ranging from 3.32 to 13.29. **(B)** An agarose source gel showing the triplicate amplicons (a-c) for each value of n. **(C)** Average fraction B (f_B) and fraction S (f_S) for each value of n. Error bars represent standard deviation of triplicate reactions (n = 3). **(D)** Log-plot constructed from XNA fraction data in panel C. $\varphi_{\langle B,S\rangle}$ value reported on the plot $\pm$ 95% CI.



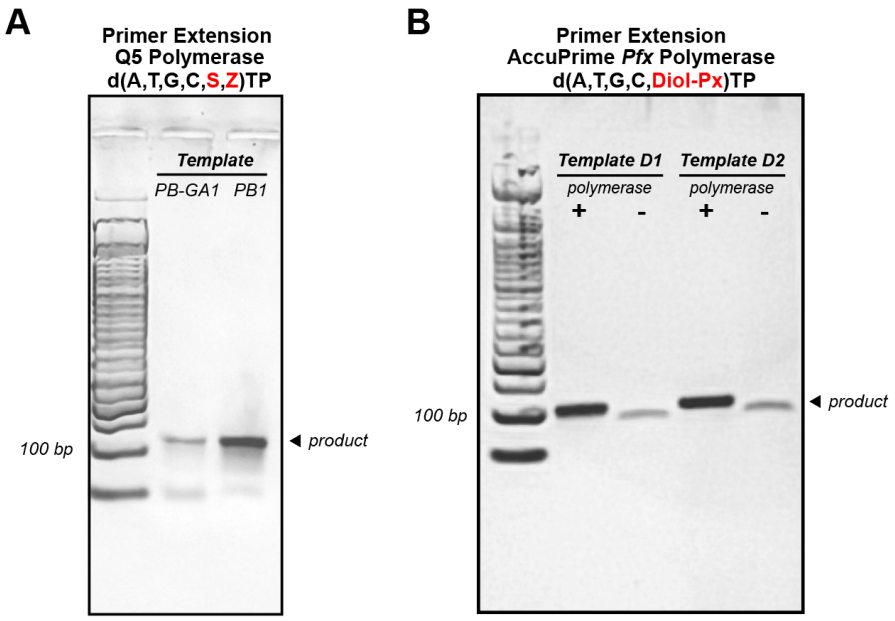
Supplementary Figure S26. Analysis and reproduction of experiments reported in Karalkar et al. 2016. **(A)** Percent cleaved data was taken from Karalkar et al. 2016 (reference table from publication: Table 2) and plotted as a function of theoretical cycle. **(B)** A log-plot was constructed using the data in panel A for linear regression. We observed that while the first three data points fall within the linear regime, the final point (n = 13.3) was in the nonlinear regime, characteristic of an incomplete PCR reaction. Using the slope of the log-plot, the fidelity was estimated to be 66.5%, compared to the 86% metric reported by the authors. We reproduced the PCR experiment described by Karalkar et al. 2016 as described (template **B2,S2**) using four theoretical cycle values that ranged from 3.32 to 13.29. **(C)** Agarose source gel showing the triplicate amplicons (a-c) for each theoretical cycle value. **(D)** Average fraction B (f_B) and fraction S (f_S) for each value of n. Error bars represent standard deviation of triplicate reactions (n = 3). **(E)** log-plot constructed from XNA fraction data in panel C. $\varphi_{\langle B,S \rangle}$ is reported on the plot $\pm$ 95% CI.



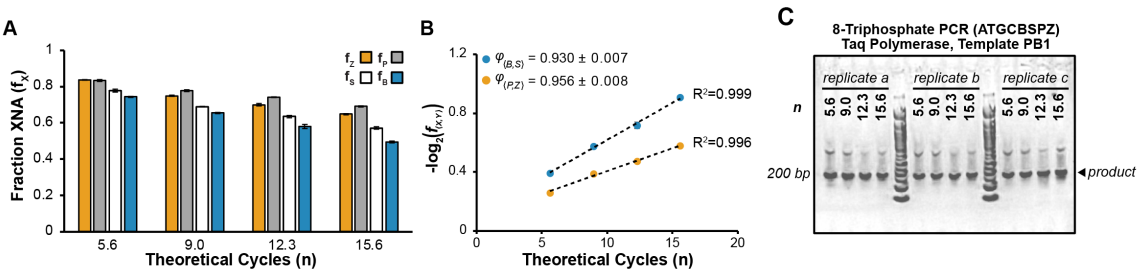
Supplementary Figure S27. Improved BS fidelity after conditions optimization. (A) The average fraction **B** (f_B) and fraction **S** (f_S) remaining after n theoretical cycles for PCR amplification of template **B2**,**S2** by Q5 polymerase using a 6X dSTP/dTTP nucleotide ratio. **(B)** Log-transformed XNA fraction data used to calculate fidelity ($\varphi_{\langle B,S\rangle}$) using WLS regression. $\varphi_{\langle B,S\rangle}$ is reported on the plot $\pm$ 95% CI. Linear regression statistics are shown in **Supplementary Table 8**. The regression fit line is displayed as a dashed line. Error bars represent standard deviation of experimental replicates (n = 3). **(C)** Agarose gel for amplification of varying template concentrations by Q5 polymerase in PCR reactions containing 6X dSTP/dTTP nucleotide ratio. **(D)** The average fraction **B** (f_B) and fraction **S** (f_S) after n theoretical cycles of template **B2**,**S2** by Q5 polymerase using a 10X dSTP/dTTP nucleotide ratio. **(E)** log-plot constructed for amplification of template **B2**/**S2** by Q5 polymerase in PCR reactions containing 10X combined dBTP/dATP and dSTP/dTTP ratios. $\varphi_{\langle B,S\rangle}$ is reported on the plot $\pm$ 95% CI. **(F)** Agarose gel for amplification of varying template concentrations by Q5 polymerase in PCR reactions containing 10X combined dBTP/dATP and dSTP/dTTP ratios.



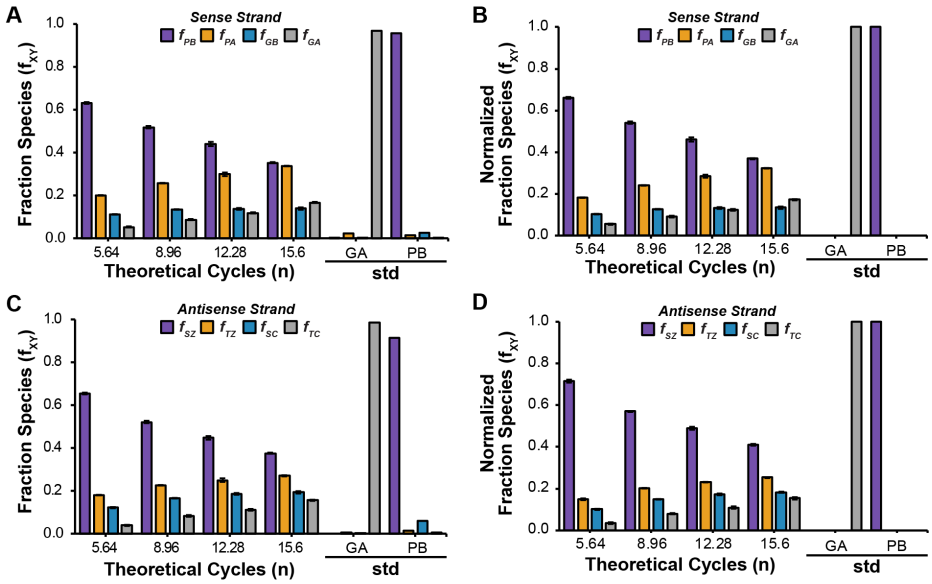
Supplementary Figure S28. Source agarose gels for PZ PCR experiments. Source agarose gels for PCR amplification of template **P1**,**Z2** at four theoretical doubling values (*n* = 9.0, 12.3, 15.6, and 18.9), performed in triplicate (labeled a-c). **(A)** Amplification using buffer pH 8.0 (top) and buffer pH 8.8 (bottom) using Takara Taq polymerase, at standard dPTP concentration of 200 µM. This gel corresponds to the data shown in **Fig. 5B**. **(B)** Amplification using buffer pH 8.0 using Takara Taq polymerase and increased dPTP concentration of 500 µM. This gel corresponds to the [dPTP] opt. series presented in **Fig. 5C**. One replicate (a) was dropped for the 18.9 theoretical cycle condition due to failed amplification. 50 bp DNA ladder (NEB) is included for reference.



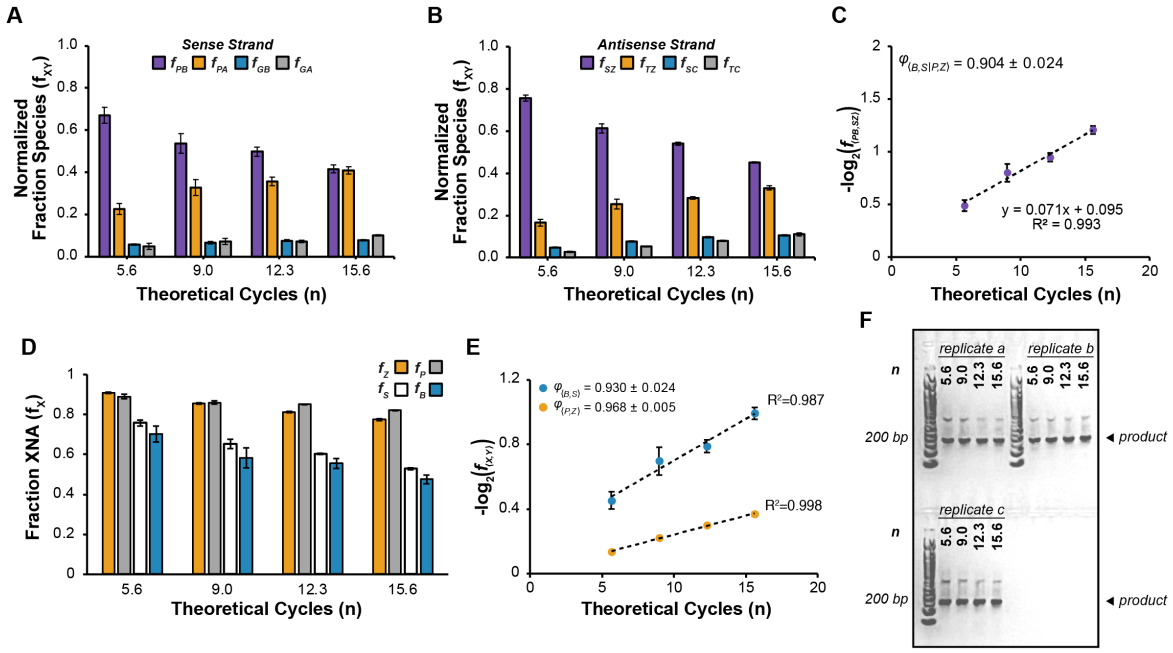
Supplementary Figure S29. Agarose gels showing primer extension products used for model training and testing. **(A)** Primer extensions of template **PB-GA1** and **PB1** (**Supplementary Table 2**) by Q5 polymerase. **(B)** Primer extensions of templates **D1** and **D2** by AccuPrime *Pfx* polymerase. Reactions are shown with '+' and without '-' polymerase added to the reaction. Ladders shown are 50 bp ladder (NEB).



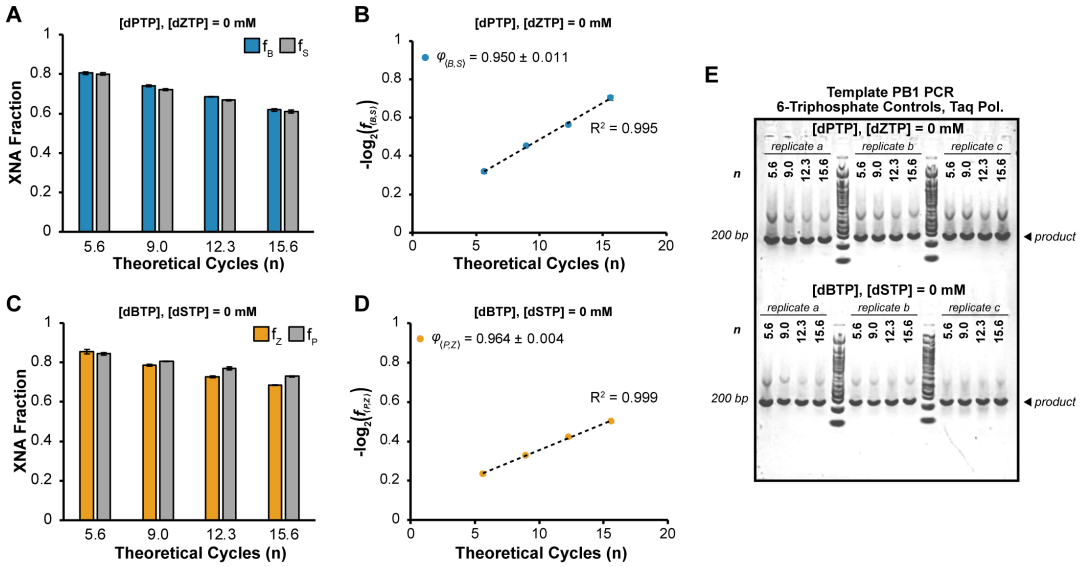
Supplementary Figure S30. Raw XNA fraction data, individual log plots, and source agarose gel for 8-letter PCR with Taq polymerase. (A) Average XNA fractions f_B, f_S, f_P, and f_Z remaining after n theoretical cycles for PCR amplification of template **PB1** by Taq polymerase with eight nucleotide triphosphates added (dATP, dTTP, dGTP, dCTP, dBTP, dSTP, dPTP, dZTP). Error bars represent standard deviation of experimental replicates (n =3). **(B)** Log-transformed XNA fraction data used to calculate fidelities $\varphi_{\langle B,S\rangle}$ and $\varphi_{\langle P,Z\rangle}$ using WLS regression. $\varphi_{\langle B,S\rangle}$ and $\varphi_{\langle P,Z\rangle}$ ± 95% CI are reported on the plot. Linear regression statistics are shown in **Supplementary Table 8**. The regression fit line is displayed as a dashed line. Error bars represent standard deviation of experimental replicates (n = 3). **(C)** Agarose gel for amplification of varying template concentrations of template **PB1** (**Supplementary Table 2**) by Taq polymerase with eight nucleotide triphosphates (dATP, dTTP, dGTP, dCTP, dBTP, dSTP, dPTP, dZTP) supplemented.



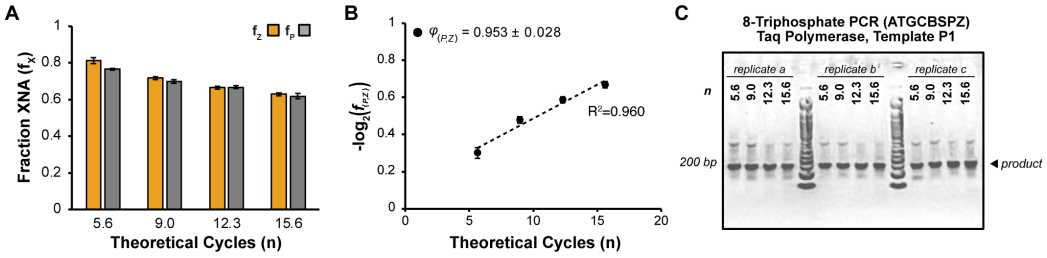
Supplementary Figure S31. Raw fraction species and normalized fraction species for 8-letter PCR with Taq polymerase. (A) Raw fraction species (f_{XY}) determined by combining per-read basecalls from models 8L-BN2 and 8L-PG1 applied to the original XNA positions of **B** and **P**, respectively, in reference sequence **PB1** (sense strand). **(B)** Raw fraction species (f_{XY}) determined by combining per-read basecalls from models 8L-SN2 and 8L-ZC1 applied to the original XNA positions of **S** and **Z**, respectively, in the reverse complement of reference **PB1** (antisense strand). **(C)** Normalized fraction species (f_{XY}) for the sense strand. (D) Normalized fraction species (f_{XY}) for the antisense strand.



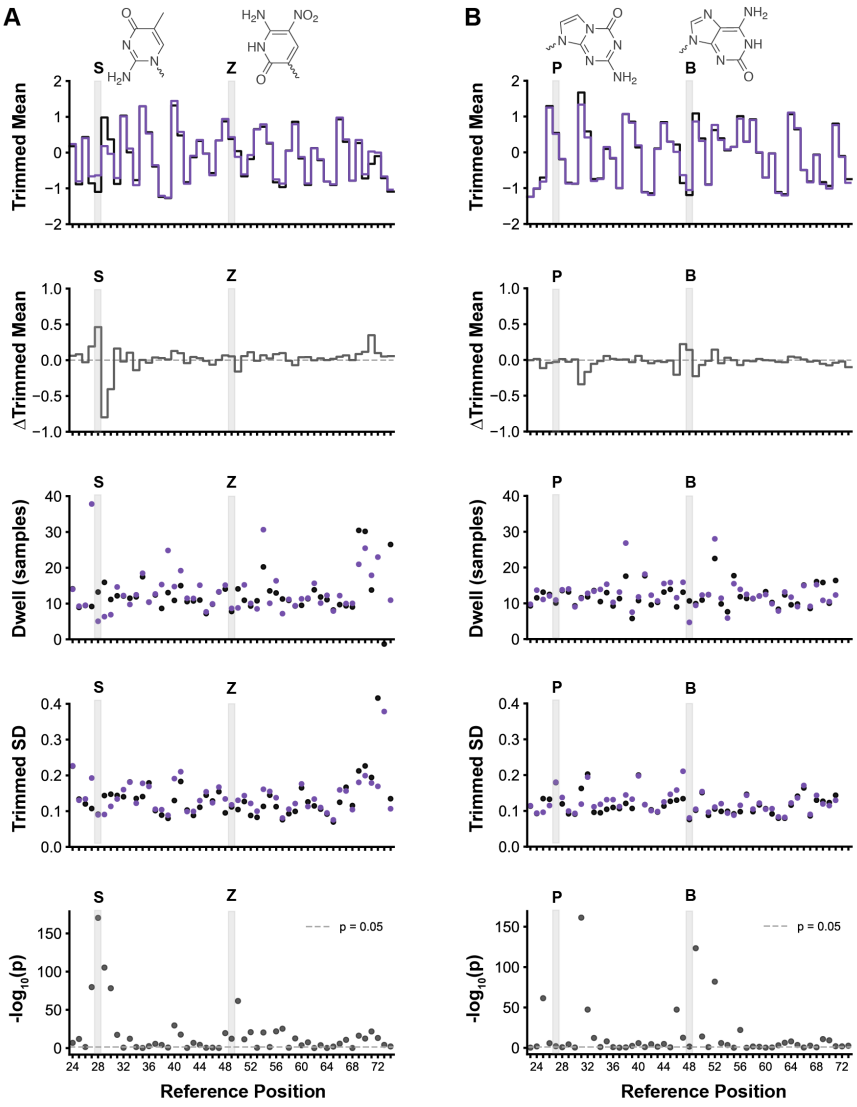
Supplementary Figure S32. Raw XNA fraction data, individual log plots, and source agarose gel for 8-letter PCR with Takara Taq polymerase. **(A)** Normalized fraction species (f_{XY}) determined by combining per-read basecalls from models 8L-BN2 and 8L-PG1 applied to the original XNA positions of **B** and **P**, respectively, in reference sequence **PB1** (sense strand). **(B)** Normalized fraction species (f_{XY}) of controls determined by combining per-read basecalls from models 8L-SN2 and 8L-ZC1 applied to the original XNA positions of **S** and **Z**, respectively, in the reverse complement of reference **PB1** (antisense strand). **(C)** Log_2-transformed average fraction of **PB** and **SZ** species ($f_{\langle PB\,|\,SZ\rangle}$). Regression fit shown as a dashed line. **(D)** Average XNA fractions f_B, f_S, f_P, and f_Z remaining after n theoretical cycles for PCR amplification of template **PB1** by Takara Taq polymerase with all eight nucleotide triphosphates added (dATP, dTTP, dGTP, dCTP, dBTP, dSTP, dPTP, dZTP). **(E)** Log_2-transformed XNA fraction data used to calculate fidelities $\varphi_{\langle B,S\rangle}$ and $\varphi_{\langle P,Z\rangle}$ using WLS regression. The regression fit lines are displayed as a dashed line. **(F)** Agarose gel for amplification of varying template concentrations of template **PB1** (**Supplementary Table 2**) by Takara Taq polymerase with 8 triphosphates supplemented. All error bars shown represent the standard deviation of experimental replicates (n = 3). Linear regression statistics are reported in **Supplementary Table 8**.



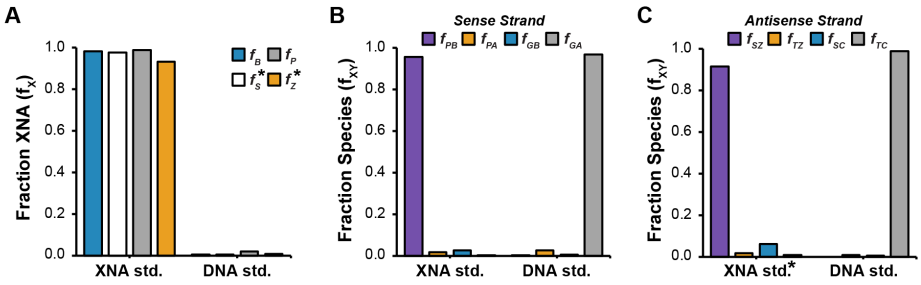
Supplementary Figure S33. Raw XNA fraction data, log plots, and source gel for 6-triphosphate PCR using 6-letter template PB1. **(A)** Average XNA fractions f_B and f_S remaining after n theoretical cycles for PCR amplification of template **PB1** by Taq polymerase with six (**ATGCBS**) nucleoside triphosphates supplemented (**P** and **Z** triphosphates excluded). **(B)** Log_2-transformed XNA fraction data used to calculate fidelity $\varphi_{\langle B,S\rangle}$ using WLS regression. $\varphi_{\langle B,S\rangle}$ ($\pm$ 95% CI, n = 3) is reported on the plot. **(C)** Average XNA fractions f_P and f_Z remaining after n theoretical cycles for PCR amplification of template **PB1** by Taq polymerase with six (**ATGCPZ**) nucleoside triphosphates supplemented (**B** and **S** triphosphates excluded). **(D)** Log_2-transformed XNA fraction data used to calculate fidelity $\varphi_{\langle P,Z\rangle}$ using WLS regression. **(E)** Agarose gel for amplification of varying template concentrations of template **PB1** by Taq polymerase with six nucleoside triphosphates supplemented (top: **P** and **Z** triphosphates excluded; bottom: **B** and **S** triphosphates excluded). All error bars shown represent standard deviation of experimental replicates (n = 3). Linear regression statistics reported in **Supplementary Table 8**.



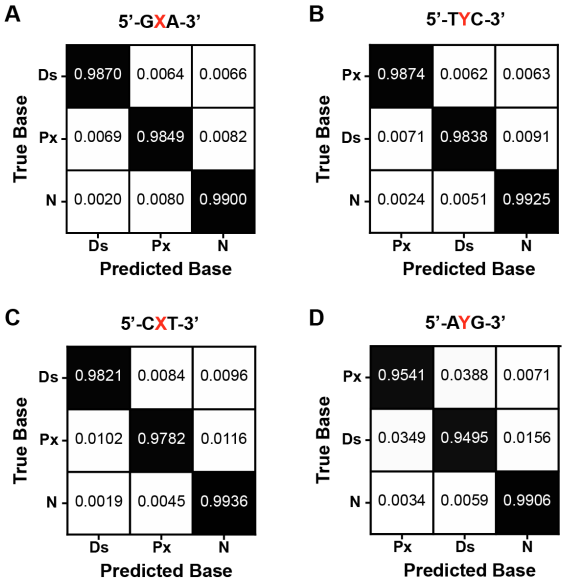
Supplementary Figure S34. Raw XNA fraction data, log plot, and source agarose gel for 8-triphosphate PCR on template P1. **(A)** Average XNA fractions f_P and f_Z remaining after n theoretical cycles for PCR amplification of ssDNA template **P1** by Taq polymerase with 8 nucleotide triphosphates supplemented (dATP, dTTP, dGTP, dCTP, dBTP, dSTP, dPTP, dZTP). **(B)** Log2-transformed XNA fraction used to calculate fidelity $\varphi_{\langle P,Z \rangle}$ using WLS regression. **(C)** Agarose source gel after amplification of varying concentrations of ssDNA template **P1** by Taq polymerase with 8 triphosphates supplemented. All error bars shown represent standard deviation of experimental replicates (n = 3). Linear regression statistics are reported in **Supplementary Table 8**.



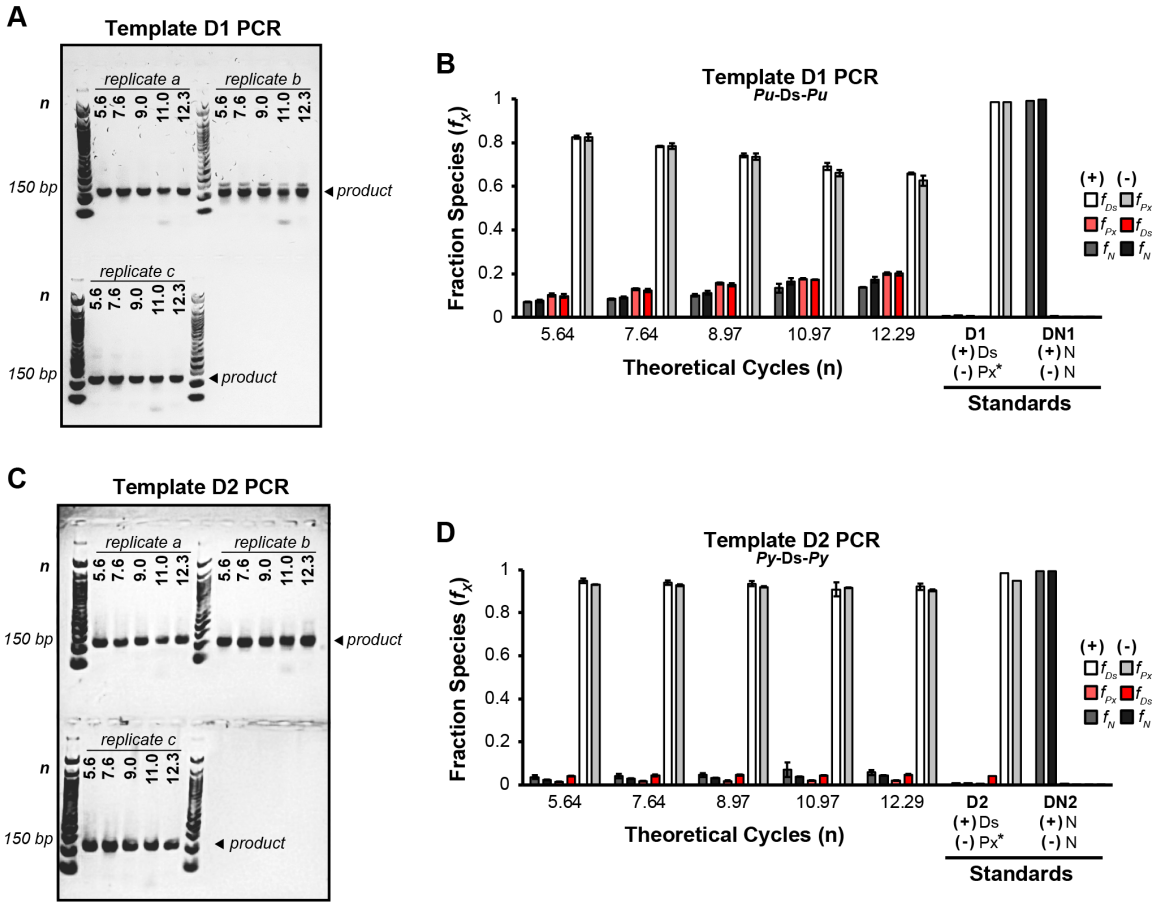
Supplementary Figure S35. Comparison of extracted signal metrics for 8-letter DNA containing the ubp XNAs B, S, P, and Z vs DNA in single-sequence contexts. Two dsDNA sequences (**PB1, PB-GA1**, **Supplementary Table 2**) differing only in two positions, **(A) S/T** and **Z/C**, and **(B) P/G** and **B/A**, were basecalled and aligned to **PB-GA1** reference sequences. Signal metrics for 500 randomly sampled reads after Remora signal refinement were extracted. Metrics calculated from normalized current values. Plots show comparison between DNA base (black) and XNA base (orange) for signal trimmed mean ionic current (Trimmed Mean), difference between DNA and XNA trimmed mean ionic current (ΔTrimmed Mean), dwell time in terms of number of average sample points collected (Dwell), standard deviation of normalized signal within the signal region (Trimmed SD), and statistical significance of a Welch's t-test performed on DNA vs XNA for trimmed mean current values at the given position ($-\log_{10}(p)$). Gray highlight denotes the location of the ubp XNAs within the original reference sequence.



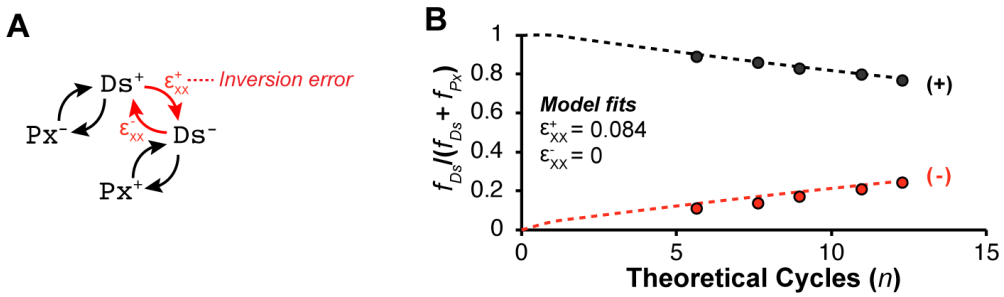
Supplementary Figure S36. Model performance on demultiplexed standards for 8-letter PCR. (A) XNA fractions f_B, f_S, f_P and f_Z called by comparative basecalling models 8L-BN2, 8L-SN2, 8L-PG1, and 8L-ZC1 (**Supplementary Table 6),** respectively, are shown for barcoded DNA standards spiked into 8-letter PCR sequencing mix. Condition labeled 'XNA std.' is the product of a primer extension on template **PB1** by Q5 polymerase (**Supplementary Table 2**, **Supplementary Fig. S29A**). Condition labeled 'DNA std.' is the product of PCR amplification of template **PB-GA1**. **(B)** Fraction species (f_{XY}) determined by combining per-read basecalls from models 8L-BN2 and 8L-PG1 applied to the original XNA positions of **B** and **P**, respectively, in template **PB1** (sense strand). **(C)** Fraction species (f_{XY}) of controls determined by combining per-read basecalls from models 8L-SN2 and 8L-ZC1 applied to the original XNA positions of **S** and **Z**, respectively, in the reverse complement of template **PB1** (*i.e.,* antisense strand). For the antisense strand, polymerase misincorporations of **S** and **Z** are possible during the initial primer extension reaction, designated by an asterisk (*).



Supplementary Figure S37. DsPx inversion multiclassification performance on test dataset. Multiclassification performance on a test dataset (withheld from training) as determined by comparing outputs of three binary comparative models to detect strand inversions. Red font '**X**' represents the XNA position on the sense strand and '**Y**' represents the XNA position on the antisense strand. Multiclassification mapping rules are shown in **Supplementary Table 10**. Confusion matrices are shown with true base (ground truth) on the vertical axis and predicted base on the horizontal axis. **Diol-Px** control was generated by primer extension from a **Ds**-containing template, leaving open the possibility of polymerase errors introduced during strand synthesis. Multiclassification performance is shown for **(A)** template **D1** sense strand, **(B)** template **D1** antisense strand, **(C)** template **D2** sense strand, **(D)** and template **D2** antisense strand.



Supplementary Figure S38. Source agarose gels and raw XNA fraction data for Ds:Diol-Px PCR. (A) Agarose gel corresponding to PCR amplification of template **D1**, with data presented in **Fig. 6D**. The number of theoretical cycles (*n*) corresponding to each replicate group is labeled on the gel. Each reaction was performed in triplicate, as labeled by replicates a, b, and c. 50 bp DNA Ladder (NEB) is included for reference. The intended product is labeled to the right of each gel. **(B)** Raw fraction species for **Ds** (f_{Ds}), **Px** (f_{Px}), and **N** (f_N; N=A,T,G,C) located on the sense (+) or antisense (-) strands, for various theoretical cycles, are shown for amplification of template **D1** by AccuPrime *Pfx* polymerase (**Fig. 6D**). Raw fractions were determined by comparing per-read basecalls from three binary comparative models to characterize inversions, following mapping rules shown in **Supplementary Table 10**. Standards include primer extension of **Ds**-containing template **D1** (dDiol-PxTP supplemented) and standard PCR amplification of template **DN1**. Asterisk (*) represents potential polymerase misincorporation of **Diol-Px** in the standard due to polymerase errors. **(C)** Agarose gel for PCR amplification of template **D2**, corresponding to data presented in **Fig. 6C**. **(D))** Raw fraction species for **Ds** (f_{Ds}), **Px** (f_{Px}), and **N** (f_N; N=A,T,G,C) located on the sense (+) or antisense (-) strands, for various theoretical cycles, are shown for amplification of template **D2** by AccuPrime *Pfx* polymerase (**Fig. 6C**).



Supplementary Figure S39. DsPx inversion state transition model fit without Px:Px self-pairing. **(A)** Schematic of simplified **DsPx** PCR inversion model including XNA inversion errors originating from **Ds**:**Ds** self-pairs. Superscript sign indicates the template strand for the error (+ for sense, - for antisense). **(B)** Fraction **Ds** (f_{Ds}) relative to the inversion-agnostic fraction of the XNA-containing species (f_{Ds}+f_{Px}) on sense (+) and antisense (-) strands over an increasing number of theoretical PCR cycles for template **D1** (*Pu*-Ds-*Pu*). The dashed line represents a fit of the state transition model with only **Ds**:**Ds** type self-pairs allowed (i.e., ε_{YY}^{+} and ε_{YY}^{-} = 0). All error bars shown represent the standard deviation of experimental replicates (n = 3).