Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Quantifying XNA replication fidelity using nanopore sequencing

Jorge Marchand

jmarcha@uw.edu

University of Washington https://orcid.org/0000-0002-7765-610X

Nicholas Kaplan

The University of Washington

Jayson Sumabat

The University of Washington

Jane McKelvey

University of Washington

Jeantine Lunshof

Harvard Medical School

Article

Keywords:

Posted Date: March 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8743774/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on August 25th, 2026. See the published version at https://doi.org/10.1038/s41467-026-76964-9.

Quantifying XNA replication fidelity using nanopore sequencing

Authors: Nicholas A. Kaplan[1], Jayson R. Sumabat[1], Jane V. McKelvey[2], Jeantine E. Lunshof[3], Jorge A. Marchand[1,2,4*]

Affiliations:

[1]Department of Chemical Engineering, University of Washington; Seattle, WA 98195, USA.

[2]Department of Chemistry, University of Washington; Seattle, WA 98195, USA.

[3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA.

[4]Molecular Engineering & Sciences Institute, University of Washington; Seattle, WA 98195, USA.

*Corresponding author. Email: jmarcha@uw.edu

Abstract

Expanded genetic alphabets built from unnatural base-pairing xenonucleic acids, or ubp XNAs, are part of an emerging frontier in biotechnology, therapeutics, and synthetic biology. However, their adoption has been constrained by replication fidelities that remain substantially lower than those of standard, natural DNA. We present a method for rapidly and accurately measuring ubp XNA replication fidelity using single-molecule nanopore sequencing. By coupling nanopore sequencing with post hoc machine-learning classifiers, we infer strand-resolved, cycle-dependent fidelity parameters directly from PCR amplification data and place these measurements on a rigorous mathematical footing that accounts for coupled replication dynamics. We apply this approach to three chemically distinct ubp systems—hydrogen-bonding **B≡S** and **P≡Z** pairs, and the hydrophobic **Ds**:**Diol-Px** pair. Using this method, we demonstrate: (i) rapid screening of reaction conditions that yields >98.6% per-cycle replication fidelity for the **B≡S** base pair, (ii) single-molecule tracking of replication outcomes in an 'Hachimoji' 8-letter system (**ATGCBSPZ**), and (iii) first estimates of context-dependent inversion error rates in the hydrophobic **Ds≡Px** system, revealing inversion frequencies of ≈8.5% in the 5`-*pur*-**Ds**-*pur*-3` context. Together, this approach provides a general route to quantitatively assess polymerase fidelity in expanded genetic alphabets, enabling systematic optimization across chemistries, enzymes, and reaction conditions.

Main Text

Introduction

Modern molecular biology widely operates on infrastructure implicitly optimized for a 4-letter genetic alphabet (**A**, **T**/**U**, **G**, **C**). Core operations such as sequencing, synthesis, and amplification are highly optimized and routinely performed at scale. Over the past several decades, however, structurally modified nucleotides have shown transformative potential across biotechnology, spanning therapeutic aptamers,[1–3] enhanced diagnostic tests,[4,5] and biomanufacturing of biologics.[6,7] This growing interest has motivated the synthesis and characterization of hundreds of nucleotide analogs, many of which are now commercially available as phosphoramidite or triphosphate monomers.[8–10]

Within this broader landscape of non-standard nucleic acids, a subset of nucleotides are capable of supporting an expanded genetic alphabet. Unnatural base-pairing xenonucleic acids, or ubp XNAs, are synthetic analogues specifically designed to expand the chemical rules that underlie genetic information storage and transmission. Ubp XNAs introduce new functionality through orthogonal base-pairing rules, allowing us to extend alphabets beyond the 4-letter limit (**A**=**T**, **G**≡**C**). Here, hydrogen-bonding ubp XNAs achieve orthogonality by adopting distinct donor–acceptor patterns at the purine–pyrimidine interface, while non-hydrogen-bonding systems rely instead on steric complementarity and entropic exclusion to maintain pairing selectivity (**Fig. 1A**).[11–13]

Despite their substantial structural divergence from canonical nucleotides, hydrogen-bonding and non-hydrogen-bonding ubp XNAs have been shown to be faithfully incorporated by polymerases.[14] Yet even with biochemical compatibility, replication fidelity of ubp XNAs is measurably lower than that of standard DNA (*in vitro*: 10^{-4} to 10^{-7}; *in vivo*: 10^{-9} to 10^{-10} errors per nucleotide replicated). Notably, estimates of *in vitro* ubp XNA replication fidelity have been reported to be as low as ≈ 85% (for the **B**≡**S**$^{\mathbf{n}}$ pair). On the high end, replication fidelity for ubp XNAs can exceed 98% for certain pairs (> 99.97% **Ds**:**Px**[15,16], ≈ 99.98% **dNaM**:**dTPT3**[17], 99.7% **P**≡2-F-**Z**($CONH_2$).[18] Far less is known about the *in vivo* replication fidelity of ubp XNAs, with estimates reported only for **dNaM**:**d5SICS** and **dNaM**:**dTPT3,** as high as 99.6% retention/doubling in certain sequence contexts using an engineered, semi-synthetic *E. coli* strain.[19,20] The low replication fidelity observed in expanded alphabets poses a significant

challenge for integrating ubp XNAs into molecular biology workflows and hinders the prospect of their usage for long-term biological information storage.

Prior research efforts have directly explored the molecular basis of ubp XNA replication error. Various error modes have since been attributed to structural features that allow mismatches in the pre-insertion complex during replication. One of the most well-studied examples is the error associated with template-**B** replication. Though the **keto-B** tautomer is commonly depicted in the **B≡S** base pair, **B** is also known to exhibit various minor tautomers, of which the **enol-B** tautomer (**Fig. 1B**) is proposed to form a Watson-Crick-Franklin pair with **T**.[21–23] The resulting **enol-B:T** pair leads to a **B→A**, **S→T** transition. Similarly, the H(1N) position in amino-5-nitro-(1H)-pyridin-2-one (**Z**) has a relatively low pK_a (≈ 7.8), which leads to a **Z(-H)≡G** mismatch and **P→G**, **Z→C** transition when deprotonated.[24–28] Finally, for hydrophobic bases, errors such as **Pa**:**A** mismatches (in the **Ds:Pa** system) and **Ds:Ds** self-pairs have been observed.[29,30] This latter mutation mode can lead to inversions during PCR (e.g., **Ds→Px**, **Px→Ds**). Lesser-known sources of error in the **Ds**:**Px** system, including **Ds**:**A, Ds**:**T, Ds**:**C**, and **Px**:**T** mismatches have also been observed, with their relative frequencies thought to be dependent on the surrounding sequence context.[12,29,31]

Understanding the origin of ubp XNA errors, including a mechanistic understanding of error modes, can help us improve ubp XNA replication fidelity. For example, lowering reaction pH and optimizing nucleotide concentrations effectively reduces **Z(-H)≡G** mismatching.[24,32] Alternatively, structural modifications have been used to discourage common error modes in various pairing systems (*e.g.*, **Px** > **Pa**, 2-F-**Z**($CONH_2$) > $\mathbf{Z^n}$).[12,18] For the **B≡S** pair, modifications to the **B** heterocycle, including 7-deaza-isoguanosine, 8-aza-7-deaza-isoguanosine, and 8-aza-isoguanosine, have also been proposed or shown to reduce mismatching by favoring the keto form.[23,33–36] Substituting **T** with 2-thio-**T** in PCR introduces steric clashes with **enol-B** and also decreases error.[37] Despite these improvements, ubp XNA replication fidelity still falls below desired thresholds for reliable amplification.

Although substantial progress has been made in the field, the fidelity landscape of ubp XNA replication remains poorly characterized, creating barriers to adoption by the wider scientific community. Sequence context effects on fidelity are known but incompletely characterized.[15,32,38] Various other error modes, including DNA → XNA forward mutations and XNA → XNA transversions, have been proposed but sparsely measured. Ubp XNA replication fidelity is also

further modulated by poorly understood factors such as transient polymerase stalling, polymerase active-site microenvironment, identity of divalent cation loaded in polymerase active site, and reaction conditions (including temperature, pH, buffer composition). New ubp XNA structures address some of the replication failure modes, but themselves remain incompletely characterized or commercially inaccessible. Efforts to consolidate these knowledge gaps face additional hurdles. Existing methods for measuring replication fidelity are bespoke and depend entirely on the XNA system being studied. Additionally, prior work has primarily used low-throughput techniques (*e.g.*, gel-based assays, mass spectrometry quantification, and Sanger sequencing-style di-deoxynucleotide termination assays) that are difficult to reproduce and have high margins of error.[15,18,19,39] Finally, the field lacks a unified mathematical descriptor of replication fidelity, which makes comparisons of reported values non-generalizable. Standardized, high-throughput fidelity-measurement strategies are needed to enable quantitative, unified benchmarking across XNA systems and to accelerate further optimization of base structure, polymerase choice, and other reaction conditions.

In this work, we introduce a nanopore sequencing-based strategy that uses deep learning to quantify ubp XNA replication fidelity at a single-molecule resolution. Although nanopore reads exhibit modest *de novo* basecalling accuracy (≈ 90-95%), post hoc RNN classifiers trained directly on specific sequences can detect modified bases with high accuracy (> 98%). We trained and benchmarked 33 models spanning three ubp systems (BS: **B≡S^{n}**, PZ: **P≡Z^{n}**, and DsPx: **Ds:Diol-Px**), demonstrating that this approach measures per-cycle XNA retention during PCR and resolves strand-specific replication error signatures. We further benchmark the method against experimental conditions and published values, proposing a unified analytical framework for ubp XNA polymerase fidelity measurements (**Supplementary Table 1**). Through a series of experimental vignettes, we use this high-throughput method to identify higher fidelity conditions for **BS** replication, quantitatively track formation of various molecular species in an 8-letter PCR system (**B**, **S**, **P**, **Z**), and provide robust evidence for unique error modes that appear under certain conditions in a hydrophobic ubp XNA system (**Ds:Diol-Px**).

Results

Training and Testing of Comparative XNA Basecalling Models

Nanopore sequencing is one of the few platforms capable of detection of modified and non-canonical nucleotides. In nanopore sequencing, an applied voltage drives single DNA molecules through a membrane-embedded pore protein, and the resulting ionic current is measured. The mobility of ions through the pore is modulated by the identity of the nucleotides occupying the pore volume, and ionic current is measured as single molecules of DNA traverse a membrane-bound pore protein. Commercial platforms for nanopore sequencing (Oxford Nanopore Technologies, ONT) pair a highly processive motor protein with a pore protein that typically interrogates a 9–10 nt sequence window (r10.4.1 chemistry). Prior studies have shown that ONT's motor proteins (in both r9.4.1 and r10.4.1 chemistries) can translocate a wide range of modified nucleotides including epigenetic marks and ubp XNAs.[40,41] Although ONT's default basecalling models are trained exclusively on canonical DNA (**A**, **T**, **G**, **C**), secondary analysis of the raw current signal can be used to detect deviations from the canonical sets. One of the most sensitive approaches for modified base detection uses deep learning to train comparative basecalling models, in which neural networks learn to distinguish modified from unmodified DNA based on raw current trajectories. When trained using a defined sequence context, comparative basecalling models can achieve high-accuracy classification.[42–44]

Building on this premise, we set out to develop a comparative basecalling framework to quantify ubp XNA replication fidelity. To generate training inputs, we designed and sequenced synthetic DNA-XNA oligos (**Supplementary Tables 2-5**) that contained a site-specific ubp substitutions for six ubp nucleotides: **B**, **S** (**S**=**S^{n}** variant), **P**, **Z** (**Z**=**Z^{n}** variant), **Ds**, **Px** (**Px** = **Diol-Px** variant). For each ubp XNA oligo, we also generated and sequenced the corresponding standard DNA oligos where the XNA base was substituted with the DNA base that would result from likely modes of replication error (**B→A**; **S→T**; **P→G**; **Z→C**). Since the error modes of the **Ds:Diol-Px** are less well defined, standard DNA-version oligos were generated that contained a mixture of all four standard bases at the XNA position (**Ds→N**, **Px →N**). Example normalized and segmented ionic-current traces for all six comparative datasets (**Fig. 2A, Supplementary Fig. S1-S4**) show clear, localized differences within ≈ 2-3 nt of the XNA substitution, with the magnitude and spatial extent of these perturbations varying across bases and sequence contexts.

For each comparative set, we trained deep-learning classifiers based on Long Short-Term Memory recurrent neural networks (LSTM RNNs) using a modified implementation of Remora (Oxford Nanopore Technologies). In this architecture, raw normalized ionic-current traces centered on the target position are first processed through a stack of one-dimensional convolutional layers to extract local signal features. In parallel, a one-hot–encoded representation of the signal-mapped, aligned reference sequence, expanded into kmer representation (9-mer), is passed through a separate convolutional stack. The resulting feature sequences are concatenated and fed into two LSTM layers, after which a fully connected head outputs logits and label probabilities. These probabilities correspond to the classification of a canonical (DNA) versus modified (XNA) state at the queried position (**Fig. 2B**). For this type of classification task, we found model training converged rapidly (typically within 20–30 epochs) and required only a modest number of reads (≈ 1,000 reads per model). Training statistics for all comparative basecalling models generated are summarized in **Supplementary Table 6**.

Representative confusion matrices and external test results for one set of **B→A** and **S→T** classifiers (**Fig. 2C**) illustrate that this strategy can reproducibly discriminate key replication-error modes with high accuracy. Across the 33 models we trained, most showed > 98% recall and specificity in benchmarking, underscoring their suitability for quantifying XNA retention within defined sequence contexts. For most models, > 95% recall and specificity could be achieved using a signal window size of ± 50 (**Fig. 2D**). However, the **P→G** classifier showed improved accuracy when the window size was increased to ± 125, suggesting that the required information was embedded in longer-range perturbations for this sequence context (**Supplementary Fig. S3 & S5**).

We evaluated quantitative response *in silico* by generating synthetic .pod5 files containing controlled mixtures of XNA (**BS**) and DNA (**AT**) reads. For each mixture, the predicted proportions closely matched the expected values, yielding linear responses with $R^2 \geq 0.999$ (**Supplementary Fig. S6**). Comprehensive benchmarking statistics for all comparative models are provided in **Supplementary Table 7**. Together, these results demonstrate that the comparative basecalling models are both accurate and generalizable within their training scope. We next applied these models to quantify PCR replication fidelity, first focusing on the **B≡S** pair, which is known to exhibit a high error rate attributed to the keto-enol tautomerization of **B**.

Quantification of B≡S Replication Fidelity

We first needed to establish the experimental and theoretical basis for making fidelity measurements from single-molecule reads. First, PCR needs to be performed under conditions where the number of effective doublings, or 'theoretical doublings', is defined stoichiometrically and is therefore independent of replication efficiency. This is achieved by running reactions in a primer-limited regime, using the primer-to-template ratio to set the theoretical cycle number 'n'. Additionally, though fidelity is often estimated using endpoint values, a more robust strategy is to measure the relative *loss* of XNA as a function of doubling.[37] Fractional loss is insensitive to absolute yield and initial template abundance, and less sensitive to basecalling model accuracy. Using the comparative basecalling models we measure $f_B(n)$ and $f_S(n)$, which are the fraction **B** and fraction **S** remaining after n theoretical cycles. In an ideally replicating system, $\log f(n)$ is linear. These measurements can then be mapped directly to the underlying average fidelity parameter, $\varphi_{\langle B,S\rangle}$, using **Eqns. 1-14**.

A graphical overview of PCR experiments to measure polymerase fidelity is shown in **Fig. 3A**. Complementary synthetic templates containing the **BS** pair (**Sequences B1** and **S1**, **Supplementary Table 2**) were amplified by Taq polymerase in PCR reactions containing equimolar amounts of dNTPs (dATP, dTTP, dGTP, dCTP, dBTP, and dSTP). Each reaction was performed in triplicate, using unique pairs of forward and reverse primers containing 24-mer DNA barcodes (**Supplementary Tables 3 & 4)**, allowing many samples to be multiplexed on a single nanopore flow cell. Pooled PCR products are sequenced, demultiplexed, and classified by our trained comparative sequencing models to obtain $f_B(n)$ and $f_S(n)$ from read data (**Supplementary Fig. S7**). Individual sequencing runs also contained barcoded synthetic DNA standards (**Supplementary Fig. S8**), used as internal controls to validate the performance of our comparative basecalling models.

Plots for $f_B(n)$ and $f_S(n)$ show exponential decay of **B** & **S** is maintained across 29 theoretical cycles, or 5 x 10^8-fold amplification (**Fig. 3B, Supplementary Fig. S9**). Notably, the measurements of $f_B(n)$ and $f_S(n)$ are strand resolved and show slight disagreement, which is consistent with asymmetric polymerase fidelities on templates **B** and **S**. We can show how this effect arises when the primary error mode is mediated by **B:T** mismatching by modeling the ODEs that govern this system (**Supplementary Fig. S10**, **Eqn. 15**). While the individual polymerase fidelities on template **B** (φ_B) and **S** (φ_S) cannot be decoupled without an estimate for either φ_B or

φ_S, the average replication fidelity, $\varphi_{\langle B,S\rangle}$, can be extracted from the slope of the log-transformed average fraction data. Under our baseline conditions with Taq polymerase, we estimate $\varphi_{\langle B,S\rangle}$ = 89.0 $\pm$ 0.4% (95% CI) per theoretical cycle (**Fig. 3C**, **Supplementary Table 8)**. φ_B can be estimated directly from $\varphi_{\langle B,S\rangle}$ measurements ($\varphi_B = \varphi^2_{\langle B,S\rangle}$) if we assume template-**B** replication errors are the primary source of mutations (*i.e.*, $\varphi_S \approx 1$). Under these assumptions, we place our own baseline estimate of $\varphi_B \approx 79\%$ per theoretical cycle.

Taken together, these results demonstrate that $f_B(n)$ and $f_S(n)$ can be measured independently and jointly used to derive $\varphi_{\langle B,S\rangle}$ with high internal consistency. Moreover, these measurements exhibit negligible flow cell-to-flow cell variability, indicating that the inferred quantities are robust to experimental replication (**Supplementary Fig. S11**). The resulting XNA fraction trajectories agree closely with the expected theoretical behavior and are obtained under an experimental design that minimizes confounding sources of error (**Supplementary Fig. S10** & **S12**). As such, deviations from linearity in the log-transformed traces can provide an early indication if model assumptions are being violated. Notably, the strand-resolved nature of the method eliminates the need for a correctly paired ubp duplex, which can become minor products (**BT** + **AS** > **BS**) due to strand-swapping as primers are consumed (**Supplementary Fig. S13**). These findings underscore the need for caution when interpreting fidelity measurements that only use an endpoint, or methods that require intact ubp duplexes (e.g., restriction digests), as such approaches may obscure the true replication dynamics of the system.

High-Throughput Screen of BS Replication Fidelity

Having validated this approach, we next took advantage of the high-throughput capacity of this strategy to improve the replication fidelity of **B**≡**S** in PCR. Though the **B**≡**S** (***S*** = iso-cytosine) pair was first reported over 30 years ago, few prior efforts have investigated improving replication fidelity during PCR. As previously mentioned, replication errors of **B**≡**S**$^{\mathbf{n}}$ base pairs have largely been attributed to base pairing between the **B** enol tautomer and **T** (**B-enol**≡**T**), with upper bound estimates of this tautomeric equilibrium constant being ≈ 10 ([keto]/[enol]), or 10% in neutral pH water.[23,35,45] During PCR, various types of additional interactions could play a major role in either shifting this equilibrium or modulating rates of dTTP extension on template **B-enol** strands, including interactions with residues in the active site, stacking interactions from adjacent nucleotides, microenvironment pH, and divalent cation (M^{2+}) loaded in polymerase.

We started this fidelity optimization campaign using the previously measured baseline conditions (Taq polymerase, equimolar dNTPs). In an initial condition screen, we performed endpoint-based fidelity measurements (12-16 theoretical cycles, **Fig. 4A**, **Supplementary Table 9**, **Supplementary Fig. S14-S18**) on the same template **BS** template (**B1**, **S1**, **Supplementary Table 2**) varying polymerase (A-family, B-family, and engineered variants), divalent cation (Mn^{2+}, Mg^{2+}, Co^{2+}), buffer pH (7.75-9.25), and buffer additives (betaine, DMSO). In a screen of 10 polymerases (**Fig. 4B**), we found Q5 had the highest end-point average fidelity (Q5: $\varphi_{\langle B,S\rangle} \approx$ 96.4%, or 3.6% error rate) compared to Taq polymerase end-point fidelity in the same screen (Taq: $\varphi_{\langle B,S\rangle} \approx 87.6\%$). No clear trends emerged for polymerase family (A vs B) or proofreading activity (± 3`-5` Exo). However, Deep Vent polymerase (DV exo+; $\varphi_{\langle B,S\rangle} \approx 92.1\%$) significantly outperformed the variant Deep Vent polymerase (DV exo−: $\varphi_{\langle B,S\rangle} \approx 87.7\%$) suggesting that 3`-5` exonuclease activity could play a role in improving net fidelity.

The nucleotide identity 3` to the template XNA has previously been shown to significantly impact the relative incorporation efficiency of dSTP opposite template-**B** by Taq polymerase. This prior work proposed that 3` neighboring nucleotide could be altering the **B** keto-enol tautomer equilibrium, with varying degrees of effect (3`-**C** >3`-**G** > 3`-**T** favoring **enol-B**).[38] However, the investigation was purely kinetic, and replication fidelity was never measured. Based on these prior findings, we measured the influence of sequence context (template `5-G**B**T-3`, 5`-G**B**C-3`, 5`-G**B**A-3`, 5`-G**B**G-3`, and 5`-C**B**G-3`) on **BS** replication fidelity for both Taq and Q5 polymerase (**Fig. 4C, Supplementary Fig. S19-S21**). For Taq polymerase, we observed $\varphi_{\langle B,S\rangle}$ trends consistent with previously reported kinetics regarding the adjacent 3`-nt (**C** > **A** = **G** > **T**). Also consistent with prior kinetic studies, we observed no significant dependence on the tested 5`-nt (**G** vs **C**). This result is mechanistically consistent with extension, in which the 5`-nucleotide is flipped out of the stacking arrangement.[38] The effects of 3`-nt were less pronounced for Q5 replication, with only slightly reduced $\varphi_{\langle B,S\rangle}$ observed for 3`-**G**.

Some of the largest improvements to $\varphi_{\langle B,S\rangle}$ have previously been achieved by tuning nucleoside triphosphate ratios in PCR. To systematically evaluate these effects, we performed a screen using Taq polymerase in which the relative amounts of dBTP/dATP and dSTP/dTTP were varied to favor formation of **B≡S** pairs and disfavor **B≡T** mismatches (**Fig. 4D**, **Supplementary Fig. S22**). The strongest gains in $\varphi_{\langle B,S\rangle}$were obtained by increasing the dSTP/dTTP ratio, consistent with **B≡T** mismatching as the dominant error pathway. Simultaneous increases to both

dSTP/dTTP and dBTP/dATP did not yield significant improvements beyond those achieved by elevating dSTP/dTTP alone. Substantial improvements in $\varphi_{\langle B,S\rangle}$ from increasing dSTP/dTTP were similarly observed in a different nucleotide context and when the PCR was performed using Q5 polymerase (**Supplementary Fig. S23 & S24**).

By combining optimal conditions for polymerase choice, sequence context, and nucleotide composition, we achieved the highest **B≡S** replication fidelity reported to date (**Supplementary Table 1**, **Supplementary Fig. S25 & S26**). Q5 polymerase in the 5′-G**B**T-3′ sequence context with a 6X dSTP/dTTP ratio produced a log-linear decay corresponding to $\varphi_{\langle B,S\rangle} = 98.3 \pm 0.2\%$ (95% CI, **Supplementary Fig. S27A-C**), consistent with > 90% **B≡S** retention over a 1,000-fold amplification (or approximately 10 theoretical cycles). Additional fidelity improvements were observed at higher nucleotide skews, with $\varphi_{\langle B,S\rangle} = 99.2 \pm 1.5\%$ (95% CI) when using 10X dSTP/dTTP and dBTP/dATP. However, agarose gel electrophoresis of these reactions revealed substantial side-product formation and incomplete amplification at later cycle numbers (**Supplementary Fig. S27D-F**), indicating diminishing returns and practical limits on fidelity improvements at extreme nucleotide ratios.

Tracing Optimization of PZ Replication Fidelity

One of the strengths of this strategy is its generalizability to different ubp XNA systems. A recent study by Wang et al. (2024) reported a replication-fidelity optimization for the **P≡Z** base pair. In their work, an enzyme-assisted sequencing strategy was developed to quantify the XNA fraction after amplification using short read (Illumina) sequencing. To reduce replication errors due to **Z** deprotonation (**Fig. 1B),** they found that lowering the buffer pH to 8.0 and increasing dPTP concentration improved **P≡Z** replication fidelity to an optimized $\varphi_{\langle P,Z\rangle}$ = 98.68% (value calculated from the reported decay rate: $(1 - \delta_{\langle P,Z\rangle})$ = 99.34%).[32]

In a single set of experiments, we were able to re-trace this prior fidelity optimization and provide an updated $\varphi_{\langle P,Z\rangle}$ estimate with higher precision and accuracy (**Fig. 5A-C**). Starting from high pH (8.8) conditions, we amplified **P≡Z** template (**P1**, **Z1**, **Supplementary Table 2**, **Supplementary Fig. S28**) using Takara Taq polymerase, which was previously found to replicate **PZ** base pair with high fidelity and amplification efficiency.[32] Consistent with the expected error mode, at this high pH we measured a low fidelity, $\varphi_{\langle P,Z\rangle} \approx 94.9\%$, or a 1 in 20 error rate. At pH 8.0, we observe the error rate decreases by half to a 1 in 40 error ($\varphi_{\langle P,Z\rangle} \approx 97.4\%$). Finally,

optimization of dPTP concentration shows further improvement to $\varphi_{\langle P,Z\rangle}$ = 97.9 ± 0.4% (95% CI). Though the original estimate for $\varphi_{\langle P,Z\rangle}$ falls within the 95% CI of the linear regression from our own measurement, the uncertainty of endpoint-based measurements is markedly larger and often not quantified (**Supplementary Table 1**). Since accurate fidelity quantification becomes increasingly challenging as replication error rates decrease, this new approach can provide the resolution needed to distinguish small, but consequential, changes in replication fidelity. High precision will inevitably be required to guide the future engineering of high-fidelity ubp systems.

Replication Fidelity of a Hachimoji DNA System

At the upper end of ubp XNA complexity are genetic alphabets composed of 8 distinct nucleotides. The Hachimoji DNA system, consisting of the canonical bases (**A**, **T**, **G**, **C**) and four ubp (**B**, **S**, **P**, **Z**), was previously shown to support both replication by DNA polymerases and transcription by a T7 RNA polymerase mutant.[11] A central challenge for such 8-letter systems is the simultaneous optimization of replication fidelity across multiple unnatural base pairs; conditions that favor high-fidelity replication of one unnatural base pair may compromise the other. Prior approaches for measuring XNA fraction are unable to measure heterogeneous populations. Nanopore sequencing, by contrast, allows us to observe changes with single-molecule resolution and provides us a general means for tracking the decay, emergence, and interconversion of various species in an 8-letter system.

To explore this capability, we performed 8-letter PCR (**Fig. 5D**) on a 6-letter starting template (**PB1**, **Supplementary Table 2**, **Supplementary Fig. S29-S36**) to generate a 'Hachimoji DNA-like' product. In this implementation, the **S** nucleotide supplied was the N-nucleoside, 5-Me-isocytosine, rather than the originally reported C-nucleoside. We trained four comparative sequencing models (**B**, **S**, **P**, **Z**) to monitor mutation of each ubp XNA to standard DNA (**B→N**, **S→N**, **P→G**, **Z→C**) in the template context of **PB1** (**Supplementary Tables 6 & 7, Supplementary Fig. S35**). Applying these models to both ubp XNA positions (**BS** and **PZ**) within an individual read revealed the progressive decay of 8-letter DNA (**BSPZ**) and appearance of both the 6-letter (**BS** & **PZ**) and 4-letter (**ATGC**-only) products (**Fig. 5E**). The XNA fraction estimates derived from the purine-containing (**BP→AG**) sense and pyrimidine-containing (**SZ→TC**) antisense strand showed strong agreement, supporting internal consistency in this approach (**Supplementary Fig. S31**). Under the tested conditions with Taq polymerase, we estimate an

effective per-cycle fidelity for joint retention of both ubp XNAs to be $\varphi_{\langle B,S|P,Z\rangle}$ = 88.8 ± 1.6% (95% CI) per theoretical cycle (**Fig. 5F**). Average replication fidelities for the individual base pairs were also calculated, with $\varphi_{\langle B,S\rangle}$ = 93.0 ± 0.7% and $\varphi_{\langle P,Z\rangle}$ = 95.6 ± 0.8% per theoretical cycle (**Supplementary Fig. S30**). While $\varphi_{\langle P,Z\rangle}$ was consistent between PCR reactions containing six (**ATGCPZ**) or eight (**ATGCBSPZ**) nucleoside triphosphates, $\varphi_{\langle B,S\rangle}$ was reduced by ≈ 2% upon the addition of dPTP and dZTP (**Supplementary Fig. S33**). Analysis of the canonical basecalls for reads classified as '**N**' by comparative **B** or **S** versus **N** models does not suggest the presence of DNA transversions indicative of 'fat' (**B**≡**P**) or 'skinny' (**S**≡**Z**) mismatches (data not shown).[46] Using Takara Taq polymerase for 8-letter PCR, we observed consistent $\varphi_{\langle B,S\rangle}$ (93.0 ± 2.4%) and improved $\varphi_{\langle P,Z\rangle}$ (96.8 ± 0.5%), resulting in an increased average joint fidelity of $\varphi_{\langle B,S|P,Z\rangle}$ = 90.4 ± 2.4% (95% CI, **Supplementary Fig. S32C**). Although the PCR conditions were not further optimized for maximum fidelity, these results demonstrate that position-resolved basecalling can quantitatively trace decay pathways for multiple molecular species, providing a practical framework to guide future analysis and optimization of increasingly complex expanded-alphabet systems.

Measuring Error Modes in Hydrophobic Base Pair Replication

Hydrophobic ubp XNAs exhibit among the highest replication fidelities for expanded genetic alphabets. For example, various prior works have reported conditions for **Ds:Px** fidelity that exceeds $\varphi_{\langle Ds,Px\rangle} > 99.7\%$ per theoretical cycle (**Supplementary Table 1**).[15,16] Despite high fidelity, hydrophobic ubp XNAs are associated with a wider diversity of replication error modes.[29,47] In particular, several hydrophobic pairs display marked sensitivity to local sequence context. For the **Ds:Px** pair, both the replication fidelity and amplification efficiency have been shown to depend on the 5` and 3` flanking nucleotides, with reduced fidelity and/or amplification efficiency reported for the purine-rich context 5`-*Pu*-**Ds**-*Pu*-3` (*Pu* = Purine).[15,16,31] In addition to XNA→DNA transversions, hydrophobic ubp XNAs have been reported to self-pair (e.g., **Ds:Ds**, *rate unknown*), leading to inversion errors in which a **Ds:Px** pair is converted to **Px:Ds**. Prior methods to measure **Ds:Px** fidelity rely on ensemble measurements and are not well-suited to distinguish between these error modes. The combination of a high apparent fidelity, strong sequence-context dependence, and multiple competing error pathways makes the **Ds:Px** system an ideal testbed for developing and validating our fidelity quantification approach.

In our experimental system for **Ds:Px**, we use **Px=Diol-Px,** a structural variant reported to display similar replication fidelity to the originally reported **Ds:Px**. We first trained two additional classes of models to discriminate between transversions, inversions, and correct incorporations. The generalized **Ds→N** and **Diol-Px→N** comparative basecalling models were trained to capture all standard DNA mutation routes by sampling all possible canonical bases at the ubp XNA position. For inversions, we trained models against synthetic DNA where the unnatural base pair was inverted at the focus position (*e.g.*, **Ds** vs **Diol-Px**). Both sets of **Ds**:**Diol-Px** models showed high accuracy in training and validation (**Supplementary Tables 6 & 7**). When applied to reads derived from PCR data, these models enable strand-specific measurements of both DNA transversion and XNA inversions (**Supplementary Table 10, Supplementary Fig. S37**).

We applied these models to PCR-amplified **Ds** templates embedded in previously-reported low-fidelity (**D1**, 5`-G**Ds**A-3`) and high-fidelity (**D2**, 5`-T**Ds**C-3`) sequence contexts (**Fig. 6A, Supplementary Table 2, Supplementary Fig. S38**).[15] Since our basecalling data is strand-resolved, the apparent fidelity values could be decomposed into their underlying error pathways. XNA → DNA transversion error terms were estimated from the observed decay of ubp XNA fraction, while XNA inversion error terms were inferred by fitting a state-space model to the residual ubp XNA decay (**Fig. 6B**). Consistent with prior reports, we observe high retention of **Ds:Diol-Px** pair in the high-fidelity context and a more rapid decay in the low-fidelity context (**Fig. 6C-D**). Notably, ubp XNA inversion events were negligible in the high-fidelity context, whereas in the low-fidelity context, ubp XNA inversions were the dominant decay product.

Fitting observations of the total fractional XNA loss, we measured fidelity accounting only for XNA → DNA transversions of $\varphi_{\langle Ds,Px\rangle} = 97.4\% \pm 1.1\%$ (95% CI) for the low-fidelity context (5`-G**Ds**A-3`) and $\varphi_{\langle Ds,Px\rangle} = 99.3\% \pm 0.3\%$ (95% CI) for the high-fidelity (5`-T**Ds**C-3`) context (**Fig. 6E**). Finally, under our tested conditions for the low-fidelity context (5`-G**Ds**A-3`), fitting observations to the state transition model suggest inversion-type errors occur in $\approx 8.4 - 8.6\%$ replication events (**Fig. 6F**, **Supplementary Fig. S39**). Since the propensity for this error is asymmetric, this leads to the formation of a new, stable population of **Ds** in a higher fidelity context (5`-T**Ds**C-3`). Taken together, these measurements represent the first quantitative decomposition of **Ds:Px** error pathways and, to our knowledge, provide the most accurate and mechanistically resolved estimates of **Ds:Px** replication fidelity reported to date. Understanding replication error

modes at this level of detail will help clarify the stability of these systems and guide the discovery or engineering of polymerases that avoid the same classes of errors.

Discussion

Despite substantial progress, replication fidelity of expanded genetic alphabets remains an active area of investigation, with current state-of-the-art systems still far from matching the fidelity of the canonical DNA alphabet. Without selective pressure, expanded alphabet bases are progressively lost over successive replication cycles. This progressive genetic erasure provides an intrinsic form of containment, operating both temporally, through loss across successive generations as a function of replication fidelity, and spatially, through dependence on access to engineered polymerases and synthetic nucleotide building blocks. While this intrinsic containment increases the biosafety of the controlled use of biological systems with expanded DNA alphabets, the use of these systems will simultaneously raise biosafety concerns, particularly regarding potential escape. While these are valid concerns in the context of engineering DNA, our work shows that the technology offers enhanced biosafety features, both in controlled settings and in the case of escape. It should be noted that either in the event of escape or in the absence of unnatural building blocks, the information encoded by ubp XNAs will ultimately be erased. This 'intrinsically' obviates the need for a precautionary approach and mitigates potential ethical concerns regarding broader application. However, as high-throughput optimization strategies, such as the approach described here, further improve the efficiency of ubp XNA integration into biological systems, it will become increasingly important to understand the long-term behavior, stability, and evolutionary persistence of expanded genetic alphabets.

Our work demonstrates that post hoc analysis of single-molecule nanopore reads can faithfully capture the fractional loss of structurally diverse ubp XNA species during replication. Under a set of assumptions, replication fidelity can be directly measured from the observed transitions between XNA and DNA states across successive replication cycles. For the broader field, the presented strategy enables high-throughput measurement of replication fidelity, systematic screening of reaction space, evaluation of polymerase variants, and rapid optimization of replication conditions for newly synthesized nucleotide structures. Additionally, since observations are state-resolved, replication fidelity can be quantified with greater robustness and

sensitivity while allowing direct decomposition of replication error modes. As expanded genetic systems approach the fidelity thresholds required for reliable amplification and information storage, such single-molecule, high-throughput strategies provide a common quantitative framework for guiding chemical and enzymatic optimization across diverse nucleotide chemistries and replication systems.

More broadly, this work illustrates how quantitative observables can be extracted from single-molecule sequencing data using widely accessible, commercial platforms. While replication fidelity is a natural and critical metric for expanded genetic alphabets, the same measurement paradigm applies to a wide range of molecular processes in which heterogeneity, rare events, and context dependence are central. Important processes, such as installation/removal of epigenetic modifications in disease models or chemically induced damage of DNA from environmental exposures, will not only be observable, but also quantifiable. As single-molecule sequencing technologies continue to mature, such approaches are expected to extend to a wider range of biochemical and biophysical measurements that are poorly captured by ensemble averages.

Acknowledgments: This work was supported by National Science Foundation Grant MCB-MCB-2419300 (NAK, JEL, JAM) and by MCB- 2440857 (JAM, JRS). Material in this work was also supported and by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-2140004 (NAK).

Contributions: Project conceptualization was performed by JAM, NAK. Methodology for this work was developed by JAM, NAK, JRS. Investigation was performed by NAK, JVM. Visualization of data and results was performed by NAK, JAM. Funding for this work was acquired by JAM, NAK. This project was administered and supervised by JAM, JEL. Writing of original draft was carried out by JAM, JEL, NAK. Reviewing and editing of the manuscript was performed by NAK, JEL, JRS, JAM.

Competing interests: The authors of this work declare no other competing financial interests.

Data availability: Raw nanopore read data (.pod5 files) analyzed in this work are archived and available upon request by contacting corresponding author. Summary of all sequencing data produced and analyzed in this work is provided in **Supplementary Data S1**.

Code availability: Code generated for analysis in this work is available on github on the xenobiolab public repository (github.com/xenobiolab).

References

1. Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S. & Hirao, I. Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol.* **31**, 453–457 (2013).
2. Biondi, E. *et al.* Laboratory evolution of artificially expanded DNA gives redesignable aptamers that target the toxic form of anthrax protective antigen. *Nucleic Acids Res.* gkw890 (2016).
3. Sawada, K. *et al.* Expanded genetic alphabet increases structural and chemical diversity of six-letter DNA for high-affinity protein-targeting aptamers. *Nat. Commun.* (2025).
4. Yaren, O. *et al.* A norovirus detection architecture based on isothermal amplification and expanded genetic systems. *J. Virol. Methods* **237**, 64–71 (2016).
5. Kawabe, H. *et al.* Harnessing Non-standard Nucleic Acids for Highly Sensitive Icosaplex (20-Plex) Detection of Microbial Threats for Environmental Surveillance. *ACS Synth. Biol.* **14**, 470–484 (2025).
6. Zhang, Y. *et al.* A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551**, 644–647 (2017).
7. Romesberg, F. E. Discovery, implications and initial use of semi-synthetic organisms with an expanded genetic alphabet/code. *Philos. Trans. R. Soc. B Biol. Sci.* **378**, 20220030 (2023).
8. Pinheiro, V. B. & Holliger, P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **16**, 245–252 (2012).
9. Ji, D., Feng, H., Liew, S. W. & Kwok, C. K. Modified nucleic acid aptamers: development, characterization, and biological applications. *Trends Biotechnol.* **41**, 1360–1384 (2023).

10. Chaput, J. C., Egli, M. & Herdewijn, P. The XNA alphabet. *Nucleic Acids Res.* **53**, gkaf635 (2025).

11. Hoshika, S. *et al.* Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* **363**, 884–887 (2019).

12. Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S. & Hirao, I. An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res.* **37**, e14–e14 (2009).

13. Noori, Z., Bermejo, A., Bofill, J. M. & Poater, J. π–π Stacking Determines the Selectivity of Unnatural DNA Base Pairs Even without Polymerase. *ACS Phys. Chem. Au* (2025).

14. Hirao, I., Kimoto, M. & Yamashige, R. Natural versus Artificial Creation of Base Pairs in DNA: Origin of Nucleobases from the Perspectives of Unnatural Base Pair Studies. *Acc. Chem. Res.* **45**, 2055–2065 (2012).

15. Yamashige, R. *et al.* Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res.* **40**, 2793–2806 (2012).

16. Okamoto, I., Miyatake, Y., Kimoto, M. & Hirao, I. High Fidelity, Efficiency and Functionalization of Ds–Px Unnatural Base Pairs in PCR Amplification for a Genetic Alphabet Expansion System. *ACS Synth. Biol.* **5**, 1220–1230 (2016).

17. Li, L. *et al.* Natural-like Replication of an Unnatural Base Pair for the Expansion of the Genetic Alphabet and Biotechnology Applications. *J. Am. Chem. Soc.* **136**, 826–829 (2014).

19. Kim, H.-J., Wenta, A. J., Dobrzycki, L. M., Biondi, E. & Benner, S. A. Improving the Fidelity of Replication of a Six-Letter DNA Alphabet. *ACS Chem. Biol.* (2025).

19. Malyshev, D. A. *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509**, 385–388 (2014).

20. Ledbetter, M. P., Karadeema, R. J. & Romesberg, F. E. Reprograming the Replisome of a Semisynthetic Organism for the Expansion of the Genetic Alphabet. *J. Am. Chem. Soc.* **140**, 758–765 (2018).
21. Blas, J. R., Luque, F. J. & Orozco, M. Unique Tautomeric Properties of Isoguanine. *J. Am. Chem. Soc.* **126**, 154–164 (2004).
22. Robinson, H. *et al.* 2‘-Deoxyisoguanosine Adopts More than One Tautomer To Form Base Pairs with Thymidine Observed by High-Resolution Crystal Structure Analysis. *Biochemistry* **37**, 10897–10905 (1998).
23. Eberlein, L. *et al.* Tautomeric Equilibria of Nucleobases in the Hachimoji Expanded Genetic Alphabet. *J. Chem. Theory Comput.* **16**, 2766–2777 (2020).
24. Yang, Z., Chen, F., Alvarado, J. B. & Benner, S. A. Amplification, Mutation, and Sequencing of a Six-Letter Synthetic Genetic System. *J. Am. Chem. Soc.* **133**, 15105–15112 (2011).
25. Reichenbach, L. F. *et al.* Structural Basis of the Mispairing of an Artificially Expanded Genetic Information System. *Chem* **1**, 946–958 (2016).
26. Smith, D. C. *et al.* Nanopores map the acid-base properties of a single site in a single DNA molecule. *Nucleic Acids Res.* **52**, 7429–7436 (2024).
27. Jerome, C. A., Hoshika, S., Bradley, K. M., Benner, S. A. & Biondi, E. In vitro evolution of ribonucleases from expanded genetic alphabets. *Proc. Natl. Acad. Sci.* **119**, e2208261119 (2022).
28. Hutter, D. & Benner, S. A. Expanding the Genetic Alphabet: Non-Epimerizing Nucleoside with the *py* DDA Hydrogen-Bonding Pattern. *J. Org. Chem.* **68**, 9839–9842 (2003).

29. Hirao, I. *et al.* An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat. Methods* **3**, 729–735 (2006).
30. Hirao, I., Mitsui, T., Kimoto, M. & Yokoyama, S. An Efficient Unnatural Base Pair for PCR Amplification. *J. Am. Chem. Soc.* **129**, 15549–15555 (2007).
31. Hamashima, K., Soong, Y. T., Matsunaga, K., Kimoto, M. & Hirao, I. DNA Sequencing Method Including Unnatural Bases for DNA Aptamer Generation by Genetic Alphabet Expansion. *ACS Synth. Biol.* **8**, 1401–1410 (2019).
32. Wang, B. *et al.* Enzyme-assisted high throughput sequencing of an expanded genetic alphabet at single base resolution. *Nat. Commun.* **15**, 4057 (2024).
33. Seela, F., Peng, X. & Li, H. Base-Pairing, Tautomerism, and Mismatch Discrimination of 7-Halogenated 7-Deaza-2‘-deoxyisoguanosine: Oligonucleotide Duplexes with Parallel and Antiparallel Chain Orientation. *J. Am. Chem. Soc.* **127**, 7739–7751 (2005).
34. Martinot, T. A. & Benner, S. A. Artificial Genetic Systems: Exploiting the “Aromaticity” Formalism To Improve the Tautomeric Ratio for Isoguanosine Derivatives. *J. Org. Chem.* **69**, 3972–3975 (2004).
35. Karalkar, N. B., Khare, K., Molt, R. & Benner, S. A. Tautomeric equilibria of isoguanine and related purine analogs. *Nucleosides Nucleotides Nucleic Acids* **36**, 256–274 (2017).
36. Karalkar, N. B., Leal, N. A., Kim, M.-S., Bradley, K. M. & Benner, S. A. Synthesis and Enzymology of 2′-Deoxy-7-deazaisoguanosine Triphosphate and Its Complement: A Second Generation Pair in an Artificially Expanded Genetic Information System. *ACS Synth. Biol.* **5**, 672–678 (2016).

37. Sismour, A. M. & Benner, S. A. The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res.* **33**, 5640–5646 (2005).

38. Maciejewska, A. M., Lichota, K. D. & Kuśmierek, J. T. Neighbouring bases in template influence base-pairing of isoguanine. *Biochem. J.* **369**, 611–618 (2003).

39. Johnson, S. C. A third base pair for the polymerase chain reaction: inserting isoC and isoG. *Nucleic Acids Res.* **32**, 1937–1941 (2004).

40. Liu, Q. *et al.* Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nat. Commun.* **10**, 2449 (2019).

41. Kawabe, H. *et al.* Enzymatic synthesis and nanopore sequencing of 12-letter supernumerary DNA. *Nat. Commun.* **14**, 6820 (2023).

42. Pratanwanich, P. N. *et al.* Identification of differential RNA modifications from nanopore direct RNA sequencing with xPore. *Nat. Biotechnol.* **39**, 1394–1402 (2021).

43. Lucas, M. C. *et al.* Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. *Nat. Biotechnol.* **42**, 72–86 (2024).

44. Alagna, N. *et al.* ModiDeC: a multi-RNA modification classifier for direct nanopore sequencing. *Nucleic Acids Res.* **53**, gkaf673 (2025).

45. Sepioł, J., Kazimierczuk, Z. & Shugar, D. Tautomerism of Isoguanosine and Solvent-Induced Keto-Enol Equilibrium. *Z. Für Naturforschung C* **31**, 361–370 (1976).

46. Hoshika, S. *et al.* "Skinny" and "Fat" DNA: Two New Double Helices. *J. Am. Chem. Soc.* **140**, 11655–11660 (2018).

47. Malyshev, D. A. *et al.* Solution Structure, Mechanism of Replication, and Optimization of an Unnatural Base Pair. *Chem. – Eur. J.* **16**, 12650–12659 (2010).

Main Figures

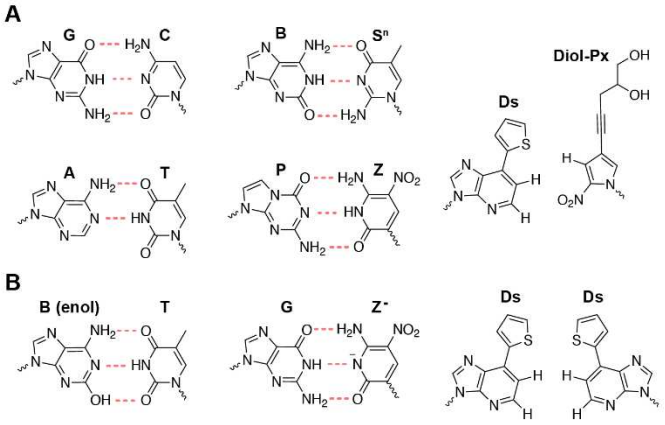


Fig. 1. Nucleobase structures and standard mutation modes of unnatural base pairing xenonucleic acids (XNAs). (A) Structures of standard DNA base pairs (**G≡C**, **A=T**), hydrogen-bonding (**B≡S**, **P≡Z**), and non-hydrogen-bonding ubp XNAs (**Ds**:**Diol-Px**) studied in this work. **S** shown is **S^{n}** (5-methyl-isocytosine). Dashed lines indicate functional groups participating in hydrogen bonding. **(B)** Putative molecular origins of replication error modes of hydrogen-bonding and non-hydrogen-bonding unnatural base pairs. During PCR, keto-enol tautomerization of **B** can lead to **B→A** and **S→T** DNA transitions while deprotonation of **Z** can lead to **P→G** and **Z→C** DNA transitions. Self-pairing of the **Ds** base can lead to XNA strand inversions (*i.e*, **Ds→Diol-Px** and **Diol-Px→ Ds**).

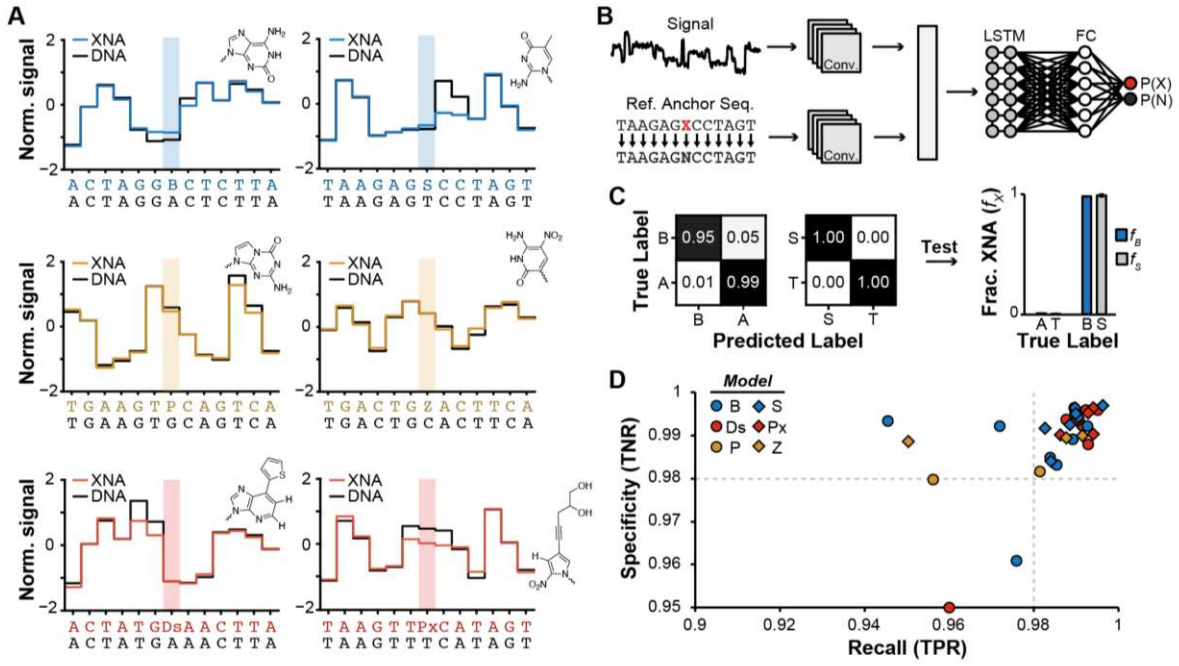


Fig. 2. Sequencing, training, and benchmarking comparative basecalling models. (A) Example segmented and normalized signal traces sequencing ubp XNAs (colored traces, top: **B**, **S**; middle: **P**, **Z**; bottom: **Ds**, **Diol-Px**) compared to a similar standard DNA sequence (black). Colored shade highlights the ubp XNA focus position in the sequence. **(B)** Overview of comparative basecalling model architecture. **(C)** Example confusion matrices generated from benchmarking comparative models for detecting **BS** error modes (**B** vs **A**; Model **BA1** and **S** vs **T**; Model **ST1**) in the 5`-G**B**C-3` context. Basecalled fraction XNA when the same models are applied to demultiplexed synthetic DNA standards containing an **AT** (barcoded **BA1**,**ST1**; **Supplementary Table 2**) or **BS** (barcoded **B1**,**S1**) base pair across five independent sequencing experiments. Error bars represent standard deviation across multiple sequencing experiments (n = 5). **(D)** Specificity (TNR, true negative rate) versus recall (TPR, true positive rate) for 33 comparative basecalling models. 98% thresholds for recall and specificity are shown with dashed lines. Full training results and benchmarking statistics for models are tabulated in **Supplementary Tables 6 & 7**.

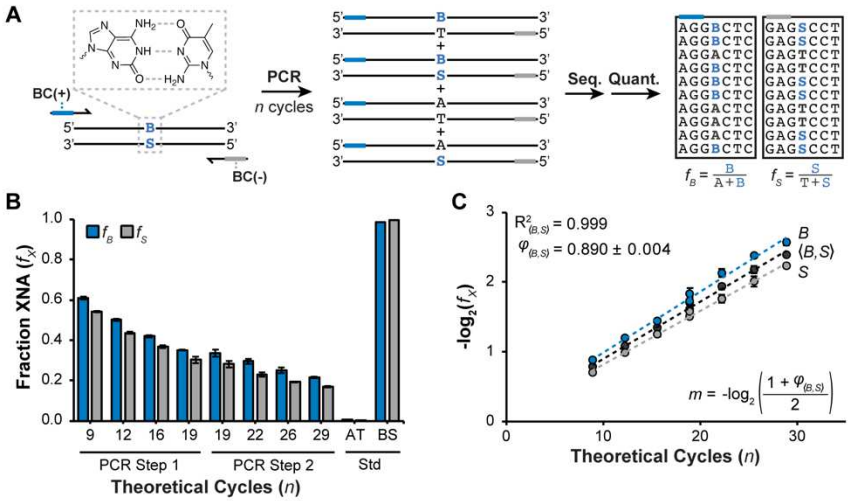


Fig. 3. Replication fidelity quantification of the BS pair. (A) Graphical overview of the procedure for performing fidelity measurements from a PCR assay. **(B)** Raw XNA fractions for **B** and **S** (f_B, f_S) are shown following PCR amplification by Taq polymerase at varying numbers of theoretical cycles. Additional data points added for theoretical cycles 18.9-28.9 were generated by 1000X dilution of PCR products from the initial round. 'Std' denotes the measured fraction for synthetic DNA standards containing an AT (barcoded **BA1 ST1**; **Supplementary Table 2**) or **BS** base pair (barcoded **B1**, **S1, Supplementary Fig. S8**). Error bars represent the standard deviation of experimental replicates (n=3). **(C)** Log transform of the ubp XNA fractions plotted on a -$\log_2$-linear scale. Dashed lines show the linear weighted least squares (WLS) regression model fits for **B** fraction (B), **S** fraction (S), and the average fraction $\langle B, S\rangle$. The average replication fidelity, $\varphi_{\langle B,S\rangle}$, is extracted from the slope of the $\langle B, S\rangle$ line (m). Error bars represent the standard deviation of experimental replicates (n =3).

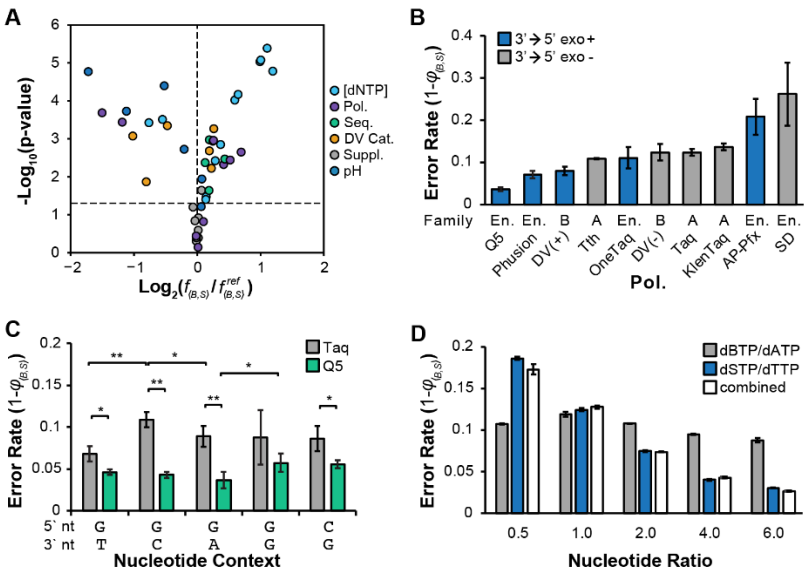


Fig. 4. High-throughput optimization of BS replication fidelity in PCR. (A) Volcano plot showing fold-change in average XNA fraction at a tested experimental condition compared to a reference condition ($f_{\langle B,S\rangle}/f^{ref}_{\langle B,S\rangle}$) after a fixed number of theoretical cycles. Tested experimental conditions varied one of: nucleotide concentration ([dNTP]), polymerase variant (Pol.), sequence context (Seq.), divalent cations (DV Cat.), buffer pH (pH), or additional PCR additives (Suppl.). P-values were calculated using two-tailed Welch's t-tests comparing screening condition and reference replicates (n = 3 per group), with significance defined as $p < 0.05$. Fold-change in XNA fraction was calculated using replicate averages. **(B)** Polymerase error rate (1- $\varphi_{\langle B,S\rangle}$) for 10 screened polymerase variants (**Supplementary Table 9A**). Error bars represent the propagated standard deviation of experimental replicates (n =3). **(C)** Polymerase error rate (1- $\varphi_{\langle B,S\rangle}$) observed for replicating **BS** pair in 5 different sequence contexts for two polymerases: Taq polymerase (Taq) and Q5 polymerase (Q5). Statistical significance is shown for selected conditions, with asterisks representing statistical significance of a two-tailed z-test (*** p<0.001, ** p<0.01, * p<0.05). Error bars represent the 95% confidence interval calculated by propagation of the standard error of the slope from WLS regression (**Supplementary Table 8**). **(D)** Observed Taq polymerase error rate for **BS** replication at varying nucleotide ratios (**Supplementary Table 9B**): dBTP/dATP, dSTP/dTTP, or both dBTP/dATP and dSTP/dTTP ('combined'). Error bars represent the propagated standard deviation of experimental replicates (n =3).

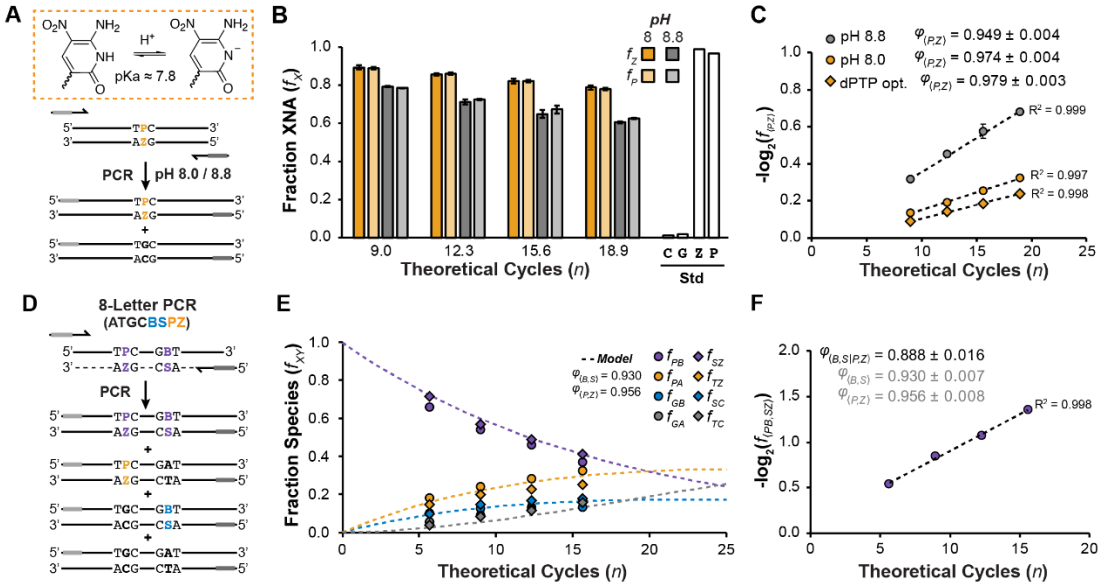


Fig. 5. Quantification of replication fidelity for 6-letter (PZ) and 8-letter (PZBS) DNA. (A) PCR performed on template **P1**,**Z1** (**Supplementary Table 2**) at pH 8.0 or pH 8.8. XNA ⟶ DNA mutation rates for the **PZ** pair are sensitive to pH due to a proposed Z deprotonation that can lead to Z(-H) ≡G pairing. **(B)** XNA fractions for **P** (f_P) and **Z** (f_Z) at pH 8.0 (orange) and pH 8.8 (gray), varying theoretical cycles, are shown following amplification by Takara Taq polymerase. XNA fractions measured for positive and negative synthetic controls are shown in white. **(C)** -Log$_2$-transformed, normalized XNA fraction data showing WLS fits and average fidelity $\varphi_{\langle P,Z\rangle}$. **(D)** 8-letter PCR performed on 6-letter DNA (**ATGCBP**) template **PB1**. **(E)** Normalized fraction species for the various 4-letter (f_{GA}, f_{TC}), 5-letter (f_{GB}, f_{SC}, f_{PA}, f_{TZ}), and 6-letter (f_{PB}, f_{SZ}), DNA molecular species over an increasing number of theoretical PCR cycles. Modeled fraction species (**Eqn. 16)** is plotted using a dashed line, with modeled average fidelities $\varphi_{\langle B,S\rangle}$ and $\varphi_{\langle P,Z\rangle}$ shown. **(F)** -Log$_2$-transformed, average fraction of 6-letter species ($f_{\langle PB,SZ\rangle}$). The average joint replication fidelity for the 8-letter species, $\varphi_{\langle B,S|P,Z\rangle}$, is extracted from the WLS regression fit. Average replication fidelities for individual ubp XNA pairs, $\varphi_{\langle B,S\rangle}$ and $\varphi_{\langle P,Z\rangle}$, were also calculated using WLS regression (values in gray; **Supplementary Fig. S30, Supplementary Table 8**). All error bars shown represent the standard deviation of experimental replicates (n = 3).

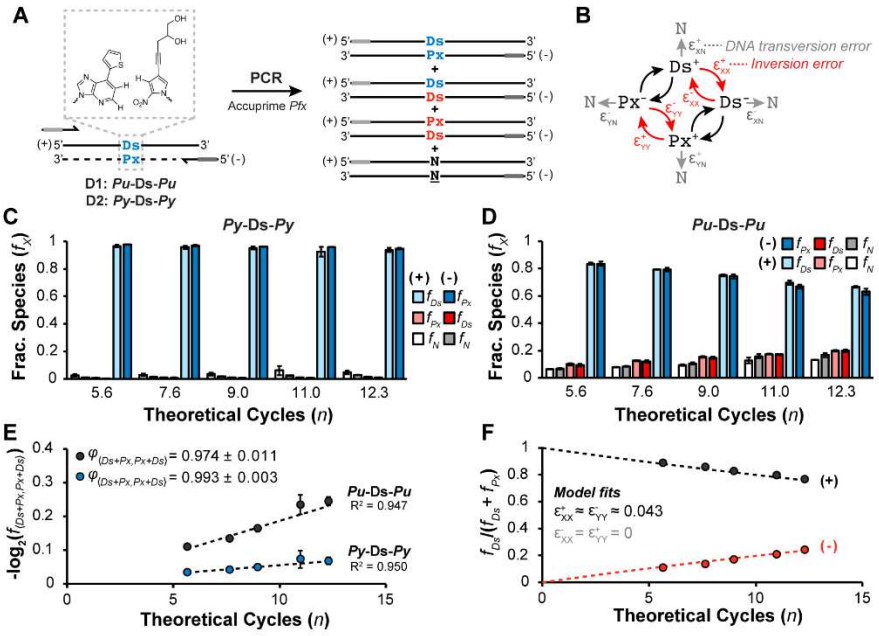


Fig. 6. Replication fidelity and strand inversion quantification DsPx PCR. (A) PCR performed on template **D1** (5`-*Pu*-**Ds**-*Pu*-3`; *Pu*=Purine) or **D2** (5`-*Py*-**Ds**-*Py*-3`; *Py*=Pyrimidine) using AccuPrime *Pfx* polymerase (**Supplementary Table 2**). **(B)** Schematic of **DsPx** PCR model including XNA → DNA transversion errors and XNA inversion errors. The superscript sign denotes the template strand for the error (+ for sense, - for antisense). **(C)** Normalized 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. **(D)** Normalized fraction species for **Ds** (f_{Ds}), **Px** (f_{Px}), and **N** (f_N) located on the sense (+) or antisense (-) strands, for various theoretical cycles, are shown for amplification of template **D1** by AccuPrime *Pfx* polymerase. **(E)** -Log_2-linear transform of the fraction of XNA-containing species (f_{Ds}+f_{Px}) averaged across strands ($f_{\langle Ds+Px,Px+Ds\rangle}$). The average transversion fidelity ($\varphi_{\langle Ds+Px,Px+Ds\rangle}$), agnostic to inversion errors, is extracted using WLS regression for amplification of templates **D1** (*Pu*-**Ds**-*Pu*) and **D2** (*Py*-**Ds**-*Py*) by AccuPrime *Pfx* polymerase. **(F)** Fraction **Ds** (f_{Ds}) relative to the 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 the fit of the state transition model with unconstrained error terms. All error bars shown represent the standard deviation of experimental replicates (n = 3).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- KaplanXNAReplicationSI.pdf