Kinetics and Fidelity of Psychrophilic DNA Polymerases

*Yuan Xue[1], Ido Braslavsky[2], Stephen R. Quake[1,3,4]**

1. Department of Bioengineering, Stanford University, Stanford, CA, USA

2. The Robert H. Smith Faculty of Agriculture, Food and Environment, Institute of Biochemistry, Food Science, and Nutrition, The Hebrew University of Jerusalem, Rehovot 7610001, Israel

3. Department of Applied Physics, Stanford University, Stanford, CA, USA.

4. Chan Zuckerberg Biohub, Mission Bay, CA, USA.

KEYWORDS: **Polymerase fidelity. Psychrophilic DNA polymerase. Temperature effect on enzymes. Polymerase error rate.**

Abstract: The discovery of extremophiles has enabled development of groundbreaking biotechnology. While most of the extremophile research is focused on thermophiles, organisms that adapt to living in cold temperature, known as psychrophiles, remain under-studied. We expressed and purified DNA polymerases PIPI and PIPB from *Psychromonas ingrahamii*, a psychrophile that can grow below water's freezing temperature. We demonstrate that they have *in vitro* DNA replication activity at temperatures as low as -19ºC. To our knowledge, this is the coolest DNA polymerization reaction ever carried out in a laboratory. In exploring the behavior of a variety of polymerases as a function of temperature, we found that reaction temperature substantially increases substitution and deletion error rates of both psychrophilic and mesophilic DNA polymerases. Motif analysis further reveals that the substitution error profiles cluster according to phylogenetic similarity of polymerases. Our results provide a useful reference for

how reaction temperature, a crucial parameter of biochemistry, can affect the fidelity of DNA polymerases adapted to a wide range of environment.

Introduction:

Life in cold environment is confronted with the thermodynamic obstacle of maintaining sufficient enzymatic activity; however, the abundance of various life-forms found in near or below freezing temperature suggests that biology has evolved ways to overcome such challenge. Contrary to what one would expect from naïve application of Arrhenius equation, life-forms that adapted to living in cold environment do not grow exponentially slower in the cold than their mesophilic cousins at room temperature[1,2]. This suggests that efficiency of psychrophilic enzymes are evolutionarily adapted to sustain relevant metabolism in low temperature[3–5]. Low temperature adaptation of enzyme is thought to increase the catalytic rate by increasing local flexibility of active site, thus lowering activation energy barrier[6–8]. This can be contrasted with high temperature adaptation which often confers global structural robustness to an enzyme by increasing intra-molecular interactions that maintain functional tertiary shape. Studies based on directed evolution have shown that structural stability and enzymatic efficiency can be optimized simultaneously and are thought to be mutually compatible properties[9,10]. Combining the observations helps to explain the general biochemical trend of cold-adapted enzymes (summarized in **Figure 1a**): an increase in enzyme's catalytic activity is often associated with a reduced binding affinity and a broader substrate specificity[11,12]. However, a direct test of these ideas on orthologous enzymes from organisms adapted to low, medium, and high temperatures has not yet been performed.

As DNA polymerase is a central component of cellular replication and biotechnological platforms[13], understanding how reaction temperature can affect fidelity is of vital interest. Polymerase error rate is tightly controlled to ensure the duplication of genetic material which can encode functional descendants. Replication error from DNA polymerase is a source of genetic variations and underlies the cause of many inheritable diseases[14–17]. Polymerase errors play a crucial role in biotechnology applications such as polymerase chain reaction (PCR) and multiple-displacement amplification (MDA)[18–20]. Polymerase errors can introduce base mismatches and indels at an exponential rate, significantly impairing downstream interpretation. In high-

53 throughput sequencing, replication errors of polymerase can reduce accuracy and precision of
54 base calling[21–23]. To test the effect of temperature on polymerase fidelity, we characterized error
55 rate of a variety of DNA polymerases as a function of temperature. Furthermore, we explored
56 how temperature adaptation affects polymerase fidelity at different reaction temperatures by
57 characterizing the error profiles of DNA polymerases of psychrophilic, mesophilic, and
58 thermophilic origins. Combining unique molecular indices (UMIs) and high-throughput
59 sequencing, we carried out a multiplexed measurement of substitution, insertion, and deletion
60 error rate at the single-molecule level for seven different DNA polymerases, some of which are
61 commonly used in molecular cloning and sequencing applications. Our data provide evidence
62 that the error rate of cold-adapted DNA polymerases are more sensitized to fluctuations in
63 reaction temperature than thermal-adapted ones. Our study reveals how temperature adaptation
64 affects psychrophilic DNA polymerase kinetics and fidelity which may enable biotechnology
65 development.

66

67 **Materials and Methods:**

68 **Polymerase Cloning.** *Psychromonas ingrahamii* 37 DNA Polymerases I and II sequences are
69 found on Joint Genome Institute, Integrated Microbial Genomes & Microbiomes (IMG/G) portal
70 (IMG id: 639633052). The sequences were codon optimized for E. coli expression. PIPI gene
71 fragment (1845 bp) was synthesized and cloned into pD454 T7 expression vector (Ampicillin
72 resistance) by ATUM (formerly DNA 2.0) to generate a maltose-binding protein (MBP) fusion
73 gene containing a single HRV 3C cleavage site. PIPB was synthesized by Integrated DNA
74 Technologies (iDT) as a single DNA fragment (2430 bp). The fragment was amplified in PCR
75 with primers containing 5'-NdeI and 3'-SalI sites, and the resulting product was digested and
76 ligated into linearized pMAL-c5x vector (New England Biolabs). All cloning products were first
77 subcloned in NEB-5α strains and then purified using ZymoPURE plasmid miniprep kits (Zymo
78 Research). The sequences of the expression constructs were all verified with Sanger sequencing
79 by Sequetech. Plasmid maps for the constructs used in this study are provided (**Supplemental
80 text 1**).

81 **Polymerase Purification.** Several expression and purification conditions had been attempted
82 to optimize yield and purity of active polymerase products. We first attempted to express a
83 glutathione S-transferase (GST)-PIPB fusion construct in BL-21 DE3 grown in Luria Broth (LB)
84 media. Isopropyl β-D-1-thiogalactopyranoside (IPTG) induction of protein expression was
85 carried out at 16ºC between 12 - 16 hours. Harvested cells were lysed with either 5 freeze-thaw
86 cycles or a sonicator. We screened a variety of lysis buffer and found that most of the protein
87 remained insoluble. We suspect that GST-tag does not promote protein solubility based on
88 previous reports[24]. In our second approach, we generated MBP-PIPI (112.4 kDa) and MBP-PIPB
89 (134.8 kDa) fusion genes on a pMAL-c5x vector expressed in Arctic Express DE3 (Agilent
90 Technologies), a bacterial expression cell line that overexpresses psychrophilic chaperones.
91 Instead of LB, we grew the bacteria in a defined media M9ZB (**Supplemental text 2**) at 37ºC
92 until OD600 of 2.5 and induced expression with 1M IPTG at 10ºC or 16ºC for 20-24 hours. Cells
93 were harvested in lysis buffer (20 mM Tris-HCl pH 7.50, 10% glycerol, 50 mM NaCl, 10 mM 2-
94 Mercaptoethanol) with the additions of 400 U of DNase I and 1X protease inhibitor cocktail (No
95 EDTA; ThermoFisher). Cells were lysed in French press homogenizer by passing through at 500
96 psi once, and at 10000 psi for three times. Clarified cell lysate was mixed with amylose resin at
97 4ºC for three hours. Resin was washed with 20 column volumes of chaperone removal buffer (20
98 mM Tris-HCl pH 7.50, 10% glycerol, 50 mM NaCl, 10 mM 2-Mercaptoethanol, 50 mM KCl, 5
99 mM ATP) and washed with 10 column volumes of lysis buffer (with no addition of DNase I or
100 protease inhibitor). Bound proteins were eluted by incubating at 4ºC for 3 hours with 10 mM of
101 maltose. MBP-PIPI was subjected to HRV-3C cleavage at 4ºC overnight. We did not cleave
102 MBP-PIPB as it was found to induce precipitation of the protein. Both proteins were further
103 purified in HiLoad 16/60 Superdex 200 preparatory grade size exclusion column (GE
104 Healthcare) to separate fractions of polymerase proteins from contaminating proteins. Purified
105 proteins were eluted in lysis buffer, concentrated in Amicon Ultra protein filters, and then stored
106 in -80ºC until experiments.

107 **Protein quantification.** Protein concentration was quantified by loading proteins on
108 denaturing SDS-PAGE gel and stained with SYPRO Red following the manufacturer's
109 recommended protocol (ThermoFisher). Linear dynamic range of the quantification was

110 established by separately staining a titration series of known amount of MBP-paramyosin-Sal. A
111 secondary quantification measurement with Bioanalyzer Protein Analysis Kits (Agilent)
112 provided similar concentration values as SYPRO Red assay (data not shown).

113 **Polymerase Activity Assay.** Polymerase activity from 5ºC to 50ºC was measured with
114 EvaEZ™ Fluorometric Polymerase Activity Assay Kit (Biotium) according to the provided
115 protocol. Briefly, assay mix and polymerase of known concentrations were equilibrated to the
116 reaction temperature separately for 5 minutes before mixing. Fluorescence signals of Eva
117 dsDNA binding dye and ROX reference dye were measured in Bio-rad 96 well qPCR machine
118 for 1 hour at constant reaction temperature with 30 seconds interval. Titrations of the polymerase
119 were performed at 20ºC to ensure linear response of measured activity at the given polymerase
120 concentration. A positive control sample with saturating amount of Klenow Exo- polymerase
121 was included to ensure that maximal fluorescence signals remained constant over the course of
122 measurement at each reaction temperature. The initial slope of the fluorescence gain for each
123 polymerase is used to estimate the steady-state speed of replication. The slopes are then
124 normalized to the maximal slope measured across all temperature range for each corresponding
125 polymerase.

126 **Gel Extension Assay.** For measurement of subzero polymerase activity, we prepared 9.5 µL of
127 primed extension mix containing a final concentration in 10 µL volume of 10 nM primed
128 template with 5'-ATTOTM 633 fluorescent label (iDT) (**Supplemental text 1**), 50 µM dNTP, 10
129 mM MOPS pH 8.50, 30% glycerol, 1.5 mM MgCl2, 0.1 mg/mL BSA, 50 mM KCl. Reaction
130 master mix and polymerase were separately equilibrated to -19ºC on a TropiCoolerTM (Boekel
131 Scientific) for 30 minutes, then they were rapidly mixed with pre-chilled pipette tips. One
132 microliter aliquot of the reaction was quenched in 9 µL of 90% formamide with 50 mM EDTA
133 pH 8.0 and heat denatured at 95ºC for 5 minutes before cooling on ice. For exonuclease assay, a
134 similar reaction was prepared but without the addition of dNTP. After thermal equilibration, the
135 polymerase was mixed with the reaction master mix and incubated on the thermal cooler. One
136 microliter of the quenched mixture was resolved on 15% polyacrylamide TBE-Urea precast gels
137 (Bio-rad) at 195 constant voltage for 50 minutes. Gels were imaged on Typhoon 9410 imager
138 (GE).

139 **Molecular Barcoding and Library Preparation.** We purified homogeneous (< 10^{-9} single
140 nucleotide polymorphism per base) pUC19 plasmid template from a single clone of NEB-5α
141 (New England Biolabs) transformed with pUC19 grown in LB supplemented with ampicillin
142 antibiotic. About 1 μg of purified pUC19 was co-digested with 10 units of SapI and NdeI at 37°C
143 for 24 minutes. Digested pUC19 template fragment was purified twice with 0.35X (V:V) of
144 Ampure XP beads. Concentration and purity of the co-digested pUC19 template was determined
145 on Bioanalyzer using high-sensitivity dsDNA quantification kit. Applying UMI barcoding
146 strategy to measure polymerase errors on single-molecule level has been reported previously by
147 the Xie group[25]. We adapted their strategy to measure polymerase replication at different reaction
148 temperatures. All reactions were conducted in a buffer made up of 10 mM MOPS pH 8.50, 30%
149 glycerol, 1.5 mM MgCl2, 0.1 mg/mL BSA, 50 mM KCl unless otherwise stated. We used MOPS
150 buffer instead of Tris-based buffer as it has a much smaller ΔpKa/ΔT scaling which minimizes
151 changes in buffer acidity at different temperatures. We increased the UMI length to 15
152 nucleotides on both ends, effectively increasing the combinatorial space to 4^{30} which minimizes
153 the barcode collision rate between UMIs (**Supplemental text 1**). After annealing UMI index
154 primer to pUC19 template (100:1 molar ratio), we pre-incubated the reaction mixture in 96 well
155 plate at the desired reaction temperature for 2 minutes prior to the addition of polymerase
156 (volume of reaction mixture to volume of polymerase is 19:1). Reaction was quenched with 50
157 mM EDTA after giving sufficient time to complete the reaction. Reaction time for each
158 polymerase and condition was determined using fluorometric activity assays and qPCR
159 measurement. We then purified the reaction mixture with Ampure beads and synthesized the
160 complementary strand using the reverse UMI primer with Q5 DNA polymerase in its native
161 reaction buffer (NEB) by incubating at 72°C for 10 minutes which is followed by EDTA
162 quenching. We performed a second round of DNA purification and quantified molecular
163 concentration of the resulting products using qPCR. A total input of 20000 barcoded molecules
164 from each reaction was amplified in two rounds of 13 PCR cycles, with Ampure XP purification
165 between each successive round of amplification. Agilent Bioanalyzer dsDNA kit was used to
166 assess purity of the amplified product.

167 **Replication Error Analysis.** Library was sequenced on MiSeq using V3 2 x 300 bp chemistry
168 kit. Each read is sorted into groups according to reaction indices. Then each read is trimmed to
169 minima of 150 length and Q score of 20 (Phred+33 quality score), ensuring that the composite
170 reads will cover the entirety of template sequence. Reads with the same reaction indices are then
171 grouped according to their UMI barcodes and the barcodes are then stripped. Reads originating
172 from the same replicated molecule would share the same UMI barcode, thus consensus for each
173 position was determined by 90% majority within the UMI family that shares the same base.
174 Bases with less than 90% majority consensus may contain PCR-induced errors and thus are not
175 included for downstream analysis. Consensus sequence with non-matching bases in the first 5
176 positions are removed to avoid analysis of mis-primed products. We observed that the apparent
177 error rate for a polymerase decreased and plateaued as consensus number increased, and we
178 picked a minimum of 5 or 10 consensus threshold for each molecule to improve accuracy in
179 replication error calling. Consensus sequence from each UMI family is then aligned to the
180 reference sequence using BWA-MEM[26]. Substitution, insertion, and deletion errors are
181 determined by comparing the consensus sequence of each UMI molecule to the pUC19 reference
182 sequence without counting ambiguous assignments.

183 **Dataset and code availability**

184 Dataset and code to reproduce the analysis will be available on public repository.

185

186 **Results:**

187 **Psychrophilic DNA polymerases can replicate nucleic acid below zero degree Celsius**

188 To better understand the effect of temperature-adaptation on polymerase fidelity, we
189 characterized replication error rates of psychrophilic, mesophilic, and thermophilic DNA
190 polymerases across a wide range of reaction temperatures. As no commercial source was
191 available for psychrophilic DNA polymerases, we determined the coding sequences
192 bioinformatically, and optimized a recombinant cloning and purification strategy for DNA

193 polymerases I and II from *Psychromonas ingrahamii*, a gram-negative bacteria previously shown
194 to grow and replicate at -12ºC in laboratory condition (**Supplementary Figure 1a**) (**Methods**)[27–
195 29]. We named the DNA polymerases I and II from this psychrophilic bacteria as *Psychromonas*
196 *ingrahamii* polymerase I *(*PIPI) and *Psychromonas ingrahamii* polymerase B *(*PIPB),
197 respectively. PIPI was purified by maltose affinity column, cleaved off its tandem-fused maltose-
198 binding protein (MBP) tag, and then re-purified with size exclusion chromatography (**Methods**).
199 As we were unable to obtain a soluble version of PIPB after protease digestion of the fusion tag,
200 we used MBP-PIPB for all of our experiments. A graphical overview of our study is summarized
201 in **Figure 1b**.

202 To determine the temperature profile of polymerase replication, we adapted a dsDNA-
203 binding fluorescent dye to detect accumulation of extended products in real-time in the presence
204 of various DNA polymerases at different reaction temperature (**Methods)**. The increase in
205 fluorescence signal over time suggests that our purified psychrophilic polymerases are indeed
206 biochemically active (left panel in **Figure 2a**). As positive controls for comparison, we show the
207 temperature profiles of Klenow fragment lacking 3'→5' exonuclease domain (KleExo-), a
208 mesophilic DNA polymerase (center panel in **Figure 2a**), and Taq polymerase, a thermophilic
209 DNA polymerase commonly used in molecular assays (right panel in **Figure 2a**). Consistent
210 with previous report, Klenow and Taq DNA polymerases are maximally active at or above 37ºC
211 [30,31]. Intriguingly, PIPI and MBP-PIPB exhibit characteristically distinct temperature profiles for
212 replication activity: PIPI reaches peak polymerase activity at 37ºC, similar to KleExo- (left panel
213 in **Figure 2a**). As observed previously, the optimal temperature of enzyme activity can differ
214 from the temperature to which the host organism is adapted to[31,32]. On the other hand, polymerase
215 activity of MBP-PIPB is optimal at 20ºC and becomes inactivated above 30ºC. Combined, our
216 results show that the temperature profiles are distinct for the two psychrophilic DNA
217 polymerases, suggesting a different physiological requirement for their catalytic rates and
218 structural robustness. This may not be surprising as family A and B polymerases are thought to
219 carry out different roles *in vivo*.

220 As *Psychromonas ingrahamii* was characterized to grow and replicate at -12ºC[28], we
221 hypothesized that its polymerases would remain active in cold environment. Using 5'-

222 fluorescently labeled, primed M13 phage template (**Methods**), we measured extension activity of
223 the two psychrophilic DNA polymerases at -19ºC in aqueous solution supplemented with 30%
224 glycerol. We discover that PIPI retains higher apparent extension activity as it completed a 5-mer
225 extension within 7 hours, while MBP-PIPB did not (center panel in **Figure 2b**). MBP-PIPB
226 failed to complete 5-mer extension at -19ºC up to 23 hours; however, the product lengths
227 noticeably increased (right panel in **Figure 2b**). As these two polymerases are predicted to
228 contain 3' → 5' exonuclease domains, we also determined their exonuclease activity (**Figure 2c**).
229 We find that MBP-PIPB can cleave single base on the timescale of minutes at 25ºC, while PIPI
230 fails to cleave at up to 2 hours, suggesting its exonuclease activity is slower than MBP-PIPB or
231 inactive (**Supplementary Figure 1b**). The exonuclease activity of family B polymerases is
232 thought to play a role in DNA repair. It may help explain the apparent discrepancy in
233 polymerization rate at sub-zero temperature between the two polymerases.

234 **Multiplexed measurement of polymerase error rate across reaction temperature**

235 In order to profile polymerase errors across a range of reaction temperature and
236 conditions, we adapted a high-throughput multiplexed sequencing approach as illustrated in
237 **Figure 3a** [25]. Briefly, a reaction mixture containing homogeneous single-stranded DNA (ssDNA)
238 phage template (NEB) is equilibrated to the desired temperature prior to the addition of DNA
239 polymerase. Unless otherwise noted, we have used 3-(*N*-morpholino)propanesulfonic acid
240 (MOPS, pH 8.5) to buffer each reaction to pH changes, which can influence enzyme fidelity
241 independently of temperature[33]. After quenching and removal of the polymerase, the
242 complementary strand is then synthesized with high-fidelity Q5 DNA polymerase to minimize
243 errors introduced in the second strand. Both the first and second strand are synthesized using
244 DNA primers that encode a predefined library index and unique molecular index (UMI) to
245 barcode the identities of each reaction and individual molecules, respectively. The dsDNA
246 products are then PCR amplified, purified, and pooled at equimolar concentrations for
247 sequencing on a MiSeq 2 x 300 bp platform. We found no runoff amplification by-products in up
248 to 26 rounds of PCR amplification, while 20 rounds of PCR amplification failed to yield
249 sufficient DNA amount for downstream purification (**Supplementary Figure 2a**). The purified
250 libraries are pooled and sequenced on Illumina MiSeq platform. We deconvolved the sequenced

251 reads by their reaction indices and UMIs, removed sequencing errors by trimming reads with low
252 quality, and called consensus sequence of molecules as described in **Methods**. Quantification of
253 base counts with high sequence consensus reveals an uniform coverage along the entire M13
254 template, indicating no substantial positional effect on measurement dropout.

255 **Reaction temperature increases error rate of psychrophilic and mesophilic DNA**
256 **polymerases**

257 To quantify the error rate of each polymerase reaction, we determined base mismatch
258 (substitution error) or indel (deletion or insertion error) in each unique molecule. As insertion
259 rate is relatively low, we focused our analysis primarily on substitution and deletion rates of
260 psychrophilic, mesophilic, and thermophilic DNA polymerases in **Figures 4a-b**. Quantification
261 of Q5 DNA polymerase error rate in its native buffer (Q5NativeBuffer) at 72ºC provides a
262 baseline for our measurement, revealing an average substitution, deletion, and insertion rates of
263 4.44×10^{-6}, 9.50×10^{-7}, and 1.37×10^{-7} per base, respectively, which agrees well with previously
264 reported values (**Table 1** and right panel in **Figure 4a**)[25,34]. Phusion (ThermoFisher), another
265 high-fidelity and commonly used thermophilic polymerase, has a similar substitution rate of 5.05
266 $\times 10^{-6}$ per base (**Table 1**) at 72ºC in its native Phusion HF buffer (PhusionNativeBuffer;
267 ThermoFisher). We discover that the substitution rate of Phusion DNA polymerase increases by
268 more than a factor of two when switched from its native buffer to MOPS, a buffer with low
269 ΔpKa/ΔT coefficient; furthermore, we notice an increased substitution rate of Phusion
270 polymerase in its native buffer (PhusionNativeBuffer) in lower reaction temperature. As the
271 composition of Phusion native buffer (ThermoFisher) is undisclosed, we postulate the ΔpKa/ΔT
272 of its buffering reagent may be high and thus its acidity can change substantially upon
273 temperature shift. which leads to changes in Phusion polymerase error rate. This would explain
274 why, relative to the Phusion native buffer, the substitution rate of Phusion in MOPS buffer
275 remains relatively invariant to temperature changes (**Table 1** and right panel in **Figure 4a**).
276 Supporting this hypothesis, we do not observe substantial differences in the substitution rate of
277 Q5 polymerase in MOPS buffer between 30ºC and 72ºC. Our finding thus suggests that
278 biological assays that involve the use of thermophilic polymerases at low temperature consider
279 the use of alternative buffering agent to ensure optimal fidelity.

280 Intriguingly, we observe a substantial temperature dependence for psychrophilic
281 polymerase and mesophilic polymerase error rate. The temperature effect is most dramatic on
282 PIPI and KleExo- whose substitution rate rose by more than 5 and 30 fold, respectively, between
283 reaction temperatures of 10°C and 50°C (**Tables 2-3**) (left and center panels in **Figures 4a-b**).
284 Similarly, the deletion rate also scales positively with reaction temperature for MBP-PIPB, PIPI,
285 and KleExo- (**Tables 2-3**) (left and center panels in **Figures 4a-b**). Curiously, the temperature
286 effect on substitution rate of KleLF and T4 polymerases is not monotonic, revealing a maxima of
287 error rate at 37°C (center panel in **Figures 2b**). In general, we find a log linear regression
288 describes well the correlation between substitution and deletion error rates across all polymerases
289 and reaction conditions (**Supplementary Figure 2b**), while insertion rate occurs at much lower
290 frequency compared to the other error types and remains largely temperature invariant (**Tables**
291 **1-3**). Next, we wondered if the substitution error spectrum is dependent on reaction temperature.
292 We show that reaction temperature had the strongest effect on substitution spectrum of KleExo-.
293 As expected, increase in reaction temperature is associated with increased frequency of
294 transversion relative to transition errors (**Figure 4c**). Meanwhile, reaction temperature had a
295 weak effect on the substitution spectrum of PIPI and MBP-PIPB and most other polymerases
296 (**Supplementary Figure 2c**).

297 **Distinct mutational profiles are driven by polymerase family type**

298 Polymerase replication error rate is context dependent[35,36]. We wondered whether
299 phylogenetically related polymerases that experienced vastly different temperature adaptation
300 histories would produce error profiles that are similar to each other across the range of reaction
301 temperature we tested. To validate the phylogenetic relations, we aligned the primary peptide
302 sequences of psychrophilic DNA polymerases to polymerases with known family assignment
303 (**Supplementary Figure 2d**). To analyze the relations between polymerase family and sequence
304 context of substitution errors, we convolved the errors by their corresponding +1 and -1 bases to
305 generate 3-mer error spectrum. We only kept 3-mer errors that occurred at greater than 20%
306 frequency in at least a single reaction, and removed the rest from further analysis. This filtering
307 step allows us to narrow down 60 distinct 3-mer errors (~32% of the entire 3-mer spectrum) and
308 focus on errors that are abundant in at least some of the reactions. We normalized the range of

error rate for each reaction: we divide the 3-mer error rate by error sum within the corresponding reaction and then apply standardization such that the mean and standard deviation of each 3-mer error rate are 0 and 1, respectively. Thus, a high Z-score indicates a particular error motif occurs more frequently in some polymerase reactions than others and *vice versa*. Hierarchical clustering of the standardized motif error rates reveals that error profiles cluster by the polymerase, indicating good concordance of error profiles for reactions at different temperatures (**Figure 5a**). Interestingly, we discover that error profiles of polymerase tended to cluster by their phylogenetic homology in which errors generated by families A and B largely separated into distinct clusters regardless of their respective temperature adaptation range. Our results thus provide evidence that conservation of polymerase sequence is coupled to the frequency of error profiles independent of temperature adaptation experience.

To identify mutational profiles that best explain the observed mutational profiles across all the reactions, we performed principal component analysis on the standardized the 3-mer error rate. This led us to identify principal components 1 and 3 that best explained the mutational profiles of the polymerase families. Projecting the data onto these two principal components reveals a natural boundary on this space that separates the mutational profiles between family A and B (**Figure 5b**). Overall, high principal component 1 score is associated with mutational profile of family B, while high principal component 3 score is associated with that of family A. To identify mutational motifs that best distinguish each polymerase family, we trained a random forest decision tree to classify origin of each 3-mer error motif (**Figure 5c**). We discover that polymerase family A more frequently produced A → G transition preceded by G/C at -1 position. On the other hand, polymerase family B more frequently produced pyrimidine (C/T) → A transversion. In addition, the error position is often preceded by a purine (A/G) in the -1 position. This suggests that conserved structural differences between the two polymerase families may underlie the propensity for polymerase to generate distinct mutational signatures.

335 **Discussion:**

336 DNA polymerases are the biological machinery underlying transmission of genetic
337 information and hold the key to many biotechnological innovations. While the majority of
338 studies have focused on polymerases of mesophilic or thermophilic origins, we are one of the
339 first to provide a comprehensive biochemical characterization of activity and fidelity of bacterial
340 DNA polymerases across a wide temperature adaptation range. To our knowledge, this is the first
341 demonstration that DNA extension can be accomplished with psychrophilic DNA polymerases
342 below water's freezing temperature (0ºC). This opens the door towards enabling more efficient
343 nucleic acid amplification techniques directly on cryopreserved samples[37]. Reaction at low
344 temperature is desirable as it can inhibit mammalian nuclease activity and reduce phototoxicity
345 and DNA damages[23,38]. Thus, the development of low temperature biotechnology should in
346 theory enable more accurate and precise measurement of biological attributes at a state that is
347 closer to the "ground-truth". Ice may also provide an unexpected environment for biological
348 reactions. As early as 1964, groups have observed that biochemical reactions can be substantially
349 catalyzed in partially frozen aqueous system due to increased concentration of reactant in the
350 aqueous portions. These observations suggest that ice slurry may be an unexpected environment
351 for the origin of living systems[39–42]. Combined with cold-active enzymes, dynamic modulation of
352 liquid-frozen phase may provide an under-exploited niche for biotechnology applications[43,44].

353 In this study, we have also explored one of the long standing questions in enzymology –
354 is there a necessary trade-off between substrate specificity and activity? To address this question,
355 we biochemically characterized bacterial DNA polymerase adapted to a wide range of
356 environmental temperatures. Our results suggest that distinct temperature adaptation can
357 differentially alter how the stability, kinetics, and fidelity of DNA polymerases change in
358 response to fluctuations of environmental temperature. Our data shows that psychrophilic
359 polymerases are more heat labile while thermophilic polymerases are largely heat resistant, as
360 consistent with previous reports of other psychrophilic enzymes[45]. Furthermore, increased
361 reaction temperature substantially increased the error rate of psychrophilic DNA polymerases
362 and *E. coli* DNA polymerase I (Klenow fragments) but not that of thermophilic polymerases
363 such as Phusion and Q5. Nucleic acid analogues are of interest to a wide range of synthetic

biology and biological engineering research as a mean to expand orthogonal signals or building parts[46]. Most often, this is achieved through a combination of rational engineering and directed evolution on mesophilic or thermophilic polymerase templates[47–50]. Our results shown here suggest that using cold-adapted polymerases as a starting template may achieve higher substrate promiscuity and thus greater incorporation rate of unnatural nucleotides. This holds the possibility to improve most sequencing-by-synthesis sequencers available on the market as well as accelerating the application of nucleic acids analogues.

Acknowledgments:

YX thanks professors James Berger and Daniel Herschlag for helpful advices and discussions. YX thanks Richard Pfuetzner, Steven Wilson, Qiangjun Zhou, and Austin Wang for sharing equipment and suggestions. YX thanks Norma Neff, Fabio Zanini, Felix Horns, Derek Croote, Mark Kowarsky, Brian Yu, Bojk Berghuis, and Nimit Jain for discussions and suggestions.

Table 1. Error rates (per base) for thermophilic DNA polymerases.

Origin	Reaction	Temperature (°C)	Substitution rate ($base^{-1}$)	Deletion rate ($base^{-1}$)	Insertion rate ($base^{-1}$)
Thermophile	Q5[NB]	72	4.44×10^{-6}	9.50×10^{-7}	1.37×10^{-7}
	Q5	30	6.04×10^{-6}	7.45×10^{-7}	5.31×10^{-7}
	Q5	72	5.38×10^{-6}	1.56×10^{-7}	5.05×10^{-8}
	Taq	30	6.32×10^{-5}	2.27×10^{-6}	2.01×10^{-6}
	Taq	50	2.18×10^{-4}	2.18×10^{-5}	3.53×10^{-6}
	Phusion[NB]	30	2.51×10^{-5}	1.82×10^{-8}	8.32×10^{-7}
	Phusion[NB]	50	8.11×10^{-6}	ND	1.95×10^{-6}
	Phusion[NB]	72	5.05×10^{-6}	ND	9.30×10^{-8}
	Phusion	30	2.02×10^{-5}	2.32×10^{-6}	4.81×10^{-7}
	Phusion	50	3.64×10^{-5}	2.46×10^{-6}	4.84×10^{-7}
	Phusion	72	1.24×10^{-5}	3.36×10^{-7}	5.11×10^{-7}

[NB]=reaction carried out in native buffer supplied by the manufacturer.

[ND]= Error frequency below the detectable error baseline.

Table 2. Error rates (per base) for psychrophilic DNA polymerases.

Origin	Reaction	Temperature (°C)	Substitution rate ($base^{-1}$)	Deletion rate ($base^{-1}$)	Insertion rate ($base^{-1}$)
Psychrophile	MBP-PIPB	10	9.52×10^{-6}	1.84×10^{-6}	1.26×10^{-6}
	MBP-PIPB	20	2.55×10^{-5}	2.20×10^{-5}	1.09×10^{-6}
	PIPI	10	3.87×10^{-5}	1.48×10^{-6}	1.62×10^{-7}
	PIPI	20	6.34×10^{-5}	4.15×10^{-6}	2.20×10^{-7}
	PIPI	37	1.83×10^{-4}	1.59×10^{-5}	ND
	PIPI	50	2.11×10^{-4}	8.47×10^{-6}	ND

[ND]= Error frequency below the detectable error baseline.

386 Table 3. Error rates (per base) for mesophilic DNA polymerases.

Origin	Reaction	Temperature (ºC)	Substitution rate ($base^{-1}$)	Deletion rate ($base^{-1}$)	Insertion rate ($base^{-1}$)
Mesophile	KleExo-	10	4.20×10^{-5}	7.65×10^{-6}	3.38×10^{-7}
	KleExo-	20	6.91×10^{-5}	8.55×10^{-6}	ND
	KleExo-	37	1.07×10^{-3}	1.75×10^{-4}	4.39×10^{-7}
	KleExo-	50	1.38×10^{-3}	1.94×10^{-4}	3.38×10^{-6}
	KleLF	10	8.53×10^{-6}	ND	2.29×10^{-8}
	KleLF	20	1.56×10^{-5}	1.12×10^{-7}	2.14×10^{-7}
	KleLF	37	2.99×10^{-5}	1.20×10^{-7}	1.43×10^{-7}
	KleLF	50	2.00×10^{-5}	ND	5.45×10^{-8}
	T4	10	3.25×10^{-5}	1.32×10^{-6}	2.37×10^{-8}
	T4	37	5.92×10^{-5}	1.17×10^{-6}	ND
	T4	50	1.45×10^{-5}	ND	ND

387 [ND]= Error frequency below the detectable error baseline.

388 **Figure 1. (a)** A schematic illustrating the apparent trade-off between enzymatic speed versus
389 fidelity in the context of temperature adaptation. **(b)** Graphical summary of the study.

390 **Figure 2. (a)** Normalized extension speed of purified psychrophilic, mesophilic, and
391 thermophilic DNA polymerases measured by a fluorometric assay. **(b)** Extension activity of
392 purified psychrophilic DNA polymerases carried out at -19ºC for 7 or 23 hours on a template
393 with 5 base gap. **(c)** Exonuclease activity of psychrophilic DNA polymerases carried out at room
394 temperature for 5, 10, or 20 minutes. Both measurements were repeated at least twice in two
395 independent experiments.

396

Figure 3. (a) A schematic illustrating the experimental overview of multiplexed single-molecule measurement of polymerase error rate at various reaction temperature. Replication errors (highlighted by blue cross) are distinguished from sequencing errors (highlighted by orange cross) by comparing consensus sequence of individual molecules to the reference sequence of M13 phage template.

402 **Figure 4.** Substitution **(a)** and deletion **(b)** error rates of psychrophilic (left), mesophilic (center),
403 and thermophilic (right) DNA polymerases as a function of reaction temperature. **(c)**
404 Temperature-dependent variation in substitution spectrum across MBP-PIPB (left), PIPI (center),
405 and KleExo- (right) polymeases.

Figure 5. (a) Heatmap of standardized 3-mer substitution rates ordered by hierarchical clustering. **(b)** Projection of polymerase reactions by principal components (PCs) 1 and 3 of standardized 3-mer substitution rates. Each dot is colored by the corresponding DNA polymerase and annotated with the reaction temperature. Dotted line is identity of PC1 and PC3 which separates substitution profiles between polymerase family A and B. **(c)** Motifs that distinguish error profiles between polymerase families A (blue) and B (red) are predicted by random forest classifier. Motifs predicted to have zero weights are removed and next we showcase here the top motifs with a difference in standard Z-score greater than 1. Logo plots reflect the frequency of non-zero weighted mutation motifs at -1, 0, and +1 positions that are enriched in polymerase families A (top) or B (bottom).

416 **Supplementary Figure 1. (a)** SDS-PAGE analysis of purified DNA polymerases. Lane L:
417 PrecisionPlus Ladder (Bio-rad). Lane 1: MBP-paramyosin protein control (NEB). Lane 2:
418 Klenow Exo- (NEB). Lane 3: Klenow LF (NEB). Lane 4: Q5 (NEB). Lane 5: Taq (NEB). Lane
419 6: MBP-PIPB (134.8 kDa). Lane 7: PIPI (69.9 kDa). **(b)** Exonuclease activity assay of PIPI and
420 MBP-PIPB at 25°C for 2 hours.

421 **Supplementary Figure 2. (a)** High-sensitivity Bioanalyzer DNA (Agilent) assay of replicated
422 products after 20 or 26 rounds of PCR amplification in the flow-through (FT) or elution fractions
423 of Ampure purification output. **(b)** Deletion and substitution rate (per base) is correlated as
424 modeled by linear regression (r^2: 0.98). **(c)** Substitution spectrum of mesophilic and thermophilic
425 DNA polymerases across reaction temperature. **(d)** Multiple sequence alignment (Clustal
426 Omega) of peptide sequences for polymerases used in this study. While Phusion DNA
427 polymerase sequence is undisclosed, it is expected to be a derivative of Pyrococcus DNA
428 polymerase II (Pfu). Family A cluster consists of Taq, PIPI, and Klenow polymerases. Family B
429 polymerases consist of T4, Pfu, and PIPB.

Corresponding Author

*Stephen R. Quake (steve@quake-lab.org)

Funding Sources

YX is supported by the Stanford Interdisciplinary Graduate Fellowship and Weiland Family Fellowship. IB acknowledges support from Stanford University and from The Hebrew University of Jerusalem. SRQ is supported by the Chan Zuckerberg Biohub.

437 **References**

438

439 1. Ratkowsky, D. A., Olley, J., McMeekin, T. A. & Ball, A. Relationship between
440 temperature and growth rate of bacterial cultures. *J. Bacteriol.* **149**, 1–5 (1982).

441 2. Ingraham, J. L. & Stokes, J. L. Psychrophilic Bacteria. *J. Bacteriol.* **76**, 75–80 (1958).

442 3. Svingor, Á., Kardos, J., Hajdú, I., Németh, A. & Závodszky, P. A better enzyme to cope
443 with cold: Comparative flexibility studies on psychrotrophic, mesophilic, and thermophilic
444 IPMDHS. *J. Biol. Chem.* **276**, 28121–28125 (2001).

445 4. Ingraham, J. L. & Bailey, G. F. Comparative Study of Effect of Temperature on
446 Metabolism of Psychrophilic and Mesophilic Bacteria. *J. Bacteriol.* **77**, 609–613 (1959).

447 5. Wintrode, P. L., Miyazaki, K. & Arnold, F. H. Cold adaptation of a mesophilic subtilisin-
448 like protease by laboratory evolution. *J. Biol. Chem.* **275**, 31635–31640 (2000).

449 6. Lonhienne, T., Gerday, C. & Feller, G. Psychrophilic enzymes: Revisiting the
450 thermodynamic parameters of activation may explain local flexibility. *Biochim. Biophys.*
451 *Acta - Protein Struct. Mol. Enzymol.* **1543**, 1–10 (2000).

452 7. Lonhienne, T. *et al.* Modular structure, local flexibility and cold-activity of a novel
453 chitobiase from a psychrophilic antarctic bacterium. *J. Mol. Biol.* **310**, 291–297 (2001).

454 8. Liang, Z. X. *et al.* Evidence for increased local flexibility in psychrophilic alcohol
455 dehydrogenase relative to its thermophilic homologue. *Biochemistry* **43**, 14676–14683
456 (2004).

457 9. Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N. & Arnold, F. H. Directed
458 evolution study of temperature adaptation in a psychrophilic enzyme. *J. Mol. Biol.* **297**,
459 1015–1026 (2000).

460 10. Nguyen, V. *et al.* Evolutionary drivers of thermoadaptation in enzyme catalysis. **294**, 289–
461 294 (2017).

462 11. Tsigos, I., Velonia, K., Smonou, I. & Bouriotis, V. Purification and characterization of an
463 alcohol dehydrogenase from the Antarctic psychrophile Moraxella sp. TAE123. *Eur. J.*
464 *Biochem.* **254**, 356–362 (1998).

465 12. D'Amico, S., Sohier, J. S. & Feller, G. Kinetics and Energetics of Ligand Binding
466 Determined by Microcalorimetry: Insights into Active Site Mobility in a Psychrophilic α-
467 Amylase. *J. Mol. Biol.* **358**, 1296–1304 (2006).

468 13. Loeb, L. A. & Monnat, R. J. DNA polymerases and human disease. *Nat. Rev. Genet.* **9**,
469 594–604 (2008).

470 14. Copeland, W. C. Inherited Mitochondrial Diseases of DNA Replication. *Annu. Rev. Med.*
471 **59**, 131–146 (2008).

472 15. Broughton, B. C. *et al.* Molecular analysis of mutations in DNA polymerase η in
473 xeroderma pigmentosum-variant patients. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 815–820
474 (2002).

475 16. Donigan, K. A. *et al.* Human POLB gene is mutated in high percentage of colorectal
476 tumors. *J. Biol. Chem.* **287**, 23830–23839 (2012).

477 17. Cabelof, D. C. *et al.* Haploinsufficiency in DNA polymerase β increases cancer risk with
478 age and alters mortality rate. *Cancer Res.* **66**, 7460–7465 (2006).

479 18. Aschenbrenner, J. & Marx, A. DNA polymerases and biotechnological applications. *Curr.*
480 *Opin. Biotechnol.* **48**, 187–195 (2017).

481 19. Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. Rapid amplification of plasmid
482 and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle
483 amplification. *Genome Res.* **11**, 1095–1099 (2001).

484 20. Chen, M. *et al.* Comparison of Multiple Displacement Amplification (MDA) and Multiple
485 Annealing and Looping-Based Amplification Cycles (MALBAC) in single-cell
486 sequencing. *PLoS One* **9**, 1–12 (2014).

487 21. Schmitt, M. W. *et al.* Detection of ultra-rare mutations by next-generation sequencing.
488 *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14508–14513 (2012).

489 22. Smith, T., Heger, A. & Sudbery, I. UMI-tools: Modeling sequencing errors in Unique
490 Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499
491 (2017).

492 23. Potapov, V. & Ong, J. L. Examining sources of error in PCR by single-molecule
493 sequencing. *PLoS One* **12**, 1–19 (2017).

494 24. Hammarstrom, M., Hellgren, N., Van den Berg, S., Berglund, H. & Hard, T. Rapid
495 screening for improved solubility of small human proteins produced as fusion proteins in
496 Escherichia coli. *Protein Sci.* xx–xxv (2001) doi:10.4018/978-1-4666-6371-8.

497 25. Lee, D. F., Lu, J., Chang, S., Loparo, J. J. & Xie, X. S. Mapping DNA polymerase errors
498 by single-molecule sequencing. *Nucleic Acids Res.* **44**, e118 (2016).

26. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **00**, 1–3 (2013).

27. Breezee, J., Cady, N. & Staley, J. T. Subfreezing growth of the sea ice bacterium 'Psychromonas ingrahamii'. *Microb. Ecol.* **47**, 300–304 (2004).

28. Auman, A. J., Breeze, J. L., Gosink, J. J., Kämpfer, P. & Staley, J. T. Psychromonas ingrahamii sp. nov., a novel gas vacuolate, psychrophilic bacterium isolated from Arctic polar sea ice. *Int. J. Syst. Evol. Microbiol.* **56**, 1001–1007 (2006).

29. Riley, M. *et al.* Genomics of an extreme psychrophile, Psychromonas ingrahamii. *BMC Genomics* **9**, 1–19 (2008).

30. Lawyer, F. C. *et al.* High-level expression, purification, and enzymatic characterization of full-length Thermus aquaticus DNA polymerase and a truncated form deficient in 5′ to 3′ exonuclease activity. *Genome Res.* **2**, 275–287 (1993).

31. Brown, H. S. & Licata, V. J. Enthalpic switch-points and temperature dependencies of DNA binding and nucleotide incorporation by Pol i DNA polymerases. *Biochim. Biophys. Acta - Proteins Proteomics* **1834**, 2133–2138 (2013).

32. Piotrowski, Y., Gurung, M. K. & Larsen, A. N. Characterization and engineering of a DNA polymerase reveals a single amino-acid substitution in the fingers subdomain to increase strand-displacement activity of A-family prokaryotic DNA polymerases. *BMC Mol. Cell Biol.* **20**, 1–11 (2019).

33. Ling, L. L., Keohavong, P., Dias, C. & Thilly, W. G. Optimization of the polymerase chain reaction with regard to fidelity: Modified T7, Taq, and vent DNA polymerases. *Genome Res.* **1**, 63–69 (1991).

34. Pezza, J. A., Kucera, R. & Sun, L. Polymerase Fidelity: What is it, and what does it mean for your PCR? *New England Biolabs (NEB) Tools & Resources.* vol. 3 4–5 (2010).

35. de Paz, A. M. *et al.* High-resolution mapping of DNA polymerase fidelity using nucleotide imbalances and next-generation sequencing. *Nucleic Acids Res.* **46**, e78 (2018).

36. Schroeder, J. W., Hirst, W. G., Szewczyk, G. A. & Simmons, L. A. The Effect of Local Sequence Context on Mutational Bias of Genes Encoded on the Leading and Lagging Strands. *Curr. Biol.* **26**, 692–697 (2016).

37. Guillaumet-Adkins, A. *et al.* Single-cell transcriptome conservation in cryopreserved cells and tissues. *Genome Biol.* **18**, 1–15 (2017).

530 38. Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce
531 single-cell RNA-seq artifacts: A molecular atlas of kidney development. *Dev.* **144**, 3625–
532 3632 (2017).

533 39. Butler, A. R. & Bruice, T. C. Catalysis in Water and Ice. A Comparison of the Kinetics of
534 Hydrolysis of Acetic Anhydride, β-Propiolactone, and p-Nitrophenyl Acetate and the
535 Dehydration of 5-Hydro-6-hydroxydeoxyuridine in Water and in Ice. *J. Am. Chem. Soc.*
536 **86**, 313–319 (1964).

537 40. Bruice, T. C. & Butler, A. R. Catalysis in Water and Ice. II.1 The Reaction of
538 Thiolactones with Morpholine in Frozen Systems. *J. Am. Chem. Soc.* **86**, 4104–4108
539 (1964).

540 41. Alburn, H. E. & Grant, N. H. Reactions in Frozen Systems. II.1 Enhanced
541 Hydroxylaminolysis of Simple Amides. *J. Am. Chem. Soc.* **87**, 4174–4177 (1965).

542 42. Grant, N. H. Biochemical Reactions In Essentially Nonaqueous Systems. **6**, 182–187
543 (1969).

544 43. Hobbs, P. V. *Ice Physics*. (Oxford University Press, 2010).

545 44. Preis, S. G. *et al.* Labyrinth ice pattern formation induced by near-infrared irradiation. *Sci.*
546 *Adv.* **5**, (2019).

547 45. Siddiqui, K. S. & Cavicchioli, R. Cold-Adapted Enzymes. *Annu. Rev. Biochem.* **75**, 403–
548 433 (2006).

549 46. Hirao, I. *et al.* An unnatural hydrophobic base pair system: Site-specific incorporation of
550 nucleotide analogs into DNA and RNA. *Nat. Methods* **3**, 729–735 (2006).

551 47. Gardner, A. F. *et al.* Therminator DNA Polymerase: Modified nucleotides and unnatural
552 substrates. *Front. Mol. Biosci.* **6**, 1–9 (2019).

553 48. Chen, T. & Romesberg, F. E. Directed polymerase evolution. *FEBS Lett.* **588**, 219–229
554 (2014).

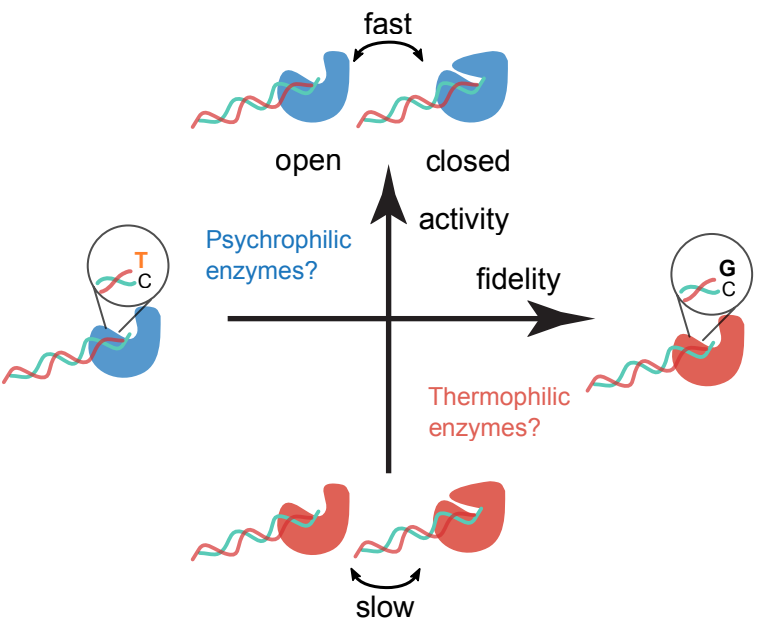
555 49. Xia, G. *et al.* Directed evolution of novel polymerase activities: mutation of a DNA
556 polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6597–
557 602 (2002).

558 50. Laos, R., Thomson, J. M. & Benner, S. A. DNA polymerases engineered by directed
559 evolution to incorporate nonstandard nucleotides. *Front. Microbiol.* **5**, 1–14 (2014).

560

Figure 1

a


fast
open
closed
activity
Psychrophilic
enzymes?
fidelity
T
C
G
C
Thermophilic
enzymes?
slow

b

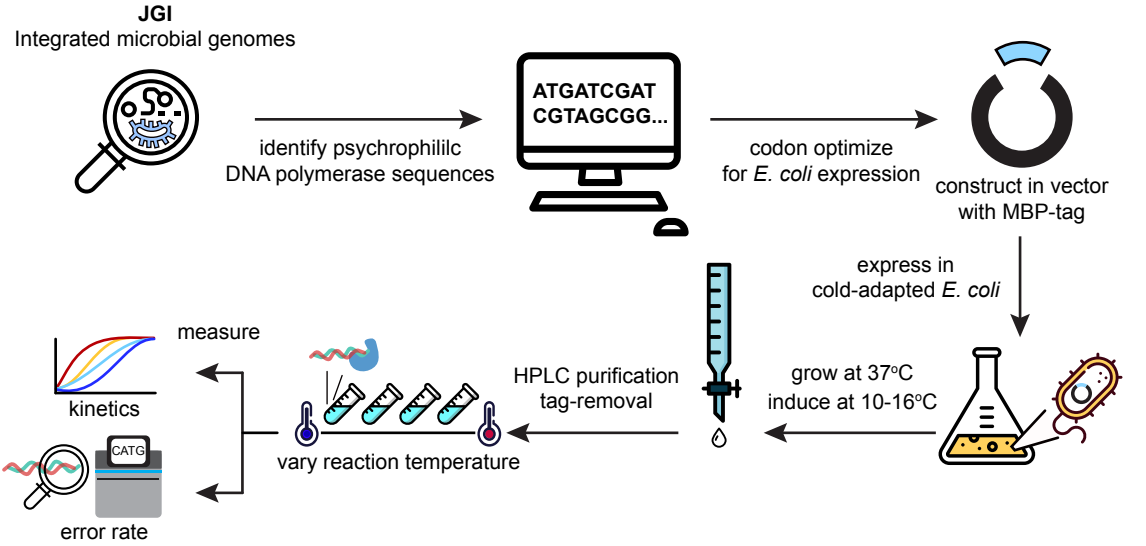
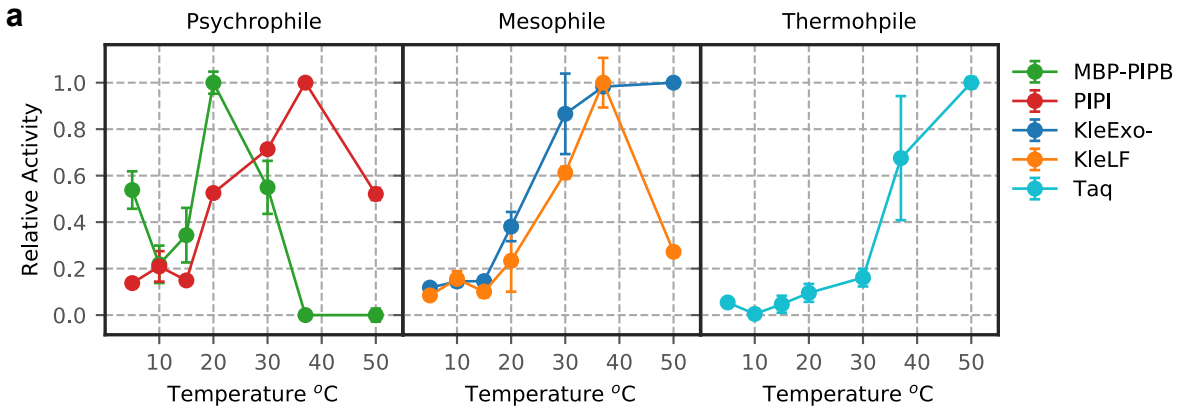
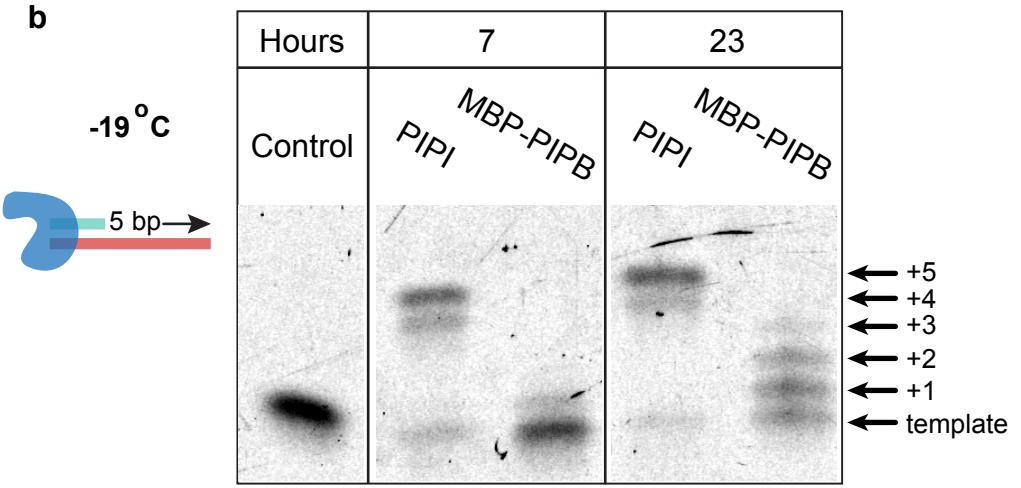

JGI
Integrated microbial genomes
identify psychrophililc
DNA polymerase sequences
ATGATCGAT
CGTAGCGG...
codon optimize
for E. coli expression
construct in vector
with MBP-tag
express in
cold-adapted E. coli
grow at 37°C
induce at 10-16°C
HPLC purification
tag-removal
vary reaction temperature
measure
kinetics
CATG
error rate

Figure 2


a
Psychrophile
Mesophile
Thermohpile
Relative Activity
1.0
0.8
0.6
0.4
0.2
0.0
10 20 30 40 50
Temperature °C
MBP-PIPB
PIPI
KleExo-
KleLF
Taq


b
-19 °C
5 bp
Hours
7
23
Control
PIPI
MBP-PIPB
PIPI
MBP-PIPB
+5
+4
+3
+2
+1
template

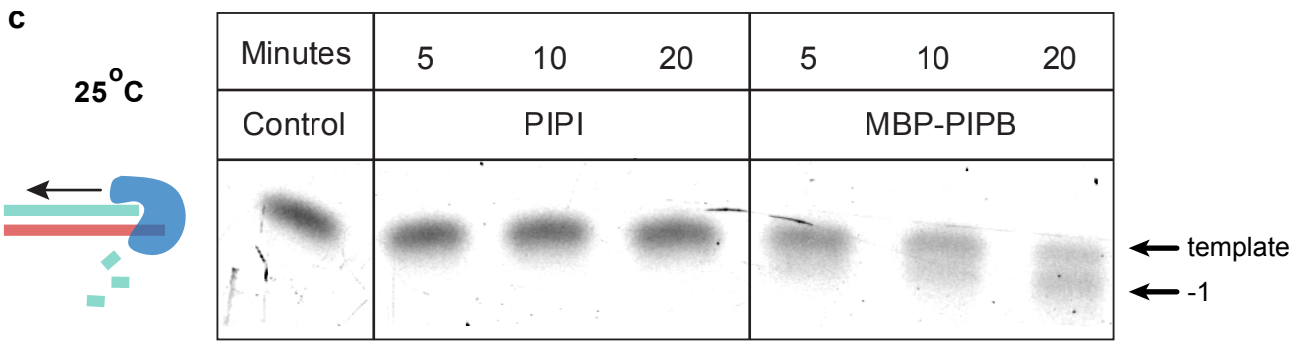

c
25 °C
Minutes
5
10
20
5
10
20
Control
PIPI
MBP-PIPB
template
-1

Figure 3

a

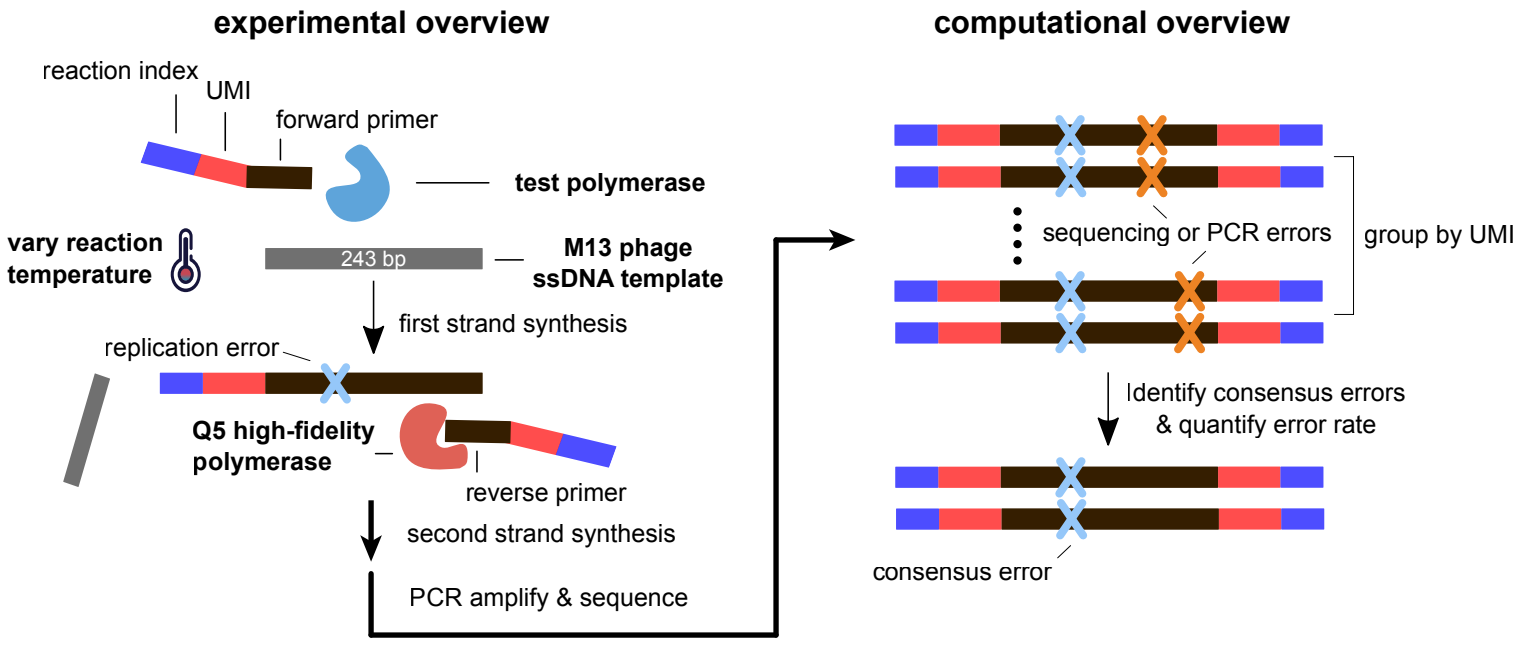

experimental overview
computational overview
reaction index
UMI
forward primer
test polymerase
vary reaction temperature
243 bp
M13 phage ssDNA template
first strand synthesis
replication error
Q5 high-fidelity polymerase
reverse primer
second strand synthesis
PCR amplify & sequence
sequencing or PCR errors
group by UMI
Identify consensus errors & quantify error rate
consensus error

Figure 4

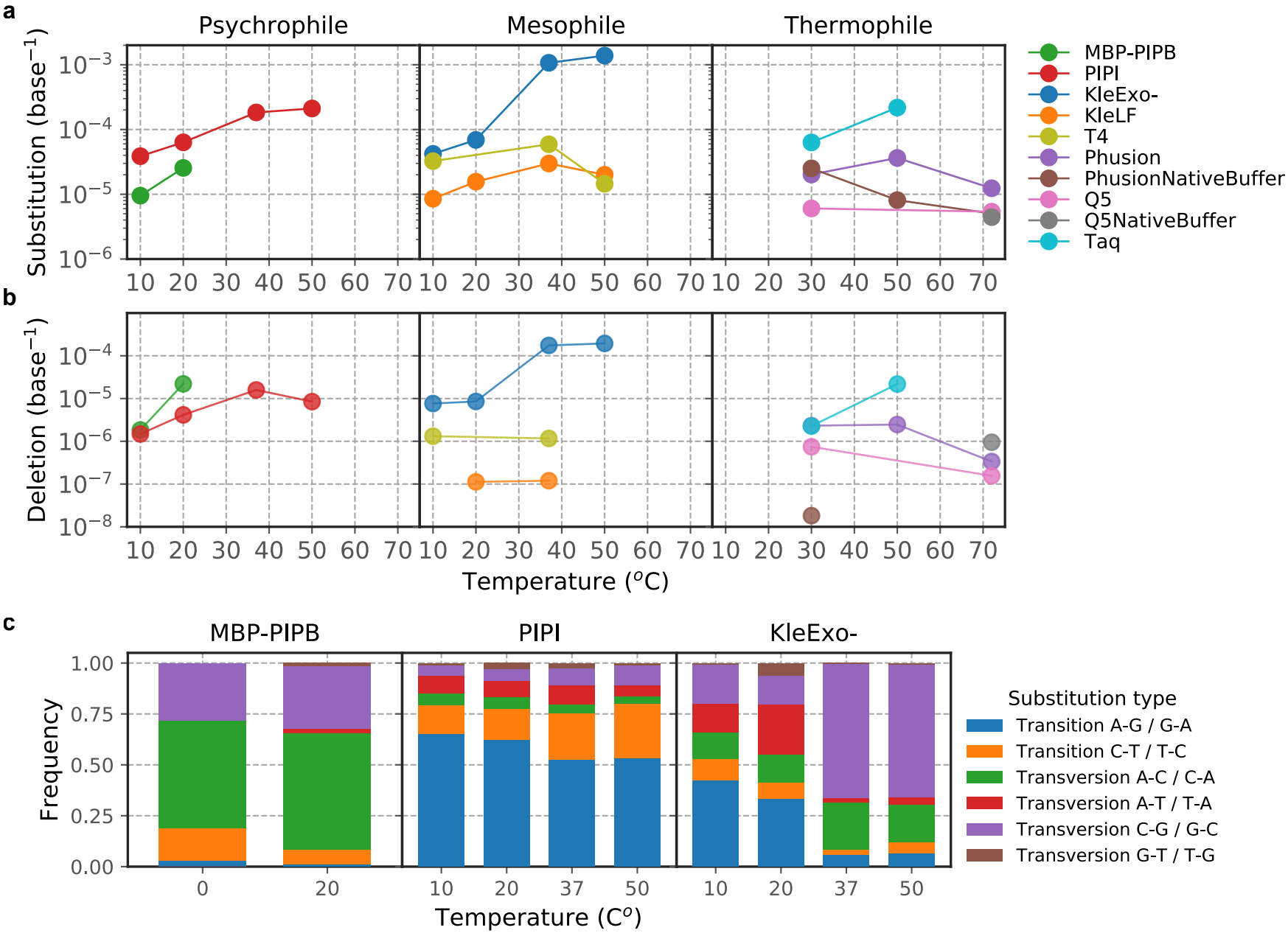
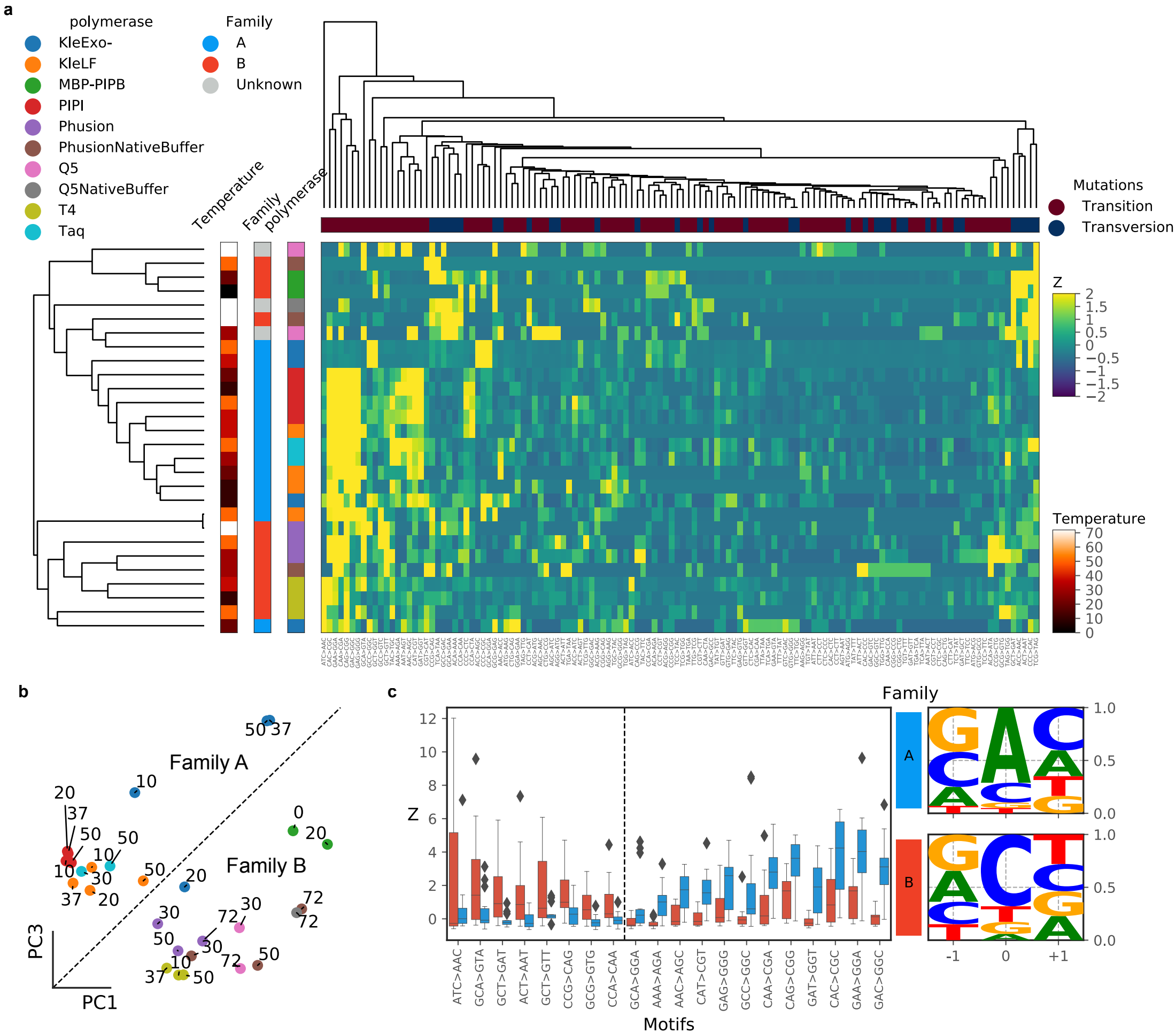
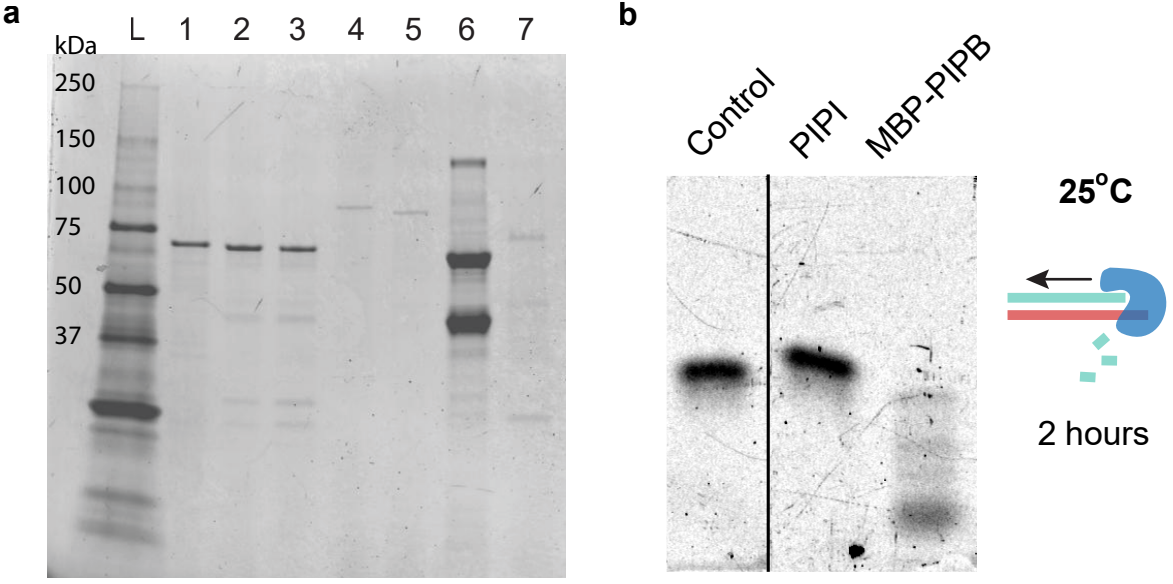

a
Psychrophile
Mesophile
Thermophile
Substitution (base^{-1})
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10 20 30 40 50 60 70
MBP-PIPB
PIPI
KleExo-
KleLF
T4
Phusion
PhusionNativeBuffer
Q5
Q5NativeBuffer
Taq
b
Deletion (base^{-1})
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
Temperature (^{o}C)
c
MBP-PIPB
PIPI
KleExo-
Frequency
1.00
0.75
0.50
0.25
0.00
0
20
10 20 37 50
Temperature (C^o)
Substitution type
Transition A-G / G-A
Transition C-T / T-C
Transversion A-C / C-A
Transversion A-T / T-A
Transversion C-G / G-C
Transversion G-T / T-G

Figure 5


a
polymerase
KleExo-
KleLF
MBP-PIPB
PIPI
Phusion
PhusionNativeBuffer
Q5
Q5NativeBuffer
T4
Taq
Family
A
B
Unknown
Temperature
Family
polymerase
Mutations
Transition
Transversion
Z
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
Temperature
70
60
50
40
30
20
10
0
b
Family A
Family B
PC3
PC1
c
Family
Z
Motifs
ATC>AAC
GCA>GTA
GCT>GAT
ACT>AAT
GCT>GTT
CCG>CAG
GCG>GTG
CCA>CAA
GCA>GGA
AAA>AGA
AAC>AGC
CAT>CGT
GAG>GGG
GCC>GGC
CAA>CGA
CAG>CGG
GAT>GGT
CAC>CGC
GAA>GGA
GAC>GGC
A
B
-1
0
+1

Supplementary Figure 1



Supplementary Figure 2

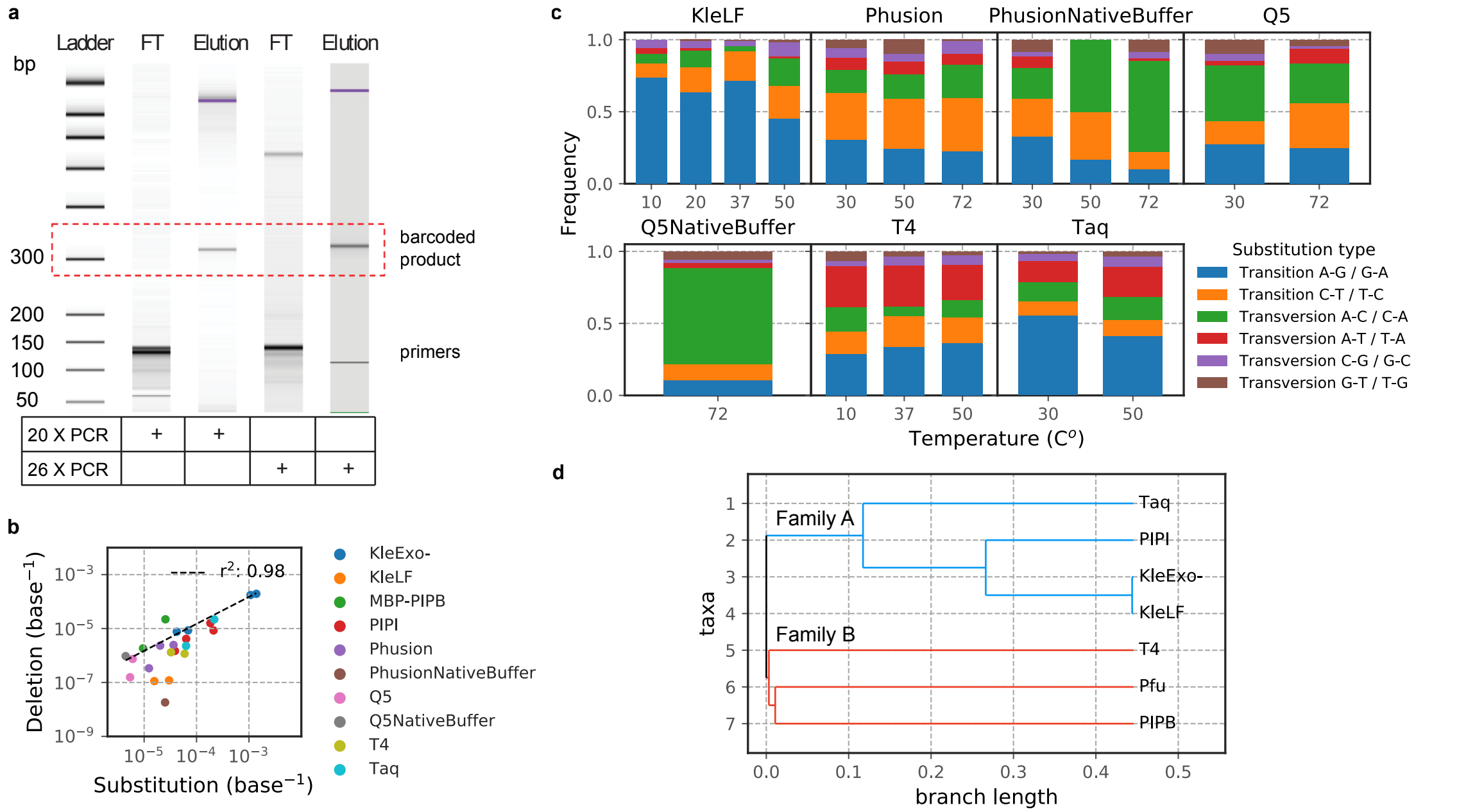


Table 1. Error rates (per base) for thermophilic DNA polymerases.

Origin	Reaction	Temperature (ºC)	Substitution rate ($base^{-1}$)	Deletion rate ($base^{-1}$)	Insertion rate ($base^{-1}$)
Thermophile	Q5[NB]	72	4.44×10^{-6}	9.50×10^{-7}	1.37×10^{-7}
	Q5	30	6.04×10^{-6}	7.45×10^{-7}	5.31×10^{-7}
	Q5	72	5.38×10^{-6}	1.56×10^{-7}	5.05×10^{-8}
	Taq	30	6.32×10^{-5}	2.27×10^{-6}	2.01×10^{-6}
	Taq	50	2.18×10^{-4}	2.18×10^{-5}	3.53×10^{-6}
	Phusion[NB]	30	2.51×10^{-5}	1.82×10^{-8}	8.32×10^{-7}
	Phusion[NB]	50	8.11×10^{-6}	ND	1.95×10^{-6}
	Phusion[NB]	72	5.05×10^{-6}	ND	9.30×10^{-8}
	Phusion	30	2.02×10^{-5}	2.32×10^{-6}	4.81×10^{-7}
	Phusion	50	3.64×10^{-5}	2.46×10^{-6}	4.84×10^{-7}
	Phusion	72	1.24×10^{-5}	3.36×10^{-7}	5.11×10^{-7}

[NB]=reaction carried out in native buffer supplied by the manufacturer.

[ND]= Error frequency below the detectable error baseline.

Table 2. Error rates (per base) for psychrophilic DNA polymerases.

Origin	Reaction	Temperature (ºC)	Substitution rate ($base^{-1}$)	Deletion rate ($base^{-1}$)	Insertion rate ($base^{-1}$)
Psychrophile	MBP-PIPB	10	9.52×10^{-6}	1.84×10^{-6}	1.26×10^{-6}
	MBP-PIPB	20	2.55×10^{-5}	2.20×10^{-5}	1.09×10^{-6}
	PIPI	10	3.87×10^{-5}	1.48×10^{-6}	1.62×10^{-7}
	PIPI	20	6.34×10^{-5}	4.15×10^{-6}	2.20×10^{-7}
	PIPI	37	1.83×10^{-4}	1.59×10^{-5}	ND
	PIPI	50	2.11×10^{-4}	8.47×10^{-6}	ND

[ND]= Error frequency below the detectable error baseline.

Table 3. Error rates (per base) for mesophilic DNA polymerases.

Origin	Reaction	Temperature (ºC)	Substitution rate ($base^{-1}$)	Deletion rate ($base^{-1}$)	Insertion rate ($base^{-1}$)
Mesophile	KleExo-	10	4.20×10^{-5}	7.65×10^{-6}	3.38×10^{-7}
	KleExo-	20	6.91×10^{-5}	8.55×10^{-6}	ND
	KleExo-	37	1.07×10^{-3}	1.75×10^{-4}	4.39×10^{-7}
	KleExo-	50	1.38×10^{-3}	1.94×10^{-4}	3.38×10^{-6}
	KleLF	10	8.53×10^{-6}	ND	2.29×10^{-8}
	KleLF	20	1.56×10^{-5}	1.12×10^{-7}	2.14×10^{-7}
	KleLF	37	2.99×10^{-5}	1.20×10^{-7}	1.43×10^{-7}
	KleLF	50	2.00×10^{-5}	ND	5.45×10^{-8}
	T4	10	3.25×10^{-5}	1.32×10^{-6}	2.37×10^{-8}
	T4	37	5.92×10^{-5}	1.17×10^{-6}	ND
	T4	50	1.45×10^{-5}	ND	ND

ND= Error frequency below the detectable error baseline.