#1114

Examining Sources of Error in PCR by Single-Molecule Sequencing

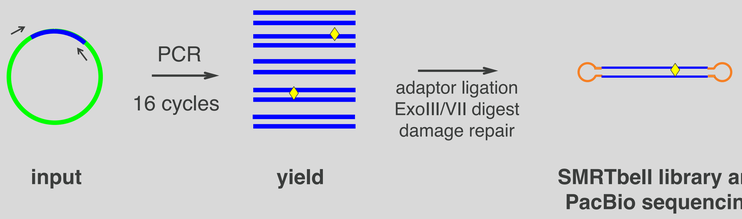
Jennifer L. Ong and Vladimir Potapov
New England Biolabs, Ipswich, MA 01938, USA

INTRODUCTION

The Polymerase Chain Reaction (PCR) is an integral part of many NGS sample preparation workflows. Mistakes made during PCR appear in sequencing data and contribute to false mutations that can ultimately confound genetic analysis. We utilized a single-molecule sequencing assay to comprehensively catalog the different types of errors introduced during PCR, including polymerase misincorporation, structure-induced template-switching, PCR-mediated recombination and DNA damage caused by thermocycling.

METHODS

EXPERIMENTAL STRATEGY



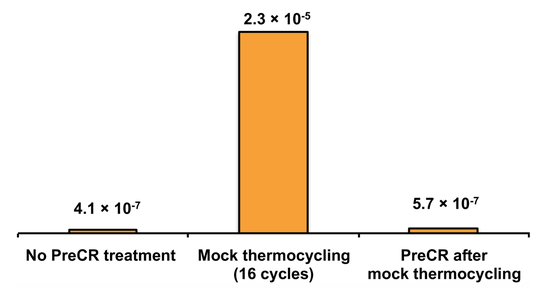
BIOINFORMATICS WORKFLOW

- For each ZMW, split all subreads into two groups (top or bottom strand) based on mapping directionality (BLASR) and build consensus reads (CCS 2)
- Map consensus reads (BLASR) and tally up all mutations by comparing consensus read to a reference
- Read filtering: number of passes >15 and Quality Value =93; length of consensus read >80% of the reference; exclude chimeric reads (BWA)
- Compute error rates and normalize to number of doubling events in PCR

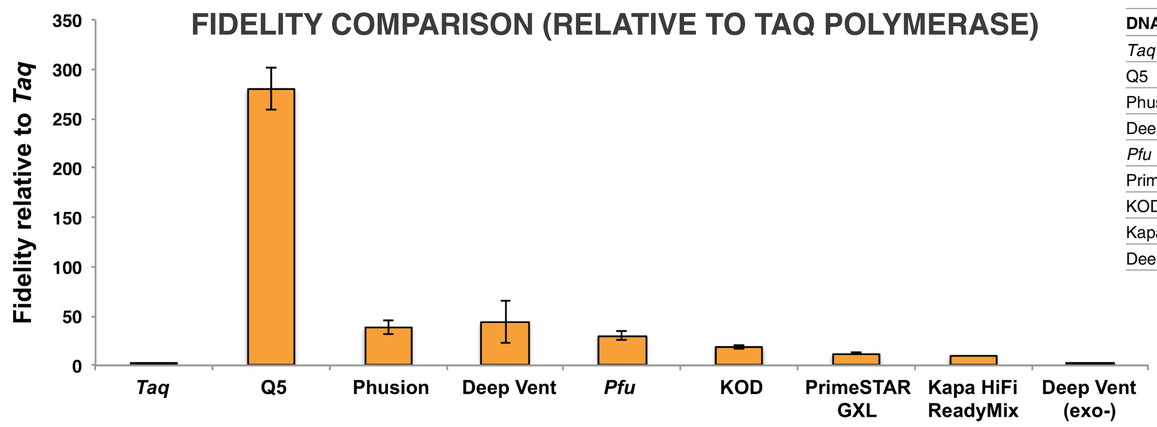
RESULTS

SEQUENCEABLE DNA DAMAGE FROM THERMOCYCLING

- Plasmid libraries were subjected to 16 mock thermocycles and sequenced before and after thermocycling,
- Per cycle, thermocycling introduced mutations at a rate of once per 714,000 bases, almost all mutations (97%) were C->T, and indicative of cytosine deamination
- PreCR treatment (DNA damage repair) after mock thermocycling reduces substitution rate to pre-thermocylcing levels



FIDELITY COMPARISON (RELATIVE TO TAQ POLYMERASE)

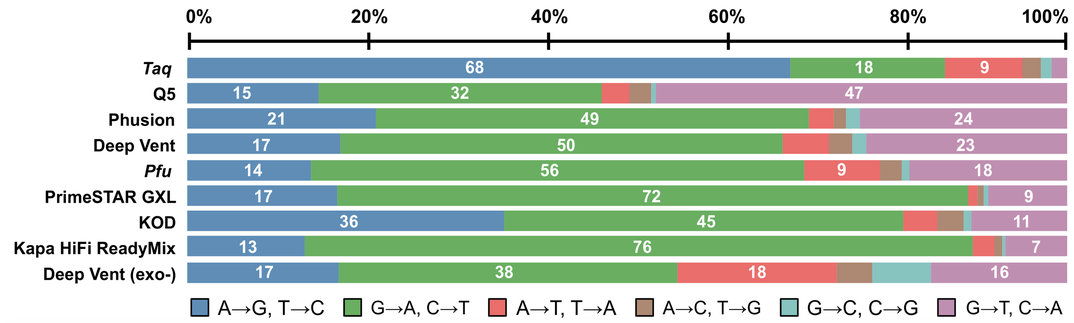


DNA Polymerase	Fidelity, rel. to *Taq* [c]
Taq	1
Q5	280
Phusion	39
Deep Vent	44
Pfu	30
PrimeSTAR GXL	18
KOD	12
Kapa HiFi HotStart ReadyMix	9.4
Deep Vent (exo-)	0.3

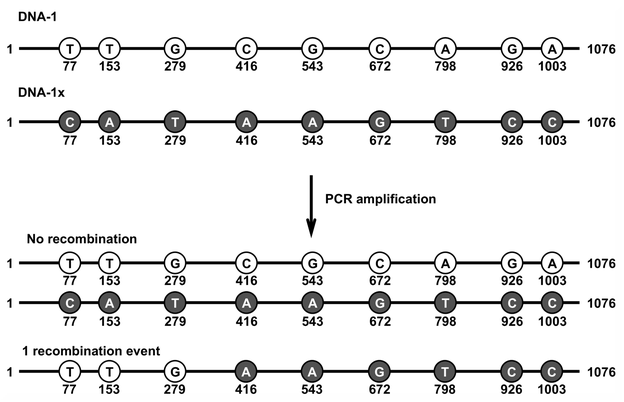
POLYMERASE BASE SUBSTITUTION ERROR RATES

DNA Polymerase	Substitution rate [a]	Accuracy [b]
Taq	1.5×10^{-4} ($\pm 0.2 \times 10^{-4}$)	6,456
Q5	5.3×10^{-7} ($\pm 0.9 \times 10^{-7}$)	1,870,763
Phusion	3.9×10^{-6} ($\pm 0.7 \times 10^{-6}$)	255,118
Deep Vent	4.0×10^{-6} ($\pm 2.0 \times 10^{-6}$)	251,129
Pfu	5.1×10^{-6} ($\pm 1.1 \times 10^{-6}$)	195,275
PrimeSTAR GXL	8.4×10^{-6} ($\pm 1.1 \times 10^{-6}$)	118,467
KOD	1.2×10^{-5} ($\pm 0.2 \times 10^{-5}$)	82,303
Kapa HiFi HotStart ReadyMix	1.6×10^{-5} ($\pm 0.3 \times 10^{-5}$)	63,323
Deep Vent (exo-)	5.0×10^{-4} ($\pm 0.1 \times 10^{-4}$)	2,020

MUTATIONAL SPECTRUM FOR VARIOUS POLYMERASES



[a] Substitution rate: substitutions/base/doubling
[b] Accuracy: number of bases over which 1 substitution error is expected



TEMPLATE-SWITCHING BY TAQ DNA POLYMERASE

A single-molecule assay to measure template-switching during amplification of a mixed population by Taq polymerase

Template pair	N_{re} [a]	N_{total} [b]	Recombination rate [c]	Strands with at least 1 recombination event
DNA-1:DNA-1x	19,943	77,725,936	9.6×10^{-5}	23%
DNA-2:DNA-2x	14,687	44,271,304	1.3×10^{-4}	28%

[a] Number of recombination events.
[b] Total number of analyzed sequenced bases.
[c] Recombination rate is per base per doubling. Recombination rate is doubled to account for "cryptic" recombination events.

REFERENCES

Potapov V, Ong JL (2017) Examining Sources of Error in PCR by Single-Molecule Sequencing. PLOS ONE 12(1): e0169774. doi: 10.1371/journal.pone.0169774

https://github.com/potapovneb/pcr-fidelity

CONCLUSIONS

We analyzed PCR products at the single-molecule level and present here a more complete picture of the types of mistakes that occur during DNA amplification. In addition to polymerase base substitution errors, other sources of error were found to be equally prevalent. For very accurate polymerases, DNA damage introduced during temperature cycling appear to be a major contributor towards mutations occurring in amplification products. PCR-mediated recombination by Taq polymerase was observed at the single-molecule level, and found to occur as frequently as polymerase base substitution errors, suggesting it may also be an underappreciated source of error for multiplex amplification reactions