Supporting Information

for

Deaminase-Driven Random Mutation Enables Efficient DNA Mutagenesis for Protein Evolution

Ying Hao,[1,#] Tong-Tong Ji,[1,#] Shu-Yi Gu,[2,#] Shan Zhang,[1] Yao-Hua Gu,[2] Xia Guo,[1] Li Zeng,[1] Fang-Yin Gang,[2] Jun Xiong,[2] Yu-Qi Feng,[1,*] Neng-Bin Xie,[2,*] Bi-Feng Yuan,[1,2,3,4,*]

[1] College of Chemistry and Molecular Sciences, Department of Radiation and Medical Oncology, Zhongnan Hospital of Wuhan University, Wuhan University, Wuhan 430071, China.

[2] Department of Occupational and Environmental Health, School of Public Health, Research Center of Public Health, Renmin Hospital of Wuhan University, Wuhan University, Wuhan 430071, China.

[3] Hubei Provincial Center for Disease Control and Prevention & NHC Specialty Laboratory of Food Safety Risk Assessment and Standard Development, Wuhan 430079, China.

[4] Hubei Key Laboratory of Biomass Resource Chemistry and Environmental Biotechnology, Wuhan University, Wuhan 430071, China.

[#] These authors contributed equally to this work.

* Corresponding author

Bi-Feng Yuan. E-mail: bfyuan@whu.edu.cn

Neng-Bin Xie. E-mail: nengbinxie@whu.edu.cn

Yu-Qi Feng. E-mail: yqfeng@whu.edu.cn

Table of Contents

Page S3-S4	Expression and purification of A3A-RL and ABE8e proteins.
Page S5	Table S1. Sequences of the synthesized MT-1 DNA and PCR primers.
Page S6	Table S2. Amino acid sequence of A3A-RL, ABE8e and EGFP protein.
Page S7	Table S3. The DNA coding sequences of A3A-RL, ABE8e, and EGFP in plasmids.
Page S8	Figure S1. Expression of A3A-RL protein.
Page S9	Figure S2. Expression of ABE8e protein.
Page S10	Figure S3. The schematic diagram of library preparation for epPCR and DRM.
Page S11	Figure S4. The schematic illustration of plasmid for the expression of EGFP protein.
Page S12	Figure S5. The sequencing depths obtained through high-throughput sequencing analysis by epPCR and DRM.
Page S13	Figure S6. The mutation distributions by epPCR.
Page S14	Figure S7. The mutation distributions by DRM.
Page S15	Figure S8. Fluorescence intensity of each EGFP mutant generated by epPCR and DRM.
Page S16	Figure S9. Comparison of the EGFP mutation frequencies produced by epPCR and DRM.
Page S17	Figure S10. Comparison of epPCR, DRM, DNA shuffling and SeSam methods.

Expression and purification of A3A-RL and ABE8e proteins

The coding sequence of A3A-RL protein was cloned into pET-41a (+) plasmid between XbaI and XhoI restriction enzyme digestion sites and an additional human rhinovirus 3C protease (HRV 3C) digestion site was inserted between the glutathione S-transferase (GST) tag and A3A-RL protein (Figure S1). The plasmids for the expression of recombinant A3A-RL protein was transformed into *Escherichia coli* (*E. coli*) BL21(DE3) pLysS strain. Transformed *E. coli* cells were cultured using LB medium (tryptone 10 g L^{-1}, yeast extract 5 g L^{-1}, and NaCl 10 g L^{-1}) supplemented with kanamycin (10 μg mL^{-1}) and chloramphenicol (10 μg mL^{-1}) at 37 °C under shaking at 180 rpm. Protein expression was started by the addition of 0.5 mM isopropyl-β-Dthiogalactoside (IPTG, Sangon) when the OD600 nm of *E. coli* cell suspension reached 0.4-0.6. The expression of recombinant proteins was induced 16 h at 16 °C under shaking at 180 rpm. The *E. coli cells* were pelleted by centrifugation at 10000 g for 5 min and cell pellets were lysed by sonication in PBS buffer with 1 mM dithiothreitol and 50 μg mL^{-1} phenylmethylsulfonyl fluoride (PMSF). The supernatant was obtained by centrifugation at 12000 g for 30 min and filtered with a 0.22 μM membrane (Jinteng, Tianjin, China). The obtained supernatant was then incubated with Glutathione Sepharose™ 4B beads (Sangon, Shanghai, China) according to the manufacturer's recommended procedure. Recombinant proteins were concentrated on the beads. After the digestion with HRV 3C protease (Sangon, Shanghai, China), A3A-RL protein was released from the beads and then further purified with a size-exclusion column (Millipore, Darmstadt, Germany) and equilibrated with a storage solution containing 50 mM Tris–HCl (pH 7.5), 50 mM NaCl, 0.01 mM EDTA, 0.5 mM dithiothreitol, and 0.01% Tween20.

As for ABE8e, the coding sequence of ABE8e protein was inserted into the pET-49b plasmid between MluI and XhoI restriction enzyme digestion sites to generate the plasmid of pET-49b (Figure S2). The pET-49b-ABE8e plasmid was transformed into *E. coli* BL21(DE3) pLysS cells and cultured at 37 °C. Protein expression was induced with 1 mM IPTG at an OD600 of 0.6-0.8, and the *E. coli* cells were then incubated at 16 °C for 16 h. Following culturing, *E. coli* cells were harvested and lysed by sonication in PBS buffer. The lysate was filtered with 0.22 μM membrane and incubated with Glutathione Sepharose™ 4B beads. Nucleic acids were removed using Cryonase Cold-active nuclease in a 1 mL solution of 80 mM

of Tris–HCl (pH 7.5) and 5 mM of $MgCl_2$, and the proteins were released using HRV 3C protease and further purified via size-exclusion column according to the manufacturer's protocol. The purified ABE8e protein was stored in a solution containing 200 mM of Tris-HCl (pH 7.5), 400 mM of KCl, 10% glycerol (v/v), and 2 mM of dithiothreitol (DTT) at −80 °C.

Table S1. Sequences of the synthesized MT-1 DNA and PCR primers.

Oligonucleotides	Sequence (5′ to 3′)
MT-1	GGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAG ACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAA GCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGC TGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGC ACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAA ATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGT TGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGC TGGCGAAAGGGGGATGTGCTGCA
MT-F	GGTGATGACGGTGAAAACCTCT
MT-R	TGCAGCACATCCCCCTTT
EGFP-F	GCCAGGATCCATATTACTTGTACAGCTCGTCCATGCCG
EGFP-R	AAAAACTAGTATGGTGAGCAAGGGCGAGGAGC
pre-T5	CACTCTTTCCCTACACGACGCTCTTCCGATCTGAGAGAGAGGA GAATATAAATGTTATTGAT
pre-T7	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTCTCCTCCTA ATATAAATATCATCA
P5-index	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCT TTCCCTACACGACGCTCTTCCGATCT
P7-index	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAG TTCAGACGTGTGCTCTTCCGATCT

Table S2. Amino acid sequence of A3A-RL, ABE8e and EGFP proteins. The red highlights indicate amino acid mutations in the proteins.

Protein name	Sequence (N-terminal to C-terminal)
A3A-RL	MEASPASGPRHLMDPHIFTSNFNNEPWVRGRHKTYLCYEVERLDN GTSVKMDQHRGFLHNQAKNLLCGFYGRHAELRFLDLVPSLQLDPA QIYRVTWFISWSPCFSWGCAGEVRAFLQENTHVRLRIFAARIYDYDP LYKEALQMLRDAGAQVSIMTYDEFKHCWDTFVDHQGCPFQPWDG LDEHSQALSGRLRAILQNQGN
ABE8e	MSEVEFSHEYWMRHALTLAKRARDEREVPVGAVLVLNNRVIGEGW NRAIGLHDPTAHAEIMALRQGGLVMQNYRLIDATLYVTFEPCVMCA GAMIHSRIGRVVFGVRNSKRGAAGSLMNVLNYPGMNHRVEITEGIL ADECAALLCDFYRMPRQVFNAQKKAQSSINSGGSSGGSSGSETPGT SESATPESSGGSSGGS
EGFP	MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLK FICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEG YVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG HKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQ QNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGI TLGMDELYK*

Table S3. The DNA coding sequences of A3A-RL, ABE8e, and EGFP in plasmids. The red highlights indicate base mutations in DNA.

Plasmids	Sequences
A3A-RL	ATGGAAGCCAGCCCAGCATCCGGGCCCAGACACTTGATGGATCCACACATCTTCAC TTCCAACTTTAACAATGAACCTTGGGTCCGCGGACGTCATAAGACCTACCTGTGCTA CGAAGTGGAGCGCCTGGACAATGGCACCTCGGTCAAGATGGACCAGCACCGTGGC TTTCTCCACAACCAGGCTAAGAATCTTCTCTGTGGCTTTTACGGCCGCCATGCGGAG CTGCGCTTCTTGGACCTGGTTCCTTCTTTGCAGTTGGACCCGGCCCAGATCTACAGG GTCACTTGGTTCATCTCCTGGAGCCCCTGCTTCTCCTGGGGCTGTGCCGGGGAAGT GCGTGCGTTCCTTCAGGAGAACACACACGTGAGACTGCGTATCTTCGCTGCCCGCA TCTATGATTACGACCCCCTATATAAGGAGGCACTGCAAATGCTGCGGGATGCTGGGG CCCAAGTCTCCATCATGACCTACGATGAATTTAAGCACTGCTGGGACACCTTTGTGG ACCACCAGGGATGTCCCTTCCAGCCCTGGGATGGACTAGATGAGCACAGCCAAGCC CTGAGTGGGAGGCTGCGGGCCATTCTCCAGAATCAGGGAAAC
ABE8e	ATGTCCGAGGTGGAATTCAGTCATGAATATTGGATGAGACACGCTCTCACACTAGCC AAAAGAGCCCGTGACGAAAGGGAAGTTCCCGTTGGCGCCGTGCTGGTCTTGAACA ACCGAGTGATTGGCGAAGGTTGGAATAGAGCTATCGGCCTGCACGATCCTACAGCT CATGCTGAAATTATGGCTCTGAGACAGGGCGGGCTCGTGATGCAGAATTACCGACT GATCGATGCTACCCTGTACGTTACTTTTGAACCTTGTGTAATGTGTGCTGGCGCAAT GATTCACAGCAGAATTGGAAGAGTGGTGTTTGGAGTGAGAAACTCCAAGAGGGGG GCCGCTGGAAGCCTGATGAATGTGCTGAATTATCCCGGAATGAACCACAGAGTGGA GATAACCGAAGGTATTCTGGCGGATGAGTGCGCCGCTCTATTATGTGATTTCTATCGG ATGCCAAGACAGGTTTTTAACGCCCAGAAAAAGGCTCAGAGTAGCATCAACAGCG GCGGGTCAAGCGGCGGCAGCAGTGGCAGCGAGACACCTGGTACATCAGAGTCTGC CACTCCTGAATCCAGTGGCGGCAGCTCCGGCGGCAGC
EGFP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCG ATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCC GTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCG CTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCT ACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGC CGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATC GACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACA GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTC AAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGC AGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAG CACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTG CTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTA A

Figure S1. Expression of A3A-RL protein. (A) The schematic illustration of plasmid for the expression of A3A-RL protein. (B) Analysis of the purified A3A-RL protein by sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

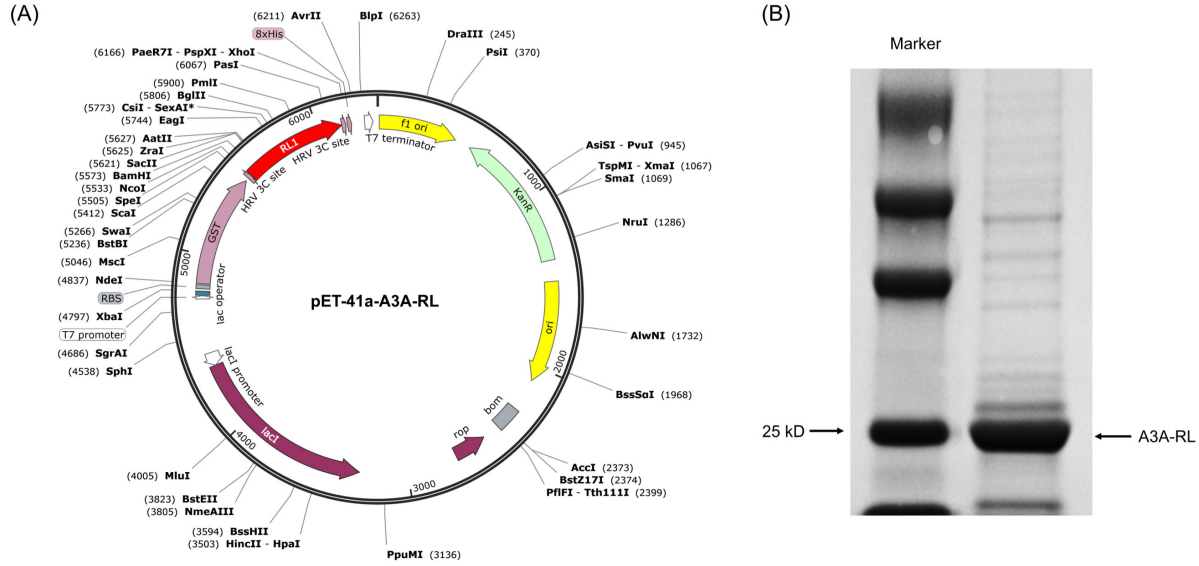

Figure S2. Expression of ABE8e protein. (A) The schematic illustration of plasmid for the expression of ABE8e protein. (B) Analysis of the purified ABE8e protein by sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

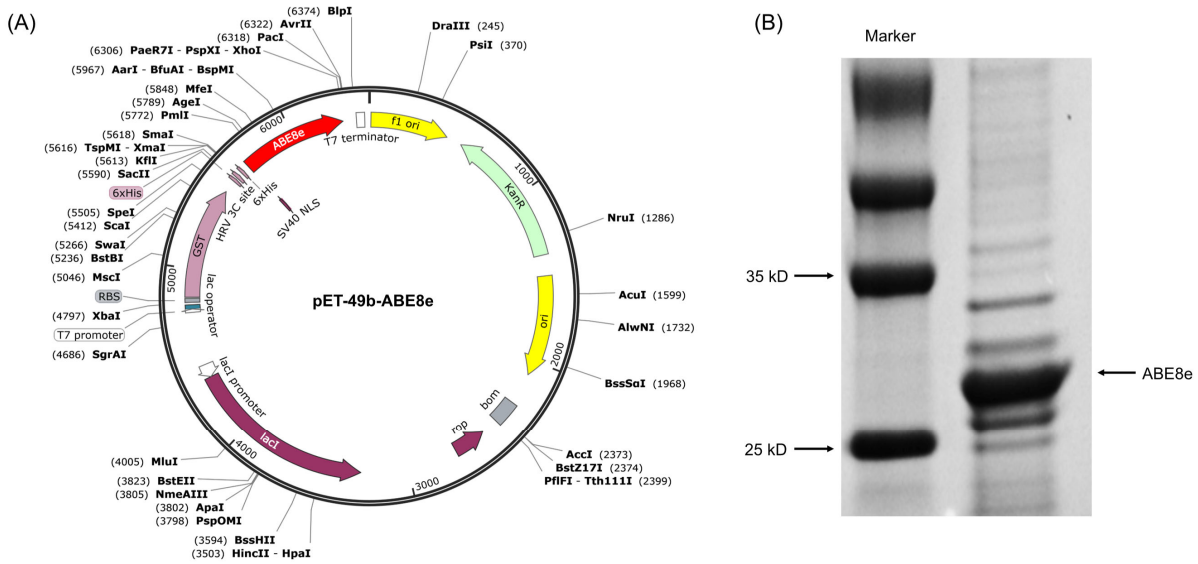

Figure S3. The schematic diagram of library preparation for epPCR and DRM.

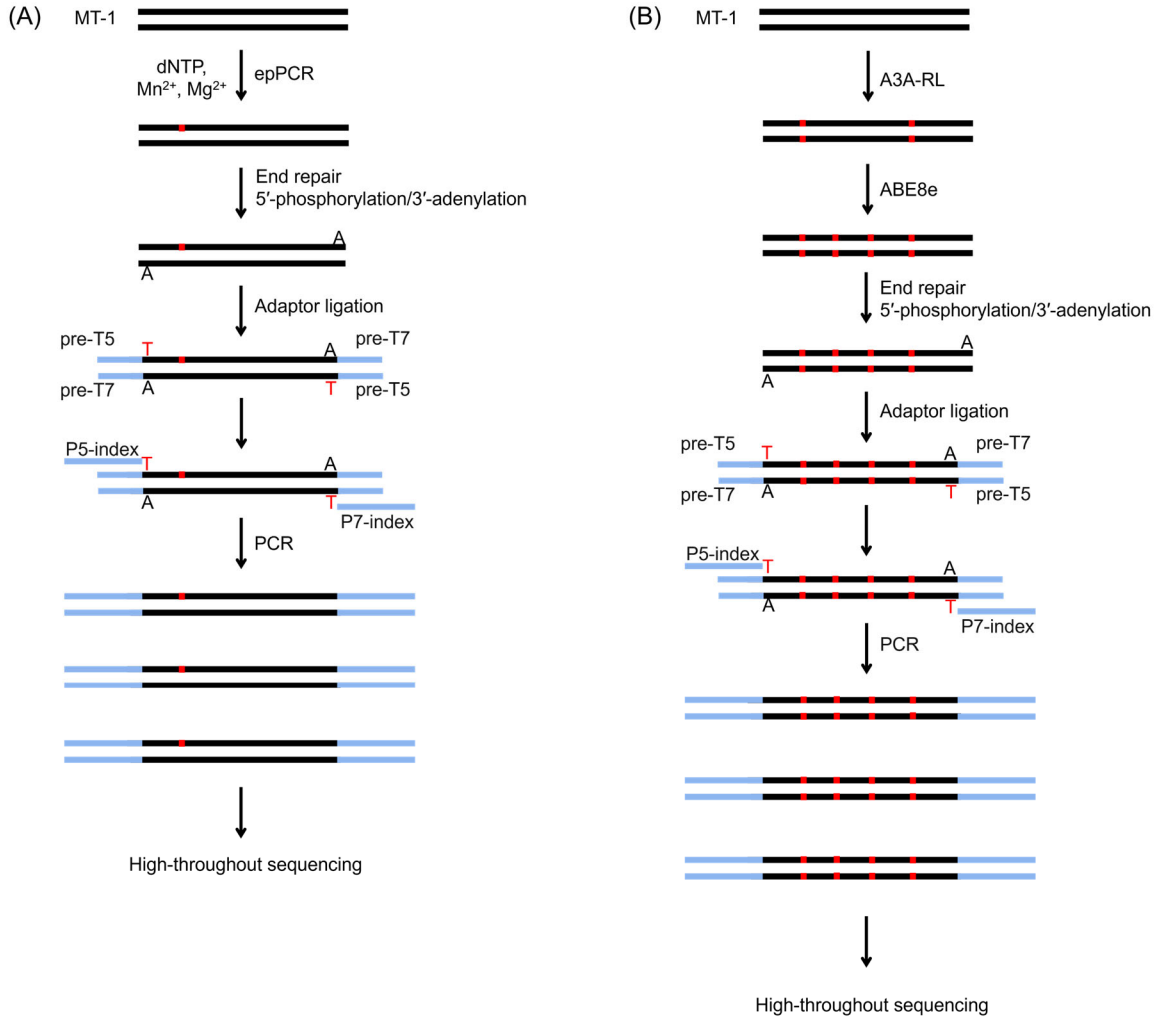

Figure S4. The schematic illustration of plasmid for the expression of EGFP protein.

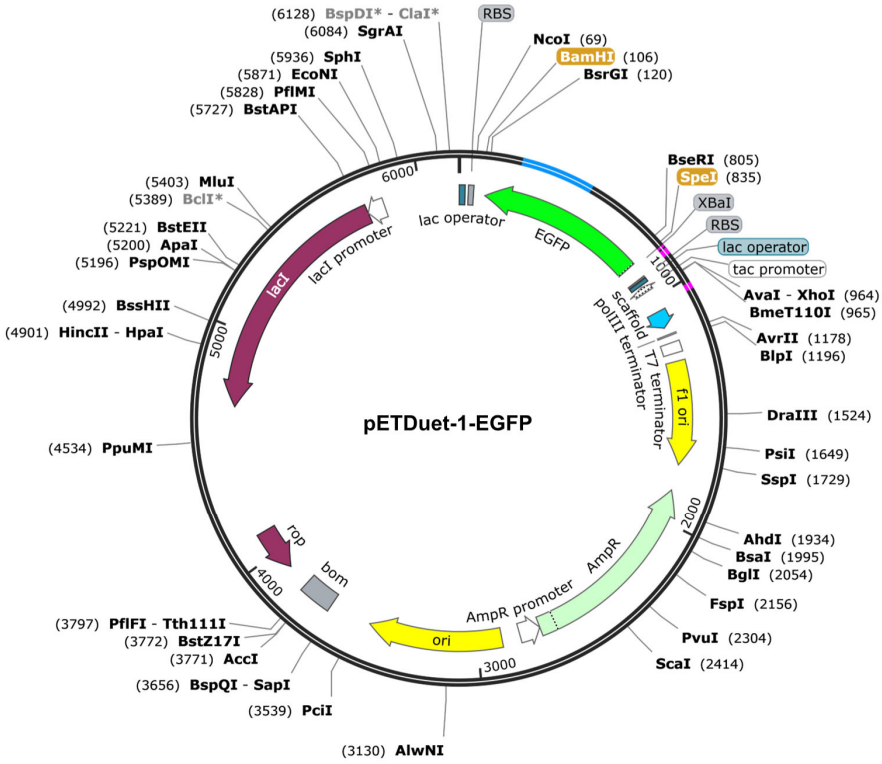

Figure S5. The sequencing depths obtained through high-throughput sequencing analysis by epPCR and DRM. (A) The sequencing depths by epPCR method. (B) The sequencing depths by DRM method. MT-1 DNA was used in the high-throughput sequencing analysis.

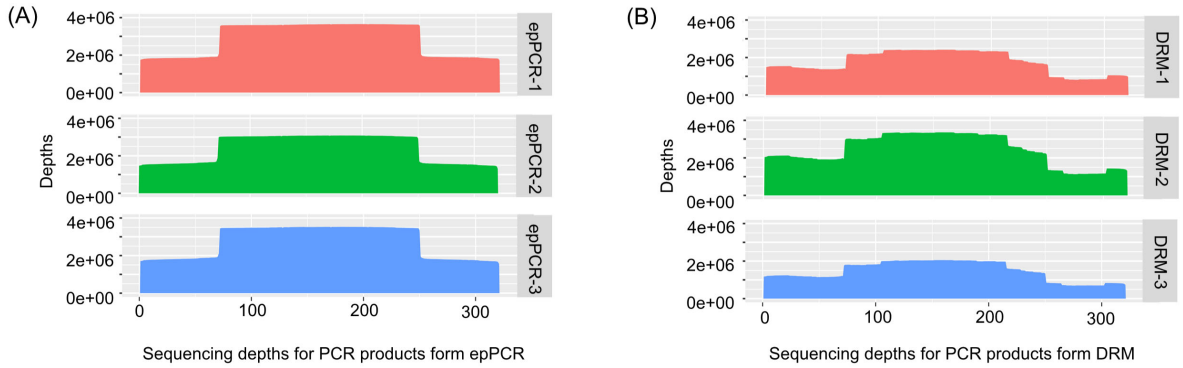

Figure S6. The mutation distributions by epPCR. The epPCR method was found to primarily introduce mutants carrying a relatively small number of base mutations, with the majority of mutants harboring between 1 and 9 base mutations. Three parallel experiments were carried out.

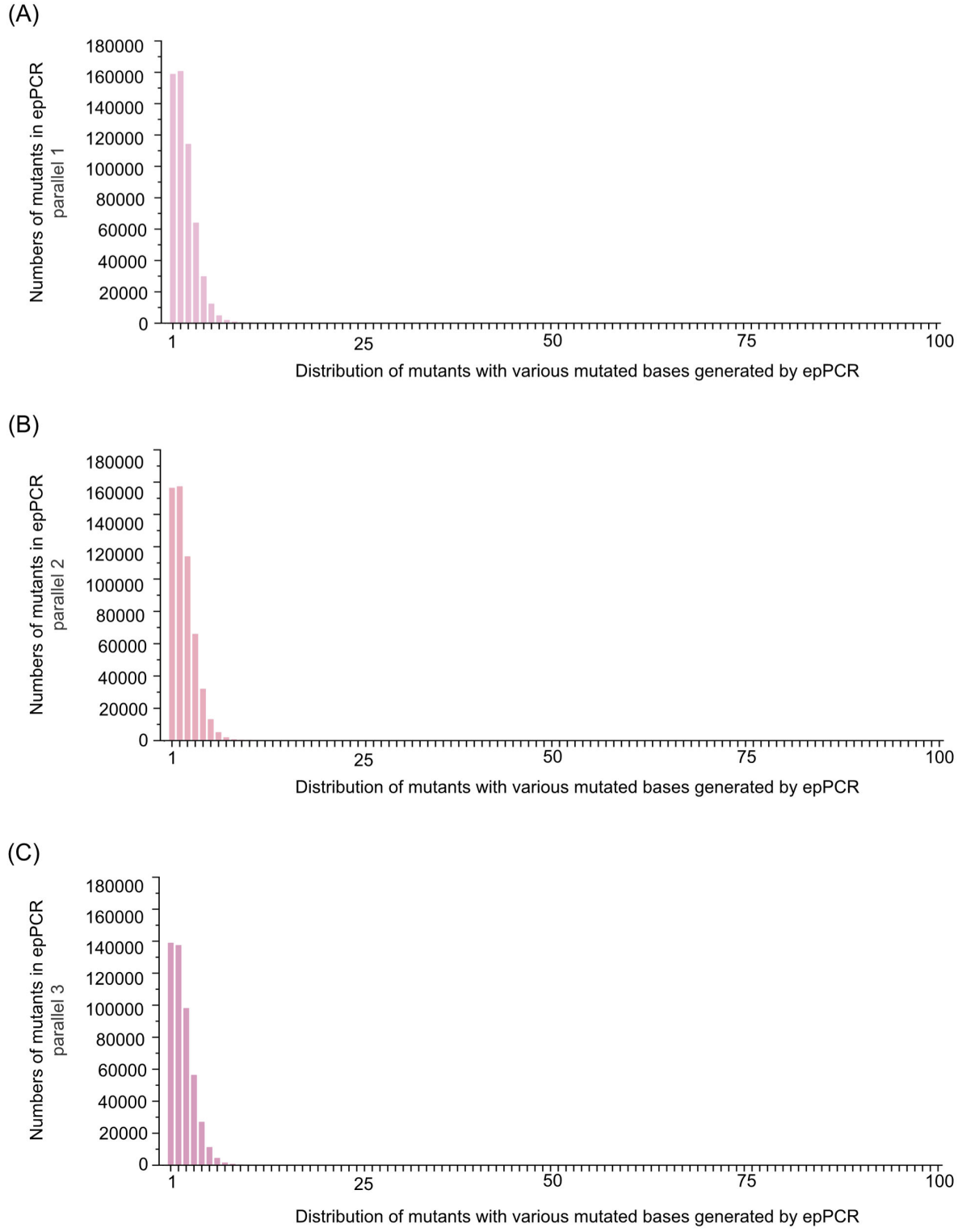

Figure S7. The mutation distributions by DRM. The DRM method introduced a much broader range of base mutations, with mutants carrying from 1 to 75 base mutations. Three parallel experiments were carried out.

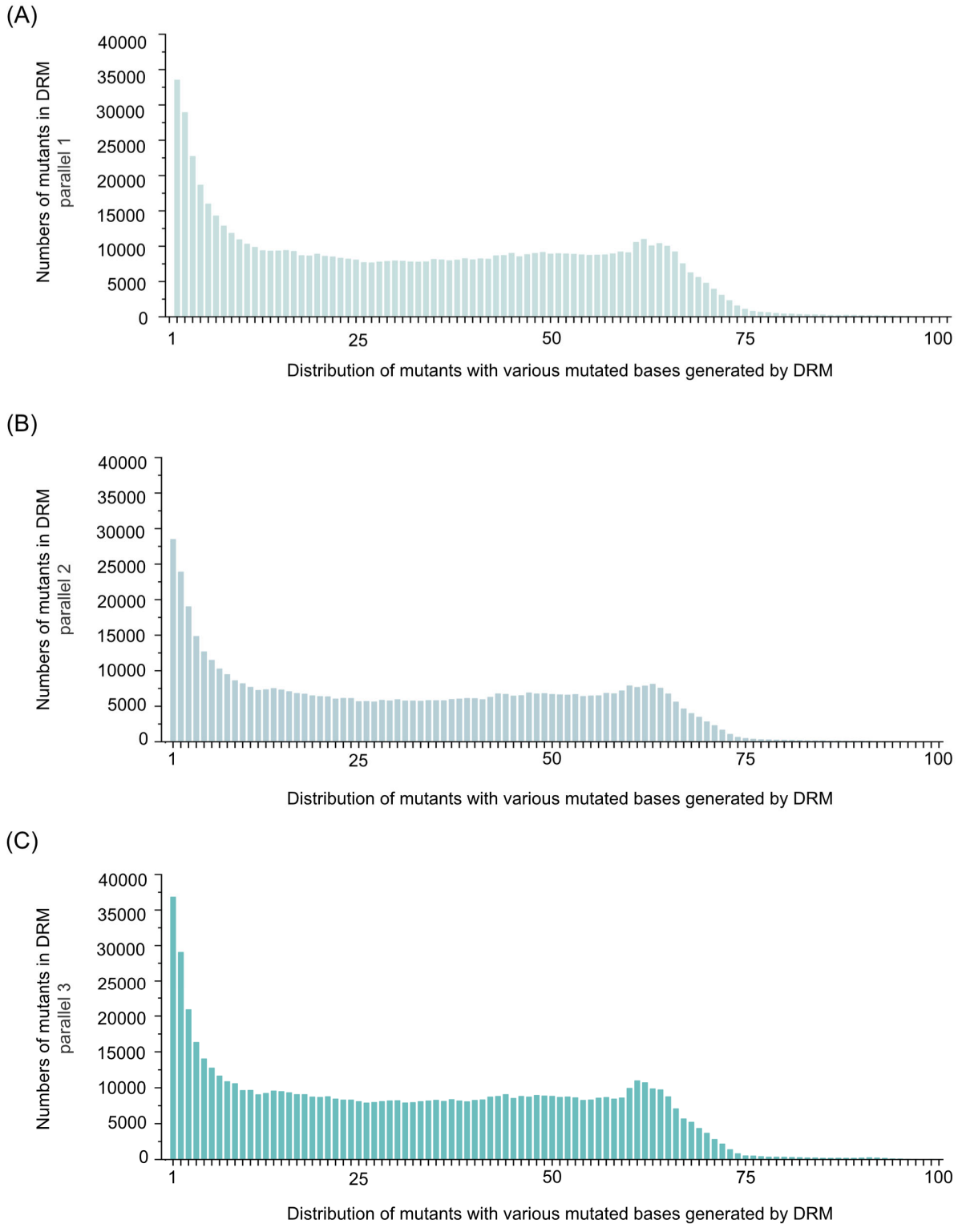

Figure S8. Fluorescence intensity of each EGFP mutant generated by epPCR and DRM. (A) Fluorescence intensity of each EGFP mutant generated by epPCR. (B) Fluorescence intensity of each EGFP mutant generated by DRM.

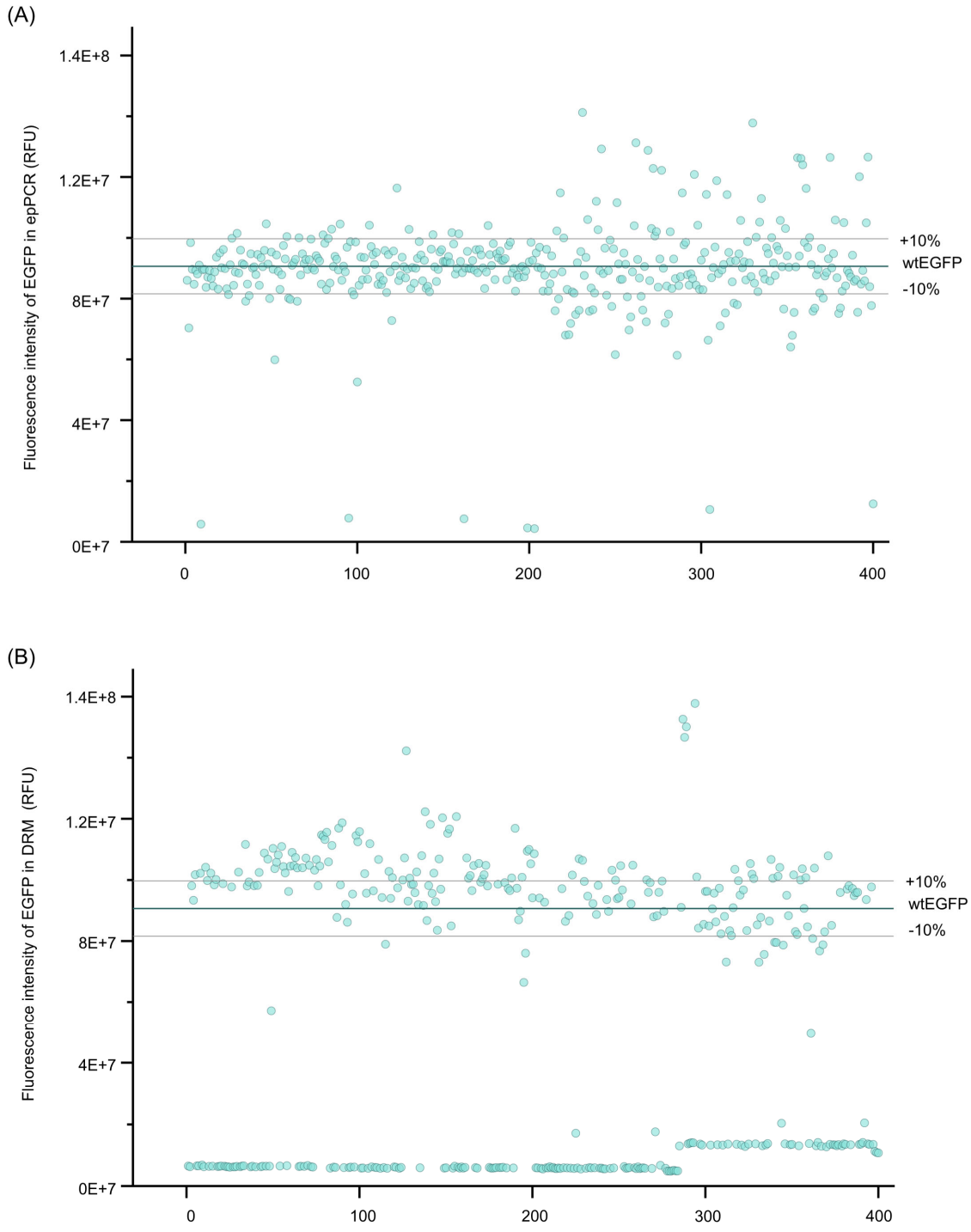

Figure S9. Comparison of the EGFP mutation frequencies produced by epPCR and DRM. (A) The EGFP mutation frequencies generated by epPCR and DRM. (B) The EGFP mutation frequencies in different types of mutants generated by epPCR. (C) The EGFP mutation frequencies in different types of mutants generated by DRM.

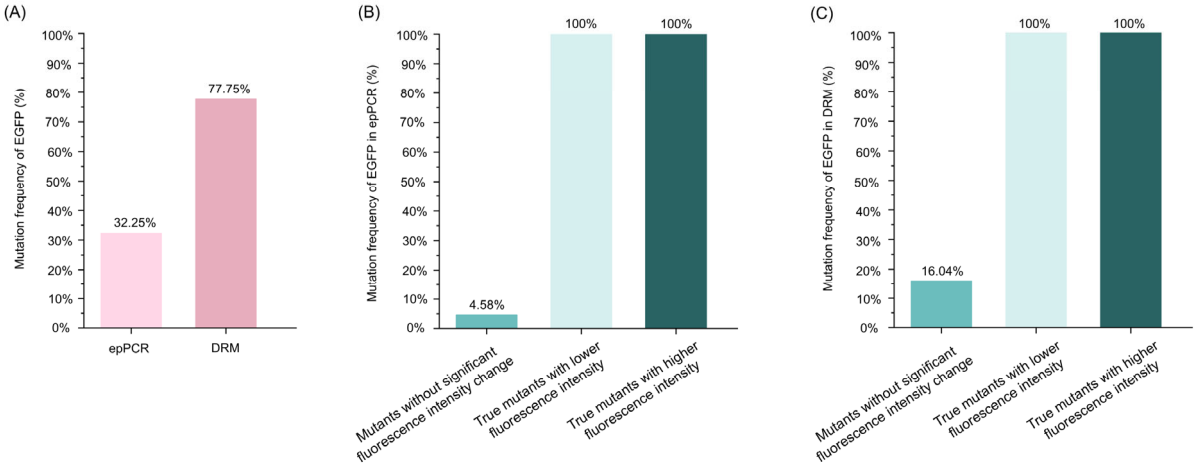

Figure S10. Comparison of epPCR, DRM, DNA shuffling and SeSam methods. (A) epPCR method. (B) DRM method. (C) DNA shuffling method. (D) SeSam method.