Premature termination codons signal nonsense-mediated gene repair in *E. coli*

Xiaolong Wang[1], Xuxiang Wang[1], Haibo Peng[1], Chunyan Li[1], Yalei Wang[1], Gang Chen[1], Jianye Zhang[1]*

College of Life Sciences, Ocean University of China, Qingdao, 266003, P. R. China

Abstract

Frameshift mutations yield truncated proteins, leading to loss-of-function, genetic disorders, or even death. Reverse mutations, which restore the wild-type phenotype of a mutant, have been assumed to be far rarer than a forward mutation. However, in this study, screening tests in *E. coli* showed that the revertants of a frameshift mutation are detected more frequently than expected. Sanger and NGS sequencing of the plasmids revealed that the reversion of the frameshift mutation is not likely caused by random mutagenesis but by an active and targeted gene repair. Molecular studies suggest that premature termination codons (PTCs) in the nonsense mRNAs signal the repair of the frameshift mutation. Genome resequencing indicates that the proofreading/mismatch repair system of the revertant is not defective but active. Transcriptome profiling identified upregulated genes/pathways involved in the frameshift repair, including DNA replication, RNA processing, RNA editing, mismatch repair, and homologous recombination. Based on these data and a review of previous studies, we hypothesized a model for frameshift repair, referred to as *nonsense-mediated gene repair* (*NMGR*). In essence, nonsense mRNAs are recognized, edited by RNA editing, and then used to direct the repair of their defective coding DNA sequence through mismatch repair and homologous recombination. Moreover, NMGR may also serve as a driving force for molecular evolution, and a new source of genetic diversity leads to the widespread existence of frameshift homologs.

[*] To whom correspondence should be addressed: Xiaolong Wang, Department of Biotechnology, Ocean University of China, No. 5 Yushan Road, Qingdao, 266003, Shandong, P. R. China, Tel: 0086-139-6969-3150, E-mail: Xiaolong@ouc.edu.cn. .

1. Introduction

DNA replication and DNA repair happen daily in cell division and multiplication. DNA damage, point mutations, insertions, and deletions (indels) can be induced by chemical and physical mutagens, such as radiation, pollutants, or toxins. Besides, because of the imperfect nature of DNA replication, mutations and indels also occur spontaneously as replication errors or slipped-strand mispairing. In a protein-coding gene, if the size of an indel is not a multiple-of-three, it causes a frameshift mutation, which results in a substantial change of its encoded amino acid sequence; also, many nonsense codons are often produced downstream the frameshift mutation, and thus, it has been generally believed that a frameshifted gene yields truncated and dysfunctional proteins [1], and therefore lead to genetic disorders or even death.

The molecular mechanisms/pathways involved in the repair of DNA damages and point mutations have been intensively studied [2], including mismatch repair, base- or nucleotide-excision, homologous recombination, and non-homologous end-joining. The molecular mechanism for the repair of frameshift mutations (hereafter *frameshift repair*), however, has not been reported. The maintenance of the reading frames of coding genes is essential for cells. A reverse mutation phenomenon was discovered as early in the 1960s [3]. Both forward and backward mutations have been explained by random mutagenesis. In principle, the rate of a reverse mutation should be much lower than that of a corresponding forward mutation. However, here we found that the reversion of a frameshift mutation occurs more frequently than that is expected, and it seems not to be random mutagenesis but involves an active and targeted gene repair.

2. Materials and Methods

2.1 Frameshift mutagenesis and back mutation screening

A plasmid containing a frameshifted β-lactamase gene (*bla*), pBR322-(bla-), was constructed by site-directed mutagenesis (Fig 1). In plasmid pBR322, a G:C base pair, located in the upstream (+136) of the wild-type *bla* gene (*bla*+), was deleted by using overlapping extension polymerase chain reaction (OE-PCR). Competent cells of *E. coli* strain DH5α were transformed with pBR322-(bla-), propagated in tetracycline

broth (TCB), and then diluted cultures were plated on tetracycline plates (TCPs) and ampicillin plates (ACPs) to screen for revertants. The recovery rates were calculated by the standard method [4]. The revertants were propagated in ampicillin broth (ACB) at 37°C with 200 rpm shaking and with 1 mL of overnight seed culture in 10 mL ACB. The growth rates of the revertants were evaluated by the doubling time during their exponential growth phase. Their plasmids DNA were extracted, and their *bla* genes were Sanger sequenced.

2.2 Construction and expression of a PTC-free frameshifted bla

A *PTC-free frameshifted bla*, denoted as *bla#*, was derived from *bla-* by replacing each nonsense codon with a sense codon according to the *readthrough rules* (Table 1), and by adding a stop codon (TAA) in the 3'-end. The *bla#* gene was chemically synthesized by *Sangon Biotech, Co. Ltd* (*Shanghai*), inserted into an expression vector, pET28a, and transformed into competent cells of *E. coli* strain BL21. Plasmid pET28a contains a kanamycin resistance gene. The transformed bacteria were therefore plated on a kanamycin plate (KCP), the transformants were propagated in kanamycin broth (KCB), plated on ACPs to screen for revertants. Then their *bla* genes were Sanger sequenced. The expression of the frameshifted BLA was induced by 0.1 mM IPTG. Their total protein samples were extracted, frameshifted BLA was purified by nickel column chromatography and analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The activity of BLA was measured by iodometry.

2.3 Genome resequencing and structure/variation analysis of the E. coli strains

Genomic DNA was extracted from wild-type (*bla+*) and revertant (*bla**) samples. Library preparation and genome sequencing were conducted by a commercial service provided by *Novogene Co. Ltd*. Paired-end reads were obtained on an Illumina HiSeq 250PE platform. Clean reads were obtained by removing adapters and low-quality sequences. For each sample, clean reads were mapped onto the reference genome of *E. coli* K12 MG1655 (NC_000913.3) and the plasmid pBR322, to identify single nucleotide polymorphisms (SNPs), indels and structure variations (SVs). Samtools (v0.1.18) was used to sort and reorder the bam alignments [5]. Genome analysis tool kit (GATK) was used for SNP calling [6]. Circos was used to display the coverage of

reads and the distribution of SNPs and indels on the ring diagram [7]. BreakDancer was used to detect structure variations [8, 9].

2.4 Transcriptome analysis of the E. coli strains

Total RNAs were extracted from a wild type (*bla*+), a frameshift (*bla*-), and two revertants (*bla**) samples. Library preparation and RNA sequencing were conducted by a commercial service (*Novogene Co. Ltd*). For each sample, the high-quality clean reads were mapped onto the annotated reference genome of *E. coli* K12 MG1655 (NC_000913.3) and plasmid pBR322. The expression levels of each gene were calculated and compared to each other to identify differentially expressed genes (DEGs), enriched gene ontology (GO) terms, and Kyoto encyclopedia of genes and genomes (KEGG) pathways.

(1) *Quantification of gene expression level*

Bowtie (v2-2.2.3) was used for index building and aligning the clean reads to the reference genome [10]. HTSeq (v0.6.1) was used to count the numbers of clean reads mapped to each gene [11]. And then, the expected number of Fragments PerKilobase of transcript sequence per million base pairs sequenced (FPKM) of each gene was calculated based on their lengths and read counts.

(2) *Differential expression analysis*

Before differential expression gene analysis, the read counts were adjusted by the edgeR program for each strain [12]. Differential expression of the read counts of two conditions was performed using the DEGSeq R package (1.20.0) [13]. The P values were adjusted by the Benjamini & Hochberg method. The threshold for significantly differential expression was set as: corrected P-value (q-value) < 0.005 and log2 (fold change) $\geqslant 1$.

(3) *GO and KEGG enrichment analysis of differentially expressed genes*

Gene Ontology (GO) enrichment analysis was implemented by the GOseq R package [14], in which the potential gene length bias was corrected. GO terms with corrected P-value less than 0.05 were considered significantly enriched by DEGs. KOBAS 2.0 [15] was used to test the statistical enrichment of DEGs in KEGG pathways (http://www.genome.jp/kegg/).

3. Results and Analysis

3.1 The growth of the frameshift mutant and the revertants

The plasmid pBR322 contains two wild-type resistance genes, a β-lactamase gene (*bla*+) and a tetracycline resistance gene (*tet*+). Using OE-PCR, one single G:C base pair in the upstream of *bla*+ was deleted (Fig 1). In the resulting frameshift *bla* (*bla*-), seventeen nonsense codons appeared. Because the active sites of BLA, including the substrate-binding site, all located downstream the deletion, this frameshift mutation is a loss-of-function. The plasmid containing wild-type or frameshifted *bla* gene was transformed into *E. coli* (strain DH5*α* and BL21), the transformants were propagated in tetracycline broth (TCB), dilution series were plated on tetracycline plate (TCPs) to estimate the total number of viable bacteria and plated on ampicillin plates (ACPs) to test their ampicillin resistance. The wild-type (*bla*+) bacterial grew well on both ACP and TCP (Fig 2A, left), and it was expected that the mutant (*bla*-) would not grow on ACP. However, a few ampicillin-resistant colonies (*bla**) were observed (Fig 2A, middle). A blank control ruled out the possibility of carryover- or cross-contamination (Fig 2A, right). So, these ampicillin-resistant colonies are considered as revertants whose *bla*- genes were repaired by reverse mutations.

The tests were repeated by different operators to measure the reproducibility. The revertants were reproduced in 23 out of 30 tests. Although most of the revertants can grow in ACB, the growth rates of the revertants are often very much slower when compared to the wild-type strain. By culturing at 37°C with 200 rpm shaking, it takes 12 to 24 hrs for the wild type but up to 36 to 48 hrs for the revertants to reach the late log phase, and many of the revertants failed to grow when they were subcultured in ACB, even without adding ampicillin.

Hitherto, it seems that there is nothing unusual because the reverse mutations are a fairly common phenomenon, arising from simple random mutations (Fig 3). In DNA replication, natural or induced mutations occur randomly, a forward mutation causes a defective mutant or a loss-of-function, and a reverse mutation restores the wild-type phenotype. Under this view, these revertants survived just because they were ‘lucky’:

their defective *bla* gene restored coincidently by a reverse mutation, while most of the other bacteria died without a reverse mutation.

3.2 Frameshift reversion is not random mutagenesis but targeted gene repair

The classical random mutagenesis model for reverse mutation sounds faultless in explaining the reversions of point mutations. However, in a thought experiment, we noticed that the model is insufficient to explain the frameshift repair observed herein: (1) it assumes that a bacterium survives if and only if the *bla* gene is restored through random mutagenesis. However, a reverse mutation must occur in a living cell, in the process of DNA replication or repair. If the mutant itself cannot live in ampicillin, the reversion must occur not after, but before they were cultured in ampicillin; (2) a frameshift mutation can only be repaired by an appropriate insertion or deletion restoring the sequence/reading frame of the defective gene. However, even if relatively small, a bacterial genome consists of millions of base pairs and thousands of genes. If the cells do not 'know' which gene is defective, there is no or little chance to repair it by random mutagenesis; (3) As determined by whole-genome sequencing [16], the baseline mutation rate of the wild-type *E. coli* is $\sim 10^{-3}$ per genome (or $\sim 10^{-9}$ per nucleotide) per generation. In *E. coli*, the rate of point mutation is meager, and the frameshift mutation rate is even ten-times lower than that of point mutations [16]. It is assumed that a reverse mutation occurs at a lower rate than a forward mutation. For a specific frameshift mutation, the reverse mutation rate should be extremely low ($<10^{-10}$) if it depends on random mutagenesis, as insertions/deletions are far rarer than base substitutions. Therefore, a frameshift revertant should be very difficult to be detected. However, through repeated screening tests, we found that the revertants of *bla*- can be detected frequently, and it seems that the reversion of a frameshift occurs much more frequently than expected. The average recovery rate of *bla*- measured in DH5α is 2.61×10^{-8} per cell division and reach up to 1.19×10^{-7} per cell division in BL21, both significantly higher than our expectation.

If a reverse mutation is caused by random mutagenesis, it is bound to be inhibited by the proofreading and mismatch repair system (PR-MRS). However, the reverse rate is not lower but 4-5 times higher in BL21 (which has a functional PR-MRS) than

that in DH5*α,* which has a defective PR-MRS, strongly suggesting that the PR-MRS system does not inhibit but promote the reversion. Therefore, an efficient mechanism specific to the repair of frameshift mutation may have been developed in the history of life. However, how a frameshifted mutation is being identified and repaired at the molecular level? It remains unknown, but the PR-MRS system may play a vital role in the process.

3.3 Sanger sequencing of the bla genes of the revertants

To investigate how the frameshift gene is repaired, the revertants were picked and subcultured in ACB. The revertants usually grow very slowly, and many of their subcultures failed to grow, but the surviving subcultures grow faster and faster. Their plasmid DNA was extracted, and their *bla* were Sanger sequenced. Unexpectedly, in these initial revertants and subcultures, the *bla* genes are often neither a wild type nor a revertant but seem to be still frameshifted (Fig 2B, top). It is puzzling how these *bla* frameshifts could be functional.

The revertants were subcultured, propagated, and their *bla* were re-sequenced for over 300 replications. In the subcultures, the sequencing results of the *bla* genes also often seem to be frameshifted. However, their Sanger diagrams often contain two sets of superimposed peaks (Fig 2B, middle), indicating that they consist of two types of *bla* sequences: one is the original frameshifted *bla*, and the other is the repaired one. By visual inspection, the repaired sequences of the revertants (*bla**) were read from the secondary peaks of the Sanger diagrams. Sometimes a sequencing diagram may contain three or more sets of peaks, but the sequences are nonreadable. In the final fast-growing subcultures, most of their *bla* copies were repaired, and their Sanger diagrams contain few superimposed peaks (Fig 2B, bottom).

As shown in Fig 2C, by aligning *bla** sequences against *bla*+ and *bla*- sequences, it was found that they contain a variety of indels and substitutions, especially G:C←→A:T transitions (Fig 2C, circles). Previously, we developed a novel tool for multiple sequence alignment, Codon-and-Aa Unified Alignment (CAUSA) [17], which can align and visualize coding sequences and their encoded protein sequences in the same user interface. As shown in Fig 2D, CAUSA demonstrates that the protein sequences

translated from the *bla** genes are all different from each other but all highly similar to that of the wild-type BLA, even when they are interrupted by some stop signals. In other words, the frameshifted genes are restored not entirely and immediately, but partially and progressively, resulting in a variety of *frameshift homologs*.

3.4 The genome of the revertant is not highly variable but very stable

As is well known, an extraordinarily high rate of spontaneous mutagenesis occurs over its whole genome in a mutator, such as *E. coli* strain mutD5, due to its deficient proofreading and mismatch repair during DNA replication [18]. The *E. coli* strain, either DH5α or BL21, is not a mutator. However, the frameshift/revertants obtained may change into mutators if their proofreading and mismatch repair system is damaged during culturing and subculturing so that high-rate of mutagenesis can occur in their genomes, not only in *bla* but also in other genes.

To clarify whether a revertant may change into a mutator, the genome sequences of a frameshift (*bla*-) and a revertant (*bla**) were sequenced by high-throughput NGS sequencing. The clean reads were mapped onto the reference genome sequence of the *E. coli* strain K12 and compared with each other (Fig 4A). As shown in Fig 4B and Table 2, the SNP/indel levels of the revertant and frameshift tested are both very low, suggesting that their proofreading/mismatch repair system is not defective. Therefore, they are not mutators but adopt a stringent proofreading and mismatch repair during the replication/repair of their DNA. Besides, if the variations observed in *bla** are caused by random mutagenesis, the number of variations observed in *bla* should be proportional to the number of variations observed in the whole genome. Therefore, the number of variations observed in the revertant genome should be higher than that in the frameshift genome. However, as shown in Fig 4B and Table 2, the numbers of SNPs/indels of the revertant are not higher but even lower than those of the frameshift. In other words, the number of *bla* variations is independent of that of the variations observed in the whole genome, suggesting that the high-level of variations of *bla** are not caused by nonspecific, whole-genome random mutagenesis, but more likely to be an active gene repair specifically targeting the frameshifted *bla*.

3.5 The other gene in the same plasmid does not change

To clarify whether tetracycline-resistant (*tet*), the other gene in the same plasmid, changes or not, it was also Sanger sequenced. However, no overlapping peaks were observed in their sequencing diagrams, suggesting that overlapping peaks observed in *bla** are not the result of nonspecific mutagenesis but more likely to be an active gene repair targeting specifically to the frameshifted *bla* gene.

To validate that the *tet* gene does not change as fast as the *bla* gene, we further mapped the NGS reads of the genome or transcriptome onto the plasmid. As shown in Fig 4D, by visual inspection, very few variations can be observed in the *bla* genes of the wild type. However, a lot of insertion, deletion, and substitutions can be observed in the *bla* genes, as well as in their transcripts, in both the frameshift and the revertant; as shown in Fig 4E, however, almost no or only a few variations were observed in the *tet* genes or their transcripts, neither in the frameshift nor in the revertant.

Such kind of random variation is hard to be distinguished from sequencing errors, filtering of the raw SNPs may discard the real changes. Therefore, we used unfiltered raw SNPs, but discarded those ambiguous ones, to count and evaluate the numbers and densities of SNPs of the two genes in the plasmid pbr322. The SNPs were called by using bcftools and summarized in Table 3. In DNA-level, the density of raw SNPs is very high in both genes. The density of *bla* variation is slightly higher than that of *tet* in the frameshift; however, is significantly higher than that of *tet* in the revertant; in RNA level, the number of raw SNPs is smaller than those in DNA, but the difference between *tet* and *bla* is more obvious: the density of variation is lower in *bla* than in *tet* in the wild type but significantly higher in *bla* than in *tet* in the revertant, In a word, DNA- and RNA-level variations are both higher in *bla* than in *tet*, either in the frameshift or the revertant but not in the wild type tested, demonstrating that these variations are likely to be caused by an active gene repair targeting specifically to the frameshifted *bla* rather than non-specific random mutagenesis.

3.6 PTCs signal the repair of the frameshift mutation

We speculated that it was the PTCs presented in the nonsense mRNAs that signal the repair of frameshift mutations. A *PTC-free bla-* (denoted as *bla#*) was derived from *bla-* by replacing each nonsense codon with an appropriate sense codon

following the *in vivo readthrough rules* (Table 1). The *bla#* gene was synthesized chemically, cloned in the plasmid vector pET-28a, and expressed in *E. coli* strain BL21. The *bla#* transformants were plated on ACPs, but no revertant was found. Their *bla#* genes were Sanger sequenced, but few overlapping peaks were observed, suggesting that the *bla#* gene is not repaired as efficiently as the *bla-* gene. Analysis of total and purified protein by SDS-PAGE revealed a protein band corresponding to the predicted molecular mass (34-kDa), representing a defective BLA expressed from *bla#* (Fig 5A). No lactamase activity was detected by iodimetry test of the defective product (Fig 5B), confirming that the frameshift is indeed a loss-of-function. In other words, the *bla#* gene is kept unchanged and expressed a dysfunctional product in its wrong reading frame when PTCs are replaced. The only difference between *bla#* and *bla-* is that: *bla-* consists of a certain number of PTCs, but *bla#* does not. Therefore, it must be the PTCs that signaled the repair of the frameshift mutation.

3.7 Identifying genes/proteins involved in frameshift repair

We speculated that frameshift repair requires the upregulation of relevant genes and pathways. By transcriptome sequencing and profiling, as shown in Fig 6A, the gene expression patterns of these strains are significantly different from each other. When compared to the wildtype (*W*), 292 / 247 genes show upregulated transcription in the frameshift (*FS*) / revertant (*R*), respectively (Fig 6A, Supplementary Data S1, S2) in the frameshift (*FS*) or revertant (*R*). In the pathway enrichment analysis, 65 or 58 pathways were identified as upregulated, respectively (Supplementary data S3, S4). Most upregulated genes/pathways do not appear to be involved in frameshift repair but more likely related to nutrition or ampicillin resistance, probably due to different media components. However, we manually selected upregulated genes/pathways that most probably involve in frameshift repair. The criterion for selection is whether a gene/pathway is involved in the processing of DNA or RNA (the storages of the genetic information). As shown in Fig 6B-6C and Table 4-5, the selected upregulated genes/pathways include *DNA replication* (Fig 6D), *RNA degradation* (Fig 6E), *mismatch repair,* and *homologous recombination* (Fig 6F).

Besides, the selenocysteinyl-tRNA-specific translation factor (*selB*) gene is found upregulated in the subculture of the revertant (Supplementary data S5). SelB is a specialized translation factor that takes the place of elongation factor EF-Tu, in the case of selenocysteines being inserted into a peptide chain at the site of a UGA codon [19, 20]. The upregulation of *selB* in the revertant suggested that the nonsense codons remained in the mRNAs might be readthrough in the revertants.

By comparing the read count and FPKM of each sample (Supplementary data S6), we found that the expression of the cytosine or isoguanine deaminase gene (*codA*) and the adenosine deaminase gene (*add*) were upregulated in the frameshift. Also, the expression of cytidine/deoxycytidine deaminase (*cdd*) and tRNA-specific adenosine deaminase (*tadA*) gene were upregulated in the revertant:

(1) In *E. coli*, cytosine or isoguanine deaminase (CodA) catalyzes the hydrolytic deamination of cytosine to uracil, and also the deamination of isoguanine, a mutagenic oxidation product of adenine and isocytosine, in DNA [21, 22]. In eukaryotes, cytosine deaminase is involved in RNA editing [23].

(2) In *E. coli*, cytidine/deoxycytidine deaminase (Cdd) scavenges exogenous and endogenous cytidine and 2'-deoxycytidine for uridine monophosphate (UMP) synthesis [24-26]. In eukaryotes, cytidine/deoxycytidine deaminase is also involved in RNA editing [27, 28].

(3) In *E. coli*, adenosine deaminase (Add) is an inducible enzyme that converts adenine, adenosine, and deoxyadenosine to guanine nucleotides [29, 30]. In eukaryotes, adenosine deaminase (ADAR2) is an important enzyme that is responsible for RNA editing [31-33].

(4) The tRNA-specific adenosine deaminase (TadA) catalyzes the deamination of adenosine to inosine at the wobble position 34 of tRNA-Arg2. TadA was first identified in yeast [34] and is the first prokaryotic RNA editing enzyme that is reported in *E. coli* [35]. TadA is related to the eukaryotic RNA editing enzymes, including ADAR1 and ADAR2, which is widely existing in many eukaryotic species, such as humans [36], mouse [37] and fruit fly [38].

As mentioned above, Sanger sequencing showed that the base substitutions in the repaired *bla* are mostly G:C←→A:T transitions, suggesting that they are derived most probably from the hydrolytic deamination of cytosine to uracil in RNA editing. As is known, deamination generates deoxyinosine from deoxyadenosine, which links DNA mismatch repair and RNA editing [39]. In the above, we show that the deamination and the mismatch repair pathways are both upregulated in the frameshift, suggesting that RNA-editing and mismatch-repair might be both involved. These genes are also upregulated in the revertant, suggesting that gene repair is not ceased but still ongoing in the revertants.

4. Discussion

4.1 A new model for the repair of frameshift mutations

The mechanisms for the repair of DNA damages and point mutations have been well-understood. However, the molecular mechanism for the recognition and repair of frameshift mutations is yet unknown. In the above, the expression of the PTC-free *bla* indicated that PTCs signaled the repair of the frameshift mutation. We postulated that a nonsense mRNA (a transcript of the frameshift gene) must be involved in the repair. Conceivably, it is not possible to recognize a PTC in a double helix DNA but more likely in a nonsense mRNA during the process of translation.

To gain further insight into this very complex problem, we conducted an in-depth survey of previous studies that covered a wide range of prokaryotic and eukaryotic organisms. By presuming that frameshift repair is widely existing and the schematism is conserved among species, here we integrate plenty of fragmentary pieces of evidence and propose a hypothetical model for frameshift repair, referred to as *Nonsense-mediated gene repair* (NMGR), consisting five significant steps (Fig 7):

Step 1. ***Nonsense mRNA recognition***: in principle, a frameshift coding gene cannot be recognized at the DNA level, and therefore must be transcribed firstly. The transcripts of the frameshift gene (nonsense mRNAs) are recognized by the RNA surveillance system [40]. When a nonsense mRNA is translated, the ribosomes are blocked when they encounter a stop codon [41].

Step 2. ***Nonsense mRNA processing***: Nonsense-mediated mRNA decay (NMD) is the best-studied nonsense mRNA processing pathway. In eukaryotes, NMD targeting mRNAs harboring PTCs for degradation enhances their decay rate and reduces the abundance of the mRNA transcribed from that gene. In bacteria, it is not yet clear how nonsense mRNAs are processed, but it is known that the quality of their mRNAs is also tightly controlled [42]. Therefore, we speculate that NMD may also exist in bacteria, although it has not yet been characterized. Nevertheless, in both prokaryotes and eukaryotes, nonsense mRNAs must be processed by one of the pathways, either degradation (step 2), translation (step 3), or editing (step 4).

Step 3. ***Nonsense mRNA translation*****:** Nonsense mRNAs are not always fully degraded but sometimes could also be translated. A frameshifted coding gene is sometimes expressed through some unique mechanisms, including programmed translational frameshifting and translational readthrough. If the target gene is essential, a cell survives if and only if it produces functional proteins by translating a nonsense transcript/mRNA and dies if no protein is produced or the products are truncated or nonfunctional.

Step 4. ***RNA editing***: nonsense mRNAs must be involved in the recognition and repair of the frameshift mutation, but a nonsense mRNA is defective by itself, and thus, cannot template the repair of its coding gene directly. Therefore, it is very likely that a nonsense mRNA must be modified by RNA editing before templating the repair. Nonsense mRNAs are recognized by PTC signaling, but it ought to be impossible for a cell to locate the frameshift-causing indel precisely, either in a genomic DNA or a transcript. Therefore, usually, it is not possible to restore the original coding sequence. Therefore, the cells have to insert, delete, and substitute nucleotide(s) in the target mRNA in or near to a PTC randomly, resulting in a variety of variants. After editing, the translation is restored, and a functional BLA is produced if the mRNA is repaired.

Step 5. ***RNA-directed DNA repair***: to repair the frameshifted coding gene, mRNAs are transported back to chromosomal DNA to localize its defective coding sequence.

The defective coding gene is repaired by using an edited mRNA as the template. A bacterium recovers if functional mRNAs and proteins are produced by a standard transcription and translation process.

4.2 *The NMGR model better explains the variations of the bla genes*

As shown by a question mark in Fig 7, there exist two possible NMGR strategies: (1) *direct NMGR*: using an original nonsense mRNA to identify and edit its defective CDS by DNA mutagenesis; however, the edited CDS must be re-transcribed and re-translated to produce any functional proteins; (2) *indirect NMGR*: the nonsense mRNAs are first edited, and then transported back to DNA to identify, and to template the repair of its defective CDS, *i.e.*, a nonsense mRNA is functionalized by RNA editing before directing the repair of its defective CDS. Compared to the direct strategy, indirect NMGR seems to be more reasonable and cost-effective, because the change of CDS happens only when a cell has obtained a functional mRNA.

As described above, distinct types of bases were deleted, inserted, or substituted in the *bla* genes of the revertants. These editing events are mostly located near to or in the downstream of the frameshift-causing indel, but not in the upstream (Fig 2B-2C). This phenomenon can be well-explained by the PTC-signaled editing of the nonsense mRNAs: PTCs presented downstream the indel may not only signal the recognition of nonsense mRNA but also serve as flags for RNA editing. Because the PTCs appear in the nonsense mRNAs only in downstream of the indel, RNA editing is restricted near to or downstream of the indel. Therefore, the change in DNA is also restricted to the corresponding region in the coding gene. Therefore, a frameshift mutation is repaired indirectly by editing nonsense mRNAs rather than the CDS directly. It is possible that RNA editing may sometimes cause even more problems in a nonsense mRNA, while these non-repaired mRNAs can be recycled or degraded. This process is repeated in a cell until the gene is repaired or PTCs are eliminated.

When the cells are cultured in ampicillin-containing media, the survival of a cell depends on whether it can produce enough functional proteins before it was killed. Through transcribing, each copy of the target gene can produce many nonsense mRNA copies, and each copy can be edited in different ways without changing their

CDSs. Therefore, the probability of generating a functional mRNA by nonsense mRNA editing is much greater than the probability of direct random mutagenesis in the coding sequences. Therefore, it may bring a greater survival rate to the cells when a fatal frameshift mutation occurs.

4.3 *Numerous previous studies are supportive of the NMGR model*

In the above, we proposed that NMGR depends on RNA-directed DNA repair but is triggered by PTCs, and the nonsense mRNAs are edited before directing the repair of the coding DNA. The proposed NMGR model well explains our observations and is consistent with numerous previous studies. The model is an integration of several key links, which have already been intensively studied:

(1). mRNA decay is linked to other nonsense mRNA processing pathways

In eukaryotes, NMD uses the presence of an exon junction complex (EJC) downstream of PTCs as a second signal to distinguish a PTC from a true stop codon [43]. In addition to NMD, the nonsense mRNAs may also be subject to some other pathways, including translational repression, transcriptional silencing, and alternative splicing, all of which are linked to NMD [44-46]. In eukaryotes, the NMD regulating factors, including three interacting proteins, UPF1, UPF2 (also known as NMD2), and UPF3, are encoded by highly conserved genes originally identified in yeast [47].

A frameshifted coding gene is sometimes expressed through some special mechanisms, including programmed translational frameshifting [48-50] and translational readthrough [51]. Ribosomal frameshifting occurs in various organisms, from *E. coli* to mammals, whereby the ribosome is guided toward a triplet codon that is shifted one basepair position to the upstream (+1) or downstream (−1) [52]. In yeast, telomere maintenance is globally controlled by the ribosomal frameshifting process and NMD [53]. Translational readthrough is a process wherein a nonsense mRNA is translated by translating the PTCs, each into a specific amino acid [54]. Mutations in *upf* genes caused nonsense suppression and subsequent readthrough in yeast and 5-20 fold decreases in the rates of NMD [55], suggesting a direct link between NMD, nonsense suppression, and readthrough. The three Upf proteins all interact with Sup35, and Upf1 interact further with a second release factor, Sup45 (eRF1), and binding of

an SMG-1-Upf1-eRF1-eRF3 (SURF) complex to EJC triggers Upf1 phosphorylation and NMD [56]. Besides, the nonsense suppression caused by *upf* mutations is additive to those by *sup35* and *sup45* mutations [57], suggesting that nonsense suppression is a nonsense-mRNA processing pathway, which is an alternative to NMD. In a word, although nonsense mRNAs usually lead to translation errors and NMD, when NMD is inhibited, they can also be translated by ribosomal frameshifting or translational readthrough.

(2). RNA can direct DNA repair

All types of RNAs are subject to translation, degradation, or processing, and a variety of molecular/cellular mechanisms are involved in RNA processing [58]. In the last few decades, studies have established a link between DNA repair and RNA processing. For example, several proteins respond to DNA damage, and base-excision repair enzymes, such as SMUG1, APE1, and PARP1, have been shown to participate in RNA surveillance and processing [59].

As is known, RNA is transcribed in the nucleus and then transported to the cytoplasm in eukaryotes [60]. Recently, there is growing evidence that RNAs can also be transported back to the nucleus to repair DNA double-strand breaks (DSBs), a process known as *RNA-directed DNA Repair* [61-63]. In yeast [63], bacteria, and human embryonic kidney cells [64], a synthetic DNA/RNA hybrid or an RNA-only oligonucleotide can serve as a template for DNA synthesis for the repair of a DSB in a homologous DNA. In *E. coli*, small RNA patches can also direct DNA modification [65]. Here we report that mRNA can be edited and direct precise repair of a frameshift mutation of its CDS by itself. Besides, we showed that many G:C→A:T substitutions were observed in the repaired bla in the revertants; four deaminase genes (*codA*, *cdd*, *add,* and *tadA*) and the mismatch repair pathways were upregulated in the frameshift or the revertant. Therefore, we speculate that mismatch repair and RNA editing are both involved in frameshift repair. In addition, mismatch repair and homologous recombination are likely to be involved, as they are both upregulated in the frameshift and the revertant tested (Fig 6F).

(3). RNA editing is linked to DNA repair

RNA editing has been studied extensively and intensively, indicating that RNA editing is widely existing in *E. coli* [27], plants [66], humans [67-69], and mammals [70]. Although there are several different RNA editing mechanisms, including A-to-I editing [71], C-to-U editing [72], and insertion/deletion of uridine [73], the basic schematism are highly conserved among species. In particular, the nonsense *apoB* mRNA editing complex suppresses its own NMD, suggesting that nonsense mRNAs can be subjected to RNA editing, which is linked to the NMD pathway [74].

Both DNA mismatch repair and RNA editing involve deamination, the removal of an amine group from a DNA/RNA molecule. Enzymes that catalyze this reaction are called DNA/RNA deaminases. Deamination of DNA deoxyadenosine bases generates deoxyinosine, create missense mutations predisposing humans to cancer, and interfere with other molecular genetic processes. Cytidine deamination was first recognized as a mechanism for RNA editing in wheat [75]. The characterization of tRNA-specific adenosine deaminases (ADATs) has led to the suggestion that these enzymes, as well as cytidine/adenosine deaminases (CDARs/ADARs), acting on pre-mRNAs, belong to a superfamily of RNA dependent deaminases [76]. The first example of C→U editing is the apolipoprotein B mRNA, in which single, site-specific cytidine deamination introduces a stop codon (UAA) into the reading frame of the edited transcript [77]. It is well known that deamination based RNA editing plays many important roles, and dysregulated RNA editing might contribute to genomic instability in cancer [78]. In eukaryotes, several studies have provided evidence that deamination is related to both DNA repair and RNA editing. For example, endonuclease V, an RNA editing enzyme that is highly conserved from *E. coli* to human [39], is involved in excision repair that removes deoxyinosine from DNA.

TadA, the first prokaryotic deaminase for RNA editing identified in *E. coli,* displays sequence similarity to the yeast tRNA deaminase subunit Tad2p [35]. Besides, cytidine deaminase (CodA) has been used as a molecular model for RNA editing and a mechanism for RNA substrate recognition [27]. These studies suggest that RNA editing is conserved between prokaryotes and eukaryotes, and a direct link between the regulation of mismatch repair and RNA editing, and therefore support our

speculation that mismatch repair and RNA editing may both involve in frameshift repair.

In a previous study, we reported that the genetic code, coding genes, and genomes are optimal for frameshift tolerance [79]. Recently, two groups of researchers reported that premature termination codons were found widespread in populations of humans [80] and fruit flies [81], suggesting ongoing evolutionary plasticity at the genetic level in modern human and other species. Moreover, reading frame transitions are widely observed in prokaryotic genomes [82], and frameshifted genes can be retained for millions of years and enable the acquisition of new functions [83]. In this study, we demonstrate that PTCs signal the repair of frameshift mutations, resulting in a variety of frameshift homologs, among which functional ones were preserved in the evolution history. In summary, the shiftability of the genetic code forms the theoretical basis of frameshift tolerance, and the NMGR mechanism explains how frameshift homologs are produced, taken together, they well explain why functional frameshift homologs are widely observed in nature [82]. In conclusion, NMGR is not only the molecular mechanism for frameshift repair but also be a driving force for molecular evolution and a source of genetic diversity.

Data accessibility. The preprint version of this article is available from bioRxiv at https://doi.org/10.1101/069971. Supplementary data are available from FigShare at https://doi.org/10.6084/m9.figshare.11497128.v1. File S1 and S2 contain the list of upregulated genes in the frameshift (FS) or the initial revertant (R) compared to the wild-type (W), respectively; File S3 and S4 contain the list of upregulated pathways in the frameshift (FS) or the initial revertant (R) compared to the wild-type (W), respectively; File S5 contains the list of upregulated pathways in the subculture of the revertant (RE) compared to the wild-type (W); File S6 contains the ReadCount and FPKM for all genes.

Author Contributions

Xiaolong Wang conceived the study, designed the experiments, analyzed the data, and wrote the paper. Xuxiang Wang, Chunyan Li, Haibo Peng, and Yalei Wang performed the

experiments. Gang Chen and Jianye Zhang provided some equipment, discussed the paper, and gave some suggestions.

Acknowledgments:

This study was supported by The *National Natural Science Foundation of China* (Grant No. 31571369).

Figure legends

Fig 1. The introduction of a frameshift mutation in the upstream of the *bla* gene in the plasmid pBR322. Sanger sequencing diagrams: (A) the wild type (*bla*+); (B) the frameshift mutant (*bla*-). (C) Alignment of the nucleotide sequence of *bla*+ and *bla*-.

Fig 2. Growth of the wild-type/frameshift *E. coli* and Sanger sequencing of the *bla* genes:

(A) Growth of the wild-type and frameshift *E. coli* on ampicillin- or tetracycline plates: *bla*+: wild type; *bla*-: frameshift; *bla**: revertants; *blank*: blank control; ACP: ampicillin-containing plates; ACP: tetracycline-containing plates.

(B) The Sanger diagram of the *bla* genes of the revertants: *Top*: in the initial revertants, most of the *bla* genes are still frameshifted; *Middle*: in the later subcultures, sequencing diagrams contains two sets of overlapping peaks, the main peaks are the frameshifted *bla*, the second peaks are the repaired *bla*; *Bottom*: in the final fast-growing subcultures, sequencing diagrams contains very few overlapping peaks.

(C) The ClustalW alignment of the wild-type, frameshifted and repaired *bla* gene sequences: ***bla*+**: wild-type; *bla*-: frameshift; *bla*A-E*: different (independent) revertants; *box*: the base G deleted in the OE-PCR; *underlined*: the bases (in blue) inserted or deleted in the revertants; *circles*: base substitutions (G:C←→A:T transitions).

(D) The codon and amino acid unified alignment (CAUSA) of *bla* and their translated protein sequences: *underlined*: the base G deleted in codon 46; *red box*: stop codons; *red star*: stop signals.

Fig 3. The traditional *random mutagenesis* model for frameshift repair: *Step 1*: a frameshift mutation occurs in a coding gene, many hidden stop codons (*red bars*) emerge in the CDS, and in the nonsense mRNAs, cause *translational termination* and resulted in truncated products; *Step 2*: the coding gene is repaired by a *reverse mutation* through random mutagenesis, the reading frame is restored immediately, stop codons are hidden (*green bars*), and translation proceeds, producing functional mRNAs and proteins, which are identical and same to the original genes and proteins.

Fig 4. NGS resequencing and variation analysis the genomes of the frameshift/revertant:

(A) The genome-wide variation map of the frameshift /revertant: to visualize the coverages and the distribution of mutations, Circos software was used to show the clean reads of each sample, the depth of coverage (number of clean reads covering a given base position of the reference sequence), and the SNP/indel, using the reference genome sequence as a ruler. The outermost circle: the coordinate position of the reference genome, and then, outside-in, the distribution of indel/SNPs (red dots), the depth of coverage (grey), the GC content (green) and the GC skew value (pink) in the reference genome, respectively. There are two samples, the frameshift (*bla*-) and the revertant (*bla**), so there are two sets of indel/SNP and reads coverage depth mapped onto the reference genome sequences.

(B) The distribution of the numbers of SNPs/indels of the frameshift/revertant on the reference genome: The abscissa axis shows the chromosome coordination of the reference genome sequence, and the ordinate axis represents the SNP/indel density on the genome sequence in per 10 kb. The left scale: SNP (red peaks); the right scale: indel (blue peaks). See Table 2 for data.
(C) Mapping of reads for *bla* coding sequences and RNA transcripts. See Table 3 for data.
(D) Mapping of reads for *tet* coding sequences and RNA transcripts. See Table 3 for data.

Fig 5. Expression of the PTC-free frameshifted *bla* (*bla#*) in *E. coli* BL21:
(A) SDS-PAGE of the lysates of pET28a-*bla#* recombinant *E. coli* cells; Lane 1: protein marker; Lane 2 and 3: whole-cell lysates of uninduced and induced cells; lane 4 and 5: the supernatant of uninduced and induced cells; lane 6 and 7: the precipitate of uninduced and induced cells;
(B) Detection of β-lactamase activity by iodimetry: (1) the revertants of *E. coli* BL21/pET28a-bla- caused the fading of the iodine/starch solution, suggesting that their *bla-* was repaired, produced active β-lactamases, and transformed ampicillin into penicillium thiazole acid, which binds with starch and competitively inhibits the binding of iodine to starch; (2) *E. coli* BL21/pET28a-bla# did not cause fading, suggesting that it was not repaired but expressed a frameshifted BLA (BLA#), which is an inactive product; (3) the negative control (*E. coli* BL21), no β-lactamase, no fading.

Fig 6. The transcriptome analysis of different *E. coli* strains: *W-1*: wildtype; *FS-2*: frameshift; *R-3*: a revertant; *RE-4*: a subculture of the revertant; (A) The number of differentially expressed genes (DEGs) and the cluster analysis of DEGs; (B) the most enriched GO terms by comparing the frameshift (FS_2) with the wild type (W_1); (C) the most enriched GO terms by comparing the revertant (R_3) with the frameshift (FS_2); (D) DNA replication pathway: (E) genes upregulated in the RNA degradation/processing pathway; (F) genes upregulated in the mismatch repair and homologous recombination pathways. *Underlined* in B, C: GO terms involved in the processing of DNA or RNA, which probably also involve in frameshift repair; in D, E, F: *green box:* genes involved in the pathways; *rex box*: genes upregulated in the frameshift or revertant.

Fig 7. The new model for frameshift repair: nonsense-mediated gene repair (NMGR): *Step 1*: when the frameshifted gene is transcribed and translated, the nonsense mRNAs containing many PTCs (*yellow bars*) are identified by mRNA surveillance through PTCs signaling; *Step 2*: some of the nonsense mRNAs causing *translational termination* are degraded by NMD; *Step 3*: some of the nonsense mRNAs are translated by translational frameshifting or readthrough; *Step 4*: some of the nonsense mRNAs are repaired by *RNA editing*, PTCs are replaced by sense codons (*green bars*) through base insertion, deletion, and substitution; *Step 5*: some of the original or repaired nonsense mRNAs are transported to identify and localize the defective coding gene; *Step 6*: the coding gene is repaired by *RNA-directed DNA repair*, the reading frame is restored partially, and the translation proceeds, the remaining stop codons were readthrough, producing functional mRNAs and proteins with many variations, and so that the host recovered gradually.

Table 1. The natural suppressor tRNA for nonsense mutations (the *readthrough rules*).

Site	tRNA (AA)	Wild type		Correction	
		Code	**Anti-code**	**Code**	**Anti-code**
supD	Ser (S)	→ UCG	CGA←	→ UAG	CUA←
supE	Gln (Q)	→ CAG	CUG←	→ UAG	CUA←
supF	Tyr (Y)	→ UAC	GUA←	→ UAG	CUA←
supG	Lys (K)	→ AAA	UUU←	→ UAA	UUA←
supU	Trp (W)	→ UGG	CCA←	→ UGA	UCA←

Table 2. The summary of SNP/indel variations in the *bla**/- genome

Sample	Indels			SNPs				
	INS	DEL	Total	ts	tv	ts/tv	Total	Density
bla-	2	4	6	43	27	1.59	70	0.02/Kb
*bla**	2	3	5	42	27	1.56	69	0.01/Kb

Note: INS: insertion; DEL: deletion; ts: transitions; tv: transversions;

Table 3. The count and density of raw SNPs of the two genes in the plasmid pbr322

	tet (1191 bp)		bla (861 bp)	
Count of SNPs	Count	Density	Count	Density
Frameshift-DNA	76	63.81/Kb	60	69.69/Kb
Revertant-DNA	70	58.77/Kb	68	78.98/Kb
Wildtype-RNA	5	4.20/Kb	2	2.32/Kb
Revertant-RNA	6	5.04/Kb	6	6.97/Kb

1

2 Table 4. Possible NMGR-relevant genes and pathways that were upregulated in the frameshift

Pathway Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
ABC transporters	KEGG	eco02010	32	171	0.0001	0.0090
RNA degradation	KEGG	eco03018	1	15	0.7419	0.9762
DNA replication	KEGG	eco03030	1	17	0.7822	0.9762
Mismatch repair	KEGG	eco03430	1	22	0.8575	0.9796
Homologous recombination	KEGG	eco03440	1	27	0.9069	0.9796

3

4

5 Table 5. Possible NMGR-relevant genes and pathways that were upregulated in the revertant

Pathway Term	Database	ID	Input number	Background number	P-Value	Corrected P-Value
ABC transporters	KEGG	eco02010	32	171	3.53E-05	0.0018
Homologous recombination	KEGG	eco03440	2	27	0.6539	0.9519
RNA degradation	KEGG	eco03018	3	15	0.1017	0.5470
DNA replication	KEGG	eco03030	1	17	0.7564	0.9519
Mismatch repair	KEGG	eco03430	1	22	0.8357	0.9519

6

7

References

1. Seligmann, H. and D.D. Pollock, *The ambush hypothesis: hidden stop codons prevent off-frame gene reading.* DNA Cell Biol, 2004. **23**(10): p. 701-5.
2. Jensen, N.M., et al., *An update on targeted gene repair in mammalian cells: methods and mechanisms.* J Biomed Sci, 2011. **18**: p. 10.
3. Streisinger, G., et al., *Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday.* Cold Spring Harb Symp Quant Biol, 1966. **31**: p. 77-84.
4. Lieb, M., *Forward and Reverse Mutation in a Histidine-Requiring Strain of Escherichia Coli.* Genetics, 1951. **36**(5): p. 460-77.
5. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
6. DePristo, M.A., et al., *A framework for variation discovery and genotyping using next-generation DNA sequencing data.* Nat Genet, 2011. **43**(5): p. 491-8.
7. Krzywinski, M., et al., *Circos: an information aesthetic for comparative genomics.* Genome Res, 2009. **19**(9): p. 1639-45.
8. Fan, X., et al., *BreakDancer: Identification of Genomic Structural Variation from Paired-End Read Mapping.* Curr Protoc Bioinformatics, 2014. **45**: p. 15 6 1-11.
9. Chen, K., et al., *BreakDancer: an algorithm for high-resolution mapping of genomic structural variation.* Nat Methods, 2009. **6**(9): p. 677-81.
10. Langmead, B., *Aligning short sequencing reads with Bowtie.* Curr Protoc Bioinformatics, 2010. **Chapter 11**: p. Unit 11 7.
11. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.
12. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
13. Wang, L., et al., *DEGseq: an R package for identifying differentially expressed genes from RNA-seq data.* Bioinformatics, 2010. **26**(1): p. 136-8.
14. Young, M.D., et al., *Gene ontology analysis for RNA-seq: accounting for selection bias.* Genome Biol, 2010. **11**(2): p. R14.
15. Xie, C., et al., *KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W316-22.
16. Lee, H., et al., *Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing.* Proc Natl Acad Sci U S A, 2012. **109**(41): p. E2774-83.
17. Wang, X. and C. Yang, *CAUSA 2.0: accurate and consistent evolutionary analysis of proteins using codon and amino acid unified sequence alignments.* PeerJ PrePrints **3**.
18. Schaaper, R.M., *Mechanisms of mutagenesis in the Escherichia coli mutator mutD5: role of DNA mismatch repair.* Proc Natl Acad Sci U S A, 1988. **85**(21): p. 8126-30.
19. Itoh, Y., S. Sekine, and S. Yokoyama, *Crystal structure of the full-length bacterial selenocysteine-specific elongation factor SelB.* Nucleic Acids Res, 2015. **43**(18): p. 9028-38.

20. Baron, C. and A. Bock, *The length of the aminoacyl-acceptor stem of the selenocysteine-specific tRNA(Sec) of Escherichia coli is the determinant for binding to elongation factors SELB or Tu.* J Biol Chem, 1991. **266**(30): p. 20375-9.

21. Ireton, G.C., et al., *The structure of Escherichia coli cytosine deaminase.* J Mol Biol, 2002. **315**(4): p. 687-97.

22. Danielsen, S., et al., *Characterization of the Escherichia coli codBA operon encoding cytosine permease and cytosine deaminase.* Mol Microbiol, 1992. **6**(10): p. 1335-44.

23. MacGinnitie, A.J., S. Anant, and N.O. Davidson, *Mutagenesis of apobec-1, the catalytic subunit of the mammalian apolipoprotein B mRNA editing enzyme, reveals distinct domains that mediate cytosine nucleoside deaminase, RNA binding, and RNA editing activity.* J Biol Chem, 1995. **270**(24): p. 14768-75.

24. Vita, A., et al., *Cytidine deaminase from Escherichia coli B. Purification and enzymatic and molecular properties.* Biochemistry, 1985. **24**(21): p. 6020-4.

25. Ashley, G.W. and P.A. Bartlett, *Purification and properties of cytidine deaminase from escherichia coli.* J Biol Chem, 1984. **259**(21): p. 13615-20.

26. Cohen, R.M. and R. Wolfenden, *Cytidine deaminase from Escherichia coli. Purification, properties and inhibition by the potential transition state analog 3,4,5,6-tetrahydrouridine.* J Biol Chem, 1971. **246**(24): p. 7561-5.

27. Navaratnam, N., et al., *Escherichia coli cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition.* J Mol Biol, 1998. **275**(4): p. 695-714.

28. Navaratnam, N., et al., *Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site.* Cell, 1995. **81**(2): p. 187-95.

29. Nygaard, P., *Adenosine deaminase from Escherichia coli.* Methods Enzymol, 1978. **51**: p. 508-12.

30. Koch, A.L. and G. Vallee, *The properties of adenosine deaminase and adenosine nucleoside phosphorylase in extracts of Escherichia coli.* J Biol Chem, 1959. **234**(5): p. 1213-8.

31. Oakes, E., et al., *Adenosine Deaminase That Acts on RNA 3 (ADAR3) Binding to Glutamate Receptor Subunit B Pre-mRNA Inhibits RNA Editing in Glioblastoma.* J Biol Chem, 2017. **292**(10): p. 4326-4335.

32. Kim, U., et al., *Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing.* Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11457-61.

33. Kim, U. and K. Nishikura, *Double-stranded RNA adenosine deaminase as a potential mammalian RNA editing factor.* Semin Cell Biol, 1993. **4**(4): p. 285-93.

34. Gerber, A., et al., *Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2.* EMBO J, 1998. **17**(16): p. 4780-9.

35. Wolf, J., A.P. Gerber, and W. Keller, *tadA, an essential tRNA-specific adenosine deaminase from Escherichia coli.* EMBO J, 2002. **21**(14): p. 3841-51.

36. Maas, S., A.P. Gerber, and A. Rich, *Identification and characterization of a human tRNA-specific adenosine deaminase related to the ADAR family of pre-mRNA editing enzymes.* Proc Natl Acad Sci U S A, 1999. **96**(16): p. 8895-900.

37. Maas, S., Y.G. Kim, and A. Rich, *Sequence, genomic organization and functional expression of the murine tRNA-specific adenosine deaminase ADAT1.* Gene, 2000. **243**(1-2): p. 59-66.

38. Keegan, L.P., et al., *The properties of a tRNA-specific adenosine deaminase from Drosophila melanogaster support an evolutionary link between pre-mRNA editing and tRNA modification.* Mol Cell Biol, 2000. **20**(3): p. 825-33.

39. Kuraoka, I., *Diversity of Endonuclease V: From DNA Repair to RNA Editing.* Biomolecules, 2015. **5**(4): p. 2194-206.

40. Niu, D.K. and J.L. Cao, *Nucleosome deposition and DNA methylation may participate in the recognition of premature termination codon in nonsense-mediated mRNA decay.* FEBS Lett, 2010. **584**(16): p. 3509-12.

41. Hwang, J. and Y.K. Kim, *When a ribosome encounters a premature termination codon.* BMB Rep, 2013. **46**(1): p. 9-16.

42. Richards, J., et al., *Quality control of bacterial mRNA decoding and decay.* Biochim Biophys Acta, 2008. **1779**(9): p. 574-82.

43. Scofield, D.G., X. Hong, and M. Lynch, *Position of the final intron in full-length transcripts: determined by NMD?* Mol Biol Evol, 2007. **24**(4): p. 896-9.

44. Isken, O., et al., *Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay.* Cell, 2008. **133**(2): p. 314-27.

45. Ward, A.J., et al., *Nonsense-mediated decay as a terminating mechanism for antisense oligonucleotides.* Nucleic Acids Res, 2014. **42**(9): p. 5871-9.

46. Lewis, B.P., R.E. Green, and S.E. Brenner, *Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 189-92.

47. He, F., A.H. Brown, and A. Jacobson, *Upf1p, Nmd2p, and Upf3p are interacting components of the yeast nonsense-mediated mRNA decay pathway.* Mol Cell Biol, 1997. **17**(3): p. 1580-94.

48. Chen, J., et al., *Dynamic pathways of -1 translational frameshifting.* Nature, 2014. **512**(7514): p. 328-32.

49. Bekaert, M., et al., *Identification of programmed translational -1 frameshifting sites in the genome of Saccharomyces cerevisiae.* Genome Res, 2005. **15**(10): p. 1411-20.

50. Klobutcher, L.A. and P.J. Farabaugh, *Shifty ciliates: frequent programmed translational frameshifting in euplotids.* Cell, 2002. **111**(6): p. 763-6.

51. Eswarappa, S.M., et al., *Programmed translational readthrough generates antiangiogenic VEGF-Ax.* Cell, 2014. **157**(7): p. 1605-18.

52. Namy, O., et al., *A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting.* Nature, 2006. **441**(7090): p. 244-7.

53. Advani, V.M., A.T. Belew, and J.D. Dinman, *Yeast telomere maintenance is globally controlled by programmed ribosomal frameshifting and the nonsense-mediated mRNA decay pathway.* Translation (Austin), 2013. **1**(1): p. e24418.

54. Kopelowitz, J., et al., *Influence of codon context on UGA suppression and readthrough.* J Mol Biol, 1992. **225**(2): p. 261-9.

55. Leeds, P., et al., *Gene products that promote mRNA turnover in Saccharomyces cerevisiae.* Mol Cell Biol, 1992. **12**(5): p. 2165-77.

56. Kashima, I., et al., *Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay.* Genes Dev, 2006. **20**(3): p. 355-67.

57. Keeling, K.M., et al., *Leaky termination at premature stop codons antagonizes nonsense-mediated mRNA decay in S. cerevisiae.* RNA, 2004. **10**(4): p. 691-703.

58. Isken, O. and L.E. Maquat, *Quality control of eukaryotic mRNA: safeguarding cells from abnormal mRNA function.* Genes Dev, 2007. **21**(15): p. 1833-56.

59. Vohhodina, J., D.P. Harkin, and K.I. Savage, *Dual roles of DNA repair enzymes in RNA biology/post-transcriptional control.* Wiley Interdiscip Rev RNA, 2016.

60. Nakielny, S. and G. Dreyfuss, *Transport of proteins and RNAs in and out of the nucleus.* Cell, 1999. **99**(7): p. 677-90.

61. Keskin, H., C. Meers, and F. Storici, *Transcript RNA supports precise repair of its own DNA gene.* RNA Biol, 2016. **13**(2): p. 157-65.

62. Keskin, H., et al., *Transcript-RNA-templated DNA recombination and repair.* Nature, 2014. **515**(7527): p. 436-9.

63. Storici, F., et al., *RNA-templated DNA repair.* Nature, 2007. **447**(7142): p. 338-41.

64. Shen, Y., et al., *RNA-driven genetic changes in bacteria and in human cells.* Mutat Res, 2011. **717**(1-2): p. 91-8.

65. Thaler, D.S., G. Tombline, and K. Zahn, *Short-patch reverse transcription in Escherichia coli.* Genetics, 1995. **140**(3): p. 909-15.

66. Takenaka, M., et al., *RNA editing in plants and its evolution.* Annu Rev Genet, 2013. **47**: p. 335-52.

67. Zougman, A., et al., *Evidence for insertional RNA editing in humans.* Curr Biol, 2008. **18**(22): p. 1760-5.

68. Athanasiadis, A., A. Rich, and S. Maas, *Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome.* PLoS Biol, 2004. **2**(12): p. e391.

69. Picardi, E., D.S. Horner, and G. Pesole, *Single cell transcriptomics reveals specific RNA editing signatures in the human brain.* RNA, 2017.

70. Slavov, D., M. Clark, and K. Gardiner, *Comparative analysis of the RED1 and RED2 A-to-I RNA editing genes from mammals, pufferfish and zebrafish.* Gene, 2000. **250**(1-2): p. 41-51.

71. Terajima, H., et al., *ADARB1 catalyzes circadian A-to-I editing and regulates RNA rhythm.* Nat Genet, 2017. **49**(1): p. 146-151.

72. Blanc, V. and N.O. Davidson, *C-to-U RNA editing: mechanisms leading to genetic diversity.* J Biol Chem, 2003. **278**(3): p. 1395-8.

73. Kable, M.L., et al., *RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA.* Science, 1996. **273**(5279): p. 1189-95.

74. Chester, A., et al., *The apolipoprotein B mRNA editing complex performs a multifunctional cycle and suppresses nonsense-mediated decay.* EMBO J, 2003. **22**(15): p. 3971-82.

75. Yu, W. and W. Schuster, *Evidence for a site-specific cytidine deamination reaction involved in C to U RNA editing of plant mitochondria.* J Biol Chem, 1995. **270**(31): p. 18227-33.

76. Gerber, A.P. and W. Keller, *RNA editing by base deamination: more enzymes, more targets, new mysteries.* Trends Biochem Sci, 2001. **26**(6): p. 376-84.

77. Young, S.G., *Recent progress in understanding apolipoprotein B.* Circulation, 1990. **82**(5): p. 1574-94.

78. Anant, S. and N.O. Davidson, *Hydrolytic nucleoside and nucleotide deamination, and genetic instability: a possible link between RNA-editing enzymes and cancer?* Trends Mol Med, 2003. **9**(4): p. 147-52.

79. Wang, X., et al., *Why are frameshift homologs widespread within and across species?* bioRxiv, 2016.

80. Fujikura, K., *Premature termination codons in modern human genomes.* Sci Rep, 2016. **6**: p. 22468.

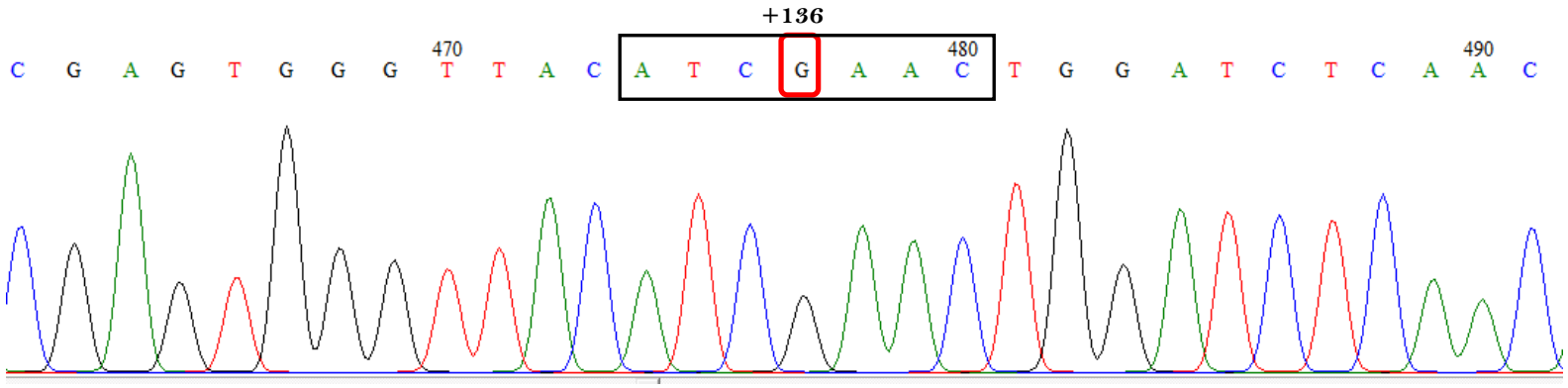
81. Yang, H., et al., *Expression profile and gene age jointly shaped the genome-wide distribution of premature termination codons in a Drosophila melanogaster population.* Mol Biol Evol, 2015. **32**(1): p. 216-28.

82. Antonov, I., et al., *Identification of the nature of reading frame transitions observed in prokaryotic genomes.* Nucleic Acids Res, 2013. **41**(13): p. 6514-30.

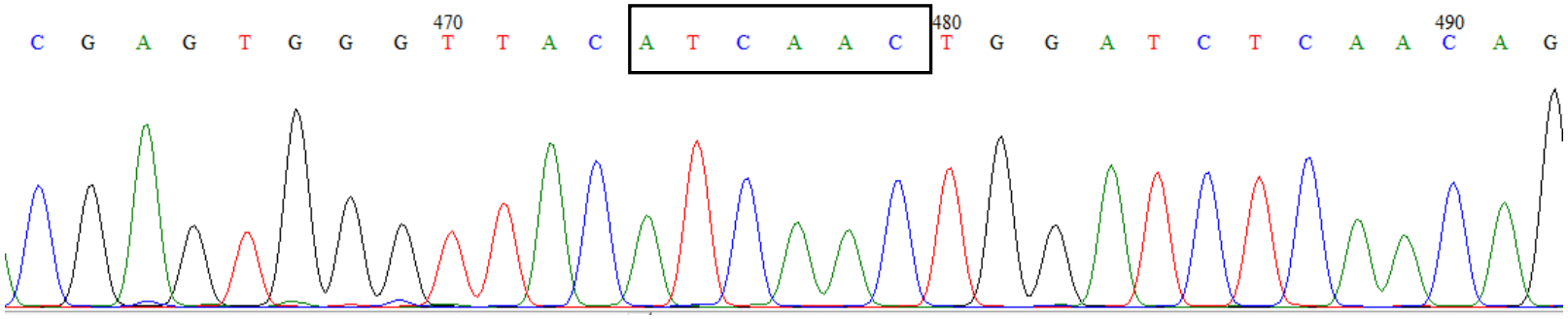
83. Raes, J. and Y. Van de Peer, *Functional divergence of proteins through frameshift mutations.* Trends Genet, 2005. **21**(8): p. 428-31.

Fig 1. The frameshift mutagenesis of the *bla* gene in the plasmid pBR322

1A



1B



1C

				136			
Wild-type (*bla* +):	**…A**	**T**	**C**	**G**	**A**	**A**	**C…**
	|	|	|		|	|	|
Frameshift (*bla* -):	**…A**	**T**	**C**	**-**	**A**	**A**	**C…**

Fig 2A

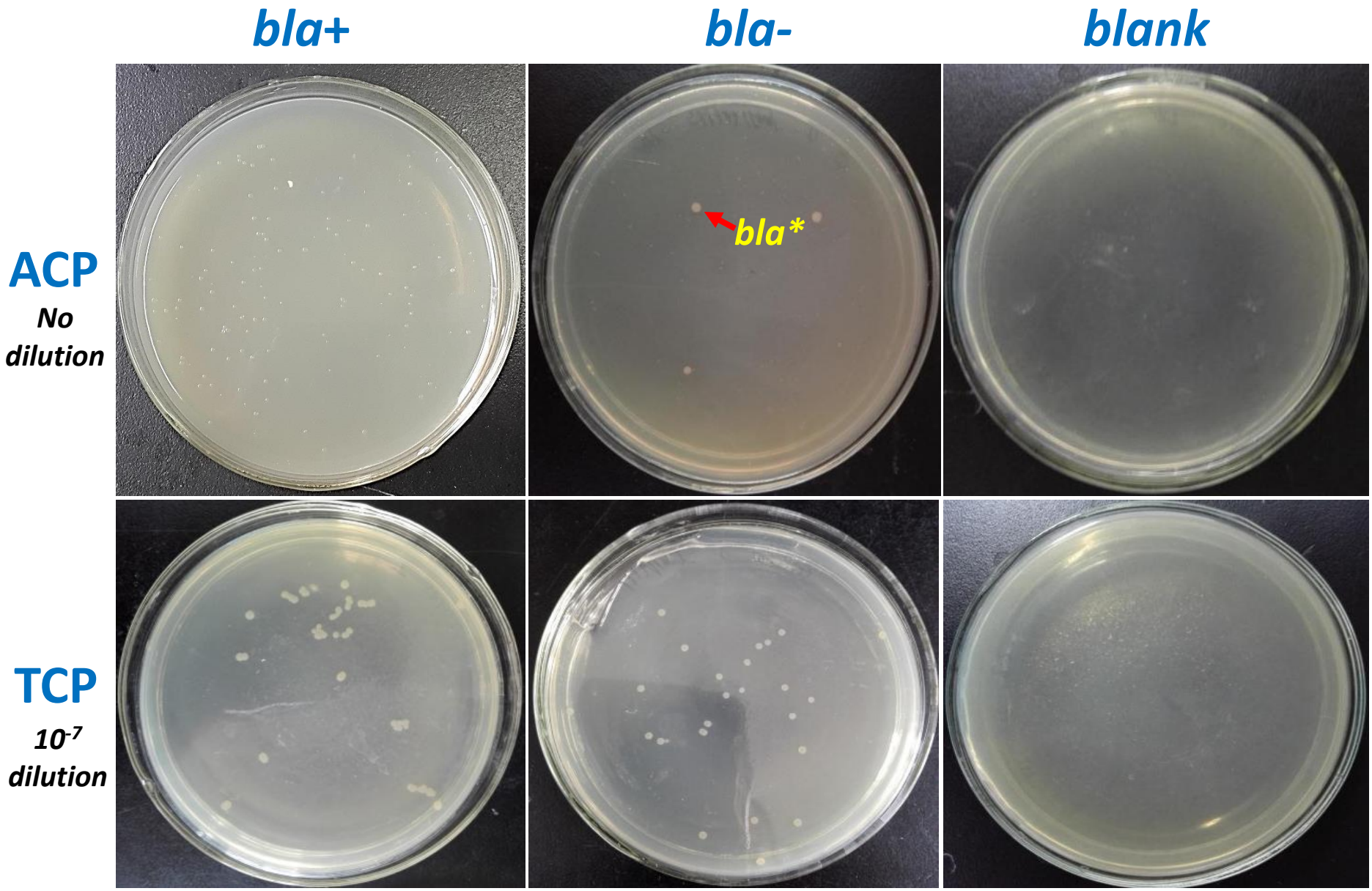
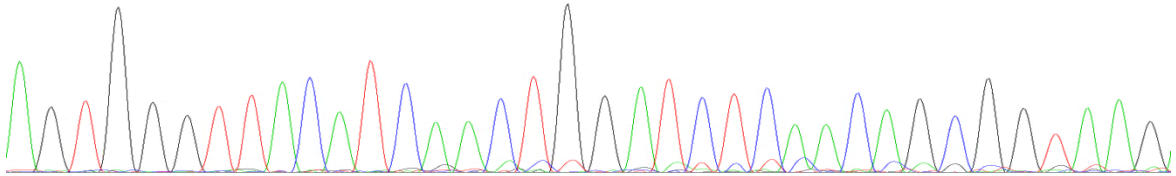


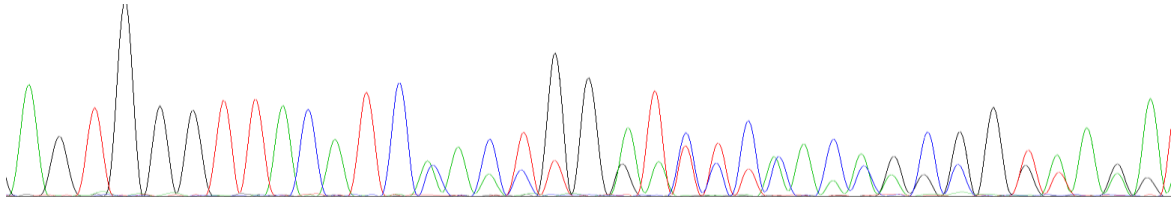
Fig 2B

Initial revertants (grow Slowly)



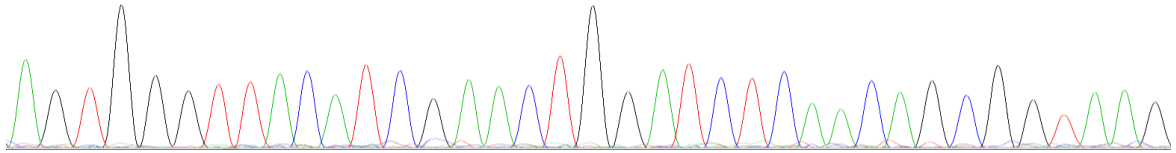
```
Frameshift: A G T G G G T T A C A T C A A C T G G A T C T C A A C A G C G G T A A G
            | | | | | | | | | | | | | \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \
Wild-type:  A G T G G G T T A C A T C G A A C T G G A T C T C A A C A G C G G T A A G
```

Later subculture (grow faster)



```
Frameshift: A G T G G G T T A C A T C A A C T G G A T C T C A A C A G C G G T A A G .
(main peak) | | | | | | | | | | | | | \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \ \
Revertants: A G T G G G T T A C A T C C A A C T G G A T C T C A A C A G C G G T A A G
(2nd peak)  | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Wild-type:  A G T G G G T T A C A T C G A A C T G G A T C T C A A C A G C G G T A A G
```

Final subculture (grow fast)



```
Revertants: A G T G G G T T A C A T C G A A C T G G A T C T C A A A A G C G G T A A G
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Wild-type:  A G T G G G T T A C A T C G A A C T G G A T C T C A A C A G C G G T A A G
```

Fig 2C

```
                          20                  40                  60
bla+   : ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA:  77
bla-   : ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA:  77
bla*A  : ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA:  77
bla*B  : ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA:  77
bla*C  : ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA:  77
bla*D  : ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA:  77
bla*E  : ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGA:  77

         80                  100                 120                 140
bla+   : AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGG-TTACATCGAACTGGAT-CTCAACAG: 152
bla-   : AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGG-TTACATC-AACTGGAT-CTCAACAG: 151
bla*A  : AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGG-TTACATC-AACTGGAT-CTCAACAA: 151
bla*B  : AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGGTACCTTC--ACTGGATTTCCAACAC: 152
bla*C  : AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGGTTACTTC--ACTGGATTCTCACCAC: 152
bla*D  : AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGG-TTACATC-AACTGGAT-CTCAACAG: 151
bla*E  : AACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGG-TTACATC-AACTGGAT-CTCAACAG: 151

           160                 180                 200                 220
bla+   : CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG: 229
bla-   : CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG: 228
bla*A  : CGGGAAAATCCTTGGAAGTTTTCCCCCCCAAAAACGGTTTCCAATGGTGAACACTTTTAAAGTTCTGGTATGGGGCG: 228
bla*B  : CGTAAAAATCCTGGAAATTTTCCCCCCAAAAAAC-TTTTCCAATGGAGAACACTTTTAAAGTTCCGGCTAGGGGCC: 228
bla*C  : CGGTAAAATCCTTGAAAGTTTTCCCCCCAAAAAACGTTTTCCAAGGATAACAACTTTTAAAGTTCGGCTATG-GGCC: 228
bla*D  : CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAACACTTTTAAAGTTCTGCTATGTGGCG: 228
bla*E  : CGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCG: 228
```

Fig 2D

Coding sequences

Protein sequences

Seq	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
bla+	atg	agt	att	caa	cat	ttc	cgt	gtc	gcc	ctt	att	ccc	ttt	ttt	gcg	gca	ttt	tgc	ctt	cct	gtt	ttt	gct	cac	cca	gaa	acg	ctg	gtg	aaa	gta	aaa	gat	gct	gaa	gat	cag	ttg	ggt	gca	cga	gtg	ggt	tac	atc	Gaa	ctg	gat
bla-	atg	agt	att	caa	cat	ttc	cgt	gtc	gcc	ctt	att	ccc	ttt	ttt	gcg	gca	ttt	tgc	ctt	cct	gtt	ttt	gct	cac	cca	gaa	acg	ctg	gtg	aaa	gta	aaa	gat	gct	gaa	gat	cag	ttg	ggt	gca	cga	gtg	ggt	tac	atc	aac	tgg	atc
bla*A	atg	agt	att	caa	cat	ttc	cgt	gtc	gcc	ctt	att	ccc	ttt	ttt	gcg	gca	ttt	tgc	ctt	cct	gtt	ttt	gct	cac	cca	gaa	acg	ctg	gtg	aaa	gta	aaa	gat	gct	gaa	gat	cag	ttg	ggt	gca	cga	gtg	ggt	tac	atc	aac	tgg	atc
bla*B	atg	agt	att	caa	cat	ttc	cgt	gtc	gcc	ctt	att	ccc	ttt	ttt	gcg	gca	ttt	tgc	ctt	cct	gtt	ttt	gct	cac	cca	gaa	acg	ctg	gtg	aaa	gta	aaa	gat	gct	gaa	gat	cag	ttg	ggt	gca	cga	gtg	ggg	tac	ctt	cac	tgg	att
bla*C	atg	agt	att	caa	cat	ttc	cgt	gtc	gcc	ctt	att	ccc	ttt	ttt	gcg	gca	ttt	tgc	ctt	cct	gtt	ttt	gct	cac	cca	gaa	acg	ctg	gtg	aaa	gta	aaa	gat	gct	gaa	gat	cag	ttg	ggt	gca	cga	gtg	ggg	tta	ctt	cac	tgg	att
bla*D	atg	agt	att	caa	cat	ttc	cgt	gtc	gcc	ctt	att	ccc	ttt	ttt	gcg	gca	ttt	tgc	ctt	cct	gtt	ttt	gct	cac	cca	gaa	acg	ctg	gtg	aaa	gta	aaa	gat	gct	gaa	gat	cag	ttg	ggt	gca	cga	gtg	ggt	tac	atc	aac	tgg	atc
bla*E	atg	agt	att	caa	cat	ttc	cgt	gtc	gcc	ctt	att	ccc	ttt	ttt	gcg	gca	ttt	tgc	ctt	cct	gtt	ttt	gct	cac	cca	gaa	acg	ctg	gtg	aaa	gta	aaa	gat	gct	gaa	gat	cag	ttg	ggt	gca	cga	gtg	ggt	tac	atc	aac	tgg	atc
bla+	M	S	I	Q	H	F	R	V	A	L	I	P	F	F	A	A	F	C	L	P	V	F	A	H	P	E	T	L	V	K	V	K	D	A	E	D	Q	L	G	A	R	V	G	Y	I	E	L	D
bla-	M	S	I	Q	H	F	R	V	A	L	I	P	F	F	A	A	F	C	L	P	V	F	A	H	P	E	T	L	V	K	V	K	D	A	E	D	Q	L	G	A	R	V	G	Y	I	N	W	I
bla*A	M	S	I	Q	H	F	R	V	A	L	I	P	F	F	A	A	F	C	L	P	V	F	A	H	P	E	T	L	V	K	V	K	D	A	E	D	Q	L	G	A	R	V	G	Y	I	N	W	I
bla*B	M	S	I	Q	H	F	R	V	A	L	I	P	F	F	A	A	F	C	L	P	V	F	A	H	P	E	T	L	V	K	V	K	D	A	E	D	Q	L	G	A	R	V	G	Y	L	H	W	I
bla*C	M	S	I	Q	H	F	R	V	A	L	I	P	F	F	A	A	F	C	L	P	V	F	A	H	P	E	T	L	V	K	V	K	D	A	E	D	Q	L	G	A	R	V	G	L	L	H	W	I
bla*D	M	S	I	Q	H	F	R	V	A	L	I	P	F	F	A	A	F	C	L	P	V	F	A	H	P	E	T	L	V	K	V	K	D	A	E	D	Q	L	G	A	R	V	G	Y	I	N	W	I
bla*E	M	S	I	Q	H	F	R	V	A	L	I	P	F	F	A	A	F	C	L	P	V	F	A	H	P	E	T	L	V	K	V	K	D	A	E	D	Q	L	G	A	R	V	G	Y	I	N	W	I

Coding sequences

Protein sequences

Seq	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87
bla+	ctc	aac	agc	ggt	aag	atc	ctt	gag	agt	ttt	cgc	ccc	gaa	gaa	cgt	ttt	cca	agt	ttt	cgc	ccc	gaa	gaa	cgt	ttt	cca	atg	atg	agc	act	ttt	aaa	gtt	ctg	cta	tgt	ggc	gcg	gta	tta	tcc	cgt	gtt	gac	gcc	ggg	caa	gag
bla-	tca	aca	gcg	gta	aga	tcc	ttg	aga	gtt	ttc	gcc	ccg	aag	aac	gtt	ttc	caa	gtt	ttc	gcc	ccg	aag	aac	gtt	ttc	caa	tga	tga	gca	ctt	tta	aag	ttc	tgc	tat	gtg	gcg	cgg	tat	tat	ccc	gtg	ttg	acg	ccg	ggc	aag	agc
bla*A	tca	aca	acg	gga	aaa	tcc	ttg	gaa	gtt	ttc	ccc	ccc	aaa	aac	ggt	ttc	caa	gtt	ttc	ccc	ccc	aaa	aac	ggt	ttc	caa	tgg	tga	aca	ctt	tta	aag	ttc	tgg	tat	ggg	gcg	cgg	tat	tat	ccc	gtg	gtg	acg	ccc	ggc	aag	aac
bla*B	tcc	aac	acc	gta	aaa	atc	ctg	gaa	att	ttc	ccc	ccc	aaa	aaa	ctt	ttc	caa	att	ttc	ccc	ccc	aaa	aaa	ctt	ttc	caa	tgg	aga	aca	ctt	tta	aag	ttc	cgg	cta	ggg	gcc	cgg	gaa	taa	ccc	ctg	gtg	aac	ccc	ggc	aaa	aag
bla*C	ctc	acc	acc	ggt	aaa	atc	ctt	gaa	agt	ttt	ccc	ccc	aaa	aaa	cgt	ttt	cca	agt	ttt	ccc	ccc	aaa	aaa	cgt	ttt	cca	agg	ata	aca	act	ttt	aaa	gtt	cgg	cta	tgg	gcc	cgg	tat	tta	tcc	cgg	gtt	gac	ccc	ggg	aaa	aac
bla*D	tca	aca	gcg	gta	aga	tcc	ttg	aga	gtt	ttc	gcc	ccg	aag	aac	gtt	ttc	caa	gtt	ttc	gcc	ccg	aag	aac	gtt	ttc	caa	tga	tga	aca	ctt	tta	aag	ttc	tgc	tat	gtg	gcg	cgg	tat	tat	ccc	gtg	ttg	acg	ccg	ggc	aag	aac
bla*E	tca	aca	gcg	gta	aga	tcc	ttg	aga	gtt	ttc	gcc	ccg	aag	aac	gtt	ttc	caa	gtt	ttc	gcc	ccg	aag	aac	gtt	ttc	caa	tga	tga	gca	ctt	tta	aag	ttc	tgc	tat	gtg	gcg	cgg	tat	tat	ccc	gtg	ttg	acg	ccg	ggc	aag	agc
bla+	L	N	S	G	K	I	L	E	S	F	R	P	E	E	R	F	P	S	F	R	P	E	E	R	F	P	M	M	S	T	F	K	V	L	L	C	G	A	V	L	S	R	V	D	A	G	Q	E
bla-	S	T	A	V	R	S	L	R	V	F	A	P	K	N	V	F	Q	V	F	A	P	K	N	V	F	Q	*	*	A	L	L	K	F	C	Y	V	A	R	Y	Y	P	V	L	T	P	G	K	S
bla*A	S	T	T	G	K	S	L	E	V	F	P	P	K	N	G	F	Q	V	F	P	P	K	N	G	F	Q	W	*	T	L	L	K	F	W	Y	G	A	R	Y	Y	P	V	V	T	P	G	K	N
bla*B	S	N	T	V	K	I	L	E	I	F	P	P	K	K	L	F	Q	I	F	P	P	K	K	L	F	Q	W	R	T	L	L	K	F	R	L	G	A	R	E	*	P	L	V	N	P	G	K	K
bla*C	L	T	T	G	K	I	L	E	S	F	P	P	K	K	R	F	P	S	F	P	P	K	K	R	F	P	R	I	T	T	F	K	V	R	L	W	A	R	Y	L	S	R	V	D	P	G	K	N
bla*D	S	T	A	V	R	S	L	R	V	F	A	P	K	N	V	F	Q	V	F	A	P	K	N	V	F	Q	*	*	T	L	L	K	F	C	Y	V	A	R	Y	Y	P	V	L	T	P	G	K	N
bla*E	S	T	A	V	R	S	L	R	V	F	A	P	K	N	V	F	Q	V	F	A	P	K	N	V	F	Q	*	*	A	L	L	K	F	C	Y	V	A	R	Y	Y	P	V	L	T	P	G	K	S

Coding sequences

Protein sequences

Seq	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
bla+	caa	ctc	ggt	cgc	cgc	ata	cac	tat	tct	cag	aat	gac	ttg	gtt	gag	tac	tca	cca	gtc	aca	gaa	aag	cat	ctt	acg	gat	ggc	atg	aca	gta	aga	gaa	tta	tgc	agt	gct	gcc	ata	acc	atg	agt	gat	aac	act	gcg	gcc	aac	tta
bla-	aac	tcg	gtc	gcc	gca	tac	act	att	ctc	aga	atg	act	tgg	ttg	agt	act	cac	cag	tca	cag	aaa	agc	atc	tta	cgg	atg	gca	tga	cag	taa	gag	aat	tat	gca	gtg	ctg	cca	taa	cca	tga	gtg	ata	aca	ctg	cgg	cca	act	tac
bla*A	aac	tcc	gtc	gcc	gca	tac	act	att	ctc	aaa	atg	act	tgg	ttg	aat	act	ccc	caa	tca	cca	aaa	aac	atc	cta	ccg	atg	gca	tgg	cag	taa	gag	aat	tat	gca	atg	gtg	cca	tta	cca	tgg	atg	ata	acc	ctt	ccg	cca	act	tac
bla*B	aac	tcc	gtc	ggc	gca	taa	cac	aat	cct	caa	att	gct	tgg	ttg	gat	aat	cca	cca	tca	caa	aaa	aac	att	cta	acg	atg	gca	tgg	caa	taa	gaa	gaa	tta	gcc	gtg	gtg	gcc	tta	cca	aga	atg	gat	acc	ctc	cgc	caa	cct	tac
bla*C	aaa	ctc	ggc	ggc	cgc	tta	ccc	aat	cct	caa	aag	gac	tgg	gtg	aag	aat	tca	cca	ttc	aca	aaa	aac	cac	ctt	acg	aat	gca	tgg	aca	gta	gaa	gaa	tta	tgc	atg	gcg	ccc	ata	ccc	tgg	atg	gat	aac	acg	gcg	gcc	aac	tta
bla*D	aac	tcg	gtc	gcc	gca	tac	act	att	ctc	aga	atg	act	tgg	ttg	aat	act	cac	cag	tca	cag	aaa	agc	atc	tta	cgg	atg	gca	tga	cag	taa	gag	aat	tat	gca	gtg	ctg	cca	taa	cca	tga	gtg	ata	aca	ctg	cgg	cca	act	tac
bla*E	aac	tcg	gtc	gcc	gca	tac	act	att	ctc	aga	atg	act	tgg	ttg	agt	act	cac	cag	tca	cag	aaa	agc	atc	tta	cgg	atg	gca	tga	cag	taa	gag	aat	tat	gca	gtg	ctg	cca	taa	cca	tga	gtg	ata	aca	ctg	cgg	cca	act	tac
bla+	Q	L	G	R	R	I	H	Y	S	Q	N	D	L	V	E	Y	S	P	V	T	E	K	H	L	T	D	G	M	T	V	R	E	L	C	S	A	A	I	T	M	S	D	N	T	A	A	N	L
bla-	N	S	V	A	A	Y	T	I	L	R	M	T	W	L	S	T	H	Q	S	Q	K	S	I	L	R	M	A	*	Q	*	E	N	Y	A	V	L	P	*	P	*	V	I	T	L	R	P	T	Y
bla*A	N	S	V	A	A	Y	T	I	L	K	M	T	W	L	N	T	P	Q	S	P	K	N	I	L	P	M	A	W	Q	*	E	N	Y	A	M	V	P	L	P	W	M	I	T	L	P	P	T	Y
bla*B	N	S	V	G	A	*	H	N	P	Q	I	A	W	L	D	N	P	P	S	Q	K	N	I	L	T	M	A	W	Q	*	E	E	L	A	V	V	A	L	P	R	M	D	T	L	R	Q	P	Y
bla*C	K	L	G	G	R	L	P	N	P	Q	K	D	W	V	K	N	S	P	F	T	K	N	H	L	T	N	A	W	T	V	E	E	L	C	M	A	P	I	P	W	M	D	N	T	A	A	N	L
bla*D	N	S	V	A	A	Y	T	I	L	R	M	T	W	L	N	T	H	Q	S	Q	K	S	I	L	R	M	A	*	Q	*	E	N	Y	A	V	L	P	*	P	*	V	I	T	L	R	P	T	Y
bla*E	N	S	V	A	A	Y	T	I	L	R	M	T	W	L	S	T	H	Q	S	Q	K	S	I	L	R	M	A	*	Q	*	E	N	Y	A	V	L	P	*	P	*	V	I	T	L	R	P	T	Y

Fig 3 *Random mutagenesis*

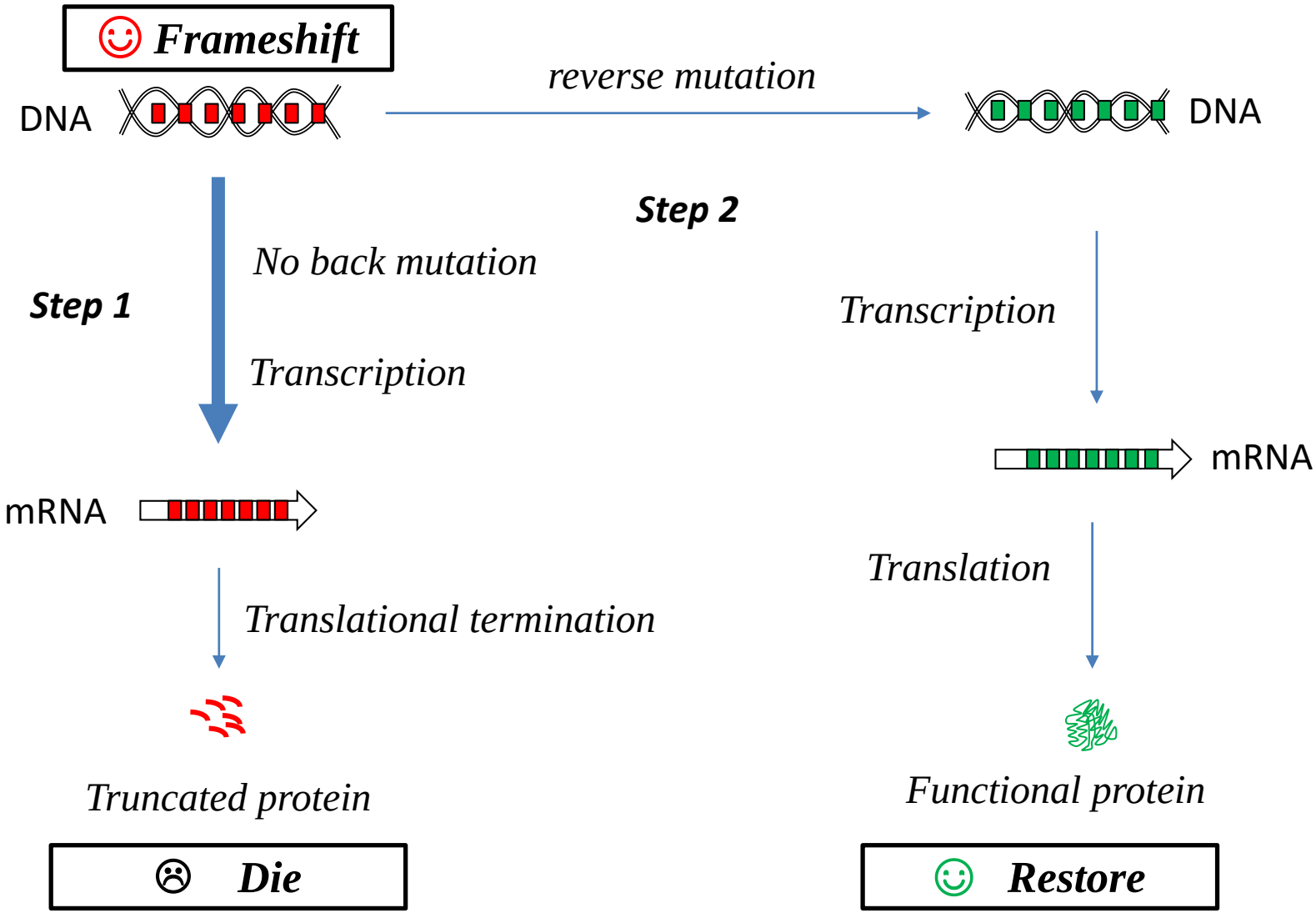


Fig 4A

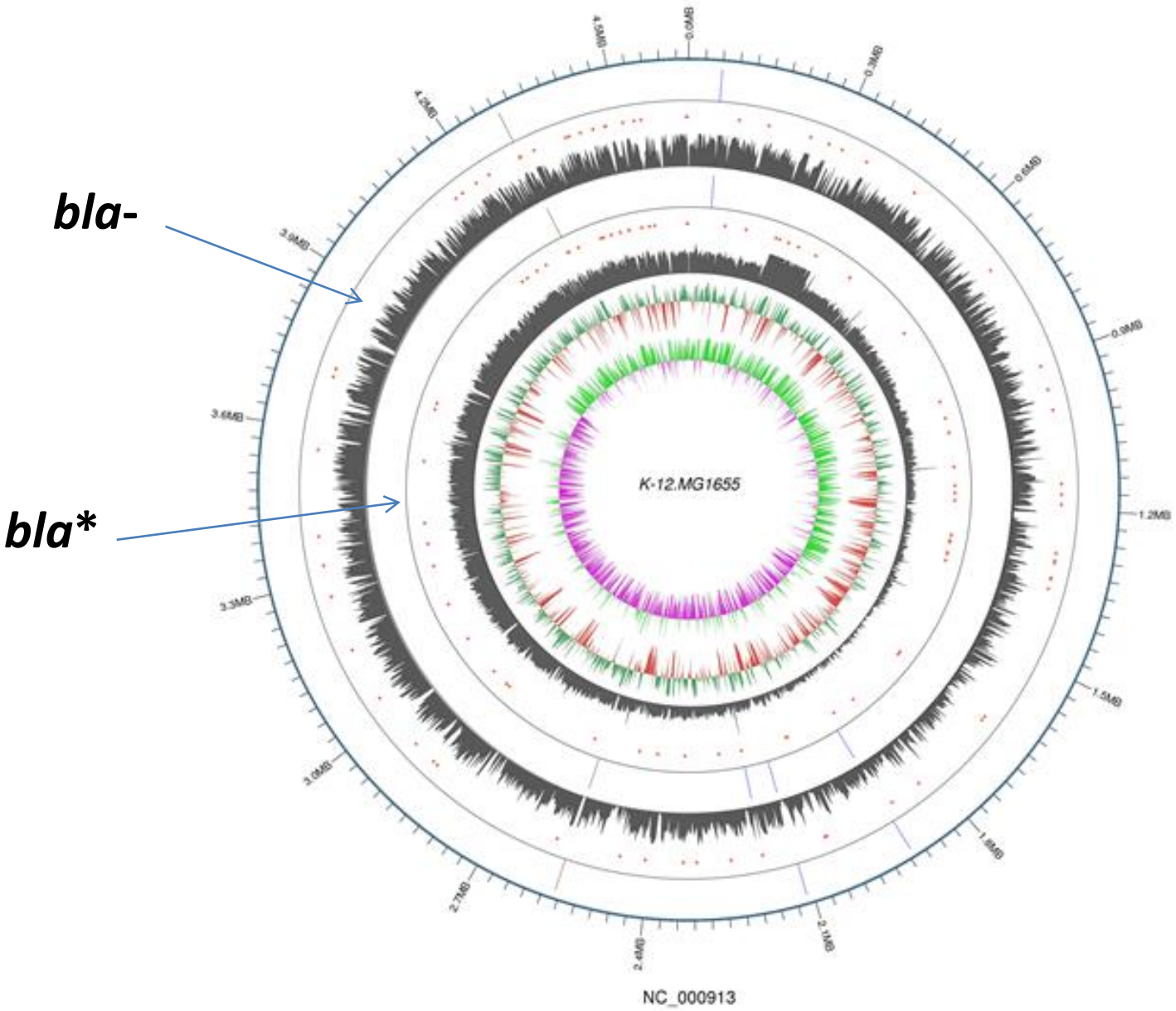
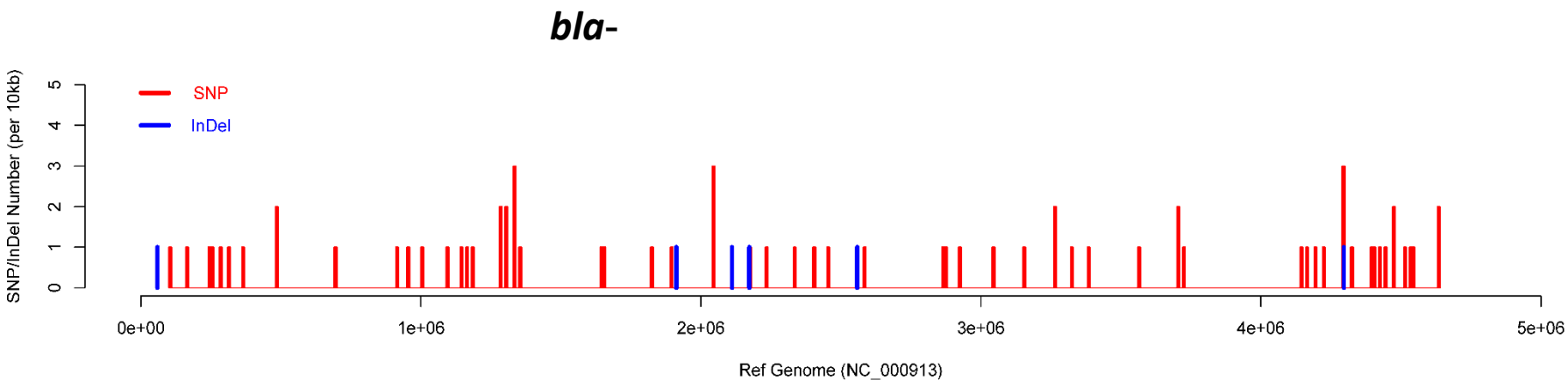


Fig 4B


bla-
SNP
InDel
SNP/InDel Number (per 10kb)
0
1
2
3
4
5
0e+00
1e+06
2e+06
3e+06
4e+06
5e+06
Ref Genome (NC_000913)

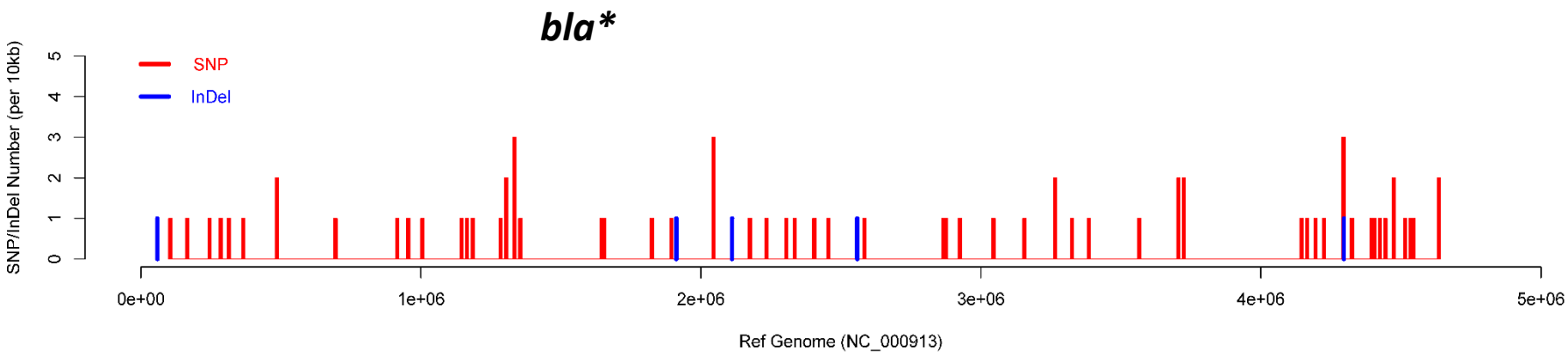

bla*
SNP
InDel
SNP/InDel Number (per 10kb)
0
1
2
3
4
5
0e+00
1e+06
2e+06
3e+06
4e+06
5e+06
Ref Genome (NC_000913)

Fig 4C

bla-

bla*

bla+ **transcripts**

bla* **transcripts**

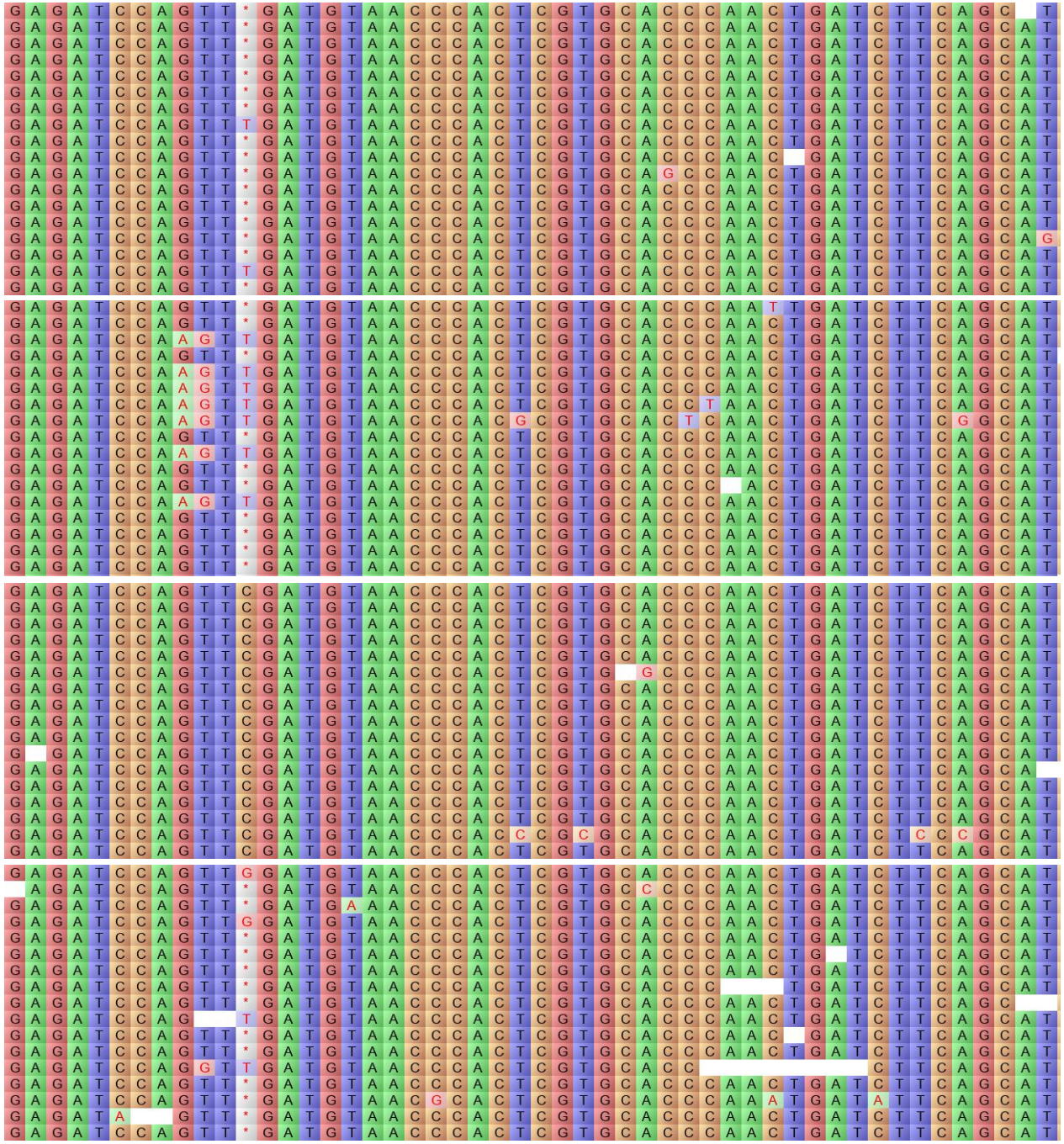


Fig 4D

***tet* CDS**
bla-

***tet* CDS**
bla*

***tet* transcripts**
bla-

***tet* transcripts**
bla*

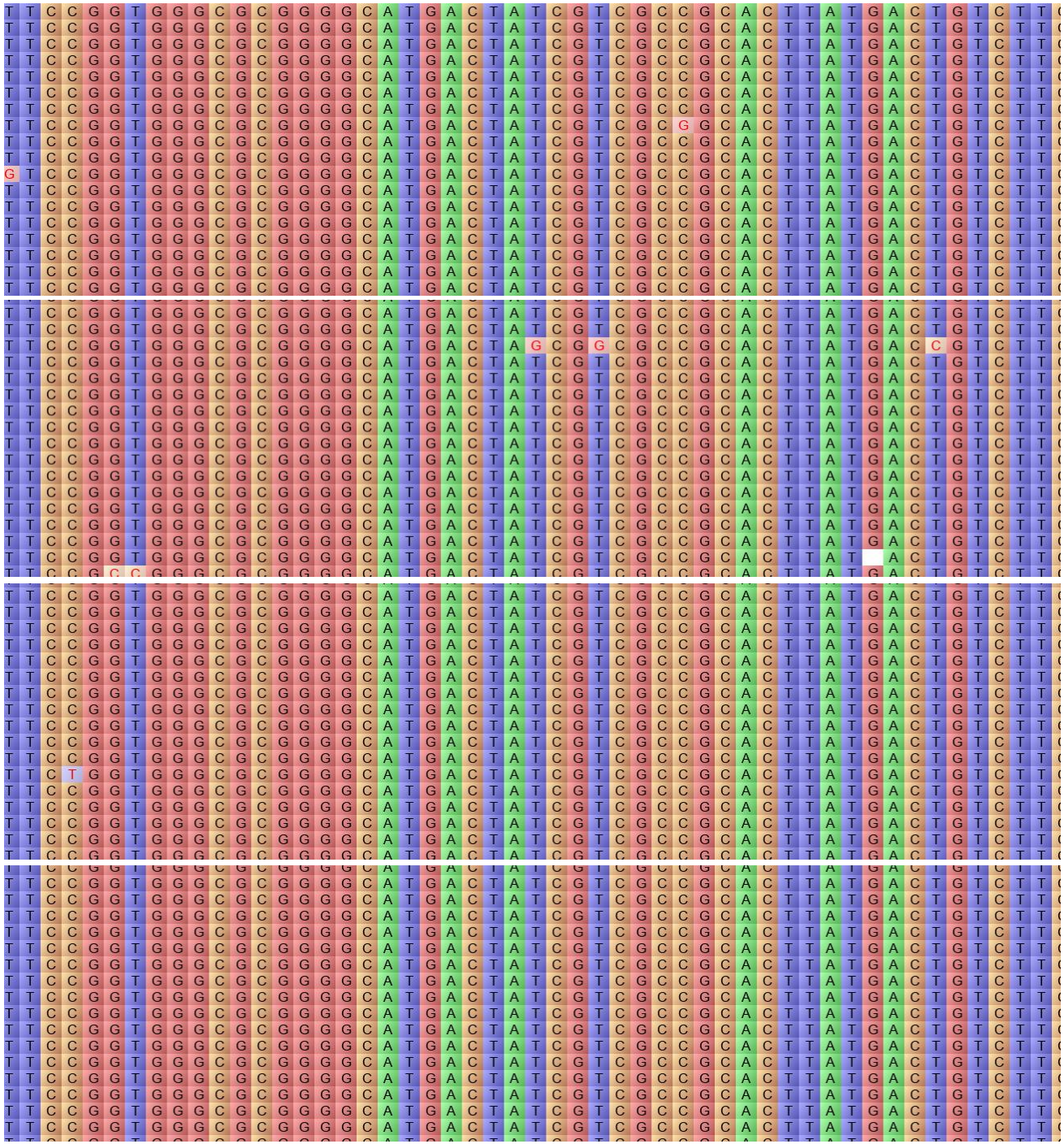


Fig 5. *Expression of the PTC-free frameshifted bla (bla#)*

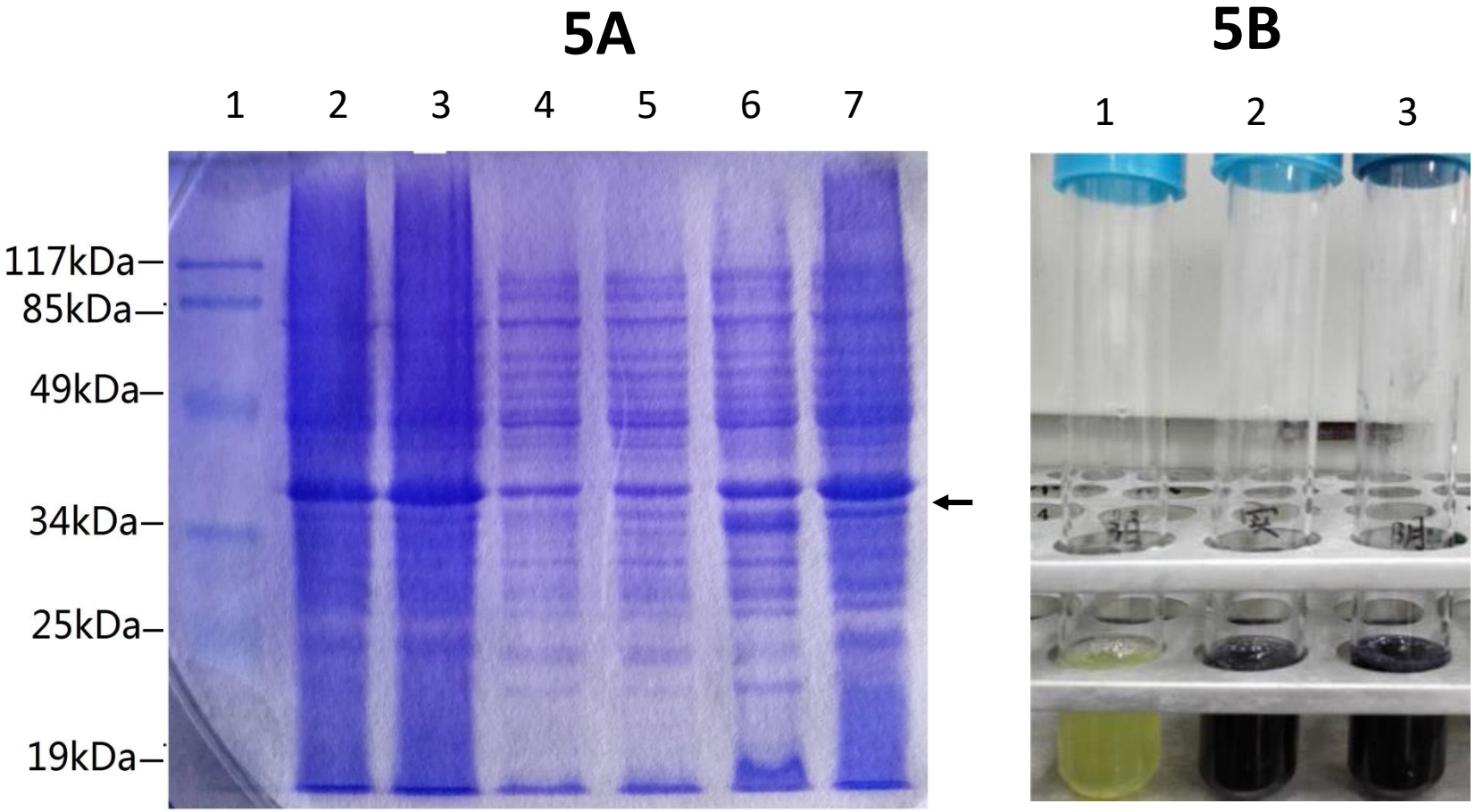
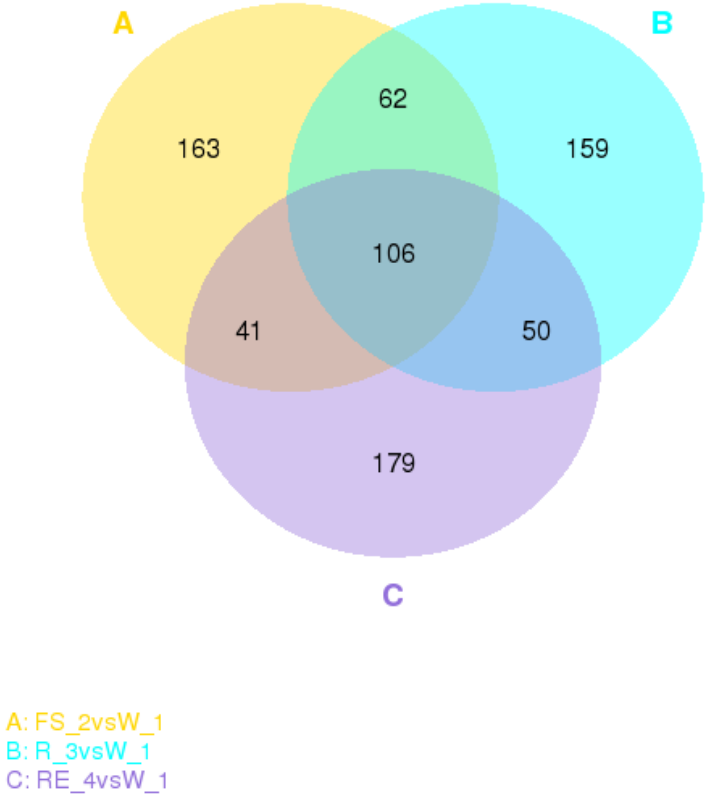
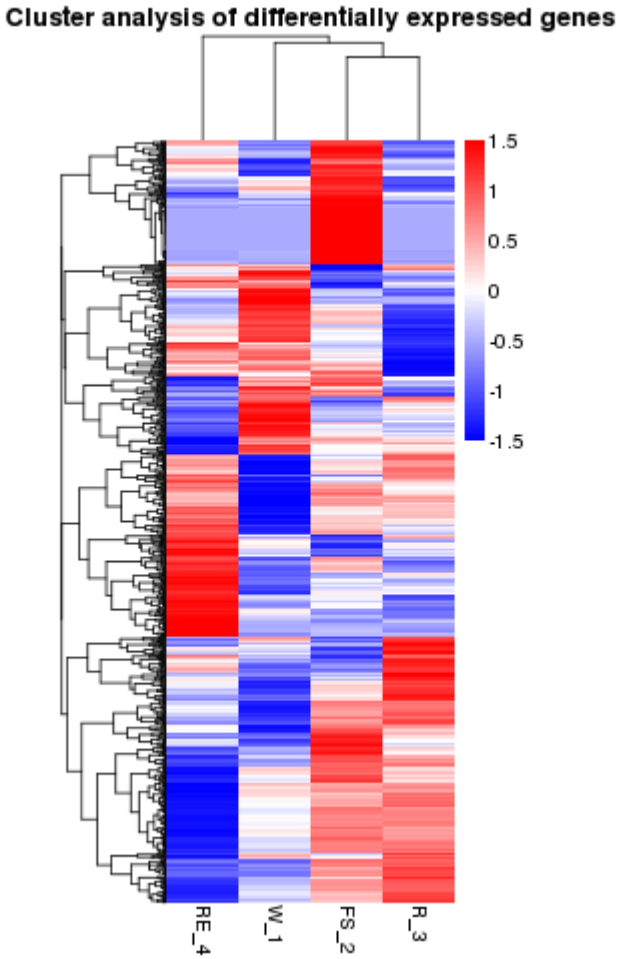


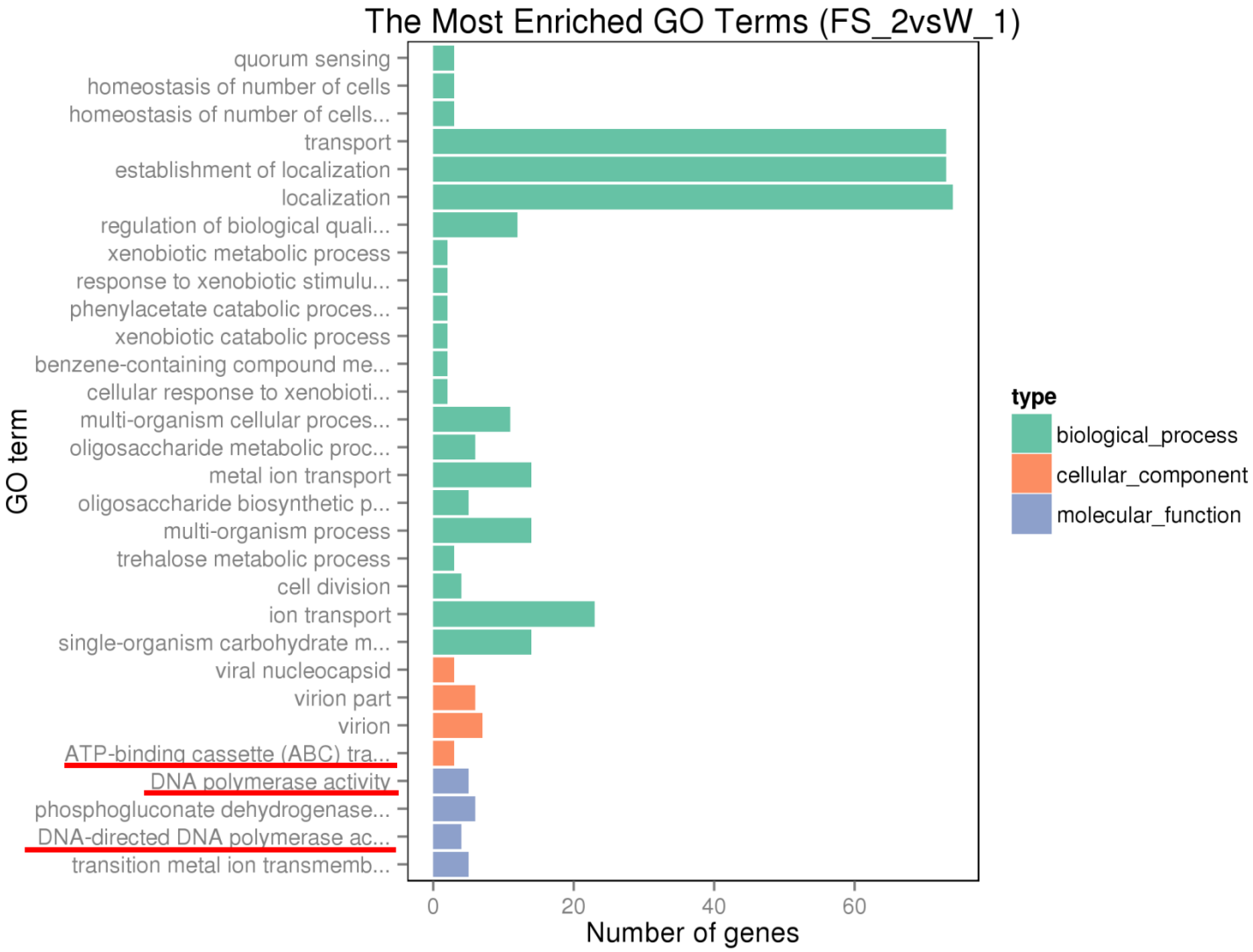
Fig 6. Transcriptome Analysis

6A Numbers of differentially expressed genes (DGE)

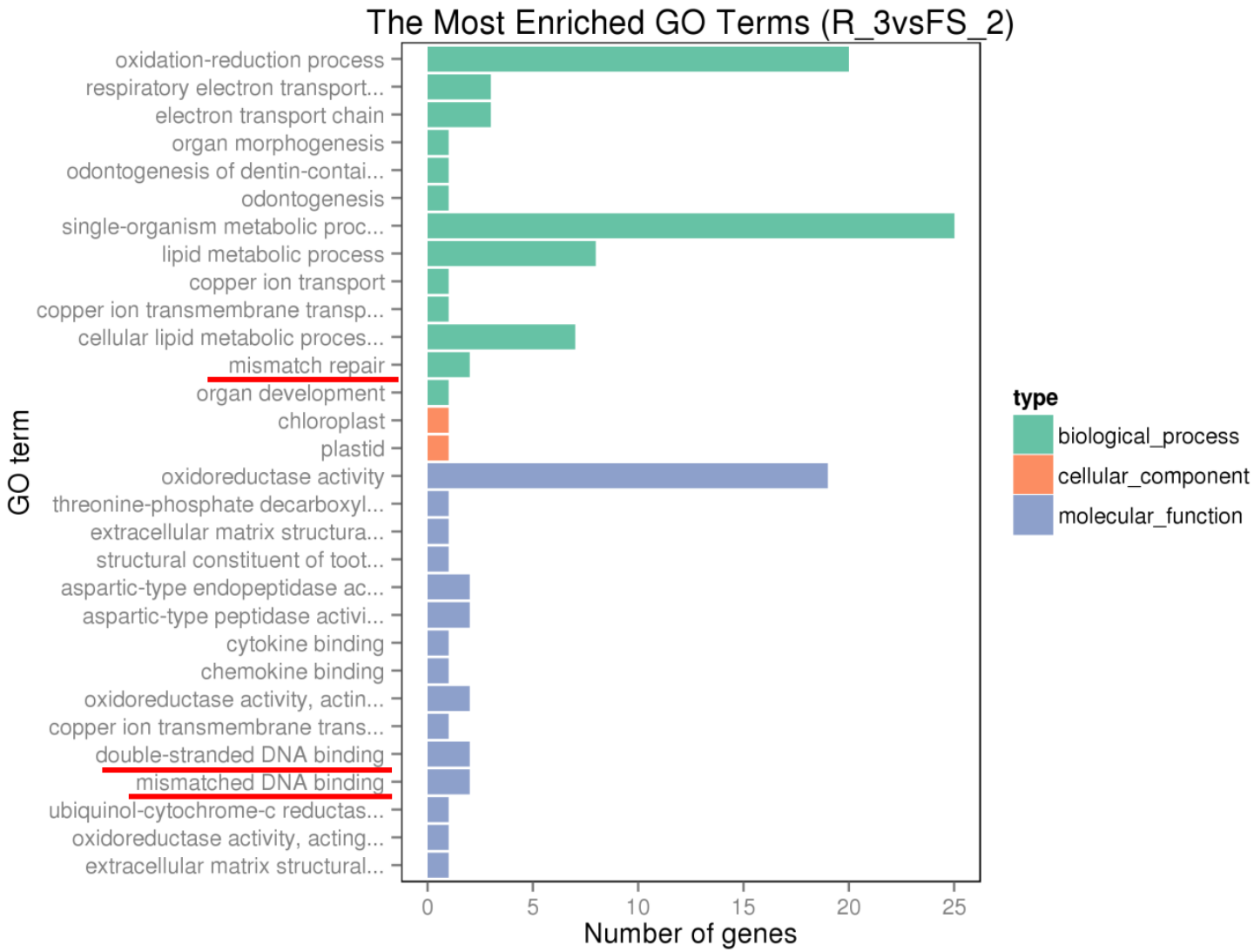




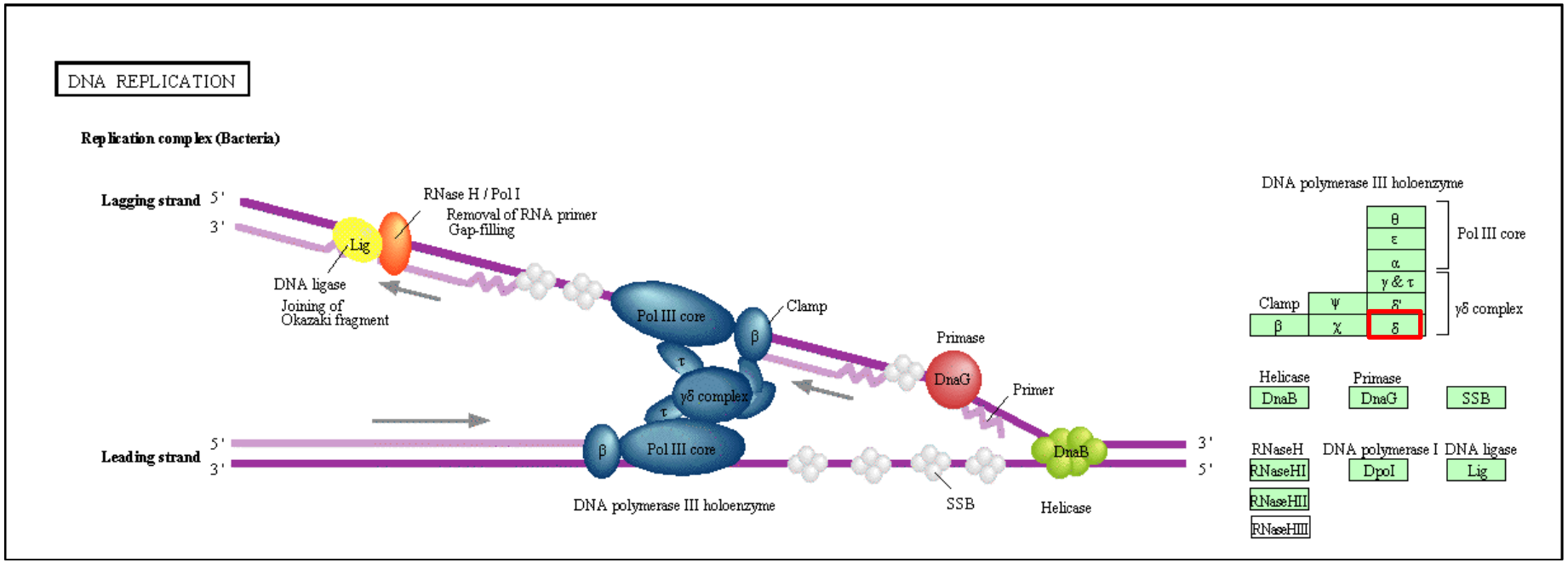
6B


The Most Enriched GO Terms (FS_2vsW_1)
GO term
quorum sensing
homeostasis of number of cells
homeostasis of number of cells...
transport
establishment of localization
localization
regulation of biological quali...
xenobiotic metabolic process
response to xenobiotic stimulu...
phenylacetate catabolic proces...
xenobiotic catabolic process
benzene-containing compound me...
cellular response to xenobioti...
multi-organism cellular proces...
oligosaccharide metabolic proc...
metal ion transport
oligosaccharide biosynthetic p...
multi-organism process
trehalose metabolic process
cell division
ion transport
single-organism carbohydrate m...
viral nucleocapsid
virion part
virion
ATP-binding cassette (ABC) tra...
DNA polymerase activity
phosphogluconate dehydrogenase...
DNA-directed DNA polymerase ac...
transition metal ion transmemb...
0
20
40
60
Number of genes
type
biological_process
cellular_component
molecular_function

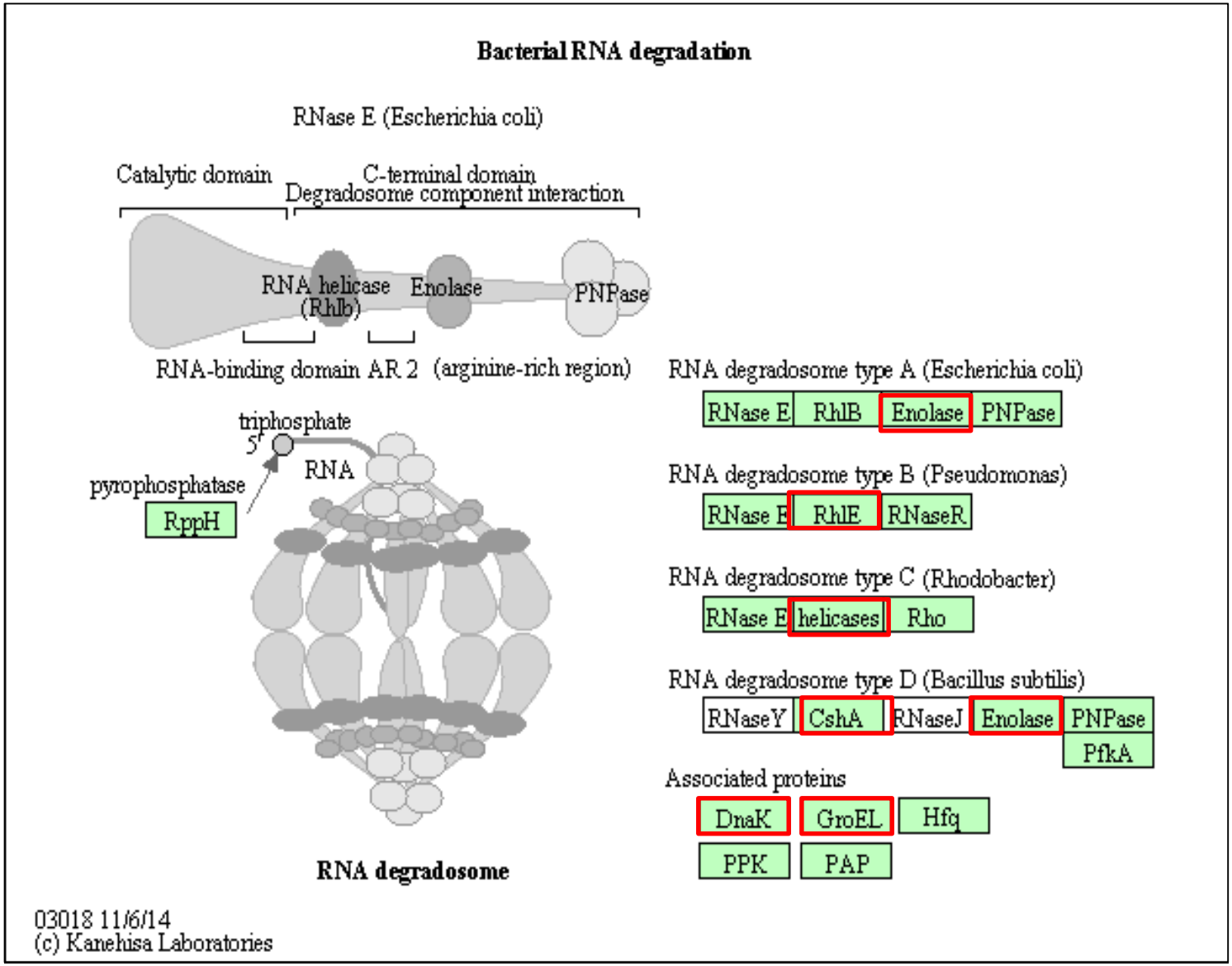
6C


The Most Enriched GO Terms (R_3vsFS_2)
oxidation-reduction process
respiratory electron transport...
electron transport chain
organ morphogenesis
odontogenesis of dentin-contai...
odontogenesis
single-organism metabolic proc...
lipid metabolic process
copper ion transport
copper ion transmembrane transp...
cellular lipid metabolic proces...
mismatch repair
organ development
chloroplast
plastid
oxidoreductase activity
threonine-phosphate decarboxyl...
extracellular matrix structura...
structural constituent of toot...
aspartic-type endopeptidase ac...
aspartic-type peptidase activi...
cytokine binding
chemokine binding
oxidoreductase activity, actin...
copper ion transmembrane trans...
double-stranded DNA binding
mismatched DNA binding
ubiquinol-cytochrome-c reductas...
oxidoreductase activity, acting...
extracellular matrix structural...
GO term
0
5
10
15
20
25
Number of genes
type
biological_process
cellular_component
molecular_function

6D genes upregulated in DNA replication



6E genes upregulated in RNA degradation



6F genes upregulated in mismatch repair/homologous recombination

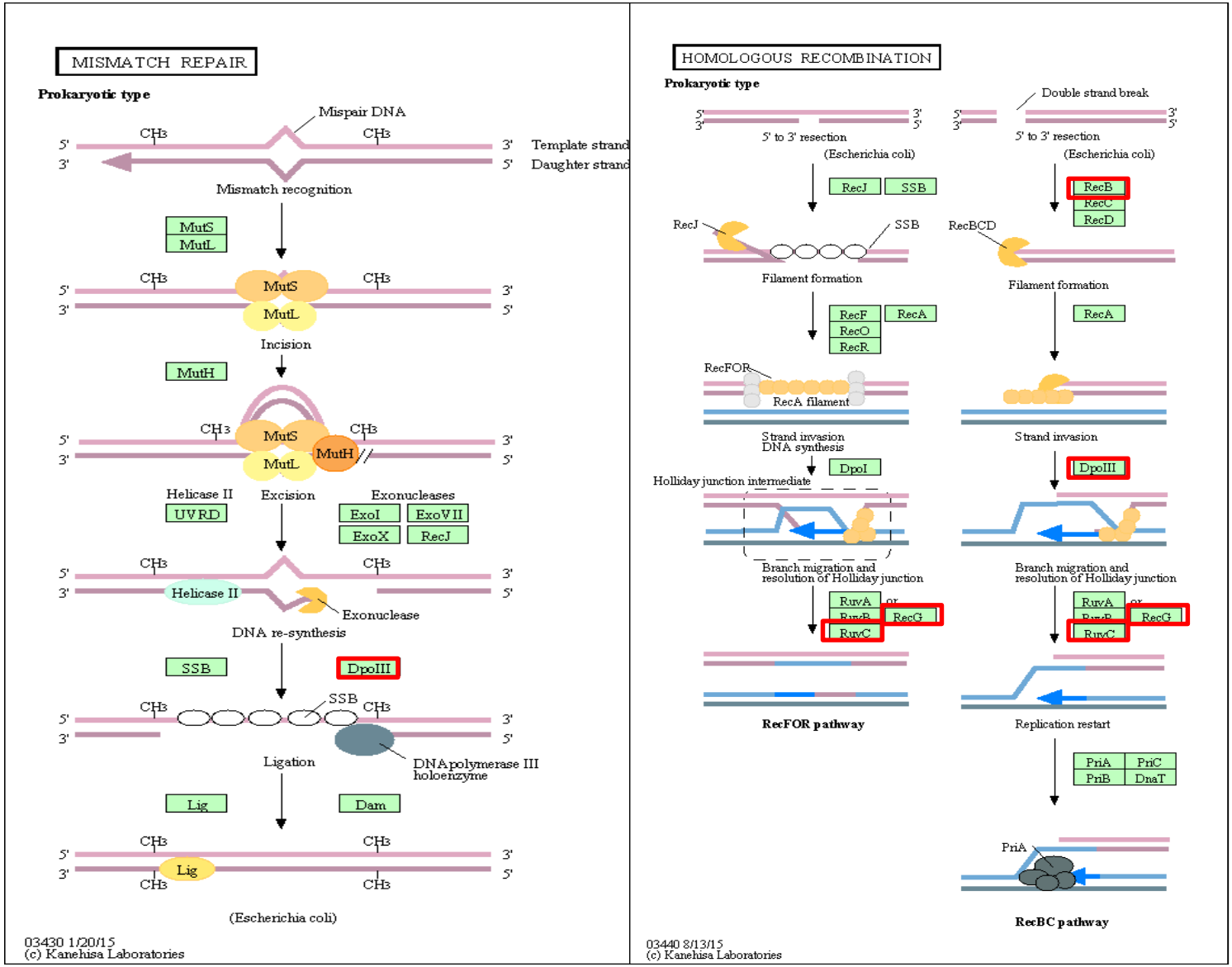
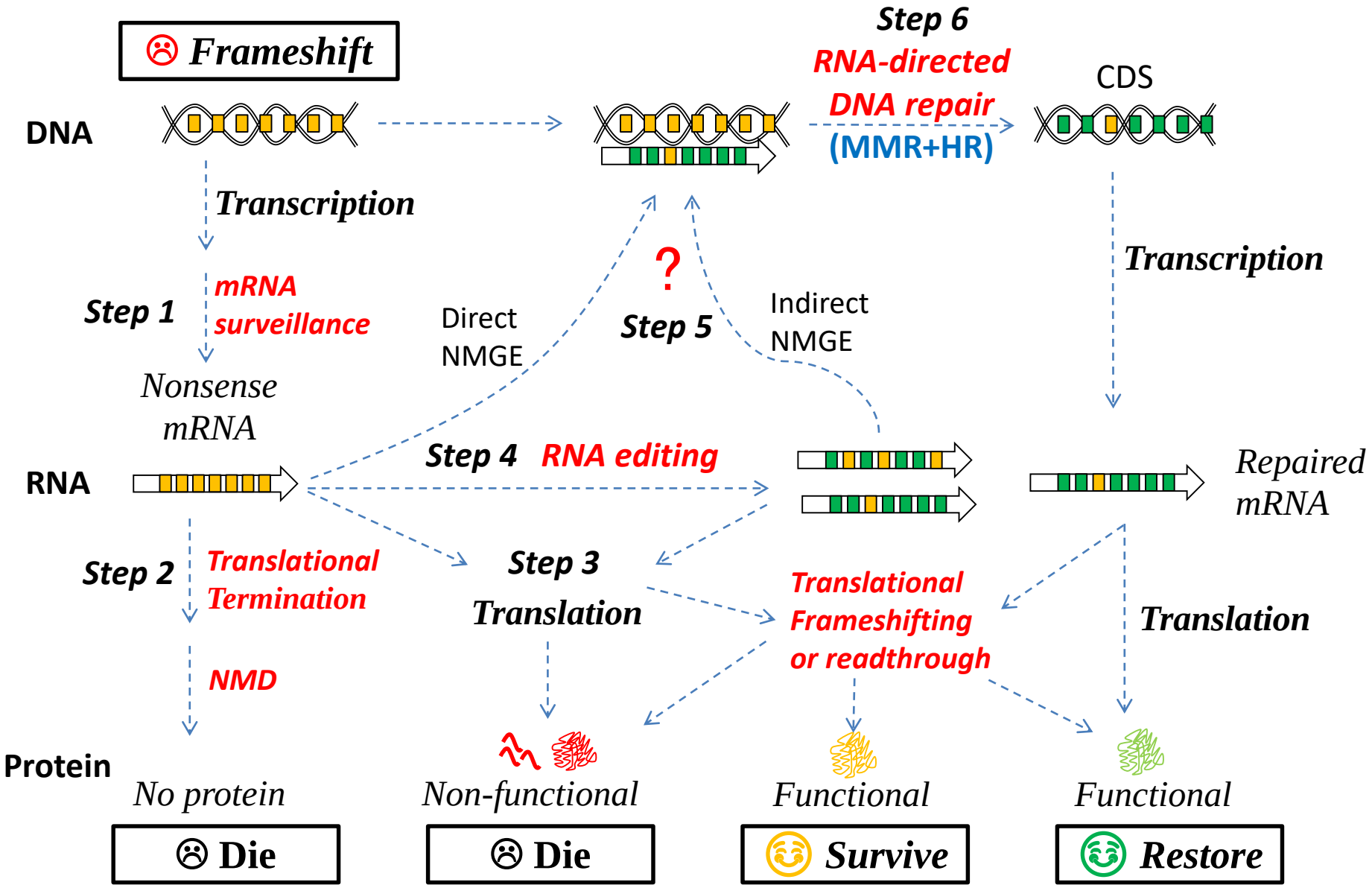



Fig 7 *Nonsense-mediated gene editing (NMGE)*