Biotechnology Reports 32 (2021) e00692

Contents lists available at ScienceDirect

Biotechnology Reports

journal homepage: www.elsevier.com/locate/btre

# A genetic toolkit for co-expression of multiple proteins of diverse physiological implication

Ali Samy Abdelaal [a,c], Syed Shams Yazdani [a,b,*]

[a] Microbial Engineering Group, International Centre for Genetic Engineering and Biotechnology, New Delhi, India
[b] DBT-ICGEB Centre for Advanced Bioenergy Research, International Centre for Genetic Engineering and Biotechnology, New Delhi, India
[c] Department of Genetics, Faculty of Agriculture, Damietta University, Damietta, Egypt

ARTICLE INFO

*Keywords:*
Plasmid construction
Co-expression
Pathway engineering
Fine-tuning gene expression
Fluorescent proteins
FACS

ABSTRACT

Construction of plasmids is crucial for expression of functional proteins of diverse physiological impact in *E. coli*. Here, we first designed and constructed a novel pair of bacterial expression vectors, i.e., pAS01 and pAS02, to be co-transformed with pQE30 for the co-expression of three target genes. The three plasmids contain ColE1, p15A and pSC101 origin of replication for high, medium and low copy plasmids, respectively, and same promoter (T5) and RBS. We then cloned genes encoding three reporter proteins (GFPuv, TurboRFP, and EYFP) in each of these plasmids and co-expressed in *E. coli* in six different combinations. Each of these reporter proteins exhibited diverse impact on growth, plasmid copy number and stability, and expression of other reporter proteins. Our results indicate that GFP and RFP were the most and the least favorable proteins for the cells, respectively, in terms of these parameters, especially on impacting expression of other co-expressed proteins.

## 1. Introduction

Nearly 50 years have passed since the first production of recombinant DNA molecules [1] and construction of bacterial plasmids *in vitro* [2]. However, there is still a shortage of manufacturing capacity due to the increasing demand of diverse recombinant proteins, which requires a much better approach in the optimization of production strategies and expression vector design. Efficient strategies for the development of recombinant proteins are becoming particularly important as more applications that demand significant amounts of high-quality proteins enter the industry. Moreover, expressing multiple proteins of diverse nature is the major thrust area to fulfill the goal of metabolic engineering and synthetic biology. Plasmids are extrachromosomal highly folded circular self-replicating cytoplasmic DNA elements present in most bacteria. Since the advent of genetic engineering, they have been used as molecular vehicles for recombinant genes. Plasmid-based expression is the most common option when expressing genes in prokaryotes [3].

Plasmids are commonly found in variable copy numbers in bacteria and they inflict a metabolic burden on the host cells, as they utilize the cellular resources for their replication and recombinant protein production. This metabolic burden correlates with plasmid copy number and toxicity of the inducers toward the host, which influence the bacterial growth rate as well as its products [4]. A common observation is that low- and medium-copy plasmids provide much more benefits over high-copy plasmids, such as tighter control of gene expression, the ability to replicate larger genes, and lower metabolic burden on the host cells [5]. Additionally, use of high-copy plasmid to enhance gene expression may rather lead to limitation in gene expression due to plasmid metabolic burden. Moreover, protein expression at very high levels might be harmful and, in this case, low- or medium-copy plasmids may be sufficient or even desirable [6]. For the production of periplasmatically secreted proteins, it was found out that the expression rate should be optimized rather than maximized [7].

There are many factors related to the expression vector that influence the recombinant protein production in *E. coli*. For example, adequate origin of replication for the desired plasmid copy number, regulation of transcription for adequate protein expression, promoter with an ease of induction, and translational features at the 5′ end of the mRNA transcript including ribose binding site (RBS) and the sequence between the RBS sequence and the start codon [4, 8, 9].

Reporter genes play a key role in gene expression studies. They are most commonly used as indicators of translational activity within cells. The cells are assayed for presence of the reporter by directly measuring the amount of the reporter protein. Fluorescent proteins are considered

* Corresponding author.
*E-mail address:* shams@icgeb.res.in (S.S. Yazdani).

https://doi.org/10.1016/j.btre.2021.e00692
Received 28 June 2021; Received in revised form 22 November 2021; Accepted 29 November 2021
Available online 3 December 2021

as ideal reporters which do not require substrates [10]. The green fluorescent protein (GFPuv) has a single excitation peak centered at about 395 nm, with an emission peak at 509 nm, while turbo red fluorescent protein (TurboRFP) has its excitation peak at around 553 nm, with an emission peak at 574 nm. The enhanced yellow fluorescent protein (EYFP) generally has an excitation peak at around 513 nm and an emission peak near 527 nm. These three fluorescent proteins are very commonly used in the laboratory studies for various applications, however, a relative efficiency of their expression and co-expression have not yet been performed.

Various approaches have been employed for the co-expression of multiple proteins in bacteria, especially when performing metabolic or pathway engineering. These include the use of operons, dual promoter systems, bidirectional promoters (BDPs), or co-transformation of more than one vector [11–14]. Each of these is associated with a number of limitations such as uneven or unreliable protein expression levels, silencing of some promoters, increased toxicity to cells when using multiple inducers, or availability/generalizability of BDPs.

The main objective of our study was to understand the behavior of multiple proteins of different physiological implications on the bacterial cells; i.e., the plasmid copy number, growth, and co-expression of other recombinant proteins, when expressed individually or together. To achieve this objective, we created a new vector system that can accept three genes/operons to be co-expressed at different expression levels for various studies. This system consists of three plasmids, each carrying the gene of interest and a different selection marker. The plasmids also have different compatible origin of replication where different cloned genes can be induced with a single inducer. To highlight the effectiveness of the toolkit, co-expression of the different florescent proteins have been demonstrated at various expression levels.

## 2. Materials and methods

### 2.1. Strain and plasmid construction

Strains, plasmids, and primers for PCR used in this study have been listed in Table 1. Plasmid construction and DNA manipulations were performed using standard molecular cloning protocols. Plasmid maintenance and propagation were performed in *E. coli* DH5*α* strain.

The vector pQE30 was purchased from Qiagen, while pZA31MCS and pZS21MCS were purchased from Expressys and used as the backbone plasmids. To create pAS01 and pAS02 plasmids, the DNA sequence containing PLtetO-1 promoter with RBS and partial MCS from pZA31MCS and pZS21MCS plasmids were replaced with the DNA sequence containing T5 promoter, RBS, His-tag, and MCS from pQE30 using *Aat*II and *Hind*III restriction enzymes. The construction of pAS01 and pAS02 has been described in detail in Fig. 1 and their sequences have been deposited to NCBI repository with GenBank accession numbers OL542782 and OL542783, respectively.

### 2.2. Cloning and transformation of genes encoding fluorescent protein

Genes encoding traffic light fluorescent proteins GFPuv, TurboRFP, and EYFP were amplified from their corresponding vectors pSS9, pTurboRFP-N, and pICH47742:PtFCP:Cas9YFP, respectively, and each gene was separately cloned into pQE30, pAS01, and pAS02 vectors to generate all possible nine combinations. All expressed proteins had 6-histidine tag at their N-terminus for detection using Western blot.

The *GFPuv* coding sequence was amplified from the pSS9 vector as template using GFPuv-F/GFPuv-R primers and inserted into the pQE30, pAS01, and pAS02 vectors to obtain pQE30_GFPuv, pAS01_GFPuv and pAS02_GFPuv, respectively. The cloning of *TurboRFP* and EYFP genes was performed in the similar way. The *TurboRFP* gene was amplified from the pTurboRFP-N vector as template using TurboRFP-F/TurboRFP-R primers and cloned in the three plasmids to obtain pQE30_turboRFP, pAS01_turboRFP and pAS02_turboRFP, respectively. The *EYFP* gene was amplified from the pICH47742:PtFCP:Cas9YFP vector as template using EYFP-F/EYFP-R primers and cloned in the three plasmids to obtain pQE30_EYFP, pAS01_EYFP and pAS02_EYFP, respectively.

**Table 1**
Strains, plasmids and primers used in this study.

| Name | Description | Reference |
|---|---|---|
| ***E. coli* Strain** | | |
| DH5α | *E. coli* K12 Φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 | Invitrogen |
| GRY | DH5α carrying pQE30_GFPuv, pAS01_turboRFP and pAS02_EYFP | This study |
| GYR | DH5α carrying pQE30_GFPuv, pAS01_EYFP and pSA02_turboRFP | This study |
| RGY | DH5α carrying pQE30_turboRFP, pAS01_GFPuv and pAS02_EYFP | This study |
| RYG | DH5α carrying pQE30_turboRFP, pAS01_EYFP and pAS02_GFPuv | This study |
| YGR | DH5α carrying pQE30_EYFP, pAS01_GFPuv and pAS02_turboRFP | This study |
| YRG | DH5α carrying pQE30_EYFP, pAS01_turboRFP and pAS02_GFPuv | This study |
| **Plasmids** | | |
| pQE30 | ColE1 ori T5 Lac O $Amp^r$ | Invitrogen |
| pZA31 MCS | p15A ori PLtetO-1 $Cm^r$ | Expressys |
| pZS21 MCS | pSC101 ori PLtetO-1 $Kan^r$ | Expressys |
| pAS01 | p15A ori T5 Lac O $Cm^r$ | This study |
| pAS02 | pSC101 ori T5 Lac O $Kan^r$ | This study |
| pSS9 | pBR322 backbone harboring GFPuv | Addgene |
| pTurboRFP-N | Mamalian expression vector harboring TurboRFP | Evrogen |
| pICH47742: PtFCP:Cas9YFP | pICH47742 backbone harboring EYFP | Addgene |
| pQE30_GFPuv | pQE30 harboring GFPuv | This study |
| pQE30_turboRFP | pQE30 harboring turboRFP | This study |
| pQE30_EYFP | pQE30 harboring EYFP | This study |
| pAS01_GFPuv | pAS01 harboring GFPuv | This study |
| pAS01_turboRFP | pAS01 harboring turboRFP | This study |
| pAS01_EYFP | pAS01 harboring EYFP | This study |
| pAS02_GFPuv | pAS02 harboring GFPuv | This study |
| pAS02_turboRFP | pAS02 harboring turboRFP | This study |
| pAS02_EYFP | pAS02 harboring EYFP | This study |
| **Primers** | | |
| GFP_F | TAGGTACCATGAGTAAAGGAGAAGAACTTTTC | This study |
| GFP_R | TATAAGCTTCTATTTGTATAGTTCATCCATGCC | This study |
| RFP_F | TAGGTACCATGAGCGAGCTGATCAAGGAG | This study |
| RFP_R | TATAAGCTTTCATCTGTGCCCCAGTTTGC | This study |
| EYFP_F | TAGGTACCATGGTGAGCAAGGGCGAGG | This study |
| EYFP_R | TATAAGCTTTCACTTGTACAGCTCGTCCATGC | This study |

Note: the enzyme sites are underlined.

All clones were screened by restriction digestion analysis and verified by gene sequencing. Competent DH5a cells were transformed with either single construct or a combination of three constructs. Six different triples of plasmids generated six different strains, i.e. GRY, GYR, RGY, RYG, YGR, and YRG. The first, second and third position in the triplet denotes high (pQE30), medium (pAS01), and low (pAS02) copy number plasmids, while letter 'G', 'R' and 'Y' denotes *GFPuv, TurboRFP* and *EYFP* gene, respectively.

### 2.3. Sample preparation

Samples were prepared for all the analysis as follows. The bacterial strains were grown in 5 ml LB medium (pH ~7) with the appropriate antibiotics at 37 °C and 180 rpm in an incubator shaker (Eppendorf) and were induced using 0.5 mM IPTG at $OD_{600}$= 0.6. Samples were collected after 4 h of induction for all analysis, except for real-time PCR where samples were collected after 2 h of induction.

### 2.4. Quantitative measurement of plasmid copy number

The plasmid copy number for each expressing clone was determined using real-time PCR as described earlier [15]. The sample for real-time PCR was prepared using the iQ™ SYBR Green Supermix (Bio-Rad)

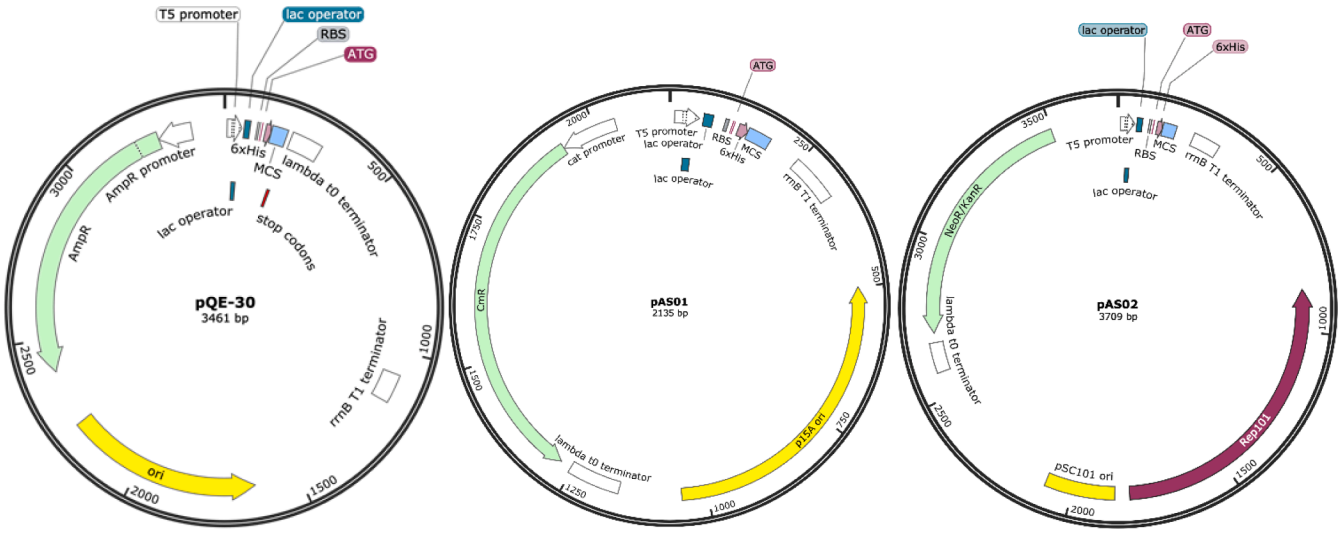


**Fig. 1.** Map of the pQE30, pAS01 and pAS02 vectors used in the study.

following the manufacturer's instructions and the PCR was performed on the iCycler Thermal Cycler (Bio-Rad) using the SYBR-490 filter for fluorescence detection. The experiment was performed in biological triplicate for analysis of plasmid copy number.

Plasmid copy number relative to the chromosome was calculated using the following equation: Plasmid copy number $= 2^{(\text{Ct chromosome} - \text{Ct plasmid})}$.

### 2.5. Standard plate count method

To obtain the population of cells bearing each of the plasmid, colony-forming units (CFU) were obtained against each antibiotic as follows. A volume of culture was aliquoted from each sample when $OD_{600}$ reached 0.4 and serial $10^5$-fold dilutions were made in LB medium and spread on 4 LB-agar plates, 3 plates having 3 different antibiotics, i.e., ampicillin, chloramphenicol and kanamycin, and 1 control plate having no antibiotic. Colonies were counted after 14 h of incubation at 37 °C.

### 2.6. Gene expression analysis

Gene expression analysis was done by real-time PCR as follows. Total RNA was isolated from each sample and the expression levels of transcripts of all florescent genes in different strains were determined using primers listed in Table 1. *rrsA* (16S ribosomal RNA) gene was used for normalizing RNA expression.

Recombinant protein production was analyzed using Western blot. Total protein was isolated from each sample and separated using SDS-PAGE. Proteins were then transferred to nitrocellullose membrane, followed by immunodetection of proteins using monoclonal anti-poly histidine−peroxidase antibody from Merck.

### 2.7. Microscopy and image generation

Samples were pelleted and properly resuspended in 1X Phosphate buffer saline (PBS), placed on glass slide under coverslip, and scanned under microscope. A series of images were digitized with a confocal laser scanning microscope (A1r from Nikon) using different filters. The images were analyzed using Nikon NIS-Elements Advanced Research.

### 2.8. Data analysis

Data were analyzed in Microsoft Excel program and are presented as mean and standard deviations (SD) based on triplicates from at least three independent experiments.

Two-way analysis of variance (ANOVA) was used to compare the effects of strains and fluorescent proteins in qPCR and FACS. The least significant difference (LSD) values were used to compare means ($p \leq 0.05$ and 0.01). The Pearson Correlation was used to measure the strength and direction of linear relationships between pairs of continuous variables. These data were analyzed using IBM SPSS Statistics, version 20.

### 2.9. Fluorescence-activated cell sorting (FACS) analysis

For FACS study, cells were harvested 4 h post-induction, pelleted and resuspended in PBS. The cells were then analyzed on a BD FACSAria III for quantitative assessment of GFPuv (405 nm laser for excitation, 450/40 filter for detection), TurboRFP (561 nm laser for excitation, 610/20 filter for detection), and EYFP (488 nm laser for excitation, 530/30 filter for detection) fluorescent-producing cell population using an appropriate compensation matrix. Final data, given as log CFU/mL, resulted from at least three independent experiments with three replicates each. Data were analyzed using FlowJo software.

## 3. Results and discussion

### 3.1. Design and features of pAS01 and pAS02 vectors

For obtaining high yield of a recombinant protein, often high-level expression of the gene is needed using high copy number plasmid. But it has also been observed that high copy plasmid is counterproductive for some proteins. Further, in metabolic engineering, some enzymes need to be produced at lower level for optimizing the stoichiometry of a pathway. In general, the host cells suffer from high expression of recombinant proteins since it drains out cellular energy, thereby disrupting essential host cell functions. Besides, the protein of interest can be toxic to the cell. Our toolkit has been designed for use in metabolic engineering experiment for stoichiometry optimization of expression and co-expression of multiple proteins.

The criteria for an effective *E. coli* co-expression vector include the existence of a strong promoter, a compatible origin of replication, a rich multiple cloning site, and a suitable antibiotic resistance. In order to achieve this goal, we constructed two vectors, pAS01 with chloramphenicol resistance and pAS02 with kanamycin resistance, to be used for co-expression with commercial vectors bearing either ampicillin or any different resistance marker.

The new vectors were constructed in the following manner. The pAS01 was obtained by combining portion of the pZA31MCS vector and

the pQE30 vector together. The resultant vector, pAS01 (2135 bp), carries p15A replication origin and the chloramphenicol resistance gene (CmR) of pZA31MCS and T5 promoter, synthetic ribosomal binding site RBSII and 6xHis-tag coding sequence of pQE30 (Fig. 1). pAS02 vector was constructed by ligating portion of the vector pZS21MCS and the vector pQE30 together. The recombinant vector, pAS02 (3709 bp), carries pSC101 replication origin and the kanamycin resistance gene (KanR) of pZS21MCS and T5 promoter, synthetic ribosomal binding site RBSII and 6xHis-tag coding sequence of pQE30 (Fig. 1).

The three vectors used in our further study, i.e., pQE30, pAS01, and pAS02, had the ability to co-exist with each other due to different origin of replications. At the same time, all the vectors contained same promoter and RBS to ensure that differences in the expression of the cloned genes is only due to differences in plasmid copy number and not due to differences in the promoter or RBS strength. This technique also avoids use of multiple inducers that might have adverse impact on the growth and physiology of the host.

### 3.2. *Generation of recombinant plasmids for expression and co-expression of traffic light fluorescent reporter proteins*

To study the usefulness of the modified plasmids and the toolkit, we considered genes encoding GFPuv, EYFP, and TurboRFP proteins that emit green, yellow and red fluorescence, respectively, and have minimal overlap in their excitation/emission wavelengths. The three genes were separately cloned into each of the three plasmids under study to get nine different constructs in order to express each gene in different expression vector with different expression level (Fig. 2). All constructs were confirmed using PCR and DNA sequence analysis. These nine constructs were subsequently employed for co-transformation in different

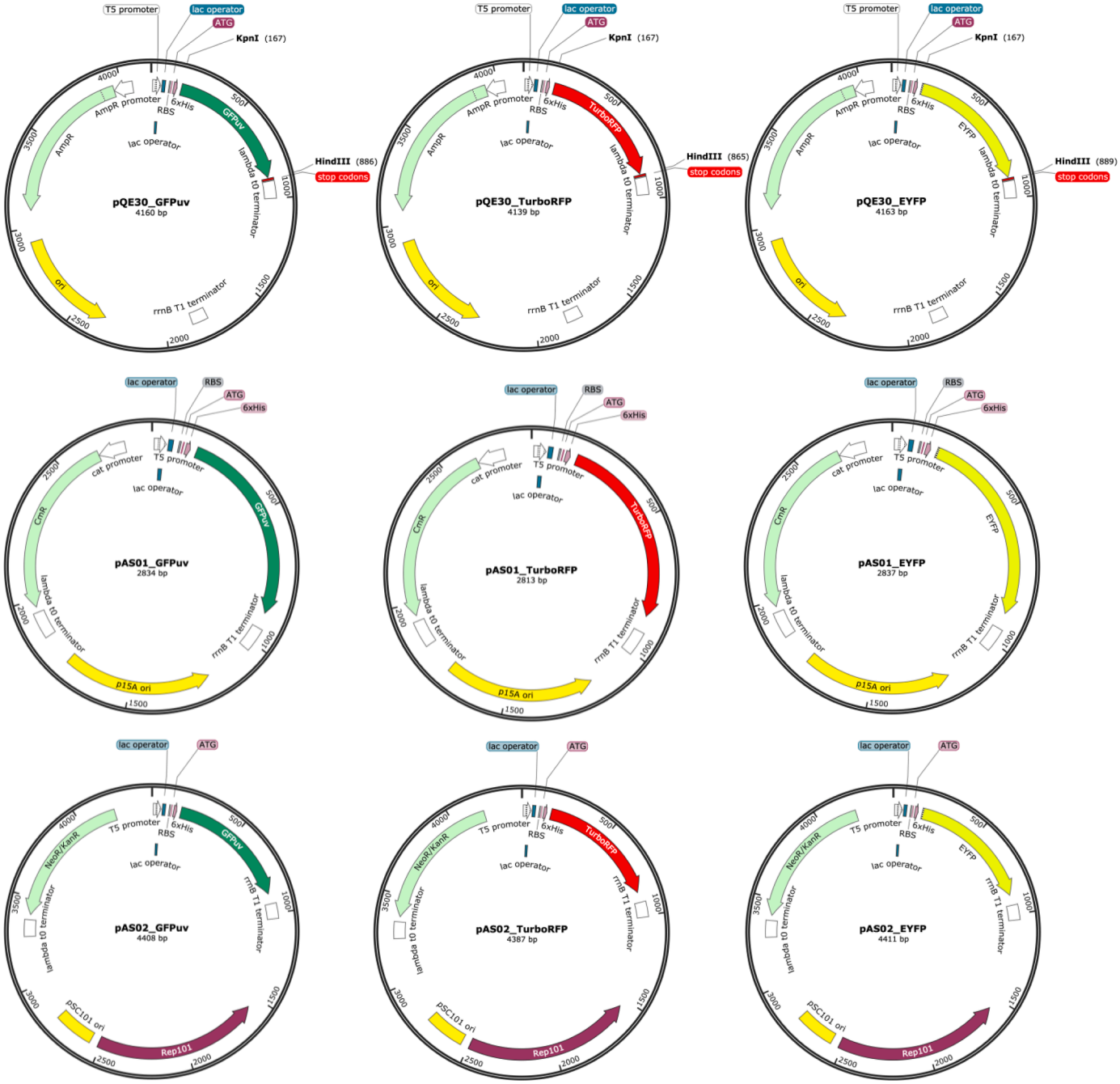


**Fig. 2.** Cloning of genes for fluorescent proteins GFPuv, TurboRFP and EYFP in three different plasmids, pQE30, pASA01 and pASA02.

combinations.

### 3.3. Impact of gene type on plasmid copy number

Plasmid copy numbers are likely to vary when genes coding for different kind of proteins are cloned and expressed. Using three traffic light reporter genes as a basis to determine the variation in the plasmid copy number, we measured the plasmid copy number using real-time qPCR method when the cells were at their early induction phase, i.e., 2 hrs post-induction. qPCR amplifications of the genes from nine strains of *E. coli* were performed using *GFPuv, TurboRFP*, and *EYFP* as plasmid genes, and *rrsA* as chromosomal gene. Relative quantification between each plasmid gene and chromosomal gene were calculated, as shown in Fig. 3A. As expected, the copy number of pQE30-based plasmid was highest in the cases where *GFPuv* and *EYFP* genes were cloned. However, the same was not true for *TurboRFP* gene where the copy number pQE30-based plasmid was much lower, close to the pAS01-based plasmid. Based on the type of the gene cloned, it appears that cells decided to control the plasmid copy number, with highest for GFP (17 copies), followed by 10 copies for YFP, and lowest (2.8 copies) for RFP. The medium copy pAS01-based plasmid remained at 2–3 copies for GFP and RFP, while it reduced to <1 copy for YFP. The low copy pAS02-based plasmid was maximum for GFP (3 copies), while it reduced to 1 copy for RFP and <1 copy for YFP. It became evident from these results that the concept of high, medium and low copy numbers of plasmid based on the types of origin of replications is not absolute and can change depending upon the kind of heterologous genes that it carries. It is also known from the previous studies that plasmid copy remains significantly lower in logarithmic phase of growth as compared to the stationary phase [16, 17], reflecting overall lower plasmid copy number in our study.

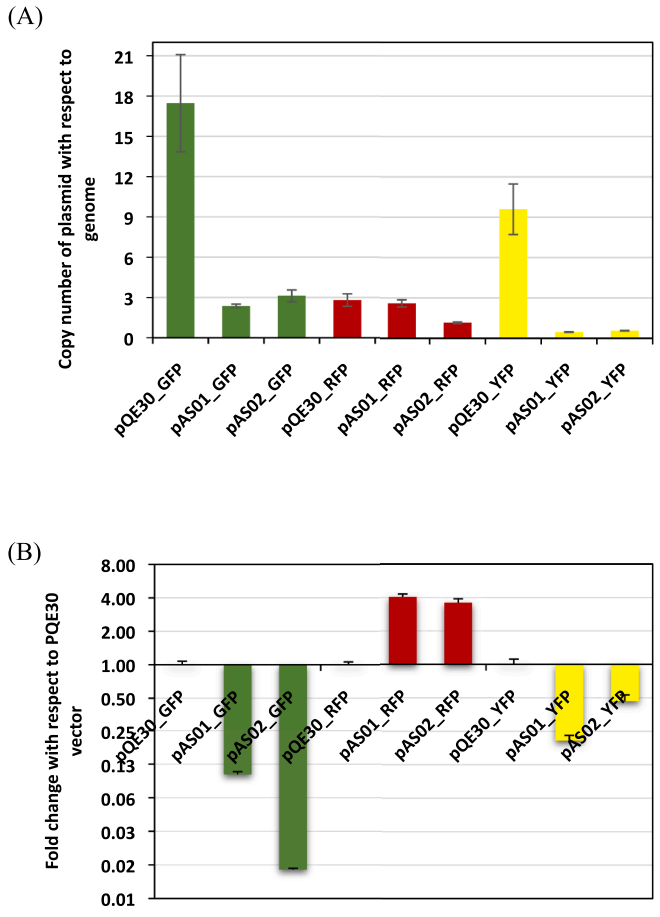


**Fig. 3.** (A) Plasmid copy numbers and (B) Gene expression detection using Real Time qPCR. Error bars represent the standard deviation ($n = 4$).

### 3.4. Profiling of gene expression upon replication origin swapping

Differences in the copy number of the plasmids because of their unique origin of replication will have implications to the expression of the gene [18]. One would like to believe that increasing the copy number of a plasmid would lead to increased gene dosage [4, 19]. We have seen in the previous section that plasmid copy number may not be uniform for all the genes. Here, we tested how the expression of the three traffic light reporter genes varied with the change in origin of replication.

The expression levels of the target genes (*GFPuv, TurboRFP*, and *EYFP*) were normalized with respect to that obtained using pQE30. When the expression levels of the *GFPuv* gene in the three constructs were compared, it was found that pAS01-based and pAS02-based plasmid expressed 10-fold less and 20-fold less GFP transcript as compared to pQE30-based plasmid (Fig. 3B). This kind of transcript profile was expected as per the origin of replication these plasmids were bearing. However, same was not true for RFP expression where the pAS01-based and pAS02-based plasmid exhibited ~4-fold higher expression than pQE30-based plasmid (Fig. 3B). This clearly indicated that high copy number plasmid was not suitable for RFP expression, possibly due to its adverse impact on the growth and physiology of the cells. YFP expression exhibited a middle trend where pQE30-based plasmid showed highest expression and other two plasmids showed only 2–5-fold lower expression. These expression patterns of the transcript was also found to be reflected at the protein level when the samples from induced culture were separated on SDS-PAGE and detected on Western-blot (Supplementary Figure S1).

We also assessed the florescence emission ability of each transformants under confocal scanning microscope. The number of cells fluorescing in the GFP transformants as well as intensity of fluorescence of these cells followed pQE30>pAS01>pAS02 pattern, as expected from their respective origin of replication and from the previous expression data (Fig. 4A). The RFP showed poor fluorescence from pQE30-based plasmid, while the medium and low copy vector exhibited relatively high fluorescence (Fig. 4B). This kind of unusual pattern was also obvious from the plasmid copy number and gene expression data. YFP on the other hand showed highest fluorescence for the pQE30-based plasmid and then a drastic drop in fluorescence for pAS01-based plasmid and a negligible fluorescence for pAS02-based plasmid (Fig. 4C). A general observation for GFP and YFP had been that their intensity of fluorescence reduced with the lower plasmid copy number, while the number of fluorescing cells remained more or less similar. On the other hand, in case of RFP, the number of fluorescing cells is much less for pQE30 while the intensity remained similar. This phenomenon indicates the issue of plasmid stability for pQE30-based construct bearing RFP gene (Fig. 4).

### 3.5. Multiple gene co-expression at three different expression levels

The three florescent protein encoding genes *GFPuv, TurboRFP* and *EYFP*, each cloned in three different plasmids having compatible origin of replications, were co-expressed at different levels depending on plasmid expression system generating six strains, i.e., GRY, GYR, RGY, RYG, YGR and YRG, as mentioned in Materials and Methods Section 2.2 and Table 1. Here, the first, second and third position in the strain nomenclature denotes high (pQE30), medium (pAS01) and low (pAS02) copy plasmid and letter 'G', 'R' and 'Y' denotes *GFPuv, TurboRFP* and *EYFP* gene, respectively. Gene expression from the generated constructs was first analyzed using real-time qPCR.

All the three genes were found to be expressed in all the six strains by real-time qPCR (Fig. 5). The overall expression level of all the three genes across all the six strains seem not to be statistically different at

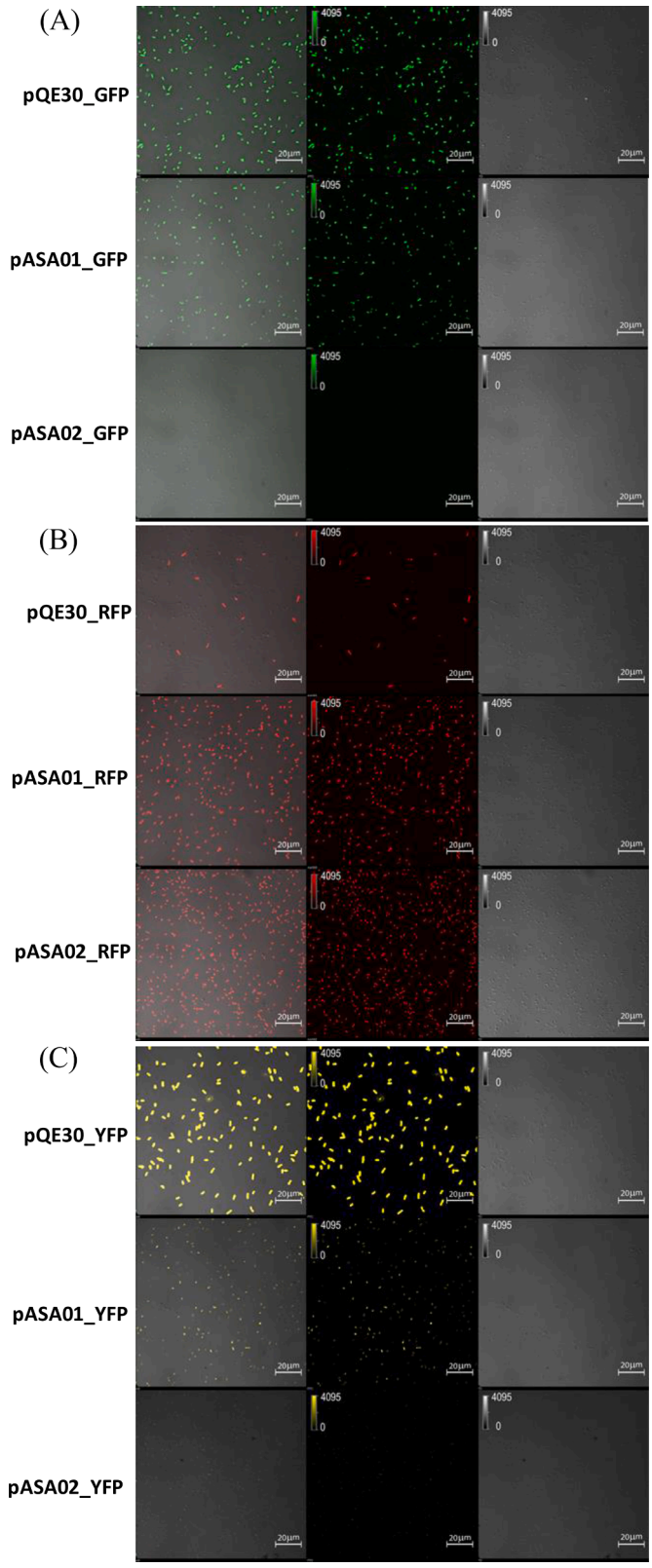


**Fig. 4.** Analysis of expression for single plasmid transformants expressing each of the three traffic signal reporter proteins GFPuv (A), TurboRFP (B) and EYFP (C) in high copy pQE30, medium copy pASA01 and low copy pASA02 plasmid using confocal microscopy. The first column is merged picture of phase-contrast with fluorescent image, second column is fluorescence image and last column is phase contrast image only.

0.05 level (except for YGR vs YRG) (Supplementary Table S1). In RGY, RYG, YGR, and YRG strains, the gene expression levels were consistent with their plasmid copy numbers (Fig. 5A). However in GRY strain, the highest level of gene expression was for *TurboRFP* present in medium copy plasmid rather than for *GFPuv* that was present in high copy plasmid. Also in GYR strain, the *EYFP* gene expression was lowest even when it was present in the medium copy plasmid, while *TurboRFP* present in the lowest copy plasmid showed moderate expression. These results suggest that while in majority of the conditions, the co-expression of multiple genes follows the expected expression pattern, there are few occasions where the expected norms are not followed.

To understand better about the occasions where the norms were not followed, we tried to understand the impact of co-expression of proteins on each other. When the expression of each gene was analyzed across all the six strains, it was observed that the genes expressed via high copy plasmid led to highest transcript level, irrespective of whether the second gene is cloned in the medium copy or the low copy plasmid (Fig. 5B). With respect to expression pattern of genes present in medium copy plasmid, we found that both GFP and YFP showed higher expression when RFP was present in high copy as compared to any of the other two. For example, GFP showed 1.16-fold expression in the RGY strain where GFP was in the medium copy and RFP was in the high copy plasmid. On the contrary, GFP showed only 0.28-fold expression in the YGR strain where YFP was in high copy. Similarly, YFP expressed better when RFP was either in the medium or higher copy plasmid. None of the other two proteins had significant impact on RFP expression. Pearson correlations between these pairs of proteins performed based on qPCR data also suggested correlation in the expression. A highly significant positive correlation was found between YFP and RFP at the 0.01 level, while significant positive correlations were found between GFP and each RFP and YFP at the 0.05 level (Supplementary Table S2).

### 3.6. *Quantification of fluorescence co-expression via cytometry*

For accurate quantification and qualification of different fluorescent proteins in all strains, FACS was used to detect different fluorophores in each strain. The cells transformed with the pQE30, pAS01, and pAS02 backbone plasmids were used as negative controls. Each strain had its positive controls corresponding to its constructs. All the three fluorescence proteins, i.e., GFP, RFP and YFP, were detected in all the six strains via fluorescence microscopy (Fig. 6), with their intensity and numbers varied between the strains as expected.

As mentioned before, the high level expression of EYFP overshadowed GFPuv expression level due to some overlap in their emission spectra. This was obvious with 97.6% cells expressing GFPuv in GRY strain, while only 37.5% cells expressed GFPuv in GYR strain. Moreover, the number of cells producing GFPuv were found to be negligible in the cases where GFPuv was expressed in the low and medium copy plasmid and YFP was co-expressed in high copy, e.g., YGR and YRG (Fig. 7, Supplementary Figure S3). Most of the cells showed RFP expression in the high (>97% cells) and medium (77–99% cells) copy plasmid range, while ~30% of the cells showed RFP expression in the low copy plasmid transformants. In case of YFP, high number of cells (94–99% cells) showed expression in all combinations, except for GRY strain where ~40% of cells showed YFP expression (Fig. 7). In terms of intensity of these fluorescence proteins, the highest florescence were detected where these proteins were expressed via high copy plasmid (Supplementary Figure S4). However, importantly, a significant difference in intensity exist between the two strains carrying high copy plasmid for each protein (eg. GRY vs GYR for GFPuv, RGY vs RYG for TurboRFP and YGR vs YRG for EYFP) (Supplementary Figure S4). The highest intensity was registered by YFP in YRG strain, followed by RFP in RYG strain. As expected, Pearson correlations between all pairs of proteins performed based on FACS data indicated a highly significant positive correlation between all proteins (Supplementary Table S3).

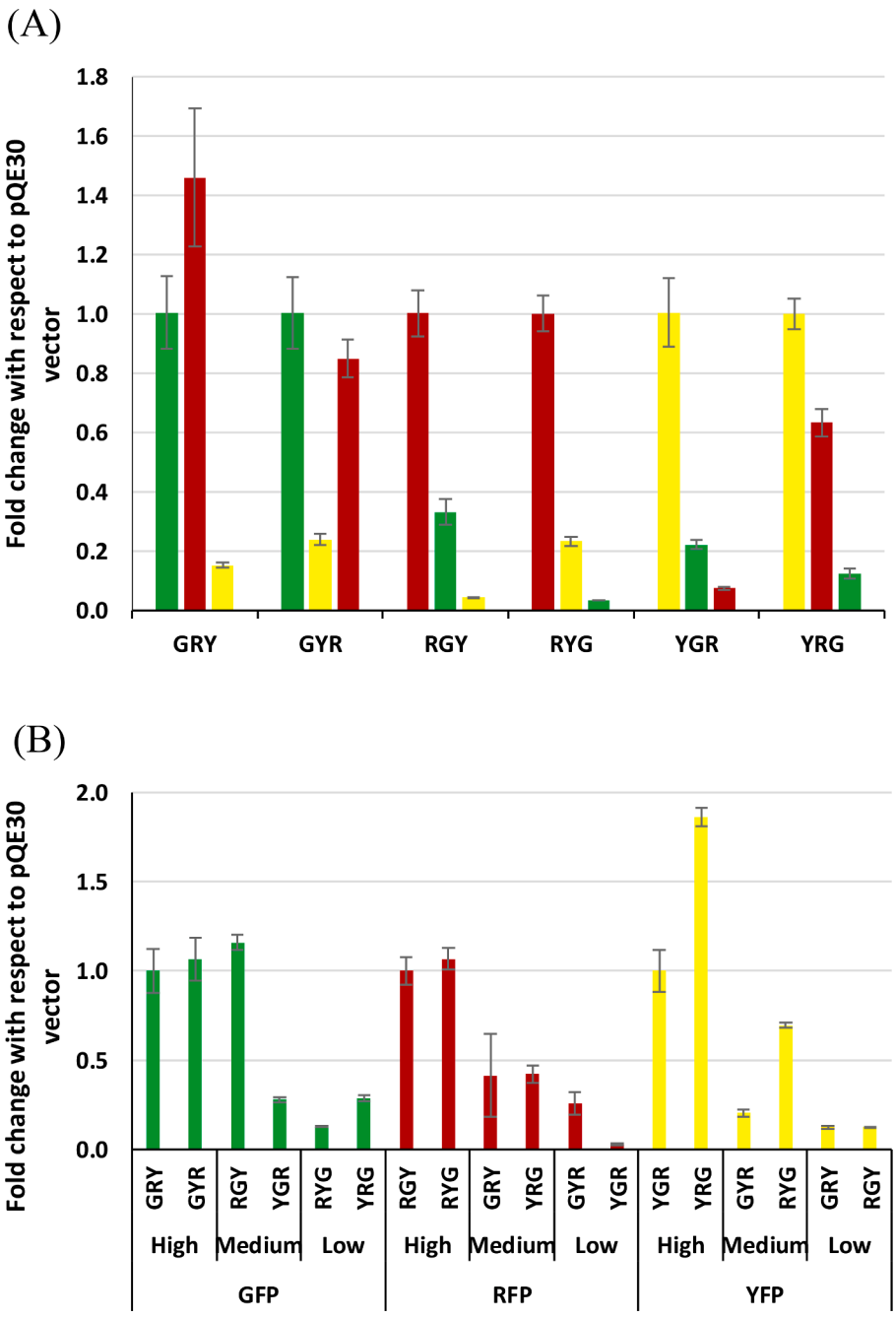


**Fig. 5.** Gene expression detection of GFP (G), RFP (R) and YFP (Y) in the triple-plasmid transformants using Real Time qPCR. (A) Comparison of expression level of the three genes within the same transformant. Details of strain name in the X-axis has been mentioned in Table 1. The first, second and third plasmid in the triplet represents pQE30, pASA01 and pASA02, respectively, and the expression of all the genes has been normalized with respect to that expressed by pQE30. (B) Comparison of expression of each gene across all the six transformants. The expression of the specific gene has been normalized with respect to that expressed by pQE30 of the first triad. The green, red and yellow bar represents transcript level of *GFPuv, TurboRFP* and *EYFP* gene, respectively. The High, Medium and Low represents the expression level from the plasmid pQE30, pASA01 and pASA02, respectively. Error bars represent the standard deviation ($n = 3$).

### *3.7. Plasmid heterogeneity and stability at the cellular level*

Plasmid loss is the principle factor of decreased recombinant protein productivity in plasmid-based systems, thus plasmid stability is necessary for stable and precise production of recombinant proteins. Although, naturally found *E. coli* plasmids are extremely stable, engineered expression plasmids are sometime lost upon co-expressed with other vectors, causing significant reductions of protein yields [20]. The six bacterial strains (GRY, GYR, RGY, RYG, YGR, and YRG) were grown, induced, appropriate dilutions were made in LB media and equal volumes were plated on 4 different LB agar plates, one was plain LB-agar plate, while three were having three different antibiotics according to the antibiotic markers present in the three plasmid (Supplementary Figure S5). Standard plate count method was used to obtain colony-forming units (CFU) from each plate for all strains. The data showed that the highest number of colonies on the non-selective plate was in GYR combination with about 100 million cells per ml of the liquid culture, while the lowest was in RGY combination with about 2 million cells (Fig. 8A). This suggest that while GFP is very well tolerated by the cells in the high copy plasmid, RFP is deleterious to the cells in the high copy. Also, the cell populations declined to less than half when GFP in high copy was associated with RFP in medium copy (in GRY) as compared to the strain GYR where YFP was in the medium copy. Similarly, the cell populations declined to less than half when YFP in high copy was associated with RFP in medium copy (in case of YRG) as compared to the strain YGR where GFP was in the medium copy. Cells with RFP in high copy preferred YFP in medium copy rather than GFP in medium copy. We further looked at whether all the three plasmids were present in the 6-co-transformed strains by plating the cells on different antibiotic plates. It was found that the GFP containing plasmids were the most stable-ones (87 – 97% positive population) irrespective of their origin of replications (Fig. 8B). The RFP containing plasmids were least

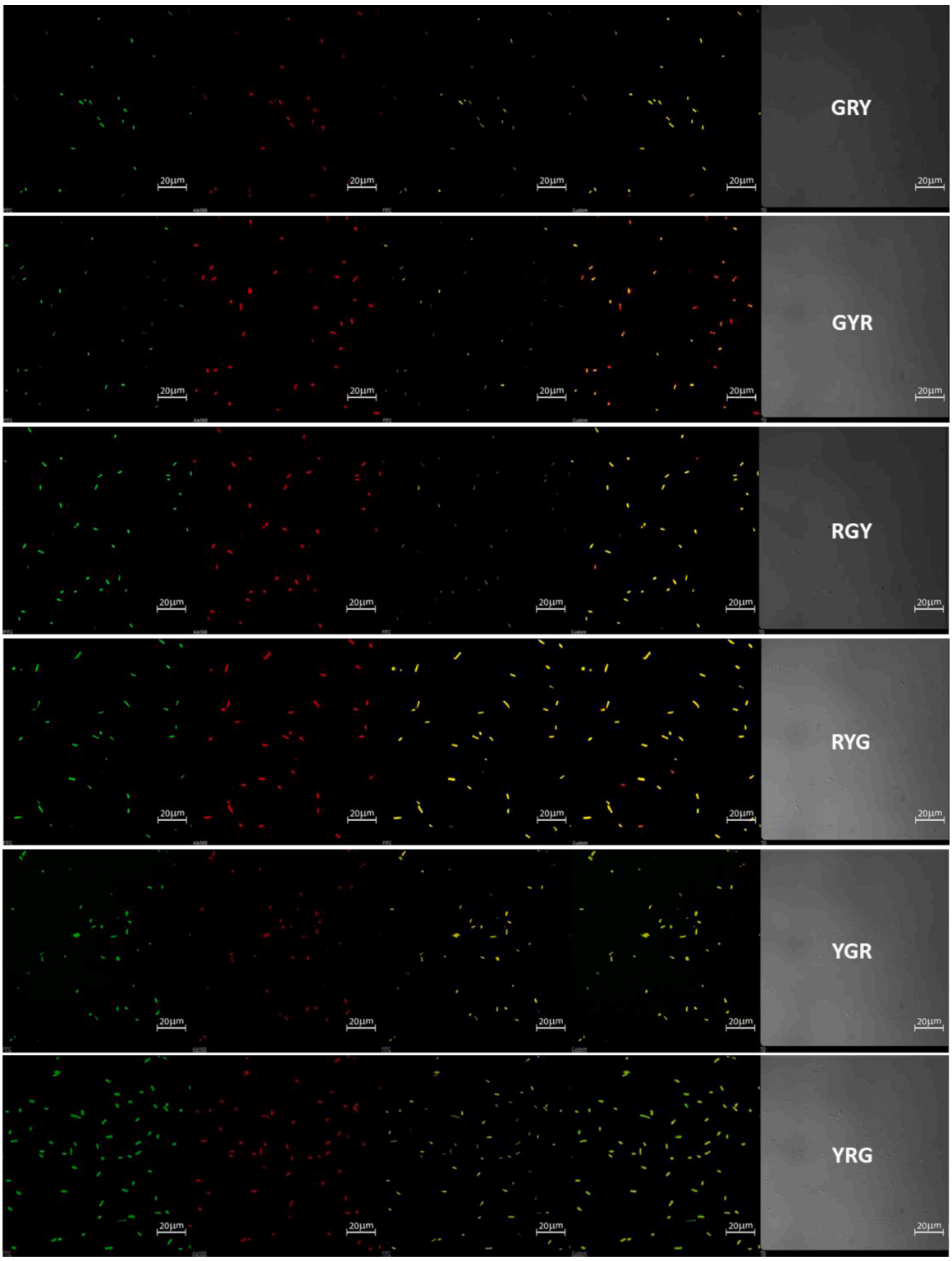


**Fig. 6.** Fluorescent micrographs showing different fluorescent proteins with different expression levels in *E. coli*. Individual cells with green, red, and yellow fluorescent proteins are shown in their colors along with merged photos and negative control.

stable in high copy plasmid (44 to 60% positive population) and most stable in low copy plasmid (68–96% positive population). YFP exhibited 75% and 92% positive population in high copy and low copy plasmid, respectively. YFP in the medium copy exhibited diverse profile with 93% positive population in strain GYR having GFP in high copy and only 29% positive population in strain RYG when RFP was in high copy.

The overall observation was that the largest number of populations in all the six co-transformants were bearing GFP containing plasmid and thus was the most favorable protein for the *E. coli*. On the other hand, RFP containing plasmids were the least suitable for the cells, indicating its possible toxicity and growth-inhibitory effect to the cells.

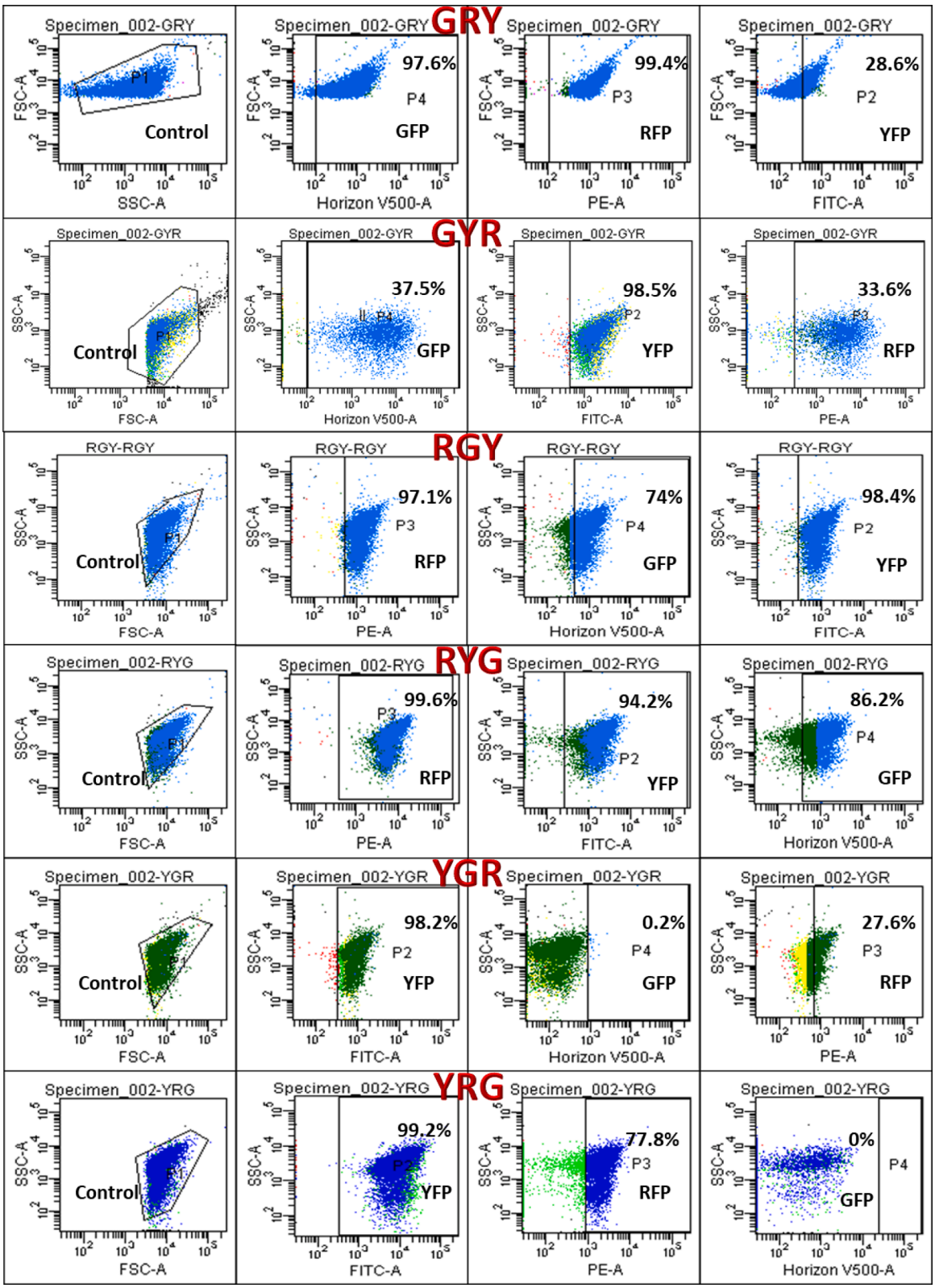


**Fig. 7.** Analysis of expression of different fluorescent proteins in different expression plasmids. Each graph shows the relative level of each florescent expression in arbitrary units per cell against forward scatter. An extension of this figure is Supplementary Figure S2 where all the controls are also included. Filters used for fluorescence detection: GFPuv – Horizon V500, TurboRFP – PE,EYFP- FITC.

## 4. Conclusions

In this study, we designed a 3-plasmid system having compatible origin of replication leading to high, medium and low copy number plasmid. All three plasmids had same promoter, RBS, and terminator for uniformity in expression in terms of inducer used. We cloned three commonly used reporter genes in the lab, i.e., GFPuv, EYFP, and TurboRFP, in all three plasmids to generate nine constructs and evaluated

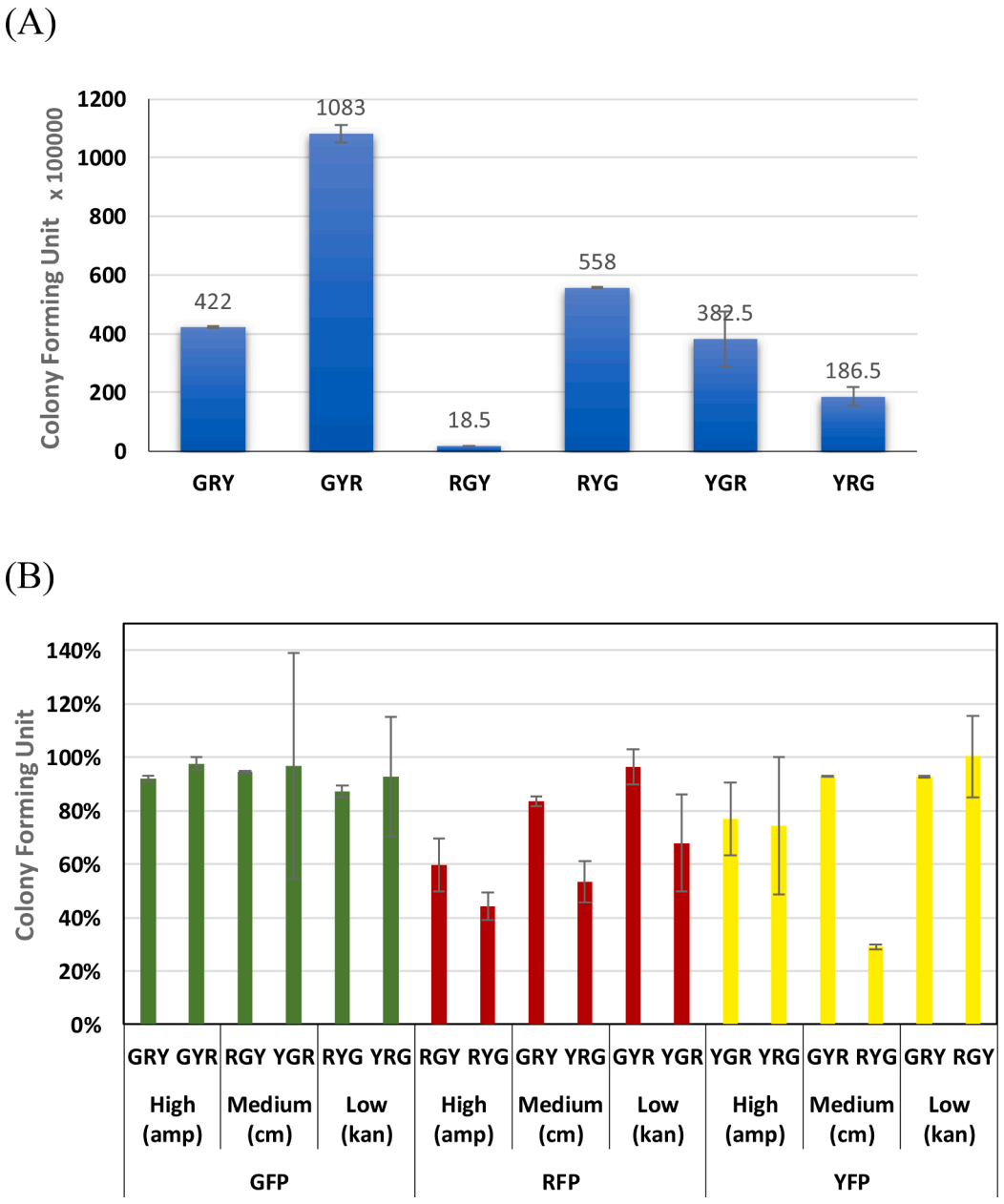


**Fig. 8.** Colony-forming units (CFU) of all strains. (A) CFU on LB plate and (B) CFU percentage on different antibiotic plate with respect to LB plate. Error bars represent the standard deviation ($n = 3$).

their expression individually and in combination. The real time PCR and emission of fluorescence of the single plasmid transformed cells indicated that GFP and YFP were expressed best in the high copy plasmid, while RFP expression declined in the high copy plasmid. When all the three plasmids were transformed in six different combination in the *E. coli* host, it was found in many of them that the co-expression of proteins led to interference in expression on each other. Importantly, a significant loss of viability was observed for the strain RGY, suggesting TurboRFP may be imparting toxic effect to the cells when expressed in high copy. Overall, our toolkit allows the expression of multiple gene/operons in different proportion. Further, our study provides new insights on the co-expression of common reporter proteins in the laboratory.

## 5. Funding

This study was supported by fund received from Department of Biotechnology, Govt of India via grant no. BT/PR/centre/03/2011-Phase II.

## 6. Author contributions

ASA and SSY designed the study, ASA performed experimental work, ASA and SSY analyzed the data and wrote the manuscript, SSY generated resources for the study.

## Declaration of Competing Interest

Authors have no competing interests to declare.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.btre.2021.e00692.

## Reference


[1] D.A. Jackson, R.H. Symons, P. Berg, Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of escherichia coli, in: Proceedings of the National Academy of Sciences 69, 1972, pp. 2904–2909.

[2] S.N. Cohen, et al., Construction of biologically functional bacterial plasmids in vitro, in: Proceedings of the National Academy of Sciences 70, 1973, pp. 3240–3244.

[3] Palomares, L.A., S. Estrada-Mondaca, and O.T. Ramírez, Production of recombinant proteins: challenges and solutions, in Recombinant Gene Expression. 2004. p. 015-052.

[4] F.J.M. Mergulhão, et al., Medium and copy number effects on the secretion of human proinsulin in Escherichia coli using the universal stress promoters uspA and uspB, Appl. Microbiol. Biotechnol. 61 (5–6) (2003) 495–501.

[5] T. Carrier, K.L. Jones, J.D. Keasling, mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system, Biotechnol. Bioeng. 59 (6) (1998) 666–672.

[6] J.R. Swartz, Advances in Escherichia coli production of therapeutic proteins, Curr. Opin. Biotechnol. 12 (2) (2001) 195–201.

[7] L.C. Simmons, D.G. Yansura, Translational level is a critical factor for the secretion of heterologous proteins in Escherichia coli, Nat. Biotechnol. 14 (5) (1996) 629–634.

[8] H. Chen, et al., Determination of the optimal aligned spacing between the shine – dalgarno sequence and the translation initiation codon of Escherichia coli m RNAs, Nucleic Acids Res. 22 (23) (1994) 4953–4957.

[9] S.K. Berwal, R.K. Sreejith, J.K. Pal, Distance between RBS and AUG plays an important role in overexpression of recombinant proteins, Anal. Biochem. 405 (2) (2010) 275–277.

[10] E. Schenborn, D. Groskreutz, Reporter gene vectors and assays, Mol. Biotechnol. 13 (1) (1999) 29–44.

[11] K.-J. Kim, et al., Two-promoter vector is highly efficient for overproduction of protein complexes, Protein Sci. 13 (6) (2004) 1698–1703.

[12] Jawed, K., et al., Improved butanol production using FASII pathway in e. coli. ACS synthetic biology, 2020. 9(9): p. 2390–2398.

[13] T. Vogl, et al., Engineered bidirectional promoters enable rapid multi-gene co-expression optimization, Nat. Commun. 9 (1) (2018).

[14] S. Yang, S.C. Sleight, H.M. Sauro, Rationally designed bidirectional promoter improves the evolutionary stability of synthetic genetic circuits, Nucleic Acids Res. 41 (1) (2013) e33-e33.

[15] M. Skulj, et al., Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes, Microb. Cell Fact. 7 (1) (2008).

[16] N. Akasaka, et al., Change in the plasmid copy number in acetic acid Bacteria in response to growth phase and acetic acid Concentration, J. Biosci. Bioeng. 119 (6) (2015) 661–668.

[17] M. Plotka, M. Wozniak, T. Kaczorowski, Quantification of plasmid copy number with single colour droplet digital PCR, PLoS ONE 12 (1) (2017) e0169846.

[18] A. Herman-Antosiewicz, M. Obuchowski, G. Wegrzyn, A plasmid cloning vector with precisely regulatable copy number in escherichia coli, Mol. Biotechnol. 17 (3) (2001) 193–200.

[19] J.W. Choi, K.S. Ra, Y.S. Lee, Enhancement of bovine growth hormone gene expression by increasing the plasmid copy number, Biotechnol. Lett. 21 (1999) 1-5.

[20] P. Balbás, Understanding the art of producing protein and nonprotein molecules in escherichia coli, Mol. Biotechnol. 19 (3) (2001) 251–268.