G&I Genomics & Informatics

eISSN 2234-0742
Genomics Inform 2015;13(3):81-85
http://dx.doi.org/10.5808/GI.2015.13.3.81

ORIGINAL ARTICLE

Efficiency to Discovery Transgenic Loci in GM Rice Using Next Generation Sequencing Whole Genome Re-sequencing

Doori Park[1], Dongin Kim[2], Green Jang[2], Jongsung Lim[2], Yun-Ji Shin[3], Jina Kim[3], Mi-Seong Seo[3], Su-Hyun Park[4], Ju-Kon Kim[4], Tae-Ho Kwon[3]*, Ik-Young Choi[1]**

[1]Institute of Green Bio Science and Technology, Seoul National University, Pyeongchang 25354, Korea,
[2]National Instrumentation Center for Environmental Management, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea,
[3]Natural Bio-Materials Inc., Wanju 55322, Korea, [4]Crop Biotech Institute, Green-Bio Science and Technology, Seoul National University, Pyeongchang 25354, Korea

Molecular characterization technology in genetically modified organisms, in addition to how transgenic biotechnologies are developed now require full transparency to assess the risk to living modified and non-modified organisms. Next generation sequencing (NGS) methodology is suggested as an effective means in genome characterization and detection of transgenic insertion locations. In the present study, we applied NGS to insert transgenic loci, specifically the epidermal growth factor (EGF) in genetically modified rice cells. A total of 29.3 Gb (~72× coverage) was sequenced with a 2 × 150 bp paired end method by Illumina HiSeq2500, which was consecutively mapped to the rice genome and T-vector sequence. The compatible pairs of reads were successfully mapped to 10 loci on the rice chromosome and vector sequences were validated to the insertion location by polymerase chain reaction (PCR) amplification. The EGF transgenic site was confirmed only on chromosome 4 by PCR. Results of this study demonstrated the success of NGS data to characterize the rice genome. Bioinformatics analyses must be developed in association with NGS data to identify highly accurate transgenic sites.

Keywords: genetically modified organisms, next generation sequencing (NGS) T-DNA, rice, risk assessment

Introduction

Genetic engineering technology is widely used in the agricultural and plant biotechnology fields, ranging from the food and feed industries to bio-pharmaceuticals and cosmetics [1, 2]. The history of genetically modified (GM) technology began with the discovery of plasmid DNA, where the plasmid could be transferred from one cell to another genome [3]. Scientists subsequently applied the basic plasmid vector system principle and developed recombinant DNA technology to create genetically engineered organisms. Today, GM techniques have been applied to various research fields, including crop sciences, drug manufacturing, and animal husbandry.

The development of transgenic biotechnologies over the last 20 years has led to safety concerns regarding genetically modified organisms (GMOs), particularly in food crops and new pharmaceuticals, which are the most controversial issues. Safety concerns regarding GMOs have resulted in research, debates, and ongoing public unease. Therefore, the European Union (EU) and National Institutes of Health (NIH) in the United States proposed an authorization process in commercial GMO use; however, public apprehension for transgenic techniques remains uncertain and controversial [4-8].

Generally, molecular characterization and identification of GMOs are performed using Southern blots and polymerase chain reaction (PCR) based detection followed by conventional sequencing methods [7]. However, these appro-

Received July 25, 2015; Revised August 24, 2015; Accepted August 24, 2015

***Corresponding author:** Tel: +82-63-223-4555, Fax: +82-63-223-4550, E-mail: thkwon@jbnu.ac.kr

****Corresponding author:** Tel: +82-33-339-5657, Fax: +82-33-339-5753, E-mail: choii@snu.ac.kr

aches are limited to evaluate whether the host genome has unintended sequence substitutions and indels [9]. Moreover, if sufficient genomic information is not available for the chosen comparative model species, it is difficult to detect the correct transgenic insert site location or sequence contamination of vector DNA [9, 10].

Recent publications of GMO molecular characterizations reported the use of next generation sequencing (NGS) approaches as an effective means to detect the precise transgenic insert location [9, 11, 12]. High-throughput DNA sequencing technologies and bioinformatics can be coupled with NGS to offer new possibilities in drawing genetic maps with feasible costs. For these reasons, researchers have tested new approaches in the molecular characterization of GMOs using NGS technologies [9, 10, 12].

Here, we examined transgenic insertion sites using paired-end whole genome re-sequencing data following Yang *et al*. with modifications [9]. Human epidermal growth factor (EGF) was inserted into GM rice cells, which could produce EGF safety without endotoxin derived from bacteria and was used as material for this study. Deep sequencing was performed with the Illumina Hiseq2500 platforms (Illumina Inc., San Diego, CA, USA). In this pilot study, we demonstrated the potential of NGS for examination of transgenic insertion loci and discuss some technical bottlenecks of this new method.

Methods

GM rice samples

The GM rice event PJKS131-2 was transformed with the *EGF* inserted pJKS131 vector, produced by Natural Bio-Materials Inc. (Jeonju, Korea). Taxonomically, the event PJKS131-2 was derived from *Oryza sativa* L. cv. Dongjin. The T-vector was transformed with rice callus as described by Chan *et al*. [13]. Transgenic rice calli were incubated with 50 mg/L of hygromycin B antibiotic (A.G. Scientific Inc., San Diego, CA, USA) for selection. The GM rice callus samples were subjected to NGS and further validated by PCR amplification.

DNA extraction and whole genome shotgun library and sequencing

The calli of GM rice event PJKS131-2 were collected and stored at −80°C. Total genomic DNA was extracted using the CTAB method in liquid nitrogen. Genomic DNA quality was evaluated by 0.5% agarose gel electrophoresis. Following the quality check, genomic DNA was sheared with average 500 bp fragment sizes. Truseq DNA PCR free Library Preparation Kit (Illumina Inc.) was used to construct the DNA library according to the manufacturer's protocol. The quality of constructed DNA libraries was confirmed by the LabChip GX system (PerkinElmer, Waltham, MA, USA). DNA libraries were sequenced with 150-bp paired-end sequencing using Illumina Hiseq2500.

Transgenic insertion analysis

Initially, paired-end reads were filtered out by phred scores < 20 and duplicate sequences were removed. After filtration, DNA fragments were consecutively mapped against the rice reference genome (phytozome v9 [14]) and T-vector sequence (Supplementary Fig. 1). The transgene insertion types were classified by adaptation and modification of the analytical strategies reported in Yang *et al*. [9]. Fig. 1 shows the workflow applied in this method. Initially, all NGS reads were individually mapped to the rice reference genome and transgenic vector (types A and C in Yang *et al*. [9]). Subsequently, these NGS reads were eliminated to conduct the following analyses. NGS reads not classified as above were classified into the following two classes: one side of the NGS read matched the reference genome, (1) the other one matched to vector (type B in Yang *et al*. [9]); or (2) one

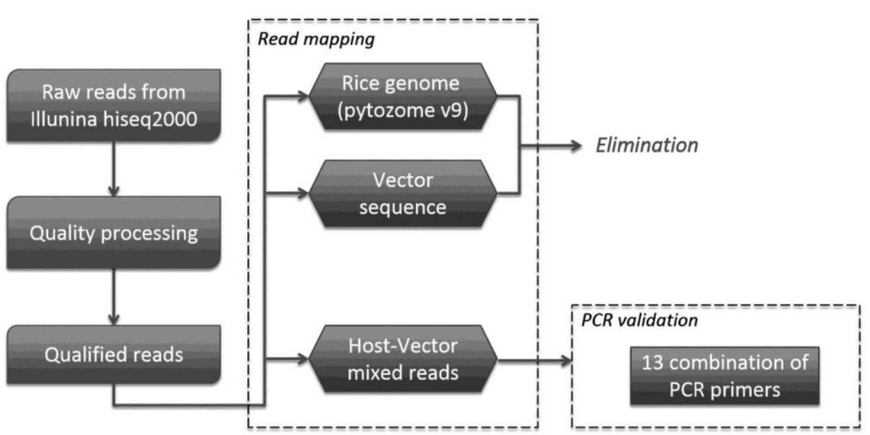


Fig. 1. Summary of the work-flow. PCR, polymerase chain reaction.

Table 1. Whole genome sequencing summary

Event	No. of reads	Total read length (bp)	Q30 (%)	GC ratio (%)
PJKS131-2	194,965,440	29,359,127,691 (72×)	71.56	41.58

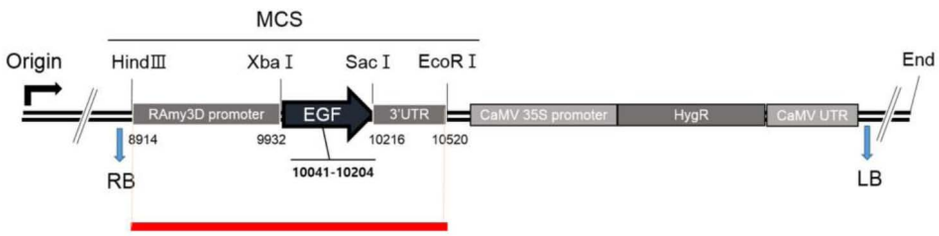


Fig. 2. Schematic drawing of the transgenic vector genome. Red line represent the region of mapped reads. MCS, multiple cloning site; RB, right border; LB, left border.

side of the NGS read exhibited both elements from the rice reference and transgenic vector (types D and E in Yang *et al*. [9]).

Experimental validation of transgenic inserts

Each of the 13 combination primer sets was designed congruent with the transgenic insertion region orientation. PCR was conducted using DNA polymerase (Solgent Co., Daejeon, Korea) following the manufacturer's instructions. The reaction was performed under the following conditions: a pre-denaturation step at 95°C for 5 min; denaturation at 95°C for 60 s; 30 amplification cycles, including annealing at 60°C for 45 s, and elongation at 72°C for 120 s; and a final elongation at 72°C for 5 min.

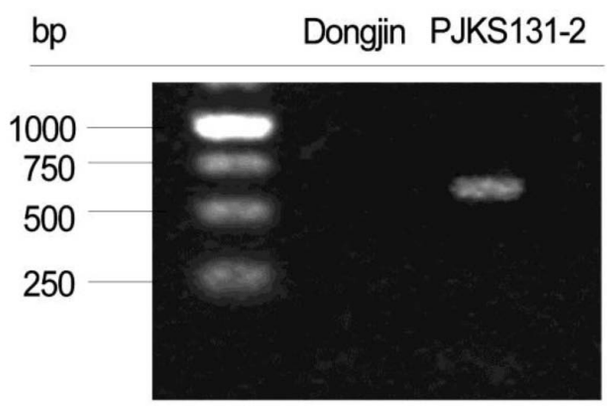


Fig. 3. Polymerase chain reaction validation of transgenic site.

Results

Whole genome re-sequencing and mapping to discover the transgenic position

The transgenic GM rice site, PJKS131-2, was detected by performing whole genome re-sequencing using callus tissue. Genomic DNA libraries were constructed with an average 500 bp and both ends were read with 150 bp paired-end sequencing methods. A total length of raw sequencing reads were 29.3 Gb (~194.9 million reads), which showed ~72× coverage in the total read length (Table 1). Following quality control processing, reads with average phred scores ≥ 30 were estimated at ~71.5% (Table 1).

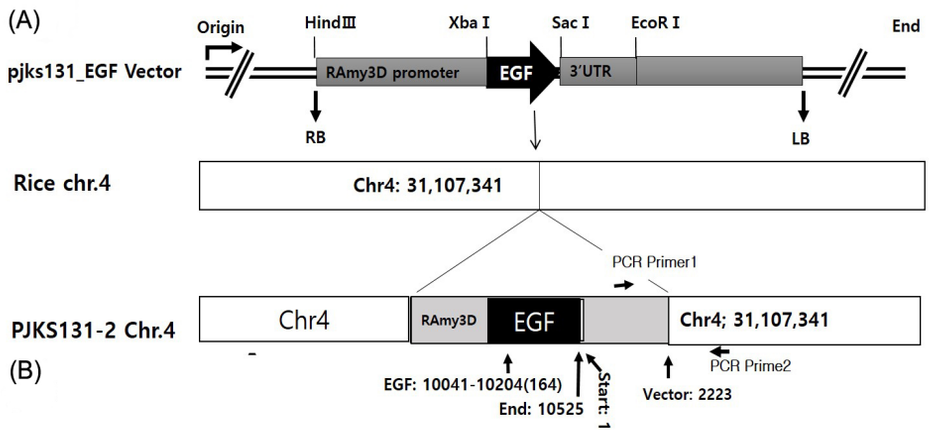
The types of mapped reads were classified by alignment of all NGS reads to the rice reference genome and transgenic vector sequences. Fig. 2 shows construction of the pJKS131 transgenic vector. Reads were aligned on the cloning vector positions 8,500 bp to 10,500 bp, similar to transgenic insert locations. Detailed mapping strategies were described in the Methods. The transgene insertion site was identified by classifying reads where one end matched the host genome and the other end matched the vector sequences (i.e., types B, D, and E) mapped back to the rice chromosome and known vector sequences. Eleven pairs of reads were identified on rice chromosome including chromosome 4. The total mapped reads described above were compatible with the transgenic vector backbone sequences.

PCR validation of mapping prediction

Thirteen PCR primers designed based on mapping direction validated the mapping results of 10 transgenic insert candidates. PCR results confirmed the target *EGF* sequence was successfully inserted on rice chromosome 4 (Figs. 3 and 4). The remaining reads were concluded to be artifacts, because all matches were not detected with PCR.

Discussion

Recent developments in NGS methods and accompanying bioinformatics tools have paved the way for ongoing genomics research widely used in the agricultural biotechnology field. Consequently, several studies reported new



GAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGAAATAGAGTAGATGCCGACCGGGATCTGTCG
ATCGACAAGCTCGAGTTTCTCCATAATAATGTGTGAGTAGTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGT
TGAGCATATAAGAAACCCTTAGTATGTGATGCAAGTAAACAATTATGGATTCTACTTATGTTTCTTTTCCTTTTTCTGACTTGGGTTGCAGG
TATGAGAAGAGGCATTCCAACATTCCGGCTCACGTCTCCCCATGCTTCCGTGTCAAGGAAGGTGACCATGTCATCATTGGCCAGTGCAGGT
AAAACTTAACTCCCATACCTTAGTTTTTGACTCCTTAGAATCATCTTAAATAACAGAGGTGCTATTAGTTCAAACATTATCAGTTCACCCAGC
TAGTACTATGAGATTTGGTTGCTGAAGTACAATTTCTGCATTTTCAACAGGCCGCTGTCGAAAACTGTGAGGTTCAACGTCCTGAAGGTCAT
CCCAGCTGGATCCACCGGCGGCAGCGGCGGCAAGAAGGCCTTCACCG

Fig. 4. Transgenic position of epidermal growth factor (*EGF*) locus on the rice chromosome 4 and polymerase chain reaction (PCR) test to identify T-DNA junction sequence. (A) The *EGF* is inserted on the position 31,104,341 of the chromosome 4. (B) The bold with underline is T-DNA sequence of the vector 2,026−2,223 bp and the next bases is rice transgenic locus chromosome 4 (31107341-31107690) in the fragment amplified by PCR test primer1 (5′ TACCTGCATGCTGCGGTGAAG 3′) and primer2 (5′AGGGCTGTGTAGAAGTACTCGC 3′).

approaches in GM crop safety assessment using NGS platforms [10-12]. In our study, we investigated EGF inserted GM rice events using NGS technology and bioinformatics to test the potential uses of this new approach in molecular assessment of transgenic organisms.

Results were successful in differentiating NGS read types using *in silico* analyses from GM rice, PJKS131-2 and hypothetically, the outcome was acceptable in terms of read classification. However, as a validation step, we experienced unexpected problems. Consistent with mapping and aligning data, we considered all possible transgenic insertion directions on the rice chromosomes and designed PCR primers based on loci information. Among the primers, except for locus specific primers on chromosome 4, results showed all matches were mismatches, which was caused by computational errors derived from analogous sequences between the rice genome and the transgenic vector. Therefore, we concluded it is essential to develop more accurate algorithms based on the transformation vector.

In addition, it is important to note our experimental sample was collected from rice callus tissues, with *Agrobacterium* co-incubation and a plant cell suspension culture system. Transgenic plant cell suspension culture system exhibits several advantages, including a low microorganism risk and chemical contamination, simple cell culture methods, economical facilities, and stable productivity. However, it is difficult to obtain pure genomic DNA of the host plant without plasmid DNA mixing using the plant cell culture method. We eliminated NGS raw reads mapped only against vector DNA (type C), however if raw reads contained too many vector backbone sequences, problems in further bioinformatics analyses would still occur. Further studies are required with appropriate controls of GM plants in cell culture environments.

In the present study, we completed a proof-of-concept experiment to examine the molecular characterization of a recombinant-protein produced GM rice event using NGS methods. New approaches have recently been reported to assess the development and release of GM crops, however these techniques are not popularized in the field of GM risk assessment. However, previous studies in other disciplines have successfully established NGS, but for practical reasons, it has not been easy to apply this new method for testing GMOs. NGS strategies largely depend on sample quality, amount of data, and subsequent bioinformatics analyses. Therefore, it is critical proper guidelines to discovery transgenic site by NGS data matched and PCR test in the GMOs established and required.

Supplementary material

Supplementary data including one figure can be found with this article online at http://www.genominfo.org/src/sm/gni-13-81-s001.pdf.

Acknowledgments

We thank members of bioinformatics team of National Instrumentation Center for Environmental management in Seoul National University. This work was supported by Next-Generation BioGreen21 program (PJ01131301), Rural Development Administration of the Korean government.

References

1. Sabalza M, Christou P, Capell T. Recombinant plant-derived pharmaceutical proteins: current technical and economic bottlenecks. *Biotechnol Lett* 2014;36:2367-2379.
2. Schürch C, Blum P, Zülli F. Potential of plant cells in culture for cosmetic application. *Phytochem Rev* 2008;7:599-605.
3. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids *in vitro*. *Proc Natl Acad Sci U S A* 1973;70:3240-3244.
4. European Parliament. Commission Implementing Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. European Union Legislation. OJ L 268, 1-23. 2003.
5. European Parliament. REGULATION (EC) No 1830/2003 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 September 2003 concerning the traceability and labelling of genetically modified organisms and the traceability of food and feed products produced from genetically modified organisms and amending Directive 2001/18/EC. European Union Legislation. OJ L 268, 24-28. 2003.
6. European Parliament. Commission Implementing Regulation (EU) No 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006. European Union Legislation. OJ L 157, 1-48. 2013.
7. Food and Agricultural Organization of the United Nations (FAO). Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants. Rep. No. CAC/GL 45-2003. Rome: FAO, 2003.
8. Food and Agricultural Organization of the United Nations (FAO). Guideline for the conduct of food safety assessment of foods derived from recombinant-DNA plants. Rep. No. CAC/GL. 68-2008. Rome: FAO, 2003.
9. Yang L, Wang C, Holst-Jensen A, Morisset D, Lin Y, Zhang D. Characterization of GM events by insert knowledge adapted re-sequencing approaches. *Sci Rep* 2013;3:2839.
10. Pauwels K, De Keersmaecker SC, De Schrijver A, du Jardin P, Roosens NH, Herman P. Next-generation sequencing as a tool for the molecular characterisation and risk assessment of genetically modified plants: added value or not? *Trends Food Sci Technol* 2015;45:319-326.
11. Kovalic D, Garnaat C, Guo L, Yan Y, Groat J, Silvanovich A, *et al*. The use of next generation sequencing and junction sequence analysis bioinformatics to achieve molecular characterization of crops improved through modern biotechnology. *Plant Genome* 2012;5:149-163.
12. Wahler D, Schauser L, Bendiek J, Grohmann L. Next-generation sequencing as a tool for detailed molecular characterisation of genomic insertions and flanking regions in genetically modified plants: a pilot study using a rice event unauthorised in the EU. *Food Anal Methods* 2013;6:1718-1727.
13. Chan MT, Chang HH, Ho SL, Tong WF, Yu SM. *Agrobacterium*-mediated production of transgenic rice plants expressing a chimeric alpha-amylase promoter/beta-glucuronidase gene. *Plant Mol Biol* 1993;22:491-506.
14. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, *et al*. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* 2012;40:D1178-D1186.

SUPPLEMENTARY INFORMATION

Efficiency to Discovery Transgenic Loci in GM Rice Using Next Generation Sequencing Whole Genome Re-sequencing

Doori Park[1], Dongin Kim[2], Green Jang[2], Jongsung Lim[2], Yun-Ji Shin[3], Jina Kim[3], Mi-Seong Seo[3], Su-Hyun Park[4], Ju-Kon Kim[4], Tae-Ho Kwon[3]*, Ik-Young Choi[1]**

[1]Institute of Green Bio Science and Technology, Seoul National University, Pyeongchang 25354, Korea,
[2]National Instrumentation Center for Environmental Management, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea,
[3]Natural Bio-Materials Inc., Wanju 55322, Korea,
[4]Crop Biotech Institute, Green-Bio Science and Technology, Seoul National University, Pyeongchang 25354, Korea

http://www.genominfo.org/src/sm/gni-13-81-s001.pdf

※ pJKS131 vector sequence
1060-2085 : HPTII
2376-2401 : LB T-DNA
8634-8661 : RB T-DNA
8907-9938 : RAmy3D promoter
9945-9965 : Hwang`s 5`UTR
9966-10040 : RAmy3D signal peptide
10041-10204 : EGF mature peptide
10211-10519 : RAmy3D 3`UTR

```
        10         20         30         40         50         60
GTAATCATGG TCATAGCTGT TTCCTGTGTG AAATTGTTAT CCGCTCACAA TTCCACACAA

        70         80         90        100        110        120
CATACGAGCC GGAAGCATAA AGTGTAAAGC CTGGGGTGCC TAATGAGTGA GCTAACTCAC

       130        140        150        160        170        180
ATTAATTGCG TTGCGCTCAC TGCCCGCTTT CCAGTCGGGA AACCTGTCGT GCCAGCTGCA

       190        200        210        220        230        240
TTAATGAATC GGCCAACGCG CGGGGAGAGG CGGTTTGCGT ATTGGCTAGA GCAGCTTGCC

       250        260        270        280        290        300
AACATGGTGG AGCACGACAC TCTCGTCTAC TCCAAGAATA TCAAAGATAC AGTCTCAGAA

       310        320        330        340        350        360
GACCAAAGGG CTATTGAGAC TTTTCAACAA AGGGTAATAT CGGGAAACCT CCTCGGATTC

       370        380        390        400        410        420
CATTGCCCAG CTATCTGTCA CTTCATCAAA AGGACAGTAG AAAAGGAAGG TGGCACCTAC

       430        440        450        460        470        480
AAATGCCATC ATTGCGATAA AGGAAAGGCT ATCGTTCAAG ATGCCTCTGC CGACAGTGGT

       490        500        510        520        530        540
CCCAAAGATG GACCCCCACC CACGAGGAGC ATCGTGGAAA AAGAAGACGT TCCAACCACG

       550        560        570        580        590        600
TCTTCAAAGC AAGTGGATTG ATGTGATAAC ATGGTGGAGC ACGACACTCT CGTCTACTCC

       610        620        630        640        650        660
AAGAATATCA AAGATACAGT CTCAGAAGAC CAAAGGGCTA TTGAGACTTT TCAACAAAGG

       670        680        690        700        710        720
GTAATATCGG GAAACCTCCT CGGATTCCAT TGCCCAGCTA TCTGTCACTT CATCAAAAGG

       730        740        750        760        770        780
ACAGTAGAAA AGGAAGGTGG CACCTACAAA TGCCATCATT GCGATAAAGG AAAGGCTATC
```

```
       790         800         810         820         830         840
GTTCAAGATG CCTCTGCCGA CAGTGGTCCC AAAGATGGAC CCCCACCCAC GAGGAGCATC

       850         860         870         880         890         900
GTGGAAAAAG AAGACGTTCC AACCACGTCT TCAAAGCAAG TGGATTGATG TGATATCTCC

       910         920         930         940         950         960
ACTGACGTAA GGGATGACGC ACAATCCCAC TATCCTTCGC AAGACCTTCC TCTATATAAG

       970         980         990        1000        1010        1020
GAAGTTCATT TCATTTGGAG AGGACACGCT GAAATCACCA GTCTCTCTCT ACAAATCTAT

      1030        1040        1050        1060        1070        1080
CTCTCTCGAG CTTTCGCAGA TCCCGGGGGG CAATGAGATA TGAAAAAGCC TGAACTCACC

      1090        1100        1110        1120        1130        1140
GCGACGTCTG TCGAGAAGTT TCTGATCGAA AAGTTCGACA GCGTCTCCGA CCTGATGCAG

      1150        1160        1170        1180        1190        1200
CTCTCGGAGG GCGAAGAATC TCGTGCTTTC AGCTTCGATG TAGGAGGGCG TGGATATGTC

      1210        1220        1230        1240        1250        1260
CTGCGGGTAA ATAGCTGCGC CGATGGTTTC TACAAAGATC GTTATGTTTA TCGGCACTTT

      1270        1280        1290        1300        1310        1320
GCATCGGCCG CGCTCCCGAT TCCGGAAGTG CTTGACATTG GGGAGTTTAG CGAGAGCCTG

      1330        1340        1350        1360        1370        1380
ACCTATTGCA TCTCCCGCCG TGCACAGGGT GTCACGTTGC AAGACCTGCC TGAAACCGAA

      1390        1400        1410        1420        1430        1440
CTGCCCGCTG TTCTACAACC GGTCGCGGAG GCTATGGATG CGATCGCTGC GGCCGATCTT

      1450        1460        1470        1480        1490        1500
AGCCAGACGA GCGGGTTCGG CCCATTCGGA CCGCAAGGAA TCGGTCAATA CACTACATGG

      1510        1520        1530        1540        1550        1560
CGTGATTTCA TATGCGCGAT TGCTGATCCC CATGTGTATC ACTGGCAAAC TGTGATGGAC

      1570        1580        1590        1600        1610        1620
GACACCGTCA GTGCGTCCGT CGCGCAGGCT CTCGATGAGC TGATGCTTTG GGCCGAGGAC

      1630        1640        1650        1660        1670        1680
TGCCCCGAAG TCCGGCACCT CGTGCACGCG GATTTCGGCT CCAACAATGT CCTGACGGAC

      1690        1700        1710        1720        1730        1740
AATGGCCGCA TAACAGCGGT CATTGACTGG AGCGAGGCGA TGTTCGGGGA TTCCCAATAC
```

```
      1750        1760        1770        1780        1790        1800
GAGGTCGCCA ACATCTTCTT CTGGAGGCCG TGGTTGGCTT GTATGGAGCA GCAGACGCGC

      1810        1820        1830        1840        1850        1860
TACTTCGAGC GGAGGCATCC GGAGCTTGCA GGATCGCCAC GACTCCGGGC GTATATGCTC

      1870        1880        1890        1900        1910        1920
CGCATTGGTC TTGACCAACT CTATCAGAGC TTGGTTGACG GCAATTTCGA TGATGCAGCT

      1930        1940        1950        1960        1970        1980
TGGGCGCAGG GTCGATGCGA CGCAATCGTC CGATCCGGAG CCGGGACTGT CGGGCGTACA

      1990        2000        2010        2020        2030        2040
CAAATCGCCC GCAGAAGCGC GGCCGTCTGG ACCGATGGCT GTGTAGAAGT ACTCGCCGAT

      2050        2060        2070        2080        2090        2100
AGTGGAAACC GACGCCCCAG CACTCGTCCG AGGGCAAAGA AATAGAGTAG ATGCCGACCG

      2110        2120        2130        2140        2150        2160
GATCTGTCGA TCGACAAGCT CGAGTTTCTC CATAATAATG TGTGAGTAGT TCCCAGATAA

      2170        2180        2190        2200        2210        2220
GGGAATTAGG GTTCCTATAG GGTTTCGCTC ATGTGTTGAG CATATAAGAA ACCCTTAGTA

      2230        2240        2250        2260        2270        2280
TGTATTTGTA TTTGTAAAAT ACTTCTATCA ATAAAATTTC TAATTCCTAA AACCAAAATC

      2290        2300        2310        2320        2330        2340
CAGTACTAAA ATCCAGATCC CCCGAATTAA TTCGGCGTTA ATTCAGTACA TTAAAAACGT

      2350        2360        2370        2380        2390        2400
CCGCAATGTG TTATTAAGTT GTCTAAGCGT CAATTTGTTT ACACCACAAT ATATCCTGCC

      2410        2420        2430        2440        2450        2460
ACCAGCCAGC CAACAGCTCC CCGACCGGCA GCTCGGCACA AAATCACCAC TCGATACAGG

      2470        2480        2490        2500        2510        2520
CAGCCCATCA GTCCGGGACG GCGTCAGCGG GAGAGCCGTT GTAAGGCGGC AGACTTTGCT

      2530        2540        2550        2560        2570        2580
CATGTTACCG ATGCTATTCG GAAGAACGGC AACTAAGCTG CCGGGTTTGA AACACGGATG

      2590        2600        2610        2620        2630        2640
ATCTCGCGGA GGGTAGCATG TTGATTGTAA CGATGACAGA GCGTTGCTGC CTGTGATCAC

      2650        2660        2670        2680        2690        2700
CGCGGTTTCA AAATCGGCTC CGTCGATACT ATGTTATACG CCAACTTTGA AAACAACTTT
```

```
      2710        2720        2730        2740        2750        2760
GAAAAAGCTG TTTTCTGGTA TTTAAGGTTT TAGAATGCAA GGAACAGTGA ATTGGAGTTC

      2770        2780        2790        2800        2810        2820
GTCTTGTTAT AATTAGCTTC TTGGGGTATC TTTAAATACT GTAGAAAAGA GGAAGGAAAT

      2830        2840        2850        2860        2870        2880
AATAAATGGC TAAAATGAGA ATATCACCGG AATTGAAAAA ACTGATCGAA AAATACCGCT

      2890        2900        2910        2920        2930        2940
GCGTAAAAGA TACGGAAGGA ATGTCTCCTG CTAAGGTATA TAAGCTGGTG GGAGAAAATG

      2950        2960        2970        2980        2990        3000
AAAACCTATA TTTAAAAATG ACGGACAGCC GGTATAAAGG GACCACCTAT GATGTGGAAC

      3010        3020        3030        3040        3050        3060
GGGAAAAGGA CATGATGCTA TGGCTGGAAG GAAAGCTGCC TGTTCCAAAG GTCCTGCACT

      3070        3080        3090        3100        3110        3120
TTGAACGGCA TGATGGCTGG AGCAATCTGC TCATGAGTGA GGCCGATGGC GTCCTTTGCT

      3130        3140        3150        3160        3170        3180
CGGAAGAGTA TGAAGATGAA CAAAGCCCTG AAAAGATTAT CGAGCTGTAT GCGGAGTGCA

      3190        3200        3210        3220        3230        3240
TCAGGCTCTT TCACTCCATC GACATATCGG ATTGTCCCTA TACGAATAGC TTAGACAGCC

      3250        3260        3270        3280        3290        3300
GCTTAGCCGA ATTGGATTAC TTACTGAATA ACGATCTGGC CGATGTGGAT TGCGAAAACT

      3310        3320        3330        3340        3350        3360
GGGAAGAAGA CACTCCATTT AAAGATCCGC GCGAGCTGTA TGATTTTTTA AAGACGGAAA

      3370        3380        3390        3400        3410        3420
AGCCCGAAGA GGAACTTGTC TTTTCCCACG GCGACCTGGG AGACAGCAAC ATCTTTGTGA

      3430        3440        3450        3460        3470        3480
AAGATGGCAA AGTAAGTGGC TTTATTGATC TTGGGAGAAG CGGCAGGGCG GACAAGTGGT

      3490        3500        3510        3520        3530        3540
ATGACATTGC CTTCTGCGTC CGGTCGATCA GGGAGGATAT CGGGGAAGAA CAGTATGTCG

      3550        3560        3570        3580        3590        3600
AGCTATTTTT TGACTTACTG GGGATCAAGC CTGATTGGGA GAAAATAAAA TATTATATTT

      3610        3620        3630        3640        3650        3660
TACTGGATGA ATTGTTTTAG TACCTAGAAT GCATGACCAA AATCCCTTAA CGTGAGTTTT
```

```
      3670        3680        3690        3700        3710        3720
CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTT

      3730        3740        3750        3760        3770        3780
TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT

      3790        3800        3810        3820        3830        3840
TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA

      3850        3860        3870        3880        3890        3900
TACCAAATAC TGTCCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG

      3910        3920        3930        3940        3950        3960
CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA

      3970        3980        3990        4000        4010        4020
AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCGG

      4030        4040        4050        4060        4070        4080
GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGAACTGA

      4090        4100        4110        4120        4130        4140
GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGGCGGACA

      4150        4160        4170        4180        4190        4200
GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA

      4210        4220        4230        4240        4250        4260
ACGCCTGGTA TCTTTATAGT CCTGTCGGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT

      4270        4280        4290        4300        4310        4320
TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTTAC

      4330        4340        4350        4360        4370        4380
GGTTCCTGGC CTTTTGCTGG CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT

      4390        4400        4410        4420        4430        4440
CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA

      4450        4460        4470        4480        4490        4500
CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAGAGCG CCTGATGCGG TATTTTCTCC

      4510        4520        4530        4540        4550        4560
TTACGCATCT GTGCGGTATT TCACACCGCA TATGGTGCAC TCTCAGTACA ATCTGCTCTG

      4570        4580        4590        4600        4610        4620
ATGCCGCATA GTTAAGCCAG TATACACTCC GCTATCGCTA CGTGACTGGG TCATGGCTGC
```

```
      4630        4640        4650        4660        4670        4680
GCCCCGACAC CCGCCAACAC CCGCTGACGC GCCCTGACGG GCTTGTCTGC TCCCGGCATC

      4690        4700        4710        4720        4730        4740
CGCTTACAGA CAAGCTGTGA CCGTCTCCGG GAGCTGCATG TGTCAGAGGT TTTCACCGTC

      4750        4760        4770        4780        4790        4800
ATCACCGAAA CGCGCGAGGC AGGGTGCCTT GATGTGGGCG CCGGCGGTCG AGTGGCGACG

      4810        4820        4830        4840        4850        4860
GCGCGGCTTG TCCGCGCCCT GGTAGATTGC CTGGCCGTAG GCCAGCCATT TTTGAGCGGC

      4870        4880        4890        4900        4910        4920
CAGCGGCCGC GATAGGCCGA CGCGAAGCGG CGGGGCGTAG GGAGCGCAGC GACCGAAGGG

      4930        4940        4950        4960        4970        4980
TAGGCGCTTT TTGCAGCTCT TCGGCTGTGC GCTGGCCAGA CAGTTATGCA CAGGCCAGGC

      4990        5000        5010        5020        5030        5040
GGGTTTTAAG AGTTTTAATA AGTTTTAAAG AGTTTTAGGC GGAAAAATCG CCTTTTTTCT

      5050        5060        5070        5080        5090        5100
CTTTTATATC AGTCACTTAC ATGTGTGACC GGTTCCCAAT GTACGGCTTT GGGTTCCCAA

      5110        5120        5130        5140        5150        5160
TGTACGGGTT CCGGTTCCCA ATGTACGGCT TTGGGTTCCC AATGTACGTG CTATCCACAG

      5170        5180        5190        5200        5210        5220
GAAAGAGACC TTTTCGACCT TTTTCCCCTG CTAGGGCAAT TTGCCCTAGC ATCTGCTCCG

      5230        5240        5250        5260        5270        5280
TACATTAGGA ACCGGCGGAT GCTTCGCCCT CGATCAGGTT GCGGTAGCGC ATGACTAGGA

      5290        5300        5310        5320        5330        5340
TCGGGCCAGC CTGCCCCGCC TCCTCCTTCA AATCGTACTC CGGCAGGTCA TTTGACCCGA

      5350        5360        5370        5380        5390        5400
TCAGCTTGCG CACGGTGAAA CAGAACTTCT TGAACTCTCC GGCGCTGCCA CTGCGTTCGT

      5410        5420        5430        5440        5450        5460
AGATCGTCTT GAACAACCAT CTGGCTTCTG CCTTGCCTGC GGCGCGGCGT GCCAGGCGGT

      5470        5480        5490        5500        5510        5520
AGAGAAAACG GCCGATGCCG GGATCGATCA AAAAGTAATC GGGGTGAACC GTCAGCACGT

      5530        5540        5550        5560        5570        5580
CCGGGTTCTT GCCTTCTGTG ATCTCGCGGT ACATCCAATC AGCTAGCTCG ATCTCGATGT
```

```
      5590        5600        5610        5620        5630        5640
ACTCCGGCCG CCCGGTTTCG CTCTTTACGA TCTTGTAGCG GCTAATCAAG GCTTCACCCT

      5650        5660        5670        5680        5690        5700
CGGATACCGT CACCAGGCGG CCGTTCTTGG CCTTCTTCGT ACGCTGCATG GCAACGTGCG

      5710        5720        5730        5740        5750        5760
TGGTGTTTAA CCGAATGCAG GTTTCTACCA GGTCGTCTTT CTGCTTTCCG CCATCGGCTC

      5770        5780        5790        5800        5810        5820
GCCGGCAGAA CTTGAGTACG TCCGCAACGT GTGGACGGAA CACGCGGCCG GGCTTGTCTC

      5830        5840        5850        5860        5870        5880
CCTTCCCTTC CCGGTATCGG TTCATGGATT CGGTTAGATG GGAAACCGCC ATCAGTACCA

      5890        5900        5910        5920        5930        5940
GGTCGTAATC CCACACACTG GCCATGCCGG CCGGCCCTGC GGAAACCTCT ACGTGCCCGT

      5950        5960        5970        5980        5990        6000
CTGGAAGCTC GTAGCGGATC ACCTCGCCAG CTCGTCGGTC ACGCTTCGAC AGACGGAAAA

      6010        6020        6030        6040        6050        6060
CGGCCACGTC CATGATGCTG CGACTATCGC GGGTGCCCAC GTCATAGAGC ATCGGAACGA

      6070        6080        6090        6100        6110        6120
AAAAATCTGG TTGCTCGTCG CCCTTGGGCG GCTTCCTAAT CGACGGCGCA CCGGCTGCCG

      6130        6140        6150        6160        6170        6180
GCGGTTGCCG GGATTCTTTG CGGATTCGAT CAGCGGCCGC TTGCCACGAT TCACCGGGGC

      6190        6200        6210        6220        6230        6240
GTGCTTCTGC CTCGATGCGT TGCCGCTGGG CGGCCTGCGC GGCCTTCAAC TTCTCCACCA

      6250        6260        6270        6280        6290        6300
GGTCATCACC CAGCGCCGCG CCGATTTGTA CCGGGCCGGA TGGTTTGCGA CCGTCACGCC

      6310        6320        6330        6340        6350        6360
GATTCCTCGG GCTTGGGGGT TCCAGTGCCA TTGCAGGGCC GGCAGACAAC CCAGCCGCTT

      6370        6380        6390        6400        6410        6420
ACGCCTGGCC AACCGCCCGT TCCTCCACAC ATGGGGCATT CCACGGCGTC GGTGCCTGGT

      6430        6440        6450        6460        6470        6480
TGTTCTTGAT TTTCCATGCC GCCTCCTTTA GCCGCTAAAA TTCATCTACT CATTTATTCA

      6490        6500        6510        6520        6530        6540
TTTGCTCATT TACTCTGGTA GCTGCGCGAT GTATTCAGAT AGCAGCTCGG TAATGGTCTT
```

```
      6550        6560        6570        6580        6590        6600
GCCTTGGCGT ACCGCGTACA TCTTCAGCTT GGTGTGATCC TCCGCCGGCA ACTGAAAGTT

      6610        6620        6630        6640        6650        6660
GACCCGCTTC ATGGCTGGCG TGTCTGCCAG GCTGGCCAAC GTTGCAGCCT TGCTGCTGCG

      6670        6680        6690        6700        6710        6720
TGCGCTCGGA CGGCCGGCAC TTAGCGTGTT TGTGCTTTTG CTCATTTTCT CTTTACCTCA

      6730        6740        6750        6760        6770        6780
TTAACTCAAA TGAGTTTTGA TTTAATTTCA GCGGCCAGCG CCTGGACCTC GCGGGCAGCG

      6790        6800        6810        6820        6830        6840
TCGCCCTCGG GTTCTGATTC AAGAACGGTT GTGCCGGCGG CGGCAGTGCC TGGGTAGCTC

      6850        6860        6870        6880        6890        6900
ACGCGCTGCG TGATACGGGA CTCAAGAATG GGCAGCTCGT ACCCGGCCAG CGCCTCGGCA

      6910        6920        6930        6940        6950        6960
ACCTCACCGC CGATGCGCGT GCCTTTGATC GCCCGCGACA CGACAAAGGC CGCTTGTAGC

      6970        6980        6990        7000        7010        7020
CTTCCATCCG TGACCTCAAT GCGCTGCTTA ACCAGCTCCA CCAGGTCGGC GGTGGCCCAT

      7030        7040        7050        7060        7070        7080
ATGTCGTAAG GGCTTGGCTG CACCGGAATC AGCACGAAGT CGGCTGCCTT GATCGCGGAC

      7090        7100        7110        7120        7130        7140
ACAGCCAAGT CCGCCGCCTG GGGCGCTCCG TCGATCACTA CGAAGTCGCG CCGGCCGATG

      7150        7160        7170        7180        7190        7200
GCCTTCACGT CGCGGTCAAT CGTCGGGCGG TCGATGCCGA CAACGGTTAG CGGTTGATCT

      7210        7220        7230        7240        7250        7260
TCCCGCACGG CCGCCCAATC GCGGGCACTG CCCTGGGGAT CGGAATCGAC TAACAGAACA

      7270        7280        7290        7300        7310        7320
TCGGCCCCGG CGAGTTGCAG GGCGCGGGCT AGATGGGTTG CGATGGTCGT CTTGCCTGAC

      7330        7340        7350        7360        7370        7380
CCGCCTTTCT GGTTAAGTAC AGCGATAACC TTCATGCGTT CCCCTTGCGT ATTTGTTTAT

      7390        7400        7410        7420        7430        7440
TTACTCATCG CATCATATAC GCAGCGACCG CATGACGCAA GCTGTTTTAC TCAAATACAC

      7450        7460        7470        7480        7490        7500
ATCACCTTTT TAGACGGCGG CGCTCGGTTT CTTCAGCGGC CAAGCTGGCC GGCCAGGCCG
```

```
      7510        7520        7530        7540        7550        7560
CCAGCTTGGC ATCAGACAAA CCGGCCAGGA TTTCATGCAG CCGCACGGTT GAGACGTGCG

      7570        7580        7590        7600        7610        7620
CGGGCGGCTC GAACACGTAC CCGGCCGCGA TCATCTCCGC CTCGATCTCT TCGGTAATGA

      7630        7640        7650        7660        7670        7680
AAAACGGTTC GTCCTGGCCG TCCTGGTGCG GTTTCATGCT TGTTCCTCTT GGCGTTCATT

      7690        7700        7710        7720        7730        7740
CTCGGCGGCC GCCAGGGCGT CGGCCTCGGT CAATGCGTCC TCACGGAAGG CACCGCGCCG

      7750        7760        7770        7780        7790        7800
CCTGGCCTCG GTGGGCGTCA CTTCCTCGCT GCGCTCAAGT GCGCGGTACA GGGTCGAGCG

      7810        7820        7830        7840        7850        7860
ATGCACGCCA AGCAGTGCAG CCGCCTCTTT CACGGTGCGG CCTTCCTGGT CGATCAGCTC

      7870        7880        7890        7900        7910        7920
GCGGGCGTGC GCGATCTGTG CCGGGGTGAG GGTAGGGCGG GGGCCAAACT TCACGCCTCG

      7930        7940        7950        7960        7970        7980
GGCCTTGGCG GCCTCGCGCC CGCTCCGGGT GCGGTCGATG ATTAGGGAAC GCTCGAACTC

      7990        8000        8010        8020        8030        8040
GGCAATGCCG GCGAACACGG TCAACACCAT GCGGCCGGCC GGCGTGGTGG TGTCGGCCCA

      8050        8060        8070        8080        8090        8100
CGGCTCTGCC AGGCTACGCA GGCCCGCGCC GGCCTCCTGG ATGCGCTCGG CAATGTCCAG

      8110        8120        8130        8140        8150        8160
TAGGTCGCGG GTGCTGCGGG CCAGGCGGTC TAGCCTGGTC ACTGTCACAA CGTCGCCAGG

      8170        8180        8190        8200        8210        8220
GCGTAGGTGG TCAAGCATCC TGGCCAGCTC CGGGCGGTCG CGCCTGGTGC CGGTGATCTT

      8230        8240        8250        8260        8270        8280
CTCGGAAAAC AGCTTGGTGC AGCCGGCCGC GTGCAGTTCG GCCCGTTGGT TGGTCAAGTC

      8290        8300        8310        8320        8330        8340
CTGGTCGTCG GTGCTGACGC GGGCATAGCC CAGCAGGCCA GCGGCGGCGC TCTTGTTCAT

      8350        8360        8370        8380        8390        8400
GGCGTAATGT CTCCGGTTCT AGTCGCAAGT ATTCTACTTT ATGCGACTAA AACACGCGAC

      8410        8420        8430        8440        8450        8460
AAGAAAACGC CAGGAAAAGG GCAGGGCGGC AGCCTGTCGC GTAACTTAGG ACTTGTGCGA
```

```
      8470        8480        8490        8500        8510        8520
CATGTCGTTT TCAGAAGACG GCTGCACTGA ACGTCAGAAG CCGACTGCAC TATAGCAGCG

      8530        8540        8550        8560        8570        8580
GAGGGGTTGG ATCAAAGTAC TTTGATCCCG AGGGGAACCC TGTGGTTGGC ATGCACATAC

      8590        8600        8610        8620        8630        8640
AAATGGACGA ACGGATAAAC CTTTTCACGC CCTTTTAAAT ATCCGTTATT CTAATAAACG

      8650        8660        8670        8680        8690        8700
CTCTTTTCTC TTAGGTTTAC CCGCCAATAT ATCCTGTCAA ACACTGATAG TTTAAACTGA

      8710        8720        8730        8740        8750        8760
AGGCGGGAAA CGACAATCTG ATCCAAGCTC AAGCTGCTCT AGCATTCGCC ATTCAGGCTG

      8770        8780        8790        8800        8810        8820
CGCAACTGTT GGGAAGGGCG ATCGGTGCGG GCCTCTTCGC TATTACGCCA GCTGGCGAAA

      8830        8840        8850        8860        8870        8880
GGGGGATGTG CTGCAAGGCG ATTAAGTTGG GTAACGCCAG GGTTTTCCCA GTCACGACGT

      8890        8900        8910        8920        8930        8940
TGTAAAACGA CGGCCAGTGC CAAGCTTGCA TGCGATCTTC AACCACCTGT GCTAGCTACT

      8950        8960        8970        8980        8990        9000
CCACTGCTCC ATAGGCAATC ATCAATCAGT AATCCGTTCT GAAAAGAAGA TATAGGTGTG

      9010        9020        9030        9040        9050        9060
CGCAATCAGG AACGTTCTAG TTCGTGCTAG AAATCAGCAG CTCCTAAGTT AGCATCTCGA

      9070        9080        9090        9100        9110        9120
TGAACTTAAA TGCTCGCTGC GGGCGTCCGG CGGAGATGAA GTTTGTGATA AACTTGGTCA

      9130        9140        9150        9160        9170        9180
TGACATTCAT ATATGTGCCT GGTGTACGGA GTAGTTCATC AGCAAACATA CACCTACTTC

      9190        9200        9210        9220        9230        9240
TACCTTATCC ATTTGGATTG CTCATGGCGG CTTTGATATG GAATTTGTAA TGAACTTGGT

      9250        9260        9270        9280        9290        9300
TATGACTTAT GACATACTGA TACTCGTAAC ATTCATAGAT ACTGACATAA ATTCATCAAC

      9310        9320        9330        9340        9350        9360
TACAATAGAT GAGATGGCTA GTCTTAGTAG AACAGTAGTC TCTCTTTCCG GCTTGCTCCA

      9370        9380        9390        9400        9410        9420
TTGGCTGATG ACGATGAACA ACTCGGACTC ATTGATTCCA GCATTATCTG ATTCTCGCAT
```

```
      9430        9440        9450        9460        9470        9480
TTCGAGGTCC GGATTAGGGT CTCACCGAGA TGTGGATAGA ATTGCCATGT CAGGAATTGA

      9490        9500        9510        9520        9530        9540
AGGAGGACGA GCCATATGTG CATATACATG ACGGGAGATC AAGCGGCCAG TCAAGAGGCT

      9550        9560        9570        9580        9590        9600
AACTGCAACC CTATTATATA CGATCAGCCT GCTAGAACAC GTAGCACTGT CTTTTTTGTC

      9610        9620        9630        9640        9650        9660
TGAACTCTGA AGATGAAAGG TTCAGAGAAA TGGCTCGCCT TATCCAAGCC GGCGATGGAT

      9670        9680        9690        9700        9710        9720
GGAGGAGGAG GTAGCCGGCG CCCGCCTCAG GCAGTCGTCG CGATCACGCC GCCGCATCCC

      9730        9740        9750        9760        9770        9780
GTCGCCTTGG AGACCGGGCC CCGACGCGGC CGACGCGGCG CCTACGTGGC CATGCTTTAT

      9790        9800        9810        9820        9830        9840
TGCCTTATCC ATATCCACGC CATTTATTGT GGTCGTCTCT CCTGATCATT CTCATTCCCC

      9850        9860        9870        9880        9890        9900
TGCCACGGTG ACCGTGCCCC CGGTGTTCTA TATATGCCCC CCGACGTCGA GGTCATTCGC

      9910        9920        9930        9940        9950        9960
CACGAACACA TCGATCATCC ATCATCTACA AGAGATCGTC TAGAATTATT ACATCAAAAC

      9970        9980        9990       10000       10010       10020
AAAAGATGAA GAATACCAGC TCGTTGTGTT TGCTTCTCCT CGTGGTGCTT TGCTCACTAA

     10030       10040       10050       10060       10070       10080
CATGCAATTC GGGACAAGCA AATTCCGATT CCGAGTGTCC GCTCAGCCAC GACGGATACT

     10090       10100       10110       10120       10130       10140
GCCTCCATGA TGGGGTCTGC ATGTACATTG AGGCCCTGGA CAAGTACGCG TGTAACTGCG

     10150       10160       10170       10180       10190       10200
TGGTTGGCTA TATCGGCGAA AGGTGCCAGT ATCGGGACTT GAAGTGGTGG GAGCTTCGCT

     10210       10220       10230       10240       10250       10260
GATAGGTACC GAGCTCGGGC TCAAGCCCTA AACTGAACGG GATAGTCATG CTCAAACCAG

     10270       10280       10290       10300       10310       10320
TTTCTACACG GCAAGAATTT ACTGATTCTT ATACTTTTGC AGTCAATTAA ATTATGGTTT

     10330       10340       10350       10360       10370       10380
TTATATATGT AATTTTGTAT CCGATTGTAG CGTTCGAATA AGTAGGCAGG CTCTCTAGCC
```

```
     10390       10400       10410       10420       10430       10440
TCTAGGTTAA TTGCGGGGCA TATGTAGCTT GCCAGTTAAT TGTGTTTGTA TCACGCAGTT

     10450       10460       10470       10480       10490       10500
TGTAACCGTT GGTGCAATAT ATAATGTCAG GTTCAGGATG CAGTAAAAAA TCATACTGCA

     10510       10520       10530       10540       10550       10560
CCGATCAGTG AGTTTTTATG AATTC..... .......... .......... ..........
```

Supplementary Fig. 1. pJKS131 vector sequence to transfer the *EGF* to the rice genome.