

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
8 July 2004 (08.07.2004)

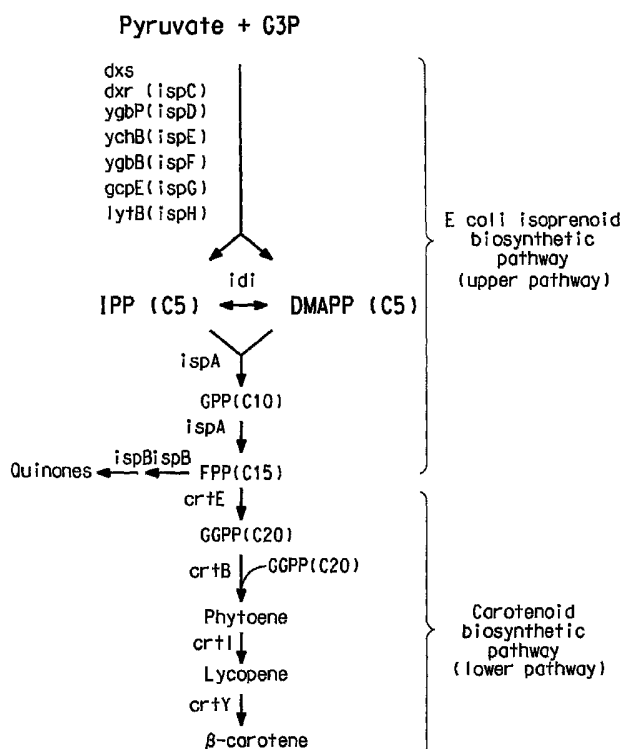
PCT

(10) International Publication Number
WO 2004/056975 A2

- (51) International Patent Classification⁷: **C12N** Road, Wilmington, DE 19803 (US). **SUH, Wonchul**; 8 Salina Court, Piersons Ridge, Hockessin, DE 19707 (US).
- (21) International Application Number: PCT/US2003/041812 (74) Agent: **FELTHAM, S., Neil**; E.I. Dupont de Nemours and Company, Legal Patent Records Center, 4417 Lancaster Pike, Wilmington, DE 19805 (US).
- (22) International Filing Date: 19 December 2003 (19.12.2003)
- (25) Filing Language: English (81) Designated States (*national*): AU, CA, JP, NO.
- (26) Publication Language: English (84) Designated States (*regional*): European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR).
- (30) Priority Data: 60/434,618 19 December 2002 (19.12.2002) US
- (71) Applicant: **E.I. DU PONT DE NEMOURS AND COMPANY** [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).
- (72) Inventors: **CHENG, Qiong**; 4 Collins Drive, Wilmington, DE 19803 (US). **ROUVIERE, Pierre, E.**; 737 Taunton
- Published:**
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: INCREASING CAROTENOID PRODUCTION IN BACTERIA VIA CHROMOSOMAL INTEGRATION

Isoprenoid Pathway in E. coli



(57) Abstract: The present invention relates to carotenoid overproducing bacteria. The genes of the isoprenoid pathway in the bacterial hosts of the invention have been engineered such that certain genes are either up-regulated or down regulated resulting in the production of carotenoid compounds at a higher level than is found in the un-modified host. Genes that may be up-regulated include the *dxs*, *idi*, *ispB*, *lytB* and *ygbBP* genes. Additionally it has been found that a partial disruption of the *yjeR* gene has the effect of enhancing carotenoid production.

WO 2004/056975 A2

TITLE

INCREASING CAROTENOID PRODUCTION IN BACTERIA VIA CHROMOSOMAL INTEGRATION

This application claims the benefit of U.S. Provisional Application
5 No. 60/434,618 filed December 19, 2002.

FIELD OF THE INVENTION

This invention is in the field of microbiology. More specifically, this
invention pertains to carotenoid overproducing bacterial strains.

BACKGROUND OF THE INVENTION

10 Carotenoids are pigments that are ubiquitous throughout nature
and synthesized by all oxygen evolving photosynthetic organisms and in
some heterotrophic growing bacteria and fungi. Industrial uses of
carotenoids include pharmaceuticals, food supplements, electro-optic
applications, animal feed additives, and colorants in cosmetics, to mention
15 a few. Because animals are unable to synthesize carotenoids *de novo*,
they must obtain them by dietary means. Thus, manipulation of
carotenoid production and composition in plants or bacteria can provide
new or improved sources of carotenoids.

Carotenoids come in many different forms and chemical structures.
20 Most naturally occurring carotenoids are hydrophobic tetraterpenoids
containing a C₄₀ methyl-branched hydrocarbon backbone derived from
successive condensation of eight C₅ isoprene units (isopentenyl
pyrophosphate, IPP). In addition, novel carotenoids with longer or shorter
backbones occur in some species of nonphotosynthetic bacteria.

25 The genetics of carotenoid pigment biosynthesis are well-known
(Armstrong et al., *J. Bact.*, 176: 4795-4802 (1994); Armstrong et al., *Annu.*
Rev. Microbiol., 51:629-659 (1997)). This pathway is extremely well-
studied in the Gram-negative, pigmented bacteria of the genera *Pantoea*,
formerly known as *Erwinia*. In both *E. herbicola* EHO-10 (ATCC 39368)
30 and *E. uredovora* 20D3 (ATCC 19321), the *crt* genes are clustered in two
operons, *crt Z* and *crt EXYIB* (US 5,656,472; US 5,545,816;
US 5,530,189; US 5,530,188; and US 5,429,939).

Isoprenoids constitute the largest class of natural products in
nature, and serve as precursors for sterols (eukaryotic membrane
35 stabilizers), gibberellins and abscisic acid (plant hormones),
menaquinone, plastoquinones, and ubiquinone (used as carriers for
electron transport), tetrapyrroles as well as carotenoids and the phytol
side chain of chlorophyll (pigments for photosynthesis). All isoprenoids

are synthesized via a common metabolic precursor, isopentenyl pyrophosphate (IPP). Until recently, the biosynthesis of IPP was generally assumed to proceed exclusively from acetyl-CoA via the classical mevalonate pathway. However, the existence of an alternative,
5 mevalonate-independent pathway for IPP formation has been characterized in eubacteria and green algae.

E. coli contains genes that encode enzymes of the mevalonate-independent pathway of isoprenoid biosynthesis (Figure 1). In this pathway, isoprenoid biosynthesis starts with the condensation of pyruvate
10 with glyceraldehyde-3-phosphate (G3P) to form deoxy-D-xylulose via the enzyme encoded by the *dxs* gene. A host of additional enzymes are then used in subsequent sequential reactions, converting deoxy-D-xylulose to the final C5 isoprene product, isopentenyl pyrophosphate (IPP). IPP is converted to the isomer dimethylallyl pyrophosphate (DMAPP) via the
15 enzyme encoded by the *idi* gene. IPP is condensed with DMAPP to form C10 geranyl pyrophosphate (GPP) which is then elongated to C15 farnesyl pyrophosphate (FPP).

FPP synthesis is common in both carotenogenic and non-carotenogenic bacteria. *E. coli* does not normally contain the genes
20 necessary for conversion of FPP to β -carotene (Figure 1). Enzymes in the subsequent carotenoid pathway generate carotenoid pigments from the FPP precursor and can be divided into two categories: carotene backbone synthesis enzymes and subsequent modification enzymes. The backbone
synthesis enzymes include geranyl geranyl pyrophosphate synthase
25 (CrtE), phytoene synthase (CrtB), phytoene dehydrogenase (CrtI) and lycopene cyclase (CrtY/L), etc. The modification enzymes include ketolases, hydroxylases, dehydratases, glycosylases, etc.

E. coli is a convenient host for heterologous carotenoid production. Most of the carotenogenic genes from bacteria, fungi and higher plants
30 can be functionally expressed in *E. coli* (Sandmann, G., *Trends in Plant Science*, 6:14-17 (2001)). Furthermore, many genetic tools are available for use in *E. coli*, a production host often used for large-scale bioprocesses.

Engineering *E. coli* for increased carotenoid production has
35 previously focused on overexpression of key isoprenoid pathway genes from multi-copy plasmids. It has been postulated that the total amount of carotenoids produced in non-carotenogenic hosts is limited by the availability of terpenoid precursors (Albrecht et al., *Biotechnol. Lett.*,

21:791-795 (1999)). Several studies have reported between a 1.5X and 50X increase in carotenoid formation in such *E. coli* systems upon cloning and transformation of plasmids encoding isopentenyl diphosphate isomerase (*idi*), deoxy-D-xylulose-5-phosphate (DXP) synthase (*dxs*), DXP reductoisomerase (*dxr*) from various sources (Kim, S., and Keasling, J., *Biotech. Bioeng.*, 72:408-415 (2001); Mathews, P., and Wurtzel, E., *Appl. Microbiol. Biotechnol.*, 53:396-400 (2000); Harker, M., and Bramley, P., *FEBS Letter.*, 448:115-119 (1999); Misawa, N., and Shimada, H., *J. Biotechnol.*, 59:169-181 (1998); Liao et al., *Biotechnol. Bioeng.*, 62:235-241 (1999); and Misawa et al., *Biochem. J.*, 324:421-426 (1997)). In addition, it has also been reported that increasing isoprenoid precursor concentration may be lethal (Sandmann, G., *supra*).

The highest level of carotenoids produced to date in *E. coli* are around 1.57 mg/g dry cell weight (DCW). In contrast, engineered strains of *Candida utilis* produce 7.8 mg of lycopene per gram of dry cell weight of lycopene (Sandmann, *supra*). It has been speculated that the limits for carotenoid production in a non-carotenogenic host, such as *E. coli*, had been reached at the level of around 1.5 mg/g DCW due to carotenoid overload of the membranes, disrupting membrane functionality. Because of this, it has been suggested that the future focus of engineering *E. coli* for high levels of carotenoid production should be on formation of additional membranes (Albrecht et al., *supra*).

Most of the work to date in the metabolic engineering of isoprenoids has been done using carotenoids primarily because of the easy color screening. Engineering an increased supply of isoprenoid precursors for increased production of carotenoids is necessary. It has been shown that a rate-limiting step in carotenoid biosynthesis is the isomerization of IPP to DMAPP (Kajiwara et al., *Biochem. J.*, 423: 421-426 (1997)). It was also found that the conversion from FPP to GGPP is the first functional limiting step for the production of carotenoids in *E. coli* (Wang et al., *Biotechnol. Prog.*, 62: 235-241 (1999)). Transformation of *E. coli* for overexpression of the *dxs*, *dxr*, and *idi* genes was found to increase production of carotenoids by a factor of 3.5 (Albrecht et al., *supra*). To avoid competition from other pathways and to relieve the limiting steps, a GGPP synthase (*gps*) from *Archaeoglobus fulgidus* was cloned in a multi-copy expression vector and over-expressed in *E. coli*, along with the *E. coli idi* gene (Wang et al., *supra*). These examples show

that a multi-copy expression vector has been widely used for the metabolic engineering for the production of carotenoids.

The problem to be solved, therefore, is to engineer and provide microbial hosts which are capable of producing increased levels of carotenoids. Applicants have solved the stated problem by making modifications to the *E. coli* chromosome, increasing β -carotene production up to 6 mg per gram dry cell weight (6000 PPM), an increase of 30-fold over initial levels; with no lethal effect.

SUMMARY OF THE INVENTION

The invention provides a carotenoid overproducing bacteria comprising the genes encoding a functional carotenoid enzymatic biosynthetic pathway wherein the *dxs*, *idi* and *ygbBP* genes are overexpressed and wherein the *yjeR* gene is down regulated.

Additionally the invention provides a carotenoid overproducing bacteria comprising the genes encoding a functional carotenoid enzymatic biosynthetic pathway wherein the *dxs*, *idi*, *ygbBP* and *ispB* genes are overexpressed. Optionally the *lytB* gene may also be overexpressed to further enhance the carotenoid production.

In a preferred embodiment, the invention provides a carotenoid overproducing bacteria selected from the group consisting of a strain having the ATCC identification number PTA-4807 and a strain having the ATCC identification number PTA-4823

In another embodiment the invention provides a method for the production of a carotenoid comprising:

- a) growing the carotenoid overproducing bacteria of the invention the bacteria overexpressing at least one gene selected from the group consisting of *dxs*, *idi*, *ygbBP*, *ispB*, *lytB*, *dxr*, wherein *yjeR* is optionally downregulated, for a time sufficient to produce a carotenoid; and
- b) optionally recovering the carotenoid from the carotenoid overproducing bacteria of step (a).

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

Figure 1 outlines the isoprenoid and carotenoid biosynthetic pathways used for production of β -carotene in *E. coli*.

Figure 2 shows the strategy for chromosomal integration of promoter or full gene sequences and stacking the strong promoter-isoprenoid gene fusions.

Figure 3 shows PCR analysis of chromosomal insertions.

Figure 4 shows PCR analysis of chromosomal insertions.

Figure 5 shows PCR analysis of chromosomal insertions.

Figure 6 shows the plasmid map of pSUH5.

5 Figure 7 shows the plasmid map of pPCB15.

Figure 8 shows the strategy for creating *E. coli* Tn5 mutants which have increased carotenoid production.

Figure 9 shows increased β -carotene production from an *E. coli* Tn5 mutant.

10 Figure 10 shows insertion site of Tn5 in the Y15; *yjeR::Tn5* mutation.

Figure 11 shows β -carotene production by the engineered *E. coli* strains of the present invention.

15 Figure 12 shows bacteriophage P1 mediated transduction and parallel combinatorial stacking used in the optimization of β -carotene production.

The invention can be more fully understood from the following detailed description and the accompanying sequence descriptions, which form a part of this application.

20 The following sequences comply with 37 C.F.R. 1.821-1.825 ("Requirements for Patent Applications Containing Nucleotide Sequences and/or Amino Acid Sequence Disclosures - the Sequence Rules") and are consistent with World Intellectual Property Organization (WIPO) Standard ST.25 (1998) and the sequence listing requirements of the EPO and PCT
25 (Rules 5.2 and 49.5(a-bis), and Section 208 and Annex C of the Administrative Instructions). The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

Gene/Protein Product	Source	Nucleotide SEQ ID NO	Amino Acid SEQ ID NO
<i>CrtE</i>	<i>Pantoea stewartii</i>	1	2
<i>CrtX</i>	<i>Pantoea stewartii</i>	3	4
<i>CrtY</i>	<i>Pantoea stewartii</i>	5	6
<i>CrtI</i>	<i>Pantoea stewartii</i>	7	8
<i>CrtB</i>	<i>Pantoea stewartii</i>	9	10
<i>CrtZ</i>	<i>Pantoea stewartii</i>	11	12

Gene/Protein Product	Source	Nucleotide SEQ ID NO	Amino Acid SEQ ID NO
<i>dxs(16a)</i>	<i>Methylobacter 16a</i>	13	14
<i>lytB(16a)</i>	<i>Methylobacter 16a</i>	15	16
<i>dxr(16a)</i>	<i>Methylobacter 16a</i>	17	18

SEQ ID NOs:19-20 are oligonucleotide primers used to amplify the carotenoid biosynthesis genes from *P. stewartii*.

SEQ ID NOs:21-32 are oligonucleotide primers used to create chromosomal integration of the *T5* strong promoter (*P_{T5}*) upstream from *E. coli* isoprenoid genes in the present invention.

SEQ ID NO:33 is the nucleotide sequence of the *P_{T5}* promoter sequence inserted in pKD4 to create pSUH5.

SEQ ID NO:34-45 are oligonucleotide primers for creating *dxs(16a)*, *dxr(16a)*, and *lytB(16a)* gene insertions in the *E. coli* chromosome.

SEQ ID NO:46-62 are oligonucleotide primers used for screening to confirm correct insertion of chromosomal integrations in the present invention.

SEQ ID NO:63 is the nucleotide sequence of the *yjeR::Tn5* mutant gene.

SEQ ID NO:64 is the nucleotide sequence for plasmid pPCB15.

SEQ ID NO:65 is the nucleotide sequence for plasmid pKD46.

SEQ ID NO:66 is the nucleotide sequence for plasmid pSUH5.

BRIEF DESCRIPTION OF BIOLOGICAL DEPOSITS

The following biological deposit have been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the purposes of Patent Procedure:

Depositor Identification	Int'l. Depository	
Reference	Designation	Date of Deposit
Plasmid pCP20	ATCC# PTA-4455	June 13, 2002
<i>Methylobacter 16a</i>	ATCC# PTA-2402	August 22, 2000
WS#124 <i>E. coli</i> strain <i>P_{T5}-dxs P_{T5}-idi P_{T5}-ygbBP yjeR::Tn5</i> , pPCB15	ATCC# PTA-4807	November 20, 2002
WS#208 <i>E. coli</i> strain <i>P_{T5}-dxs P_{T5}-idi P_{T5}-ygbBP P_{T5}-ispB</i> , pDCQ108	ATCC# PTA-4823	November 26, 2002

As used herein, "ATCC" refers to the American Type Culture Collection International Depository Authority located at ATCC, 10801

University Blvd., Manassas, VA 20110-2209, USA. The "International Depository Designation" is the accession number to the culture on deposit with ATCC.

5 The listed deposits will be maintained in the indicated international depository for at least thirty (30) years and will be made available to the public upon the grant of a patent disclosing it. The availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by government action.

DETAILED DESCRIPTION OF THE INVENTION

10 In this disclosure, a number of terms and abbreviations are used. The following definitions are provided.

"Open reading frame" is abbreviated ORF.

"Polymerase chain reaction" is abbreviated PCR.

15 As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

20 The term "isoprenoid" or "terpenoid" refers to the compounds and any molecules derived from the isoprenoid pathway including 10 carbon terpenoids and their derivatives, such as carotenoids and xanthophylls.

25 A "carotene" refers to a hydrocarbon carotenoid. Carotene derivatives that contain one or more oxygen atoms, in the form of hydroxy-, methoxy-, oxo-, epoxy-, carboxy-, or aldehydic functional groups, or within glycosides, glycoside esters, or sulfates, are collectively known as "xanthophylls". Carotenoids are furthermore described as being acyclic, monocyclic, or bicyclic depending on whether the ends of the hydrocarbon backbones have been cyclized to yield aliphatic or cyclic ring structures (G. Armstrong, (1999) In Comprehensive Natural Products
30 Chemistry, Elsevier Press, volume 2, pp 321-352).

The terms " λ -Red recombination system", " λ -Red system" and " λ -Red recombinase" are used interchangeably to describe a group of enzymes encoded by the bacteriophage λ genes *exo*, *bet*, and *gam*. The enzymes encoded by the three genes work together to increase the rate of
35 homologous recombination in *E. coli*, an organism generally considered to have a relatively low rate of homologous recombination; especially when using linear integration cassettes. The λ -Red system facilitates the ability to use short regions of homology (10-50 bp) flanking linear double-

stranded (ds) DNA fragments for homologous recombination. In the present method, the λ -Red genes are expressed on helper plasmid pKD46 (Datsenko and Wanner, *PNAS*, 97:6640-6645 (2000); SEQ ID NO:65).

5 The terms "*Methylobacter* 16a strain" and "*Methylobacter* 16a" are used interchangeably and refer to a bacterium (ATCC PTA-2402) of a physiological group of bacteria known as methylotrophs, which are unique in their ability to utilize methane as a sole carbon and energy source.

 The term "*yjeR*" refers to the oligo-ribonuclease gene locus.

10 The term "Dxs" refers to the enzyme D-1-deoxyxylulose 5-phosphate encoded by the *dxs* gene which catalyzes the condensation of pyruvate and D-glyceraldehyde 3-phosphate to D-1-deoxyxylulose 5-phosphate (DOXP).

 The terms "Dxr" or "IspC" refer to the enzyme DOXP reductoisomerase encoded by the *dxr* or *ispC* gene that catalyzes the simultaneous reduction and isomerization of DOXP to 2-C-methyl-D-erythritol-4-phosphate. The names of the gene, *dxr* or *ispC*, are used interchangeably in this application. The names of gene product, Dxr or IspC are used interchangeably in this application.

20 The term "YgbP" or "IspD" and refers to the enzyme encoded by the *ygbB* or *ispD* gene that catalyzes the CTP-dependent cytidylation of 2-C-methyl-D-erythritol-4-phosphate to 4-diphosphocytidyl-2C-methyl-D-erythritol. The names of the gene, *ygbP* or *ispD*, are used interchangeably in this application. The names of gene product, YgbP or IspD are used interchangeably in this application.

25 The term "YchB" or "IspE" and refers to the enzyme encoded by the *ychB* or *ispE* gene that catalyzes the ATP-dependent phosphorylation of 4-diphosphocytidyl-2C-methyl-D-erythritol to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate. The names of the gene, *ychB* or *ispE*, are used interchangeably in this application. The names of gene product, YchB or IspE are used interchangeably in this application.

30 The term "YgbB" or "IspF" refers to the enzyme encoded by the *ygbB* or *ispF* gene that catalyzes the cyclization with loss of CMP of 4-diphosphocytidyl-2C-methyl-D-erythritol to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate to 2C-methyl-D-erythritol-2,4-cyclodiphosphate. The names of the gene, *ygbB* or *ispF*, are used interchangeably in this application. The names of gene product, YgbB or IspF are used interchangeably in this application.

The term "GcpE" or "IspG" refers to the enzyme encoded by the *gcpE* or *ispG* gene that is involved in conversion of 2C-methyl-D-erythritol-2,4-cyclodiphosphate to 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate. The names of the gene, *gcpE* or *ispG*, are used interchangeably in this application. The names of gene product, GcpE or IspG are used interchangeably in this application.

The term "LytB" or "IspH" refers to the enzyme encoded by the *lytB* or *ispH* gene and is involved in conversion of 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate to isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). The names of the gene, *lytB* or *ispH*, are used interchangeably in this application. The names of gene product, LytB or IspH are used interchangeably in this application.

The term "Idi" refers to the enzyme isopentenyl diphosphate isomerase encoded by the *idi* gene that converts isopentenyl diphosphate to dimethylallyl diphosphate.

The term "IspA" refers to the enzyme farnesyl pyrophosphate (FPP) synthase encoded by the *ispA* gene.

The term "IspB" refers to the enzyme octaprenyl diphosphate synthase, which supplies the precursor of the side chain of the isoprenoid quinones encoded by the *ispB* gene.

The term "pPCB15" refers to the plasmid (Figure 7; SEQ ID NO:64) containing β -carotene synthesis genes *Pantoea crtEXYIB*, using as a reporter plasmid for monitoring β -carotene production in *E. coli* genetically engineered via the present method.

The term "pKD46" refers to the plasmid (SEQ ID NO:65; Datsenko and Wanner, *supra*) having GenBank® Accession number AY048746. Plasmid pKD46 expresses the components of the λ -Red Recombinase system.

The term "pSUH5" refers to the plasmid (Figure 6; SEQ ID NO:66) that was constructed by cloning a phage *T5* promoter (*P_{T5}*) region into the *NdeI* restriction endonuclease site of pKD4 (Datsenko and Wanner, *supra*). It was used as a template plasmid for PCR amplification of a fused kanamycin selectable marker/phage *T5* promoter linear DNA nucleotide.

The term "triple homologous recombination" in the present invention refers to a genetic recombination between two linear (PCR-generated) DNA fragments and the target chromosome via their

homologous sequences resulting in chromosomal integration of the two linear nucleic acid fragments into the target chromosome.

5 The term "homology arm" refers to a nucleotide sequence which enables homologous recombination between two nucleic acids having substantially the same nucleotide sequence in a particular region of two different nucleic acids. The preferred size range of the nucleotide sequence of the homology arm is from about 10 to about 100 nucleotides.

10 The term "site-specific recombinase" is used in the present invention to describe a system comprised of one or more enzymes which recognize specific nucleotide sequences (recombination target sites) and which catalyze recombination between the recombination target sites. Site-specific recombination provides a method to rearrange, delete, or introduce exogenous DNA. Examples of site-specific recombinases and their associated recombination target sites are: *Cre-lox*, *FLP/FRT*, *R/RS*,
15 *Gin/gix*, *Xer/dif*, *Int/att*, a pSR1 system, a *cer* system, and a *fim* system. The present invention illustrates the use of a site-specific recombinase to remove selectable markers. Antibiotic resistance markers, flanked on both sides by *FRT* recombination target sites, are removed by expression of the FLP site-specific recombinase.

20 The terms "stacking", "combinatorial stacking", "chromosomal stacking", and "trait stacking" are used interchangeably and refer to the repeated process of stacking multiple genetic traits into one *E. coli* host using bacteriophage P1 transduction in combination with the site-specific recombinase system for removal of selection markers (Figure 12).

25 The term "parallel combinatorial fashion" refers to the P1 transduction with the P1 lysate mixture made from various donor cells, so that multiple genetic traits can move the recipient cell in parallel.

The term "integration cassette" and "recombination element" refers to a linear nucleic acid construct useful for the transformation of a
30 recombination proficient bacterial host. Recombination elements of the invention may include a variety of genetic elements such as selectable markers, expressible DNA fragments, and recombination regions having homology to regions on a bacterial chromosome or on other recombination elements. Expressible DNA fragments can include
35 promoters, coding sequences, genes, and other regulatory elements specifically engineered into the recombination element to impart a desired phenotypic change upon recombination.

The term "expressible DNA fragment" means any DNA that influences phenotypic changes in the host cell. An "expressible DNA fragment" may include for example, DNA comprising regulatory elements, isolated promoters, open reading frames, coding sequences, genes, or combinations thereof.

The term "pDCQ108" refers to the plasmid containing β -carotene synthesis genes *Pantoea crtEXYIB* used as a reporter plasmid for monitoring β -carotene production in *E. coli* that were genetically engineered via the present method (ATCC PTA-4823).

The terms " P_{T5} promoter" and "phage T5 promoter" are used interchangeably and refer to the nucleotide sequence that comprises the -10 and -35 consensus sequences, lactose operator (*lacO*), and ribosomal binding site (*rbs*) from phage T5 (SEQ ID NO:33).

The term "helper plasmid" refers to either pKD46 encoding λ -Red recombinase or pCP20 encoding FLP site-specific recombinase (ATCC PTA-4455; Datsenko and Wanner, *supra*; and Cherepanov and Wackernagel, *Gene*, 158:9-14 (1995)).

The term "carotenoid overproducing bacteria" refers to a bacteria of the invention which has been genetically modified by the up-regulation or down-regulation of various genes to produce a carotenoid compound a levels greater than the wildtype or unmodified host.

The term "*E. coli*" refers to *Escherichia coli* strain K-12 derivatives, such as MG1655 (ATCC 47076) and MC1061 (ATCC 53338).

The term "*Pantoea stewartii* subsp. *stewartii*" is abbreviated as "*Pantoea stewartii*" and is used interchangeably with *Erwinia stewartii* (Mergaert et al., *Int J. Syst. Bacteriol.*, 43:162-173 (1993)).

The term "*Pantoea ananatas*" is used interchangeably with *Erwinia uredovora* (Mergaert et al., *supra*).

The term "*Pantoea crtEXYIB* cluster" refers to a gene cluster containing carotenoid synthesis genes *crtEXYIB* amplified from *Pantoea stewartii* ATCC 8199. The gene cluster contains the genes *crtE*, *crtX*, *crtY*, *crtI*, and *crtB*. The cluster also contains a *crtZ* gene organized in opposite orientation and adjacent to *crtB* gene.

The term "CrtE" refers to geranylgeranyl pyrophosphate synthase enzyme encoded by *crtE* gene which converts trans-trans-farnesyl diphosphate + isopentenyl diphosphate to pyrophosphate + geranylgeranyl diphosphate.

The term "CrtY" refers to lycopene cyclase enzyme encoded by *crtY* gene which converts lycopene to β -carotene.

The term "CrtI" refers to phytoene dehydrogenase enzyme encoded by *crtI* gene which converts phytoene into lycopene via the intermediaries of phytofluene, zeta-carotene and neurosporene by the introduction of 4 double bonds

The term "CrtB" refers to phytoene synthase enzyme encoded by *crtB* gene which catalyzes reaction from prephytoene diphosphate (geranylgeranyl pyrophosphate) to phytoene.

The term "CrtX" refers to zeaxanthin glucosyl transferase enzyme encoded by *crtX* gene which converts zeaxanthin to zeaxanthin- β -diglucoside.

The term "CrtZ" refers to the β -carotene hydroxylase enzyme encoded by *crtZ* gene which catalyses hydroxylation reaction from β -carotene to zeaxanthin.

The term "carotenoid biosynthetic pathway" refers to those genes comprising members of the upper and/or lower isoprenoid pathways of the present invention as illustrated in Figure 1. In the present invention, the terms "upper isoprenoid pathway" and "upper pathway" will be used interchangeably and will refer the enzymes involved in converting pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate (FPP). These enzymes include, but are not limited to Dxs, Dxr (IspC), YgpP (IspD), YchB (IspE), YgbB (IspF), GcpE (IspG), LytB (IspH), Idi, IspA, and optionally IspB. In the present invention, the terms "lower carotenoid pathway" and "lower pathway" will be used interchangeably and refer to those enzymes which convert FPP to carotenoids, especially β -carotene (Figure 1). The enzymes in this pathway include, but are not limited to CrtE, CrtY, CrtI, CrtB, CrtX, and CrtZ. In the present invention, the "lower pathway" genes are expressed on reporter plasmids pPCB15 or pDCQ108.

The term "carotenoid biosynthetic enzyme" is an inclusive term referring to any and all of the enzymes encoded by the *Pantoea crtEXYIB* cluster. The enzymes include CrtE, CrtY, CrtI, CrtB, and CrtX.

The terms "P1 donor cell" and "donor cell" are used interchangeably in the present invention and refer to a bacterial strain susceptible to infection by a bacteriophage or virus, and which serves as a source for the nucleic acid fragments packaged into the transducing particles. Typically the genetic make up of the donor cell is similar or

identical to the "recipient cell" which serves to receive P1 lysate containing transducing particles or virus produced by the donor cell.

The terms "P1 recipient cell" and "recipient cell" are used interchangeably in the present invention and refer to a bacterial strain
5 susceptible to infection by a bacteriophage or virus and which serves to receive lysate containing transducing particles or virus produced by the donor cell.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form
10 gene segments which are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well-established
15 procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon
20 usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding
25 sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise
30 regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene
35 not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A

"transgene" is a gene that has been introduced into the genome by a transformation procedure.

The term "genetic end product" means the substance, chemical or material (i.e. isoprenoids, carotenoids) that is produced as the result of the activity of a gene product. Typically a gene product is an enzyme and a genetic end product is the product of that enzymatic activity on a specific substrate. A genetic end product may be the result of a single enzyme activity or the result of a number of linked activities, such as found in a biosynthetic pathway (several enzyme activities).

"Operon", in bacterial DNA, is a cluster of contiguous genes transcribed from one promoter that gives rise to a polycistronic mRNA.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site(s), effector binding site(s), and stem-loop structure(s).

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions ("inducible promoters"). Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". Promoters can be further classified by the relative strength of expression observed by their use (i.e. weak, moderate, or strong). It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include regulatory signals capable of affecting mRNA processing or gene expression.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to an RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (US 5,107,065; WO 99/28508). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic", "recombinant" or "transformed" organisms.

The terms "transduction" and "generalized transduction" are used interchangeably and refer to a phenomenon in which bacterial DNA is transferred from one bacterial cell (the donor) to another (the recipient) by

a phage particle containing bacterial DNA (Figure 12). The bacterial DNA fragment from the donor can undergo homologous recombination with the recipient cell's chromosome, stably integrating the donor cell's DNA fragment into the recipient's chromosome.

5 The terms "plasmid", "vector" and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or
10 nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated
15 sequence into a cell. "Transformation cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitates transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced
20 expression of that gene in a foreign host.

 The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence
25 analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR, Inc. 1228 S. Park St. Madison, WI 53715 USA), and the FASTA program incorporating the
30 Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-20. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY. Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based
35 on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters which originally load with the software when first initialized.

The present invention relates to carotenoid overproducing bacteria. The genes of the isoprenoid pathway in the bacterial hosts of the invention have been engineered such that certain genes are either up-regulated or down regulated resulting in the production of carotenoid compounds at a higher level than is found in the unmodified host. In some instances the genes that are regulated are directly involved in the carotenoid biosynthetic pathway. In other instances the genes involved are chromosomal genes that have no understood relationship to the carotenoid biosynthetic pathway.

It has been found that over-expression of certain combinations of carotenoid biosynthetic genes will give an unexpectedly high level of carotenoid production. Examples of genes useful in this manner which are part of the carotenoid biosynthetic pathway are the *dxs* gene, (catalyzing the condensation of pyruvate and D-glyceraldehyde 3-phosphate to D-1-deoxyxylulose 5-phosphate), the *idi* gene (converting isopentenyl diphosphate to dimethylallyl diphosphate), the *ygbB* (*ispF*) gene (catalyzing the cyclization with loss of CMP of 4-diphosphocytidyl-2C-methyl-D-erythritol to 4-diphosphocytidyl-2C-methyl-D-erythritol-2-phosphate to 2C-methyl-D-erythritol-2,4-cyclodiphosphate), the *ygbP* (*ispD*) gene (catalyzing the CTP-dependent cytidylation of 2-C-methyl-D-erythritol-4-phosphate to 4-diphosphocytidyl-2C-methyl-D-erythritol) and together referred to as the *ygbBP* gene, the *lytB* (*ispH*) gene (involved in conversion of 2C-methyl-D-erythritol-2,4-cyclodiphosphate to dimethylallyl diphosphate and isopentenyl diphosphate), and the *ispB* gene encoding the enzyme octaprenyl diphosphate synthase. When these genes are selectively over expressed under the control of a strong promoter the result is an unexpectedly high level of carotenoid production. It is important to note that it is the combination of the over-expression of these genes that has been shown to give the desired effect.

Alternatively, it has also been found that certain essential chromosomal genes, when mutated, will alter the output of the carotenoid biosynthetic pathway. One such gene is the *yjeR* gene (defining a oligo-ribonuclease locus). It has been found that a partial mutation in this gene will unexpectedly increase carotenoid production in a host cell capable of carotenoid biosynthesis.

Genes Involved in Carotenoid Production.

The enzyme pathway involved in the biosynthesis of carotenoids can be conveniently viewed in two parts, the upper isoprenoid pathway

providing for the conversion of pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate (FPP) and the lower carotenoid biosynthetic pathway, which provides for the synthesis of phytoene and all subsequently produced carotenoids. The upper pathway is ubiquitous in many non-carotogenic microorganisms and in these cases it will only be necessary to introduce genes that comprise the lower pathway for the biosynthesis of the desired carotenoid. The key division between the two pathways concerns the synthesis of farnesyl pyrophosphate. Where FPP is naturally present, only elements of the lower carotenoid pathway will be needed. However, it will be appreciated that for the lower pathway carotenoid genes to be effective in the production of carotenoids, it will be necessary for the host cell to have suitable levels of FPP within the cell. Where FPP synthesis is not provided by the host cell, it will be necessary to introduce the genes necessary for the production of FPP. Each of these pathways will be discussed below in detail.

The Upper Isoprenoid Pathway

Isoprenoid biosynthesis occurs through either of two pathways, generating the common C5 isoprene sub-unit, isopentenyl pyrophosphate (IPP). First, IPP may be synthesized through the well-known acetate/mevalonate pathway. However, recent studies have demonstrated that the mevalonate-dependent pathway does not operate in all living organisms. An alternate mevalonate-independent pathway for IPP biosynthesis has been characterized in bacteria and in green algae and higher plants (Horbach et al., *FEMS Microbiol. Lett.*, 111:135-140 (1993); Rohmer et al., *Biochem.*, 295: 517-524 (1993); Schwender et al., *Biochem.*, 316: 73-80 (1996); and Eisenreich et al., *Proc. Natl. Acad. Sci. USA*, 93: 6431-6436 (1996)).

Many steps in the mevalonate-independent isoprenoid pathway are known (Figure 1). For example, the initial steps of the alternate pathway leading to the production of IPP have been studied in *Mycobacterium tuberculosis* by Cole et al. (*Nature*, 393:537-544 (1998)). The first step of the pathway involves the condensation of two 3-carbon molecules (pyruvate and D-glyceraldehyde 3-phosphate) to yield a 5-carbon compound known as D-1-deoxyxylulose-5-phosphate. This reaction occurs by the DXS enzyme, encoded by the *dxs* gene. Next, the isomerization and reduction of D-1-deoxyxylulose-5-phosphate yields 2-C-methyl-D-erythritol-4-phosphate. One of the enzymes involved in the isomerization and reduction process is D-1-deoxyxylulose-5-phosphate

reductoisomerase (DXR), encoded by the gene *dxr* (*ispC*). 2-C-methyl-D-erythritol-4-phosphate is subsequently converted into 4-diphosphocytidyl-2C-methyl-D-erythritol in a CTP-dependent reaction by the enzyme encoded by the non-annotated gene *ygbP*. Recently, however, the *ygbP* gene was renamed as *ispD* as a part of the *isp* gene cluster (SwissProtein Accession #Q46893).

Next, the 2nd position hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol can be phosphorylated in an ATP-dependent reaction by the enzyme encoded by the *ygbB* gene. YchB phosphorylates 4-diphosphocytidyl-2C-methyl-D-erythritol, resulting in 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate. The *ygbB* gene was renamed as *ispE*, also as a part of the *isp* gene cluster (SwissProtein Accession #P24209). YgbB converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate in a CTP-dependent manner. This gene has also been recently renamed, and belongs to the *isp* gene cluster. Specifically, the new name for the *ygbB* gene is *ispF* (SwissProtein Accession #P36663).

The enzymes encoded by the *gcpE* (*ispG*) and *lytB* (*ispH*) genes (and perhaps others) are thought to participate in the reactions leading to formation of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). IPP may be isomerized to DMAPP via IPP isomerase, encoded by the *idi* gene. However, this enzyme is not essential for survival and may be absent in some bacteria using 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. Recent evidence suggests that the MEP pathway branches before IPP and separately produces IPP and DMAPP via the *lytB* gene product. A *lytB* knockout mutation is lethal in *E. coli* except in media supplemented with both IPP and DMAPP.

The synthesis of FPP occurs via the isomerization of IPP to dimethylallyl pyrophosphate. This reaction is followed by a sequence of two prenyltransferase reactions catalyzed by *ispA*, leading to the creation of geranyl pyrophosphate (GPP; a 10-carbon molecule) and farnesyl pyrophosphate (FPP; a 15-carbon molecule).

Genes encoding elements of the upper pathway are known from a variety of plant, animal, and bacterial sources, as shown in Table 1.

Table 1
Sources of Genes Encoding the Upper Isoprene Pathway

Gene	GenBank Accession Number and Source Organism
<i>dxs</i> (D-1-deoxyxylulose 5-phosphate synthase)	AF035440, <i>Escherichia coli</i> Y18874, <i>Synechococcus</i> PCC6301 AB026631, <i>Streptomyces</i> sp. CL190 AB042821, <i>Streptomyces griseolosporeus</i> AF111814, <i>Plasmodium falciparum</i> AF143812, <i>Lycopersicon esculentum</i> AJ279019, <i>Narcissus pseudonarcissus</i> AJ291721, <i>Nicotiana tabacum</i>
<i>dxr (ispC)</i> (1-deoxy-D-xylulose 5-phosphate reductoisomerase)	AB013300, <i>Escherichia coli</i> AB049187, <i>Streptomyces griseolosporeus</i> AF111813, <i>Plasmodium falciparum</i> AF116825, <i>Mentha x piperita</i> AF148852, <i>Arabidopsis thaliana</i> AF182287, <i>Artemisia annua</i> AF250235, <i>Catharanthus roseus</i> AF282879, <i>Pseudomonas aeruginosa</i> AJ242588, <i>Arabidopsis thaliana</i> AJ250714, <i>Zymomonas mobilis</i> strain ZM4 AJ292312, <i>Klebsiella pneumoniae</i> , AJ297566, <i>Zea mays</i>
<i>ygbP (ispD)</i> (2-C-methyl-D-erythritol 4-phosphate cytidyltransferase)	AB037876, <i>Arabidopsis thaliana</i> AF109075, <i>Clostridium difficile</i> AF230736, <i>Escherichia coli</i> AF230737, <i>Arabidopsis thaliana</i>
<i>ychB (ispE)</i> (4-diphosphocytidyl-2-C-methyl-D-erythritol kinase)	AF216300, <i>Escherichia coli</i> AF263101, <i>Lycopersicon esculentum</i> AF288615, <i>Arabidopsis thaliana</i>
<i>ygbB (ispF)</i> (2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase)	AB038256, <i>Escherichia coli</i> mecs gene AF230738, <i>Escherichia coli</i> AF250236, <i>Catharanthus roseus</i> (MECS) AF279661, <i>Plasmodium falciparum</i> AF321531, <i>Arabidopsis thaliana</i>
<i>gcpE (ispG)</i> (1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase)	O67496, <i>Aquifex aeolicus</i> P54482, <i>Bacillus subtilis</i> Q9pky3, <i>Chlamydia muridarum</i> Q9Z8H0, <i>Chlamydophila pneumoniae</i> O84060, <i>Chlamydia trachomatis</i> P27433, <i>Escherichia coli</i> P44667, <i>Haemophilus influenzae</i>

Gene	GenBank Accession Number and Source Organism
	Q9ZLL0, <i>Helicobacter pylori</i> J99 O33350, <i>Mycobacterium tuberculosis</i> S77159, <i>Synechocystis</i> sp. Q9WZZ3, <i>Thermotoga maritima</i> O83460, <i>Treponema pallidum</i> Q9JZ40, <i>Neisseria meningitidis</i> Q9PPM1, <i>Campylobacter jejuni</i> Q9RXC9, <i>Deinococcus radiodurans</i> AAG07190, <i>Pseudomonas aeruginosa</i> Q9KTX1, <i>Vibrio cholerae</i>
lytB (ispH)	AF027189, <i>Acinetobacter</i> sp. BD413 AF098521, <i>Burkholderia pseudomallei</i> AF291696, <i>Streptococcus pneumoniae</i> AF323927, <i>Plasmodium falciparum</i> gene M87645, <i>Bacillus subtilis</i> U38915, <i>Synechocystis</i> sp. X89371, <i>C. jejuni</i> sp O67496
IspA (FPP synthase)	AB003187, <i>Micrococcus luteus</i> AB016094, <i>Synechococcus elongatus</i> AB021747, <i>Oryza sativa</i> FPPS1 gene for farnesyl diphosphate synthase AB028044, <i>Rhodobacter sphaeroides</i> AB028046, <i>Rhodobacter capsulatus</i> AB028047, <i>Rhodovulum sulfidophilum</i> AF112881 and AF136602, <i>Artemisia annua</i> AF384040, <i>Mentha x piperita</i> D00694, <i>Escherichia coli</i> D13293, <i>B. stearothermophilus</i> D85317, <i>Oryza sativa</i> X75789, <i>A. thaliana</i> Y12072, <i>G. arboreum</i> Z49786, <i>H. brasiliensis</i> U80605, <i>Arabidopsis thaliana</i> farnesyl diphosphate synthase precursor (FPS1) mRNA, complete cds X76026, <i>K. lactis</i> FPS gene for farnesyl diphosphate synthetase, QCR8 gene for bc1 complex, subunit VIII X82542, <i>P. argentatum</i> mRNA for farnesyl diphosphate synthase (FPS1) X82543, <i>P. argentatum</i> mRNA for farnesyl diphosphate synthase (FPS2) BC010004, <i>Homo sapiens</i> , farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase), clone MGC 15352 IMAGE, 4132071, mRNA, complete cds AF234168, <i>Dictyostelium discoideum</i> farnesyl diphosphate synthase (Dfps)

Gene	GenBank Accession Number and Source Organism
	L46349, <i>Arabidopsis thaliana</i> farnesyl diphosphate synthase (FPS2) mRNA, complete cds L46350, <i>Arabidopsis thaliana</i> farnesyl diphosphate synthase (FPS2) gene, complete cds L46367, <i>Arabidopsis thaliana</i> farnesyl diphosphate synthase (FPS1) gene, alternative products, complete cds M89945, Rat farnesyl diphosphate synthase gene, exons 1-8 NM_002004, <i>Homo sapiens</i> farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (FDPS), mRNA U36376, <i>Artemisia annua</i> farnesyl diphosphate synthase (fps1) mRNA, complete cds XM_001352, <i>Homo sapiens</i> farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (FDPS), mRNA XM_034497, <i>Homo sapiens</i> farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (FDPS), mRNA XM_034498, <i>Homo sapiens</i> farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (FDPS), mRNA XM_034499, <i>Homo sapiens</i> farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (FDPS), mRNA XM_0345002, <i>Homo sapiens</i> farnesyl diphosphate synthase (farnesyl pyrophosphate synthetase, dimethylallyltranstransferase, geranyltranstransferase) (FDPS), mRNA

The most preferred source of genes for the upper isoprene pathway in the present invention is from *Methylobacter* 16a (ATCC PTA-2402). *Methylobacter* 16a is particularly well-suited for the present invention, as the methanotroph is naturally pink-pigmented, producing a 30-carbon carotenoid. Thus, the organism possesses the genes of the upper isoprene pathway. Sequences of these preferred genes are presented as the following SEQ ID numbers: the *dxs*(16a) gene (SEQ ID

NO:13), the *dxr(16a)* gene (SEQ ID NO:17), and the *lytB(16a)* gene (SEQ ID NO:15).

The Lower Carotenoid Biosynthetic Pathway

The division between the upper isoprenoid pathway and the lower carotenoid pathway is somewhat subjective. Because FPP synthesis is common in both carotenogenic and non-carotenogenic bacteria, the first step in the lower carotenoid biosynthetic pathway is considered to begin with the prenyltransferase reaction converting farnesyl pyrophosphate (FPP) to geranylgeranyl pyrophosphate (GGPP). The gene *crtE*, encoding GGPP synthetase, is responsible for this prenyltransferase reaction which adds IPP to FPP to produce the 20-carbon molecule GGPP. A condensation reaction of two molecules of GGPP occurs to form phytoene (PPPP), the first 40-carbon molecule of the lower carotenoid biosynthesis pathway. This enzymatic reaction is catalyzed by *crtB*, encoding phytoene synthase.

Lycopene, which imparts a "red" colored spectra, is produced from phytoene through four sequential dehydrogenation reactions by the removal of eight atoms of hydrogen, catalyzed by the gene *crtI* (encoding phytoene desaturase). Intermediaries in this reaction are phytofluene, zeta-carotene, and neurosporene.

Lycopene cyclase (*crtY*) converts lycopene to β -carotene. In the present invention, a reporter plasmid is used which produces β -carotene as the genetic end product. However, additional genes may be used to create a variety of other carotenoids. For example, β -carotene is converted to zeaxanthin via a hydroxylation reaction resulting from the activity of β -carotene hydroxylase (encoded by the *crtZ* gene). β -cryptoxanthin is an intermediate in this reaction.

β -carotene is converted to canthaxanthin by β -carotene ketolase encoded by either the *crtW* or *crtO* gene. Echinenone is an intermediate in this reaction. Canthaxanthin can then be converted to astaxanthin by β -carotene hydroxylase encoded by the *crtZ* or *crtR* gene. Adonirubrin is an intermediate in this reaction.

Zeaxanthin can be converted to zeaxanthin- β -diglucoside. This reaction is catalyzed by zeaxanthin glucosyl transferase (*crtX*).

Zeaxanthin can be converted to astaxanthin by β -carotene ketolase encoded by *crtW*, *crtO* or *bkt*. The BKT/CrtW enzymes synthesized canthaxanthin via echinenone from β -carotene and 4-ketozeaxanthin. Adonixanthin is an intermediate in this reaction.

Spheroidene can be converted to spheroidenone by spheroidene monooxygenase encoded by *crtA*.

Neurosporene can be converted spheroidene and lycopene can be converted to spirilloxanthin by the sequential actions of
 5 hydroxyneurosporene synthase, methoxyneurosporene desaturase and hydroxyneurosporene-O-methyltransferase encoded by the *crtC*, *crtD* and *crtF* genes, respectively.

β -carotene can be converted to isorenieratene by β -carotene desaturase encoded by *crtU*.

10 Genes encoding elements of the lower carotenoid biosynthetic pathway are known from a variety of plant, animal, and bacterial sources, as shown in Table 2.

Table 2

15 Sources of Genes Encoding the Lower Carotenoid Biosynthetic Pathway

Gene	GenBank Accession Number and Source Organism
<i>crtE</i> (GGPP Synthase)	AB000835, <i>Arabidopsis thaliana</i> AB016043 and AB019036, <i>Homo sapiens</i> AB016044, <i>Mus musculus</i> AB027705 and AB027706, <i>Daucus carota</i> AB034249, <i>Croton sublyratus</i> AB034250, <i>Scoparia dulcis</i> AF020041, <i>Helianthus annuus</i> AF049658, <i>Drosophila melanogaster</i> signal recognition particle 19kDa protein (<i>srp19</i>) gene, partial sequence; and geranylgeranyl pyrophosphate synthase (<i>quemao</i>) gene, complete cds AF049659, <i>Drosophila melanogaster</i> geranylgeranyl pyrophosphate synthase mRNA, complete cds AF139916, <i>Brevibacterium linens</i> AF279807, <i>Penicillium paxilli</i> geranylgeranyl pyrophosphate synthase (<i>ggs1</i>) gene, complete AF279808, <i>Penicillium paxilli</i> dimethylallyl tryptophan synthase (<i>paxD</i>) gene, partial cds; and cytochrome P450 monooxygenase (<i>paxQ</i>), cytochrome P450 monooxygenase (<i>paxP</i>), PaxC (<i>paxC</i>), monooxygenase (<i>paxM</i>), geranylgeranyl pyrophosphate synthase (<i>paxG</i>), PaxU (<i>paxU</i>), and metabolite transporter (<i>paxT</i>) genes, complete cds AJ010302, <i>Rhodobacter sphaeroides</i> AJ133724, <i>Mycobacterium aurum</i> AJ276129, <i>Mucor circinelloides f. lusitanicus carG</i>

Gene	GenBank Accession Number and Source Organism
	gene for geranylgeranyl pyrophosphate synthase, exons 1-6 D85029, <i>Arabidopsis thaliana</i> mRNA for geranylgeranyl pyrophosphate synthase, partial cds L25813, <i>Arabidopsis thaliana</i> L37405, <i>Streptomyces griseus</i> geranylgeranyl pyrophosphate synthase (<i>crtB</i>), phytoene desaturase (<i>crtE</i>) and phytoene synthase (<i>crtI</i>) genes, complete cds U15778, <i>Lupinus albus</i> geranylgeranyl pyrophosphate synthase (<i>ggps1</i>) mRNA, complete cds U44876, <i>Arabidopsis thaliana</i> pregeranylgeranyl pyrophosphate synthase (<i>GGPS2</i>) mRNA, complete cds X92893, <i>C. roseus</i> X95596, <i>S. griseus</i> X98795, <i>S. alba</i> Y15112, <i>Paracoccus marcusii</i>
crtX (Zeaxanthin glucosylase)	D90087, <i>E. uredovora</i> M87280 and M90698, <i>Pantoea agglomerans</i>
crtY (Lycopene- β -cyclase)	AF139916, <i>Brevibacterium linens</i> AF152246, <i>Citrus x paradisi</i> AF218415, <i>Bradyrhizobium</i> sp. ORS278 AF272737, <i>Streptomyces griseus</i> strain IFO13350 AJ133724, <i>Mycobacterium aurum</i> AJ250827, <i>Rhizomucor circinelloides</i> f. <i>lusitanicus</i> <i>carRP</i> gene for lycopene cyclase/phytoene synthase, exons 1-2 AJ276965, <i>Phycomyces blakesleeianus</i> <i>carRA</i> gene for phytoene synthase/lycopene cyclase, exons 1-2 D58420, <i>Agrobacterium aurantiacum</i> D83513, <i>Erythrobacter longus</i> L40176, <i>Arabidopsis thaliana</i> lycopene cyclase (<i>LYC</i>) mRNA, complete cds M87280, <i>Pantoea agglomerans</i> U50738, <i>Arabidopsis thaliana</i> lycopene epsilon cyclase mRNA, complete cds U50739, <i>Arabidopsis thaliana</i> lycopene β cyclase mRNA, complete cds U62808, <i>Flavobacterium</i> ATCC21588 X74599, <i>Synechococcus</i> sp. <i>lcy</i> gene for lycopene cyclase X81787, <i>N. tabacum</i> <i>CrtL-1</i> gene encoding lycopene cyclase X86221, <i>C. annuum</i> X86452, <i>L. esculentum</i> mRNA for lycopene β-cyclase

Gene	GenBank Accession Number and Source Organism
	X95596, <i>S. griseus</i> X98796, <i>N. pseudonarcissus</i>
crtI (Phytoene desaturase)	AB046992, <i>Citrus unshiu</i> <i>CitPDS1</i> mRNA for phytoene desaturase, complete cds AF039585, <i>Zea mays</i> phytoene desaturase (<i>pds1</i>) gene promoter region and exon 1 AF049356, <i>Oryza sativa</i> phytoene desaturase precursor (<i>Pds</i>) mRNA, complete cds AF139916, <i>Brevibacterium linens</i> AF218415, <i>Bradyrhizobium</i> sp. ORS278 AF251014, <i>Tagetes erecta</i> AF364515, <i>Citrus x paradisi</i> D58420, <i>Agrobacterium aurantiacum</i> D83514, <i>Erythrobacter longus</i> L16237, <i>Arabidopsis thaliana</i> L37405, <i>Streptomyces griseus</i> geranylgeranyl pyrophosphate synthase (<i>crtB</i>), phytoene desaturase (<i>crtE</i>) and phytoene synthase (<i>crtI</i>) genes, complete cds L39266, <i>Zea mays</i> phytoene desaturase (<i>Pds</i>) mRNA, complete cds M64704, Soybean phytoene desaturase M88683, <i>Lycopersicon esculentum</i> phytoene desaturase (<i>pds</i>) mRNA, complete cds S71770, carotenoid gene cluster U37285, <i>Zea mays</i> U46919, <i>Solanum lycopersicum</i> phytoene desaturase (<i>Pds</i>) gene, partial cds U62808, <i>Flavobacterium</i> ATCC21588 X55289, <i>Synechococcus pds</i> gene for phytoene desaturase X59948, <i>L. esculentum</i> X62574, <i>Synechocystis</i> sp. <i>pds</i> gene for phytoene desaturase X68058, <i>C. annuum pds1</i> mRNA for phytoene desaturase X71023, <i>Lycopersicon esculentum pds</i> gene for phytoene desaturase X78271, <i>L. esculentum</i> (Ailsa Craig) PDS gene X78434, <i>P. blakesleeianus</i> (NRRL1555) <i>carB</i> gene X78815, <i>N. pseudonarcissus</i> X86783, <i>H. pluvialis</i> Y14807, <i>Dunaliella bardawil</i> Y15007, <i>Xanthophyllomyces dendrorhous</i> Y15112, <i>Paracoccus marcusii</i> Y15114, <i>Anabaena</i> PCC7210 <i>crtP</i> gene Z11165, <i>R. capsulatus</i>

Gene	GenBank Accession Number and Source Organism
crtB (Phytoene synthase)	AB001284, <i>Spirulina platensis</i> AB032797, <i>Daucus carota</i> PSY mRNA for phytoene synthase, complete cds AB034704, <i>Rubrivivax gelatinosus</i> AB037975, <i>Citrus unshiu</i> AF009954, <i>Arabidopsis thaliana</i> phytoene synthase (PSY) gene, complete cds AF139916, <i>Brevibacterium linens</i> AF152892, <i>Citrus x paradisi</i> AF218415, <i>Bradyrhizobium</i> sp. ORS278 AF220218, <i>Citrus unshiu</i> phytoene synthase (Psy1) mRNA, complete cds AJ010302, <i>Rhodobacter</i> AJ133724, <i>Mycobacterium aurum</i> AJ278287, <i>Phycomyces blakesleeianus</i> carRA gene for lycopene cyclase/phytoene synthase, AJ304825, <i>Helianthus annuus</i> mRNA for phytoene synthase (psy gene) AJ308385, <i>Helianthus annuus</i> mRNA for phytoene synthase (psy gene) D58420, <i>Agrobacterium aurantiacum</i> L23424, <i>Lycopersicon esculentum</i> phytoene synthase (PSY2) mRNA, complete cds L25812, <i>Arabidopsis thaliana</i> L37405, <i>Streptomyces griseus</i> geranylgeranyl pyrophosphate synthase (crtB), phytoene desaturase (crtE) and phytoene synthase (crtI) genes, complete cds M38424, <i>Pantoea agglomerans</i> phytoene synthase (crtE) gene, complete cds M87280, <i>Pantoea agglomerans</i> S71770, Carotenoid gene cluster U32636, <i>Zea mays</i> phytoene synthase (Y1) gene, complete cds U62808, <i>Flavobacterium</i> ATCC21588 U87626, <i>Rubrivivax gelatinosus</i> U91900, <i>Dunaliella bardawil</i> X52291, <i>Rhodobacter capsulatus</i> X60441, <i>L. esculentum</i> GTom5 gene for phytoene synthase X63873, <i>Synechococcus</i> PCC7942 pys gene for phytoene synthase X68017, <i>C. annuum</i> psy1 mRNA for phytoene synthase X69172, <i>Synechocystis</i> sp. pys gene for phytoene synthase X78814, <i>N. pseudonarcissus</i>

Gene	GenBank Accession Number and Source Organism
crtZ (β -carotene hydroxylase)	D58420, <i>Agrobacterium aurantiacum</i> D58422, <i>Alcaligenes</i> sp. D90087, <i>E. uredovora</i> M87280, <i>Pantoea agglomerans</i> U62808, <i>Flavobacterium</i> ATCC21588 Y15112, <i>Paracoccus marcusii</i>
crtW (β -carotene ketolase)	AF218415, <i>Bradyrhizobium</i> sp. ORS278 D45881, <i>Haematococcus pluvialis</i> D58420, <i>Agrobacterium aurantiacum</i> D58422, <i>Alcaligenes</i> sp. X86782, <i>H. pluvialis</i> Y15112, <i>Paracoccus marcusii</i>
crtO (β -C4-ketolase)	X86782, <i>H. pluvialis</i> Y15112, <i>Paracoccus marcusii</i>
crtU (β -carotene dehydrogenase)	AF047490, <i>Zea mays</i> AF121947, <i>Arabidopsis thaliana</i> AF139916, <i>Brevibacterium linens</i> AF195507, <i>Lycopersicon esculentum</i> AF272737, <i>Streptomyces griseus</i> strain IFO13350 AF372617, <i>Citrus x paradisi</i> AJ133724, <i>Mycobacterium aurum</i> AJ224683, <i>Narcissus pseudonarcissus</i> D26095 and U38550, <i>Anabaena</i> sp. X89897, <i>C. annuum</i> Y15115, <i>Anabaena</i> PCC7210 crtQ gene
crtA (spheroidene monooxygenase)	AJ010302, <i>Rhodobacter sphaeroides</i> Z11165 and X52291, <i>Rhodobacter capsulatus</i>
crtC (hydroxyneurosporene synthase)	AB034704, <i>Rubrivivax gelatinosus</i> AF195122 and AJ010302, <i>Rhodobacter sphaeroides</i> AF287480, <i>Chlorobium tepidum</i> U73944, <i>Rubrivivax gelatinosus</i> X52291 and Z11165, <i>Rhodobacter capsulatus</i> Z21955, <i>M. xanthus</i>
crtD (carotenoid 3,4-desaturase)	AJ010302 and X63204, <i>Rhodobacter sphaeroides</i> U73944, <i>Rubrivivax gelatinosus</i> X52291 and Z11165, <i>Rhodobacter capsulatus</i>
crtF (1-OH-carotenoid methylase)	AB034704, <i>Rubrivivax gelatinosus</i> AF288602, <i>Chloroflexus aurantiacus</i> AJ010302, <i>Rhodobacter sphaeroides</i> X52291 and Z11165, <i>Rhodobacter capsulatus</i>

The most preferred source of *crt* genes is from *Pantoea stewartii*. Sequences of these preferred genes are presented as the following SEQ ID numbers: the *crtE* gene (SEQ ID NO:1), the *crtX* gene (SEQ ID NO:3),

crtY (SEQ ID NO:5), the *crtI* gene (SEQ ID NO:7), the *crtB* gene (SEQ ID NO:9) and the *crtZ* gene (SEQ ID NO:11).

By using various combinations of the genes presented in Table 2 and the preferred genes of the present invention, innumerable different carotenoids and carotenoid derivatives could be made using the methods of the present invention, provided that sufficient sources of FPP are available in the host organism. For example, the gene cluster *crtEXYIB* enables the production of β -carotene. Addition of the *crtZ* to *crtEXYIB* enables the production of zeaxanthin.

It is envisioned that useful products of the present invention will include any carotenoid compound as defined herein including, but not limited to antheraxanthin, adonixanthin, astaxanthin, canthaxanthin, capsorubrin, β -cryptoxanthin, didehydrolycopene, didehydrolycopene, β -carotene, ζ -carotene, δ -carotene, γ -carotene, keto- γ -carotene, ψ -carotene, ϵ -carotene, β,ψ -carotene, torulene, echinenone, gamma-carotene, zeta-carotene, alpha-cryptoxanthin, diatoxanthin, 7,8-didehydroastaxanthin, fucoxanthin, fucoxanthinol, isorenieratene, β -isorenieratene, lactucaxanthin, lutein, lycopene, neoxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, siphonaxanthin, spheroidene, spheroidenone, spirilloxanthin, uriolide, uriolide acetate, violaxanthin, zeaxanthin- β -diglucoside, zeaxanthin, and C30-carotenoids.

Methods for Optimizing the Carotenoid Biosynthetic Pathway

Metabolic engineering generally involves the introduction of new metabolic activities into the host organism or the improvement of existing processes by engineering changes such as adding, removing, or modifying genetic elements (Stephanopoulos, G., *Metab. Eng.*, 1: 1-11 (1999)). One such modification is genetically engineering modulations to the expression of relevant genes in a metabolic pathway.

There are a variety of ways to modulate gene expression. Microbial metabolic engineering generally involves the use of multi-copy vectors to express a gene of interest under the control of a constitutive or inducible promoter. This method of metabolic engineering for industrial use has several drawbacks. It is sometimes difficult to maintain the vectors due to segregational instability. Deleterious effects on cell viability and growth are often observed due to the vector burden. It is also difficult to control the optimal expression level of desired genes on a vector. To avoid the undesirable effects of using a multi-copy vector, a chromosomal

integration approach using homologous recombination via a single insertion of bacteriophage λ , transposons, or other suitable vectors containing the gene of interest has been used. However, this method also has drawbacks such as the need for multiple cloning steps in order to get the gene of interest into a suitable vector prior to recombination. Another drawback is the instability associated with the inserted genes, which can be lost due to excision. Lastly, these methods have a limitation associated with the number of possible insertions and the inability to control the location of the insertion site on a chromosome.

Several processes are involved in the regulation of gene expression. The main steps are (1) the initiation of transcription, (2) the termination of transcription, (3) the processing of transcripts, and (4) translation. Among these, the transcription initiation is a major step for controlling gene expression. The transcription initiation is determined by the sequence of the promoter region that includes a binding site for RNA polymerase together with possible binding sites for one or more transcription factors.

Strong promoters are widely used for constitutive overexpression of key genes in a metabolic pathway. Strong and moderately strong promoters that are useful for expression in *E. coli* include *lac*, *trp*, λP_L , λP_R , *T7*, *tac*, *T5* (P_{T5}), and *trc*. A conventional way to regulate the amount and the timing of protein expression is to use an inducible promoter. An inducible promoter is not always active the way constitutive promoters are (e.g. viral promoters). Inducible promoters are normally activated in response to certain environmental or chemical stimuli (i.e. heat shock promoter, isopropyl- β -thiogalactopyranoside (IPTG) responsive promoters, and tetracycline (tet) responsive promoters, to name a few).

Promoters of the stationary phase σS regulon, which are active under stress conditions and at the onset of the stationary phase, control expression of about 100 genes involved in the protection of the cell against various stresses. The promoters of the σS regulon genes may also be useful for the expression of the desired genes when the metabolite products inhibit a cell growth. The σS -dependent stationary phase promoters includes *rpoS*, *bolA*, *appY*, *dps*, *cyxAB-appA*, *csgA*, *treA*, *osmB*, *katE*, *xthA*, *otsBA*, *glgS*, *osmY*, *pex*, and *mcc*, to name a few.

Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary, however, it is most preferred if included.

Alternatively, it may be necessary to reduce or eliminate the expression of certain genes in the target pathway or in competing pathways that may serve as competing sinks for energy or carbon. Methods of down-regulating genes for this purpose have been explored.

5 Where the sequence of the gene to be disrupted is known, one of the most effective methods of gene down-regulation is targeted gene disruption, a process where foreign DNA is inserted into a structural gene so as to disrupt transcription. This can be effected by the creation of genetic cassettes comprising the DNA to be inserted (often a genetic
10 marker) flanked by sequence having a high degree of homology to a portion of the gene to be disrupted. Introduction of the cassette into the host cell results in insertion of the foreign DNA into the structural gene via the native DNA replication mechanisms of the cell or by the λ -Red recombination system used in the present invention. (See for example
15 Hamilton et al., *J. Bacteriol.*, 171:4617-4622 (1989); Balbas et al., *Gene*, 136:211-213 (1993); Gueldener et al., *Nucleic Acids Res.*, 24:2519-2524 (1996); and Smith et al., *Methods Mol. Cell. Biol.*, 5:270-277 (1996))

Antisense technology is another method of down regulating genes where the sequence of the target gene is known. To accomplish this, a
20 nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the anti-sense strand of RNA will be transcribed. This construct is then introduced into the host cell and the antisense strand of RNA is produced. Antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the protein of
25 interest. A person of skill in the art will know that special considerations are associated with the use of antisense technologies in order to reduce expression of particular genes. For example, the proper level of expression of antisense genes may require the use of different chimeric genes utilizing different regulatory elements known to the skilled artisan.

30 Although targeted gene disruption and antisense technology offer effective means of down regulating genes where the sequence is known, other less specific methodologies have been developed that are not sequence based. For example, cells may be exposed to UV radiation and then screened for the desired phenotype. Mutagenesis with chemical
35 agents is also effective for generating mutants and commonly used substances include chemicals that affect non-replicating DNA such as HNO_2 and NH_2OH , as well as agents that affect replicating DNA such as acridine dyes, notable for causing frame-shift mutations. Specific

methods for creating mutants using radiation or chemical agents are well documented in the art. See for example Thomas D. Brock in
Biotechnology: A Textbook of Industrial Microbiology, Second Edition
(1989) Sinauer Associates, Inc., Sunderland, MA., or Deshpande, Mukund
5 V., *Appl. Biochem. Biotechnol.*, 36, 227, (1992).

Another non-specific method of gene disruption is the use of transposable elements or transposons. Transposons are genetic elements that insert randomly into DNA but can be later retrieved on the basis of sequence to determine where the insertion has occurred. Both
10 *in vivo* and *in vitro* transposition methods are known. Both methods involve the use of a transposable element in combination with a transposase enzyme. When the transposable element or transposon is contacted with a nucleic acid fragment in the presence of the transposase, the transposable element will randomly insert into the nucleic acid
15 fragment. The technique is useful for random mutagenesis and for gene isolation, since the disrupted gene may be identified on the basis of the sequence of the transposable element. Kits for *in vitro* transposition are commercially available (see for example The Primer Island Transposition Kit, available from Perkin Elmer Applied Biosystems, Branchburg, NJ,
20 based upon the yeast Ty1 element; The Genome Priming System, available from New England Biolabs, Beverly, MA; based upon the bacterial transposon Tn7; and the EZ::TN Transposon Insertion Systems, available from Epicentre Technologies, Madison, WI, based upon the Tn5 bacterial transposable element). Transposon-mediated random insertion
25 in the chromosome can be used for isolating mutants for any number of applications including enhanced production of any number of desired products including enzymes or other proteins, amino acids, or small organic molecules including alcohols.

The present invention has made use of this last method of pathway
30 modulation to cause mutations in various essential genes to test whether there was any effect on the output of the carotenoid biosynthetic pathway. Transposon mutagenesis was used to create an *E. coli* mutant having a partial disruption in the *yjeR* gene. The precise sequence of the mutated gene is given as SEQ ID NO:63. This *yjeR* mutation (*yjeR*::*Tn5*) resulted in
35 increased β -carotene production through an increase in plasmid copy number of the carotenoid producing plasmid (pPCB15 or pDCW108). The effect of mutation of this locus on plasmids is novel and could not have been predicted from known studies. Stacking the *yjeR* mutation

(*yjeR::Tn5*) into the engineered *E. coli* strains that were made by chromosomal engineering of a non-endogenous promoter upstream of isoprenoid genes and chromosomally integrating non-endogenous isoprenoid pathway genes allowed further increases of β -carotene production.

The general methods described herein for pathway modulation are useful and enable the skilled person to practice the present invention. It will be appreciated that other, less traditional methods may be envisioned that will allow the practitioner to make the necessary modifications in the isoprenoid pathway. One such method involving chromosomal promoter replacement using a bacteriophage transduction system was used herein to good effect and is described below.

Optimization of Carotenoid Production in *E. coli* by Bacteriophage Transduction.

The present method combines promoter replacement via homologous recombination (in a recombination proficient host) with a bacteriophage transducing system. The method allows for the rapid insertion of strong promoters upstream of desired elements for increased gene expression. The method also facilitates the production of libraries to assess which combinations of expressible genetic elements will optimize production of the desired genetic end product (Figure 12). In this way, genes not normally associated with a particular biosynthetic pathway may be identified which unexpectedly have significant effects on the production of the desired genetic end product.

Integration Cassettes

One aspect of the promoter replacement method is the use of an integration cassette. As used in the present invention, "integration cassettes" are the linear double-stranded DNA fragments chromosomally integrated by homologous recombination via the use of two PCR-generated fragments or one PCR-generated fragment as seen in Figure 2. The integration cassette comprises a nucleic acid integration fragment that contains an expressible DNA fragment and a selectable marker bounded by specific recombinase sites responsive to a site-specific recombinase, and homology arms having homology to different portions of the host cell's chromosome. Typically, the integration cassette will have the general structure: 5'-RR1-RS-SM-RS-Y-RR2-3' wherein

(i) RR1 is a first homology arm ;

(ii) RS is a recombination site responsive to a site-specific recombinase;

(iii) SM is a DNA fragment encoding a selectable marker;

(iv) Y is a first expressible DNA fragment; and

5 (v) RR2 is a second homology arm.

Expressible DNA fragments of the invention are those that will be useful in genetically engineering biosynthetic pathways. For example, it may be useful to engineer a strong promoter in place of a native promoter in certain pathways. Virtually any promoter is suitable for the present
10 invention including, but not limited to *lac*, *ara*, *tet*, *trp*, λP_L , λP_R , *T7*, *tac*, *P_{T5}*, and *trc* (useful for expression in *Escherichia coli*) as well as the *amy*, *apr*, *npr* promoters and various phage promoters useful for expression in *Bacillus*, for example.

Alternatively, different coding regions may be introduced
15 downstream of existing native promoters. In this manner, new coding regions comprising a biosynthetic pathway may be introduced that either complete or enhance a pathway already in existence in the host cell. These coding regions may be genes which retain their native promoters or may be chimeric genes operably linked to an inducible or constitutive
20 strong promoter for increased expression of the genes in the targeted biosynthetic pathway. Preferred in the present invention are the genes of the isoprenoid/carotenoid biosynthetic pathway, which include *dxs*, *dxr*, *ygbP*, *ychB*, *ygbB*, *idi*, *ispA*, *lytB*, *gcpE*, *ispB*, *gps*, *crtE*, *crtY*, *crtI*, *crtB*, *crtX*, and *crtZ*, as defined above and illustrated in Figure 1. In the present
25 invention, it is preferred if the expressible DNA fragment is a promoter or a coding region useful for modulation of a biosynthetic pathway. Exemplified in the present invention is the phage *T5* strong promoter used for the modulation of the isoprenoid biosynthetic pathway in a recombinant proficient *E. coli* host. In some situations the expressible DNA fragment
30 may be in antisense orientation where it is desired to down-regulate certain elements of the pathway.

Generally, the preferred length of the homology arms is about 10 to about 100 base pairs in length. Given the relatively short lengths of the homology arms used in the present invention for homologous
35 recombination, one would expect that the level of acceptable mismatched sequences should be kept to an absolute minimum for efficient recombination, preferably using sequences which are identical to those targeted for homologous recombination. From 20 to 40 base pairs of

homology, the efficiency of homologous recombination increases by four orders of magnitude (Yu et al. *PNAS*. 97:5978-5983. (2000)). Therefore, multiple mismatching within homology arms may decrease the efficiency of homologous recombination; however, one skilled in the art can easily ascertain the acceptable level of mismatching.

The present invention makes use of a selectable marker on one of the two recombination elements (integration cassettes). Selectable markers are known in the art including, but are not limited to antibiotic resistance markers such as ampicillin, kanamycin, and tetracycline resistance. Selectable markers may also include amino acid biosynthesis enzymes (for selection of auxotrophs normally requiring the exogenously supplied amino acid of interest) and enzymes which catalyze visible changes in appearance such as β -galactosidase in *lac*⁻ bacteria. As used herein, the markers are flanked by site-specific recombinase recognition sequences. After selection and construct verification, a site-specific recombinase is used to remove the marker. The steps of the present invention can then be repeated with additional *in vivo* chromosomal modifications. The integration cassette used to engineer the chromosomal modification includes a promoter and/or gene, and a selection marker flanked by site-specific recombinase sequences. Site-specific recombinases, such as the use of flippase (FLP) recombinase in the present invention, recognize specific recombination sequences (i.e. *FRT* sequences) and allow for the excision of the selectable marker. This aspect of the invention enables the repetitive use of the present process for multiple chromosomal modifications. The invention is not limited to the FLP-*FRT* recombinase system as several examples of site specific recombinases and their associated specific recognition sequences are known in the art. Examples of other suitable site-specific recombinases and their corresponding recognition sequences include: *Cre-lox*, *R/RS*, *Gin/gix*, *Xer/dif*, *Int/att*, a pSR1 system, a *cer* system, and a *fim* system.

Recombination Proficient Host Cells

The present invention makes use of a recombination proficient host cell that is able to mediate efficient homologous recombination between the integration cassettes and the host cell chromosome. Some organisms mediate homologous recombination very effectively (yeast for example) while others require genetic intervention. For example *E. coli*, a host generally considered as one which does not undergo efficient transformation via homologous recombination naturally, may be altered to

make it a recombination proficient host. Transformation with a helper plasmid containing the λ -Red recombinase system increases the rate of homologous recombination several orders of magnitude (Murphy et al., *Gene*, 246:321-330 (2000); Murphy, K., *J. Bacteriol.*, 180:2063-2071; 5 Poteete and Fenton, *J. Bacteriol.*, 182:2336-2340 (2000); Poteete, A., *FEMS Microbiology Lett.*, 201:9-14 (2001); Datsenko and Wanner, *supra*; Yu et al., *supra*; Chaverroche et al., *Nucleic Acids Research*, 28:e97:1-6 (2000); US 6,355,412; US 6,509,156; and US SN 60/434602). The λ -Red system can also be chromosomally integrated into the host. The λ -Red 10 system contains three genes (*exo*, *bet*, and *gam*) which change the normally recombination deficient *E. coli* into a recombination proficient host.

Normally, *E. coli* efficiently degrades linear double stranded DNA via its RecBCD endonuclease, resulting in transformation efficiencies not 15 useful for chromosomal engineering. The *gam* gene encodes for a protein that binds to the *E.coli* RecBCD complex, inhibiting endonuclease activity. The *exo* gene encodes for a λ -exonuclease which processively degrades the 5' end strand of double stranded DNA and creates 3' single stranded overhangs. The protein encoded by *bet* complexes with the λ - 20 exonuclease and binds to the single-stranded DNA overhangs and promotes renaturation of complementary strands and is capable of mediating exchange reactions. The λ -Red recombinase system enables the use of homologous recombination as a tool for *in vivo* chromosomal engineering in hosts, such as *E. coli*, normally considered difficult to 25 transform by homologous recombination. The λ -Red system works in other bacteria as well (Poteete, A., *supra*, 2001). Use of the λ -Red recombinase system should be applicable to other hosts generally used for industrial production. These additional hosts include, but are not limited to *Agrobacterium*, *Erythrobacter*, *Chlorobium*, *Chromatium*, 30 *Flavobacterium*, *Cytophaga*, *Rhodobacter*, *Rhodococcus*, *Streptomyces*, *Brevibacterium*, *Corynebacteria*, *Mycobacterium*, *Deinococcus*, *Paracoccus*, *Escherichia*, *Bacillus*, *Myxococcus*, *Salmonella*, *Yersinia*, *Erwinia*, *Pantoea*, *Pseudomonas*, *Sphingomonas*, *Methylomonas*, *Methylobacter*, *Methylococcus*, *Methylosinus*, *Methylomicrobium*, 35 *Methylocystis*, *Alcaligenes*, *Synechocystis*, *Synechococcus*, *Anabaena*, *Thiobacillus*, *Methanobacterium*, *Klebsiella*, and *Myxococcus*. Preferred hosts are selected from the group consisting of *Escherichia*, *Bacillus*, and *Methylomonas*.

λ -Red Recombinase System

The λ -Red recombinase system used in the present invention is contained on a helper plasmid (pKD46) and is comprised of three essential genes, *exo*, *bet*, and *gam* (Datsenko and Wanner, *supra*). The *exo* gene encodes an λ -exonuclease, which processively degrades the 5' end strand of double-stranded (ds) DNA and creates 3' single-stranded overhangs. *Bet* encodes for a protein which complexes with the λ -exonuclease and binds to the single stranded DNA and promotes renaturation of complementary strands and is capable of mediating exchange reactions. *Gam* encodes for a protein that binds to the *E.coli*'s RecBCD complex and blocks the complex's endonuclease activity.

The λ -Red system is used in the present invention because homologous recombination in *E.coli* occurs at a very low frequency and usually requires extensive regions of homology. The λ -Red system facilitates the ability to use short regions of homology (10-100 bp) flanking linear dsDNA fragments for homologous recombination. Additionally, the RecBCD complex normally expressed in *E.coli* prevents the use of linear dsDNA for transformation as the complex's exonuclease activity efficiently degrades linear dsDNA. Inhibition of the RecBCD complex's endonuclease activity by *gam* is essential for efficient homologous recombination using linear dsDNA fragments.

Combinatorial P1 Transduction System

Transduction is a phenomenon in which bacterial DNA is transferred from one bacterial cell (the donor) to another (the recipient) by a phage particle containing bacterial DNA. When a population of donor bacteria is infected with a phage, the events of the phage lytic cycle may be initiated. During lytic infection, the enzymes responsible for packaging viral DNA into the bacteriophage sometimes package host DNA. The resulting particle is called a transducing particle. Upon lysis of the cell, a mixture ("P1 lysate") of transducing particles and normal virions are released. When this lysate is used to infect a population of recipient cells, most of the cells become infected with normal virus. However, a small proportion of the population receives transducing particles that inject the DNA they received from the previous host bacterium. This DNA can undergo genetic recombination with the DNA of the other host. Conventional P1 transduction can move only one genetic trait (i.e. gene) at a time (donor to recipient cell).

It will be appreciated that a number of host systems may be used for purposes of the present invention including, but not limited to those with known transducing phages such as *Agrobacterium*, *Erythrobacter*, *Chlorobium*, *Chromatium*, *Flavobacterium*, *Cytophaga*, *Rhodobacter*, *Rhodococcus*, *Streptomyces*, *Brevibacterium*, *Corynebacteria*, *Mycobacterium*, *Deinococcus*, *Paracoccus*, *Escherichia*, *Bacillus*, *Myxococcus*, *Salmonella*, *Yersinia*, *Erwinia*, *Pantoea*, *Pseudomonas*, *Sphingomonas*, *Methylomonas*, *Methylobacter*, *Methylococcus*, *Methylosinus*, *Methylomicrobium*, *Methylocystis*, *Alcaligenes*, *Synechocystis*, *Synechococcus*, *Anabaena*, *Thiobacillus*, *Methanobacterium*, *Klebsiella*, and *Myxococcus*. Phages suitable for use in the present method may include, but are not limited to P1, P2, lambda, ϕ 80, ϕ 3538, T1, T4, P22, P22 derivatives, ES18, Felix "o", P1-CmCs, Ffm, PY20, Mx4, Mx8, PBS-1, PMB-1, and PBT-1.

The present method provides a system for moving multiple genetic traits into a single *E. coli* host in a parallel combinatorial fashion using the bacteriophage P1 mixtures in combination with the site-specific recombinase system for removal of selection markers (Figure 12). After P1 transduction with the P1 lysate mixture made from various donor cells, the transduced recipient cells are screened for antibiotic resistance and assayed for increased production of the desired genetic end product. After selection for the optimized transductants, the antibiotic resistance marker is removed by a site-specific recombinase. The selected transductants can be used again as a recipient cell in additional rounds of P1 transduction in order to engineer multiple chromosomal modifications, optimizing the production of the desired genetic end product. The present combinatorial P1 transduction method enables quick and easy chromosomal trait stacking for optimal production of the desired genetic end product.

Using the method described above, the promoters of the key isoprenoid genes that encode for rate-limiting enzymes involved in the isoprenoid pathway were engineered. Replacement of the endogenous promoters with a strong promoter (P_{T5}) resulted in increased β -carotene production.

An advantage of the present method of promoter replacement is that it allows for multiple chromosomal modifications within the host cell. The system is a means for moving multiple genetic traits into a single host

cell using the bacteriophage P1 transduction in combination with a site-specific recombinase for removal of selection markers (Figures 2 and 12).

The present combinatorial P1 transduction method for promoter replacement enabled isolation and identification of the *ispB* gene and its effect on increasing the production of β -carotene when placed under the control of the strong promoter. The effect of *ispB* on increasing the production of β -carotene was an unexpected and non-obvious result. *IspB* (octaprenyl diphosphate synthase), which synthesizes the precursor of the side chain of the isoprenoid quinones, drains away the FPP substrate from the carotenoid biosynthetic pathway (Figure 1). The mechanism of how overexpression of *ispB* gene under the control of phage T5 strong promoter increases the β -carotene production is not clear yet. However, the result suggests that *IspB* may increase the flux of the carotenoid biosynthetic pathway. Stacking the *ispB* gene under the control of a strong promoter into the chromosome of the engineered *E. coli* strains facilitated a further increase in β -carotene production (Figure 11).

Measurement of the Carotenoid End Product

If the desired genetic end product is a colored product then transformants can be selected for on the basis of colored colonies, and the product can be quantitated by UV/vis spectrometry at the product's characteristic $\lambda_{\max}$ peaks. Alternative analytical methods can also be used including, but not limited to HPLC, CE, GC and GC-MS.

In the present invention, β -carotene was measured by UV/vis spectrometry at β -carotene's characteristic $\lambda_{\max}$ peaks at 425, 450 and 478 nm. The carotenoid was extracted by acetone from the cell pellet. The host strain included a reporter plasmid for the expression of genes involved in the synthesis of β -carotene. The reporter plasmid (pPCB15 or pDCQ108) carried the *Pantoea stewartii crtEXYIB* gene cluster. The gene cluster facilitated the production of β -carotene. Therefore, an increase of carbon flux through the isoprenoid upper pathway will result in an increase in the amount of β -carotene produced; resulting in colonies with more intense color on agar plates when compared to the strain that does not have T5 promoters engineered upstream of the isoprenoid genes. The amount of carotenoid produced was measured by HPLC analysis. Detection of β -carotene was measured by absorption at 450 nm at its respective retention time using HPLC under particular solvent conditions. Quantitative analysis was carried out by comparing the peak area for β -carotene to a known β -carotene standard.

Description of the Preferred Embodiments

E. coli has been genetically modified to create several strains capable of enhanced production of β -carotene. One of the strains has been shown to produce up to 6 mg β -carotene per gram of dry cell weight.

5 Promoter replacement was accomplished using an easy one-step method of bacterial *in vivo* chromosomal engineering using two linear (PCR-generated) DNA fragments in order to increase carotenoid production in a host cell. The fragments were designed to contain short flanking regions of homology between the fragments and the target site on
10 the host (*E. coli*) chromosome. The phage λ -Red recombinase system was expressed on a helper plasmid and under control of an arabinose-inducible promoter for controllable and efficient *in vivo* triple homologous recombination between the two PCR-generated DNA fragments and the host cell's chromosome. At least one of the two linear double stranded
15 (ds) DNA fragments used during recombination was designed to contain a selective marker (kanamycin) flanked by site-specific recombinase sequences (*FRT*)(Example 1). The selectable marker permitted the identification and selection of the cells that had undergone the desired recombination event. The constructs of the selected recombinants were
20 verified by sequence analysis. The selective marker was excised by a second helper plasmid (pCP20) containing the site-specific recombinase gene under the control of the P_R promoter of λ phage (Examples 6-12 and 17).

A strong promoter (phage P_{T5}) was placed upstream of the *E. coli*
25 target genes *dxs*, *idi*, *ygbBygbP*, *ispB*, *ispAdxs* (Example 1) via triple homologous recombination using two (PCR-generated) linear dsDNA fragments and the targeted chromosomal DNA (Figures 2). In each example, one of the two fragments contained a kanamycin resistance marker flanked by site-specific *FRT* recombinase sequences. Flanking
30 the site-specific recombinase sequences were homology arms which contained short (approximately 10-50 bp) regions of homology. A first recombination region (homology arm #1) was linked to the 5'-end of the first fragment. A second recombination region (homology arm #2) was linked to the 3'-end of the first fragment. The second PCR generated
35 linear dsDNA fragment contained the P_{T5} strong promoter. The third recombination region (homology arm #3) was linked to the 3'-end of the second fragment. The first recombination region (homology arm #1) had homology to an upstream portion of the native bacterial chromosomal

promoter targeted for replacement. The second recombination region (homology arm #2 located on the 3'-end of the first fragment) had homology to the 5'-end portion of the second fragment. The third recombination region (homology arm #3) had homology to a downstream portion of the native bacterial chromosomal promoter targeted for replacement (Figure 2).

The recombination proficient *E. coli* host (containing the λ -Red recombination system on the helper plasmid pKD46) was transformed with the two PCR-generated fragments resulting in the chromosomal replacement of the targeted native promoter with the construct containing the kanamycin selectable marker of the first fragment and the P_{T5} strong promoter of the second fragment (Examples 1 and 6-12, Figure 2). The promoter replacement resulted in the formation of an augmented *E. coli* chromosomal gene (either *dxs*, *idi*, *ygbBygbP*, *ispB* or *ispAdxs* genes), operably linked to the introduced non-native promoter. The bacterial host cells that had undergone the desired recombination event were selected according to the expression of the selectable marker and their ability to grow in selected media. The selected recombinants were then transformed with a second helper plasmid, pCP20 (Cherepanov and Wackernagel, *supra*), expressing the flippase (Flp) site-specific recombinase which excised the selectable marker (Examples 6-12). The constructs were confirmed via PCR fragment analysis (Figures 3-5). The recombinant bacterial host cell containing the augmented isoprenoid genes (*dxs*, *idi*, *ygbBygbP*, *ispB* or *ispAdxs*) and the carotenoid reporter plasmid (pPCB15) was then tested for increased production of β -carotene. Placement of one or more of the *E. coli* *dxs*, *idi*, *ygbBygbP*, *ispB* or *ispAdxs* genes (normally expressed at very low levels) under control of the strong P_{T5} promoter resulted in significant increases in β -carotene production (Examples 18-19, Figure 11).

In another embodiment, the method was used to simultaneously add a foreign gene and promoter. The first of the two PCR-generated fragments was designed so that it contained the fusion product of a selectable marker (kanamycin) and promoter (P_{T5}) (Example 2, Figure 2)). The second PCR-generated fragment contained the fusion product of a selectable marker (*kan-P_{T5}*) and the *Methylobionas* 16a *dxs*(16a) (SEQ ID NO:13), *dxr*(16a) (SEQ ID NO:17) or *lytB*(16a) (SEQ ID NO:15) genes (foreign to *E. coli*). Once again, homology arms were designed to allow for precise incorporation into the host bacterial chromosome. The desired

recombinants were selected by methods previously described. The selectable marker was then removed by a site-specific recombinase as previously described. The recombinant constructs were confirmed by PCR fragment analysis. β -carotene production in the transformed *E. coli* reporter strain was measured as previously described. Cells containing the *Methylomonas* 16a *dxs*(16a) and/or *lytB*(16a) genes (homologous to the *E. coli* *dxs* and *lytB* genes) under the control of the P_{T5} promoter exhibited an increase in β -carotene production (Figure 11). The present method was useful in the simultaneous addition of a foreign promoter and gene. Subsequent removal of the selectable marker is required so that the process can be repeated, if desired, to engineer bacterial biosynthetic pathways for increased production of the desired product.

In another embodiment, the bacterial host strain was engineered to contain multiple chromosomal modifications, including multiple promoter and gene additions or replacements so that the production efficiency of the desired final product is increased. In a preferred embodiment, the incorporated or augmented chromosomal genes encode for enzymes useful for the production of carotenoids.

In another preferred embodiment the constructs made by chromosomal engineering of non-endogenous promoters upstream of isoprenoid genes and chromosomally integrating non-endogenous isoprenoid pathway genes into the host chromosome are combined into a single strain. The phage *T5* strong promoter (P_{T5})-*ispAdxs* P_{T5} -*idi*, P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a), P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a), P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a) P_{T5} -*idi*, P_{T5} -*dxs* P_{T5} -*idi*, P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBygbP*, P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBygbP* P_{T5} -*lytB*(16a), P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBygbP* *yjeR*::*Tn5*, and P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBygbP* P_{T5} -*ispB* were constructed by combinatorial stacking. Stacking of these constructs in a combinatorial manner facilitated the development of engineered host strains capable of significantly increased carotenoid production.

In another embodiment, gene loci carrying transposon insertions that confer the ability to increase carotenoid production were engineered into the host chromosome. The *E. coli* *yjeR* gene carrying a *Tn5* transposon insertion sequence (*yjeR*::*Tn5*; SEQ ID NO:63) was stacked in combination with P_{T5} -*dxs*, P_{T5} -*idi* and P_{T5} -*ygbBygbP* to create a strain producing 19-fold higher levels of β -carotene (ATCC PTA-4807).

In another embodiment, an *E. coli* reporter strain was constructed for assaying β -carotene production. Briefly, the reporter strain was created by cloning the gene cluster *crtEXYIB* from *Pantoea stewartii* into a reporter plasmid (pPCB15) that was subsequently used to transform the *E. coli* host (Figure 7). The cluster contained many of the genes required for the synthesis of carotenoids, producing β -carotene in the transformed *E. coli*. It should be noted that the *crtZ* gene (β -carotene hydroxylase) was included in the gene cluster. However, since no promoter was present to express the *crtZ* gene (organized in opposite orientation and adjacent to *crtB* gene), no zeaxanthin was produced. The zeaxanthin glucosyl transferase enzyme (encoded by the *crtX* gene located within the gene cluster) had no substrate for its reaction. Increases in β -carotene production were reported as increases relative to the control strain production (Figure 11).

In another embodiment, a new reporter plasmid was created. Reporter plasmid pPCB15, used for many of the experiments, is considered a low copy number plasmid. A new medium-copy number reporter plasmid was generated, (pDCQ108) that also contained the *Pantoea stewartii crtEXYIB* gene cluster (Example 19). Plasmid pDCQ108 was then used as the reporter plasmid in *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBygbP* P_{T5} -*ispB* leading to an approximately 30-fold increase in β -carotene production when compared to the control strain (Figure 11; Examples 20 and 21; Table 9)).

It has been speculated that the limits for carotenoid production in non-carotenogenic host such as *E. coli* had been reached at the level of around 1.5 mg/g cell dry weight (1,500 ppm) due to overload of the membranes and blocking of membrane functionality (Albrecht et al., *supra*). The present method has solved the stated problem by making modifications on the *E. coli* chromosome that resulted in increased β -carotene production of up to 6 mg per gram dry cell weight (6,000 ppm), an increase of 30-fold over initial levels with no lethal effect. The bacterial production of 6,000 ppm carotenoids is much higher than the maximum accepted limit (1,600 ppm) for carotenoid production in bacteria.

One of skill in the art will recognize that the present method can be applied to a variety of hosts in addition to *E. coli*. Use of the present method in other hosts is supported by the fact that: 1) the isoprenoid pathway is common in bacteria, 2) the λ -Red system has been reported to

work in a variety of hosts, and 3) phage transduction is known to occur in many hosts.

EXAMPLES

The present invention is further defined in the following Examples.

5 It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and
10 conditions.

GENERAL METHODS

Standard recombinant DNA and molecular cloning techniques used in the Examples are well known in the art and are described by Sambrook,
15 J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, (1989) (Maniatis) and by T. J. Silhavy, M. L. Bannan, and L. W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) and by Ausubel, F. M. et al., Current Protocols in Molecular Biology, pub. by Greene Publishing Assoc. and Wiley-
20 Interscience (1987).

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for
25 General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994)) or by Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, MA
30 (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee; WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

35 Manipulations of genetic sequences were accomplished using the suite of programs available from the Genetics Computer Group Inc. (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI). Where the GCG program "Pileup" was used the gap

creation default value of 12, and the gap extension default value of 4 were used. Where the CGC "Gap" or "Bestfit" programs were used the default gap creation penalty of 50 and the default gap extension penalty of 3 were used. Multiple alignments were created using the FASTA program
5 incorporating the Smith-Waterman algorithm (W. R. Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.] (1994), Meeting Date 1992, 111-120. Editor(s): Suhai, Sandor. Publisher: Plenum, New York, NY). In any case where program parameters were not prompted for, in these or any other programs, default values were used.

10 The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "μL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), and "rpm" means revolutions per minute.

EXAMPLE 1

15 Construction of *E. coli* Strains with the phage P_{T5} Promoter Chromosomally-integrated Upstream of the Isoprenoid Genes (Promoter Replacement)

The native promoters of the *E. coli* isoprenoid genes *dxs*, *idi*, *ygbBygbP*, *ispB*, and *ispAdxs*, (Figure 1) were replaced with the (P_{T5})
20 promoter using two PCR-fragments chromosomal integration method as described in Figure 2. The method for replacement is based on homologous recombination via the λ -Red recombinase encoded on a helper plasmid. Recombination occurs between the *E. coli* chromosome and two PCR fragments that contain 20-50 bp homology patches at both
25 ends of PCR fragments (Figure 2). For integration of the P_{T5} promoter upstream of these genes, a two PCR fragment method was employed. In this method, the two linear fragments included a DNA fragment (1489 bp) containing a kanamycin selectable marker (*kan*) flanked by site-specific recombinase target sequences (*FRT*) and a DNA fragment (154 bp)
30 containing a phage *T5* promoter (P_{T5}) comprising the -10 and -35 consensus promoter sequences, lac operator (*lacO*), and a ribosomal binding site (*rbs*).

By using the two PCR fragment method, the kanamycin selectable marker and P_{T5} promoter (*kan-P_{T5}*) were integrated upstream of the *dxs*,
35 *idi*, *ygbBP*, *ispB*, and *ispAdxs* genes, yielding *kan-P_{T5}-dxs*, *kan-P_{T5}-idi*, *kan-P_{T5}-ygbBP*, *kan-P_{T5}-ispB*, and *kan-P_{T5}-ispAdxs*. The linear DNA fragment (1489 bp) containing a kanamycin selectable marker was

synthesized by PCR from plasmid pKD4 (Datsenko and Wanner, *supra*) with primer pairs as follows in Table 3.

TABLE 3
Primers for Amplification of the Kanamycin Selectable Marker

Primer Name	Primer Sequence	SEQ ID NO:
5'-kan(dxs)	<u>TGGAAGCGCTAGCGGACTACATCATCCA</u> <u>GCGTAATAAATAACGTCTTGAGCGATTGT</u> <u>GTAG</u> ¹	21
5'-kan(idi)	<u>TCTGATGCGCAAGCTGAAGAAAAATGAGC</u> <u>ATGGAGAATAATATGACGTCTTGAGCGAT</u> <u>TGTGTAG</u> ¹	22
5'-kan(ygbBP)	<u>GACGCGTCGAAGCGCGCACAGTCTGCGG</u> <u>GGCAAAACAATCGATAACGTCTTGAGCGA</u> <u>TTGTGTAG</u> ¹	23
5'-kan(ispAdxs)	<u>ACCATGACGGGGCGAAAAATATTGAGAG</u> <u>TCAGACATTCATGTGTAGGCTGGAGCTGC</u> <u>TTC</u> ¹	24
3'-kan	<u>GAAGACGAAAGGGCCTCGTGATACGCCT</u> <u>ATTTTATAGGTTATATGAATATCCTCCTT</u> <u>AGTTCC</u> ²	25

¹ The underlined sequences illustrate each respective homology arm chosen to match sequences in the upstream region of the chromosomal integration site, while the remainder is the priming sequence

² The underlined sequences illustrate homology arm chosen to match sequences in the 5'-end region of the *T5* promoter DNA fragment

The second linear DNA fragment (154 bp) containing the *P*_{T5} promoter was synthesized by PCR from pQE30 (QIAGEN, Inc. Valencia, CA) with primer pairs as follows in Table 4.

TABLE 4
Primers for Amplification of the *P*_{T5} Promoter

Primer Name	Primer Sequence	SEQ ID NO:
5'-T5	<u>CTAAGGAGGATATTCATATAACCTATAAAA</u> <u>ATAGGCGTATCACGAGGCCC</u> ¹	26
3'-T5(dxs)	<u>GGAGTCGACCAGTGCCAGGGTCGGGTATT</u> <u>TGGCAATATCAAACTCATAGTTAATTTCTC</u> <u>CTCTTTAATG</u> ²	27
3'-T5(idi)	<u>TGGGAACTCCCTGTGCATTCAATAAAATGA</u> <u>CGTGTTCCGTTTGCATAGTTAATTTCTCCT</u>	28

<u>Primer Name</u>	<u>Primer Sequence</u>	<u>SEQ ID NO:</u>
	CTTTAATG ²	
3'- T5(ygbBP)	<u>CGGCCGCCGGAACACGGCGCAAACATC</u> <u>CAAATGAGTGGTTGCCATAGTTAATTTCTC</u> CTCTTTAATG ²	29
3'- T5(ispAdxs)	<u>CCTGCTTAACGCAGGCTTCGAGTTGCTGC</u> <u>GGAAAGTCCATAGTTAATTTCTCCTCTTTA</u> ATG ²	30

¹ The underlined sequences illustrate homology arm chosen to match sequences in the 3'-end region of the kanamycin DNA fragment

² The underlined sequences illustrate each respective homology arm chosen to match sequences in the downstream region of the chromosomal integration site

5

The linear DNA fragment (1,647 bp) containing fused kanamycin selectable marker-phage T5 promoter is synthesized by PCR from pSUH5 with primer pairs as follows in Table 5. The pSUH5 plasmid (Figure 6; SEQ ID NO:66) was constructed by cloning a phage T5 promoter (P_{T5}) region (SEQ ID NO:33) into the *NdeI* restriction endonuclease site of pKD4 (Datsenko and Wanner, *supra*).

10

TABLE 5

Primers for Amplification of the Fused Kanamycin Selectable Marker-
Phage P_{T5} Promoter

15

<u>Primer Name</u>	<u>Primer Sequence</u>	<u>SEQ ID NO:</u>
5'- kanT5(ispB)	<u>ACCATAAACCCCTAAGTTGCCTTTGTTTACA</u> <u>GTAAGGTAATCGGGGCGTCTTGAGCGATT</u> GTGTAG ¹	31
3'- kanT5(ispB)	<u>CGCCATATCTTGC GCGGTAACTCATTGA</u> <u>TTTTTTCTAAATTCATAGTTAATTTCTCCTC</u> TTTAATG ²	32

¹ The underlined sequences illustrate each respective homology arm chosen to match sequences in the upstream region of the chromosomal integration site.

² The underlined sequences illustrate each respective homology arm chosen to match sequences in the downstream region of the chromosomal integration site.

20

Standard PCR conditions were used to amplify the linear DNA fragments with AmpliTaq Gold® polymerase (Applied Biosystems, Foster City, CA) as follows:

<u>PCR reaction:</u>	<u>PCR reaction mixture:</u>
Step1 94°C 3 min	0.5 μ L plasmid DNA
Step2 93°C 30 sec	5 μ L 10X PCR buffer
Step3 55°C 1 min	1 μ L dNTP mixture (10 mM)
5 Step4 72°C 3 min	1 μ L 5'-primer (20 μ M)
Step5 Go To Step2, 30 cycles	1 μ L 3'-primer (20 μ M)
Step6 72°C 5 min	0.5 μ L AmpliTaq Gold®
polymerase	
	41 μ L sterilized dH ₂ O

10

After completing the PCR reactions, 50 μ L of each PCR reaction mixture was run on a 1% agarose gel and the PCR products were purified using the QIAquick Gel Extraction Kit™ as per the manufacturer's instructions (Cat. # 28704, QIAGEN Inc., Valencia, CA). The PCR products were eluted with 10 μ L of distilled water. The DNA Clean & Concentrator™ kit (Zymo Research, Orange, CA) was used to further purify the PCR product fragments as per the manufacturer's instructions. The PCR products were eluted with 6-8 μ L of distilled water to a concentration of 0.5-1.0 μ g/ μ L.

20

The *E. coli* MC1061 strain, carrying the λ -Red recombinase expression plasmid pKD46 (amp^R) (SEQ ID NO:65) was used as a host strain for the chromosomal integration of the PCR fragments. The strain was constructed by transformation of *E. coli* strain MC1061 with the λ -Red recombinase expression plasmid, pKD46 (amp^R). Transformants were selected on 100 μ g/mL ampicillin LB plates at 30°C.

25

For transformation, electroporation was performed using 1-5 μ g of the purified PCR products carrying the kanamycin marker and *P*_{T5} promoter. Approximately one-half of the cells transformed were spread on LB plates containing 25 μ g/mL kanamycin in order to select antibiotic-resistant transformants. After incubating the plate at 37°C overnight, antibiotic-resistance transformants were selected as follows: 10 colonies of *kan-P*_{T5}-*dxs*, 12 colonies of *kan-P*_{T5}-*idi*, 10 colonies of *kan-P*_{T5}-*ygbBP*, 3 colonies of *kan-P*_{T5}-*ispB*, and 19 colonies of *kan-P*_{T5}-*ispA*.

30

PCR analysis was used to confirm the integration of both the kanamycin selectable marker and the *P*_{T5} promoter in the correct location on the *E. coli* chromosome. For PCR, a colony was resuspended in 50 μ L of PCR reaction mixture containing 200 μ M dNTPs, 2.5 U AmpliTaq™ (Applied Biosystems), and 0.4 μ M of specific primer pairs. Test primers

35

were chosen to match sequences of the regions located in the kanamycin (5'-primer) and the early coding-region of each isoprenoid gene (3'-primer) (Figure 3). Sequences of these primers are listed in Tables 3, 4, and 5 above and the PCR reaction was performed as described above. The resultant *E. coli* strains carrying each *kan-P_{T5}*-isoprenoid gene fusion on the chromosome were used for stacking multiple *kan-P_{T5}*-isoprenoid gene fusions on the chromosome to construct *E. coli* strain for increasing β -carotene production as described in Examples 6-12 and 17.

EXAMPLE 2

Construction of *E. coli* Strains with *Methylobionas* 16A *dxs*(16A), *dxr*(16A) and *lytB*(16A) Genes Chromosomally-Integrated

Methylobionas 16a (ATCC PTA-2402) isoprenoid genes *dxs*, *dxr* and *lytB* (WO 02/20733 A2), with *dxs* (denoted as "*dxs*(16a)" and described as SEQ ID NO:13), *dxr* (denoted as "*dxr*(16a)" and described as SEQ ID NO:17), and *lytB* (denoted as "*lytB*(16a)" and described by SEQ ID NO:15), and the fused *kan-P_{T5}* promoter were co-integrated into the inter-operon regions located at 30.9, 78.6 and 18.1 min, respectively, of the *E. coli* chromosome using the two PCR-fragments chromosomal integration method as described in Figure 2. The principle for chromosomal integration of foreign gene is same as described in Example 1.

The linear DNA fragment (1,647 bp) containing fused kanamycin selectable marker- *P_{T5}* promoter was synthesized by PCR from pSUH5 with primer pairs as follows in Table 6. The pSUH5 plasmid (Figure 6) was constructed by cloning a *P_{T5}* promoter region (SEQ ID NO:33) into the *Nde*I restriction endonuclease site of pKD4 (Datsenko and Wanner, *supra*).

TABLE 6

Primers for Amplification of the Fused Kanamycin Selectable Marker- *P_{T5}* Promoter

Primer Name	Primer Sequence	SEQ ID NO:
5'- kanT5(<i>dxs</i> 16a)	CACTAACGCCCCGCACATTGCTGCGGGC TTTTTGATTCATTTTCGCACGTCTTGAGC GATTGTGTAG ¹	34
5'- kanT5(<i>dxr</i> 16a)	TAAAGGGCTAAGAGTAGTGTGCTCTTA GCCCTTAATTACGTTTCCCGTCTTGAGC	35

<u>Primer Name</u>	<u>Primer Sequence</u>	<u>SEQ ID NO:</u>
	GATTGTGTAG ¹	
5'- kanT5(lytB16a)	<u>CTACAACTGGCGAGATGCATAGCGAGT</u> <u>ATAATTTGTATTTTTCGTCGTCTTGAGC</u> GATTGTGTAG ¹	36
3'- kanT5(dxs16a)	<u>AGTAGAGGGAAGTCTTTGGAAAGAGCC</u> <u>ATAGTTAATTTCTCCTCTTTAATG</u> ²	37
3'- kanT5(dxr16a)	<u>ACGGTGCCGCCGCAATGATGCTGTCCA</u> <u>CCAGTTAATTTCTCCTCTTTAATG</u> ²	38
3'- kanT5(lytB16a)	<u>CCACGGGGGTTTGCGAGTACGATTTC</u> <u>ATAGTTAATTTCTCCTCTTTAATG</u> ²	39

¹ The underlined sequences illustrate each respective homology arm chosen to match sequences in the upstream region of the chromosomal integration site, while the remainder is the priming sequence

² The underlined sequences illustrate homology arm chosen to match sequences in the 5'-end region of the foreign gene DNA fragment

The linear DNA fragment containing *Methylomonas* 16a *dxs*, *dxr* or *lytB* gene was synthesized by PCR from *Methylomonas* 16a (ATCC PTA-2402) genomic DNA with primer pairs as follows in Table 7.

TABLE 7
Primers for Amplification of the Foreign Gene

<u>Primer Name</u>	<u>Primer Sequence</u>	<u>SEQ ID NO:</u>
5'-dxs16a	<u>ACAGAATTCATTAAAGAGGAGAAATTA</u> <u>ACT</u> ATGGCTCTTTCCAAAGAC TTCCCTC ¹	40
5'-dxr16a	<u>ACAGAATTCATTAAAGAGGAGAAATTA</u> <u>ACT</u> GGTGGACAGCATCATTGCGGCGGCA ¹	41
5'-lytB16a	<u>ACAGAATTCATTAAAGAGGAGAAATTA</u> <u>ACT</u> ATGCAAATCGTACTCGCAAACCC ¹	42
3'-dxs16a	<u>AGGAGCGAAGTGATTATCAGTATGCTGTT</u> <u>C</u> ATATAGCCTCGAATTATCAAGCGCAAACT GTTGATG ²	43
3'-dxr16a	<u>GGCATTTTCACTCTGGCAATGCGCATAA</u> <u>AC</u> GCTTTCAAAGTCCTGTTAAGCTACCAAGG TTGATG ²	44
3'-lytB16a	<u>AGTGGCGGACGGGCAAACAAGGGTAAC</u> <u>AT</u> <u>AGGATCAATGAGGGTTATTGATCACGCT</u> <u>TG</u> CATATGTTT ²	45

¹ The underlined sequences illustrate homology arm chosen to match sequences in the 3'-end region of the fused kanamycin-phage *P_{T5}* promoter DNA fragment

² The underlined sequences illustrate each respective homology arm chosen to match sequences in the downstream region of the chromosomal integration site

The PCR reaction, purification and electro-transformation were performed as described in Example 1. Kanamycin-resistance transformants were selected including 7 colonies of *E. coli kan-P_{T5}-dxs(16a)*, 3 colonies of *E. coli kan-P_{T5}-dxr(16a)* and 12 colonies of *E. coli kan-P_{T5}-lytB(16a)*. Among these, the colonies that have a correct integration of *kan-P_{T5}-dxs(16a)*, *kan-P_{T5}-dxr(16a)* or *kan-P_{T5}-lytB(16a)* into the target site of *E. coli* chromosome was selected by PCR analysis (Figure 3, 4, and 5).

EXAMPLE 3

Cloning of β -Carotene Production Genes from *Pantoea stewartii*

Primers were designed using the sequence from *Erwinia uredovora* to amplify a fragment by PCR containing the *crt* genes. These sequences included 5'-3':

ATGACGGTCTGCGCAAAAAACACG SEQ ID NO:19
GAGAAATTATGTTGTGGATTGGAATGC SEQ ID NO:20

Chromosomal DNA was purified from *Pantoea stewartii* (ATCC no. 8199) and *Pfu* Turbo polymerase (Stratagene, La Jolla, CA) was used in a PCR amplification reaction under the following conditions: 94°C, 5 min; 94°C (1 min)-60°C (1 min)-72°C (10 min) for 25 cycles, and 72°C for 10 min. A single product of approximately 6.5 kb was observed following gel electrophoresis. *Taq* polymerase (Perkin Elmer, Foster City, CA) was used in a ten minute 72°C reaction to add additional 3' adenosine nucleotides to the fragment for TOPO cloning into pCR4-TOPO (Invitrogen, Carlsbad, CA) to create the plasmid pPCB13. Following transformation to *E. coli* DH5 α (Life Technologies, Rockville, MD) by electroporation, several colonies appeared to be bright yellow in color indicating that they were producing a carotenoid compound. Following plasmid isolation as instructed by the manufacturer using the Qiagen (Valencia, CA) miniprep kit, the plasmid containing the 6.5 kb amplified fragment was transposed with pGPS1.1 using the GPS-1 Genome Priming System kit (New England Biolabs, Inc., Beverly, MA). A number of these transposed plasmids were sequenced from each end of the transposon. Sequence was generated on an ABI Automatic sequencer using dye terminator technology (US 5,366,860; EP 272007) using transposon specific primers. Sequence assembly was performed with the Sequencher program (Gene Codes Corp., Ann Arbor MI).

EXAMPLE 4

Identification and Characterization of Bacterial Genes

Genes encoding *crtE*, *X*, *Y*, *I*, *B*, and *Z* were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., *J. Mol. Biol.* 215:403-410 (1993)) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank® CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The sequences obtained in Example 3 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D., *Nature Genetics*, 3:266-272 (1993)) provided by the NCBI.

All comparisons were done using either the BLASTNnr or BLASTXnr algorithm. The results of the BLAST comparison are given in Table 7 which summarize the sequences to which they have the most similarity. Table 7 displays data based on the BLASTXnr algorithm with values reported in expect values. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

TABLE 8

ORF Name	Gene Name	Similarity Identified	SEQ ID No. base	SEQ ID No. Peptide	% Identity ^a	% Similarity ^b	E-value ^c	Citation
1	<i>crtE</i>	Geranylgeranyl pyrophosphate synthetase (or GGPP synthetase, or farnesyltransferase) EC 2.5.1.29 gi 117509 sp P21684 CRTE_PANAN GERANYLGERANYL PYROPHOSPHATE SYNTHETASE (GGPP SYNTHETASE) (FARNESYLTRANSFERASE) Zeaxanthin glucosyl transferase EC 2.4.1.-	1	2	83	88	e-137	Misawa et al., <i>J. Bacteriol.</i> 172 (12), 6704-6712 (1990)
2	<i>crtX</i>	gi 1073294 pir S52583 crtX protein - <i>Erwinia herbicola</i> Lycopene cyclase gi 1073295 pir S52585 lycopene cyclase - <i>Erwinia herbicola</i> Phytoene desaturase EC 1.3.-.- gi 1073299 pir S52586 phytoene dehydrogenase (EC 1.3.-.-) - <i>Erwinia herbicola</i> Phytoene synthase EC 2.5.1.-	3	4	75	79	0.0	Lin et al., <i>Mol. Gen. Genet.</i> 245 (4), 417-423 (1994)
3	<i>crtY</i>	gi 1073294 pir S52583 crtX protein - <i>Erwinia herbicola</i> Lycopene cyclase gi 1073295 pir S52585 lycopene cyclase - <i>Erwinia herbicola</i> Phytoene desaturase EC 1.3.-.- gi 1073299 pir S52586 phytoene dehydrogenase (EC 1.3.-.-) - <i>Erwinia herbicola</i> Phytoene synthase EC 2.5.1.-	5	6	83	91	0.0	Lin et al., <i>Mol. Gen. Genet.</i> 245 (4), 417-423 (1994)
4	<i>crtI</i>	gi 1073294 pir S52583 crtX protein - <i>Erwinia herbicola</i> Lycopene cyclase gi 1073295 pir S52585 lycopene cyclase - <i>Erwinia herbicola</i> Phytoene desaturase EC 1.3.-.- gi 1073299 pir S52586 phytoene dehydrogenase (EC 1.3.-.-) - <i>Erwinia herbicola</i> Phytoene synthase EC 2.5.1.-	7	8	89	91	0.0	Lin et al., <i>Mol. Gen. Genet.</i> 245 (4), 417-423 (1994)
5	<i>crtB</i>	gi 1073294 pir S52583 crtX protein - <i>Erwinia herbicola</i> Lycopene cyclase gi 1073295 pir S52585 lycopene cyclase - <i>Erwinia herbicola</i> Phytoene desaturase EC 1.3.-.- gi 1073299 pir S52586 phytoene dehydrogenase (EC 1.3.-.-) - <i>Erwinia herbicola</i> Phytoene synthase EC 2.5.1.-	9	10	88	92	e-150	Lin et al., <i>Mol. Gen. Genet.</i> 245 (4), 417-423 (1994)

6	<i>crtZ</i>	Beta-carotene hydroxylase gil117526 sp P21688 CRTZ_PANAN BETA-CAROTENE HYDROXYLASE	11	12	88	91	3e-88	Misawa et al., J. Bacteriol. 172 (12), 6704-6712 (1990)
---	-------------	---	----	----	----	----	-------	---

a% Identity is defined as percentage of amino acids that are identical between the two proteins.
b% Similarity is defined as percentage of amino acids that are identical or conserved between the two proteins.
cExpect value. The Expect value estimates the statistical significance of the match, specifying the number of matches, with a given score, that are expected in a search of a database of this size absolutely by chance.

EXAMPLE 5

Analysis of Gene Function by Transposon Mutagenesis

Several plasmids carrying transposons which were inserted into each coding region including *crtE*, *crtX*, *crtY*, *crtI*, *crtB*, and *crtZ* were
5 chosen using sequence data generated in Example 3. These plasmid variants were transformed to *E. coli* MG1655 and grown in 100 mL Luria-Bertani broth in the presence of 100 µg/mL ampicillin. Cultures were grown for 18 hr at 26°C, and the cells were harvested by centrifugation. Carotenoids were extracted from the cell pellets using 10 mL of acetone.
10 The acetone was dried under nitrogen and the carotenoids were resuspended in 1 mL of methanol for HPLC analysis. A Beckman System Gold® HPLC with Beckman Gold Nouveau Software (Columbia, MD) was used for the study. The crude extraction (0.1 mL) was loaded onto a 125 x 4 mm RP8 (5 µm particles) column with corresponding guard
15 column (Hewlett-Packard, San Fernando, CA). The flow rate was 1 mL/min, while the solvent program used was: 0-11.5 min 40% water/60% methanol; 11.5-20 min 100% methanol; 20-30 min 40% water/60% methanol. The spectrum data were collected by the Beckman photodiode array detector (model 168).

20 In the clone with wild type *crtEXYIBZ*, the carotenoid was found to have a retention time of 15.8 min and an absorption spectra of 450 nm, 475 nm. This was the same value observed in comparison to the β-carotene standard. This suggested that *crtZ* gene organized in the opposite orientation was not expressed in this construct. The transposon
25 insertion in *crtZ* had no effect as expected (data not shown).

HPLC spectral analysis also revealed that a clone with transposon insertion in *crtX* also produced β-carotene. This is consistent with the proposed function of *crtX* encoding a zeaxanthin glucosyl transferase enzyme at a later step of the carotenoid pathway following synthesis of β-carotene.
30

The transposon insertion in *crtY* did not produce β-carotene. The carotenoid's elution time (15.2 min) and absorption spectra (443 nm, 469 nm, 500 nm) agree with those of the lycopene standard. Accumulation of lycopene in the *crtY* mutant confirmed the role *crtY* as a
35 lycopene cyclase encoding gene.

The *crtI* extraction, when monitored at 286 nm, had a peak with retention time of 16.3 min and with absorption spectra of 276 nm, 286 nm, 297 nm, which agrees with the reported spectrum for phytoene. Detection

of phytoene in the *crtI* mutant confirmed the function of the *crtI* gene as one encoding a phytoene dehydrogenase enzyme.

The acetone extracted from the *crtE* mutant or *crtB* mutant was clear. Loss of pigmented carotenoids in these mutants indicated that both the *crtE* gene and *crtB* genes are essential for carotenoid synthesis. No carotenoid was observed in either mutant, which is consistent with the proposed function of *crtB* encoding a prephytoene pyrophosphate synthase and *crtE* encoding a geranylgeranyl pyrophosphate synthetase. Both enzymes are required for β -carotene synthesis.

Results of the transposon mutagenesis experiments are shown below in Table 9. The site of transposon insertion into the gene cluster *crtEXYIB* is recorded, along with the color of the *E. coli* colonies observed on LB plates, the identity of the carotenoid compound (as determined by HPLC spectral analysis), and the experimentally assigned function of each gene.

Table 9
Transposon Insertion Analysis of Carotenoid Gene Function

Transposon insertion site	Colony color	Carotenoid observed by HPLC	Assigned gene function
Wild Type (with no transposon insertion)	Yellow	β -carotene	
<i>crtE</i>	White	None	Geranylgeranyl pyrophosphate synthetase
<i>crtB</i>	White	None	Prephytoene pyrophosphate synthase
<i>crtI</i>	White	Phytoene	Phytoene dehydrogenase
<i>crtY</i>	Pink	Lycopene	Lycopene cyclase
<i>crtZ</i>	Yellow	β -carotene	β -carotene hydroxylase
<i>crtX</i>	Yellow	β -carotene	Zeaxanthin glucosyl transferase

EXAMPLE 6
Construction of *E. coli* P_{T5} -*ispAdxs* P_{T5} -*idi* Strain for Increased β -Carotene Production

In order to characterize the effect of the chromosomal integration of the P_{T5} promoter in the front of the isoprenoid genes on β -carotene production, a strain (*E. coli* P_{T5} -*ispAdxs* P_{T5} -*idi*) containing a chromosomally integrated P_{T5} promoter upstream from *ispAdxs* and *idi* genes and capable of producing β -carotene was constructed.

First, P1 lysate of the *E. coli kan-P_{T5}-ispAdxs* strain was prepared by infecting a growing culture of bacteria with the P1 phage and allowing the cells to lyse. For P1 infection, *E. coli kan-P_{T5}-ispAdxs* strain was inoculated in 4 mL LB medium with 25 µg/mL kanamycin, grown at 37°C overnight, and then sub-cultured with 1:100 dilution of an overnight culture in 10 mL LB medium containing 5 mM CaCl₂. After 20-30 min of growth at 37°C, 10⁷ P1_{vir} phages were added. The cell-phage mixture was aerated for 2-3 h at 37°C until lysed, several drops of chloroform were added and the mixture vortexed for 30 sec and incubated for an additional 30 min at room temp. The mixture was then centrifuged for 10 min at 4500 rpm, and the supernatant transferred into a new tube to which several drops of chloroform were added.

Second, P1 lysate made on *E. coli kan-P_{T5}-ispAdxs* strain was transduced into the recipient strain, *E. coli* MG1655 containing a β-carotene biosynthesis expression plasmid pPCB15 (cam^R) (Figure 6). The plasmid pPCB15 (cam^R) encodes the carotenoid biosynthesis gene cluster (*crtEXYIB*) from *Pantoea Stewartii* (ATCC no. 8199). The pPCB15 plasmid was constructed from ligation of *Sma*I digested pSU18 (Bartolome et al., *Gene*, 102:75-78 (1991)) vector with a blunt-ended *Pme*I/*Not*I fragment carrying *crtEXYIB* from pPCB13 (Example 3). The *E. coli* MG1655 pPCB15 recipient cells were grown to mid-log phase (1-2 x 10⁸ cells/ml) in 4 mL LB medium with 25 µg/mL chloramphenicol at 37°C. Cells were spun down for 10 min at 4500 rpm and resuspended in 2 mL of 10 mM MgSO₄ and 5 mM CaCl₂. Recipient cells (100 µL) were mixed with 1 µL, 10 µL, or 100 µL of P1 lysate stock (10⁷ pfu/µL) made from the *E. coli kan-P_{T5}-ispAdxs* strain and incubated at 30°C for 30 min. The recipient cell-lysate mixture was spun down at 6500 rpm for 30 sec, resuspended in 100 µL of LB medium with 10 mM of sodium citrate, and incubated at 37°C for 1 h. Cells were plated on LB plates containing both 25 µg/mL kanamycin and 25 µg/mL chloramphenicol in order to select for antibiotic-resistant transductants and incubated at 37°C for 1 or 2 days. Six kanamycin-resistance transductants were selected.

To eliminate kanamycin selectable marker from the chromosome, a FLP recombinase expression plasmid pCP20 (amp^R) (ATCC PTA-4455) (Cherepanov and Wackernagel, *supra*), which has a temperature-sensitive replication of origin, was transiently transformed into one of the kanamycin-resistant transductants by electroporation. Cells were spread onto LB agar containing 100 µg/mL ampicillin and 25 µg/mL

chloramphenicol LB plates, and grown at 30°C for 1 day. Colonies were picked and streaked on 25 µg/mL chloramphenicol LB plates without ampicillin antibiotics and incubated at 43°C overnight. Plasmid pCP20 has a temperature sensitive origin of replication and was cured from the host cells by culturing cells at 43°C. The colonies were tested for ampicillin and kanamycin sensitivity to test loss of pCP20 and kanamycin selectable marker by streaking colonies on 100 µg/mL ampicillin LB plate or 25 µg/mL kanamycin LB plate. In this manner the *E. coli P_{T5}-ispAdxs* strain was constructed

In order to further stack *kan-P_{T5}-idi* on chromosome of *E. coli P_{T5}-ispAdxs*, P1 lysate made on *E. coli kan-P_{T5}-idi* strain was transduced into the recipient strain, *E. coli P_{T5}-ispAdxs*, as described above.

Approximately 85 transductants were selected. After transduction, the kanamycin selectable marker was eliminated from the chromosome as described above, yielding *E. coli P_{T5}-ispAdxs P_{T5}-idi* strain.

For the *E. coli P_{T5}-ispAdxs P_{T5}-idi* strain, the correct integration of the *P_{T5}* promoter in the front of *ispAdxs* and *idi* genes, and elimination of the kanamycin selectable marker from the *E. coli* chromosome were confirmed by PCR analysis. A colony of the *E. coli P_{T5}-ispAdxs P_{T5}-idi* strain was resuspended in 50 µL of PCR reaction mixture containing 200 µM dNTPs, 2.5 U AmpliTaq™ (Applied Biosystems), and 0.4 µM of different combination of specific primer pairs, T-kan (5'-ACCGGATATCACCCTTAT CTGCTC-3'; SEQ ID NO:46) and B-ispA (5'-CCTAATAATGCGCCATACTGCATGG-3'; SEQ ID NO:47), T-T5 (5'-TAACCTATAAAAATAGGCGTATCACGAGGCCC-3'; SEQ ID NO:48) and B-ispA, T-kan and B-idi (5'-CAGCCAACTGGAGAACGCGAGATGT-3'; SEQ ID NO:49), T-T5 and B-idi. Test primers were chosen to amplify regions located either in the kanamycin marker or the *P_{T5}* promoter and the early region of *ispAdxs* or *idi* gene (Figure 3). The PCR reaction was performed as described in Example 1. The PCR results indicated the elimination of the kanamycin selectable marker from the *E. coli* chromosome (Figure 3, lane 2 and 4). The chromosomal integration of the *P_{T5}* promoter fragment upstream of the *ispAdxs* and *idi* gene was confirmed based on the expected sizes of PCR products, 285 bp and 274 bp, respectively (Figure 3, lane 1 and 3).

EXAMPLE 7Construction of *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) Strain for Increased β -Carotene Production

In order to construct the *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) strain
 5 containing a chromosomally-integrated P_{T5} promoter upstream from
ispAdxs genes and *Methylobacter* 16a *dxs* (*dxs*(16a)), P1 lysate made on
E. coli *kan*- P_{T5} -*dxs*(16a) strain was transduced into the recipient strain,
E. coli *kan*- P_{T5} -*ispAdxs* containing a β -carotene biosynthesis expression
 plasmid pPCB15 (*cam*^R), described in Example 3. Seventy-eight
 10 kanamycin-resistance transductants were selected. The kanamycin
 selectable marker was eliminated from the chromosome of the
 transductants using a FLP recombinase expression system as described
 in Example 3, yielding the *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) strain.

In the *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) strain the correct
 15 integration of the phage *T5* promoter in the front of *ispAdxs* genes and
 P_{T5} -*dxs*(16a) at inter-operon region located at 30.9 min on the *E. coli*
 chromosome, and elimination of the kanamycin selectable marker were
 confirmed by PCR analysis. A colony of the *E. coli* P_{T5} -*ispAdxs* P_{T5} -
dxs(16a) strain was tested by PCR with different combination of specific
 20 primer pairs, T-kan and B-*ispA*, T-*T5* and B-*ispA*, T-kan and B-*dxs*(16a)
 (5'-GCGATATTGTATGTCTGATTTCAGGA-3'; SEQ ID NO:50), T-*T5* and B-
dxs(16a). Test primers were chosen to amplify regions located either in
 the kanamycin resistance gene or the P_{T5} promoter and the downstream
 region of the chromosomal integration site (Figure 3). The PCR reaction
 25 was performed as described in Example 1. The PCR results indicated the
 elimination of the kanamycin selectable marker from the *E. coli*
 chromosome (Figure 3, lane 6 and 8). The chromosomal integration of
 the P_{T5} promoter fragment upstream of the *ispAdxs* gene and the
 integration of the P_{T5} -*dxs*(16a) gene at the inter-operon region was
 30 confirmed based on the expected sizes of PCR products, 285 bp and
 2184 bp, respectively (Figure 3, lane 5 and 7).

EXAMPLE 8Construction of *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a) Strain for Increased β -Carotene Production

35 In order to create a bacterial strain capable of increased carotenoid
 production, the *Methylobacter* 16a *lytB* (*lytB*(16a)) gene under the control
 of a P_{T5} promoter was further stacked into the *E. coli* P_{T5} -*ispAdxs* P_{T5} -
dxs(16a) strain by P1 transduction in combination with the FLP

recombination system. P1 lysate made on *E. coli kan-P_{T5}-lytB(16a)* strain was transduced into the recipient strain, *E. coli kan-P_{T5}-ispAdxs kan-P_{T5}-dxs(16a)* containing the β -carotene biosynthesis expression plasmid pPCB15 (cam^R). Forty-two kanamycin-resistance transductants were selected. The kanamycin selectable marker was eliminated from the chromosome of the transductants using a FLP recombinase expression system as described in Example 6, yielding *E. coli P_{T5}-ispAdxs P_{T5}-dxs(16a) P_{T5}-lytB(16a)*.

For the *E. coli P_{T5}-ispAdxs P_{T5}-dxs(16a) P_{T5}-lytB(16a)* strain, the correct integration of the *P_{T5}* promoter upstream of *ispAdxs* genes and the addition of the *P_{T5}-dxs(16a)* and *P_{T5}-lytB(16a)* genes at inter-operon region located at 30.9 min and 18.1 min, respectively, on the *E. coli* chromosome, and elimination of the kanamycin selectable marker were confirmed by PCR analysis. A colony of the *E. coli P_{T5}-ispAdxs P_{T5}-dxs(16a) P_{T5}-lytB(16a)* strain was tested by PCR with different combination of specific primer pairs, T-kan and B-ispA, T-T5 and B-ispA, T-kan and B-dxs(16a), T-T5 and B-dxs(16a), T-kan and B-lytB(16a) (5'-TCCACTGGATGCGGGAAGCTGGCAG-3'; SEQ ID NO:51), T-T5 and B-lytB(16a). Test primers were chosen to amplify regions located either in the kanamycin resistance gene or the *P_{T5}* promoter and the downstream region of the chromosomal integration site (Figure 3). The PCR reaction was performed as described in Example 1. The PCR results indicated the elimination of the kanamycin selectable marker from the *E. coli* chromosome (Figure 3, lane 10, 12 and 14). The chromosomal integration of the *P_{T5}* promoter fragment upstream of the *ispAdxs* gene and integration of the *P_{T5}-dxs(16a)* and *P_{T5}-lytB(16a)* genes at the inter-operon region was confirmed based on the expected sizes of PCR products, 285 bp, 2184 bp, and 1282 bp, respectively (Figure 3, lane 9, 11 and 13).

EXAMPLE 9

Construction of *E. coli P_{T5}-ispAdxs P_{T5}-dxs(16a) P_{T5}-lytB(16a) P_{T5}-idi* Strain for Increased β -Carotene Production

In order to create a bacterial strain capable of increased carotenoid production, the *P_{T5}-idi* gene was further stacked into the *E. coli P_{T5}-ispAdxs P_{T5}-dxs(16a) P_{T5}-lytB(16a)* strain by P1 transduction in combination with the FLP recombination system. P1 lysate made from *E. coli kan-P_{T5}-idi* strain was transduced into the recipient strain, *E. coli kan-P_{T5}-ispAdxs kan-P_{T5}-dxs(16a) P_{T5}-lytB(16a)* containing the β -

carotene biosynthesis expression plasmid pPCB15. Approximately 450 kanamycin-resistance transductants were selected. The kanamycin selectable marker was eliminated from the chromosome of the transductants using a FLP recombinase expression system as described in Example 6, yielding *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a) P_{T5} -*idi*.

For the *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a) P_{T5} -*idi* strain, the correct integration of the P_{T5} promoter upstream of *ispAdxs* and *idi* genes and the integration of the P_{T5} -*dxs*(16a) and P_{T5} -*lytB*(16a) genes at inter-operon region located at 30.9 min and 18.1 min, respectively, on the *E. coli* chromosome, and elimination of the kanamycin selectable marker were confirmed by PCR analysis. A colony of the *E. coli* P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a) P_{T5} -*idi* strain was tested by PCR with different combination of specific primer pairs, T-kan and B-*ispA*, T-T5 and B-*ispA*, T-kan and B-*dxs*(16a), T-T5 and B-*dxs*(16a), T-kan and B-*lytB*(16a), T-T5 and B-*lytB*(16a), T-kan and B-*idi*, T-T5 and B-*idi*. Test primers were chosen to amplify regions located either in the kanamycin resistance gene or the P_{T5} promoter and the downstream region of the chromosomal integration site (Figure 3). The PCR reaction was performed as described in Example 1. The PCR results indicated the elimination of the kanamycin selectable marker from the *E. coli* chromosome (Figure 4, lane 16, 18, 20, and 22). The chromosomal integration of the P_{T5} promoter fragment upstream of the *ispAdxs* and *idi* genes and the integration of the P_{T5} -*dxs*(16a) and P_{T5} -*lytB*(16a) constructs at the inter-operon region was confirmed based on the expected sizes of PCR products, 285 bp, 274 bp, 2184 bp, and 1282 bp, respectively (Figure 4, lane 15, 17, 19 and 21).

EXAMPLE 10

Construction of *E. coli* P_{T5} -*dxs* P_{T5} -*idi* Strain for Increased β -Carotene Production

In order to characterize the effect of the chromosomal integration of P_{T5} strong promoter in the front of the *dxs* and *idi* genes on β -carotene production, *E. coli* P_{T5} -*dxs* P_{T5} -*idi*, capable of producing β -carotene, was constructed.

P1 lysate made with the *E. coli* *kan*- P_{T5} -*dxs* strain was transduced into the recipient strain, *E. coli* MG1655 containing a β -carotene biosynthesis expression plasmid pPCB15 (*cam*^R) as described in Example 6. Sixteen kanamycin-resistance transductants were selected. The

kanamycin selectable marker was eliminated from the chromosome of the transductants using a FLP recombinase expression system, yielding *E. coli P_{T5}-dxs* strain.

In order to stack *kan-P_{T5}-idi* on chromosome of *E. coli P_{T5}-dxs*, P1 lysate made on *E. coli kan-P_{T5}-idi* strain was transduced into the recipient strain, *E. coli P_{T5}-dxs*, as described above. Approximately 450 kanamycin-resistance transductants were selected. After transduction, the kanamycin selectable marker was eliminated from the chromosome as described above, yielding *E. coli P_{T5}-dxs P_{T5}-idi* strain.

For the *E. coli P_{T5}-dxs P_{T5}-idi* strain, the correct integration of the phage *P_{T5}* promoter upstream of *dxs* and *idi* genes on the *E. coli* chromosome, and elimination of the kanamycin selectable marker were confirmed by PCR analysis. A colony of the *E. coli P_{T5}-dxs P_{T5}-idi* strain was tested by PCR with different combination of specific primer pairs, T-kan and B-dxs (5'-TGGCAACA GTCGTAGCTCCTGGGTGG-3'; SEQ ID NO:52), T-T5 and B-dxs, T-kan and B-idi, T-T5 and B-idi. Test primers were chosen to amplify regions located either in the kanamycin or the *P_{T5}* promoter and the downstream region of the chromosomal integration site (Figure 3). The PCR reaction was performed as described in Example 1. The PCR results indicated the elimination of the kanamycin selectable marker from the *E. coli* chromosome (Figure 4, lane 24 and 26). The chromosomal integration of the *P_{T5}* promoter fragment upstream of the *dxs* and *idi* gene was confirmed based on the expected sizes of PCR products, 229 bp and 274 bp, respectively (Figure 4, lane 23 and 25).

EXAMPLE 11

Construction of *E. coli P_{T5}-dxs P_{T5}-idi P_{T5}-ygbBP* Strain for Increased β -Carotene Production

In order to create a bacterial strain capable of increased carotenoid production, *P_{T5}-ygbBP* gene was further stacked into the *E. coli P_{T5}-dxs P_{T5}-idi* strain by P1 transduction in combination with the FLP recombination system. P1 lysate was with *E. coli kan-P_{T5}-ygbBP* strain was transduced into the recipient strain, *E. coli kan-P_{T5}-dxs kan-P_{T5}-idi* containing the β -carotene biosynthesis expression plasmid pPCB15 (*cam*^R), as described above. Twenty-one kanamycin-resistance transductants were selected. The kanamycin selectable marker was eliminated from the chromosome of the transductants using a FLP recombinase expression system, yielding *E. coli P_{T5}-dxs P_{T5}-idi P_{T5}-ygbBP* strain.

For the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* strain, the correct integration of the P_{T5} promoter upstream of *dxs*, *idi* and *ygbBP* genes on the *E. coli* chromosome, and elimination of the kanamycin selectable marker were confirmed by PCR analysis. A colony of the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* strain was tested by PCR with different combination of specific primer pairs, T-kan and B-*dxs*, T- P_{T5} and B-*dxs*, T-kan and B-*idi*, T- P_{T5} and B-*idi*, T-kan and B-*ygb* (5'-CCAGCAGCGCATGCACCGAGTGTTC-3')(SEQ ID NO:53), T- P_{T5} and B-*ygb*. Test primers were chosen to amplify regions located either in the kanamycin resistance marker or the P_{T5} promoter and the downstream region of the chromosomal integration site (Figure 3). The PCR reaction was performed as described in Example 1. The PCR results indicated the elimination of the kanamycin selectable marker from the *E. coli* chromosome (Figure 4, lane 28, 30 and 32). The chromosomal integration of the P_{T5} promoter fragment upstream of the *dxs*, *idi* and *ygbBP* gene was confirmed based on the expected sizes of PCR products, 229 bp, 274 bp, and 296 bp, respectively (Figure 4, lane 27, 29, and 31).

EXAMPLE 12

20 Construction of *E. coli* P_{T5} -DXS P_{T5} -IDI P_{T5} -ygbBP P_{T5} -lytB(16a) Strain for Increased β -carotene Production

In order to create a bacterial strain capable of increased carotenoid production, the *Methylobacter* 16a *lytB* (*lytB*(16a)) gene under the control of a P_{T5} promoter was further stacked into the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* strain by P1 transduction in combination with the FLP recombination system. P1 lysate made with *E. coli* *kan*- P_{T5} -*lytB*(16a) strain was transduced into the recipient strain, *E. coli* *kan*- P_{T5} -*dxs* *kan*- P_{T5} -*idi* P_{T5} -*ygbBP* containing the β -carotene biosynthesis expression plasmid pPCB15 (*cam*^R), described previously. Approximately 300 kanamycin-resistance transductants were selected. The kanamycin selectable marker was eliminated from the chromosome of the transductants using a FLP recombinase expression system, yielding *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*lytB*(16a) strain.

For the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*lytB*(16a) strain, the correct integration of the P_{T5} promoter upstream of *dxs*, *idi* and *ygbBP* genes and integration of the P_{T5} -*lytB*(16a) gene at inter-operon region located at 18.1 min on the *E. coli* chromosome, and elimination of the kanamycin selectable marker were confirmed by PCR analysis. A colony

of the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*lytB*(16a) strain was tested by PCR with different combination of specific primer pairs, T-kan and B-*dxs*, T-T5 and B-*dxs*, T-kan and B-*idi*, T-T5 and B-*idi*, T-kan and B-*ygb*, T-T5 and B-*ygb*, T-kan and B-*lytB*(16a), T-T5 and B-*lytB*(16a). Test primers
5 were chosen to amplify regions located either in the kanamycin resistance marker or the phage P_{T5} promoter and the downstream region of the chromosomal integration site (Figure 3). The PCR reaction was performed as described in Example 1. The PCR results indicated the elimination of the kanamycin selectable marker from the *E. coli*
10 chromosome (Figure 4, lane 34, 36, 38 and 40). The chromosomal integration of the P_{T5} promoter fragment upstream of the *dxs*, *idi* and *ygbBP* gene and the integration of P_{T5} -*lytB*(16a) gene was confirmed based on the expected sizes of PCR products, 229 bp, 274 bp, 296 bp, and 1282 bp, respectively (Figure 4, lane 33, 35, 37, and 39).

15 EXAMPLE 13

Isolation of Chromosomal Mutations that Increase Carotenoid Production

Wild type *E. coli* is non-carotenogenic and synthesizes only the farnesyl pyrophosphate precursor for carotenoids. When the *crtEXYIB* gene cluster from *Pantoea stewartii* was introduced into *E. coli*, β -carotene
20 was synthesized and the cells exhibit a yellow color characteristic of β -carotene. *E. coli* chromosomal mutations which increase carotenoid production should result in colonies that have are more intensely pigmented or deeper yellow in color (Figure 8).

The plasmid pPCB15 (cam^R) encodes the carotenoid biosynthesis
25 gene cluster (*crtEXYIB*) from *Pantoea Stewartii* (ATCC no. 8199). The pPCB15 plasmid was constructed from ligation of *Sma*I digested pSU18 (Bartolomeet al., *Gene*, 102:75-78 (1991)) vector with a blunt-ended *Pme*I/*Not*I fragment carrying *crtEXYIB* from pPCB13 (Example 3). *E. coli* MG1655 transformed with pPCB15 was used for transposon mutagenesis.
30 Mutagenesis was performed using EZ:TNTM <KAN-2>Tnp TransposomeTM kit (Epicentre Technologies, Madison, WI) according to manufacture's instructions. The transposon (1 μ L) was electroporated into 50 μ L of highly electro-competent MG1655 (pPCB15) cells. The mutant cells were spread onto LB-Noble Agar (Difco laboratories, Detroit, MI)
35 plates with 25 μ g/mL kanamycin and 25 μ g/mL chloramphenicol, and grown at 37°C overnight. Tens of thousands of mutant colonies were visually examined for production of increased levels of β -carotene as evaluated by deeper yellow color development. The candidate mutants

were re-streaked to fresh LB-Noble Agar plates and glycerol frozen stocks made for further characterization.

EXAMPLE 14

Quantitation of Carotenoid Production

5 To confirm that the mutants selected for increased production β -carotene by visually screening for deeper yellow colonies in Example 13 indeed produced more β -carotene, the carotenoids were extracted from cultures grown from each mutant strain and quantified spectrophotometrically. Each candidate mutant strain was cultured in 10 mL LB medium with 25 μ g/mL chloramphenicol in 50 mL flasks overnight shaking at 250 rpm. MG1655 (pPCB15) was used as the control. Carotenoids were extracted from each cell pellet for 15 min into 1 mL acetone, and the amount of β -carotene produced was measured at 455 nm. Cell density was measured at 600 nm. The ratio OD455/OD600 15 was used to normalize β -carotene production for different cultures. β -carotene production was also verified by HPLC. Among the mutant clones tested, eight showed increased β -carotene production (Figure 9). Mutant Y15 showed almost two-fold increase in β -carotene production as shown in Figure 8 which represents the averages of three independent 20 measurements with standard deviations calculated and indicated as standard deviation bars.

EXAMPLE 15

Mapping of the Transposon Insertions on the *E. coli* Chromosome

The transposon insertion site in each mutant was identified by PCR 25 and sequencing directly from chromosomal DNA of the mutant strains. A modified single-primer PCR method (Karlyshev et al., *BioTechniques*, 28:1078-82, 2000) was used. For this method, a 100 μ L volume of overnight culture was heated at 99°C for 10 min in a PCR machine. Cell debris was removed by centrifugation at 4000 g for 10 min. A 1 μ L 30 volume of supernatant was used in a 50 μ L PCR reaction using either Tn5PCRF (5'-GCTGAGTTGAAGGATCAGATC-3';SEQ ID NO:54) or Tn5PCRR (5'-CGAGCAAGACGTTTCCCGTTG-3';SEQ ID NO:55) primer. PCR was carried out as follows: 5 min at 95°C; 20 cycles of 92°C for 30 sec, 60°C for 30 sec, 72°C for 3 min; 30 cycles of 92°C for 30 sec, 40°C 35 for 30 sec, 72°C for 2 min; 30 cycles of 92°C for 30 sec, 60°C for 30 sec, 72°C for 2 min. A 10- μ L volume of each PCR product was electrophoresed on an agarose gel to evaluate product length. A 40 μ L volume of each PCR product was purified using the Qiagen PCR cleanup

kit, and sequenced using sequencing primers Kan-2 FP-1 (5'-ACCTACAACAAAGCTCTCATCAACC-3';SEQ ID NO:56) or Kan-2 RP-1 (5'-GCAATGTAACATCAGAGATTTTGAG-3';SEQ ID NO:57) provided by the EZ:TN™ <KAN-2>Tnp Transposome™ kit. The chromosomal
5 insertion site of the transposon was identified as the junction between the Tn5 transposon and MG1655 chromosome DNA by aligning the sequence obtained from each mutant with the *E. coli* MG1655 genomic sequence. Mutant Y15 carried a Tn5 insertion in *yjeR* (Ghosh, S., *PNAS*, 96:4372-4377 (1999)). The Tn5 cassette was located very close to the carboxy
10 terminal end of the gene (Figure 10) and most likely resulted in functional although truncated protein product.

EXAMPLE 16

Confirmation of transposon insertions in *E. coli* chromosome

To confirm the transposon insertion sites in Example 15,
15 chromosome specific primers were designed 400-800bp upstream and downstream from the transposon insertion site for each mutant. Primers Y15_F (5'-GGATCGATCTTGAGATGACC-3';SEQ ID NO:58) and Y15_R (5'-GCTTTCGTAATTTTCGCATTCTG-3';SEQ ID NO:59) were used to screen the Y15 mutant. Three sets of PCR reactions were performed for
20 each mutant. The first set (named as PCR 1) uses a chromosome specific upstream primer with a chromosome specific downstream primer. The second set (PCR 2) uses a chromosome specific upstream primer with a transposon specific primer (either Kan-2 FP-1 or Kan-2 RP-1, depending on the orientation of the transposon in the chromosome). The
25 third set (PCR 3) uses a chromosome specific downstream primer with a transposon specific primer. PCR conditions are: 5 min at 95°C; 30 cycles of 92°C for 30 sec, 55°C for 30 sec, 72°C for 1 min; then 5 min at 72°C. Wild type MG1655 (pPCB15) cells served as control cells. For the control cells, the expected wild type bands were detected in PCR1, and no mutant
30 band was detected in PCR2 or PCR3. For all the eight mutants, no wild type bands were detected in PCR1, and the expected mutant bands were detected in both PCR2 and PCR3. The size of the products in PCR2 and PCR3 correlated well with the insertion sites in each specific gene. Therefore, the mutants contained the transposon insertions as indicated in
35 Example 15.

EXAMPLE 17Construction of *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP yjeR::Tn5* Strain for Increased β -Carotene Production

In order to create a bacterial strain capable of increased carotenoid
 5 production, a gene, *yjeR::Tn5* (SEQ ID NO:63) partially knocked-out by
 transposon (Tn5) (*kan^R*) as discovered by experiments outlined in
 Examples 13-16, was further stacked into the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -
ygbBP strain by P1 transduction. The *yjeR* gene encoding
 oligoribonuclease that has a 3'-to-5' exoribonuclease activity for small
 10 oligoribonucleotides has been isolated by random transposon (Tn5)-
 insertional mutagenesis for increasing β -carotene production. P1 lysate
 made on *E. coli yjeR::Tn5* strain was transduced into the recipient strain,
E. coli P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* containing the β -carotene biosynthesis
 expression plasmid pPCB15 (*cam^R*), described previously. Six
 15 kanamycin-resistance transductants were selected.

For the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP yjeR::Tn5* strain, the
 correct integration of the P_{T5} promoter upstream of *dxs*, *idi* and *ygbBP*
 genes and integration of the *yjeR::Tn5* gene on the *E. coli* chromosome
 was confirmed by PCR fragment analysis. A colony of the *E. coli* P_{T5} -*dxs*
 20 P_{T5} -*idi* P_{T5} -*ygbBP yjeR::Tn5* strain was tested by PCR with different
 combination of specific primer pairs, T-kan and B-*dxs*, T-T5 and B-*dxs*, T-
 kan and B-*idi*, T-T5 and B-*idi*, T-kan and B-*ygb*, T-T5 and B-*ygb*, T-
 Tn5*yjeR* (5'-GCAATGTAACATCAGAGATTTTGAG-3'; SEQ ID NO:60)
 and B-*yjeR* (5'-GCTTTCGTAATTTTCGCATTCTG-3'; SEQ ID NO:61).
 25 Test primers were chosen to amplify regions located either in the
 kanamycin selection marker or the P_{T5} promoter and the downstream
 region of the chromosomal integration site (Figure 3). The PCR reaction
 was performed as described in Example 1. The PCR results indicated the
 elimination of the kanamycin selectable marker from the *E. coli*
 30 chromosome (Figure 4, lane 42, 44, and 46). The chromosomal
 integration of the P_{T5} promoter fragment upstream of the *dxs*, *idi* and
ygbBP genes and the integration of the transposon (Tn5) into *yjeR* gene
 (*yjeR::Tn5*) was confirmed based on the expected sizes of PCR products,
 229 bp, 274 bp, 296 bp, and 285 bp, respectively (Figure 4, lane 41, 43,
 35 45, and 47).

EXAMPLE 18Construction of *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* Strain for Increased β -Carotene Production

The *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* strain was
 5 constructed by P1 transduction in the combination of the Flp site-specific
 recombinase for marker removal. P1 lysate made from *E. coli* *kan*- P_{T5} -
ispB strain was transduced into the recipient strain, *E. coli* P_{T5} -*dxs* P_{T5} -*idi*
 P_{T5} -*ygbBP* containing the β -carotene biosynthesis expression plasmid
 pPCB15 (*cam*^R). Thirty-six kanamycin-resistance transductants were
 10 selected. A kanamycin selectable marker was eliminated from the
 chromosome as described at Example 6, yielding *E. coli* P_{T5} -*dxs* P_{T5} -*idi*
 P_{T5} -*ygbBP* P_{T5} -*ispB*.

The stacking of *ispB* gene under the control of the P_{T5} strong
 promoter resulted in unexpected increase of β -carotene production. This
 15 was a non-obvious result because *IspB* (octaprenyl diphosphate
 synthase), which supplies the precursor of the side chain of the isoprenoid
 quinones, drains away the FPP precursor from the carotenoid biosynthetic
 pathway (Figure 1). The mechanism of how overexpression of *ispB* gene
 under the control of P_{T5} promoter increases the β -carotene production is
 20 not clear yet. However, the result suggests that *IspB* may increase the
 flux of the carotenoid biosynthetic pathway.

For the *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* strain the
 correct integration of the phage P_{T5} promoter in the front of *dxs*, *idi*,
ygbBP, and *ispB* genes, and elimination of the kanamycin selectable
 25 marker were confirmed by PCR analysis. A colony of the *E. coli* P_{T5} -*dxs*
 P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* was tested by PCR with different combination
 of specific primer pairs, T-T5 and B-*dxs*, T-*kan* and B-*dxs*, T-T5 and B-*idi*,
 T-*kan* and B-*idi*, T-T5 and B-*ygb*, T-*kan* and B-*ygb*, T-T5 and B-*ispB* (5'-
 AGTACAGCAATCATCGGACGAATACG-3'; SEQ ID NO:62), and T-*kan*
 30 and B-*ispB*. Test primers were chosen to amplify regions located either in
 the kanamycin selectable marker or the P_{T5} promoter and the
 downstream region of the chromosomal integration site (Figure 3). The
 PCR reaction was performed as described in Example 1. The PCR
 results indicated the elimination of the kanamycin selectable marker from
 35 the *E. coli* chromosome (Figure 5, lane 49, 51, 53, and 55). The
 chromosomal integration of the P_{T5} promoter upstream of the *dxs*, *idi*,
ygbBP and *ispB* genes was confirmed based on the expected sizes of

PCR products, 229 bp, 274 bp, 296 bp, and 318 bp, respectively (Figure 5, lane 48, 50, 52, and 54).

EXAMPLE 19

Transformation of pDCQ108 into *E. coli* P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB Strain

The low copy number plasmid pPCB15 (containing the β -carotene synthesis genes *Pantoea crtEXYIB*) used as a reporter plasmid for monitoring β -carotene production in *E. coli* P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB was replaced with the medium copy number plasmid pDCQ108 (ATCC PTA-4823) containing β -carotene synthesis genes *Pantoea crtEXYIB*. The plasmid pPCB15 was eliminated from the *E. coli* P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB strain by streaking on LB plate, incubating at 37 °C for 2 d, and picking up a white-colored colony.

The plasmid pDCQ108 (tet^R) was transformed into *E. coli* P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB strain (white colony lacking a carotenoid reporter plasmid). Electro-transformation was performed as described in Example 1. Transformants were selected on 25 μ g/mL of tetracycline LB plates at 37°C. The resultant transformants were the *E. coli* P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB strain carrying pDCQ108.

EXAMPLE 20

Measurement of β -Carotene Production in *E. coli* Strains with Chromosomal Integrations

β -carotene production of the 9 chromosomally engineered *E. coli* strains, *E. coli* pPCB15 P_{T5} -ispAdxs P_{T5} -idi, *E. coli* pPCB15 P_{T5} -ispAdxs P_{T5} -dxs(16a), *E. coli* pPCB15 P_{T5} -ispAdxs P_{T5} -dxs(16a) P_{T5} -lytB(16a), *E. coli* pPCB15 P_{T5} -ispAdxs P_{T5} -dxs(16a) P_{T5} -lytB(16a) P_{T5} -idi, *E. coli* pPCB15 P_{T5} -dxs P_{T5} -idi, *E. coli* pPCB15 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP, *E. coli* pPCB15 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -lytB(16a), *E. coli* pPCB15 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP *yjeR::Tn5*, and *E. coli* pDCQ108 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB was quantified by the following spectrophotometric method. The quantitative analysis of β -carotene production was achieved by measuring the spectra of β -carotene's characteristic $\lambda_{\max}$ peaks at 425, 450 and 478 nm. The 8 chromosomally-engineered *E. coli* control strains were grown in 5 mL LB containing 25 μ g/mL of chloramphenicol at 37°C for 24 h, and then harvested by centrifugation at 4000 rpm for 10 min. The β -carotene pigment was extracted by resuspending cell pellet in 1 mL of acetone with vortexing for 1 min and then rocking the sample for 1 h at room temperature. Following

centrifugation at 4000 rpm for 10 min, the absorption spectrum of the acetone layer containing β -carotene was measured at 450 nm using an Ultrospec 3000 spectrophotometer (Amersham Biosciences, Piscataway, NJ). The production of β -carotene in *E. coli* pPCB15 P_{T5} -*ispAdxs* P_{T5} -*idi* and *E. coli* pPCB15 P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) was approximately 3.5-fold and 4.3-fold higher than that of the control strain, *E. coli* pPCB15, respectively (Figure 11). Additional stacking of P_{T5} -*lytB*(16a) and P_{T5} -*idi* in *E. coli* pPCB15 P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a) and *E. coli* pPCB15 P_{T5} -*ispAdxs* P_{T5} -*dxs*(16a) P_{T5} -*lytB*(16a) P_{T5} -*idi* didn't increase the production of β -carotene significantly. The production of β -carotene in *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* was approximately 4.4-fold higher than that of the *E. coli* pPCB15 control strain. Additional stacking of P_{T5} -*ygbBP* and P_{T5} -*lytB*(16a) in *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* and *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* increased production of β -carotene 41 % and 45 %, respectively compared to that of *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* (Figure 11). The production of β -carotene in the *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* *yjeR::Tn5*, was approximately 19-fold higher than that of the *E. coli* pPCB15 control strain. The *E. coli* pDCQ108 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* strain showed the best titer of β -carotene production, approximately 30-fold higher than the *E. coli* pPCB15 control strain.

EXAMPLE 21

Determination of β -Carotene Content in *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* *yjeR::Tn5* and *E. coli* P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB*

Example 20 demonstrated that the *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* *yjeR::Tn5* (ATCC PTA-4807) and *E. coli* pDCQ108 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* (ATCC PTA-4823) strains in this invention produces high levels of β -carotene, showing deep orange colored colony on LB plate. The content of β -carotene in the *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* *yjeR::Tn5* and *E. coli* pDCQ108 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* strains also was quantified by HPLC analysis. The *E. coli* pPCB15 control, *E. coli* pPCB15 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* *yjeR::Tn5* and *E. coli* pDCQ108 P_{T5} -*dxs* P_{T5} -*idi* P_{T5} -*ygbBP* P_{T5} -*ispB* strains were grown in 50 mL LB containing 25 μ g/mL of chloramphenicol at 37°C for 24 h with 250 rpm agitation. Twenty mL of the culture cells was filtered on 37 mm diameter cellulose filter (0.2 μ m) (Millipore, Bedford, MA) that was pre-weighted after drying at 95 °C oven for 24 h. After washing with 10 mL of sterile water, the cells on the pre-weighted filter were completely

dried at 95 °C oven for 24 h until its weight did not change. The dry cell weight was determined by subtracting the weight of filter itself from the total weight.

Twenty mL of the culture cells was harvested by centrifugation at 4000 rpm for 10 min for carotenoid extraction and analysis. The β -carotene pigment was extracted as described in Example 20. The carotene extract obtained was analyzed for the β -carotene content by a high performance liquid chromatography (HPLC). A 125 x 4 mm RP8 (5 μ m particles) column (Hewlett-Packard, San Fernando, CA) was used for HPLC analysis of β -carotene. The flow rate was 1 mL/min and the solvent program was as follows: 0 - 11.5 min linear gradient from 40% water/60% methanol to 100% methanol, 11.5 - 20 min 100% methanol, 20-30 min 40% water/60% methanol. Detection of β -carotene was measured by absorption at 450 nm and quantitative analysis was carried out by comparing an area of the peak of β -carotene to a known β -carotene standard (Sigma, Saint Louis, MO).

E. coli pPCB15 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP *yjeR*::Tn5 and *E. coli* pDCQ108 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB strains produced 3.8 mg of β -carotene per gram of dry cell weight (3,800 ppm) and 6.0 mg of β -carotene /g of dry cell weight (6,000 ppm) β -carotene, respectively, while *E. coli* pPCB15 control strain produces 0.2 mg of β -carotene/g of dry cell weight (200 ppm) (Table 10). The HPLC analysis for the β -carotene content also showed that the chromosomally engineered *E. coli* pPCB15 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP *yjeR*::Tn5 and *E. coli* pDCQ108 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB strains produced β -carotene 19-fold and 30-fold higher than the control strain, respectively.

It has been speculated that the limits for carotenoid production in non-carotenogenic host such as *E. coli* had been reached at the level of around 1.5 mg/g cell dry weight (1,500 ppm) due to overload of the membranes and blocking of membrane functionality (Albrecht et al., *supra*). The present method has solved the stated problem by making modifications to the *E. coli* chromosome allowing β -carotene production of 6 mg per g dry weight (6,000 ppm), an increase of 30-fold over initial levels in *E. coli* pDCQ108 P_{T5} -dxs P_{T5} -idi P_{T5} -ygbBP P_{T5} -ispB.

35

TABLE 10
 β -carotene Production

Strain	β -Carotene (mg/g dcw ¹)
<i>E. coli</i> MG1655 pPCB15 ²	0.2
<i>E. coli</i> MG1655 pPCB15 ² <i>P_{T5}-dxs P_{T5}-idi P_{T5}-ygbBP yjeR::Tn5</i>	3.8
<i>E. coli</i> MG1655 pDCQ108 ³ <i>P_{T5}-dxs P_{T5}-idi P_{T5}-ygbBP P_{T5}-ispB</i>	6.0

¹ Dry Cell Weight

5 ² pPCB15 contains the carotenoid biosynthesis gene cluster (*crtEXYIB*) from *Pantoea Stewartii* (ATCC no. 8199).

³ pDCQ108 contains the carotenoid biosynthesis gene cluster (*crtEXYIB*) from *Pantoea Stewartii* (ATCC no. 8199).

CLAIMS

What is claimed is:

1. A carotenoid overproducing bacteria comprising the genes encoding a functional carotenoid enzymatic biosynthetic pathway wherein the *dxs*, *idi* and *ygbBP* genes are overexpressed and wherein the *yjeR* gene is down regulated.
2. A carotenoid overproducing bacteria comprising the genes encoding a functional carotenoid enzymatic biosynthetic pathway wherein the *dxs*, *idi*, *ygbBP* and *ispB* genes are overexpressed.
3. The carotenoid overproducing bacteria of Claim 1 or 2 wherein the *lytB* and *dxr* gene is optionally overexpressed.
ispB *lytB* and *dxr* *yjeR*
4. The carotenoid overproducing bacteria of Claim 1 or 2 wherein the carotenoid enzymatic biosynthetic pathway consists of the genes *dxs*, *dxr*, *ygpP*, *ychB*, *ygbB*, *lytB*, *idi*, *ispA*, *ispB* *crtE*, *crtB*, *crtI*, and *crtY*.
5. The carotenoid overproducing bacteria of Claim 4 wherein the carotenoid enzymatic biosynthetic pathway optionally additionally comprises the *crtZ* and *crtW* genes.
6. The carotenoid overproducing bacteria of any of Claims 1-5 wherein the bacteria is selected from the group consisting *Agrobacterium*, *Erythrobacter*, *Chlorobium*, *Chromatium*, *Flavobacterium*, *Cytophaga*, *Rhodobacter*, *Rhodococcus*, *Streptomyces*, *Brevibacterium*, *Corynebacteria*, *Mycobacterium*, *Deinococcus*, *Paracoccus*, *Escherichia*, *Bacillus*, *Myxococcus*, *Salmonella*, *Yersinia*, *Erwinia*, *Pantoea*, *Pseudomonas*, *Sphingomonas*, *Methylomonas*, *Methylobacter*, *Methylococcus*, *Methylosinus*, *Methylomicrobium*, *Methylocystis*, *Alcaligenes*, *Synechocystis*, *Synechococcus*, *Anabaena*, *Thiobacillus*, *Methanobacterium*, *Klebsiella*, and *Myxococcus*.
7. The carotenoid overproducing bacteria of Claim 6 wherein the bacteria is *E. coli*.
8. The carotenoid overproducing bacteria of Claims 1-3 wherein the *dxs*, *dxr*, *ygpP*, *ychB*, *ygbB*, *lytB*, *idi*, *ispA*, *ispB* are derived from a *Methylomonas* sp..
9. The carotenoid overproducing bacteria of any of Claims 1 – 3 wherein the *dxs*, *idi*, *ispB* and *ygbBP* genes are under the control of a strong promoter.

10. The carotenoid overproducing bacteria of Claim 9 wherein the strong promoter is selected from the group consisting of *lac*, *ara*, *tet*, *trp*, λP_L , λP_R , *T7*, *tac*, P_{T5} , and *trc*.
11. The carotenoid overproducing bacteria of any of Claims 1-3 wherein the *dxs*, *idi*, *ispB* and *ygbBP* genes are integrated in multicopy in the bacterial chromosome.
12. The carotenoid overproducing bacteria of any of Claims 1-3 wherein the *dxs*, *idi*, *ispB* and *ygbBP* genes are present in multicopy in the bacteria on one or more plasmids.
13. The carotenoid overproducing bacteria of of Claim 7 wherein the *yjeR* gene is down regulated by gene disruption.
14. The carotenoid overproducing bacteria of Claim 13 wherein the disrupted *yjeR* gene has the nucleotide sequence as set forth in SEQ ID NO:63.
15. The carotenoid overproducing bacteria of either of any of Claims 1 –3 wherein the *dxs*, *idi*, *ispB* *ygbBP* and *lytB* genes are chromosomally integrated into the host cell genome.
16. A carotenoid overproducing bacteria selected from the group consisting of: a strain having the ATCC identification number PTA-4807 and a strain having the ATCC identification number PTA-4823.
17. A method for the production of a carotenoid comprising:
 - a) growing the carotenoid overproducing bacteria of any of Claims 1 –5, the bacteria overexpressing at least one gene selected from the group consisting of *dxs*, *idi* *ygbBP*, *ispB*, *lytB*, *dxr*, wherein *yjeR* is optionally downregulated, for a time sufficient to produce a carotenoid; and
 - b) optionally recovering the carotenoid from the carotenoid overproducing bacteria of step (a).
18. A method according to Claim 17 wherein the carotenoid is selected from the group consisting of antheraxanthin, adonixanthin, astaxanthin, canthaxanthin, capsorubrin, β -cryptoxanthin, dihydrolycopene, dihydrolycopene, β -carotene, ζ -carotene, δ -carotene, γ -carotene, keto- γ -carotene, ψ -carotene, ϵ -carotene, β,ψ -carotene, torulene, echinenone, gamma-carotene, zeta-carotene, alpha-cryptoxanthin, diatoxanthin, 7,8-didehydroastaxanthin, fucoxanthin, fucoxanthinol, isorenieratene, β -isorenieratene lactucaxanthin, lutein, lycopene, neoxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene,

rhodopin, rhodopin glucoside, siphonaxanthin, spheroidene, spheroidenone, spirilloxanthin, uriolide, uriolide acetate, violaxanthin, zeaxanthin- β -diglucoside, zeaxanthin, and C30-carotenoids.

19. A method according to Claim 18 wherein the carotenoid is
5 produced at a level of at least about 6 mg per gram dry cell weight.

20. A method according to Claim 18 wherein the bacteria is selected from the group consisting *Agrobacterium*, *Erythrobacter*, *Chlorobium*, *Chromatium*, *Flavobacterium*, *Cytophaga*, *Rhodobacter*, *Rhodococcus*, *Streptomyces*, *Brevibacterium*, *Corynebacteria*,
10 *Mycobacterium*, *Deinococcus*, *Paracoccus*, *Escherichia*, *Bacillus*, *Myxococcus*, *Salmonella*, *Yersinia*, *Erwinia*, *Pantoea*, *Pseudomonas*, *Sphingomonas*, *Methylomonas*, *Methylobacter*, *Methylococcus*, *Methylosinus*, *Methylobacterium*, *Methylocystis*, *Alcaligenes*, *Synechocystis*, *Synechococcus*, *Anabaena*, *Thiobacillus*,
15 *Methanobacterium*, *Klebsiella*, and *Myxococcus*.

21. A method according to Claim 20 wherein the bacteria is *E. coli*.

22. A method according to Claim 17 wherein the *dxs*, *idi*, *ygbBP*, *ispB* and *lytB* genes are under the control of a promoter selected from the group consisting of *lac*, *ara*, *tet*, *trp*, λP_L , λP_R , *T7*, *tac*, P_{T5} , and *trc*.

20 23. A method according to Claim 17 wherein the *dxs*, *idi*, *ispB*, *ygbBP* and *lytB* genes are integrated in multicopy in the bacterial chromosome.

24. A method according to Claim 17 wherein the *dxs*, *idi*, *ispB*, *ygbBP* and *lytB* genes are in multicopy in the bacteria on one or more
25 plasmids.

25. A method according to Claim 17 wherein the *yjeR* gene is down regulated by gene disruption.

26. A method according to Claim 25 wherein the disrupted *yjeR* gene has the nucleotide sequence as set forth in SEQ ID NO:63.

30 27. A method according to Claim 17 wherein the *dxs*, *idi*, *ispB*, *ygbBP* and *lytB* genes are chromosomally integrated into the host cell genome.

1/15

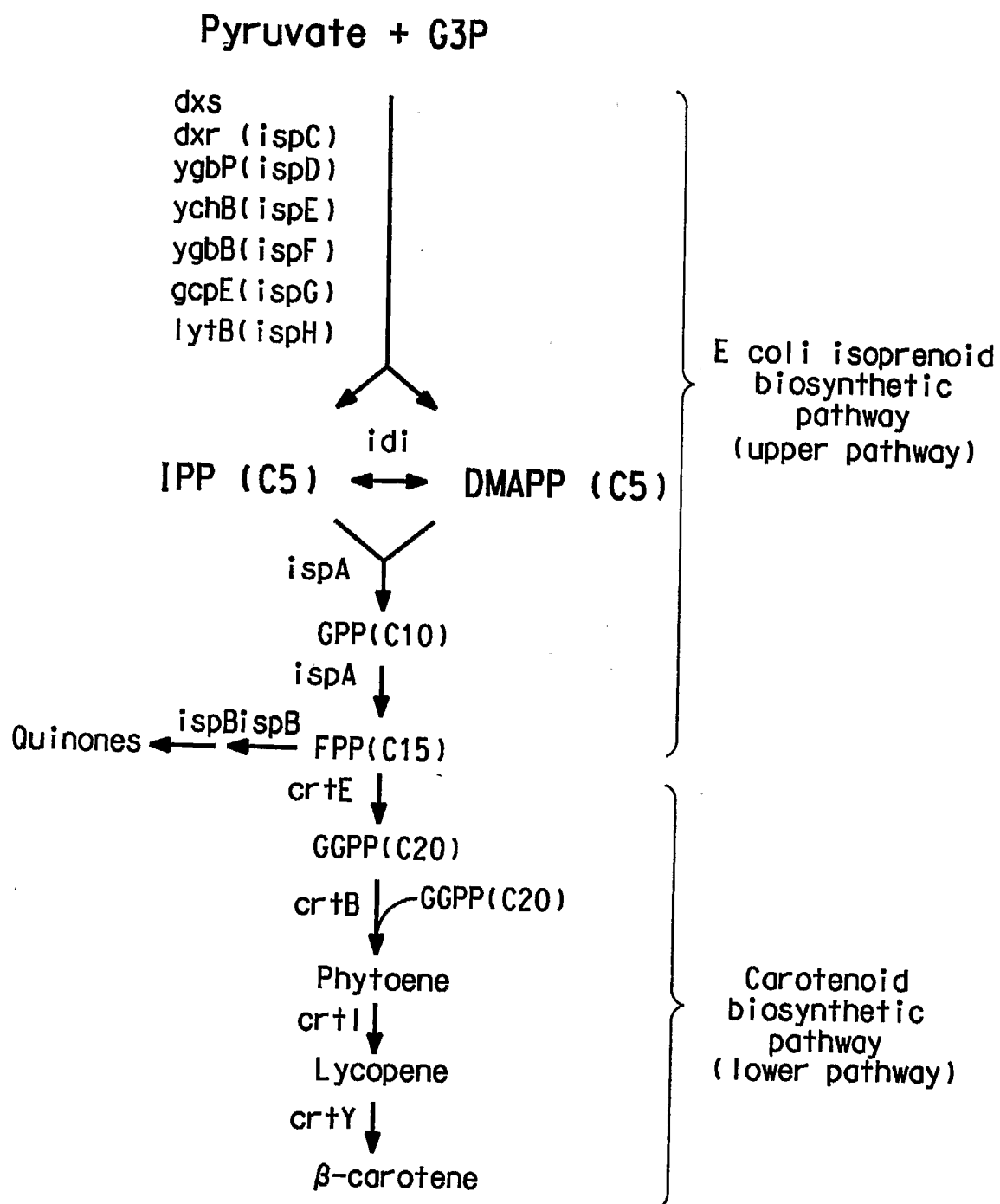
Isoprenoid Pathway in *E. coli*

FIG. 1

2/15

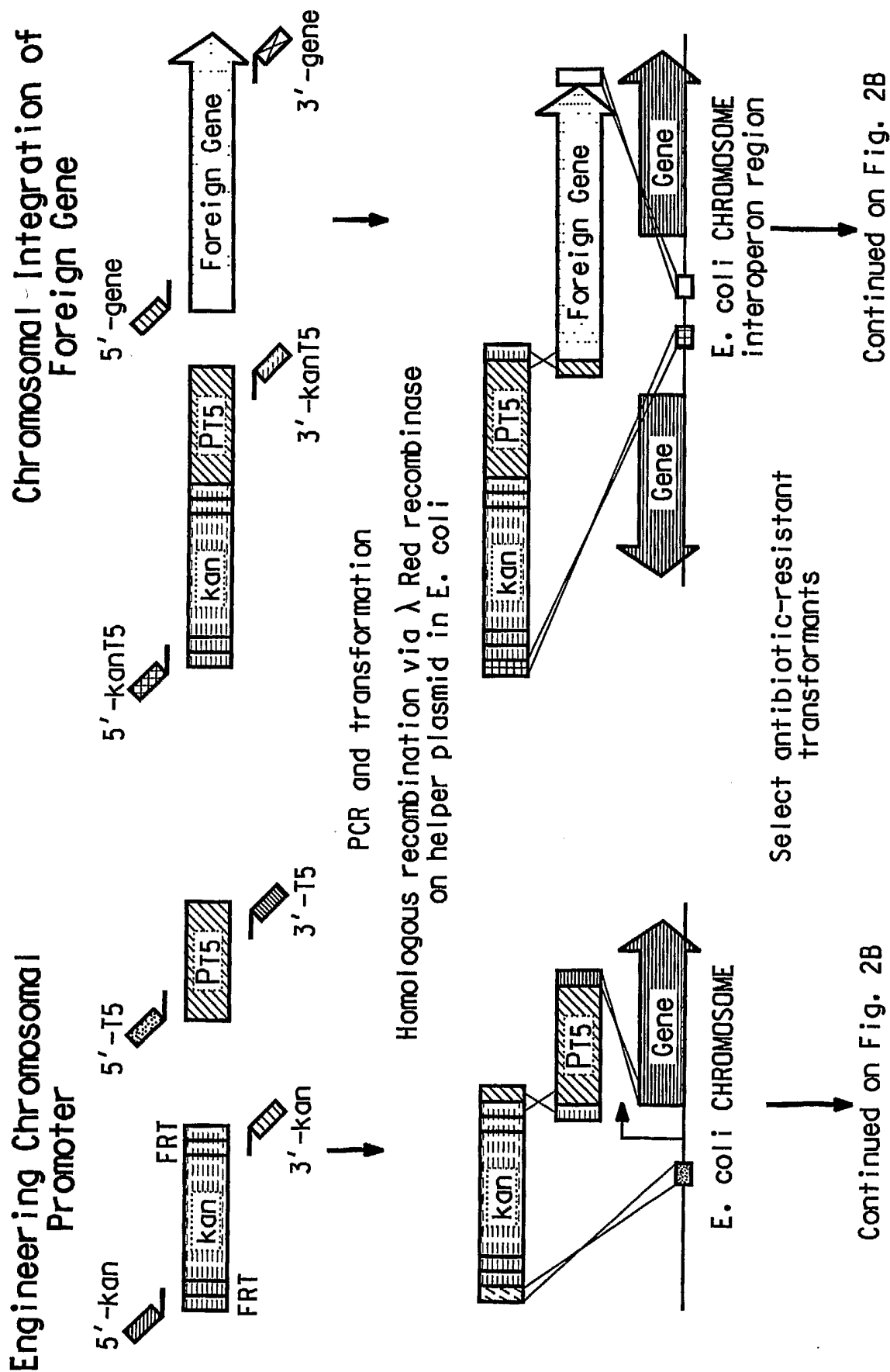


FIG. 2A

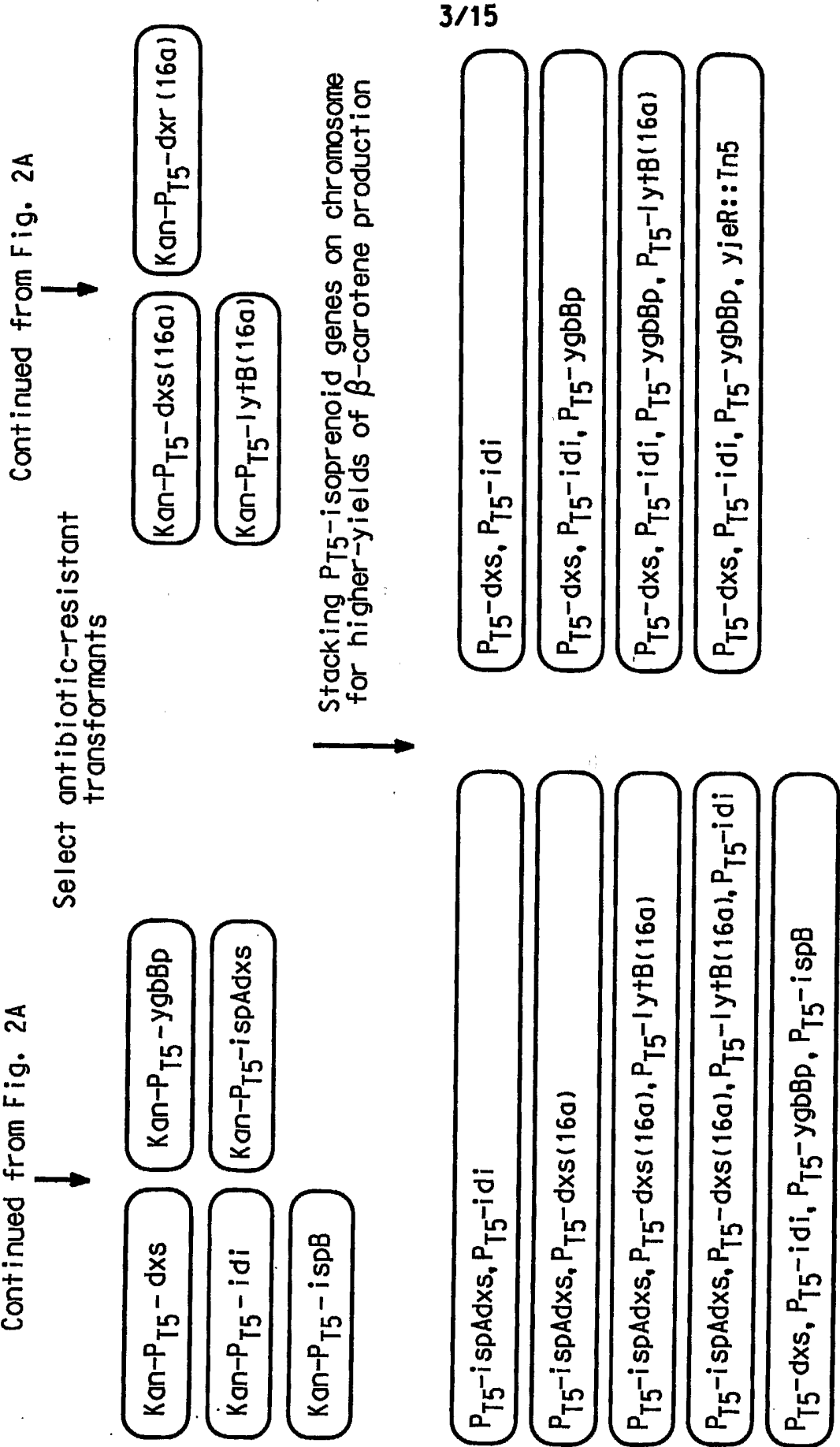


FIG. 2B

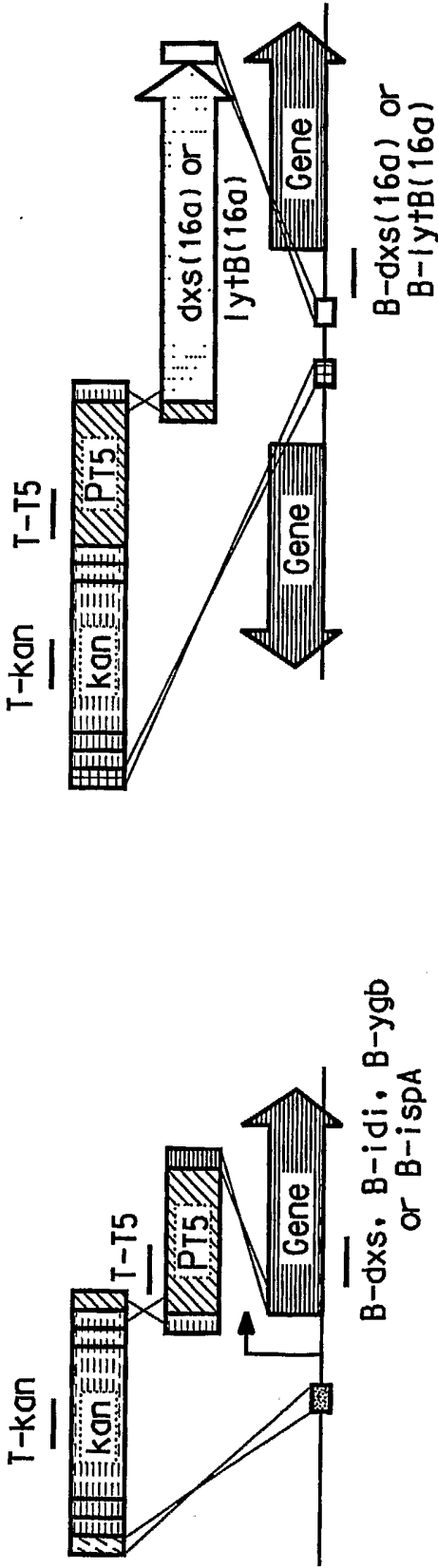


FIG. 3A

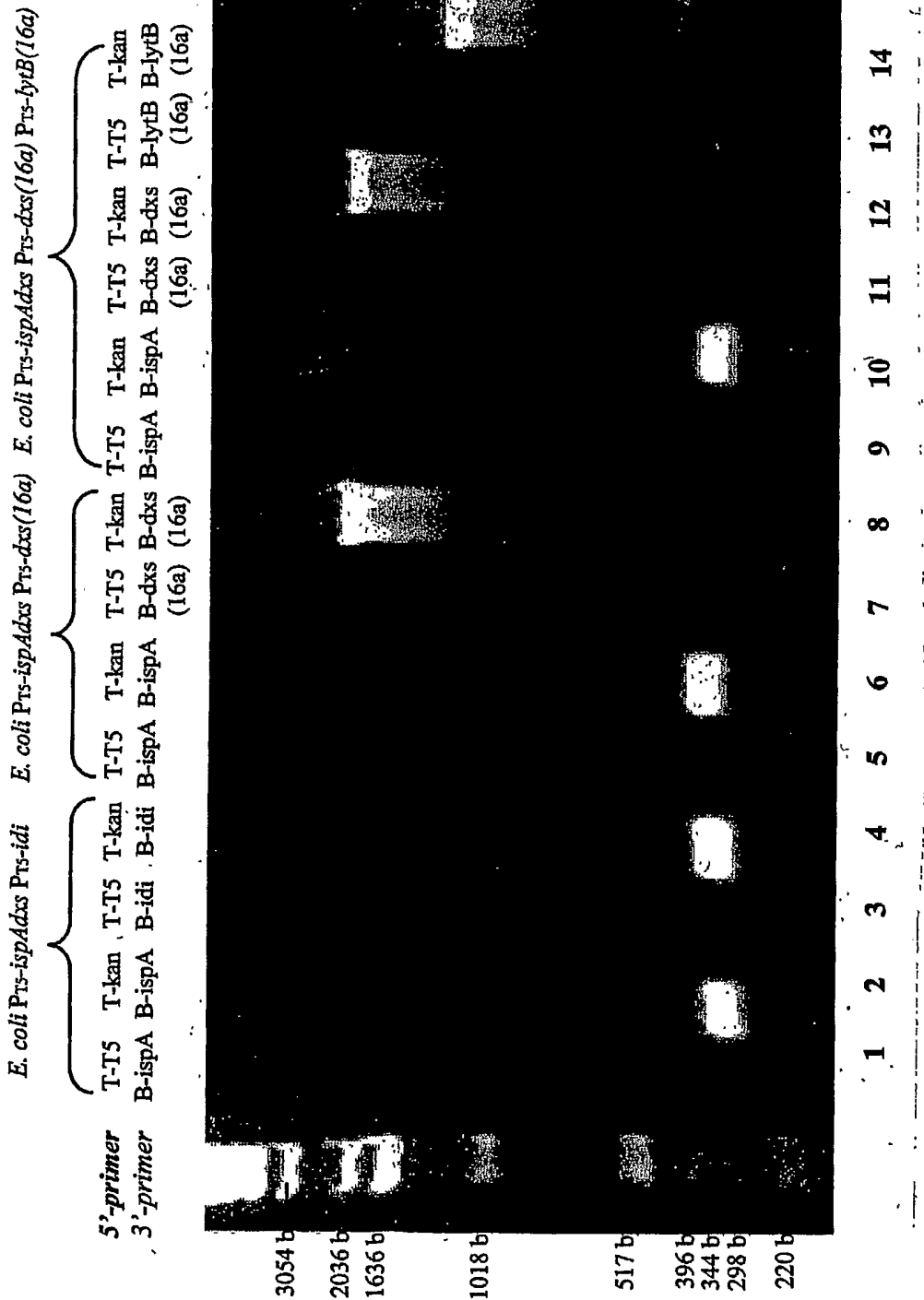


FIG. 3B

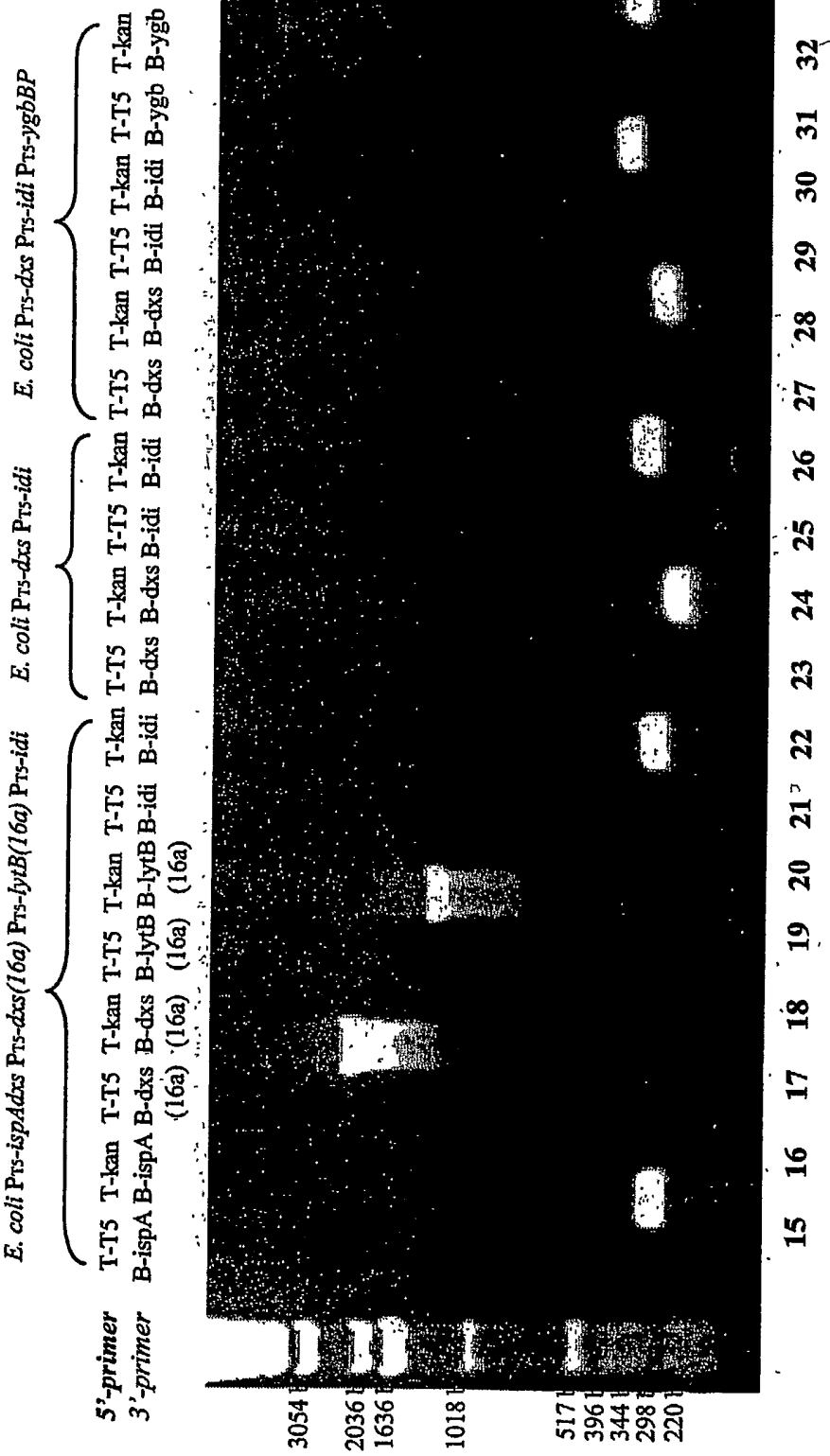


FIG. 4A

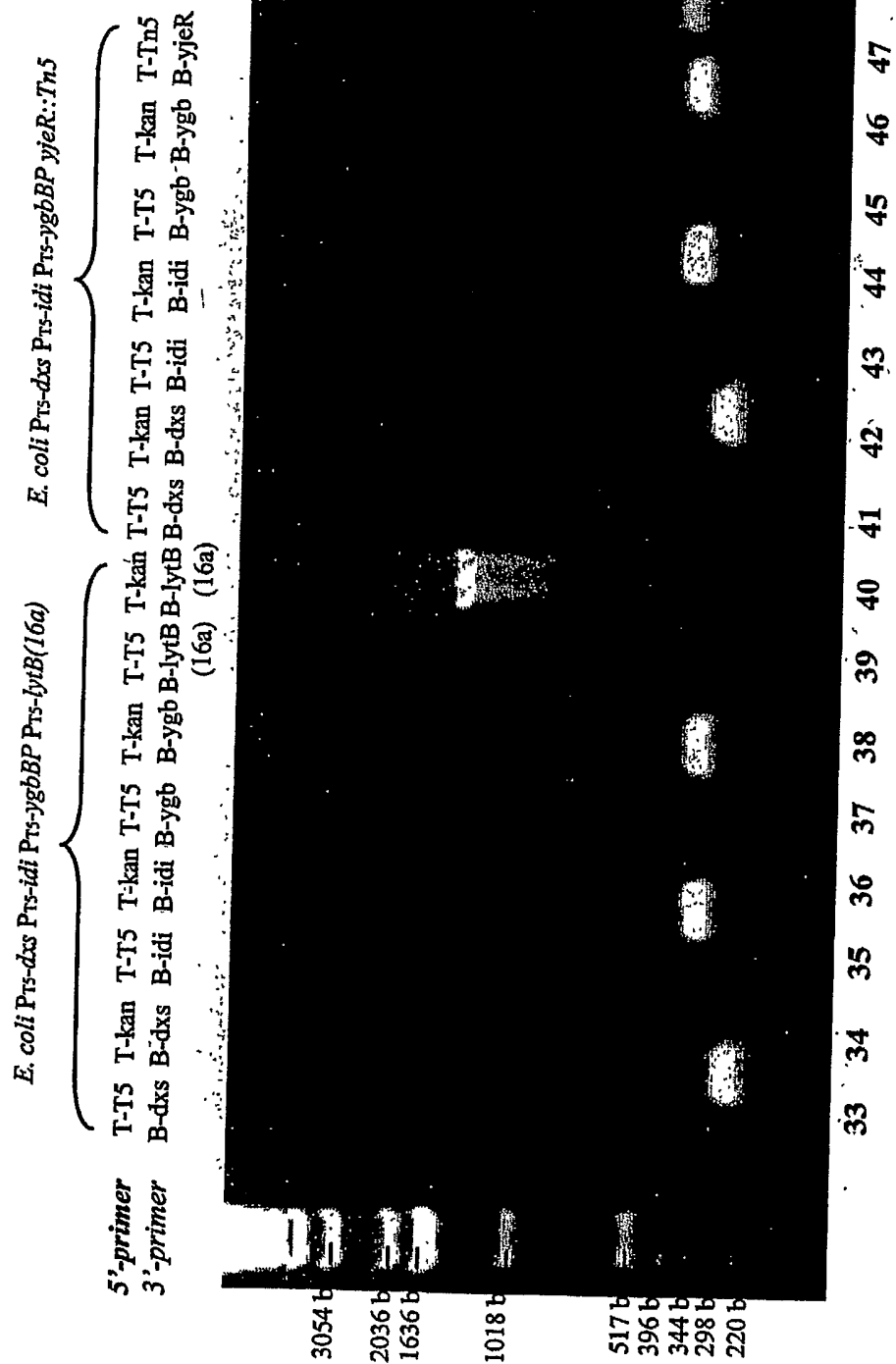


FIG. 4B

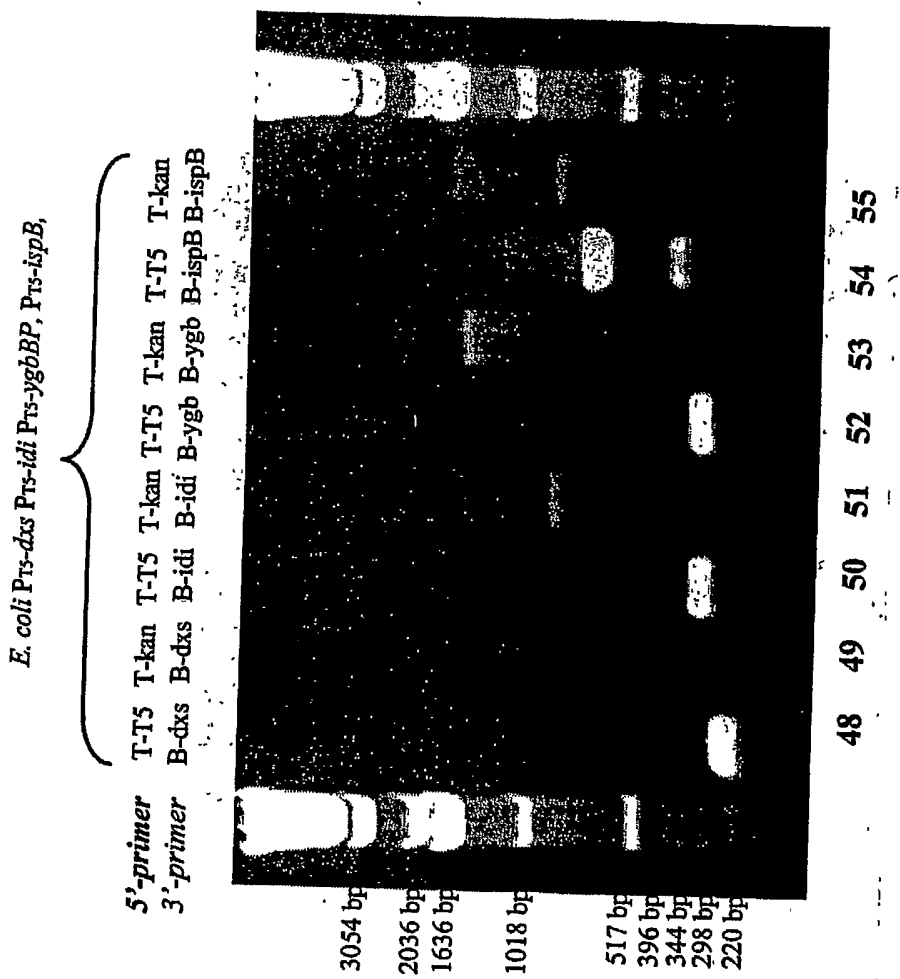
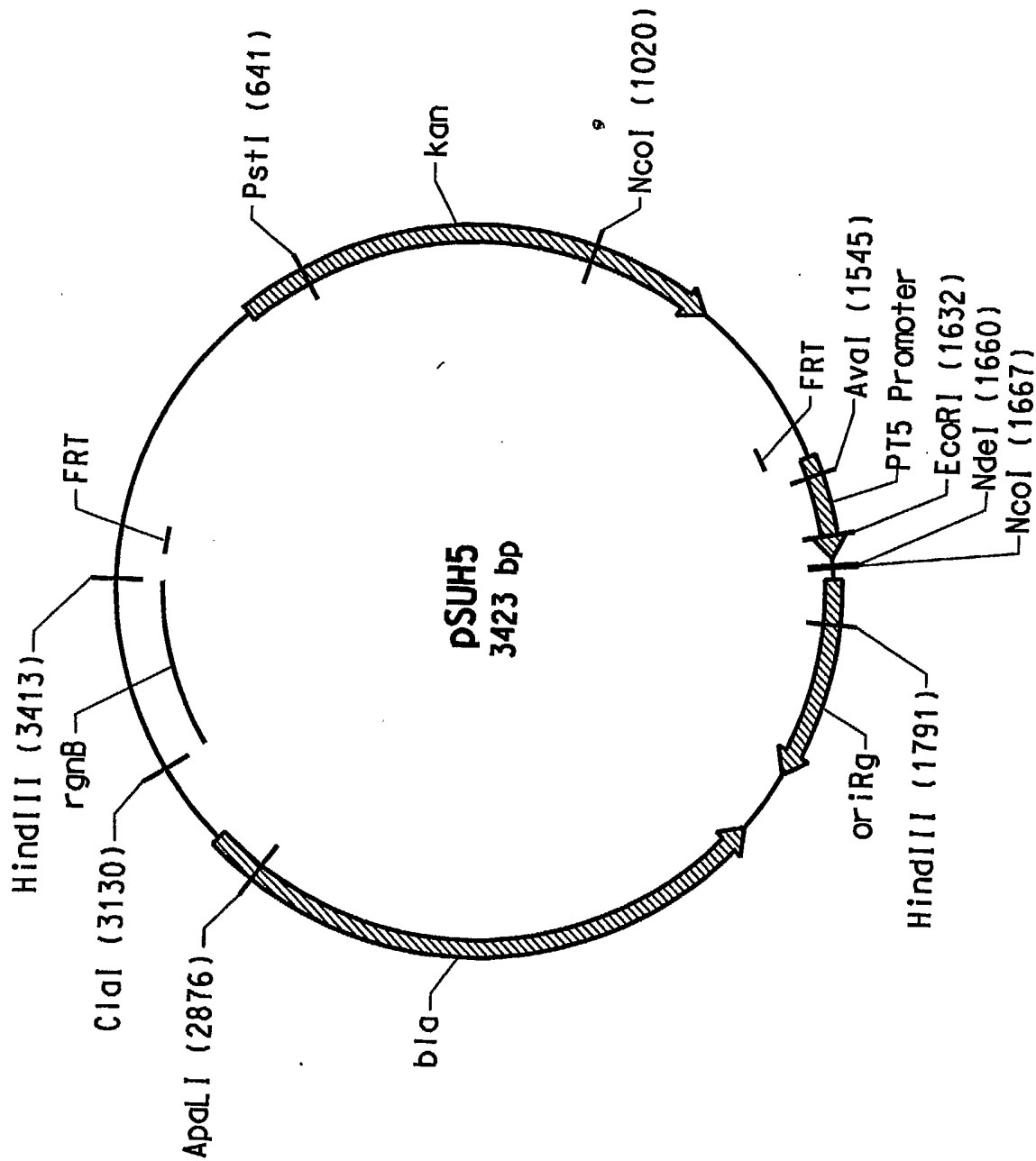


FIG. 5

FIG. 6



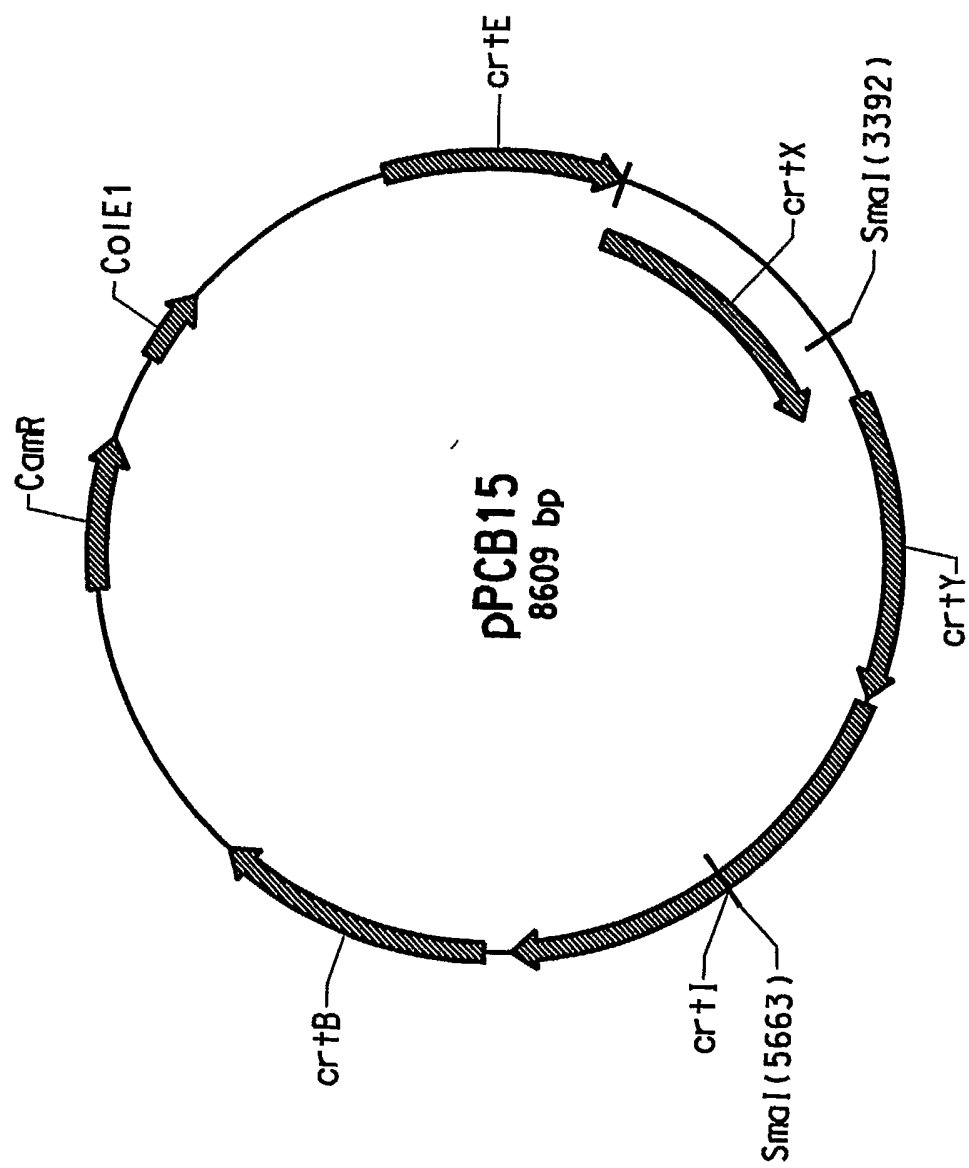


FIG. 7

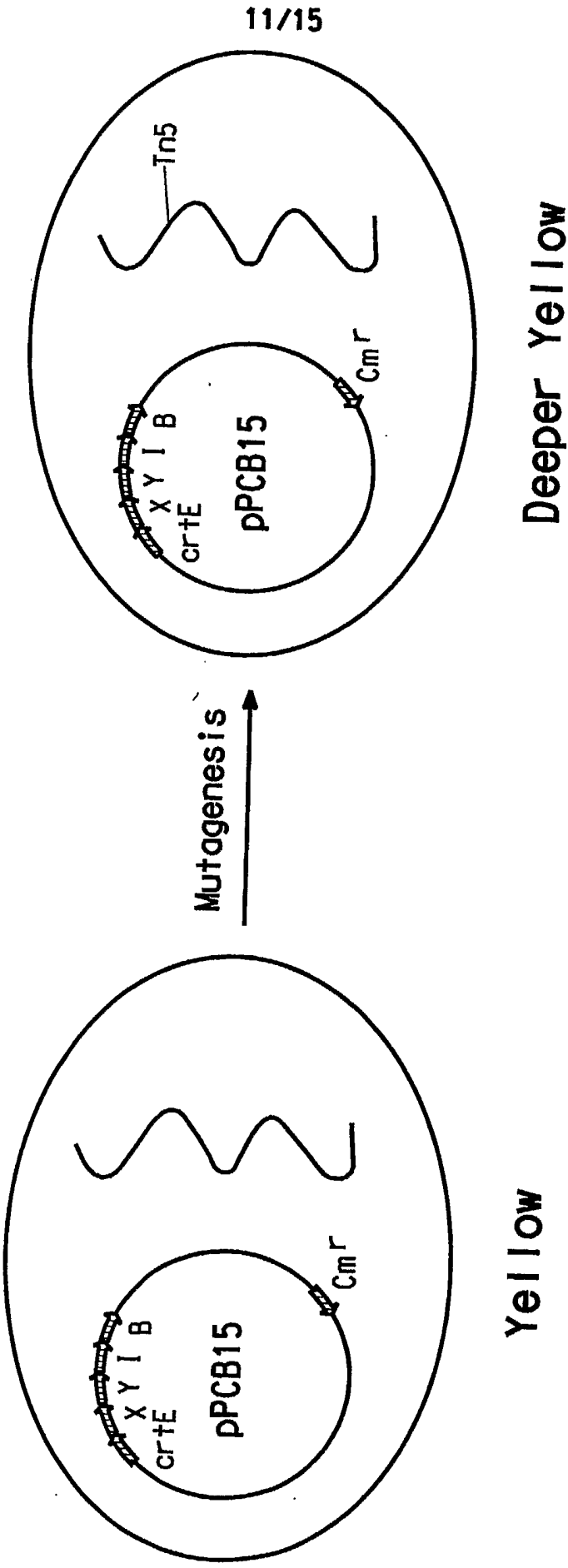


FIG. 8

12/15

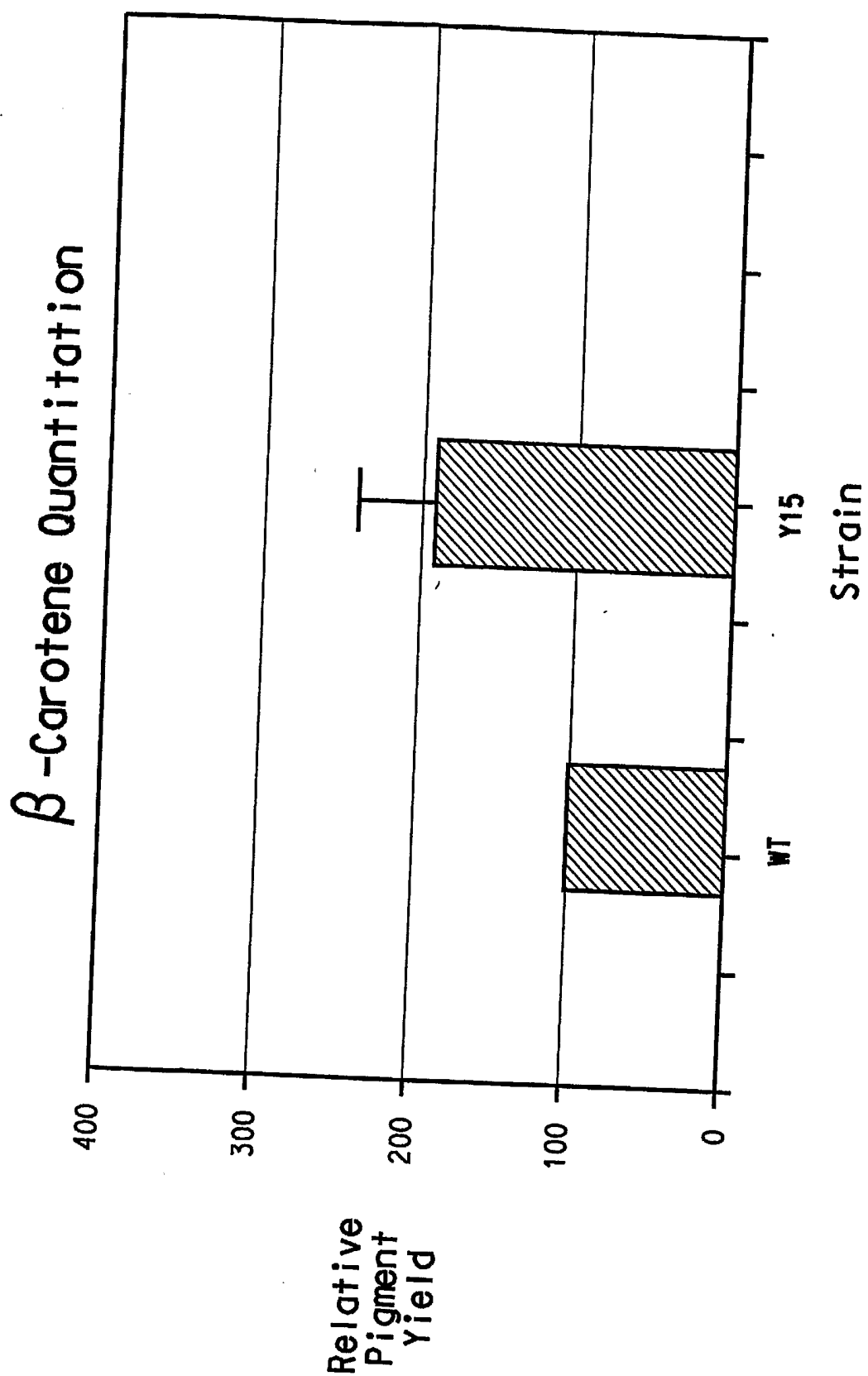


FIG. 9

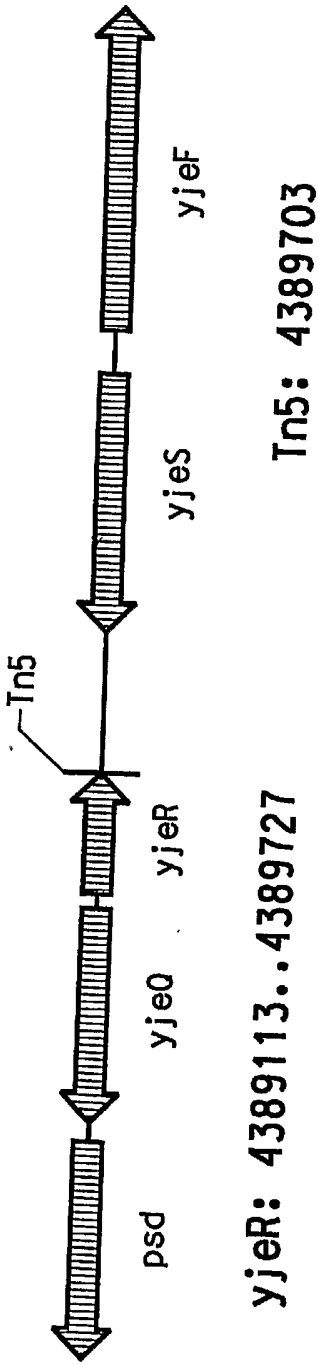


FIG. 10

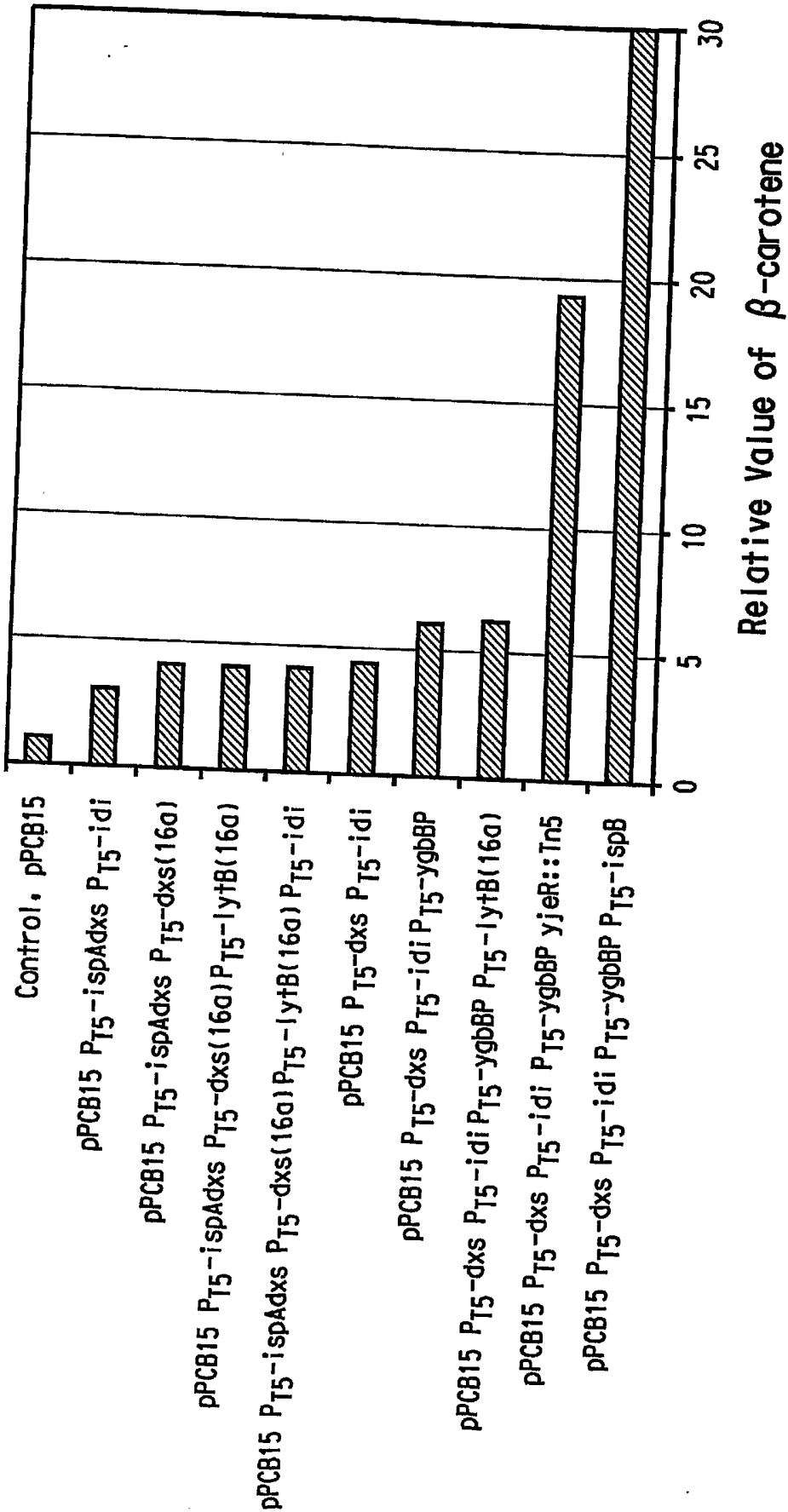


FIG. 11

Combinatorial Approach to Increase β -carotene

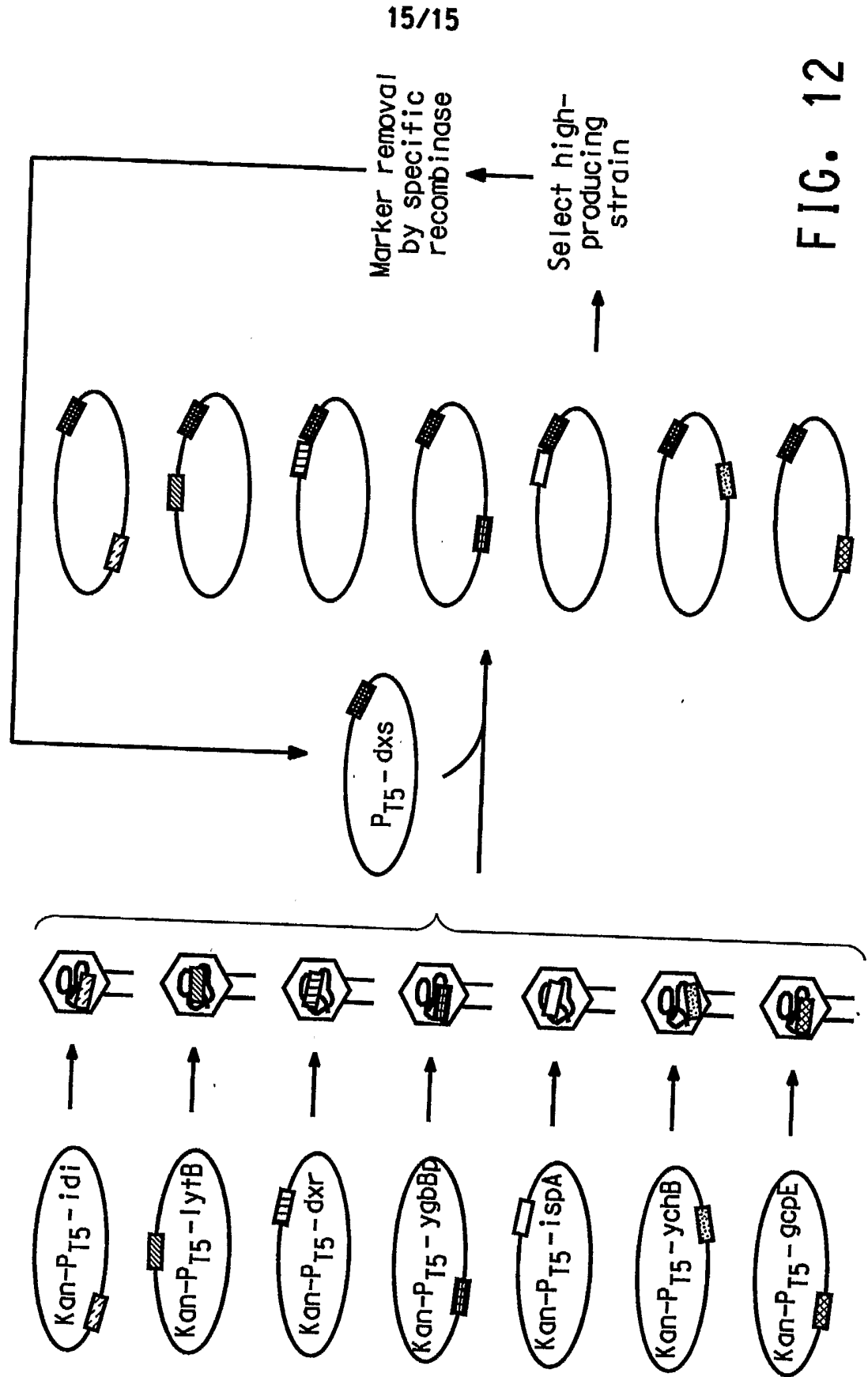


FIG. 12

SEQUENCE LISTING

<110> E. I. duPont de Nemours and Company, Inc.

<120> Increasing Carotenoid Production in Bacteria Via Chromosomal Integration

<130> CL2027 PCT

<150> US 60/434618

<151> 2002-12-19

<160> 66

<170> PatentIn version 3.2

<210> 1

<211> 912

<212> DNA

<213> *Pantoea stewartii*

<220>

<221> misc_feature

<222> (1)..(3)

<223> ttg alternative start codon used to encode methionine

<400> 1

```

ttgacggtct gcgcaaaaaa acacgttcac cttactggca tttcggctga gcagttgctg      60
gctgatatcg atagccgcct tgatcagtta ctgccggttc agggtgagcg ggattgtgtg      120
ggtgccgcga tgcgtgaagg cacgctggca ccgggcaaac gtattcgtcc gatgctgctg      180
ttattaacag cgcgcatct tggctgtgcg atcagtcacg ggggattact ggatttagcc      240
tgcgcggttg aaatggtgca tgctgcctcg ctgattctgg atgatatgcc ctgcatggac      300
gatgcgcaga tgcgtcgggg gcgtcccacc attcacacgc agtacgggtga acatgtggcg      360
attctggcgg cggtcgcttt actcagcaaa gcgtttgggg tgattgccga ggctgaaggt      420
ctgacgccga tagccaaaac tcgcgcggtg tcggagctgt cactgcatg tggcatgcag      480
ggtctggttc agggccagtt taaggacctc tcggaaggcg ataaaccccg cagcgccgat      540
gccatactgc taaccaatca gtttaaaacc agcacgctgt tttgcgcgtc aacgcaaatg      600
gcgtccattg cggccaacgc gtctcgcaa gcgcgtgaga acctgcatcg tttctcgctc      660
gatctcggcc aggcctttca gttgcttgac gatcttaccg atggcatgac cgataccggc      720
aaagacatca atcaggatgc aggtaaatca acgctgggtca atttattagg ctgaggcgcg      780
gtcgaagaac gcctgcgaca gcatttgcg ctggccagtg aacacctttc cgcggcacgc      840
caaaacggcc attccaccac ccaacttttt attcaggcct ggtttgacaa aaaactcgct      900
gccgtcagtt aa                                                    912

```

<210> 2

<211> 303

<212> PRT

<213> *Pantoea stewartii*

<400> 2

Met Thr Val Cys Ala Lys Lys His Val His Leu Thr Gly Ile Ser Ala

1	5	10	15
Glu Gln Leu	Leu Ala Asp Ile Asp	Ser Arg Leu Asp Gln	Leu Leu Pro
	20	25	30
Val Gln Gly	Glu Arg Asp Cys Val Gly Ala Ala Met	Arg Glu Gly Thr	
	35	40	45
Leu Ala Pro Gly Lys Arg	Ile Arg Pro Met Leu Leu Leu Thr Ala		
	50	55	60
Arg Asp Leu Gly Cys	Ala Ile Ser His Gly Gly Leu Leu Asp Leu Ala		
	65	70	75
Cys Ala Val Glu	Met Val His Ala Ala Ser Leu Ile Leu Asp Asp Met		
	85	90	95
Pro Cys Met Asp	Asp Ala Gln Met Arg Arg Gly Arg Pro Thr Ile His		
	100	105	110
Thr Gln Tyr	Gly Glu His Val Ala Ile Leu Ala Ala Val Ala Leu Leu		
	115	120	125
Ser Lys Ala Phe Gly	Val Ile Ala Glu Ala Glu Gly Leu Thr Pro Ile		
	130	135	140
Ala Lys Thr Arg Ala	Val Ser Glu Leu Ser Thr Ala Ile Gly Met Gln		
	145	150	155
Gly Leu Val Gln	Gly Gln Phe Lys Asp Leu Ser Glu Gly Asp Lys Pro		
	165	170	175
Arg Ser Ala Asp	Ala Ile Leu Leu Thr Asn Gln Phe Lys Thr Ser Thr		
	180	185	190
Leu Phe Cys	Ala Ser Thr Gln Met Ala Ser Ile Ala Ala Asn Ala Ser		
	195	200	205
Cys Glu Ala Arg Glu	Asn Leu His Arg Phe Ser Leu Asp Leu Gly Gln		
	210	215	220
Ala Phe Gln Leu Leu	Asp Asp Leu Thr Asp Gly Met Thr Asp Thr Gly		
	225	230	235
Lys Asp Ile Asn	Gln Asp Ala Gly Lys Ser Thr Leu Val Asn Leu Leu		
	245	250	255
Gly Ser Gly Ala	Val Glu Glu Arg Leu Arg Gln His Leu Arg Leu Ala		
	260	265	270
Ser Glu His	Leu Ser Ala Ala Cys Gln Asn Gly His Ser Thr Thr Gln		
	275	280	285

Leu phe Ile Gln Ala Trp Phe Asp Lys Lys Leu Ala Ala Val Ser
 290 295 300

<210> 3
 <211> 1296
 <212> DNA
 <213> Pantoea stewartii

<220>
 <221> CDS
 <222> (1)..(1296)

<400> 3
 atg agc cat ttt gcg gtg atc gca ccg ccc ttt ttc agc cat gtt cgc 48
 Met Ser His Phe Ala Val Ile Ala Pro Phe Phe Ser His Val Arg
 1 5 10 15
 gct ctg caa aac ctt gct cag gaa tta gtg gcc cgc ggt cat cgt gtt 96
 Ala Leu Gln Asn Leu Ala Gln Glu Leu Val Ala Arg Gly His Arg Val
 20 25 30
 acg ttt ttt cag caa cat gac tgc aaa gcg ctg gta acg ggc agc gat 144
 Thr Phe Phe Gln Gln His Asp Cys Lys Ala Leu Val Thr Gly Ser Asp
 35 40 45
 atc gga ttc cag acc gtc gga ctg caa acg cat cct ccc ggt tcc tta 192
 Ile Gly Phe Gln Thr Val Gly Leu Gln Thr His Pro Pro Gly Ser Leu
 50 55 60
 tcg cac ctg ctg cac ctg gcc gcg cac cca ctc gga ccc tcg atg tta 240
 Ser His Leu Leu His Leu Ala Ala His Pro Leu Gly Pro Ser Met Leu
 65 70 75 80
 cga ctg atc aat gaa atg gca cgt acc agc gat atg ctt tgc cgg gaa 288
 Arg Leu Ile Asn Glu Met Ala Arg Thr Ser Asp Met Leu Cys Arg Glu
 85 90 95
 ctg ccc gcc gct ttt cat gcg ttg cag ata gag ggc gtg atc gtt gat 336
 Leu Pro Ala Ala Phe His Ala Leu Gln Ile Glu Gly Val Ile Val Asp
 100 105 110
 caa atg gag ccg gca ggt gca gta gtc gca gaa gcg tca ggt ctg ccg 384
 Gln Met Glu Pro Ala Gly Ala Val Val Ala Glu Ala Ser Gly Leu Pro
 115 120 125
 ttt gtt tcg gtg gcc tgc gcg ctg ccg ctc aac cgc gaa ccg ggt ttg 432
 Phe Val Ser Val Ala Cys Ala Leu Pro Leu Asn Arg Glu Pro Gly Leu
 130 135 140
 cct ctg gcg gtg atg cct ttc gag tac ggc acc agc gat gcg gct cgg 480
 Pro Leu Ala Val Met Pro Phe Glu Tyr Gly Thr Ser Asp Ala Ala Arg
 145 150 155 160
 gaa cgc tat acc acc agc gaa aaa att tat gac tgg ctg atg cga cgt 528
 Glu Arg Tyr Thr Thr Ser Glu Lys Ile Tyr Asp Trp Leu Met Arg Arg
 165 170 175
 cac gat cgt gtg atc gcg cat cat gca tgc aga atg ggt tta gcc ccg 576
 His Asp Arg Val Ile Ala His His Ala Cys Arg Met Gly Leu Ala Pro
 180 185 190
 cgt gaa aaa ctg cat cat tgt ttt tct cca ctg gca caa atc agc cag 624
 Arg Glu Lys Leu His His Cys Phe Ser Pro Leu Ala Gln Ile Ser Gln
 195 200 205

ttg atc ccc gaa ctg gat ttt ccc cgc aaa gcg ctg cca gac tgc ttt Leu Ile Pro Glu Leu Asp Phe Pro Arg Lys Ala Leu Pro Asp Cys Phe 210 215 220	672
cat gcg gtt gga ccg tta cgg caa ccc cag ggg acg ccg ggg tca tca His Ala Val Gly Pro Leu Arg Gln Pro Gln Gly Thr Pro Gly Ser Ser 225 230 235 240	720
act tct tat ttt ccg tcc ccg gac aaa ccc cgt att ttt gcc tcg ctg Thr Ser Tyr Phe Pro Ser Pro Asp Lys Pro Arg Ile Phe Ala Ser Leu 245 250 255	768
ggc acc ctg cag gga cat cgt tat ggc ctg ttc agg acc atc gcc aaa Gly Thr Leu Gln Gly His Arg Tyr Gly Leu Phe Arg Thr Ile Ala Lys 260 265 270	816
gcc tgc gaa gag gtg gat gcg cag tta ctg ttg gca cac tgt ggc ggc Ala Cys Glu Glu Val Asp Ala Gln Leu Leu Leu Ala His Cys Gly Gly 275 280 285	864
ctc tca gcc acg cag gca ggt gaa ctg gcc cgg ggc ggg gac att cag Leu Ser Ala Thr Gln Ala Gly Glu Leu Ala Arg Gly Gly Asp Ile Gln 290 295 300	912
gtt gtg gat ttt gcc gat caa tcc gca gca ctt tca cag gca cag ttg Val Val Asp Phe Ala Asp Gln Ser Ala Ala Leu Ser Gln Ala Gln Leu 305 310 315 320	960
aca atc aca cat ggt ggg atg aat acg gta ctg gac gct att gct tcc Thr Ile Thr His Gly Gly Met Asn Thr Val Leu Asp Ala Ile Ala Ser 325 330 335	1008
cgc aca ccg cta ctg gcg ctg ccg ctg gca ttt gat caa cct ggc gtg Arg Thr Pro Leu Leu Ala Leu Pro Leu Ala Phe Asp Gln Pro Gly Val 340 345 350	1056
gca tca cga att gtt tat cat ggc atc ggc aag cgt gcg tct cgg ttt Ala Ser Arg Ile Val Tyr His Gly Ile Gly Lys Arg Ala Ser Arg Phe 355 360 365	1104
act acc agc cat gcg ctg gcg cgg cag att cga tcg ctg ctg act aac Thr Thr Ser His Ala Leu Ala Arg Gln Ile Arg Ser Leu Leu Thr Asn 370 375 380	1152
acc gat tac ccg cag cgt atg aca aaa att cag gcc gca ttg cgt ctg Thr Asp Tyr Pro Gln Arg Met Thr Lys Ile Gln Ala Ala Leu Arg Leu 385 390 400	1200
gca ggc ggc aca cca gcc gcc gcc gat att gtt gaa cag gcg atg cgg Ala Gly Gly Thr Pro Ala Ala Ala Asp Ile Val Glu Gln Ala Met Arg 405 410 415	1248
acc tgt cag cca gta ctc agt ggg cag gat tat gca acc gca cta tga Thr Cys Gln Pro Val Leu Ser Gly Gln Asp Tyr Ala Thr Ala Leu 420 425 430	1296

<210> 4
 <211> 431
 <212> PRT
 <213> Pantoea stewartii

<400> 4

Met Ser His Phe Ala Val Ile Ala Pro Pro Phe Phe Ser His Val Arg
 1 5 10 15

Ala Leu Gln Asn Leu Ala Gln Glu Leu Val Ala Arg Gly His Arg Val
 Page 4

20					25					30					
Thr	Phe	Phe	Gln	Gln	His	Asp	Cys	Lys	Ala	Leu	Val	Thr	Gly	Ser	Asp
		35					40					45			
Ile	Gly	Phe	Gln	Thr	Val	Gly	Leu	Gln	Thr	His	Pro	Pro	Gly	Ser	Leu
	50					55					60				
Ser	His	Leu	Leu	His	Leu	Ala	Ala	His	Pro	Leu	Gly	Pro	Ser	Met	Leu
65					70					75					80
Arg	Leu	Ile	Asn	Glu	Met	Ala	Arg	Thr	Ser	Asp	Met	Leu	Cys	Arg	Glu
				85					90					95	
Leu	Pro	Ala	Ala	Phe	His	Ala	Leu	Gln	Ile	Glu	Gly	Val	Ile	Val	Asp
			100					105					110		
Gln	Met	Glu	Pro	Ala	Gly	Ala	Val	Val	Ala	Glu	Ala	Ser	Gly	Leu	Pro
		115					120					125			
Phe	Val	Ser	Val	Ala	Cys	Ala	Leu	Pro	Leu	Asn	Arg	Glu	Pro	Gly	Leu
	130					135					140				
Pro	Leu	Ala	Val	Met	Pro	Phe	Glu	Tyr	Gly	Thr	Ser	Asp	Ala	Ala	Arg
145					150					155					160
Glu	Arg	Tyr	Thr	Thr	Ser	Glu	Lys	Ile	Tyr	Asp	Trp	Leu	Met	Arg	Arg
				165					170					175	
His	Asp	Arg	Val	Ile	Ala	His	His	Ala	Cys	Arg	Met	Gly	Leu	Ala	Pro
			180					185					190		
Arg	Glu	Lys	Leu	His	His	Cys	Phe	Ser	Pro	Leu	Ala	Gln	Ile	Ser	Gln
		195					200					205			
Leu	Ile	Pro	Glu	Leu	Asp	Phe	Pro	Arg	Lys	Ala	Leu	Pro	Asp	Cys	Phe
	210					215					220				
His	Ala	Val	Gly	Pro	Leu	Arg	Gln	Pro	Gln	Gly	Thr	Pro	Gly	Ser	Ser
225					230					235					240
Thr	Ser	Tyr	Phe	Pro	Ser	Pro	Asp	Lys	Pro	Arg	Ile	Phe	Ala	Ser	Leu
				245					250					255	
Gly	Thr	Leu	Gln	Gly	His	Arg	Tyr	Gly	Leu	Phe	Arg	Thr	Ile	Ala	Lys
			260					265					270		
Ala	Cys	Glu	Glu	Val	Asp	Ala	Gln	Leu	Leu	Leu	Ala	His	Cys	Gly	Gly
		275					280					285			
Leu	Ser	Ala	Thr	Gln	Ala	Gly	Glu	Leu	Ala	Arg	Gly	Gly	Asp	Ile	Gln
	290					295					300				

Val Val Asp Phe Ala Asp Gln Ser Ala Ala Leu Ser Gln Ala Gln Leu
305 310 315 320

Thr Ile Thr His Gly Gly Met Asn Thr Val Leu Asp Ala Ile Ala Ser
325 330 335

Arg Thr Pro Leu Leu Ala Leu Pro Leu Ala Phe Asp Gln Pro Gly Val
340 345 350

Ala Ser Arg Ile Val Tyr His Gly Ile Gly Lys Arg Ala Ser Arg Phe
355 360 365

Thr Thr Ser His Ala Leu Ala Arg Gln Ile Arg Ser Leu Leu Thr Asn
370 375 380

Thr Asp Tyr Pro Gln Arg Met Thr Lys Ile Gln Ala Ala Leu Arg Leu
385 390 395 400

Ala Gly Gly Thr Pro Ala Ala Ala Asp Ile Val Glu Gln Ala Met Arg
405 410 415

Thr Cys Gln Pro Val Leu Ser Gly Gln Asp Tyr Ala Thr Ala Leu
420 425 430

<210> 5
<211> 1149
<212> DNA
<213> Pantoea stewartii

<220>
<221> CDS
<222> (1)..(1149)

<400> 5	
atg caa ccg cac tat gat ctc att ctg gtc ggt gcc ggt ctg gct aat	48
Met Gln Pro His Tyr Asp Leu Ile Leu Val Gly Ala Gly Leu Ala Asn	
1 5 10 15	
ggc ctt atc gcg ctc cgg ctt cag caa cag cat ccg gat atg cgg atc	96
Gly Leu Ile Ala Leu Arg Leu Gln Gln His Pro Asp Met Arg Ile	
20 25 30	
ttg ctt att gag gcg ggt cct gag gcg gga ggg aac cat acc tgg tcc	144
Leu Leu Ile Glu Ala Gly Pro Glu Ala Gly Gly Asn His Thr Trp Ser	
35 40 45	
ttt cac gaa gag gat tta acg ctg aat cag cat cgc tgg ata gcg ccg	192
Phe His Glu Glu Asp Leu Thr Leu Asn Gln His Arg Trp Ile Ala Pro	
50 55 60	
ctt gtg gtc cat cac tgg ccc gac tac cag gtt cgt ttc ccc caa cgc	240
Leu Val Val His His Trp Pro Asp Tyr Gln Val Arg Phe Pro Gln Arg	
65 70 75 80	
cgt cgc cat gtg aac agt ggc tac tac tgc gtg acc tcc cgg cat ttc	288
Arg Arg His Val Asn Ser Gly Tyr Tyr Cys Val Thr Ser Arg His Phe	
85 90 95	

gcc Ala	ggg Gly	ata Ile	ctc Leu 100	cgg Arg	caa Gln	cag Gln	ttt Phe	gga Gly 105	caa Gln	cat His	tta Leu	tgg Trp	ctg Leu 110	cat His	acc Thr	336
gcg Ala	gtt Val	tca Ser 115	gcc Ala	gtt Val	cat His	gct Ala	gaa Glu 120	tcg Ser	gtc Val	cag Gln	tta Leu	gcg Ala 125	gat Asp	ggc Gly	cgg Arg	384
att Ile	att Ile 130	cat His	gcc Ala	agt Ser	aca Thr	gtg Val 135	atc Ile	gac Asp	gga Gly	cgg Arg	ggt Gly 140	tac Tyr	acg Thr	cct Pro	gat Asp	432
tct Ser 145	gca Ala	cta Leu	cgc Arg	gta Val	gga Gly 150	ttc Phe	cag Gln	gca Ala	ttt Phe 155	atc Ile	ggt Gly	cag Gln	gag Glu	tgg Trp	caa Gln 160	480
ctg Leu	agc Ser	gcg Ala	ccg Pro	cat His 165	ggt Gly	tta Leu	tcg Ser	tca Ser	ccg Pro 170	att Ile	atc Ile	atg Met	gat Asp	gcg Ala 175	acg Thr	528
gtc Val	gat Asp	cag Gln	caa Gln 180	aat Asn	ggc Gly	tac Tyr	cgc Arg	ttt Phe 185	gtt Val	tat Tyr	acc Thr	ctg Leu 190	ccg Pro	ctt Leu	tcc Ser	576
gca Ala	acc Thr	gca Ala 195	ctg Leu	ctg Leu	atc Ile	gaa Glu	gac Asp 200	aca Thr	cac His	tac Tyr	att Ile	gac Asp 205	aag Lys	gct Ala	aat Asn	624
ctt Leu	cag Gln 210	gcc Ala	gaa Glu	cgg Arg	gcg Ala	cgt Arg 215	cag Gln	aac Asn	att Ile	cgc Arg	gat Asp 220	tat Tyr	gct Ala	gcg Ala	cga Arg	672
cag Gln 225	ggt Gly	tgg Trp	ccg Pro	tta Leu	cag Gln 230	acg Thr	ttg Leu	ctg Leu	cgg Arg	gaa Glu 235	gaa Glu	cag Gln	ggt Gly	gca Ala	ttg Leu 240	720
ccc Pro	att Ile	acg Thr	tta Leu	acg Thr 245	ggc Gly	gat Asp	aat Asn	cgt Arg	cag Gln 250	ttt Phe	tgg Trp	caa Gln	cag Gln	caa Gln 255	ccg Pro	768
caa Gln	gcc Ala	tgt Cys	agc Ser 260	gga Gly	tta Leu	cgc Arg	gcc Ala	ggg Gly 265	ctg Leu	ttt Phe	cat His	ccg Pro	aca Thr 270	acc Thr	ggc Gly	816
tac Tyr	tcc Ser	cta Leu 275	ccg Pro	ctc Leu	gcg Ala	gtg Val 280	gcg Ala	ctg Leu	gcc Ala	gat Asp	cgt Arg	ctc Leu 285	agc Ser	gcg Ala	ctg Leu	864
gat Asp	gtg Val 290	ttt Phe	acc Thr	tct Ser	tcc Ser	tct Ser 295	gtt Val	cac His	cag Gln	acg Thr	att Ile 300	gct Ala	cac His	ttt Phe	gcc Ala	912
cag Gln 305	caa Gln	cgt Arg	tgg Trp	cag Gln	caa Gln 310	cag Gln	ggg Gly	ttt Phe	ttc Phe	cgc Arg 315	atg Met	ctg Leu	aat Asn	cgc Arg	atg Met 320	960
ttg Leu	ttt Phe	tta Leu	gcc Ala	gga Gly 325	ccg Pro	gcc Ala	gag Glu	tca Ser	cgc Arg 330	tgg Trp	cgt Arg	gtg Val	atg Met	cag Gln 335	cgt Arg	1008
ttc Phe	tat Tyr	ggc Gly	tta Leu 340	ccc Pro	gag Glu	gat Asp	ttg Leu	att Ile 345	gcc Ala	cgc Arg	ttt Phe	tat Tyr 350	gcg Ala	gga Gly	aaa Lys	1056
ctc Leu	acc Thr	gtg Val 355	acc Thr	gat Asp	cgg Arg	cta Leu	cgc Arg 360	att Ile	ctg Leu	agc Ser	ggc Gly	aag Lys 365	ccg Pro	ccc Pro	gtt Val	1104
ccc	gtt	ttc	gcg	gca	ttg	cag	gca	att	atg	acg	act	cat	cgt	tga		1149

Pro Val Phe Ala Ala Leu Gln Ala Ile Met Thr Thr His Arg
370 375 380

<210> 6
<211> 382
<212> PRT
<213> Pantoea stewartii

<400> 6

Met Gln Pro His Tyr Asp Leu Ile Leu Val Gly Ala Gly Leu Ala Asn
1 5 10 15

Gly Leu Ile Ala Leu Arg Leu Gln Gln Gln His Pro Asp Met Arg Ile
20 25 30

Leu Leu Ile Glu Ala Gly Pro Glu Ala Gly Gly Asn His Thr Trp Ser
35 40 45

Phe His Glu Glu Asp Leu Thr Leu Asn Gln His Arg Trp Ile Ala Pro
50 55 60

Leu Val Val His His Trp Pro Asp Tyr Gln Val Arg Phe Pro Gln Arg
65 70 75 80

Arg Arg His Val Asn Ser Gly Tyr Tyr Cys Val Thr Ser Arg His Phe
85 90 95

Ala Gly Ile Leu Arg Gln Gln Phe Gly Gln His Leu Trp Leu His Thr
100 105 110

Ala Val Ser Ala Val His Ala Glu Ser Val Gln Leu Ala Asp Gly Arg
115 120 125

Ile Ile His Ala Ser Thr Val Ile Asp Gly Arg Gly Tyr Thr Pro Asp
130 135 140

Ser Ala Leu Arg Val Gly Phe Gln Ala Phe Ile Gly Gln Glu Trp Gln
145 150 155 160

Leu Ser Ala Pro His Gly Leu Ser Ser Pro Ile Ile Met Asp Ala Thr
165 170 175

Val Asp Gln Gln Asn Gly Tyr Arg Phe Val Tyr Thr Leu Pro Leu Ser
180 185 190

Ala Thr Ala Leu Leu Ile Glu Asp Thr His Tyr Ile Asp Lys Ala Asn
195 200 205

Leu Gln Ala Glu Arg Ala Arg Gln Asn Ile Arg Asp Tyr Ala Ala Arg
210 215 220

Gln Gly Trp Pro Leu Gln Thr Leu Leu Arg Glu Glu Gln Gly Ala Leu
225 230 235 240

Pro Ile Thr Leu Thr Gly Asp Asn Arg Gln Phe Trp Gln Gln Gln Pro
 245 250 255

Gln Ala Cys Ser Gly Leu Arg Ala Gly Leu Phe His Pro Thr Thr Gly
 260 265 270

Tyr Ser Leu Pro Leu Ala Val Ala Leu Ala Asp Arg Leu Ser Ala Leu
 275 280 285

Asp Val Phe Thr Ser Ser Ser Val His Gln Thr Ile Ala His Phe Ala
 290 295 300

Gln Gln Arg Trp Gln Gln Gln Gly Phe Phe Arg Met Leu Asn Arg Met
 305 310 315 320

Leu Phe Leu Ala Gly Pro Ala Glu Ser Arg Trp Arg Val Met Gln Arg
 325 330 335

Phe Tyr Gly Leu Pro Glu Asp Leu Ile Ala Arg Phe Tyr Ala Gly Lys
 340 345 350

Leu Thr Val Thr Asp Arg Leu Arg Ile Leu Ser Gly Lys Pro Pro Val
 355 360 365

Pro Val Phe Ala Ala Leu Gln Ala Ile Met Thr Thr His Arg
 370 375 380

<210> 7
 <211> 1479
 <212> DNA
 <213> Pantoea stewartii

<220>
 <221> CDS
 <222> (1)..(1479)

<400> 7
 atg aaa cca act acg gta att ggt gcg ggc ttt ggt ggc ctg gca ctg 48
 Met Lys Pro Thr Thr Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
 1 5 10 15

gca att cgt tta cag gcc gca ggt att cct gtt ttg ctg ctt gag cag 96
 Ala Ile Arg Leu Gln Ala Ala Gly Ile Pro Val Leu Leu Leu Glu Gln
 20 25 30

cgc gac aag ccg ggt ggc cgg gct tat gtt tat cag gag cag ggc ttt 144
 Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Tyr Gln Glu Gln Gly Phe
 35 40 45

act ttt gat gca ggc cct acc gtt atc acc gat ccc agc gcg att gaa 192
 Thr Phe Asp Ala Gly Pro Thr Val Ile Thr Asp Pro Ser Ala Ile Glu
 50 55 60

gaa ctg ttt gct ctg gcc ggt aaa cag ctt aag gat tac gtc gag ctg 240
 Glu Leu Phe Ala Leu Ala Gly Lys Gln Leu Lys Asp Tyr Val Glu Leu
 65 70 75 80

ttg	ccg	gtc	acg	ccg	ttt	tat	cgc	ctg	tgc	tgg	gag	tcc	ggc	aag	gtc	288
Leu	Pro	Val	Thr	Pro	Phe	Tyr	Arg	Leu	Cys	Trp	Glu	Ser	Gly	Lys	Val	
				85					90					95		
ttc	aat	tac	gat	aac	gac	cag	gcc	cag	tta	gaa	gcg	cag	ata	cag	cag	336
Phe	Asn	Tyr	Asp	Asn	Asp	Gln	Ala	Gln	Leu	Glu	Ala	Gln	Ile	Gln	Gln	
			100					105					110			
ttt	aat	ccg	cgc	gat	gtt	gcg	ggg	tat	cga	gcg	ttc	ctt	gac	tat	tcg	384
Phe	Asn	Pro	Arg	Asp	Val	Ala	Gly	Tyr	Arg	Ala	Phe	Leu	Asp	Tyr	Ser	
		115					120					125				
cgt	gcc	gta	ttc	aat	gag	ggc	tat	ctg	aag	ctc	ggc	act	gtg	cct	ttt	432
Arg	Ala	Val	Phe	Asn	Glu	Gly	Tyr	Leu	Lys	Leu	Gly	Thr	Val	Pro	Phe	
	130					135					140					
tta	tcg	ttc	aaa	gac	atg	ctt	cgg	gcc	gcg	ccc	cag	ttg	gca	aag	ctg	480
Leu	Ser	Phe	Lys	Asp	Met	Leu	Arg	Ala	Ala	Pro	Gln	Leu	Ala	Lys	Leu	
145					150					155					160	
cag	gca	tgg	cgc	agc	gtt	tac	agt	aaa	gtt	gcc	ggc	tac	att	gag	gat	528
Gln	Ala	Trp	Arg	Ser	Val	Tyr	Ser	Lys	Val	Ala	Gly	Tyr	Ile	Glu	Asp	
				165					170					175		
gag	cat	ctt	cgg	cag	gcg	ttt	tct	ttt	cac	tcg	ctc	tta	gtg	ggg	ggg	576
Glu	His	Leu	Arg	Gln	Ala	Phe	Ser	Phe	His	Ser	Leu	Leu	Val	Gly	Gly	
			180					185					190			
aat	ccg	ttt	gca	acc	tcg	tcc	att	tat	acg	ctg	att	cac	gcg	tta	gaa	624
Asn	Pro	Phe	Ala	Thr	Ser	Ser	Ile	Tyr	Thr	Leu	Ile	His	Ala	Leu	Glu	
		195					200					205				
cgg	gaa	tgg	ggc	gtc	tgg	ttt	cca	cgc	ggg	gga	acc	ggg	gcg	ctg	gtc	672
Arg	Glu	Trp	Gly	Val	Trp	Phe	Pro	Arg	Gly	Gly	Thr	Gly	Ala	Leu	Val	
	210					215					220					
aat	ggc	atg	atc	aag	ctg	ttt	cag	gat	ctg	ggc	ggc	gaa	gtc	gtg	ctt	720
Asn	Gly	Met	Ile	Lys	Leu	Phe	Gln	Asp	Leu	Gly	Gly	Glu	Val	Val	Leu	
225					230					235					240	
aac	gcc	cgg	gtc	agt	cat	atg	gaa	acc	gtt	ggg	gac	aag	att	cag	gcc	768
Asn	Ala	Arg	Val	Ser	His	Met	Glu	Thr	Val	Gly	Asp	Lys	Ile	Gln	Ala	
				245					250					255		
gtg	cag	ttg	gaa	gac	ggc	aga	cgg	ttt	gaa	acc	tgc	gcg	gtg	gcg	tcg	816
Val	Gln	Leu	Glu	Asp	Gly	Arg	Arg	Phe	Glu	Thr	Cys	Ala	Val	Ala	Ser	
			260					265					270			
aac	gct	gat	gtt	gta	cat	acc	tat	cgc	gat	ctg	ctg	tct	cag	cat	ccc	864
Asn	Ala	Asp	Val	Val	His	Thr	Tyr	Arg	Asp	Leu	Leu	Ser	Gln	His	Pro	
		275					280					285				
gca	gcc	gct	aag	cag	gcg	aaa	aaa	ctg	caa	tcc	aag	cgt	atg	agt	aac	912
Ala	Ala	Ala	Lys	Gln	Ala	Lys	Lys	Leu	Gln	Ser	Lys	Arg	Met	Ser	Asn	
	290					295					300					
tca	ctg	ttt	gta	ctc	tat	ttt	ggg	ctc	aac	cat	cat	cac	gat	caa	ctc	960
Ser	Leu	Phe	Val	Leu	Tyr	Phe	Gly	Leu	Asn	His	His	His	Asp	Gln	Leu	
305					310					315					320	
gcc	cat	cat	acc	gtc	tgt	ttt	ggg	cca	cgc	tac	cgt	gaa	ctg	att	cac	1008
Ala	His	His	Thr	Val	Cys	Phe	Gly	Pro	Arg	Tyr	Arg	Glu	Leu	Ile	His	
				325				330						335		
gaa	att	ttt	aac	cat	gat	ggg	ctg	gct	gag	gat	ttt	tcg	ctt	tat	tta	1056
Glu	Ile	Phe	Asn	His	Asp	Gly	Leu	Ala	Glu	Asp	Phe	Ser	Leu	Tyr	Leu	
			340					345					350			
cac	gca	cct	tgt	gtc	acg	gat	ccg	tca	ctg	gca	ccg	gaa	ggg	tgc	ggc	1104

His Ala Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Glu Gly Cys Gly
 355 360 365
 agc tat tat gtg ctg gcg cct gtt cca cac tta ggc acg gcg aac ctc 1152
 Ser Tyr Tyr Val Leu Ala Pro Val Pro His Leu Gly Thr Ala Asn Leu
 370 375 380
 gac tgg gcg gta gaa gga ccc cga ctg cgc gat cgt att ttt gac tac 1200
 Asp Trp Ala Val Glu Gly Pro Arg Leu Arg Asp Arg Ile Phe Asp Tyr
 385 390 395 400
 ctt gag caa cat tac atg cct ggc ttg cga agc cag ttg gtg acg cac 1248
 Leu Glu Gln His Tyr Met Pro Gly Leu Arg Ser Gln Leu Val Thr His
 405 410 415
 cgt atg ttt acg ccg ttc gat ttc cgc gac gag ctc aat gcc tgg caa 1296
 Arg Met Phe Thr Pro Phe Asp Phe Arg Asp Glu Leu Asn Ala Trp Gln
 420 425 430
 ggt tcg gcc ttc tcg gtt gaa cct att ctg acc cag agc gcc tgg ttc 1344
 Gly Ser Ala Phe Ser Val Glu Pro Ile Leu Thr Gln Ser Ala Trp Phe
 435 440 445
 cga cca cat aac cgc gat aag cac att gat aat ctt tat ctg gtt ggc 1392
 Arg Pro His Asn Arg Asp Lys His Ile Asp Asn Leu Tyr Leu Val Gly
 450 455 460
 gca ggc acc cat cct ggc gcg ggc att ccc ggc gta atc ggc tcg gcg 1440
 Ala Gly Thr His Pro Gly Ala Gly Ile Pro Gly Val Ile Gly Ser Ala
 465 470 475 480
 aag gcg acg gca ggc tta atg ctg gag gac ctg att tga 1479
 Lys Ala Thr Ala Gly Leu Met Leu Glu Asp Leu Ile
 485 490

<210> 8
 <211> 492
 <212> PRT
 <213> Pantoea stewartii

<400> 8

Met Lys Pro Thr Thr Val Ile Gly Ala Gly Phe Gly Gly Leu Ala Leu
 1 5 10 15
 Ala Ile Arg Leu Gln Ala Ala Gly Ile Pro Val Leu Leu Leu Glu Gln
 20 25 30
 Arg Asp Lys Pro Gly Gly Arg Ala Tyr Val Tyr Gln Glu Gln Gly Phe
 35 40 45
 Thr Phe Asp Ala Gly Pro Thr Val Ile Thr Asp Pro Ser Ala Ile Glu
 50 55 60
 Glu Leu Phe Ala Leu Ala Gly Lys Gln Leu Lys Asp Tyr Val Glu Leu
 65 70 75 80
 Leu Pro Val Thr Pro Phe Tyr Arg Leu Cys Trp Glu Ser Gly Lys Val
 85 90 95
 Phe Asn Tyr Asp Asn Asp Gln Ala Gln Leu Glu Ala Gln Ile Gln Gln
 100 105 110

Phe Asn Pro Arg Asp Val Ala Gly Tyr Arg Ala Phe Leu Asp Tyr Ser
 115 120 125
 Arg Ala Val Phe Asn Glu Gly Tyr Leu Lys Leu Gly Thr Val Pro Phe
 130 135 140
 Leu Ser Phe Lys Asp Met Leu Arg Ala Ala Pro Gln Leu Ala Lys Leu
 145 150 155 160
 Gln Ala Trp Arg Ser Val Tyr Ser Lys Val Ala Gly Tyr Ile Glu Asp
 165 170 175
 Glu His Leu Arg Gln Ala Phe Ser Phe His Ser Leu Leu Val Gly Gly
 180 185 190
 Asn Pro Phe Ala Thr Ser Ser Ile Tyr Thr Leu Ile His Ala Leu Glu
 195 200 205
 Arg Glu Trp Gly Val Trp Phe Pro Arg Gly Gly Thr Gly Ala Leu Val
 210 215 220
 Asn Gly Met Ile Lys Leu Phe Gln Asp Leu Gly Gly Glu Val Val Leu
 225 230 235 240
 Asn Ala Arg Val Ser His Met Glu Thr Val Gly Asp Lys Ile Gln Ala
 245 250 255
 Val Gln Leu Glu Asp Gly Arg Arg Phe Glu Thr Cys Ala Val Ala Ser
 260 265 270
 Asn Ala Asp Val Val His Thr Tyr Arg Asp Leu Leu Ser Gln His Pro
 275 280 285
 Ala Ala Ala Lys Gln Ala Lys Lys Leu Gln Ser Lys Arg Met Ser Asn
 290 295 300
 Ser Leu Phe Val Leu Tyr Phe Gly Leu Asn His His His Asp Gln Leu
 305 310 315 320
 Ala His His Thr Val Cys Phe Gly Pro Arg Tyr Arg Glu Leu Ile His
 325 330 335
 Glu Ile Phe Asn His Asp Gly Leu Ala Glu Asp Phe Ser Leu Tyr Leu
 340 345 350
 His Ala Pro Cys Val Thr Asp Pro Ser Leu Ala Pro Glu Gly Cys Gly
 355 360 365
 Ser Tyr Tyr Val Leu Ala Pro Val Pro His Leu Gly Thr Ala Asn Leu
 370 375 380

Asp Trp Ala Val Glu Gly Pro Arg Leu Arg Asp Arg Ile Phe Asp Tyr
385 390 395 400

Leu Glu Gln His Tyr Met Pro Gly Leu Arg Ser Gln Leu Val Thr His
405 410 415

Arg Met Phe Thr Pro Phe Asp Phe Arg Asp Glu Leu Asn Ala Trp Gln
420 425 430

Gly Ser Ala Phe Ser Val Glu Pro Ile Leu Thr Gln Ser Ala Trp Phe
435 440 445

Arg Pro His Asn Arg Asp Lys His Ile Asp Asn Leu Tyr Leu Val Gly
450 455 460

Ala Gly Thr His Pro Gly Ala Gly Ile Pro Gly Val Ile Gly Ser Ala
465 470 475 480

Lys Ala Thr Ala Gly Leu Met Leu Glu Asp Leu Ile
485 490

<210> 9
<211> 891
<212> DNA
<213> *Pantoea stewartii*

<220>
<221> CDS
<222> (1)..(891)

<400> 9
atg gcg gtt ggc tcg aaa agc ttt gcg act gca tcg acg ctt ttc gac 48
Met Ala Val Gly Ser Lys Ser Phe Ala Thr Ala Ser Thr Leu Phe Asp
1 5 10 15
gcc aaa acc cgt cgc agc gtg ctg atg ctt tac gca tgg tgc cgc cac 96
Ala Lys Thr Arg Arg Ser Val Leu Met Leu Tyr Ala Trp Cys Arg His
20 25 30
tgc gac gac gtc att gac gat caa aca ctg ggc ttt cat gcc gac cag 144
Cys Asp Asp Val Ile Asp Asp Gln Thr Leu Gly Phe His Ala Asp Gln
35 40 45
ccc tct tcg cag atg cct gag cag cgc ctg cag cag ctt gaa atg aaa 192
Pro Ser Ser Gln Met Pro Glu Gln Arg Leu Gln Gln Leu Glu Met Lys
50 55 60
acg cgt cag gcc tac gcc ggt tcg caa atg cac gag ccc gct ttt gcc 240
Thr Arg Gln Ala Tyr Ala Gly Ser Gln Met His Glu Pro Ala Phe Ala
65 70 75 80
gcg ttt cag gag gtc gcg atg gcg cat gat atc gct ccc gcc tac gcg 288
Ala Phe Gln Glu Val Ala Met Ala His Asp Ile Ala Pro Ala Tyr Ala
85 90 95
ttc gac cat ctg gaa ggt ttt gcc atg gat gtg cgc gaa acg cgc tac 336
Phe Asp His Leu Glu Gly Phe Ala Met Asp Val Arg Glu Thr Arg Tyr
100 105 110
ctg aca ctg gac gat acg ctg cgt tat tgc tat cac gtc gcc ggt gtt 384
Page 13

Leu Thr Leu Asp Asp Thr Leu Arg Tyr Cys Tyr His Val Ala Gly Val
 115 120 125
 gtg ggc ctg atg atg gcg caa att atg ggc gtt cgc gat aac gcc acg 432
 Val Gly 130 Leu Met Met Ala Gln Ile Met Gly Val Arg Asp Asn Ala Thr
 140
 ctc gat cgc gcc tgc gat ctc ggg ctg gct ttc cag ttg acc aac att 480
 Leu Asp Arg Ala Cys Asp Leu Gly Leu Ala Phe Gln Leu Thr Asn Ile
 145 150 155 160
 gcg cgt gat att gtc gac gat gct cag gtg ggc cgc tgt tat ctg cct 528
 Ala Arg Asp Ile Val Asp Asp Ala Gln Val Gly Arg Cys Tyr Leu Pro
 165 170 175
 gaa agc tgg ctg gaa gag gaa gga ctg acg aaa gcg aat tat gct gcg 576
 Glu Ser Trp Leu Glu Glu Glu Gly Leu Thr Lys Ala Asn Tyr Ala Ala
 180 185 190
 cca gaa aac cgg cag gcc tta agc cgt atc gcc ggg cga ctg gta cgg 624
 Pro Glu Asn Arg Gln Ala Leu Ser Arg Ile Ala Gly Arg Leu Val Arg
 195 200 205
 gaa gcg gaa ccc tat tac gta tca tca atg gcc ggt ctg gca caa tta 672
 Glu Ala Glu Pro Tyr Tyr Val Ser Ser Met Ala Gly Leu Ala Gln Leu
 210 215 220
 ccc tta cgc tcg gcc tgg gcc atc gcg aca gcg aag cag gtg tac cgt 720
 Pro Leu Arg Ser Ala Trp Ala Ile Ala Thr Ala Lys Gln Val Tyr Arg
 225 230 235 240
 aaa att ggc gtg aaa gtt gaa cag gcc ggt aag cag gcc tgg gat cat 768
 Lys Ile Gly Val Lys Val Glu Gln Ala Gly Lys Gln Ala Trp Asp His
 245 250 255
 cgc cag tcc acg tcc acc gcc gaa aaa tta acg ctt ttg ctg acg gca 816
 Arg Gln Ser Thr Ser Thr Ala Glu Lys Leu Thr Leu Leu Thr Ala
 260 265 270
 tcc ggt cag gca gtt act tcc cgg atg aag acg tat cca ccc cgt cct 864
 Ser Gly Gln Ala Val Thr Ser Arg Met Lys Thr Tyr Pro Pro Arg Pro
 275 280 285
 gct cat ctc tgg cag cgc ccg atc tag 891
 Ala His Leu Trp Gln Arg Pro Ile
 290 295

<210> 10

<211> 296

<212> PRT

<213> Pantoea stewartii

<400> 10

Met Ala Val Gly Ser Lys Ser Phe Ala Thr Ala Ser Thr Leu Phe Asp
 1 5 10 15

Ala Lys Thr Arg Arg Ser Val Leu Met Leu Tyr Ala Trp Cys Arg His
 20 25 30

Cys Asp Asp Val Ile Asp Asp Gln Thr Leu Gly Phe His Ala Asp Gln
 35 40 45

Pro Ser Ser Gln Met Pro Glu Gln Arg Leu Gln Gln Leu Glu Met Lys
 50 55 60

Thr Arg Gln Ala Tyr Ala Gly Ser Gln Met His Glu Pro Ala Phe Ala
 65 70 75 80
 Ala Phe Gln Glu Val Ala Met Ala His Asp Ile Ala Pro Ala Tyr Ala
 85 90 95
 Phe Asp His Leu Glu Gly Phe Ala Met Asp Val Arg Glu Thr Arg Tyr
 100 105 110
 Leu Thr Leu Asp Asp Thr Leu Arg Tyr Cys Tyr His Val Ala Gly Val
 115 120 125
 Val Gly Leu Met Met Ala Gln Ile Met Gly Val Arg Asp Asn Ala Thr
 130 135 140
 Leu Asp Arg Ala Cys Asp Leu Gly Leu Ala Phe Gln Leu Thr Asn Ile
 145 150 155 160
 Ala Arg Asp Ile Val Asp Asp Ala Gln Val Gly Arg Cys Tyr Leu Pro
 165 170 175
 Glu Ser Trp Leu Glu Glu Glu Gly Leu Thr Lys Ala Asn Tyr Ala Ala
 180 185 190
 Pro Glu Asn Arg Gln Ala Leu Ser Arg Ile Ala Gly Arg Leu Val Arg
 195 200 205
 Glu Ala Glu Pro Tyr Tyr Val Ser Ser Met Ala Gly Leu Ala Gln Leu
 210 215 220
 Pro Leu Arg Ser Ala Trp Ala Ile Ala Thr Ala Lys Gln Val Tyr Arg
 225 230 235 240
 Lys Ile Gly Val Lys Val Glu Gln Ala Gly Lys Gln Ala Trp Asp His
 245 250 255
 Arg Gln Ser Thr Ser Thr Ala Glu Lys Leu Thr Leu Leu Leu Thr Ala
 260 265 270
 Ser Gly Gln Ala Val Thr Ser Arg Met Lys Thr Tyr Pro Pro Arg Pro
 275 280 285
 Ala His Leu Trp Gln Arg Pro Ile
 290 295

<210> 11
 <211> 528
 <212> DNA
 <213> Pantoea stewartii

<220>

<221> CDS
<222> (1)..(528)

<400> 11
 atg ttg tgg att tgg aat gcc ctg atc gtg ttt gtc acc gtg gtc ggc 48
 Met Leu Trp Ile Trp Asn Ala Leu Ile Val Phe Val Thr Val Val Gly
 1 5 10 15
 atg gaa gtg gtt gct gca ctg gca cat aaa tac atc atg cac ggc tgg 96
 Met Glu Val Val Ala Ala Leu Ala His Lys Tyr Ile Met His Gly Trp
 20 25 30
 ggt tgg ggc tgg cat ctt tca cat cat gaa ccg cgt aaa ggc gca ttt 144
 Gly Trp Gly Trp His Leu Ser His His Glu Pro Arg Lys Gly Ala Phe
 35 40 45
 gaa gtt aac gat ctc tat gcc gtg gta ttc gcc att gtg tcg att gcc 192
 Glu Val Asn Asp Leu Tyr Ala Val Val Phe Ala Ile Val Ser Ile Ala
 50 55 60
 ctg att tac ttc ggc agt aca gga atc tgg ccg ctc cag tgg att ggt 240
 Leu Ile Tyr Phe Gly Ser Thr Gly Ile Trp Pro Leu Gln Trp Ile Gly
 65 70 75 80
 gca ggc atg acc gct tat ggt tta ctg tat ttt atg gtc cac gac gga 288
 Ala Gly Met Thr Ala Tyr Gly Leu Leu Tyr Phe Met Val His Asp Gly
 85 90 95
 ctg gta cac cag cgc tgg ccg ttc cgc tac ata ccg cgc aaa ggc tac 336
 Leu Val His Gln Arg Trp Pro Phe Arg Tyr Ile Pro Arg Lys Gly Tyr
 100 105 110
 ctg aaa cgg tta tac atg gcc cac cgt atg cat cat gct gta agg gga 384
 Leu Lys Arg Leu Tyr Met Ala His Arg Met His His Ala Val Arg Gly
 115 120 125
 aaa gag ggc tgc gtg tcc ttt ggt ttt ctg tac gcg cca ccg tta tct 432
 Lys Glu Gly Cys Val Ser Phe Gly Phe Leu Tyr Ala Pro Pro Leu Ser
 130 135 140
 aaa ctt cag gcg acg ctg aga gaa agg cat gcg gct aga tcg ggc gct 480
 Lys Leu Gln Ala Thr Leu Arg Glu Arg His Ala Ala Arg Ser Gly Ala
 145 150 155 160
 gcc aga gat gag cag gac ggg gtg gat acg tct tca tcc ggg aag taa 528
 Ala Arg Asp Glu Gln Asp Gly Val Asp Thr Ser Ser Ser Gly Lys
 165 170 175

<210> 12
 <211> 175
 <212> PRT
 <213> Pantoea stewartii

<400> 12
 Met Leu Trp Ile Trp Asn Ala Leu Ile Val Phe Val Thr Val Val Gly
 1 5 10 15
 Met Glu Val Val Ala Ala Leu Ala His Lys Tyr Ile Met His Gly Trp
 20 25 30
 Gly Trp Gly Trp His Leu Ser His His Glu Pro Arg Lys Gly Ala Phe
 35 40 45
 Glu Val Asn Asp Leu Tyr Ala Val Val Phe Ala Ile Val Ser Ile Ala
 Page 16

50

55

60

Leu Ile Tyr Phe Gly Ser Thr Gly Ile Trp Pro Leu Gln Trp Ile Gly
65 70 75 80

Ala Gly Met Thr Ala Tyr Gly Leu Leu Tyr Phe Met Val His Asp Gly
85 90 95

Leu Val His Gln Arg Trp Pro Phe Arg Tyr Ile Pro Arg Lys Gly Tyr
100 105 110

Leu Lys Arg Leu Tyr Met Ala His Arg Met His His Ala Val Arg Gly
115 120 125

Lys Glu Gly Cys Val Ser Phe Gly Phe Leu Tyr Ala Pro Pro Leu Ser
130 135 140

Lys Leu Gln Ala Thr Leu Arg Glu Arg His Ala Ala Arg Ser Gly Ala
145 150 155 160

Ala Arg Asp Glu Gln Asp Gly Val Asp Thr Ser Ser Ser Gly Lys
165 170 175

<210> 13
<211> 1860
<212> DNA
<213> Methylobionas 16a

<220>
<221> CDS
<222> (1)..(1860)

<400> 13
atg gct ctt tcc aaa gac ttc cct cta ctc aat tcc atc cac acc cca 48
Met Ala Leu Ser Lys Asp Phe Pro Leu Leu Asn Ser Ile His Thr Pro
1 5 10 15
gcg gac ata cgc gcg ctg tcc aag gac cag ctc cag caa ctg gct gac 96
Ala Asp Ile Arg Ala Leu Ser Lys Asp Gln Leu Gln Gln Leu Ala Asp
20 25 30
gag gtg cgc ggc tat ctg acc cac acg gtc agc att tcc ggc ggc cat 144
Glu Val Arg Gly Tyr Leu Thr His Thr Val Ser Ile Ser Gly Gly His
35 40 45
ttt gcg gcc ggc ctc ggc acc gtg gaa ctg acc gtg gcc ttg cat tat 192
Phe Ala Ala Gly Leu Gly Thr Val Glu Leu Thr Val Ala Leu His Tyr
50 55 60
gtg ttc aat acc ccc gtc gat cag ttg gtc tgg gac gtg ggc cat cag 240
Val Phe Asn Thr Pro Val Asp Gln Leu Val Trp Asp Val Gly His Gln
65 70 75 80
gcc tat ccg cac aag att ctg acc ggt cgc aag gag cgc atg ccg acc 288
Ala Tyr Pro His Lys Ile Leu Thr Gly Arg Lys Glu Arg Met Pro Thr
85 90 95
att cgc acc ctg ggc ggc gtg tca gcc ttt ccg gcg cgg gac gag agc 336
Ile Arg Thr Leu Gly Gly Val Ser Ala Phe Pro Ala Arg Asp Glu Ser
100 105 110

gaa tac gat gcc ttc ggc gtc ggc cat tcc agc acc tcg atc agc gcg Glu Tyr Asp Ala Phe Gly Val Gly His Ser Ser Thr Ser Ile Ser Ala	384
gca ctg ggc atg gcc att gcg tcg cag ctg cgc ggc gaa gac aag aag Ala Leu Gly Met Ala Ile Ala Ser Gln Leu Arg Gly Glu Asp Lys Lys	432
atg gta gcc atc atc ggc gac ggt tcc atc acc ggc ggc atg gcc tat Met Val Ala Ile Ile Gly Asp Gly Ser Ile Thr Thr Gly Gly Met Ala Tyr	480
gag gcg atg aat cat gcc ggc gat gtg aat gcc aac ctg ctg gtg atc Glu Ala Met Asn His Ala Gly Asp Val Asn Ala Asn Leu Leu Val Ile	528
ttg aac gac aac gat atg tcg atc tcg ccg ccg gtc ggc gcg atg aac Leu Asn Asp Asn Asp Met Ser Ile Ser Pro Pro Val Gly Ala Met Asn	576
aat tat ctg acc aag gtg ttg tcg agc aag ttt tat tcg tcg gtg cgg Asn Tyr Leu Thr Lys Val Leu Ser Ser Lys Phe Tyr Ser Ser Val Arg	624
gaa gag agc aag aaa gct ctg gcc aag atg ccg tcg gtg tgg gaa ctg Glu Glu Ser Lys Lys Ala Leu Ala Lys Met Pro Ser Val Trp Glu Leu	672
gcg cgc aag acc gag gaa cac gtg aag ggc atg atc gtg ccc ggt acc Ala Arg Lys Thr Glu Glu His Val Lys Gly Met Ile Val Pro Gly Thr	720
ttg ttc gag gaa ttg ggc ttc aat tat ttc ggc ccg atc gac ggc cat Leu Phe Glu Glu Leu Gly Phe Asn Tyr Phe Gly Pro Ile Asp Gly His	768
gat gtc gag atg ctg gtg tcg acc ctg gaa aat ctg aag gat ttg acc Asp Val Glu Met Leu Val Ser Thr Leu Glu Asn Leu Lys Asp Leu Thr	816
ggg ccg gta ttc ctg cat gtg gtg acc aag aag ggc aaa ggc tat gcg Gly Pro Val Phe Leu His Val Val Thr Lys Lys Gly Lys Gly Tyr Ala	864
cca gcc gag aaa gac ccg ttg gcc tac cat ggc gtg ccg gct ttc gat Pro Ala Glu Lys Asp Pro Leu Ala Tyr His Gly Val Pro Ala Phe Asp	912
ccg acc aag gat ttc ctg ccc aag gcg gcg ccg tcg ccg cat ccg acc Pro Thr Lys Asp Phe Leu Pro Lys Ala Ala Pro Ser Pro His Pro Thr	960
tat acc gag gtg ttc ggc cgc tgg ctg tgc gac atg gcg gct caa gac Tyr Thr Glu Val Phe Gly Arg Trp Leu Cys Asp Met Ala Ala Gln Asp	1008
gag cgc ttg ctg ggc atc acg ccg gcg atg cgc gaa ggc tct ggt ttg Glu Arg Leu Leu Gly Ile Thr Pro Ala Met Arg Glu Gly Ser Gly Leu	1056
gtg gaa ttc tca cag aaa ttt ccg aat cgc tat ttc gat gtc gcc atc Val Glu Phe Ser Gln Lys Phe Pro Asn Arg Tyr Phe Asp Val Ala Ile	1104
gcc gag cag cat gcg gtg acc ttg gcc gcc ggc cag gcc tgc cag ggc Ala Glu Gln His Ala Val Thr Leu Ala Ala Gly Gln Ala Cys Gln Gly	1152

gcc aag ccg gtg gtg gcg att tat tcc acc ttc ctg caa cgc ggt tac Ala Lys Pro Val Val Ala Ile Tyr Ser Thr Phe Leu Gln Arg Gly Tyr 385 390 395 400	1200
gat cag ttg atc cac gac gtg gcc ttg cag aac tta gat atg ctc ttt Asp Gln Leu Ile His Asp Val Ala Leu Gln Asn Leu Asp Met Leu Phe 405 410 415	1248
gca ctg gat cgt gcc ggc ttg gtc ggc ccg gat gga ccg acc cat gct Ala Leu Asp Arg Ala Gly Leu Val Gly Pro Asp Gly Pro Thr His Ala 420 425 430	1296
ggc gcc ttt gat tac agc tac atg cgc tgt att ccg aac atg ctg atc Gly Ala Phe Asp Tyr Ser Tyr Met Arg Cys Ile Pro Asn Met Leu Ile 435 440 445	1344
atg gct cca gcc gac gag aac gag tgc agg cag atg ctg acc acc ggc Met Ala Pro Ala Asp Glu Asn Glu Cys Arg Gln Met Leu Thr Thr Gly 450 455 460	1392
ttc caa cac cat ggc ccg gct tcg gtg cgc tat ccg cgc ggc aaa ggg Phe Gln His His Gly Pro Ala Ser Val Arg Tyr Pro Arg Gly Lys Gly 465 470 475 480	1440
ccc ggg gcg gca atc gat ccg acc ctg acc gcg ctg gag atc ggc aag Pro Gly Ala Ala Ile Asp Pro Thr Leu Thr Ala Leu Glu Ile Gly Lys 485 490 495	1488
gcc gaa gtc aga cac cac ggc agc cgc atc gcc att ctg gcc tgg ggc Ala Glu Val Arg His His Gly Ser Arg Ile Ala Ile Leu Ala Trp Gly 500 505 510	1536
agc atg gtc acg cct gcc gtc gaa gcc ggc aag cag ctg ggc gcg acg Ser Met Val Thr Pro Ala Val Glu Ala Gly Lys Gln Leu Gly Ala Thr 515 520 525	1584
gtg gtg aac atg cgt ttc gtc aag ccg ttc gat caa gcc ttg gtg ctg Val Val Asn Met Arg Phe Val Lys Pro Phe Asp Gln Ala Leu Val Leu 530 535 540	1632
gaa ttg gcc agg acg cac gat gtg ttc gtc acc gtc gag gaa aac gtc Glu Leu Ala Arg Thr His Asp Val Phe Val Thr Val Glu Glu Asn Val 545 550 555 560	1680
atc gcc ggc ggc gct ggc agt gcg atc aac acc ttc ctg cag gcg cag Ile Ala Gly Gly Ala Gly Ser Ala Ile Asn Thr Phe Leu Gln Ala Gln 565 570 575	1728
aag gtg ctg atg ccg gtc tgc aac atc ggc ctg ccc gac cgc ttc gtc Lys Val Leu Met Pro Val Cys Asn Ile Gly Leu Pro Asp Arg Phe Val 580 585 590	1776
gag caa ggt agt cgc gag gaa ttg ctc agc ctg gtc ggc ctc gac agc Glu Gln Gly Ser Arg Glu Glu Leu Leu Ser Leu Val Gly Leu Asp Ser 595 600 605	1824
aag ggc atc ttc gcc acc atc gaa cag ttt tgc gct Lys Gly Ile Phe Ala Thr Ile Glu Gln Phe Cys Ala 610 615 620	1860

<210> 14

<211> 620

<212> PRT

<213> Methylobionas 16a

<400> 14

Met Ala Leu Ser Lys Asp Phe Pro Leu Leu Asn Ser Ile His Thr Pro
Page 19

1	5	10	15
Ala Asp Ile Arg	Ala Leu Ser Lys	Asp Gln Leu Gln Gln	Leu Ala Asp
	20	25	30
Glu Val Arg	Gly Tyr Leu Thr	His Thr Val Ser	Ile Ser Gly Gly His
	35	40	45
Phe Ala Ala Gly	Leu Gly Thr Val	Glu Leu Thr Val	Ala Leu His Tyr
	50	55	60
Val Phe Asn Thr	Pro Val Asp Gln	Leu Val Trp Asp	Val Gly His Gln
	65	70	75
Ala Tyr Pro His	Lys Ile Leu Thr	Gly Arg Lys Glu	Arg Met Pro Thr
	85	90	95
Ile Arg Thr	Leu Gly Gly Val	Ser Ala Phe Pro	Ala Arg Asp Glu Ser
	100	105	110
Glu Tyr Asp	Ala Phe Gly Val	Gly His Ser Ser	Thr Ser Ile Ser Ala
	115	120	125
Ala Leu Gly Met	Ala Ile Ala Ser	Gln Leu Arg Gly	Glu Asp Lys Lys
	130	135	140
Met Val Ala Ile	Ile Gly Asp Gly	Ser Ile Thr Gly	Gly Met Ala Tyr
	145	150	155
Glu Ala Met Asn	His Ala Gly Asp	Val Asn Ala Asn	Leu Leu Val Ile
	165	170	175
Leu Asn Asp	Asn Asp Met Ser	Ile Ser Pro Pro	Val Gly Ala Met Asn
	180	185	190
Asn Tyr Leu Thr	Lys Val Leu Ser	Ser Lys Phe Tyr	Ser Ser Val Arg
	195	200	205
Glu Glu Ser Lys	Lys Ala Leu Ala	Lys Met Pro Ser	Val Trp Glu Leu
	210	215	220
Ala Arg Lys Thr	Glu Glu His Val	Lys Gly Met Ile	Val Pro Gly Thr
	225	230	235
Leu Phe Glu Glu	Leu Gly Phe Asn	Tyr Phe Gly Pro	Ile Asp Gly His
	245	250	255
Asp Val Glu Met	Leu Val Ser Thr	Leu Glu Asn Leu	Lys Asp Leu Thr
	260	265	270
Gly Pro Val Phe	Leu His Val Val	Thr Lys Lys Gly	Lys Gly Tyr Ala
	275	280	285

Pro Ala Glu Lys Asp Pro Leu Ala Tyr His Gly Val Pro Ala Phe Asp
 290 295 300
 Pro Thr Lys Asp Phe Leu Pro Lys Ala Ala Pro Ser Pro His Pro Thr
 305 310 315 320
 Tyr Thr Glu Val Phe Gly Arg Trp Leu Cys Asp Met Ala Ala Gln Asp
 325 330 335
 Glu Arg Leu Leu Gly Ile Thr Pro Ala Met Arg Glu Gly Ser Gly Leu
 340 345 350
 Val Glu Phe Ser Gln Lys Phe Pro Asn Arg Tyr Phe Asp Val Ala Ile
 355 360 365
 Ala Glu Gln His Ala Val Thr Leu Ala Ala Gly Gln Ala Cys Gln Gly
 370 375 380
 Ala Lys Pro Val Val Ala Ile Tyr Ser Thr Phe Leu Gln Arg Gly Tyr
 385 390 395 400
 Asp Gln Leu Ile His Asp Val Ala Leu Gln Asn Leu Asp Met Leu Phe
 405 410 415
 Ala Leu Asp Arg Ala Gly Leu Val Gly Pro Asp Gly Pro Thr His Ala
 420 425 430
 Gly Ala Phe Asp Tyr Ser Tyr Met Arg Cys Ile Pro Asn Met Leu Ile
 435 440 445
 Met Ala Pro Ala Asp Glu Asn Glu Cys Arg Gln Met Leu Thr Thr Gly
 450 455 460
 Phe Gln His His Gly Pro Ala Ser Val Arg Tyr Pro Arg Gly Lys Gly
 465 470 475 480
 Pro Gly Ala Ala Ile Asp Pro Thr Leu Thr Ala Leu Glu Ile Gly Lys
 485 490 495
 Ala Glu Val Arg His His Gly Ser Arg Ile Ala Ile Leu Ala Trp Gly
 500 505 510
 Ser Met Val Thr Pro Ala Val Glu Ala Gly Lys Gln Leu Gly Ala Thr
 515 520 525
 Val Val Asn Met Arg Phe Val Lys Pro Phe Asp Gln Ala Leu Val Leu
 530 535 540
 Glu Leu Ala Arg Thr His Asp Val Phe Val Thr Val Glu Glu Asn Val
 545 550 555 560

Ile Ala Gly Gly Ala Gly Ser Ala Ile Asn Thr Phe Leu Gln Ala Gln
565 570 575

Lys Val Leu Met Pro Val Cys Asn Ile Gly Leu Pro Asp Arg Phe Val
580 585 590

Glu Gln Gly Ser Arg Glu Glu Leu Leu Ser Leu Val Gly Leu Asp Ser
595 600 605

Lys Gly Ile Phe Ala Thr Ile Glu Gln Phe Cys Ala
610 615 620

<210> 15

<211> 982

<212> DNA

<213> Methylobionas 16a

<220>

<221> CDS

<222> (22)..(975)

<400> 15

cccagtaaaa cactcaagaa t atg caa atc gta ctc gca aac ccc cgt gga 51
Met Gln Ile Val Leu Ala Asn Pro Arg Gly 10
1 5

ttc tgt gcc ggc gtg gac cgg gcc att gaa att gtc gat caa gcc atc 99
Phe Cys Ala Gly Val Asp Arg Ala Ile Glu Ile Val Asp Gln Ala Ile 15 20 25

gaa gcc ttt ggt gcg ccg att tat gtg cgg cac gag gtg gtg cat aac 147
Glu Ala Phe Gly Ala Pro Ile Tyr Val Arg His Glu Val Val His Asn 30 35 40

cgc acc gtg gtc gat gga ctg aaa caa aaa ggt gcg gtg ttc atc gag 195
Arg Thr Val Val Asp Gly Leu Lys Gln Lys Gly Ala Val Phe Ile Glu 45 50 55

gaa cta agc gat gtg ccg gtg ggt tcc tac ttg att ttc agc gcg cac 243
Glu Leu Ser Asp Val Pro Val Gly Ser Tyr Leu Ile Phe Ser Ala His 60 65 70

ggc gta tcc aag gag gtg caa cag gaa gcc gag gag cgc cag ttg acg 291
Gly Val Ser Lys Glu Val Gln Gln Glu Ala Glu Glu Arg Gln Leu Thr 75 80 85 90

gta ttc gat gcg act tgt ccg ctg gtg acc aaa gtg cac atg cag gtt 339
Val Phe Asp Ala Thr Cys Pro Leu Val Thr Lys Val His Met Gln Val 95 100 105

gcc aag cat gcc aaa cag ggc cga gaa gtg att ttg atc ggc cac gcc 387
Ala Lys His Ala Lys Gln Gly Arg Glu Val Ile Leu Ile Gly His Ala 110 115 120

ggt cat ccg gaa gtg gaa ggc acg atg ggc cag tat gaa aaa tgc acc 435
Gly His Pro Glu Val Glu Gly Thr Met Gly Gln Tyr Glu Lys Cys Thr 125 130 135

gaa ggc ggc ggc att tat ctg gtc gaa act ccg gaa gac gta cgc aat 483
Glu Gly Gly Gly Ile Tyr Leu Val Glu Thr Pro Glu Asp Val Arg Asn 140 145 150

ttg aaa gtc aac aat ccc aat gat ctg gcc tat gtg acg cag acg acc 531
Page 22

Leu Lys Val Asn Asn Pro Asn Asp Leu Ala Tyr Val Thr Gln Thr Thr
 155 160 165 170
 ttg tcg atg acc gac acc aag gtc atg gtg gat gcg tta cgc gaa caa 579
 Leu Ser Met Thr Asp Thr Lys Val Met Val Asp Ala Leu Arg Glu Gln
 175 180 185
 ttt ccg tcc att aag gag caa aaa aag gac gat att tgt tac gcg acg 627
 Phe Pro Ser Ile Lys Glu Gln Lys Lys Asp Asp Ile Cys Tyr Ala Thr
 190 195 200
 caa aac cgt cag gat gcg gtg cat gat ctg gcc aag att tcc gac ctg 675
 Gln Asn Arg Gln Asp Ala Val His Asp Leu Ala Lys Ile Ser Asp Leu
 205 210 215
 att ctg gtt gtc ggc tct ccc aat agt tcg aat tcc aac cgt ttg cgt 723
 Ile Leu Val Val Gly Ser Pro Asn Ser Ser Asn Ser Asn Arg Leu Arg
 220 225 230
 gaa atc gcc gtg caa ctc ggt aaa ccc gct tat ttg atc gat act tac 771
 Glu Ile Ala Val Gln Leu Gly Lys Pro Ala Tyr Leu Ile Asp Thr Tyr
 235 240 245 250
 cag gat ttg aag caa gat tgg ctg gag gga att gaa gta gtc ggg gtt 819
 Gln Asp Leu Lys Gln Asp Trp Leu Glu Gly Ile Glu Val Val Gly Val
 255 260 265
 acc gcg ggc gct tcg gcg ccg gaa gtg ttg gtg cag gaa gtg atc gat 867
 Thr Ala Gly Ala Ser Ala Pro Glu Val Leu Val Gln Glu Val Ile Asp
 270 275 280
 caa ctg aag gca tgg ggc ggc gaa acc act tcg gtc aga gaa aac agc 915
 Gln Leu Lys Ala Trp Gly Gly Glu Thr Thr Ser Val Arg Glu Asn Ser
 285 290 295
 ggc atc gag gaa aag gta gtc ttt tcg att ccc aag gag ttg aaa aaa 963
 Gly Ile Glu Glu Lys Val Val Phe Ser Ile Pro Lys Glu Leu Lys Lys
 300 305 310
 cat atg caa gcg tgatcaa 982
 His Met Gln Ala
 315

<210> 16
 <211> 318
 <212> PRT
 <213> Methylomonas 16a

<400> 16

Met Gln Ile Val Leu Ala Asn Pro Arg Gly Phe Cys Ala Gly Val Asp
 1 5 10 15

Arg Ala Ile Glu Ile Val Asp Gln Ala Ile Glu Ala Phe Gly Ala Pro
 20 25 30

Ile Tyr Val Arg His Glu Val Val His Asn Arg Thr Val Val Asp Gly
 35 40 45

Leu Lys Gln Lys Gly Ala Val Phe Ile Glu Glu Leu Ser Asp Val Pro
 50 55 60

Val Gly Ser Tyr Leu Ile Phe Ser Ala His Gly Val Ser Lys Glu Val
 65 70 75 80

Gln Gln Glu Ala Glu Glu Arg Gln Leu Thr Val Phe Asp Ala Thr Cys
85 90 95

Pro Leu Val Thr Lys Val His Met Gln Val Ala Lys His Ala Lys Gln
100 105 110

Gly Arg Glu Val Ile Leu Ile Gly His Ala Gly His Pro Glu Val Glu
115 120 125

Gly Thr Met Gly Gln Tyr Glu Lys Cys Thr Glu Gly Gly Ile Tyr
130 135 140

Leu Val Glu Thr Pro Glu Asp Val Arg Asn Leu Lys Val Asn Asn Pro
145 150 155 160

Asn Asp Leu Ala Tyr Val Thr Gln Thr Thr Leu Ser Met Thr Asp Thr
165 170 175

Lys Val Met Val Asp Ala Leu Arg Glu Gln Phe Pro Ser Ile Lys Glu
180 185 190

Gln Lys Lys Asp Asp Ile Cys Tyr Ala Thr Gln Asn Arg Gln Asp Ala
195 200 205

Val His Asp Leu Ala Lys Ile Ser Asp Leu Ile Leu Val Val Gly Ser
210 215 220

Pro Asn Ser Ser Asn Ser Asn Arg Leu Arg Glu Ile Ala Val Gln Leu
225 230 235 240

Gly Lys Pro Ala Tyr Leu Ile Asp Thr Tyr Gln Asp Leu Lys Gln Asp
245 250 255

Trp Leu Glu Gly Ile Glu Val Val Gly Val Thr Ala Gly Ala Ser Ala
260 265 270

Pro Glu Val Leu Val Gln Glu Val Ile Asp Gln Leu Lys Ala Trp Gly
275 280 285

Gly Glu Thr Thr Ser Val Arg Glu Asn Ser Gly Ile Glu Glu Lys Val
290 295 300

Val Phe Ser Ile Pro Lys Glu Leu Lys Lys His Met Gln Ala
305 310 315

<210> 17
<211> 1254
<212> DNA
<213> Methylobionas 16a

<220>

<221> CDS

<222> (73)..(1254)

<400> 17

ggtggacagc atcattgctgg cggcaccgtt tttctatgcc ggtatcgtgc tgatcggacg 60

gagcgtattc ga atg aaa ggt att tgc ata ttg ggc gct acc ggt tgc atc 111
Met Lys Gly Ile Cys Ile Leu Gly Ala Thr Gly Ser Ile
1 5 10ggt gtc agc acg ctg gat gtc gtt gcc agg cat ccg gat aaa tat caa 159
Gly Val Ser Thr Leu Asp Val Val Ala Arg His Pro Asp Lys Tyr Gln
15 20 25gtc gtt gcg ctg acc gcc aac ggc aat atc gac gca ttg tat gaa caa 207
Val Val Ala Leu Thr Ala Asn Gly Asn Ile Asp Ala Leu Tyr Glu Gln
30 35 40 45tgc ctg gcc cac cat ccg gag tat gcg gtg gtc atg gaa agc aag 255
Cys Leu Ala His His Pro Glu Tyr Ala Val Val Met Glu Ser Lys
50 55 60gta gca gag ttc aaa cag cgc att gcc gct tgc ccg gta gcg gat atc 303
Val Ala Glu Phe Lys Gln Arg Ile Ala Ala Ser Pro Val Ala Asp Ile
65 70 75aag gtc ttg tgc ggt agc gag gcc ttg caa cag gtg gcc acg ctg gaa 351
Lys Val Leu Ser Gly Ser Glu Ala Leu Gln Gln Val Ala Thr Leu Glu
80 85 90aac gtc gat acg gtg atg gcg gct atc gtc ggc gcg gcc gga ttg ttg 399
Asn Val Asp Thr Val Met Ala Ala Ile Val Gly Ala Ala Gly Leu Leu
95 100 105ccg acc ttg gcc gcg gcc aag gcc ggc aaa acc gtg ctg ttg gcc aac 447
Pro Thr Leu Ala Ala Lys Ala Gly Lys Thr Val Leu Leu Ala Asn
110 115 120 125aag gaa gcc ttg gtg atg tgc gga caa atc ttc atg cag gcc gtc agc 495
Lys Glu Ala Leu Val Met Ser Gly Gln Ile Phe Met Gln Ala Val Ser
130 135 140gat tcc ggc gct gtg ttg ctg ccg ata gac agc gag cac aac gcc atc 543
Asp Ser Gly Ala Val Leu Leu Pro Ile Asp Ser Glu His Asn Ala Ile
145 150 155ttt cag tgc atg ccg gcg ggt tat acg cca ggc cat aca gcc aaa cag 591
Phe Gln Cys Met Pro Ala Gly Tyr Thr Pro Gly His Thr Ala Lys Gln
160 165 170gcg cgc cgc att tta ttg acc gct tcc ggt ggc cca ttt cga cgg acg 639
Ala Arg Arg Ile Leu Leu Thr Ala Ser Gly Gly Pro Phe Arg Arg Thr
175 180 185ccg ata gaa acg ttg tcc agc gtc acg ccg gat cag gcc gtt gcc cat 687
Pro Ile Glu Thr Leu Ser Ser Val Thr Pro Asp Gln Ala Val Ala His
190 195 200 205cct aaa tgg gac atg ggg cgc aag att tgc gtc gat tcc gcc acc atg 735
Pro Lys Trp Asp Met Gly Arg Lys Ile Ser Val Asp Ser Ala Thr Met
210 215 220atg aac aaa ggt ctc gaa ctg atc gaa gcc tgc ttg ttg ttc aac atg 783
Met Asn Lys Gly Leu Glu Leu Ile Glu Ala Cys Leu Leu Phe Asn Met
225 230 235gag ccc gac cag att gaa gtc gtc att cat ccg cag agc atc att cat 831
Glu Pro Asp Gln Ile Glu Val Val Ile His Pro Gln Ser Ile Ile His
240 245 250

tcg atg gtg gac tat gtc gat ggt tcg gtt ttg gcg cag atg ggt aat 879
 Ser Met Val Asp Tyr Val Asp Gly Ser Val Leu Ala Gln Met Gly Asn
 255 260 265

ccc gac atg cgc acg ccg ata gcg cac gcg atg gcc tgg ccg gaa cgc 927
 Pro Asp Met Arg Thr Pro Ile Ala His Ala Met Ala Trp Pro Glu Arg
 270 275 280 285

ttt gac tct ggt gtg gcg ccg ctg gat att ttc gaa gta ggg cac atg 975
 Phe Asp Ser Gly Val Ala Pro Leu Asp Ile Phe Glu Val Gly His Met
 290 295 300

gat ttc gaa aaa ccc gac ttg aaa cgg ttt cct tgt ctg aga ttg gct 1023
 Asp Phe Glu Lys Pro Asp Leu Lys Arg Phe Pro Cys Leu Arg Leu Ala
 305 310 315

tat gaa gcc atc aag tct ggt gga att atg cca acg gta ttg aac gca 1071
 Tyr Glu Ala Ile Lys Ser Gly Gly Ile Met Pro Thr Val Leu Asn Ala
 320 325 330

gcc aat gaa att gct gtc gaa gcg ttt tta aat gaa gaa gtc aaa ttc 1119
 Ala Asn Glu Ile Ala Val Glu Ala Phe Leu Asn Glu Glu Val Lys Phe
 335 340 345

act gac atc gcg gtc atc atc gag cgc agc atg gcc cag ttt aaa ccg 1167
 Thr Asp Ile Ala Val Ile Ile Glu Arg Ser Met Ala Gln Phe Lys Pro
 350 355 360 365

gac gat gcc ggc agc ctc gaa ttg gtt ttg cag gcc gat caa gat gcg 1215
 Asp Asp Ala Gly Ser Leu Glu Leu Val Leu Gln Ala Asp Gln Asp Ala
 370 375 380

cgc gag gtg gct aga gac atc atc aag acc ttg gta gct 1254
 Arg Glu Val Ala Arg Asp Ile Ile Lys Thr Leu Val Ala
 385 390

<210> 18
 <211> 394
 <212> PRT
 <213> Methylobionas 16a

<400> 18

Met Lys Gly Ile Cys Ile Leu Gly Ala Thr Gly Ser Ile Gly Val Ser
 1 5 10 15

Thr Leu Asp Val Val Ala Arg His Pro Asp Lys Tyr Gln Val Val Ala
 20 25 30

Leu Thr Ala Asn Gly Asn Ile Asp Ala Leu Tyr Glu Gln Cys Leu Ala
 35 40 45

His His Pro Glu Tyr Ala Val Val Val Met Glu Ser Lys Val Ala Glu
 50 55 60

Phe Lys Gln Arg Ile Ala Ala Ser Pro Val Ala Asp Ile Lys Val Leu
 65 70 75 80

Ser Gly Ser Glu Ala Leu Gln Gln Val Ala Thr Leu Glu Asn Val Asp
 85 90 95

Thr Val Met Ala Ala Ile Val Gly Ala Ala Gly Leu Leu Pro Thr Leu
 100 105 110
 Ala Ala Ala Lys Ala Gly Lys Thr Val Leu Leu Ala Asn Lys Glu Ala
 115 120 125
 Leu Val Met Ser Gly Gln Ile Phe Met Gln Ala Val Ser Asp Ser Gly
 130 135 140
 Ala Val Leu Leu Pro Ile Asp Ser Glu His Asn Ala Ile Phe Gln Cys
 145 150 155 160
 Met Pro Ala Gly Tyr Thr Pro Gly His Thr Ala Lys Gln Ala Arg Arg
 165 170 175
 Ile Leu Leu Thr Ala Ser Gly Gly Pro Phe Arg Arg Thr Pro Ile Glu
 180 185 190
 Thr Leu Ser Ser Val Thr Pro Asp Gln Ala Val Ala His Pro Lys Trp
 195 200 205
 Asp Met Gly Arg Lys Ile Ser Val Asp Ser Ala Thr Met Met Asn Lys
 210 215 220
 Gly Leu Glu Leu Ile Glu Ala Cys Leu Leu Phe Asn Met Glu Pro Asp
 225 230 235 240
 Gln Ile Glu Val Val Ile His Pro Gln Ser Ile Ile His Ser Met Val
 245 250 255
 Asp Tyr Val Asp Gly Ser Val Leu Ala Gln Met Gly Asn Pro Asp Met
 260 265 270
 Arg Thr Pro Ile Ala His Ala Met Ala Trp Pro Glu Arg Phe Asp Ser
 275 280 285
 Gly Val Ala Pro Leu Asp Ile Phe Glu Val Gly His Met Asp Phe Glu
 290 295 300
 Lys Pro Asp Leu Lys Arg Phe Pro Cys Leu Arg Leu Ala Tyr Glu Ala
 305 310 315 320
 Ile Lys Ser Gly Gly Ile Met Pro Thr Val Leu Asn Ala Ala Asn Glu
 325 330 335
 Ile Ala Val Glu Ala Phe Leu Asn Glu Glu Val Lys Phe Thr Asp Ile
 340 345 350
 Ala Val Ile Ile Glu Arg Ser Met Ala Gln Phe Lys Pro Asp Asp Ala
 355 360 365
 Gly Ser Leu Glu Leu Val Leu Gln Ala Asp Gln Asp Ala Arg Glu Val
 Page 27

370

375

380

Ala Arg Asp Ile Ile Lys Thr Leu Val Ala
 385 390

<210> 19
 <211> 25
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer #1 for amplification of crt gene cluster

<400> 19
 atgacggtct gcgcaaaaaa acacg 25

<210> 20
 <211> 28
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer #2 for amplification of crt gene cluster

<400> 20
 gagaaattat gttgtggatt tggaatgc 28

<210> 21
 <211> 61
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'kan(dxs)

<400> 21
 tggaagcgct agcggactac atcatccagc gtaataaata acgtcttgag cgattgtgta 60
 g 61

<210> 22
 <211> 65
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'kan(idi)

<400> 22
 tctgatgcgc aagctgaaga aaaatgagca tggagaataa tatgacgtct tgagcgattg 60
 tgtag 65

<210> 23
 <211> 65
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'kan(ygbBP)

<400> 23
 gacgcgtcga agcgcgcaca gtctgcgggg caaaacaatc gataacgtct tgagcgattg 60

tgtag 65

<210> 24
 <211> 60
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'kan(ispAdxs)

<400> 24
 accatgacgg ggcgaaaaat attgagagtc agacattcat gtgtaggctg gagctgcttc 60

<210> 25
 <211> 64
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'kan

<400> 25
 gaagacgaaa gggcctcgtg atacgcctat ttttataggt tatatgaata tcctccttag 60
 ttcc 64

<210> 26
 <211> 50
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-T5

<400> 26
 ctaaggagga tattcatata acctataaaa ataggcgtat cacgaggccc 50

<210> 27
 <211> 70
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-T5(dxs)

<400> 27
 ggagtcgacc agtgccaggg tcgggtattt ggcaatatca aaactcatag ttaatttctc 60
 ctctttaatg 70

<210> 28
 <211> 68
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-T5(idi)

<400> 28
 tgggaactcc ctgtgcattc aataaaatga cgtgttccgt ttgcatagtt aatttctcct 60
 ctttaatg 68

<210> 29
 <211> 68
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-T5(ygbBP)

<400> 29
 cggccgcccg aaccacggcg caaacatcca aatgagtggg tgccatagtt aatttctcct 60
 ctttaatg 68

<210> 30
 <211> 62
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-T5(ispAdxs)

<400> 30
 cctgcttaac gcaggcttcg agttgctgcg gaaagtccat agttaatttc tcctctttaa 60
 tg 62

<210> 31
 <211> 65
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-kanT5(ispB)

<400> 31
 accataaacc ctaagttgcc tttgttcaca gtaaggtaat cggggcgtct tgagcgattg 60
 tgtag 65

<210> 32
 <211> 67
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-kanT5(ispB)

<400> 32
 cgccatatct tgcgcgggta actcattgat tttttctaaa ttcatagtta atttctcctc 60
 tttaatg 67

<210> 33
 <211> 156
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Phage T5 promoter sequence

<400> 33
 ctataaaaat aggcgtatca cgaggccctt tcgtcttcac ctcgagaaat cataaaaaat 60
 ttatttgctt tgtgagcgga taacaattat aatagattca attgtgagcg gataacaatt 120

tcacacagaa ttcattaaag aggagaaatt aactca 156

<210> 34
 <211> 65
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-kanT5(dxsl6a)

<400> 34
 cactaacgcc cgcacattgc tgcgggcttt ttgattcatt tcgcacgtct tgagcgattg 60
 tgtag 65

<210> 35
 <211> 65
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-kanT5(dxrl6a)

<400> 35
 taaagggcta agagtagtgt gctcttagcc ctttaattacg tttcccgctct tgagcgattg 60
 tgtag 65

<210> 36
 <211> 65
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-kanT5(lytB16a)

<400> 36
 ctacaactgg cgagatgcat agcgagtata atttgtattt tgcgtcgtct tgagcgattg 60
 tgtag 65

<210> 37
 <211> 51
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-kanT5(dxsl6a)

<400> 37
 agtagaggga agtctttgga aagagccata gtttaatttct cctctttaat g 51

<210> 38
 <211> 51
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-kanT5(dxrl6a)

<400> 38
 acggtgccgc cgcaatgatg ctgtccacca gtttaatttct cctctttaat g 51

<210> 39
 <211> 51
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-kanT5(lytB16a)

<400> 39
 ccacgggggt ttgcgagtac gatttgcata gttaatttct cctctttaat g 51

<210> 40
 <211> 55
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-(dxs16a)

<400> 40
 acagaattca ttaaagagga gaaattaact atggctcttt ccaaagactt ccctc 55

<210> 41
 <211> 55
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-(dxr16a)

<400> 41
 acagaattca ttaaagagga gaaattaact ggtggacagc atcattgcgg cggca 55

<210> 42
 <211> 55
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 5'-(lytB16a)

<400> 42
 acagaattca ttaaagagga gaaattaact atgcaaactg tactcgcaaa ccccc 55

<210> 43
 <211> 68
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer 3'-(dxs16a)

<400> 43
 aggagcgaag tgattatcag tatgctgttc atatagcctc gaattatcaa gcgcaaaact 60
 gttcgaatg 68

<210> 44
 <211> 67
 <212> DNA
 <213> Artificial sequence

<220>

<223> Primer 3'-(dxr16a)

<400> 44
ggcattttca ctctggcaat gcgcataaac gctttcaaag tcctgttaag ctaccaaggt 60
cttgatg 67

<210> 45
<211> 68
<212> DNA
<213> Artificial sequence

<220>
<223> Primer 3'-(lytB16a)

<400> 45
agtggcggac gggcaaaca gggtaacata ggatcaatga gggttattga tcacgcttgc 60
atatgttt 68

<210> 46
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Primer T-kan

<400> 46
accggatatc accacttatc tgctc 25

<210> 47
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Primer B-ispA

<400> 47
cctaataatg cgccatactg catgg 25

<210> 48
<211> 32
<212> DNA
<213> Artificial sequence

<220>
<223> Primer T-T5

<400> 48
taacctataa aaataggcgt atcacgaggc cc 32

<210> 49
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Primer B-idi

<400> 49
tcatgctgac ctggtgaagg aatcc 25

<210> 50
 <211> 25
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer B-dxs(16a)

<400> 50
 gcgatattgt atgtctgatt cagga

25

<210> 51
 <211> 25
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer B-lytB(16a)

<400> 51
 tccactggat gcgggaagct ggcag

25

<210> 52
 <211> 26
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer B-dxs

<400> 52
 tggcaacagt cgtagctcct gggtagg

26

<210> 53
 <211> 25
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer B-ygb

<400> 53
 ccagcagcgc atgcaccgag tgttc

25

<210> 54
 <211> 21
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer Tn5PCRF

<400> 54
 gctgagttga aggatcagat c

21

<210> 55
 <211> 21
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer Tn5PCRR

<400> 55
cgagcaagac gtttcccgtt g 21

<210> 56
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Primer Kan-2 FP-1

<400> 56
acctacaaca aagctctcat caacc 25

<210> 57
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Primer Kan-2 RP-1

<400> 57
gcaatgtaac atcagagatt ttgag 25

<210> 58
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> Primer Y15_F

<400> 58
ggatcgatct tgagatgacc 20

<210> 59
<211> 24
<212> DNA
<213> Artificial sequence

<220>
<223> Primer Y15_R

<400> 59
gctttcgtaa ttttcgcatt tctg 24

<210> 60
<211> 25
<212> DNA
<213> Artificial sequence

<220>
<223> Primer T-Tn5yjeR

<400> 60
gcaatgtaac atcagagatt ttgag 25

<210> 61
<211> 24
<212> DNA
<213> Artificial sequence

<220>
 <223> Primer B-yjeR

<400> 61
 gctttcgtaa ttttcgcatt tctg 24

<210> 62
 <211> 26
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Primer B-ispB

<400> 62
 agtacagcaa tcatcggacg aatacg 26

<210> 63
 <211> 1845
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Sequence yjeR::Tn5 mutant gene (transposon disrupted yjeR)

<400> 63
 atgggcaaaa catctatgat acacgcaatt gtggatcaat atagtcaactg tgaatgggtg 60
 gaaaatagca tgagtgcgaa tgaaaacaac ctgatttggga tcgatcttga gatgaccggt 120
 ctggatcccg agcgcgatcg cattattgag attgccacgc tggtagaccga tgccaacctg 180
 aatatttctgg cagaagggcc gaccattgca gtacaccagt ctgatgaaca gctggcgctg 240
 atggatgact ggaacgtgcg caccataacc gccagcgggc tggtagagcg cgtgaaagcg 300
 agcacgatgg gcgatcggga agctgaactg gcaacgctcg aatttttaaa acagtgggtg 360
 cctgcgggaa aatcgccgat ttgcggtaac agcatcggtc aggaccgtcg tttcctgttt 420
 aaatacatgc cggagctgga agcctacttc cactaccgtt atctcgatgt cagcaccctg 480
 aaagagctgg cgcgccgctg gaagccggaa attctggatg gttttaccaa gcaggggacg 540
 catcaggcga tggatgatat ccgtgaatcg gtggcggagc tggcttacta cctgtctctt 600
 atacacatct caaccctgaa gcttgcattg ctgcaggctg actctagagg atccccgcca 660
 cggttgatga gagctttgtt gtaggtggac cagttgggtga ttttgaactt ttgctttgcc 720
 acggaacggt ctgcgttgtc ggggaagatgc gtgatctgat ccttcaactc agcaaaaagt 780
 cgattttattc aacaaagccg ccgtcccgtc aagtcagcgt aatgctctgc cagtgttaca 840
 accaattaac caattctgat tagaaaaact catcgagcat caaatgaaac tgcaatttat 900
 tcatatcagg attatcaata ccatatTTTT gaaaaagccg tttctgtaat gaaggagaaa 960
 actcaccgag gcagttccat aggatggcaa gatcctggta tcggtctgcg attccgactc 1020
 gtccaacatc aatacaacct attaatTTCC cctcgtcaaa aataaggtta tcaagtgaga 1080
 aatcaccatg agtgacgact gaatccggtg agaatggcaa aagtttatgc atttctttcc 1140
 agacttgttc aacaggccag ccattacgct cgtcatcaaa atcactcgca tcaaccaaac 1200
 cgttattcat tcgtgattgc gcctgagcga gacgaaatac gcgatcgtg ttaaaaggac 1260

```

aattacaaac aggaatcgaa tgcaaccggc gcaggaacac tgccagcgca tcaacaatat 1320
tttcacctga atcaggatat tcttctaata cctggaatgc tgtttttccg gggatcgcag 1380
tggtgagtaa ccatgcatca tcaggagtac ggataaaatg cttgatggtc ggaagaggca 1440
taaattccgt cagccagttt agtctgacca tctcatctgt aacatcattg gcaacgctac 1500
ctttgccatg tttcagaaac aactctggcg catcgggctt cccatacaat cgatagattg 1560
tcgcacctga ttgcccgaca ttatcgcgag cccatttata cccatataaa tcagcatcca 1620
tgttggaatt taatcgcggc ctcgagcaag acgtttcccg ttgaatatgg ctcataacac 1680
cccttgattt actgtttatg taagcagaca gttttattgt tcatgatgat atatttttat 1740
cttgatgcaat gtaacatcag agattttgag acacaattca tcgatgatgg ttgagatgtg 1800
tataagagac aggcttacta ccgcgagcat tttatcaagc tgtaa 1845

```

```

<210> 64
<211> 8609
<212> DNA
<213> Artificial sequence

```

```

<220>
<223> Plasmid pPCB15

```

```

<400> 64
cgtatggcaa tgaaagacgg tgagctggtg atatgggata gtgttcaccc ttgttacacc 60
gttttccatg agcaaactga aacgttttca tcgctctgga gtgaatacca cgacgatttc 120
cggcagtttc tacacatata ttcgcaagat gtggcggtgt acggtgaaaa cctggcctat 180
ttccctaaag ggtttattga gaatatgttt ttcgtctcag ccaatccctg ggtgagtttc 240
accagttttg atttaaactg ggccaatatg gacaacttct tcgccccgt tttcaccatg 300
ggcaaatatt atacgcaagg cgacaagggt ctgatgccgc tggcgattca ggttcatcat 360
gccgtctgtg atggcttcca tgtcggcaga atgcttaatg aattacaaca gtactgcgat 420
gagtggcagg gcggggcgta attttttttaa ggcagttatt ggtgcctaga aatattttat 480
ctgattaata agatgatctt cttgagatcg ttttggctcg cgcgtaatct cttgctctga 540
aaacgaaaaa accgccttgc agggcggttt ttcgaagggt ctctgagcta ccaactcttt 600
gaaccgaggt aactggcttg gaggagcgca gtcacaaaaa cttgtccttt cagtttagcc 660
ttaaccggcg catgacttca agactaactc ctctaaatca attaccagtg gctgctgcca 720
gtggtgcttt tgcattgtct tccgggttgg actcaagacg atagttaccg gataaggcgc 780
agcggctcga ctgaacgggg ggttcgtgca tacagtccag cttggagcga actgcctacc 840
cggaactgag tgtcaggcgt ggaatgagac aaacgcggcc ataacagcgg aatgacaccg 900
gtaaaccgaa aggcaggaac aggagagcgc acgaggggag cgccagggga aacgcctggt 960
atctttatag tcctgtcggg tttcgccacc actgatttga gcgtcagatt tcgtgatgct 1020
tgtcaggggg gcggagccta tggaaaaacg gctttgccgc ggccctctca cttccctggt 1080
aagtatcttc ctggcatctt ccaggaaatc tccgccccgt tcgtaagcca tttccgctcg 1140

```

ccgcagtcga	acgaccgagc	gtagcgagtc	agtgagcgag	gaagcggaat	atatcctgta	1200
tcacatattc	tgctgacgca	ccggtgcagc	cttttttctc	ctgccacatg	aagcacttca	1260
ctgacaccct	catcagtgcc	aacatagtaa	gccagtatat	acactccgct	agcgcccaat	1320
acgcaaaccg	cctctccccg	cgcgttggcc	gattcattaa	tgcagctggc	acgacagggt	1380
tcccgactgg	aaagcgggca	gtgagcgcaa	cgcaattaat	gtgagtttagc	tcactcatta	1440
ggcaccccag	gctttacact	ttatgcttcc	ggctcgtatg	ttgtgtggaa	ttgtgagcgg	1500
ataacaattt	cacacaggaa	acagctatga	ccatgattac	gaattcgagc	tcggtacca	1560
aacgaattcg	cccttttgac	ggctcgcgca	aaaaaacacg	ttcaccttac	tggcatttcg	1620
gctgagcagt	tgctggctga	tatcgatagc	cgcttgatc	agttactgcc	ggttcagggt	1680
gagcgggatt	gtgtgggtgc	cgcgatgcgt	gaaggcacgc	tggcaccggg	caaacgtatt	1740
cgtccgatgc	tgctgttatt	aacagcgcg	gatcttggt	gtgcgatcag	tcacggggga	1800
ttactggatt	tagcctgcgc	ggttgaaatg	gtgcatgctg	cctcgctgat	tctggatgat	1860
atgccctgca	tggaacgatgc	gcagatgcgt	cgggggcgtc	ccaccattca	cacgcagtac	1920
gggtgaacatg	tggcgattct	ggcggcggtc	gctttactca	gcaaagcggt	tgggggtgatt	1980
gccgaggctg	aaggctcgac	gccgatagcc	aaaactcgcg	cgggtgtcga	gctgtccact	2040
gcgattggca	tgcagggtct	ggttcagggc	cagtttaagg	acctctcgga	aggcgataaa	2100
ccccgcagcg	ccgatgccat	actgctaacc	aatcagttta	aaaccagcac	gctgttttgc	2160
gcgtcaacgc	aaatggcgtc	cattgcggcc	aacgcgtcct	gcgaagcgcg	tgagaacctg	2220
catcgtttct	cgctcgatct	cggccaggcc	tttcagttgc	ttgacgatct	taccgatggc	2280
atgaccgata	ccggcaaaga	catcaatcag	gatgcaggta	aatcaacgct	gggtcaattta	2340
ttaggctcag	gcgcggtcga	agaacgcctg	cgacagcatt	tgcgcctggc	cagtgaacac	2400
ctttccgcgg	catgccaaaa	cggccattcc	accacccaac	tttttattca	ggcctgggtt	2460
gacaaaaaac	tcgctgccgt	cagttaagga	tgctgcatga	gccattttgc	ggatgatcgca	2520
ccgccctttt	tcagccatgt	tcgcgctctg	caaaaccttg	ctcaggaatt	agtggcccgc	2580
ggatcatcgtg	ttacgttttt	tcagcaacat	gactgcaaag	cgctggtaac	gggcagcgat	2640
atcggattcc	agaccgtcgg	actgcaaacg	catcctccc	gttccttata	gcacctgctg	2700
cacctggccg	cgcaccact	cggacctctg	atgttacgac	tgatcaatga	aatggcacgt	2760
accagcgata	tgctttgccg	ggaactgccc	gccgcttttc	atgcgttgca	gatagagggc	2820
gtgatcgttg	atcaaatgga	gccggcaggt	gcagtagtcg	cagaagcgtc	aggtctgccg	2880
tttgtttcgg	tggcctgcgc	gctgccgctc	aaccgcgaac	cgggtttgcc	tctggcggtg	2940
atgcctttcg	agtacggcac	cagcgatgcg	gctcgggaac	gctataccac	cagcgaaaaa	3000
atztatgact	ggctgatgcg	acgtcacgat	cgtgtgatcg	cgcacatgc	atgcagaatg	3060
ggtttagccc	cgctgaaaaa	actgcatcat	tgtttttctc	cactggcaca	aatcagccag	3120
ttgatccccg	aactggattt	tccccgcaaa	gcgctgccag	actgctttca	tgcggttgga	3180
ccgttacggc	aaccccgagg	gacgccgggg	tcacaaactt	cttattttcc	gtccccggac	3240

aaaccccgta	tttttgcctc	gctgggcacc	ctgcagggac	atcgttatgg	cctgttcagg	3300
accatcgcca	aagcctgcga	agaggtggat	gcgcagttac	tgttggcaca	ctgtggcggc	3360
ctctcagcca	cgcaggcagg	tgaactggcc	cggggcgggg	acattcagg	tgtggatttt	3420
gccgatcaat	ccgcagcact	ttcacaggca	cagttgacaa	tcacacatgg	tgggatgaat	3480
acggtactgg	acgctattgc	ttcccgcaca	ccgctactgg	cgctgccgct	ggcatttgat	3540
caacctggcg	tggcatcacg	aattgtttat	catggcatcg	gcaagcgtgc	gtctcggttt	3600
actaccagcc	atgcgctggc	gcggcagatt	cgatcgctgc	tgactaacac	cgattacccg	3660
cagcgtatga	caaaaattca	ggccgcattg	cgtctggcag	gcggcacacc	agccgccgcc	3720
gatattgttg	aacaggcgat	gcggacctgt	cagccagtag	tcagtgggca	ggattatgca	3780
accgcactat	gatctcattc	tggtcgggtg	cggctctggc	aatggcctta	tcgcgctccg	3840
gcttcagcaa	cagcatccgg	atatgcggat	cttgcttatt	gaggcgggtc	ctgaggcggg	3900
agggaaacct	acctggtcct	ttcacgaaga	ggatttaacg	ctgaatcagc	atcgctggat	3960
agcgcgcgtt	gtgggtccatc	actggcccga	ctaccagggt	cgtttcccc	aacgccgtcg	4020
ccatgtgaac	agtggctact	actgcgtgac	ctcccggcat	ttcgccggga	tactccggca	4080
acagtttggg	caacatttat	ggctgcatac	cgcggtttca	gccgttcattg	ctgaatcggg	4140
ccagtttagcg	gatggccgga	ttattcatgc	cagtacagtg	atcgacggac	ggggttacac	4200
gcctgattct	gcactacgcg	taggattcca	ggcatttatc	ggtcaggagt	ggcaactgag	4260
cgcgccgcat	ggtttatcgt	caccgattat	catggatgcg	acggtcgatc	agcaaaatgg	4320
ctaccgcttt	gtttataccc	tgccgctttc	cgcaaccgca	ctgctgatcg	aagacacaca	4380
ctacattgac	aaggctaata	ttcaggccga	acgggcgcgt	cagaacattc	gcgattatgc	4440
tgcgcgacag	ggttggccgt	tacagacggt	gctgcgggaa	gaacagggtg	cattgcccatt	4500
tacgttaacg	ggcgataatc	gtcagttttg	gcaacagcaa	ccgcaagcct	gtagcggatt	4560
acgcgccggg	ctgtttcatc	cgacaaccgg	ctactcccta	ccgctcgcgg	tggcgctggc	4620
cgatcgtctc	agcgcgctgg	atgtgtttac	ctcttcctct	gttcaccaga	cgattgctca	4680
ctttgccag	caacgttggc	agcaacagg	gtttttccgc	atgctgaatc	gcatgttggt	4740
tttagccgga	ccggccgagt	cacgctggcg	tgtgatgcag	cgtttctatg	gcttaccgga	4800
ggatttgatt	gcccgccttt	atgcgggaaa	actcaccgtg	accgatcggc	tacgcattct	4860
gagcggcaag	ccgcccgttc	ccgttttcgc	ggcattgcag	gcaattatga	cgactcatcg	4920
ttgaagagcg	actacatgaa	accaactacg	gtaattggtg	cgggctttgg	tggcctggca	4980
ctggcaattc	gtttacaggc	cgcagggtatt	cctgttttgc	tgcttgagca	gcgcgacaag	5040
ccgggtggcc	gggcttatgt	ttatcaggag	cagggtttta	cttttgatgc	aggccctacc	5100
gttatcaccg	atcccagcgc	gattgaagaa	ctgtttgctc	tggccggtaa	acagcttaag	5160
gattacgtcg	agctgttgcc	ggcacgccc	ttttatcgcc	tgtgctggga	gtccggcaag	5220
gtcttcaatt	acgataacga	ccaggcccag	ttagaagcgc	agatacagca	gtttaatccg	5280

cgcgatgttg	cgggttatcg	agcgttcctt	gactattcgc	gtgccgtatt	caatgagggc	5340
tatctgaagc	tcggcactgt	gcctttttta	tcgttcaaag	acatgcttcg	ggccgcgccc	5400
cagttggcaa	agctgcaggc	atggcgcagc	gtttacagta	aagttgccgg	ctacattgag	5460
gatgagcatc	ttcggcaggc	gttttctttt	cactcgctct	tagtgggggg	gaatccgttt	5520
gcaacctcgt	ccatttatac	gctgattcac	gcgttagaac	gggaatgggg	cgtctggttt	5580
ccacgcggtg	gaaccggtgc	gctggtcaat	ggcatgatca	agctgtttca	ggatctgggc	5640
ggcgaagtcg	tgcttaacgc	ccgggtcagt	catatggaaa	ccgttgggga	caagattcag	5700
gccgtgcagt	tggaagacgg	cagacggttt	gaaacctgcg	cggtggcgct	gaacgctgat	5760
gttgtagata	cctatcgcg	tctgctgtct	cagcatcccc	cagccgctaa	gcaggcgaaa	5820
aaactgcaat	ccaagcgtat	gagtaactca	ctgtttgtac	tctattttgg	tctcaaccat	5880
catcacgatc	aactcgccca	tcataccgtc	tgttttgggc	cacgctaccg	tgaactgatt	5940
cacgaaattt	ttaaccatga	tggtctggct	gaggattttt	cgctttattt	acacgcacct	6000
tgtgtcacgg	atccgtcact	ggcaccggaa	gggtgcggca	gctattatgt	gctggcgcct	6060
gttccacact	taggcacggc	gaacctcgac	tgggcggtag	aaggaccccg	actgcgcgat	6120
cgtatttttg	actaccttga	gcaacattac	atgcctggct	tgcaagcca	gttggtgacg	6180
caccgtatgt	ttacgccgtt	cgatttccgc	gacgagctca	atgcctggca	aggttcggcc	6240
ttctcggttg	aacctattct	gacctagagc	gcctggttcc	gaccacataa	ccgcgataag	6300
cacattgata	atctttatct	ggttggcgca	ggcacccatc	ctggcgcggg	cattcccggc	6360
gtaatcggct	cggcgaaggc	gacggcaggc	ttaatgctgg	aggacctgat	ttgacgaata	6420
cgtcattact	gaatcatgcc	gtcgaaacca	tggcggttgg	ctcgaaaagc	tttgcgactg	6480
catcgacgct	tttcgacgcc	aaaaccgcgc	gcagcgtgct	gatgctttac	gcatggtgcc	6540
gccactgcga	cgacgtcatt	gacgatcaaa	cactgggctt	tcatgccgac	cagccctctt	6600
cgcagatgcc	tgagcagcgc	ctgcagcagc	ttgaaatgaa	aacgcgtcag	gcctacgccg	6660
gttcgcaaat	gcacgagccc	gcttttgccg	cgtttcagga	ggtcgcgatg	gcgcatgata	6720
tcgctcccgc	ctacgcgttc	gaccatctgg	aaggttttgc	catggatgtg	cgcgaaacgc	6780
gctacctgac	actggacgat	acgctgcgtt	attgctatca	cgtcgccggg	gttggtgggcc	6840
tgatgatggc	gcaaattatg	ggcgttcgcg	ataacgccac	gctcgatcgc	gcctgcgatc	6900
tcgggctggc	tttcagttg	accaacattg	cgcgtgatat	tgtcgacgat	gctcaggtgg	6960
gccgctgtta	tctgcctgaa	agctggctgg	aagaggaagg	actgacgaaa	gcgaattatg	7020
ctgcgccaga	aaaccggcag	gccttaagcc	gtatcgccgg	gcgactggta	cgggaagcgg	7080
aaccctatta	cgtatcatca	atggccggtc	tggcacaatt	acccttacgc	tcggcctggg	7140
ccatcgcgac	agcgaagcag	gtgtaccgta	aaattggcgt	gaaagttgaa	caggccggta	7200
agcaggcctg	ggatcatcgc	cagtccacgt	ccaccgccga	aaaattaacg	cttttgctga	7260
cggcatccgg	tcaggcagtt	acttcccgga	tgaagacgta	tccaccccgt	cctgctcatc	7320
tctggcagcg	cccgatctag	ccgcatgcct	ttctctcagc	gtcgccctgaa	gtttagataa	7380

cgggtggcgcg tacagaaaac caaaggacac gcagccctct tttccctta cagcatgatg 7440
 catacgggtgg gccatgtata accgtttcag gtagcctttg cgcggtatgt agcgggaacgg 7500
 ccagcgctgg tgtaccagtc cgtcgtggac cataaaatac agtaaaccat aagcgggtcat 7560
 gcctgcacca atccactgga gcggccagat tcctgtactg ccgaagtaaa tcagggcaat 7620
 cgacacaatg gcgaatacca cggcatagag atcgttaact tcaaatacgc ctttacgcgg 7680
 ttcattgatgt gaaagatgcc agccccaacc ccagccgtgc atgatgtatt tatgtgccag 7740
 tgcagcaacc acttccatgc cgaccacggg gacaaacacg atcagggcat tccaaatcca 7800
 caacataatt tctcaagggc gaattcgcgg ggatcctcta gagtcgacct gcaggcatgc 7860
 aagcttggca ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca 7920
 acttaatcgc cttgcagcac atcccccttt cgccagctgg cgtaatagcg aagaggcccg 7980
 caccgatcgc ctttccaac agttgcgcag cctgaatggc gaatggcgct gatgtccggc 8040
 ggtgcttttg ccgttacgca ccaccccgtc agtagctgaa caggagggac agctgataga 8100
 aacagaagcc actggagcac ctcaaaaaca ccatcatata ctaaatacagt aagttggcag 8160
 catcaccgga cgcactttgc gccgaataaa tacctgtgac ggaagatcac ttcgcagaat 8220
 aaataaatcc tgggtgtccct gttgataccg ggaagccctg ggccaacttt tggcgaaaat 8280
 gagacgttga tcggcacgta agaggttcca actttcacca taatgaaata agatcactac 8340
 cgggctgatt ttttgagtta tcgagatfff caggagctaa ggaagctaaa atggagaaaa 8400
 aaatcactgg atataccacc gttgatatat cccaatggca tcgtaaagaa cattttgagg 8460
 catttcagtc agttgctcaa tgtacctata accagaccgt tcagctggat attacggcct 8520
 ttttaagac cgtaaagaaa aataagcaca agttttatcc ggcctttatt cacattcttg 8580
 cccgcctgat gaatgctcat ccggaattt 8609

<210> 65
 <211> 6329
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Plasmid pKD46

<400> 65
 catcgattta ttatgacaac ttgacggcta catcattcac tttttcttca caaccggcac 60
 ggaactcgct cgggctggcc ccgggtgcatt ttttaaatac ccgcgagaaa tagagttgat 120
 cgtcaaaacc aacattgcga ccgacgggtg cgatagggcat ccgggtggtg ctcaaaagca 180
 gcttcgcctg gctgatacgt tggtcctcgc gccagcttaa gacgctaata cctaactgct 240
 ggcggaaaag atgtgacaga cgcgacggcg acaagcaaac atgctgtgcg acgctggcga 300
 tatcaaaatt gctgtctgcc aggtgatcgc tgatgtactg acaagcctcg cgtacccgat 360
 tatccatcgg tggatggagc gactcgttaa tcgcttccat gcgccgcagt aacaattgct 420
 caagcagatt tatcgccagc agctccgaat agcgcccttc cccttgcccg gcgttaatga 480

tttgcccaaa	caggtcgctg	aaatgcggct	ggtgcgcttc	atccgggcga	aagaacccccg	540
tattggcaaa	tattgacggc	cagttaagcc	attcatgcc	gtaggcgcgc	ggacgaaagt	600
aaaccactg	gtgataccat	tcgcgagcct	ccggatgacg	accgtagtga	tgaatctctc	660
ctggcgggaa	cagcaaaata	tcacccggtc	ggcaaacaaa	ttctcgtccc	tgatTTTTca	720
ccacccccctg	accgcgaatg	gtgagattga	gaatataacc	tttcattccc	agcggtcgggt	780
cgataaaaaa	atcgagataa	ccgttggcct	caatcggcgt	taaacccgcc	accagatggg	840
cattaaacga	gtatcccggc	agcaggggat	cattttgctc	ttcagccata	cttttcatac	900
tcccgccatt	cagagaagaa	accaattgtc	catattgcat	cagacattgc	cgtcactgctg	960
tcttttactg	gctcttctcg	ctaaccaaac	cggtaacccc	gcttattaaa	agcattctgt	1020
aacaaagcgg	gaccaaagcc	atgacaaaaa	cgcgtaacaa	aagtgtctat	aatcacggca	1080
gaaaagtcca	cattgattat	ttgcacggcg	tcacactttg	ctatgccata	gcatttttat	1140
ccataagatt	agcggatcct	acctgacgct	ttttatcgca	actctctact	gtttctccat	1200
acccgttttt	ttgggaattc	gagctctaag	gaggttataa	aaaatggata	ttaatactga	1260
aactgagatc	aagcaaaagc	attcactaac	cccctttcct	gttttcctaa	tcagcccggc	1320
atttcgcggg	cgatattttc	acagctattt	caggagtcca	gccatgaacg	cttattacat	1380
tcaggatcgt	cttgaggctc	agagctgggc	gcgtcactac	cagcagctcg	cccgtgaaga	1440
gaaagaggca	gaactggcag	acgacatgga	aaaaggcctg	ccccagcacc	tgtttgaatc	1500
gctatgcac	gatcatttgc	aacgccacgg	ggccagcaaa	aaatccatta	cccgtgcgtt	1560
tgatgacgat	gttgagtttc	aggagcgc	ggcagaacac	atccggtaca	tggttgaaac	1620
cattgctcac	caccaggttg	atattgattc	agaggtataa	aacgaatgag	tactgcactc	1680
gcaacgctgg	ctgggaagct	ggctgaacgt	gtcggcatgg	attctgtcga	cccacaggaa	1740
ctgatcacca	ctcttcgcca	gacggcattt	aaaggtgatg	ccagcgatgc	gcagttcatc	1800
gcattactga	tcgttgccaa	ccagtacggc	cttaatccgt	ggacgaaaga	aattttacgcc	1860
tttctgata	agcagaatgg	catcgttccg	gtggtgggcg	ttgatggctg	gtcccgcac	1920
atcaatgaaa	accagcagtt	tgatggcatg	gactttgagc	aggacaatga	atcctgtaca	1980
tgccggattt	accgcaagga	ccgtaatcat	ccgatctgcg	ttaccgaatg	gatggatgaa	2040
tgccgccgcg	aaccattcaa	aactcgcgaa	ggcagagaaa	tcacggggcc	gtggcagtcg	2100
catcccaaac	ggatgttacg	tcataaagcc	atgattcagt	gtgcccgtct	ggccttcgga	2160
tttgctggta	tctatgacaa	ggatgaagcc	gagcgcattg	tcgaaaatac	tgcatacact	2220
gcagaacgct	agccggaacg	cgacatcact	ccggttaacg	atgaaaccat	gcaggagatt	2280
aacactctgc	tgatgcgcct	ggataaaaca	tgggatgacg	acttattgcc	gctctgttcc	2340
cagatatttc	gccgcgacat	tcgtgcatcg	tcagaactga	cacaggccga	agcagtaaaa	2400
gctcttggtg	tcctgaaaca	gaaagccgca	gagcagaagg	tggcagcatg	acaccggaca	2460
ttatcctgca	gcgtaccggg	atcgatgtga	gagctgtcga	acagggggat	gatgcgtggc	2520
acaaattacg	gctcggcgtc	atcaccgctt	cagaagttca	caacgtgata	gcaaaacccc	2580

gctccggaaa gaagtggcct gacatgaaaa tgtcctactt ccacaccctg cttgctgagg	2640
tttgcaccgg tgtggctccg gaagttaacg ctaaagcact ggcctgggga aaacagtacg	2700
agaacgacgc cagaaccctg tttgaattca cttccggcgt gaatgttact gaatccccga	2760
tcatctatcg cgacgaaagt atgcgtaccg cctgctctcc cgatggttta tgcagtgacg	2820
gcaacggcct tgaactgaaa tgcccgttta cctcccggga tttcatgaag ttccggctcg	2880
gtggtttcga ggccataaag tcagcttaca tggcccaggt gcagtacagc atgtgggtga	2940
cgcgaaaaaa tgcctggtag tttgccaact atgaccgcg tatgaagcgt gaaggcctgc	3000
attatgtcgt gattgagcgg gatgaaaagt acatggcgag ttttgacgag atcgtgccgg	3060
agttcatcga aaaaatggac gaggcactgg ctgaaattgg ttttgtatgt ggggagcaat	3120
ggcgtgacg catcctcagc ataatatccg ggtaggcgca atcactttcg tctactccgt	3180
tacaaagcga ggctgggtat ttcccggcct ttctgttatc cgaaatccac tgaaagcaca	3240
gcggctggct gaggagataa ataataaacg aggggctgta tgcacaaagc atcttctgtt	3300
gagttaagaa cgagtatcga gatggcacat agccttgctc aaattggaat caggtttgtg	3360
ccaataccag tagaaacaga cgaagaatcc atgggtatgg acagttttcc ctttgatatg	3420
taacggtgaa cagttgttct acttttgtt gttagtcttg atgcttctact gatagataca	3480
agagccataa gaacctcaga tccttccgta tttagccagt atgttctcta gtgtggttcg	3540
ttgtttttgc gtgagccatg agaacgaacc attgagatca tacttacttt gcatgtcact	3600
caaaaatfff gcctcaaaac tggtagactg aatttttgca gttaaagcat cgtgtagtgt	3660
ttttcttagt ccgttacgta ggtaggaatc tgatgtaatg gttgttggtta ttttgtcacc	3720
attcattfff atctggttgt tctcaagttc ggttacgaga tccatttgtc tatctagttc	3780
aacttgaaa atcaacgtat cagtcgggcg gcctcgctta tcaaccacca atttcatatt	3840
gctgtaagt tttaaatctt tacttattgg tttcaaaacc cattgggtta gccttttaaa	3900
ctcatggtag ttattttcaa gcattaacat gaacttaaat tcatcaaggc taatctctat	3960
atttgccttg tgagttttct tttgtgttag ttcttttaaat aaccactcat aaatcctcat	4020
agagtatttg ttttcaaaag acttaacatg ttccagatta tattttatga atttttttaa	4080
ctggaaaaga taaggcaata tctcttctact aaaaactaat tctaattttt cgcttgagaa	4140
cttggcatag tttgtccact ggaaaatctc aaagccttta accaaaggat tcctgatttc	4200
cacagttctc gtcacagct ctctgggtgc tttagctaata acaccataag cattttccct	4260
actgatgttc atcatctgag cgtattgggt ataagtgaac gataccgtcc gttctttcct	4320
tgtagggttt tcaatcgtgg gggtgagtag tgccacacag cataaaatta gcttggtttc	4380
atgctccgtt aagtcatagc gactaatcgc tagttcattt gctttgaaaa caactaattc	4440
agacatacat ctcaattggc ctaggtgatt ttaatcacta taccaattga gatgggctag	4500
tcaatgataa ttactagtcc ttttcctttg agttgtgggt atctgtaaat tctgctagac	4560
ctttgctgga aaacttgtaa attctgctag accctctgta aattccgcta gacctttgtg	4620

```

tgttttttttt gtttatattc aagtggttat aatttataga ataaagaaag aataaaaaaa 4680
gataaaaaga atagatccca gccctgtgta taactcacta ctttagtcag ttccgcagta 4740
ttacaaaagg atgtcgcaaa cgctgtttgc tcctctacaa aacagacctt aaaaccctaa 4800
aggcttaagt agcacctcgc caagctcggg tgccggccgca atcgggcaaa tcgctgaata 4860
ttccttttgt ctccgaccat caggcacctg agtcgctgtc tttttcgtga cattcagttc 4920
gctgcgctca cggctctggc agtgaatggg ggtaaattggc actacaggcg ctttttatgg 4980
attcatgcaa ggaaactacc cataatacaa gaaaagcccg tcacgggctt ctcagggcgt 5040
tttatggcgg gtctgctatg tgggtgctatc tgactttttg ctgttcagca gttcctgccc 5100
tctgattttc cagtctgacc acttcggatt atcccgtgac aggtcattca gactggctaa 5160
tgcacccagt aaggcagcgg tatcatcaac ggggtctgac gctcagtgga acgaaaactc 5220
acgttaaggg attttgggtca tgagattatc aaaaaggatc ttcacctaga tccttttaaa 5280
ttaaaaatga agttttaaat caatctaaag tatatatgag taaacttggt ctgacagtta 5340
ccaatgctta atcagtgagg cacctatctc agcgatctgt ctatttcgtt catccatagt 5400
tgccctgactc cccgtcgtgt agataactac gatacgggag ggcttaccat ctggccccag 5460
tgctgcaatg ataccgcgag acccacgctc accggctcca gatttatcag caataaacca 5520
gccagccgga agggccgagc gcagaagtgg tcctgcaact ttatccgcct ccatccagtc 5580
tattaattgt tgccgggaag ctagagtaag tagttcgcca gttaatagtt tgcgcaacgt 5640
tgttgccatt gctacaggca tcgtgggtgtc acgctcgtcg tttgggtatgg cttcattcag 5700
ctccggttcc caacgatcaa ggcgagttac atgatcccc atgttgtgca aaaaagcggg 5760
tagctccttc ggtcctccga tcgttgtcag aagtaagttg gccgcagtgt tatcactcat 5820
ggttatggca gactgcata attctcttac tgtcatgcca tccgtaagat gcttttctgt 5880
gactggtgag tactcaacca agtcattctg agaatagtgt atgcggcgac cgagttgctc 5940
ttgcccggcg tcaatacggg ataataccgc gccacatagc agaactttaa aagtgctcat 6000
cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc ttaccgctgt tgagatccag 6060
ttcgaatgaa cccactcgtg cacccaactg atcttcagca tcttttactt tcaccagcgt 6120
ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa aagggaataa gggcgacacg 6180
gaaatgttga atactcatac tcttcctttt tcaatattat tgaagcattt atcaggggta 6240
ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa taggggttcc 6300
gcgcacattt ccccgaaaag tgccacctg 6329

```

<210> 66
 <211> 3423
 <212> DNA
 <213> Artificial sequence

<220>
 <223> Plasmid pSUH5

<400> 66
 agattgcagc attacacgtc ttgagcgatt gtgtaggctg gagctgcttc gaagttccta 60

tactttctag agaataggaa cttcggaata ggaacttcaa gatccccctca cgctgccgca	120
agcactcagg gcgcaagggc tgctaaagga agcggaacac gtagaaagcc agtccgcaga	180
aacggtgctg accccggatg aatgtcagct actgggctat ctggacaagg gaaaacgcaa	240
gcgcaaagag aaagcaggta gcttgcagtg ggcttacatg gcgatatgcta gactgggcgg	300
ttttatggac agcaagcgaa ccggaattgc cagctggggc gccctctggt aagggttggga	360
agccctgcaa agtaaactgg atggctttct tgccgccaag gatctgatgg cgcaggggat	420
caagatctga tcaagagaca ggatgaggat cgtttcgcat gattgaacaa gatggattgc	480
acgcagggttc tccggccgct tgggtggaga ggctattcgg ctatgactgg gcacaacaga	540
caatcggtctg ctctgatgcc gccgtgttcc ggctgtcagc gcagggggcg ccggttcttt	600
ttgtcaagac cgacctgtcc ggtgccctga atgaactgca ggacgaggca gcgcggctat	660
cgtaggctggc cacgacgggc gttccttgcg cagctgtgct cgacgttgc actgaagcgg	720
gaagggactg gctgctattg ggcgaagtgc cggggcagga tctcctgtca tctcaccttg	780
ctcctgccga gaaagtatcc atcatggctg atgcaatgcg gcggctgcat acgcttgatc	840
cggctacctg cccattcgac caccaagcga aacatcgcat cgagcgagca cgtactcgga	900
tggaagccgg tcttgtcgat caggatgatc tggacgaaga gcatcagggg ctgcgccag	960
ccgaactgtt cgccaggctc aaggcgcgca tgcccgacgg cgaggatctc gtcgtgaccc	1020
atggcgatgc ctgcttgccg aatatcatgg tggaaaatgg ccgcttttct ggattcatcg	1080
actgtggccg gctgggtgtg gcggaccgct atcaggacat agcgttggt acccgtgata	1140
ttgctgaaga gcttggcggc gaatgggctg accgcttcct cgtgctttac ggtatcgccg	1200
ctcccgatcc gcagcgcac gccttctatc gccttcttga cgagttcttc tgagcgggac	1260
tctggggttc gaaatgaccg accaagcgac gcccaacctg ccatcacgag atttcgatcc	1320
caccgccgcc ttctatgaaa ggttgggctt cggaatcggt ttccgggacg ccggctggat	1380
gatcctccag cgcggggatc tcatgctgga gttcttcgcc caccacagct tcaaaagcgc	1440
tctgaagtcc ctatactttc tagagaatag gaacttcgga ataggaacta aggaggatat	1500
tcactataaa aataggcgta tcacgaggcc ctttcgtctt cacctcgaga aatcataaaa	1560
aatttatttg ctttgtgagc ggataacaat tataatagat tcaattgtga gcggataaca	1620
atttcacaca gaattcatta aagaggagaa attaaactcat atggaccatg gctaattccc	1680
atgtcagccg ttaagtgttc ctgtgtcact gaaaattgct ttgagaggct ctaagggtt	1740
ctcagtgcgt tacatccctg gcttgttgtc cacaaccgtt aaaccttaaa agctttaaaa	1800
gccttatata ttcttttttt tcttataaaa cttaaaacct tagaggctat ttaagttgct	1860
gatttatatt aattttattg ttcaaacatg agagcttagt acgtgaaaca tgagagctta	1920
gtacgttagc catgagagct tagtacgtta gccatgaggg tttagttcgt taaacatgag	1980
agcttagtac gttaaacadg agagcttagt acgtgaaaca tgagagctta gtacgtacta	2040
tcaacagggt gaactgcgga tcttgcggcc gcaaaaatta aaaatgaagt tttaaatcaa	2100

tctaaagtat	atatgagtaa	acttgggtctg	acagttacca	atgcttaatc	agtgaggcac	2160
ctatctcagc	gatctgtcta	tttcgttcat	ccatagttgc	ctgactcccc	gtcgtgtaga	2220
taactacgat	acgggagggc	ttaccatctg	gccccagtg	tgcaatgata	ccgcgagacc	2280
cacgctcacc	ggctccagat	ttatcagcaa	taaaccagcc	agccggaagg	gccgagcgca	2340
gaagtgggtcc	tgcaacttta	tccgcctcca	tccagtctat	taattgttgc	cgggaagcta	2400
gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	acaggcatcg	2460
tgggtgtcacg	ctcgtcgttt	ggatatggctt	cattcagctc	cggttcccaa	cgatcaaggc	2520
gagttacatg	atcccccatg	ttgtgcaaaa	aagcggttag	ctccttcggt	cctccgatcg	2580
ttgtcagaag	taagtggcc	gcagtgttat	cactcatggt	tatggcagca	ctgcataatt	2640
ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	tgggtgagtac	tcaaccaagt	2700
cattctgaga	atagtgtatg	cggcgaccga	gttgtctttg	cccggcgctca	atacgggata	2760
ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	tggaacacgt	tcttcggggc	2820
gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	actcgtgcac	2880
ccaactgatc	ttcagcatct	tttactttca	ccagcgtttc	tgggtgagca	aaaacaggaa	2940
ggcaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	ctcatactct	3000
tcctttttca	atattattga	agcatttatc	agggttattg	tctcatgagc	ggatacatat	3060
ttgaatgtat	ttagaaaaat	aaacaaatag	gggttccgcg	cacatttccc	cgaaaagtgc	3120
cacctgcac	gatggcccc	cgatggtagt	gtggggtctc	cccatgagag	agtagggaac	3180
tgccaggcat	caaataaaac	gaaaggctca	gtcgaaagac	tgggcctttc	gttttatctg	3240
ttgtttgtcg	gtgaacgctc	tcctgagtag	gacaaatccg	ccgggagcgg	atgtgaacgt	3300
tgcgaaagcaa	cggcccgag	gggtggcggc	aggacggccg	ccataaactg	ccaggcatca	3360
aattaagcag	aaggccatcc	tgacggatgg	cctttttg	tgccagtg	caagcttgca	3420
tgc						3423