**Title:** OPTIMIZED BACTERIAL HOST STRAINS OF METHYLOMONAS SP.16A

**Abstract:** Methanotrophic bacterial strains are provided that have been optimized for the production of carotenoid compounds through the down-regulation of one or more of the *crtN1*, *ald*, *crtN2* and *crtN3* genes of the carotenoid biosynthetic pathway. The resulting strains lack pigmented C₄₀ carotenoid compounds, and show an increase in the production of C₄₀ carotenoids. The use of the optimized host strains for the production of the C₄₀ carotenoids canthaxanthin and astaxanthin is also described.

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.
TITLE
OPTIMIZED BACTERIAL HOST STRAINS OF METHYLOMONAS SP. 16A

This application claims the benefit of U.S. Provisional Application No. 60/527,083 filed December 3, 2003 and U.S. Provisional Application No. 60/527,877 filed December 8, 2003.

FIELD OF THE INVENTION
The invention relates to the field of microbiology and molecular biology. More specifically, the invention relates to methanotrophic bacterial host strains optimized for the production of carotenoid compounds through the down-regulation of various genes in the carotenoid biosynthetic pathway.

BACKGROUND OF THE INVENTION
There are a number of microorganisms that utilize single carbon substrates as their sole energy source. Such microorganisms are referred to herein as “C1 metabolizers”. These organisms are characterized by the ability to use carbon substrates lacking carbon to carbon bonds as a sole source of energy and biomass. All C1 metabolizing microorganisms are generally classified as methylo trophs. Methylo trophs may be defined as any organism capable of oxidizing organic compounds that do not contain carbon-carbon bonds. However, facultative methylo trophs, obligate methylo trophs, and obligate methanotrophs are all various subsets of methylo trophs. Specifically:

- Facultative methylo trophs have the ability to oxidize organic compounds which do not contain carbon-carbon bonds, but may also use other carbon substrates such as sugars and complex carbohydrates for energy and biomass;
- Obligate methylo trophs are those organisms which are limited to the use of organic compounds that do not contain carbon-carbon bonds for the generation of energy; and
- Obligate methanotrophs are those obligate methylo trophs that have the distinct ability to oxidize methane.

The ability of obligate methanotrophic bacteria to use methane as their sole source of carbon and energy under ambient conditions, in conjunction with the abundance of methane, makes the biotransformation of methane a potentially unique and valuable process. As such, several have attempted to harness the unique natural abilities of these organisms for commercial applications. For example, the commercial applications for
the biotransformation of methane have historically fallen broadly into three categories: 1.) production of single cell protein; 2.) epoxidation of alkenes for production of chemicals; and 3.) biodegradation of chlorinated pollutants. Of these, only epoxidation of alkenes has experienced some commercial success; however, the success has been limited due to low product yields, toxicity of products, and the large amount of cell mass required to generate products. Large-scale protein production from methane, termed single cell protein (or "SCP"), has been technically feasible and commercialized at large scale (Villadsen, J., *Recent Trends Chem. React. Eng.*, [Proc. Int. Chem. React. Eng. Conf.], 2nd ed.; Kulkarni, B. D., Mashelkar, R. A., and Sharma, M. M., Eds.; Wiley East: New Delhi, India (1987); Vol 2, pp 320-33). However, SCP has not been economically successful thus far due to the relatively high cost of producing microbial protein, as compared to agriculturally derived protein (i.e., soy protein).

This makes SCP a relatively low value product whose economic production cannot tolerate heavy bioprocessing costs. Thus, the yield of the methanotrophic strain used for producing SCP may be critical to the overall economic viability of the process. Microbial biomass produced by methanotrophic bacteria is typically very high in protein content (~70-80% by weight), which can restrict the direct use of this protein to certain types of animal feed.

In addition to the synthesis of SCP, methanotrophic cells can further build the oxidation products of methane (i.e., methanol and formaldehyde) into complex molecules such as carbohydrates and lipids. For example, under certain conditions methanotrophs are known to produce exopolysaccharides (US 6,537,786; US 6,689,601; Ivanova et al., *Mikrobiologiya*, 57(4):600-5 (1988); Kilbane, John J., II, Gas, *Oil, Coal, Environ. Biotechnol.* 3, [Pap. IGT's Int. Symp.], 3rd ed., Meeting Date 1990; Akin, C. and J. Smith, Eds; IGT: Chicago, IL (1991); pp 207-26). Similarly, methanotrophs are known to accumulate both isoprenoid compounds and carotenoid pigments of various carbon lengths (US 6,660,507; US 6,689,601; Urakami et al., *J. Gen. Appl. Microbiol.*, 32(4):317-41 (1986)).

Most recently, the natural abilities of methanotrophic organisms have been extended by the advances of genetic engineering. Odom et al. have investigated *Methylomonas* sp. 16a as a microbial platform of choice for production of a variety of materials beyond single cell protein, including carbohydrates, pigments, terpenoid compounds and aromatic compounds (US 6,689,601 and US 09/941947, herein incorporated entirely by
This particular pink-pigmented methanotrophic bacterial strain is capable of efficiently using either methanol or methane as a carbon substrate, is metabolically versatile in that it contains multiple pathways for the incorporation of carbon from formaldehyde into 3-carbon units, and is capable of genetic exchange with donor species such as *Escherichia coli* via bacterial conjugation. Thus, *Methylomonas* sp. 16a can be engineered to produce new classes of products other than those naturally produced from methane. Further advancement in the metabolic engineering of this particular host organism for production of various commercial products on an economic scale, however, requires some optimization of the host organism. Specifically, it would be desirable to knockout the native carotenoid pathway of the organism leading to the production of pink-pigmented C\textsubscript{30} carotenoids, increasing the available carbon flux directed toward the products of interest. These modified host organisms should preferably lack antibiotic markers, since the presence of antibiotic resistance genes in the modified host organism could be undesirable in many food and feed applications. The problem to be solved, therefore, is to develop an optimized non-pigmented *Methylomonas* sp. 16a bacterial host organism lacking antibiotic markers for production of various commercial products on an economic scale.

The present problem has been solved through the development of a suite of optimized non-pigmented *Methylomonas* sp. 16a bacterial host organisms, each lacking antibiotic markers. These bacterial hosts were created by investigation of allelic exchange mutations within the native *crt* gene cluster (comprising the *crtN1*, *ald*, and *crtN2* genes) and the *crtN3* gene of *Methylomonas* sp. 16a, each of which is associated in the biosynthesis of native C\textsubscript{30} carotenoids in the organism. An efficient means of generating defined mutants by homologous recombination permitted transformants that have undergone allelic exchange to be selected based on a positive selection strategy. This methodology also enabled production of "markerless" transformants and permitted multiple rounds of mutation to be performed.

**SUMMARY OF THE INVENTION**

The present invention provides optimized C1 metabolizing host cells useful for the production of carotenoid compounds and particularly C\textsubscript{40} carotenoids. In one embodiment, the optimized host cells of the invention are preferably methanotrophs comprised of a functional Embden-Meyerhof carbon pathway, making them particularly effective in
the use of carbon for carotenoid production. Additionally, the host cells have mutations in at least one of the \textit{crtN1}, \textit{ald}, \textit{crtN2}, and \textit{crtN3} genes, resulting in the down-regulation and/or complete disruption of the genes. Disruption of one or more of these genes increased production of downstream carotenoids through the re-allocation of carbon in the cells.

Accordingly, the invention provides a high growth methanotrophic bacterial strain which:

a) grows on a C1 carbon substrate selected from the group consisting of methanol and methane;

b) comprises a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate dependent phosphofructokinase enzyme; and

c) possesses a disruption such that at least one endogenous gene selected from the group consisting of \textit{crtN1}, \textit{ald}, \textit{crtN2}, and \textit{crtN3} is down-regulated.

In a specific embodiment, the invention provides a \textit{Methylomonas} sp. strain optimized for the production of C₄₀ carotenoids selected from the group consisting of: MWM1000 (\textit{Δald (crtN1)}), MWM1100 (\textit{Δcrt cluster promoter}), MWM1400 (\textit{ΔcrtN1}), MWM1200 (\textit{Δcrt cluster promoter + ΔcrtN3}), MWM1600 (\textit{Δcrt cluster}), MWM1800 (\textit{ΔcrtN1 + ΔcrtN3}), MWM1900 (\textit{Δcrt cluster + ΔcrtN3}), and MWM1300 (\textit{Δald(crtN1) + ΔcrtN3}).

In a preferred embodiment, the invention provides a method for the production of C₄₀ carotenoids, comprising:

a) providing the high growth methanotrophic bacterial strain of the invention comprising at least one gene encoding an enzyme of the C₄₀ carotenoid biosynthesis pathway;

b) growing the bacterial strain of step (a) under conditions wherein the gene encoding an enzyme of the C₄₀ carotenoid biosynthesis pathway is expressed producing a C₄₀ carotenoid;

and

c) optionally recovering the C₄₀ carotenoid.

In an alternate embodiment, the invention provides a method for the production of canthaxanthin comprising:

(a) providing a high growth methanotrophic bacterial strain comprising:

i) a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate dependent phosphofructokinase enzyme; and
ii) a disruption in at least one endogenous gene selected from the group consisting of \textit{crtN1}, \textit{ald}, \textit{crtN2}, and \textit{crtN3};

iii) a \textit{crtE-idi-crtY-crtI-crtB} gene cluster having the sequence as set forth in SEQ ID NO:6; and

iv) a codon-optimized \( \beta \)-carotene ketolase gene having the sequence as set forth in SEQ ID NO:7; wherein said high growth methanotrophic bacterial strain grows on a C1 carbon substrate selected from the group consisting of methanol and methane;

(b) growing the high growth methanotrophic bacterial strain of step (a) under conditions wherein the \textit{crtE-idi-crtY-crtI-crtB} gene cluster genes and the codon-optimized \( \beta \)-carotene ketolase gene are expressed producing canthaxanthin; and

(c) optionally recovering the canthaxanthin.

Similarly, the invention provides a method for the production of astaxanthin comprising:

(a) providing a high growth methanotrophic bacterial strain comprising:

i) a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate dependent phosphofructokinase enzyme; and

ii) a disruption in at least one endogenous gene selected from the group consisting of \textit{crtN1}, \textit{ald}, \textit{crtN2}, and \textit{crtN3};

iii) a \textit{crtE-idi-crtY-crtI-crtB} gene cluster having the sequence as set forth in SEQ ID NO:6;

iv) a codon-optimized \( \beta \)-carotene ketolase gene having the sequence as set forth in SEQ ID NO:7; and

(v) a codon-optimized \( \beta \)-carotene hydroxylase gene having the sequence as set forth in SEQ ID NO:8; wherein said high growth methanotrophic bacterial strain grows on a C1 carbon substrate selected from the group consisting of methanol and methane;

(b) growing the high growth methanotrophic bacterial strain of step (a) under conditions wherein the \textit{crtE-idi-crtY-crtI-crtB} gene cluster genes, the codon-optimized \( \beta \)-carotene
ketolase gene, and the codon-optimized β-carotene hydroxylase gene are expressed producing astaxanthin; and
(c) optionally recovering the astaxanthin.

Additionally the invention provides an isolated nucleic acid molecule selected from the group consisting of:
(a) an isolated nucleic acid molecule as set forth in SEQ ID NO:7 which encodes a β-carotene ketolase enzyme; and
(b) an isolated nucleic acid molecule that is completely complementary to (a).

Similarly, the invention provides an isolated nucleic acid molecule which encodes a β-carotene ketolase enzyme as set forth in SEQ ID NO:33 wherein at least 99 codons are codon-optimized for expression in Methylomonas sp.

In another embodiment the invention provides an isolated nucleic acid molecule selected from the group consisting of:
(a) an isolated nucleic acid molecule as set forth in SEQ ID NO:8 which encodes a β-carotene hydroxylase enzyme; and
(b) an isolated nucleic acid molecule that is completely complementary to (a).

Additionally, the invention provides an isolated nucleic acid molecule which encodes a β-carotene hydroxylase enzyme as set forth in SEQ ID NO:34 wherein at least 73 codons are codon-optimized for expression in Methylomonas sp.

BRIEF DESCRIPTION OF THE FIGURES, SEQUENCE DESCRIPTIONS, AND BIOLOGICAL DEPOSITS

Figure 1 shows the upper isoprenoid and lower carotenoid biosynthetic pathways.

Figure 2 illustrates the position of the crtN3 gene within Methylomonas' genome, with respect to other genes.

Figure 3 shows results of a Clustal W analysis of the Methylomonas sp. 16a carotenoid genes crtN1, crtN2, crtN3, and sqs.

Figure 4 is a schematic diagram illustrating the two-step selection protocol used for identifying "markerless" double-crossover mutants in C1 metabolizing bacteria.

Figure 5 is a plasmid map of pGP704.
Figure 6 shows a comparison of the codon-optimized \textit{crtW} gene (SEQ ID NO:7) with the wild-type \textit{crtW} gene (SEQ ID NO:31) from \textit{Agrobacterium aurantiacum}.

Figure 7 shows a comparison of the codon-optimized \textit{crtZ} gene (SEQ ID NO:8) with the wild-type \textit{crtZ} gene (SEQ ID NO:32) from \textit{Agrobacterium aurantiacum}.

Figure 8 shows representations of the carotenoid biosynthesis gene clusters cloned in plasmids pDCQ307, pDCQ333, pDCQ324, and pDCQ334.

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

The following sequences conform 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 (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.

SEQ ID NO:1 is the nucleotide sequence of the 16S rRNA gene from \textit{Methylomonas sp.} 16a.

SEQ ID NO:2 is the nucleotide sequence of the \textit{crtN1} gene from \textit{Methylomonas sp.} 16a.

SEQ ID NO:3 is the nucleotide sequence of the \textit{ald} gene from \textit{Methylomonas sp.} 16a.

SEQ ID NO:4 is the nucleotide sequence of the \textit{crtN2} gene from \textit{Methylomonas sp.} 16a.

SEQ ID NO:5 is the nucleotide sequence of the \textit{crtN3} gene from \textit{Methylomonas sp.} 16a.

SEQ ID NO:6 is the nucleotide sequence of the \textit{crtE-idi-crtY-crtL-crtB} gene cluster from \textit{Pantoea agglomerans} DC404 (US 10/808807).

SEQ ID NO:7 is the nucleotide sequence of the codon-optimized \(\beta\)-carotene ketolase gene from \textit{Agrobacterium aurantiacum}.

SEQ ID NO:8 is the nucleotide sequence of the codon-optimized \(\beta\)-carotene hydroxylase gene from \textit{Agrobacterium aurantiacum}.
SEQ ID NOs:9 and 10 are the nucleotide sequences of primers Drdl/npr-sacB and ThhIII/npr-sacB, respectively, used for amplification of the npr-sacB cassette from plasmid pBE83, as described in Example 2.

SEQ ID NOs:11-30 and 48-53 are the nucleotide sequences of primers used for cloning of the carotenoid deletion fragments, as described in Example 3 and for confirmation of the mutant constructs as describe in Example 5.

SEQ ID NO:31 is the nucleotide sequence of the wild-type β-carotene ketolase gene from Agrobacterium aurantiacum.

SEQ ID NO:32 is the nucleotide sequence of the wild-type β-carotene hydroxylase gene from Agrobacterium aurantiacum.

SEQ ID NO:33 is the amino acid sequence of the β-carotene ketolase enzyme from Agrobacterium aurantiacum.

SEQ ID NO:34 is the amino acid sequence of the β-carotene hydroxylase enzyme from Agrobacterium aurantiacum.

SEQ ID NO:35 is the nucleotide sequence of the crtE gene from Pantoea stewartii.

SEQ ID NO:36 is the nucleotide sequence of the crtYIB gene cluster from Pantoea stewartii.

SEQ ID NOs:37-40 are the nucleotide sequences of primers used to construct the canthaxanthin expression plasmid pDCQ307, as described in Example 9.

SEQ ID NO:41 is the nucleotide sequence of the crtEidiYIBZ gene cluster from Pantoea agglomerans.

SEQ ID NOs:42 and 43 are the nucleotide sequences of primers used to construct the canthaxanthin expression plasmid pDCQ333, as described in Example 11.

SEQ ID NOs:44-47 are the nucleotide sequences of primers used to construct the astaxanthin expression plasmids pDCQ324 and pDCQ334, as described in Example 12.

Applicants made the following biological deposit under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purposes of Patent Procedure:

<table>
<thead>
<tr>
<th>Depositor Identification Reference</th>
<th>International Depository Designation</th>
<th>Date of Deposit</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Methylomonas</em> 16a</td>
<td>ATCC PTA 2402</td>
<td>August 22, 2000</td>
</tr>
</tbody>
</table>
As used herein, “ATCC” refers to the American Type Culture Collection International Depository Authority located at ATCC, 10801 University Blvd., Manassas, VA 20110-2209, U.S.A. The “International Depository Designation” is the accession number to the culture on deposit with ATCC.

The listed deposit 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**

**Definitions**

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

As used herein, the term “comprising” means the presence of the stated features, integers, steps, or components as referred to in the claims, but that it does not preclude the presence or addition of one or more other features, integers, steps, components or groups thereof.

“Open reading frame” is abbreviated ORF.

“Polymerase chain reaction” is abbreviated PCR.

“High Performance Liquid Chromatography” is abbreviated HPLC.

“Kanamycin” is abbreviated Kan.

“Ampicillin” is abbreviated Amp.

The term “isoprenoid compound” refers to compounds formally derived from isoprene (2-methylbuta-1,3-diene; CH$_2$=C(CH$_3$)CH=CH$_2$), the skeleton of which can generally be discerned in repeated occurrence in the molecule. These compounds are produced biosynthetically via the isoprenoid pathway beginning with isopentenyl pyrophosphate (IPP) and formed by the head-to-tail condensation of isoprene units, leading to molecules which may be, for example, of 5, 10, 15, 20, 30, or 40 carbons in length.

The term “carotenoid biosynthetic pathway” refers to those genes comprising members of the upper isoprenoid pathway and/or lower carotenoid biosynthetic pathway, as illustrated in Figure 1.

The terms “upper isoprenoid pathway” and “upper pathway” are used interchangeably and refer to enzymes involved in converting pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate (FPP). Genes
encoding these enzymes include, but are not limited to: the “dxs” gene (encoding 1-deoxyxylulose-5-phosphate synthase); the “dxr” gene (encoding 1-deoxyxylulose-5-phosphate reductoisomerase); the “ispD” gene (encoding a 2C-methyl-D-erythritol cytidyltransferase enzyme; also known as ygbP); the “ispE” gene (encoding 4- diphosphocytidyl-2-C- methylerythritol kinase; also known as ychB); the “ispF” gene (encoding a 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; also known as ygbB); the “pyrG” gene (encoding a CTP synthase); the “lytB” gene involved in the formation of dimethylallyl diphosphate; the “gcpE” gene involved in the synthesis of 2-C-methyl-D-erythritol 4-phosphate; the “idi” gene (responsible for the intramolecular conversion of IPP to dimethylallyl pyrophosphate); and the “ispA” gene (encoding geranyltransferase or farnesyl diphosphate synthase) in the isoprenoid pathway.

The terms “lower carotenoid biosynthetic pathway” and “lower pathway” will be used interchangeably and refer to those enzymes which convert FPP to a suite of carotenoids. These include those genes and gene products that are involved in the immediate synthesis of either diapophytoene (whose synthesis represents the first step unique to biosynthesis of C₃₀ carotenoids) or phytoene (whose synthesis represents the first step unique to biosynthesis of C₄₀ carotenoids). All subsequent reactions leading to the production of various C₃₀-C₄₀ carotenoids are included within the lower carotenoid biosynthetic pathway. These genes and gene products comprise all of the “crt” genes including, but not limited to: crtM, crtN, crtN2, crtE, crtX, crtY, crtI, crtB, crtR, crtZ, crtW, crtO, crtA, crtC, crtD, crtF, and crtU. Finally, the term “lower carotenoid biosynthetic enzyme” is an inclusive term referring to any and all of the enzymes in the present lower pathway including, but not limited to: CrtM, CrtN, CrtN2, CrtE, CrtX, CrtY, CrtI, CrtB, CrtR, CrtZ, CrtW, CrtO, CrtA, CrtC, CrtD, CrtF, and CrtU.

The term “carotenoid” refers to a class of hydrocarbons having a conjugated polyene carbon skeleton formally derived from isoprene. This class of molecules is composed of C₃₀ diapocarotenoids and C₄₀ carotenoids and their oxygenated derivatives; and, these molecules typically have strong light absorbing properties.

“C₃₀ diapocarotenoids” consist of six isoprenoid units joined in such a manner that the arrangement of isoprenoid units is reversed at the center of the molecule so that the two central methyl groups are in a 1,6-positional relationship and the remaining nonterminal methyl groups are in
a 1,5-positional relationship. All C₃₀ carotenoids may be formally derived from the acyclic C₃₀H₄₂ structure (Formula I below, hereinafter referred to as "diapophytoene"), having a long central chain of conjugated double bonds, by: (i) hydrogenation (ii) dehydrogenation, (iii) cyclization, (iv) oxidation, (v) esterification/ glycosylation, or any combination of these processes.

Formula I

"Tetraterpenes" or "C₄₀ carotenoids consist of eight isoprenoid units joined in such a manner that the arrangement of isoprenoid units is reversed at the center of the molecule so that the two central methyl groups are in a 1,6-positional relationship and the remaining nonterminal methyl groups are in a 1,5-positional relationship. All C₄₀ carotenoids may be formally derived from the acyclic C₄₀H₅₂ structure. Non-limiting examples of C₄₀ carotenoids include: phytoene, lycopene, β-carotene, zeaxanthin, astaxanthin, and canthaxanthin.

The term "CrtE" refers to a geranylgeranyl pyrophosphate synthase enzyme encoded by the crtE gene and which converts trans-trans-farnesyl diphtophosphate and isopentenyl diphosphate to pyrophosphate and geranylgeranyl diphtophosphate.

The term "Idi" refers to an isopentenyl diphosphate isomerase enzyme (E.C. 5.3.3.2) encoded by the idi gene.

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

The term "CrtI" refers to a phytoene desaturase enzyme encoded by the crtI gene. CrtI converts phytoene into lycopene via the intermediaries of phytofluene, ζ-carotene and neurosporene by the introduction of 4 double bonds.

The term "CrtB" refers to a phytoene synthase enzyme encoded by the crtB gene which catalyzes the reaction from prephytoene diphosphate to phytoene.

The term "CrtZ" refers to a carotenoid hydroxylase enzyme (e.g. β-carotene hydroxylase) encoded by the crtZ gene which catalyzes a hydroxylation reaction. The oxidation reaction adds a hydroxyl group to cyclic carotenoids having a β-ionone type ring. This reaction converts
cyclic carotenoids, such as β-carotene or canthaxanthin, into the
hydroxylated carotenoids zeaxanthin or astaxanthin, respectively.
Intermediates in the process typically include β-cryptoxanthin and
adonirubin. It is known that CrtZ hydroxylases typically exhibit substrate
flexibility, enabling production of a variety of hydroxylated carotenoids
depending upon the available substrates.

The term "CrtW" refers to a β-carotene ketolase enzyme encoded
by the crtW gene which catalyzes an oxidation reaction where a keto
group is introduced on the β-ionone type ring of cyclic carotenoids. This
reaction converts cyclic carotenoids, such as β-carotene or zeaxanthin,
into the ketocarotenoids canthaxanthin or astaxanthin, respectively.
Intermediates in the process typically include echinenone and
adonixanthin. It is known that CrtW ketolases typically exhibit substrate
flexibility.

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

The term "crtE-idi-crtY-crtl-crtB" or "crtE-idi-crtYIB" gene
cluster refers to a molecule having the following genetic organization: the
crtE, idi, crtY, crtl, and crtB genes are clustered in the order stated.

As used here, the term “C₁ carbon substrate” refers to any carbon-
containing molecule that lacks a carbon-carbon bond. Non-limiting
examples are methane, methanol, formaldehyde, formic acid, formate,
methylated amines (e.g., mono-, di-, and tri-methyl amine), methylated
thiols, and carbon dioxide. In another embodiment, the C₁ carbon
substrates is selected from the group consisting of methanol and/or
methane.

The term “C₁ metabolizer” refers to a microorganism that has the
ability to use a single carbon substrate as its sole source of energy and
biomass. C₁ metabolizers will typically be methylotrophs and/or
methanotrophs.

The term “C₁ metabolizing bacteria” refers to bacteria that have the
ability to use a single carbon substrate as their sole source of energy and
biomass. C₁ metabolizing bacteria, a subset of C₁ metabolizers, will
typically be methylotrophs and/or methanotrophs.

The term “methylotroph” means an organism capable of oxidizing
organic compounds that do not contain carbon-carbon bonds. Where the
methylotroph is able to oxidize CH₄, the methylotroph is also a
methanotroph. In one embodiment, the methylotroph uses methanol and/or methane as its primary carbon source.

The term "methanotroph" or "methanotrophic bacteria" means a prokaryote capable of utilizing methane as its primary source of carbon and energy. Complete oxidation of methane to carbon dioxide occurs by aerobic degradation pathways. Typical examples of methanotrophs useful in the present invention include (but are not limited to) the genera *Methylococcales, Methylobacter, Methylococcus*, and *Methylomonas*. In one embodiment, the methanotrophic bacteria uses methane and/or methanol as its primary carbon source.

The term "high growth methanotrophic bacterial strain" refers to a bacterium capable of growth with methane and/or methanol as the sole carbon and energy source and which possesses a functional Embden-Meyerhof carbon flux pathway, resulting in a high rate of growth and yield of cell mass per gram of C1 substrate metabolized (US 6,689,601). The specific "high growth methanotrophic bacterial strain" described herein is referred to as "*Methylomonas* 16a", "16a" or "*Methylomonas* sp. 16a", which terms are used interchangeably and which refer to the *Methylomonas* strain used in the present invention.

The terms "crtN1 gene cluster", "C30 crt gene cluster", "crt gene cluster", and "endogenous *Methylomonas* crt gene cluster" refer to an operon comprising *crtN1*, *ald*, and *crtN2* genes that is active in the native carotenoid biosynthetic pathway of *Methylomonas* sp. 16a.

The term "CrtN1" refers to an enzyme encoded by the *crtN1* gene, active in the native carotenoid biosynthetic pathway of *Methylomonas sp. 16a*. This gene is located within an operon comprising *crtN2* and *ald*.

The term "ALD" refers to an enzyme (an aldehyde dehydrogenase) encoded by the *ald* gene, active in the native carotenoid biosynthetic pathway of *Methylomonas sp. 16a*. This gene is located within an operon comprising *crtN1* and *crtN2*. As used herein, the gene and gene product of the *ald* gene may be optionally referred to as "aldehyde dehydrogenase".

The term "CrtN2" refers to an enzyme encoded by the *crtN2* gene, active in the native carotenoid biosynthetic pathway of *Methylomonas sp. 16a*. This gene is located within an operon comprising *crtN1* and *ald*.

The term "CrtN3" refers to an enzyme encoded by the *crtN3* gene, active in the native carotenoid biosynthetic pathway of *Methylomonas sp. 16a*. This gene is not located within the *crt* gene cluster; instead this gene
is present in a different location within the *Methylomonas* genome, as illustrated in Figure 2.

The term "Sqs" refers to the squalene dehydrogenase enzyme encoded by the sqs gene.

The term "pigmentless" or "white mutant" refers to a *Methylomonas* sp. 16a bacterium wherein the native pink pigment (e.g., a C<sub>30</sub> carotenoid) is not produced. Thus, the bacterial cells appear white in color, as opposed to pink.

The term "positive selection" means a selection method that enables only those cells that carry a DNA insert integrated at a specific chromosomal location to grow under particular conditions. In contrast, negative selection is based on selection methods whereby only those individuals that do not possess a certain character (e.g., cells that do not carry a DNA insert integrated at a specific chromosomal location) are selected.

The term "homologous recombination" refers to the exchange of DNA fragments between two DNA molecules (during cross over). The fragments that are exchanged are flanked by sites of identical nucleotide sequences between the two DNA molecules (i.e., homologous DNA regions). Homologous recombination is the most common means for generated genetic diversity in microbes.

The term "chromosomal integration" means that a chromosomal integration vector becomes congruent with the chromosome of a microorganism through recombination between homologous DNA regions on the chromosomal integration vector and within the chromosome.

The term "chromosomal integration vector" means an extra-chromosomal vector that is capable of integrating into the host's genome through homologous recombination.

The term "suicide vector" or "positive selection vector" refers to a type of chromosomal integration vector that is capable of replicating in one host but not in another. Thus, the vector is conditional for its replication.

The terms "single-crossover event" and "plasmid integration" are used interchangeably and mean the incorporation of a chromosomal integration vector into the genome of a host via homologous recombination between regions of homology between DNA present within the chromosomal integration vector and the host's chromosomal DNA. A "single-crossover mutant" refers to a cell that has undergone a single-crossover event.
The terms “double-crossover event”, “allelic exchange” and “gene replacement” are used interchangeably and mean the homologous recombination between a DNA region within the chromosomal integration vector and a region within the chromosome that results in the replacement of the functional chromosomal nucleotide sequence of interest (i.e., chr-NSI) with a homologous plasmid region (i.e., the replacement nucleotide sequence of interest, or re-NSI) (Figure 4). A “double-crossover mutant” or “allelic exchange mutant” is the result of a double-crossover event. This mutant can be generated by two simultaneous reciprocal breakage and reunion events between the same two DNA fragments; alternatively, a double-crossover mutant can be the result of two single-crossovers that occur non-simultaneously.

The term “chromosomal nucleotide sequence of interest” or “chr-NSI” refers to a specific chromosomal sequence that is targeted for homologous recombination. In a preferred embodiment, the chr-NSI encodes one or more of the native \textit{crtN1}, \textit{ald}, \textit{crtN2}, or \textit{crtN3} genes of \textit{Methylomonas sp. 16a}; alternatively, the chr-NSI corresponds to the promoter driving the \textit{Methylomonas sp. 16a crt} gene cluster.

The term “replacement nucleotide sequence of interest” or “re-NSI” refers to a nucleotide sequence of interest that is cloned into a chromosomal integration vector for the purpose of inducing homologous recombination with a chromosomal sequence. The re-NSI is modified with respect to chr-NSI by the addition, deletion, or substitution of at least one nucleotide. Sufficient homology must exist, however, between the two nucleotide sequences of interest to enable homologous recombination to occur. For the purposes herein, re-NSI will enable production of a transformed \textit{Methylomonas sp. 16a} having a deletion in at least one of the \textit{crtN1}, \textit{ald}, \textit{crtN2}, or \textit{crtN3} genes, or a deletion of the promoter driving the endogenous \textit{crt} gene cluster.

The term “genetic marker” or “selectable marker” means a phenotypic trait that can be visualized under special conditions. For example, an antibiotic resistance marker serves as a useful selectable marker, since it enables detection of cells which are resistant to the antibiotic, when cells are grown on media containing that particular antibiotic.

The term “markerless mutants” or “markerless transformants” refers to an allelic exchange mutant, wherein the mutant allele does not carry a genetic marker.
The term "SacB" means a *Bacillus* encoded protein that catalyzes the conversion of sucrose into levan, a product that is toxic to most Gram-negative microorganisms. The term "sacB" means a gene that encodes the "SacB" protein.

A "nucleic acid" is a polymeric compound comprised of covalently linked subunits called nucleotides. Nucleic acids include polyribonucleic acid (RNA) and polynucleotides (DNA), both of which may be single-stranded or double-stranded. DNA includes cDNA, genomic DNA, synthetic DNA, and semi-synthetic DNA.

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.

A nucleic acid fragment is "hybridizable" to another nucleic acid fragment, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid fragment can anneal to the other nucleic acid fragment under the appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd* ed., Cold Spring Harbor Laboratory: Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference).

The term "oligonucleotide" refers to a nucleic acid, generally of about at least 18 nucleotides, that is hybridizable to a genomic DNA molecule, a cDNA molecule, or an mRNA molecule.

The term "complementary" is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid fragments that are complementary to the complete sequences as reported in the accompanying Sequence Listing.

"Gene" refers to a nucleic acid fragment that expresses a specific protein. It may or may not include regulatory sequences preceding (5' non-coding 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 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 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.

“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 procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines.

“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 gene segments that are then enzymatically assembled to construct the entire gene. 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 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.

The term “homolog”, as applied to a gene, means any gene derived from the same or a different microbe having the same function. A homologous gene may have significant sequence similarity.

“Coding sequence” or “coding region of interest” refers to a DNA sequence that codes for a specific amino acid sequence.

The term “codon optimized” as it refers to genes or coding regions of nucleic acid molecules for transformation of various hosts, refers to the alteration of codons in the gene or coding regions of the nucleic acid molecules to reflect the typical codon usage of the host organism without altering the polypeptide for which the DNA codes. Within the context of
the present invention genes and DNA coding regions are codon optimized for optimal expression in *Methylomonas* sp.16a.

“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, RNA processing sites, effector binding sites, and stem-loop structures.

“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 at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. 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 sequences encoding regulatory signals capable of affecting mRNA processing or gene expression.

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., 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.

“Transformation” refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. In the present invention, the host cell’s genome is comprised of chromosomal and extrachromosomal (e.g. plasmid) genes. Host
organisms containing the transformed nucleic acid fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

"Conjugation" refers to a particular type of transformation in which a unidirectional transfer of DNA (e.g., from a bacterial plasmid) occurs from one bacterium cell (i.e., the "donor") to another (i.e., the "recipient"). The process involves direct cell-to-cell contact.

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 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.

"Transformation cassette" refers to a specific vector containing a foreign gene and having elements (in addition to the foreign gene) that facilitate 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 expression of that gene in a foreign host.

The term "altered biological activity" refers to an activity, associated with a protein encoded by a nucleotide sequence which can be measured by an assay method, where that activity is either greater than or less than the activity associated with the native sequence. "Enhanced biological activity" refers to an altered activity that is greater than that associated with the native sequence. "Diminished biological activity" is an altered activity that is less than that associated with the native sequence.

As used herein, the term "down-regulated" refers to a gene that has been mutated and/or disrupted such that the expression of the gene is less than that associated with the native sequence. In another embodiment, down-regulated includes elimination of the gene's expression (i.e. gene knockout). As used herein, the symbol "Δ" will be used to denote a mutation in the specified coding sequence and/or promoter wherein at least a portion (up to and including all) of said coding sequence and/or promoter has been disrupted by a deletion. In another embodiment, the disruption can occur by optionally inserting a nucleic acid molecule into the native sequence whereby the expression of one or more genes is down-regulated (either partially or completely).
The term "MWM1000 (Δald/crtN1)" refers to a mutant of Methylomonas sp. 16a in which the ald and crtN1 genes have been disrupted.

The term "MWM1100 (Δcrt cluster promoter)" refers to a mutant of Methylomonas sp. 16a in which the crt gene cluster promoter has been disrupted.

The term "MWM1200 (Δcrt cluster promoter + ΔcrtN3 )" refers to a mutant of Methylomonas sp. 16a in which the crt gene cluster promoter and the crtN3 gene have been disrupted.

The term "MWM1300 (Δald/crtN1 + ΔcrtN3 )" refers to a mutant of Methylomonas sp. 16a in which the ald, crtN1 and the crtN3 genes have been disrupted.

The term "MWM1400 (ΔcrtN1)" refers to a mutant of Methylomonas sp. 16a in which the crtN1 gene has been disrupted.

The term "MWM1600 (Δ ct cluster)" refers to a mutant of Methylomonas sp. 16a in which the ct gene cluster comprising the ald, crtN1, and crtN2 genes has been disrupted.

The term "MWM1800 (Δ crtN1 + ΔcrtN3)" refers to a mutant of Methylomonas sp. 16a in which the crtN1 and crtN3 genes have been disrupted.

The term "MWM1900 (Δ ct cluster + ΔcrtN3)" refers to a mutant of Methylomonas sp. 16a in which the ald, crtN1, crtN2, and crtN3 genes have been disrupted.

The term "MPM1000 (ΔcrtN3)" refers to a mutant of Methylomonas sp. 16a in which the crtN3 gene has been disrupted.

The term "MPM1200 (ΔcrtN2)" refers to a mutant of Methylomonas sp. 16a in which the crtN2 gene has been disrupted.

The term "MPM1300 (Δald)" refers to a mutant of Methylomonas sp. 16a in which the ald gene has been disrupted.

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 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)); DNASTAR (DNASTAR, Inc., Madison, WI); and the FASTA program incorporating the Smith-Waterman algorithm (W. R.
Pearson, *Comput. Methods Genome Res.*, [Proc. Int. Symp.], Meeting Date 1992, 111-20. Suhai, Sandor, Ed.; Plenum: New York, NY (1994). Within the context of this application it will be understood that where sequence analysis software is used for analysis, the results of the analysis are based on the “default values” of the program referenced, unless otherwise specified. As used herein “default values” will mean any set of values or parameters (set by the software manufacturer) which originally load with the software when first initialized.


Of particular interest in the present invention are high growth obligate methanotrophs having an energetically favorable carbon flux pathway. For example, a specific strain of methanotroph has been discovered having several pathway features that makes it particularly useful for carbon flux manipulation. This strain is known as *Methyloomonas 16a* (ATCC PTA 2402) (US 6,689,601; hereby incorporated by reference); and, this particular strain and other related methylotrophs are preferred microbial hosts for expression of numerous gene products. This strain may be identified by the characteristic 16S rRNA sequence as set forth in SEQ ID NO:1.

The present strain contains several anomalies in the carbon utilization pathway. For example, based on genome sequence data, the strain is shown to contain genes for two pathways of hexose metabolism. The Entner-Douderoff Pathway (which utilizes the keto-deoxy phosphogluconate aldolase enzyme) is present in the strain. It is generally well accepted that this is the operative pathway in obligate methanotrophs.

Also present, however, is the Embden-Meyerhof Pathway (which utilizes the fructose bisphosphate aldolase enzyme). It is well known that this pathway is either not present, or not operative, in obligate methanotrophs. Energetically, the latter pathway is most favorable and allows greater yield.
of biologically useful energy, ultimately resulting in greater yield production of cell mass and other cell mass-dependent products in *Methylomonas* 16a. The activity of this pathway in the present 16a strain has been confirmed through microarray data and biochemical evidence measuring the reduction of ATP. Although the 16a strain has been shown to possess both the Embden-Meyerhof and the Entner-Douderoff pathway enzymes, the data suggests that the Embden-Meyerhof pathway enzymes are more strongly expressed than the Entner-Douderoff pathway enzymes. This result is surprising and counter to existing beliefs concerning the glycolytic metabolism of methanotrophic bacteria. Additional methanotrophic bacteria having this characteristic were subsequently discovered including, for example, *Methylomonas clara* and *Methylosinus sporium*. It is likely that this activity has remained undiscovered in methanotrophs due to the lack of activity of the enzyme with ATP, the typical phosphoryl donor for the enzyme in most bacterial systems.

A particularly novel and useful feature of the Embden-Meyerhof pathway in strain 16a is that the key phosphofructokinase step is pyrophosphate-dependent instead of ATP-dependent. This feature adds to the energy yield of the pathway by using pyrophosphate instead of ATP.

Because of its significance in providing an energetic advantage to the strain, this gene in the carbon flux pathway is considered diagnostic for the present strain.

In methanotrophic bacteria, methane is converted to biomolecules via a cyclic set of reactions known as the ribulose monophosphate pathway or RuMP cycle. This pathway is comprised of three phases, each phase being a series of enzymatic steps. The first step is "fixation" or incorporation of C-1 (formaldehyde) into a pentose to form a hexose or six-carbon sugar. This occurs via a condensation reaction between a 5-carbon sugar (pentose) and formaldehyde and is catalyzed by hexulose monophosphate synthase. The second phase is termed "cleavage" and results in splitting of that hexose into two 3-carbon molecules. One of those 3-carbon molecules is recycled back through the RuMP pathway and the other 3-carbon fragment is utilized for cell growth. In methanotrophs and methylotrophs the RuMP pathway may occur as one of three variants.

However, only two of these variants are commonly found: the FBP/TA (fructose bisphosphatase/transaldolase) pathway or the KDPG/TA (keto deoxy phosphogluconate/transaldolase) pathway (Dijkstra C., G.E. Devries. "The Physiology and biochemistry of aerobic methanol-utilizing

The present strain is unique in the way it handles the "cleavage" steps where genes were found that carry out this conversion via fructose bisphosphate as a key intermediate. The genes for fructose bisphosphate aldolase and transaldolase were found clustered together on one piece of DNA. Secondly, the genes for the other variant involving the keto deoxy phosphogluconate intermediate were also found clustered together. Available literature teaches that these organisms (obligate methylotrophs and methanotrophs) rely solely on the KDPG pathway and that the FBP-dependent fixation pathway is utilized by facultative methylotrophs (Dijkhuizen et al., *supra*). Therefore the latter observation is expected, whereas the former is not. The finding of the FBP genes in an obligate methane-utilizing bacterium is both surprising and suggestive of utility. The FBP pathway is energetically favorable to the host microorganism due to the fact that more energy (ATP) is utilized than is utilized in the KDPG pathway. Thus, organisms that utilize the FBP pathway may have an energetic advantage and growth advantage over those that utilize the KDPG pathway. This advantage may also be useful for energy-requiring production pathways in the strain. By using this pathway, a methane-utilizing bacterium may have an advantage over other methane-utilizing organisms as production platforms for either single cell protein or for any other product derived from the flow of carbon through the RuMP pathway (e.g., carotenoids).

Thus, it is expected that high growth, energetically favorable *Methylomonas* strain which:

(a) grows on a C1 carbon substrate selected from the group consisting of methane and/or methanol; and

(b) comprises a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate-dependent phosphofructokinase enzyme;

will be a particularly preferred microbial host organism that is well suited for the production of various comical products of interest, following appropriate genetic engineering.

The Native C\(_{30}\) Carotenoid Pathway of *Methylomonas* sp. 16a

*Methylomonas* sp. 16a naturally produces a C\(_{30}\) carotenoid in very high concentrations within the cell, which is responsible for the organism's natural pink pigment upon visual inspection of the bacterial cells.
Production of this pigment is indicative of naturally high carbon flow through the isoprenoid pathway. Isoprenoids are an extremely large and diverse group of natural products found in all living organisms that have a common biosynthetic origin, based on a single metabolic precursor known as isopentenyl diphosphate (IPP). The group of natural products known as isoprenoids includes all substances that are derived biosynthetically from the 5-carbon compound isopentenyl diphosphate, and includes steroids, carotenoids, and squalene.

**Genes Involved in Carotenoid Production**

The enzyme pathway involved in the biosynthesis of carotenoid compounds can be conveniently viewed in two parts, the upper isoprenoid pathway (providing for the conversion of pyruvate and glyceraldehyde-3-phosphate to farnesyl pyrophosphate) and the lower carotenoid biosynthetic pathway (which provides for the synthesis of either diapophytoene (C_{30}) or phytoene (C_{40}) and all subsequently produced carotenoids) (see Figure 1).

The upper isoprenoid biosynthetic pathway leads to the production of a C_{5} isoprene subunit, isopentenyl pyrophosphate (IPP); however, this biosynthetic process may occur through either of two pathways. 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)). This mevalonate-independent pathway (shown in Figure 1) is characterized by, but not limited to, the enzymes encoded by the following genes: the “dxx” gene (encoding 1-deoxyxylulose-5-phosphate synthase); the “dxr” gene (encoding 1-deoxyxylulose-5-phosphate reductoisomerase; also known as *ispC*); the “ispD” gene (encoding a 2C-methyl-D-erythritol cytidylytransferase enzyme; also known as *ygbP*); the “ispE” gene (encoding 4-diphosphocytidyl-2-C-methylerythritol kinase; also known as *ychB*); the “ispF” gene (encoding a 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; also known as *ygbB*); the “pyrG” gene (encoding a CTP synthase); the “lytB” gene (also known as *ispH*) involved in the formation of dimethylallyl diphosphate; the “gcpE” gene (also known as *ispG*) involved
in the synthesis of 2-C-methyl-D-erythritol 4-phosphate; the "idi" gene (responsible for the intramolecular conversion of IPP to dimethylallyl pyrophosphate); and the "ispA" gene (encoding geranyltransferase or farnesyl diphosphate synthase). The synthesis of FPP occurs via the isomerization of IPP to dimethylallyl pyrophosphate (DMAPP). 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).

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, Applicants consider the first step in the lower carotenoid biosynthetic pathway to begin with the conversion of farnesyl pyrophosphate (FPP) to compounds of two divergent pathways, which lead to the formation of either C_{30} diapocarotenoids or C_{40} carotenoids.

For the biosynthesis of C_{40} carotenoids, a series of enzymatic reactions catalyzed by CrtE and CrtB occur to convert FPP to geranylgeranyl pyrophosphate (GGPP) to phytoene, the first 40-carbon molecule of the lower carotenoid biosynthesis pathway. From the compound phytoene, a spectrum of C_{40} carotenoids are produced by subsequent hydrogenation, dehydrogenation, cyclization, oxidation, or any combination of these processes. For example, 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 phytoene desaturase (encoded by the gene crtI). Lycopene cyclase (encoded by the gene crtY) converts lycopene to β-carotene. β-carotene is converted to zeaxanthin via a hydroxylation reaction resulting from the activity of β-carotene hydroxylase (encoded by the crtZ gene).

These examples are not limiting and many other carotenoid genes and products (e.g., crtX, crtW/O) exist within this C_{40} lower carotenoid biosynthetic pathway. For example, β-carotene can be converted to canthaxanthin by β-carotene ketolases encoded by crtW, bkt or crtO genes, canthaxanthin can be converted to astaxanthin by β-carotene hydroxylase encoded by the crtZ gene, and zeaxanthin can be converted to astaxanthin by β-carotene ketolases encoded by crtW, bkt, or crtO genes.
Within the C\textsubscript{30} pathway, the first unique step in the biosynthetic pathway begins with the conversion of FPP to diapophytoene. However, details concerning the conversion of FPP to diapophytoene and subsequent reactions to produce various C\textsubscript{30} carotenoids are not particularly well-understood or generalized.

In *Staphylococcus aureus*, it has been determined that the first committed reaction toward C\textsubscript{30} carotenoid biosynthesis is the head-to-head condensation of two molecules of FPP by CrtM, forming dehydrodrosqualene (Wieland, B. et al., *J. Bacteriol.*, 176(24): 7719-7726 (1994)). Subsequently, dehydrodrosqualene desaturase (encoded by *crtN*) is successively dehydrogenated in three steps to produce 4,4'-diaponeurosoporene (Wieland et al., supra). However, at present time public databases include only one single gene (GenBank® Accession Number X73889) and 4 genomic sequences (NC002745, NC002758, AP003137, AP003365) of *crtN* and *crtM*, isolated from *S. aureus* strains N315 and Mu50. A single report exists concerning the heterologous overexpression of *crtN* from *S. aureus* in *E. coli* (Raisig, A., and G. Sandmann., *J. Bacteriol.*, 181(19):6184-6187 (1999)). Based on identification of carotenoid compounds, it is known that the next stages in the C\textsubscript{30} metabolic pathway for *S. aureus* involve introduction of oxygen functions on the terminal methyl group to produce aldehyde and carboxylic acid forms of the carotenoid (Marshall, J.H., and G.J. Wilmoth., *J. Bacteriol.*, 147: 900-913 (1981) and 147: 914-919 (1981)); however, the genes responsible for this functionality have not been clearly identified.

In *Methylomonas 16a*, two operons have been identified within the genomic sequence containing carotenoid biosynthetic genes. The first biosynthetic operon (referred to herein as the "*crt* gene cluster"), encodes three genes, each of which is described below:

- The first gene (designated *crtN1*; SEQ ID NO:2) encodes a putative diapophytoene dehydrogenase with the highest BLAST hit to a diapophytoene dehydrogenase from *Heliobacillus mobilis* (34% identity and 58% similarity);
- The middle gene (designated *ald*; SEQ ID NO:3) encodes a putative aldehyde dehydrogenase with the highest BLAST hit to a betaine aldehyde dehydrogenase from *Arabidopsis thaliana* (33% identity and 50% similarity); and
- The third gene (designated *crtN2*; SEQ ID NO:4) also encodes a putative diapophytoene dehydrogenase with the highest BLAST
hit to a hypothetical protein of phytoene dehydrogenase family from *Staphylococcus aureus* (51% identity and 67% similarity).

The second biosynthetic operon encodes a fourth gene designated as *crtN3* (SEQ ID NO:5; Figure 2). “Clustal W” analysis conducted using the sequences for *crtN1, crtN2, crtN3* and *sqs* revealed that *crtN3* is not closely linked to *crtN1* and *crtN2* (Figure 3). “Clustal W” is a multiple sequence alignment program for DNA or proteins that produces biologically meaningful multiple sequence alignments of divergent sequences. This program calculates the best match for a selected sequence, and lines them up so that the identities, similarities and differences can be visualized (D. Higgins et. al., *Nucleic Acid Res.*, 22:4673-4680 (1994)). When the *crtN3* (which contains sequences that are homologous to domains of other FAD-dependent oxidoreductases) was viewed in context of its surrounding ORFs, it was observed that *crtN3* is located at the end of a cluster of ORFs that have high homology to proteins that play a role in fatty acid metabolism (Figure 2). The *crtN3* gene encodes a hypothetical protein with the highest BLAST hit to an unknown conserved protein family from *Bacillus halodurans* (31% identity and 48% similarity).

The interaction between these proteins of the carotenoid biosynthetic pathway in *Methylomonas sp. 16a* is not well understood at this time.

**Creation of Allelic Exchange Mutants via Homologous Recombination and Positive Selection**

The ability to produce specific defined mutations in a microorganism frequently relies on exploitation of the native homologous recombination properties of the cell to replace a nucleotide sequence of interest with a modified copy. Most frequently, the nucleotide sequence of interest is a particular functional gene of interest, which is then disrupted by the insertion of an antibiotic-resistance marker. In theory, this type of recombination event is easily detected on a selective medium; however, performing allelic exchange in C1 metabolizing microorganisms has been relatively cumbersome due to the organisms’ slow growth rates and the rarity of double-crossover events (which require extensive screening to isolate an allelic-exchange mutant). Despite these difficulties, a novel positive selection method for the identification of allelic exchange mutants obtained by targeted homologous recombination has been developed, as
described in co-pending U.S. Patent Application No. 60/527877, which is incorporated herein by reference.

Briefly, the positive selection (or direct genetic selection) of mutant bacteria is possible whenever survival of the recombinant bacteria depends upon the presence or absence of a particular function encoded by the DNA that is introduced into the organism. The advantage of a selection method over a screening method is that growth of bacteria with the specific desired mutation is greatly favored over bacteria lacking that specific mutation, thus facilitating the identification of the preferred mutants.

Direct or positive selection vectors containing genes that convey lethality to the host are well known. For example, expression of the *Bacillus subtilis* or the *B. amylolovfaciens* sacB genes in the presence of sucrose is lethal to *E. coli* and a variety of other Gram-negative and Gram-positive bacteria. The sacB gene encodes levansucrase, which catalyzes both the hydrolysis of sucrose and the polymerization of sucrose to form the lethal product levan. Although the basis for the lethality of levansucrase in the presence of sucrose is not fully understood, the inability of *E. coli* and many other gram negative bacteria to grow when sacB is expressed can be exploited to directly select for cells that have lost the sacB gene via homologous recombination. Numerous methods have been developed for the selection of various bacterial mutants, based on sacB. See for example: US 6,048,694 (Bramucci et al.) concerning *Bacillus*; US 5,843,664 (Pellicc et al.) concerning mycobacterium; US 5,380,657 (Schaefer et al.) concerning *Coryneform* bacteria; Hoang et al. (Gene, 212(1):77-86 (1998)) concerning *Pseudomonas aeruginosa*; Copass et al. (*Infection and Immun.*, 65(5):1949-1952 (1997)) concerning *Helicobacter pylori*; and Kamoun et al. (*Mol. Microbiol.*, 6(6):809-816 (1992)) concerning *Xanthomonas*.

The principle of the two-step positive selection strategy based on use of sacB for C1 metabolizing bacteria is illustrated in Figure 4. This method relies on the application of a positive selection vector, in a preferred embodiment derived from the suicide delivery vector pGP704, which is able to integrate into the chromosome of C1 metabolizing bacteria to produce mutations that are the result of both single- or double-crossover events (Figure 5). Specifically, the positive selection vector comprises:

(i) at least one gene encoding resistance to a first selectable marker (e.g., Amp);
(ii) a sacB coding region encoding a levansucrase enzyme under the control of a suitable promoter; and

(iii) a replacement nucleotide sequence of interest (i.e., re-NSI), which one desires to insert into the chromosome of the C1 metabolizing bacteria as a replacement to an existing nucleotide sequence of interest in the bacterial chromosome (i.e., chr-NSI). Thus, re-NSI is modified with respect to chr-NSI by the addition, substitution, or deletion of at least one nucleotide.

Upon transformation of C1 metabolizing bacteria with the positive selection vector described above, a single-crossover event by homologous recombination occurs between chr-NSI and re-NSI, such that the entire positive selection vector is integrated into the bacterial chromosome at the site of crossover. These events can be selected by growth on the first selectable marker (e.g., Amp), whereby a complete copy of chr-NSI and a complete copy of re-NSI are present in the chromosome. Upon removal of selection by the first selectable marker, a second crossover event may occur, resulting in the “looping out” of the positive selection vector, to yield transformants containing either the chr-NSI or the re-NSI in the chromosome (Figure 4). Direct selection of these allelic exchange transformants is possible by growing the transformants in the presence of sucrose, since single-crossover mutants will be killed under these conditions.

**Screening Methods**

Methods of screening in microbiology are discussed at length in Brock, *supra*. In the present invention, a two-step selection process permits the identification of double-crossover mutations in C1 metabolizing bacterial cells by applying positive selection pressure. Using this strategy, the positive selection vector should comprise a first selectable marker and a sacB marker. Selection involves first growing the transformants on media containing the first selectable marker, to identify those cells that have undergone a single-crossover (i.e., wherein the entire chromosomal integration vector has integrated into the host cell’s genome). Then, the selection pressure is removed and a second crossover event may occur.

Selection for allelic exchange mutants requires growth of the cells on sucrose, since SacB expression will be lethal to all single-crossover mutants. Differentiation between allelic exchange mutants containing the wildtype and mutant allele is then possible using standard molecular
techniques (e.g., PCR), well known to one of skill in the art. One preferred advantage of the two-step selection strategy described above is that allelic exchange transformants that are produced are markerless (i.e., lacking any antibiotic or other genetic marker indicative of the allelic exchange).

Host Strain Optimization via Allelic Exchange within the C₃₀ Carotenoid Biosynthetic Pathway

Four non-pigmented *Methylomonas* sp. 16a bacterial host organisms, each lacking any antibiotic markers, are provided in the instant invention. These optimized host strains comprise a deletion in the *aldl/crtN1* genes, a deletion of the *crt* gene cluster promoter, a deletion in the *aldl/crtN1* genes and the *crtN3* gene, and a deletion in the *crt* gene cluster promoter and the *crtN3* gene. Three other strains may convey additional advantages for a *Methylomonas* production host organism having knockouts in one or more of the *crtN1*, *crtN2*, and *crtN3* genes.

Each of these bacterial strains was created by investigation of allelic exchange mutations within the native *crt* gene cluster (comprising the *crtN1*, *ald*, and *crtN2* genes) and the *crtN3* gene of *Methylomonas* sp. 16a.

The process by which these allelic exchange mutants were created is shown in Figure 4, and requires a re-NSI that is modified with respect to chr-NSI by the addition, substitution, or deletion of at least one nucleotide. For the purposes herein, the chr-NSI encodes at least one of the native *crtN1*, *ald*, *crtN2*, and *crtN3* genes of *Methylomonas* sp. 16a; alternatively, the chr-NSI corresponds to the promoter driving the *Methylomonas* sp. 16a *crt* gene cluster. The re-NSI will enable production of a transformed *Methylomonas* sp. 16a having a deletion at least one of the *crtN1*, *ald*, *crtN2*, and *crtN3* genes, or a deletion of the promoter driving the *crt* gene cluster. As will be obvious to one of skill in the art, the advantage of the two-step selection methodology described herein is that the double-crossover mutant thus generated is "markerless"; enabling subsequent mutations using the same technique (i.e., since there is no need for a different selectable marker corresponding to each mutation created).

One factor to consider regardless of the specific type of re-NSI generated is the overall homology between the re-NSI and the chr-NSI. In general, it is well known in the art that homologous recombination requires a minimum of 50 nucleotides of homology on each side of the site of a crossover. When preparing a re-NSI for use in the selection processes described herein, it is preferable to have regions homologous to the chr-NSI flanking (both 5' and 3') the site of the addition, substitution, or
deletion. More preferably, a 1 kB region of homology is preferred on both sides of the addition, substitution, or deletion. In contrast, re-NSI is not expected to be limited in length, beyond the limitations inherent to homologous recombination.

Generation of a re-NSI containing an addition, substitution, or deletion of at least one nucleotide with respect to the chr-NSI can be accomplished using numerous techniques known to a skilled artisan in the field of molecular biology. Although not intended to be limiting, deletions and additions may be generated by the use of restriction endonucleases, in vitro transposition reactions, or PCR methodologies; all techniques well known to one of skill in the art.

A preferred method for generation of re-NSI is via PCR methodologies, as utilized herein. Alternatively, substitutions may be generated by mutagenesis of the re-NSI. Two suitable approaches include error-prone PCR (Leung et al., Technique, 1:11-15 (1989); Zhou et al., Nucleic Acids Res., 19:6052-6052 (1991); and Spee et al., Nucleic Acids Res., 21:777-778 (1993)) and in vivo mutagenesis. The principal advantage of error-prone PCR is that all mutations introduced by this method will be within the re-NSI, and any change may be easily controlled by changing the PCR conditions. Alternatively, in vivo mutagenesis may be employed using commercially available materials such as E. coli XL1-Red strain, and the Epicurian coli XL1-Red mutator strain from Stratagene (La Jolla, CA; Greener and Callahan, Strategies, 7:32-34 (1994)). This strain is deficient in three of the primary DNA repair pathways (mutS, mutD, and mutT), resulting in a mutation rate 5000-fold higher than that of wild-type. In vivo mutagenesis does not depend on ligation efficiency (as with error-prone PCR); however, a mutation may occur at any region of the vector and the mutation rates are generally much lower.

It is also contemplated that it may be desirable to replace a wild-type gene of interest (i.e., chr-NSI) in the C1 metabolizing bacteria with a mutant gene (i.e., re-NSI) that has been constructed using the method of “gene shuffling” (US 5,605,793; US 5,811,238; US 5,830,721; and US 5,837,458). The method of gene shuffling is particularly attractive due to its facile implementation, and high rate of mutagenesis. The process of gene shuffling involves the restriction of a gene of interest into fragments of specific size in the presence of additional populations of DNA regions of both similarity to, or difference to, the gene of interest. This pool of
fragments is then denatured and reannealed to create a mutated gene. The mutated gene is subsequently screened for altered activity.

One factor to consider during the preparation of a re-NSI for use in the two-step selection strategy concerns the placement of the addition, deletion, or substitution within the sequence of interest. Specifically, the re-NSI is first inserted into the chromosome by integration of the chromosomal integration vector (a single-crossover event). The second crossover event that occurs can result in either a mutant or wildtype sequence in the chromosome, since the single-crossover contains two copies of the nucleotide sequence of interest. In order to increase the percentage of segregants that retain the re-NSI, as opposed to reverting to the wildtype encoded by the chr-NSI, it is desirable to "center" the mutation with respect to the flanking DNA that has homology to the chr-NSI. For example, if a point mutation was perfectly centered within a re-NSI, about 50% of the segregants would be expected to retain the mutation in the chromosome (thus producing a 1:1 ratio of double-crossover mutants to wild-type cells).

**Transformation of C1 Metabolizing Bacteria**

Techniques for the transformation of C1 metabolizing bacteria are not well developed, although general methodology that is utilized for other bacteria, which is well known to those of skill in the art, may be applied.


Bacterial conjugation, relying on the direct contact of donor and recipient cells, is frequently more readily amenable for the transfer of genes into C1 metabolizing bacteria. Simplistically, this bacterial conjugation process involves mixing together “donor” and “recipient” cells in close contact with one another. Conjugation occurs by formation of cytoplasmic connections between donor and recipient bacteria, with direct transfer of newly synthesized donor DNA into the recipient cells. As is well known in the art, the recipient in a conjugation is defined as any cell that can accept DNA through horizontal transfer from a donor bacterium. The
donor in conjugative transfer is a bacterium that contains a conjugative plasmid, conjugative transposon, or mobilizable plasmid. The physical transfer of the donor plasmid can occur in one of two fashions, as described below:

1. In some cases, only a donor and recipient are required for conjugation. This occurs when the plasmid to be transferred is a self-transmissible plasmid that is both conjugative and mobilizable (i.e., carrying both tra genes and genes encoding the Mob proteins). In general, the process involves the following steps: 1.) Double-strand plasmid DNA is nicked at a specific site in oriT; 2.) A single-strand DNA is released to the recipient through a pore or pilus structure; 3.) A DNA relaxase enzyme cleaves the double-strand DNA at oriT and binds to a released 5' end (forming a relaxosome as the intermediate structure); and 4.) Subsequently, a complex of auxiliary proteins assemble at oriT to facilitate the process of DNA transfer.

2. Alternatively, a “triparental” conjugation is required for transfer of the donor plasmid to the recipient. In this type of conjugation, donor cells, recipient cells, and a “helper” plasmid participate. The donor cells carry a mobilizable plasmid or conjugative transposon. Mobilizable vectors contain an oriT, a gene encoding a nickase, and have genes encoding the Mob proteins; however, the Mob proteins alone are not sufficient to achieve the transfer of the genome. Thus, mobilizable plasmids are not able to promote their own transfer unless an appropriate conjugation system is provided by a helper plasmid (located within the donor or within a “helper” cell). The conjugative plasmid is needed for the formation of the mating pair and DNA transfer, since the plasmid encodes proteins for transfer (Tra) that are involved in the formation of the pore or pilus.


As is well known to those of skill in the art, efforts to genetically engineer a microorganism for high-level production of a specific product
frequently require substantial manipulation to the native host machinery and biosynthetic pathways. Historically, metabolic engineering of methanotrophs has remained relatively undeveloped due to the general lack of efficient genetic engineering tools, as compared to other industrial bacteria such as *E. coli* and the yeasts. Most methanotrophic molecular biology has focused on the engineering of the methane monooxygenase, such that it is directed toward more useful co-metabolic products (and wherein methane is not directly incorporated into the product molecule). In contrast, pathway engineering for net synthesis of carbon-containing compounds from methane requires both: 1) the ability to inactivate genes already present in the methanotroph; and 2) the ability to introduce foreign genes into the organism, to create new metabolic capabilities. In this way, new metabolic networks can be constructed to produce products that could not be made by the genetic engineering of methane monooxygenase alone.

The present invention represents tremendous progress in the genetic engineering of methanotrophic bacteria. Specifically, the *Methylomonas* sp. 16a optimized hosts provided herein comprise deletions in one or more of the *crtN1, ald, crtN2*, and *crtN3* genes and the *crt* gene cluster promoter will be useful host organisms for the production of a variety of commercial products. In another embodiment, the optimized host provided herein comprise deletions selected from the group consisting of *Δald(crtN1), ΔcrtN1, Δ cort* cluster promoter, and *Δ cort* gene cluster (*Δald + ΔcrtN1 + ΔcrtN2*). In a further embodiment, the optimized hosts provided herein comprise deletions selected from the group consisting of *Δald(crtN1) + ΔcrtN3, ΔcrtN1 + ΔcrtN3, Δ cort* cluster promoter + *ΔcrtN3*, and *Δ cort* gene cluster + *ΔcrtN3*. Specifically, advantages in the host organism are incurred by knocking out the native carotenoid pathway of the organism that lead to the production of pink-pigmented C\textsubscript{30} carotenoids, increasing the available carbon flux directed toward the products of interest. For example, products include, but are not limited to C\textsubscript{40} carotenoids, such as antheraxanthin, adonirubin, adonixanthin, astaxanthin, canthaxanthin, capsorubrin, β-cryptoxanthin, α-carotene, β-carotene, epsilon-carotene, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone, γ-carotene, 4-keto-γ-carotene, ζ-carotene, α-cryptoxanthin, deoxyflexaxanthin, diatoxanthin, 7,8-didehydroastaxanthin, fucoxanthin, fucoxanthinol, isorenieratene, luctucaxanthin, lutein, lycopene, myxobactone, neoxanthin, neurosporene,
hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, 4-keto-rubixanthin, siphonaxanthin, spheroidene, spheroidenone, spirilloxanthin, 4-keto-torulene, 3-hydroxy-4-keto-torulene, urolide, urolide acetate, violaxanthin, zeaxanthin-β-diglucoside, and zeaxanthin.

Additionally, these "markerless" mutants may offer significant advantages during industrial development of a C1 metabolizing host bacterium where the presence of antibiotic resistance genes would be undesirable, in terms of required regulatory approvals.

Recombinant Expression in the Optimized Hosts

Methods for introducing genes into a suitable microbial host, such as the optimized Methylomonas sp. 16a hosts described herein, are common. As will be obvious to one skilled in the art, the particular functionalities required to be introduced into a host organism for production of a particular product will depend on the host cell, the availability of substrate, and the desired end product(s).

The preferred genes to introduce into the optimized Methylomonas sp. 16a hosts of the present invention are various combinations of crt genes to produce C40 carotenoids. Examples of suitable crt genes for use in the present invention include, but are not limited to crtE, crtB, crtL, crtY, crtZ and crtX genes isolated from Pectobacterium cypripedii, as described by Cheng et al. in copending US 10/804677, incorporated herein by reference; crtE, crtB, crtL, crtY, crtZ and crtX genes isolated from a member of the Enterobacteriaceae family, as described by Cheng et al. in copending US 10/808979, incorporated herein by reference; crtE, idi, crtB, crtL, crtY, crtZ genes isolated from Pantoaea agglomerans, as described by Cheng et al. in copending US 10/8088073, incorporated herein by reference; and crtE, idi, crtB, crtL, crtY, crtZ and crtX genes isolated from Pantoaea stewartii, as described by Cheng et al. in copending US 10/810733, incorporated herein by reference. More preferably, the crtE-idi-crtY-crtL-crtB gene cluster, given as SEQ ID NO:6, derived from the crtE-idi-crtY-crtL-crtB-crtZ gene cluster (SEQ ID NO:41) isolated from Pantoaea agglomerans, described by Cheng et al. in copending US 10/808807, is used.

For the purposes of the present invention, it was desirable to modify a portion of the codons encoding polypeptides having ketolase or hydroxylase activity, respectively, to enhance the expression of genes encoding those polypeptides in Methylomonas sp. 16a and derivatives thereof. Thus, the nucleic acid sequence of the native genes, e.g., the β-
carotene ketolase (crtW) and the β-carotene hydroxylase (crtZ) genes from Agrobacterium aurantiacum, were modified to employ host preferred codons, as described in Example 8. In general, host preferred codons can be determined from the codons of highest frequency in the proteins (preferably expressed in the largest amount) in a particular host species of interest. Thus, the coding sequence for a polypeptide having ketolase or hydroxylase activity can be synthesized in whole or in part using the codons preferred in the host species. All (or portions) of the DNA also can be synthesized to remove any destabilizing sequences or regions of secondary structure which would be present in the transcribed mRNA. All (or portions) of the DNA also can be synthesized to alter the base composition to one more preferable in the desired host cell.

In one preferred embodiment, the crtE-idi-crtY-crtl-crtB gene cluster (SEQ ID NO:6) from Pantoea agglomerans is used in conjunction with the codon-optimized crtW (β-carotene ketolase) gene given as SEQ ID NO:7 to produce the C₄₀ carotenoid canthaxanthin.

In another preferred embodiment, the crtE-idi-crtY-crtl-crtB gene cluster is used in conjunction with the codon-optimized crtW gene and the codon-optimized crtZ (β-carotene hydroxylase) gene given as SEQ ID NO:8 to produce the C₄₀ carotenoid astaxanthin.

Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct chimeric genes for production of a variety of gene products.

These chimeric genes could then be introduced into the optimized hosts of the present invention via transformation to provide high level expression of the required enzymes.

Vectors or cassettes useful for the transformation of the optimized hosts of the present invention are well known in the art. Typically the vector or cassette contains sequences directing transcription and translation of the relevant gene(s), a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcriptional termination. It is most preferred when both control regions are derived from genes homologous to the transformed host cell, although it is to be understood that such control regions need not be derived from the genes native to the specific species chosen as a production host.
Initiation control regions or promoters which are useful to drive expression of the desired ORFs in the optimized host cells of the present invention are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving the desired genes is suitable for the present invention including, but not limited to: CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI, AOX1, lac, ara, tet, trp, IP, IP, T7, tac, trc, amy, apr, npr and various phage promoters. Additionally, the deoxy-xylulose phosphate synthase or methanol dehydrogenase operon promoter (Springer et al., FEMS Microbiol Lett, 160:119-124 (1998)), the promoter for polyhydroxyalkanoic acid synthesis (Foellner et al., Appl. Microbiol. Biotechnol., 40:284-291 (1993)), promoters identified from native plasmids in methylo trophs (EP 296484), Piac (Toyama et al., Microbiology, 143:595-602 (1997); EP 62971), Ptrc (Brosius et al., Gene, 27:161-172 (1984)), and promoters associated with antibiotic resistance [e.g., kanamycin (Springer et al., FEMS Microbiol Lett, 160:119-124 (1998); Ueda et al., Appl. Environ. Microbiol., 57:924-926 (1991)], tetracycline (US 4,824,786) or chloramphenicol are suitable for expression in the optimized Methylomonas sp. 16a host cells of the present invention.

It is necessary to include an artificial ribosomal binding site ("RBS") upstream of a gene to be expressed, when the RBS is not provided by the vector. This is frequently required for the second, third, etc. gene(s) of an operon to be expressed, when a single promoter is driving the expression of a first, second, third, etc. group of genes. Methodology to determine the preferred sequence of a RBS in a particular host organism will be familiar to one of skill in the art, as are means for creation of this synthetic site.

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.

Merely inserting a gene into a cloning vector does not ensure that it will be successfully expressed at the level needed. In response to the need for a high expression rate, many specialized expression vectors have been created by manipulating a number of different genetic elements that control aspects of transcription, translation, protein stability, oxygen limitation, and secretion from the host cell. More specifically, the molecular features that have been manipulated to control gene expression include: 1.) the nature of the relevant transcriptional promoter and terminator sequences; 2.) the strength of the ribosome binding site; 3.) the
number of copies of the cloned gene and whether the gene is plasmid-
borne or integrated into the genome of the host cell; 4.) the final cellular
location of the synthesized foreign protein; 5.) the efficiency of translation
in the host organism; 6.) the intrinsic stability of the cloned gene protein
within the host cell; and 7.) the codon usage within the cloned gene, such
that its frequency approaches the frequency of preferred codon usage of
the host cell. Each of these types of modifications are encompassed in
the present invention, as means to further optimize expression of preferred
products within the optimized hosts of the present invention.

Finally, to promote accumulation of a preferred product, 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 sinks for
energy or carbon. Alternatively, it may be useful to over-express various
genes upstream of desired carotenoid intermediates to enhance
production. Methods of manipulating genetic pathways for the purposes
described above are common and well known in the art.

**Industrial Production Methodologies**

For commercial production of the desired product, e.g., C_{40}
carotenoids, a variety of culture methodologies may be applied. For
example, large-scale production of a specific gene product overexpressed
from the optimized *Methylomonas sp.* 16a bacterial host organisms of the
present invention may be produced by batch or continuous culture
methodologies.

A classical batch culturing method is a closed system where the
composition of the media is set at the beginning of the culture and not
subject to artificial alterations during the culturing process. Thus, at the
beginning of the culturing process the media is inoculated with the desired
organism or organisms and growth or metabolic activity is permitted to
occur while adding nothing to the system. Typically, however, a "batch"
culture is batch with respect to the addition of carbon source and attempts
are often made at controlling factors such as pH and oxygen
concentration. In batch systems the metabolite and biomass compositions
of the system change constantly up to the time the culture is terminated.
Within batch cultures cells moderate through a static lag phase to a high
growth log phase and finally to a stationary phase where growth rate is
diminished or halted. If untreated, cells in the stationary phase will
eventually die. Cells in log phase are often responsible for the bulk of
production of end product or intermediate in some systems. Stationary or post-exponential phase production can be obtained in other systems.

A variation on the standard batch system is the Fed-Batch system. Fed-Batch culture processes are also suitable in the present invention and comprise a typical batch system with the exception that the substrate is added in increments as the culture progresses. Fed-Batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the media. Measurement of the actual substrate concentration in Fed-Batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO₂. Batch and Fed-Batch culturing methods are common and well known in the art and examples may be found in Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, 2nd ed. (1989) Sinauer Associates: Sunderland, MA, or Deshpande, Mukund V., *Appl. Biochem. Biotechnol.*, 36:227 (1992).

Commercial production of the desired product, e.g., C₄₀ carotenoids may also be accomplished with a continuous culture. Continuous cultures are an open system where a defined culture media is added continuously to a bioreactor and an equal amount of conditioned media is removed simultaneously for processing. Continuous cultures generally maintain the cells at a constant high liquid phase density where cells are primarily in log phase growth. Alternatively continuous culture may be practiced with immobilized cells where carbon and nutrients are continuously added, and valuable products, by-products or waste products are continuously removed from the cell mass. Cell immobilization may be performed using a wide range of solid supports composed of natural and/or synthetic materials.

Continuous or semi-continuous culture allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to media being drawn off must be balanced against the cell growth rate in the culture. Methods of modulating nutrients and growth factors for
continuous culture processes, as well as techniques for maximizing the rate of product formation, are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

Fermentation media in the present invention must contain suitable carbon substrates. Suitable carbon substrates for the optimized *Methylomonas* sp. 16a host cells of the present invention include methane and methanol for which metabolic conversion into key biochemical intermediates has been demonstrated.

**DESCRIPTION OF PREFERRED EMBODIMENTS**

*Methylomonas* sp. 16a is a C1 metabolizing bacteria that is capable of efficiently using either methanol or methane as a carbon substrate. *Methylomonas* is also metabolically versatile in that it contains multiple pathways for the incorporation of carbon from formaldehyde into 3-carbon units, and is capable of genetic exchange with donor species such as *Escherichia coli* via bacterial conjugation. Despite these advantageous qualities, considerable optimization of the host organism for use as a microbial platform of choice for the production of a variety of materials was required. Specifically, it was desirable to reduce and/or eliminate the amount of carbon flow channeled toward the native C30 carotenoids produced by the organism, such that this carbon would be available for more desirable products.

Following the recent development of a method for the generation and identification of defined mutations within the organism by homologous recombination, based on the positive selection vector pGP704::sacB, it was possible in the work described herein to methodically produce gene and promoter knockouts within the native C30 carotenoid biosynthetic pathway. Specifically, five different allelic exchange mutations were investigated: \( \Delta \text{crtN1}, \Delta \text{ald}/\text{crtN1}, \Delta \text{crtN2}, \Delta \text{crtN3}, \Delta \text{crt} \) gene cluster (\( \Delta \text{ald}, \Delta \text{crtN1}, \text{and } \Delta \text{crtN2} \)), and a deletion of the *crt* gene cluster promoter (thereby creating a \( \Delta \text{crtN1}, \Delta \text{ald}, \text{and } \Delta \text{crtN2 knockout} \)). Additionally, several double mutants comprising a deletion in one or more of the C30 carotenoid genes (i.e. \( \Delta \text{crtN1}, \Delta \text{ald}(<\text{crtN1}), \Delta \text{crt} \) gene cluster promoter, and \( \Delta \text{crt} \) gene cluster) in combination with a deletion of the *crtN3* gene were also investigated. The results of this work produced four optimized host strains that were non-pigmented, since the wildtype C30 carotenoid “pink” pigment was no longer produced in the organism.

It is expected that these optimized host strains will be useful for increased expression of a variety of heterologous proteins in
Methylopectos sp. 16a. For example, there is a general practical utility for microbial production of C_{40} carotenoid compounds. This practical utility results since these compounds are very difficult to make chemically (Nelis and Leenheer, *Appl. Microbiol.*, 70:181-191 (1991)). Industrially, only a few carotenoids are used for food colors, animal feeds, pharmaceuticals, and cosmetics, despite the existence of more than 600 different carotenoids identified in nature. Most carotenoids have strong color and can be viewed as natural pigments or colorants. Furthermore, many carotenoids have potent antioxidant properties and thus inclusion of these compounds in the diet is thought to be healthful.

A variety of methods for carotenoid production based on microbial platforms have recently been described in the art [e.g., *E. coli* and *Candia utilis* for production of lycopene (Farmer W.R. and J.C. Liao., *Biotechnol. Prog.*, 17: 57-61 (2001); Wang C. et al., *Biotechnol. Prog.*, 16: 922-926 (2000); Misawa, N. and H. Shimada., *J. Biotechnol.*, 59: 169-181 (1998); Shimada, H. et al., *Appl. Environ. Microbiol.*, 64:2676-2680 (1998)); *E. coli, Candia utilis* and *Paffia rhodozyma* for production of β-carotene (Albrecht, M. et al., *Biotechnol. Lett.*, 21: 791-795 (1999); Miura, Y. et al., *Appl. Environ. Microbiol.*, 64:1226-1229 (1998); US 5,691,190; *E. coli* and *Candia utilis* for production of zeaxanthin (Albrecht, M. et al., *supra*; Miura, Y. et al., *supra*); *E. coli* and *Paffia rhodozyma* for production of astaxanthin (US 5,466,599; US 6,015,684; US 5,182,208; US 5,972,642); see also: US 5,656,472, US 5,545,816, US 5,530,189, US 5,530,188, US 5,429,939, and US 6,124,113]. However, these methods of producing carotenoids suffer from low yields and reliance on relatively expensive feedstocks. Thus, it would be desirable to identify a method that produces higher yields of carotenoids in a microbial host, such as Methylopectos sp. 16a (or a derivative thereof), from an inexpensive feedstock, such as methane or methanol.

Odom et al. have previously demonstrated that the C1 metabolizing bacterium Methylopectos sp. 16a can be engineered for production of various C_{40} carotenoids (US 09/941947 corresponding to WO 02/218617), by the introduction of one or more of the lower C_{40} carotenoid biosynthetic pathway genes (i.e., *crtE, crtX, crtY, ctrl, crtB, ctw, cto, cra, ctc, crtD, ctf, and ctwU*). Despite the previous demonstration of β-carotene production in this unique microbial host, however, further advancement towards creation of a recombinant host suitable for high-level production of a specific C_{40} carotenoid product will require significant metabolic
engineering to the native host machinery and biosynthetic pathways. The present invention will be particularly useful for efforts targeted toward the genetic engineering of this unique C1 metabolizing bacteria, since it is expected that the absence of the native C₃₀ carotenoid production will provide substantial advantages for gene expression leading to C₄₀ carotenoids, such as antheraxanthin, adonirubin, adonixanthin, astaxanthin, canthaxanthin, capsorubin, β-cryptoxanthin, α-carotene, β-carotene, epsilon-carotene, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone, γ-carotene, 4-keto-γ-carotene, ζ-carotene, α-cryptoxanthin, deoxyflexixanthin, diatoxanthin, 7,8-didehydroastaxanthin, fucoxanthin, fucoxanthinol, isorenieratene, luctucaxanthin, lutein, lycopene, myxobactone, neoxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, 4-keto-rubixinanthe, siphonaxanthin, spheroidene, spheroidene, spirilloxanthin, 4-keto-torulene, 3-hydroxy-4-keto-torulene, uriolide, uriolide acetate, violaxanthin, zeaxanthin-β-diglucoside, and zeaxanthin.

The present invention also provides a codon-optimized β-carotene ketolase gene (crtW) and a β-carotene hydroxylase gene (crtZ) which were used to demonstrate the production of canthaxanthin and astaxanthin in the Methylomonas sp. 16a deletion mutants.

EXAMPLES

The present invention is further defined in the following Examples. 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 uses and conditions.

GENERAL METHODS


The meaning of abbreviations is as follows: "sec" means second(s), "min" means minute(s), "hr" means hour(s), "d" means day(s), "µL" means microliter(s), "mL" means milliliter(s), "L" means liter(s), "µM" means micromolar, "mM" means millimolar, "M" means molar, "mmol" means millimole(s), "µmol" mean micromole(s), "nmol" means nanomole(s), "g" means gram(s), "µg" means microgram(s), "ng" means nanogram(s), "nm" means nanometers, "U" means unit(s), "ppm" means parts per million, "bp" means base pair(s), "rpm" means revolutions per minute, "kB" means kilobase(s), "g" means the gravitation constant, "~" means approximately, "OD_{600}" means the optical density measured at 600 nm, "OD_{280}/OD_{280}" means the ratio of the optical density measured at 260 nm to the optical density measured at 280 nm, and "mAU" means milliabsorbance units.

**Molecular Biology Techniques:**

Methods for agarose gel electrophoresis were performed as described in Maniatis (supra). Polymerase Chain Reactions (PCR) techniques were found in White, B., *PCR Protocols: Current Methods and Applications*, Humana: Totowa, NJ (1993), Vol. 15.

**Media and Culture Conditions:**

EXAMPLE 1

GROWTH OF METHYLOMONAS SP. 16a

Example 1 summarizes the standard conditions used for growth of Methylomonas sp. 16a (ATCC# PTA-240), as described in US 6,689,601; hereby incorporated by reference.

*Methylomonas* Strain and Culture Media

The growth conditions described below were used throughout the following experimental Examples for treatment of *Methylomonas* 16a, unless conditions were specifically described otherwise.

*Methylomonas* sp. 16a was typically grown in serum stoppered Wheaton bottles (Wheaton Scientific; Wheaton, IL) using a gas/liquid ratio of at least 8:1 (i.e., 20 mL of ammonium liquid “BTZ” growth medium in a Wheaton bottle of 160 mL total volume). The composition of the BTZ growth medium is given below. The standard gas phase for cultivation contained 25% methane in air, although methane concentrations can vary ranging from about 5-50% by volume of the culture headspace. These conditions comprise growth conditions and the cells are referred to as growing cells. In all cases, the cultures were grown at 30°C with constant shaking in a rotary shaker (Lab-Line, Barnstead/Thermolyne; Dubuque, IA) unless otherwise specified.

BTZ Media for *Methylomonas* 16a

*Methylomonas* 16a typically grows in a defined medium composed of only minimal salts; no organic additions such as yeast extract or vitamins are required to achieve growth. This defined medium known as BTZ medium (also referred to herein as “ammonium liquid medium”) consisted of various salts mixed with Solution 1, as indicated in Tables 1 and 2. Alternatively, the ammonium chloride was replaced with 10 mM sodium nitrate to give “BTZ (nitrate) medium”, where specified. Solution 1 provides the composition for a 100-fold concentrated stock solution of trace minerals.

<table>
<thead>
<tr>
<th></th>
<th>Molecular Weight (mM)</th>
<th>Conc. (mM)</th>
<th>g per L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitriloacetic acid</td>
<td>191.10</td>
<td>66.90</td>
<td>12.80</td>
</tr>
<tr>
<td>CuCl₂ x 2H₂O</td>
<td>170.48</td>
<td>0.15</td>
<td>0.0254</td>
</tr>
</tbody>
</table>
Table 2

<table>
<thead>
<tr>
<th>Ammonium Liquid Medium (BTZ)**</th>
</tr>
</thead>
<tbody>
<tr>
<td>MW</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>NH₄Cl</td>
</tr>
<tr>
<td>KH₂PO₄</td>
</tr>
<tr>
<td>Na₂SO₄</td>
</tr>
<tr>
<td>MgCl₂ x 6H₂O</td>
</tr>
<tr>
<td>CaCl₂ x 2H₂O</td>
</tr>
<tr>
<td>1 M HEPES (pH 7.0)</td>
</tr>
<tr>
<td>Solution 1</td>
</tr>
</tbody>
</table>

**Dissolve in 900 mL H₂O. Adjust to pH=7.0, and add H₂O to give a final volume of 1 L. For agar plates: Add 15 g of agarose in 1 L of medium, autoclave, cool liquid solution to 50 °C, mix, and pour plates.

EXAMPLE 2

CONSTRUCTION OF A POSITIVE-SELECTION SUICIDE VECTOR FOR

*METHYLOMONAS SP. STRAIN 16a*

The construction of chromosomal mutations within the *Methylomonas* genome required the use of suicide vectors. Thus, a modified version of the conditional replication vector pGP704 was created, comprising a *npr-sacB* cassette.

pGP704 as a Vector Backbone for the C1 Chromosomal Integration Vector

The plasmid pGP704 (Miller and Mekalanos, *J. Bacteriol.*, (170): 2575-2583 (1988); Figure 5) was chosen as a suitable vector backbone for the C1 chromosomal integration vector, since it could be used as a vehicle to transfer replacement nucleotide sequences of interest into...
Methylomonas sp. 16a via conjugation. Plasmid pGP704 is a derivative of pBR322 that is AmpR but has a deletion of the pBR322 origin of replication (oriE1). Instead, the plasmid contains a cloned fragment containing the origin of replication of plasmid R6K. The R6K origin of replication (oriR6K) requires the II protein, encoded by the pir gene. In E. coli, the II protein can be supplied in trans by a prophage (λ pir) that carries a cloned copy of the pir gene. The pGP704 plasmid also contains a 1.9 kB BamHI fragment encoding the mob region of RP4. Thus, pGP704 can be mobilized into recipient strains by transfer functions provided by a derivative of RP4 integrated in the chromosome of E. coli strain SM10 or SY327. Once the plasmid is transferred, however, it is unable to replicate in recipients that lack the II protein (e.g., recipients such as Methylomonas and other C1 metabolizing bacteria). This inability permits homologous recombination to occur between nucleotide sequences of interest on pGP704 and the intact chromosomal nucleotide sequences of interest.

Thus, on the basis of the above characteristics, the pGP704 vector backbone met the following conditions for a chromosomal integration vector suitable for C1 metabolizing bacteria: 1.) it was conditional for replication, thus allowing selection for integration into the chromosome; 2.) it possessed at least one selectable marker; 3.) it had an origin of transfer that was expected to be suitable for C1 metabolizing bacteria; 4.) it possessed mobilization genes; and 5.) it contained a variety of unique cloning sites. Other alternative chromosomal integration vectors having the characteristics listed above are expected to be suitable for use in the present invention, as described herein.

Plasmid pGP704 did not, however, permit easy detection and identification of clones that had undergone allelic exchange. Thus, pGP704 was modified to permit the positive selection of double-crossover events within Methylomonas and other C1 metabolizing bacteria.

Cloning of the npr-sacB Cassette

Plasmid pBE83 contained a Bacillus amyloyoliquifaciens sacB gene under the control of the neutral protease (npr) promoter (gift from V. Nagarajan, E.I. du Pont de Nemours and Co., Inc., Wilmington, DE). The npr-sacB cassette was PCR amplified from pBE83 using DNA primers DrdI/npr-sacB and TthIII/npr-sacB. The DNA primers were constructed to include unique restriction sites at each terminus of the PCR product to facilitate subsequent cloning (as indicated by the underlined sequences below):
Drdl/npr-sacB: (SEQ ID NO:9)
5'-GACATCGATGTCGAATTGCTCGGTACCGTCGATC-3'

TthIII/npr-sacB: (SEQ ID NO:10)
5'-GACCTCGTCTCGTTATTAGTTGACTGTCAGC-3'

The PCR reaction mixture was composed of the following: 10 µL of 10X PCR buffer; 16 µL (4 µL each) of dNTPs (320 mM stock); 1 µL of *Methylomonas* chromosomal DNA solution (~500 ng/µL); 8 µL of MgCl₂ solution (25 mM); 0.5 µL of Taq polymerase (5 U/µL); 1 µL of Drdl/npr-sacB primer (~36 nmol); 1 µL of TthIII/npr-sacB primer (~35 nmol); and 71 µL of sterile deionized water (NANOpure® Water System, Barnstead/Thermolyne). The PCR protocol was then performed on a 9600 GeneAmp® PCR System (Perkin Elmer), according to the thermocycling parameters below:

- 1 cycle: 94°C (5 min);
- 1 cycle: 94°C (5 min), 60°C (2 min), 72°C (3 min);
- 35 cycles: 94°C (1 min), 60°C (2 min), 72°C (3 min);
- 1 cycle: 94°C (1 min), 60°C (2 min), 72°C (10 min); and
- Hold – 4°C (∞).

Afterward, the PCR product was ligated into the pCR2.1TOPO vector per the manufacturer's instructions (Invitrogen; Carlsbad, CA). The ligation mixture was transformed into TOP10 One Shot™ calcium chloride competent cells and transformants were screened as recommended by Invitrogen.

Plasmid DNA was isolated from positive clones (white colonies in a blue/white screen) using the QIAprep® Spin Mini-prep Kit (Qiagen; Valencia, CA) and the DNA was digested according to the manufacturer's instructions with restriction endonucleases Drdl and TthIII (New England Biolabs; Boston, MA). Initially, this PCR product was to be inserted into pGP704 digested with Drdl and TthIII; however, there were difficulties in cloning the Drdl/TthIII PCR product.

A modified cloning strategy was adopted, such that the PCR reaction described above was "repeated" using the *Pfu* DNA polymerase (Stratagene; LaJolla, CA). Specifically, the PCR reaction and protocol were performed exactly as described above, with the exception that *Pfu* polymerase and buffers from Stratagene were used. A PCR product having flush or blunt ends was produced. This PCR product was ligated directly into the XcaI site (GTA/ATC) of pGP704. The ligation mixture was

The transformants were screened using the *Drdl/npr-sacB* and *TthIII/npr-sacB* PCR primers (SEQ ID NOs:9 and 10, respectively) to identify vectors containing the *npr-sacB* insert. The PCR products were analyzed on a 0.8% agarose gel. Plasmid DNA was isolated from cells containing the pGP704::sacB vector.

**EXAMPLE 3**

**CONSTRUCTION OF pGP704::sacB::△CAROTENOID GENE OF INTEREST**

The present Example describes the creation of *crt* integration vectors that enable production of deletions within the native *C₃₀* biosynthetic pathway of the *Methylomonas* genome. Specifically, 5 constructs were made based on the positive selection vector pGP704::sacB that enable chromosomal deletions within the *crtN1*, *crtN1/ald*, *crtN2* and *crtN3* genes; additionally, since the *crtN1*, *ald*, *crtN2* genes (*crt cluster*) exist in an operon and these genes are co-transcribed from the same promoter, an additional construct was created that would permit deletion of the promoter for the *crt cluster*. These constructs (i.e., pGP704::sacB::△crtN1, pGP704::sacB::△ald(crtN1), pGP704::sacB::△ald, pGP704::sacB::△crtN2, pGP704::sacB::△crtN3, and pGP704::sacB::△ promoter *crt cluster*, pGP704::sacB::△ *crt cluster*) were generated using standard PCR and cloning methods, as described below.

**PCR Amplification and Cloning of the Carotenoid Deletion DNA Fragments into pGP704::sacB.**

For amplification of the subsequent PCR fragments [crtN1 deletion fragment (~3.5kb), *crtN1/ald* deletion fragment #2 (~1.1 kb), *crtN1/ald* deletion fragment #2 (~1.1 kb), *ald* deletion fragment #1 (~1.1 kb), *crtN2* deletion fragment #1 (~1.1 kb), *crtN2* deletion fragment #2 (~1.0kb), *crtN3* deletion fragment #1 (~1.2 kb), *crtN3* deletion fragment #1 (~1.1kb), *crt* cluster promoter deletion fragment #1 (~2.6 kb) *crt* cluster promoter deletion fragment #2 (1.1 kb), *crt* cluster deletion fragment #1 (~2.1 kb), and *crt* cluster deletion #2 (2.0 kb)], the following DNA primers (Table 3) were used, along with *Methylomonas* sp. 16a chromosomal DNA as template. The methodology used for PCR reactions and cloning into *E. coli* TOP10 One Shot™ cells were the same as previously described in Example 2. Several colonies from each transformation were screened for
the proper insert DNA fragments using the QIAprep® Spin Mini-prep Kit for plasmid isolation.

Table 3

<table>
<thead>
<tr>
<th>Deletion Fragment</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Size of PCR Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>crtN1 deletion fragment</strong></td>
<td><em>BglI/crtN1</em> (deletion): 5’AGATCTCGGACGCT ACCTCCACAT3’ (SEQ ID NO:11)</td>
<td><em>Sphi/crtN1</em> (deletion): 5’GCATGGCGGAGGTAGA CGTCAACCAC3’ (SEQ ID NO:12)</td>
<td>~3.5 kB</td>
</tr>
<tr>
<td><strong>crtN1/ald deletion fragment #1</strong></td>
<td><em>BglI/ald</em> (deletion) #1: 5’AGATCTTTGCCAACG GGTATTCGAGGAAGG 3’ (SEQ ID NO:13)</td>
<td>*Sphi/NotI/ ald (deletion)#1: 5’GCATGGCGGCGGCGG CCGATGTTTTGCGGAA TCAGC3’ (SEQ ID NO:14)</td>
<td>~1.1 kB</td>
</tr>
<tr>
<td><strong>crtN1/ald deletion fragment #2</strong></td>
<td><em>NotI/ald</em> (deletion) #2: 5’GCGGCGGCCACGACGC AGGCGCGCAG3’ (SEQ ID NO:15)</td>
<td><em>Sphi/ald</em> (deletion) #2: 5’GCATGGTTGCGGAC TACAGCGGTCC3’ (SEQ ID NO:16)</td>
<td>~1.1 kB</td>
</tr>
<tr>
<td><strong>ald deletion fragment #1</strong></td>
<td><em>BglI/ald</em> (deletion) #1: 5’AGATCTTTGCCAACG GGTATTCGAGGAAGG 3’ (SEQ ID NO:13)</td>
<td>*Sphi/Xhol/ ald (deletion) #1: 5’GCGATGGCCTGAGTGCT ATCGTGCTAATCTGAC GCTTT3’ (SEQ ID NO:17)</td>
<td>~1.1 kB</td>
</tr>
<tr>
<td><strong>ald deletion fragment #2</strong></td>
<td><em>Xhol/ald</em> (deletion) #2: 5’CTCGAGAATACGCA AGGCGCTGCAAGC3’ (SEQ ID NO:18)</td>
<td><em>Sphi/ald</em> (deletion) #2: 5’GCATGGCTTGGCGGAC TACAGCGGTCC3’ (SEQ ID NO:16)</td>
<td>~1.0 kB</td>
</tr>
<tr>
<td><strong>crtN2 deletion fragment #1</strong></td>
<td><em>BglI/crtN2</em> (deletion) #1: 5’AGATCTCTCAGGATCC CGTCGATGCTGGA3’ (SEQ ID NO:19)</td>
<td><em>Sphi/crtN2</em> (deletion) #1: 5’GCATGGGCTTGGGTGAT GATCGTGGACG3’ (SEQ ID NO:20)</td>
<td>~1.1 kB</td>
</tr>
<tr>
<td><strong>crtN2 deletion fragment #2</strong></td>
<td><em>MluI/crtN2</em> (deletion) #2 5’ACGCCGGCTGTAGGCT ACGAATTTGGG3’ (SEQ ID NO:21)</td>
<td><em>Sphi/crtN2</em> (deletion) #2 5’GCATGGCTTGGTGGT GATCGAAGG3’ (SEQ ID NO:22)</td>
<td>~1.0 kB</td>
</tr>
<tr>
<td><strong>crtN3 deletion fragment #1</strong></td>
<td><em>BglI/crtN3</em> (deletion) #1 5’AGATCTCGGTCTGAC TACCTGATCC G3’ (SEQ ID NO:23)</td>
<td><em>BglI-NotI/crtN3</em> (deletion) #1 5’AGATCTCGGCGCGCCA TTTGTGGTGTAGAATCT CGG3’ (SEQ ID NO:24)</td>
<td>~1.2 kB</td>
</tr>
<tr>
<td><strong>crtN3 deletion fragment #2</strong></td>
<td>5’NotI/crtN3 (deletion) #2 5’GGGCGCCGAAG CCAGAAGGTGAGGAT TCC3’ (SEQ ID NO:25)</td>
<td>3’NotI/crtN3 (deletion)#2 5’GCCGCGGCGGCGAATTC TCGACATTCAAGC 3’ (SEQ ID NO:26)</td>
<td>~1.1 kB</td>
</tr>
<tr>
<td><strong>BglI(truncated crtN1):</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deletion Fragment</td>
<td>Forward Primer</td>
<td>Reverse Primer</td>
<td>Size of PCR Fragment</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-------------------------------------------</td>
<td>-------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td><em>crt</em> cluster promoter deletion fragment #1</td>
<td>5'AGATCTAATGCTGC GAGGCGCGTGCAGC3' (SEQ ID NO:27)</td>
<td>5'GCATGCGGGCATCTGCTAG TTGCTGAGC3' (SEQ ID NO:28)</td>
<td>~2.6</td>
</tr>
<tr>
<td><em>crt</em> cluster promoter deletion fragment #2</td>
<td>BglII(promoter deletion): 5'AGATCTTGGCGGTG ATGCGAATCCTGTCG3' (SEQ ID NO:29)</td>
<td>NolI(promoter deletion): 5'GGCGGCGCTGTGCATG CGAATGCTCAGC3' (SEQ ID NO:30)</td>
<td>~1.1 kB</td>
</tr>
<tr>
<td><em>crt</em> cluster deletion fragment #1</td>
<td>BglII/crt cluster (deletion) #1 5'AGATCTTGGCGGTG ATGCGAATCCTGTCG3' (SEQ ID NO:48)</td>
<td>Sphl-MluI/crtN1 (deletion) 5'GCATGCGACGCGGCTGTGCATG GCGAACGATGCATCAGCAGA CGTGGCAATGCTCAGC3' (SEQ ID NO:49)</td>
<td>~2.1 kB</td>
</tr>
<tr>
<td><em>crt</em> cluster deletion fragment #2</td>
<td>MluI/crtN2 (deletion) 5'ACGCGTGTGGATGAGG ACAAGATTGTGG3' (SEQ ID NO:50)</td>
<td>Sphl/crt cluster (deletion) #2 5'GCATGCCATCTAAGTC CCAGGCGCCCTA3' (SEQ ID NO:51)</td>
<td>~2.0 kB</td>
</tr>
</tbody>
</table>

*Underlined sequences represent restriction endonuclease recognition sites.

pGP704::sacB::crtN1

Specifically, the digestion and re-ligation of a single DNA fragment that resulted in the internal deletion of the *crtN1* gene from the *Methylophanas sp.* genome generated the re-NSI for the deletion of the *crtN1* gene.

The *crtN1* deletion fragment was digested with *HindIII* and *Xhol*. The restriction digestion products were separated on a 0.8% agarose gel, subsequently excised from the agarose gel, and the DNA was extracted using the Qiaquick® Gel Extraction Kit (Qiagen). Subsequently, the *HindIII/Xhol crtN1* fragment was digested with *BglII* and *Sphl* (~3.5 kB) and was ligated into PGP704::sacB digested with *BglII* and *Sphl*.

Following overnight room temperature incubation, the ligation mixture was used to transform calcium chloride competent *E. coli* SY327 cells (Miller, V., L., and Mekalanos, J., J., *supra*, 1984). The transformation mixture was plated onto LB+Amp^25 agar plates. Transformants were screened for the appropriate insert DNA using PCR methodology and PCR primers *Sphl/crtN1*(deletion) (SEQ ID NO:12) and *BglII/crtN1* (deletion) (SEQ ID NO:11). A plasmid containing the correct insert DNA was digested with *NolI* and re-ligated. Since intra-molecular ligation events occur more frequently than inter-molecular ligation events, the ligation reaction was carried out without purifying the DNA fragment. The ligation mixture was used to transform calcium chloride competent *E.
coli SY327 cells as described above. Transformants were screened for the correct insert DNA fragment (~2.3 kb) by restriction digestion with BglII and SphI. E. coli cells containing pGP704::sacB::ΔcrtN1 were streaked for isolated colonies.

5 pGP704::sacB::Δald(ΔcrtN1)

The re-NSI assembled for the deletion of the Methylo monas sp. ald and the carboxyl terminus of the crtN1 gene was prepared by ligation of two PCR fragments flanking ald (i.e., crtN1/ald deletion fragment #1 and crtN1/ald deletion fragment #2) into pGP704::sacB. Specifically, crtN1/ald deletion fragment #1 was digested with BglII and SphI and crtN1/ald deletion fragment #2 was digested with NotI and SphI. As described previously, the restriction digestion products were separated on a 0.8% agarose gel, excised from the agarose gel, and the DNA was extracted using the Qiaquick® Gel Extraction Kit (Qiagen).

Purified crtN1/ald deletion fragment #1 was ligated into pGP704::sacB digested with BglII and SphI. PCR amplification using PCR primers BglII/ald (deletion)#1 (SEQ ID NO:13) and SphI-NotI/ald (deletion)#1 (SEQ ID NO:14) was used to detect E. coli transformants containing the correct DNA insert fragment. Plasmid DNA was purified from positive clones using the QIA prep® Spin Mini-prep Kit (Qiagen) and the resulting vector was digested with NotI and SphI.

Purified crtN1/ald deletion fragment #2 was ligated into the linearized pGP704::sacB::ald (ΔcrtN1) deletion fragment #1 vector described above. After overnight incubation at room temperature, the ligation mixture was used to transform E. coli SY327 cells and transformants were selected on LB+Amp25 agar plates. PCR amplification using PCR primers BglII/ald (deletion)#1 (SEQ ID NO:13) and SphII/ald (deletion)#2 (SEQ ID NO:16) was used to detect E. coli transformants containing the appropriate DNA insert fragment. Subsequently, E.coli cells containing pGP704::sacB::Δald(ΔcrtN1) were streaked onto LB+Amp25 agar plates to obtain isolated colonies.

pGP704::sacB::Δald

The re-NSI assembled for the deletion of the Methylo monas sp. ald gene was prepared by ligating two PCR fragments flanking ald (i.e., ald deletion fragment #1 and ald deletion fragment #2) into pGP704::sacB. Specifically, ald deletion fragment #1 was digested with BglII and SphI and ald deletion fragment #2 was digested with Xhol and SphI. As described previously, the restriction digestion products was separated on a 0.8%
agarose gel, excised from the agarose gel, and the DNA was extracted using the Qiaquick® Gel Extraction Kit (Qiagen).

Purified ald deletion fragment #1 was ligated into pGP704::sacB digested with BgIII and SphI. PCR amplification using PCR primers BgIII/ald (deletion)#1 (SEQ ID NO:13) and SphI-XhoI/ald (deletion)#1 (SEQ ID NO:17) was used to detect E. coli transformants containing the correct DNA insert fragment. Plasmid DNA was purified from positive clones using the QIAprep® Spin Mini-prep Kit (Qiagen) and the resulting vector will be digested with XhoI and SphI.

Purified ald deletion fragment #2 was ligated into the linearized pGP704::sacB::ald deletion fragment #1 vector described above. After overnight incubation at room temperature, the ligation mixture was used to transform E. coli SY327 cells and transformants was selected on LB+Amp25 agar plates. PCR amplification using PCR primers BgIII/ald (deletion)#1 (SEQ ID NO:13) and SphI/ald (deletion)#2 (SEQ ID NO:16) was used to detect E. coli transformants containing the appropriate DNA insert fragment. Subsequently, E.coli cells containing pGP704::sacB::Δald was streaked onto LB+Amp25 agar plates to obtain isolated colonies.

The construction of the re-NSI for the deletion of the crtN2 gene from the Methylomonas genome was carried out through the ligation of two PCR fragments flanking the crtN2 gene (i.e., crtN2 deletion fragment #1 and crtN2 deletion fragment #2) into integration vector pGP704::sacB.

The crtN2 deletion fragment #1 was removed from the TOPO TA vector by digestion with BgIII and SphI and the crtN2 deletion fragment #2 was excised from the TOPO TA vector by digestion with MluI and SphI. The restriction digestion mixture was separated by electrophoresis on a 0.8% agarose gel and the desired DNA fragments were excised and the DNA was extracted using the Qiaquick® Gel Extraction Kit (Qiagen) as described above.

The purified crtN2 deletion fragment #1 was ligated into pGP704::sacB digested with BgIII and SphI and the ligation mixture was used to transform calcium chloride competent E. coli SY327 cells as described above. PCR methodology using PCR primers BgIII/crtN2 (deletion) #1 (SEQ ID NO:19) and SphI/crtN2 (deletion) #1 (SEQ ID NO:20) was used to detect transformants that contained the desired DNA insert fragment. Plasmid DNA was purified from a positive clone using the QIAprep® Spin Mini-prep Kit (Qiagen), the pGP704::sacB::crtN2 deletion...
fragment #1 was digested with MluI and SphI. The MluI restriction site was located in the terminal portion of the crtN2 gene.

The purified crtN2 deletion fragment #2 was ligated into MluI and SphI digested pGP704::sacB::crtN2 deletion fragment #1 vector DNA. Following an overnight incubation at room temperature, the ligation mixture was transformed into E. coli SY327 cells and transformants were selected by plating onto LB+Amp25 agar plates. PCR amplification using PCR primers BglII/crtN2 (deletion) #1 (SEQ ID NO:19) and SphI/crtN2 (deletion) #2 (SEQ ID NO:22) was used to detect transformants containing the correct DNA insert fragment. Afterwards, E. coli cells containing pGP704::sacB::ΔcrtN2 were streaked onto fresh medium to obtain isolated colonies.

pGP704::sacB::ΔcrtN3

The re-NSI used to delete the crtN3 gene from the Methylophanas genome was generated by ligating two PCR fragments (i.e., crtN3 deletion fragment #1 and crtN3 deletion fragment #2) into pGP704::sacB.

The crtN3 deletion fragment #1 (~1.1 kb) was excised from pCR2.1 (TOPO TA vector) by restriction digestion with BamHI and XhoI. The restriction digestion mixture was separated on a 0.8% agarose gel and the crtN3 deletion fragment #1 was extracted using the Qiaquick® Gel Extraction Kit (Qiagen). This BamHI and XhoI fragment was then digested with BglII and was ligated into the BglII site of the dephosphorylated pGP704::sacB. After an overnight incubation at room temperature, the ligation mixture was used to transform calcium chloride competent E. coli SY327. The transformation mixture was plated onto LB+Amp25 agar plates. Individual colonies were screened for the appropriate insert DNA using PCR methodology and PCR primers BglII/crtN3 (deletion) #1 (SEQ ID NO:23) and BglII-NotI/crtN3 (deletion) #1 (SEQ ID NO:24) with Methylophanas sp. 16a plasmid DNA as template. Plasmid DNA was isolated from the positive clones, pGP704::sacB::crtN3 deletion fragment #1.

The crtN3 deletion fragment #2 was isolated from the TOPO TA vector by digestion with EcoRI and was separated on a 0.8% agarose gel. The ~1.1 kb DNA fragment was extracted from the gel using the Qiaquick® Gel Extraction Kit. The crtN3 deletion fragment #2 was digested with NotI and ligated into the dephosphorylated NotI site of pGP704::sacB::crtN3 deletion fragment #1. The ligation mixture was used to transform E. coli SY327. Several colonies were screened using PCR methodology (Perkin
Elmer AmpliTaq® and Epicentre Fail-Safe™ enzymes) with the BglII/crtN3 (deletion) #1 (SEQ ID NO:23) and the 3’NotI/crtN3 (deletion) #2 (SEQ ID NO:26) and plasmid template DNA. Plasmid DNA was isolated from the positive clone and digested with MluI and NdeI to confirm the presence of the correct insert DNA fragment. *E. coli* cells containing pGP704::sacB::ΔcrtN3 were streaked onto fresh medium to obtain isolated colonies.

pGP704::sacB::Δ promoter crt cluster

To prepare for the construction of the *crt* cluster promoter deletion vector (pGP704::sacB::Δ promoter *crt* cluster), an intermediary vector was generated, pGP704::sacB::hybrid. The components of pGP704::sacB::hybrid were pGP704::sacB, (*ald*) deletion fragment #1 and *crtN3* deletion fragment #2. The purpose of this vector was to make it easier to distinguish between fragments that had been cut with two restriction endonucleases as opposed to only one. This can be visualized on an agarose gel with the presence of an ~1.1 kB fragment when digested with BglII and Sphi.

The BglII and Sphi digested pGP704::sacB::hybrid was ligated with the *crt* cluster promoter deletion fragment #1 (~2.6 kB) which had been prepared using methods similar to those described above. The ligation mixture was used to transform *E. coli* SY327 and the transformation mixture was plated onto LB+Amp25 agar plates. Colonies containing the correct insert DNA fragment were screened using plasmid isolation, restriction digestion and agarose gel electrophoresis.

The pGP704::sacB::crt cluster promoter deletion fragment #1 was digested with BglII and NotI, separated on a 0.8% agarose gel and extracted from the agarose gel using the Qiaquick® Gel Extraction Kit. The BglII and NotI digested pGP704::sacB::crt cluster promoter deletion fragment #1 was ligated with the *crt* cluster promoter deletion fragment #2.

The ligation mixture was used to transform *E. coli* SY327 and was plated onto LB+Amp25 agar plates as described above. Colonies containing the correct insert DNA fragment were identified by plasmid isolation and restriction digestions using methods similar to those described above. Cells containing positive vectors (pGP704::sacB::Δ *crt* cluster promoter) were streaked for isolated colonies.

pGP704::sacB::Δcrt cluster

The construction of the re-NS1 for the deletion of the *crt* cluster from the *Methylomonas* genome was carried out through the ligation of two
PCR fragments flanking the *crt* cluster (i.e. *crt* cluster deletion fragment #1 and *crt* cluster deletion fragment #2) into integration vector pGP704::*sacB*::CDstuffer1 (contains two 1 kb DNA fragments that flank the *crt* cluster). This vector was originally constructed for the purpose of deleting the *crt* cluster. However, the identification of trans-conjugates in which the *crt* cluster had been deleted were not identified from the *Methylomonas* genome. Therefore, a different approach was used by extending the size of the flanking DNA fragments to ~2 kb to increase the probability of identifying cells that have undergone a double crossover event leading to the deletion of the *crt* cluster.

The *crt* deletion fragment #1 was excised from the TOPO TA vector by digestion with *BglII* and *MluI* and the *crt* deletion fragment #2 was removed from the TOPO TA vector by digestion with *MluI* and *SpHl*. The restriction digestion mixture was separated by electrophoresis on a 0.8% agarose gel. The correct DNA fragments were excised and the DNA was extracted using the Qiaquick Gel Extraction Kit (Qiagen) using the same method as described above.

The purified *crt* cluster deletion fragment #2 as ligated into pGP704::*sacB*::CDstuffer1 digested with *MluI* and *SpHl* and the ligation mixture was used to transform calcium chloride competent *E. coli* SY327 cells as described above. Plasmid DNA was purified from several transformants using the QIAprep Spin Mini-prep Kit (Qiagen). Restriction digestions analysis using restriction endonucleases *MluI* and *SpHl* was used to identify transformant that contained the *crt* cluster deletion fragment #2. Plasmid DNA that contain the correct insert fragment were subsequently digested with *BglII* and *MluI*.

The purified *crt* cluster deletion fragment #1 was ligated into *BglII* and *MluI* digested pGP704::*sacB*::*crt* cluster deletion fragment #2 vector DNA. Following an overnight incubation at room temperature, the ligation mixture was transformed into *E. coli* SY327 cells and transformants were selected by plating onto LB/Amp^60_ agar plates. Plasmid DNA was isolated from several colonies and restriction digestion with *BglII* and *MluI* was used to identify plasmid DNA contain the correct insert DNA.

The pGP704::*sacB*::*crt* cluster deletion vector also contains a gene that confers resistance to ampicillin. However, ampicillin selection is not straight forward when culturing *Methylomonas*. *In vitro* transposition was used to inactivate the ampicillin resistance gene and to insert a kanamycin resistance gene to the vector. The pGP704::*sacB*::*crt* cluster deletion
vector severed as the recipient DNA in the \textit{in vitro} transposition reaction (Epicentre, Madison, WI). Following the manufacturer's protocol, the EZ::TN<Kan-2> transposon was used to randomly insert the kanamycin resistance gene into the pGP704::sacB::crt cluster deletion vector. The transposition reaction was heat inactivated at 70°C for 10 minutes and the transposition mixture was used to transform calcium chloride competent \textit{E. coli} SY327 cells. The transformation mixture was plated onto LB + Kan^{50} plates. The colonies were patched on LB+Amp^{50} and LB+Kan^{50} to identify cells that contained the pGP704::sacB::crt cluster deletion vector that had received the EZ::TN<Kan-2> insertion into the ampicillin resistance gene. \textit{E. coli} cells that were Kan^{R} and Amp^{S} were streaked onto fresh medium and severed as the donor cells in the tri-parental matings.

\textbf{EXAMPLE 4}

\textbf{TRI-PARENTAL CONJUGATION OF THE VARIOUS CRT INTEGRATION VECTORS INTO MYTHYLOMONAS SP. 16A}

Each of the \textit{crt} integration vectors from Example 3 (i.e., pGP704::sacB::ΔcrtN1, pGP704::sacB::Δald(crtN1), pGP704::sacB::Δald, pGP704::sacB::ΔcrtN2, pGP704::sacB::ΔcrtN3, pGP704::sacB::Δcrt) cluster promoter, and pGP704::sacB::crt cluster deletion) were transferred into \textit{Methylomonas sp.} 16a via triparental conjugation. Specifically, the following were used as recipient, donor, and helper, respectively: \textit{Methylomonas sp.} 16a, \textit{E. coli} SY327 containing the \textit{crt} integration vectors, and \textit{E. coli} containing pRK2013.

\textbf{Theory of the Conjugation}

The mobilization of vector DNA into \textit{Methylomonas} occurs through conjugation (tri-parental mating). The pGP704::sacB vector used to make chromosomal mutations in \textit{Methylomonas} has a R6K origin of replication, which requires the \textit{II} protein. This vector can replicate in \textit{E. coli} strain SY327, which expresses the \textit{II} protein. However, this protein is not present in the \textit{Methylomonas} genome. Therefore, once the vector DNA has entered into \textit{Methylomonas}, it is unable to duplicate itself. If the vector also contains a DNA segment that shares homology to a region of the \textit{Methylomonas} genome, the vector can be integrated into the host's genome through homologous recombination. The homologous recombination system of \textit{Methylomonas} appears to be similar to that of other Gram-negative organisms.
In the case of *Methylomonas*, the mobilizable plasmid (pGP704::sacB) was used to transfer the re-NSI into this bacterium. The conjugative plasmid (pRK2013), which resided in a strain of *E. coli*, facilitated the DNA transfer.

**Growth of *Methylomonas sp. 16a***

The growth of *Methylomonas sp. 16a* for tri-parental mating initiated with the inoculation of an -80°C frozen stock culture into 20 mL of BTZ medium containing 25% methane, as described in Example 1. The culture was grown at 30°C with aeration until the density of the culture was saturated. This saturated culture was in turn used to inoculate 100 mL of fresh BTZ medium containing 25% methane. The 100 mL culture was grown at 30°C with aeration until the culture reached an OD₆₀₀ between 0.7 to 0.8. To prepare the cells for the tri-parental mating, the *Methylomonas sp. 16a* cells were washed twice in an equal volume of BTZ medium. The *Methylomonas* cell pellets were re-suspended in the minimal volume needed (approximately 200 to 250 μL). Approximately 40 μL of the re-suspended *Methylomonas* cells were used in each tri-parental mating experiment.

**Growth of the *Escherichia coli* Donor and Helper Cells**

Isolated colonies of the *E. coli* donor (pGP704::sacB::NSI) and helper (containing conjugative plasmid pRK2013) cells were used to inoculate 5 mL of LB broth containing 25 μg/μL Kan; these cultures were grown overnight at 30°C with aeration. The following day, the *E. coli* donor and helper cells were mixed together and incubated at 30°C for ~2 hours. Subsequently, the cells were washed twice in equal volumes of fresh LB broth to remove the antibiotics.

**Tri-parental Mating: Mobilization of the Donor Plasmid into *Methylomonas* Strain 16a**

Approximately 40 μL of the re-suspended *Methylomonas* cells were used to re-suspend the combined *E. coli* donor and helper cell pellets. After thoroughly mixing the cells, the cell suspension was spotted onto BTZ agar plates containing 0.05% yeast extract. The plates were incubated at 30°C for 3 days in a jar containing 25% methane.

Following the third day of incubation, the cells were scraped from the plate and re-suspended in BTZ broth. The entire cell suspension was plated onto several BTZ agar plates containing Amp³⁶. The plates were incubated at 30°C in a jar containing 25% methane until colonies were visible (~4-7 days).
Individual colonies were streaked onto fresh BTZ + Amp\textsuperscript{35} agar plates and incubated 1-2 days at 30°C in the presence of 25% methane. These cells were used to inoculate bottles containing 20 mL of BTZ and 25% methane. After overnight growth, 5 mL of the culture was concentrated by centrifugation using a tabletop centrifuge. Then, to rid the cultures of *E. coli* cells that were introduced during the tri-parental mating, the cells were inoculated into 20 mL of BTZ liquid medium containing nitrate (10 mM) as the nitrogen source, methanol (200 mM), and 25% methane and grown overnight at 30°C with aeration. Cells from the BTZ (nitrate) cultures were again inoculated into BTZ and 25% methane and grown overnight at 30°C with aeration. The cultures were monitored for *E. coli* growth by plating onto LB agar plates to verify the success of the *E. coli* elimination.

**EXAMPLE 5**

**EVALUATION OF METHYLOMONAS TRANSCONJUGANTS CONTAINING THE VARIOUS CRT INTEGRATION VECTORS**

Following the mobilization of the various *crt* integration constructs into *Methyloomonas* sp. 16a, as described in Example 4, a two-step selection strategy was applied as described below to identify the \( \Delta \text{crtN1} \), \( \Delta \text{ald(crtN1)} \), \( \Delta \text{ald} \), \( \Delta \text{crtN2} \), \( \Delta \text{crtN3} \), \( \Delta \text{crt} \) cluster promoter, and \( \Delta \text{crt} \) cluster allelic exchange mutants (see also Figure 4). Four "white" or "pigmentless" mutants were produced comprising the \( \Delta \text{crtN1} \), \( \Delta \text{ald(crtN1)} \), \( \Delta \text{crt} \) cluster promoter, or \( \Delta \text{crt} \) cluster. In contrast, successful deletion of \( \text{ald} \), \( \text{crtN2} \) or \( \text{crtN3} \) genes resulted in *Methyloomonas* transformants that still possessed the ability to produce a pink pigment. However, the carotenoid(s) that is produced in each mutant is distinct, indicating that the enzymes act at different stages along the pathway for C\textsubscript{30} carotenoid synthesis.

**Preliminary Screening for Allelic Exchange Mutants**

Cultures free of *E. coli* cells were passaged several times in fresh medium (1 mL of culture into 20 mL of fresh BTZ medium), to increase the probability of occurrence of a second crossover event. Subsequently, cells were plated onto BTZ and sucrose (5%) agar plates. Those cells grown on plates containing sucrose had lost the integration vector, which contained the \( \text{sacB} \) gene. However, the loss of the vector sequences could be due to the second crossover event occurring either on the same or opposite side of the re-NSI that was present on the insert DNA. If the second crossover event had occurred on the same side of the re-NSI as
the first crossover event, the wildtype gene of interest would be regenerated. In contrast, if the second crossover event occurred on the opposite side of the re-NSI as the first crossover event, the deletion of the gene of interest would be established in the *Methylomonas* genome.

**Verification of the Chromosomal Deletion of the *Methylomonas sp. 16a* Carotenoid Genes**

Chromosomal DNA was purified from several cultures that had grown on the sucrose plates using the MasterPure™ DNA Purification Kit (EPICENTRE®, Madison, WI). Then, PCR amplification methods were applied to confirm each suspected deletion, using the primers described below in Table 4.
## Table 4
Primers Used to Verify the Deletion of the *Methylomonas sp.* 16a Carotenoid Genes

<table>
<thead>
<tr>
<th>Carotenoid Gene(s)</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
<th>Intact Fragment</th>
<th>Deletion Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>crtN1</strong></td>
<td><em>BglII/crtN1</em> (deletion) 5'-AGATCTGCACGTCATCATC TCCCATC-3' (SEQ ID NO:11)</td>
<td><em>SphII/crtN1</em> (deletion) 5'-GCATGGCGAGA GGTAGACGTAGACAC AC-3' (SEQ ID NO:12)</td>
<td>~3.5kB</td>
<td>~2.2 kB</td>
</tr>
<tr>
<td><strong>crtN1</strong></td>
<td><em>NcoI/crtN1</em> 5'-CCATGGCCAACA CCAACACATCATC ACGCAGG-3' (SEQ ID NO:52)</td>
<td><em>ScaI/Ndel/crtN1</em> 5'-AGTACTCCGATAT GTCGGGTTTGGGTT TTGCTTTTTCCACG CAGGC3' (SEQ ID NO:53)</td>
<td>1.7 kB</td>
<td>0.3 kB</td>
</tr>
<tr>
<td><strong>crtN1/ald</strong></td>
<td><em>BglII/ald</em> (deletion) #1: 5'-AGATCTCTTGCA GCTATATC TCGGAGG-3' (SEQ ID NO:13)</td>
<td><em>SphII/ald</em> (deletion) #2: 5'-GCATGCCGTT GCGGATACAGCTG TCC-3' (SEQ ID NO:16)</td>
<td>~3.7 kB</td>
<td>~2.1 kB</td>
</tr>
<tr>
<td><strong>crtN2</strong></td>
<td><em>BglII/crtN2</em> (deletion) #1: 5'-AGATCTCTACG ATTCCGTCATT CGGAGG-3' (SEQ ID NO:19)</td>
<td><em>SphII/crtN2</em> (deletion) #2: 5'-GCATGCTGCG TTTGATAGCAGTAGE CG-3' (SEQ ID NO:22)</td>
<td>~3.4 kB</td>
<td>~2.1 kB</td>
</tr>
<tr>
<td><strong>ald</strong></td>
<td><em>BglII/ald</em> (deletion) #1 5'-AGATCTGGAACA GGGTAATCCGAGC AAGG3' (SEQ ID NO:13)</td>
<td><em>SphII/ald</em> (deletion) #2 5'-GCATGCCGTTGGGGCA TACAGGCGTACC3' (SEQ ID NO:16)</td>
<td>3.7 kB</td>
<td>2.2 kB</td>
</tr>
<tr>
<td><strong>crtN3</strong></td>
<td><em>BglII/crtN3</em> (deletion) #1: 5'-AGATCCTGGAT CTGTGACTGATC TCC-3' (SEQ ID NO:23)</td>
<td><em>NotI/crtN3</em> (deletion)#2: 5'-GGGCCGCGCCG AATACCTGACATT CAGG-3' (SEQ ID NO:26)</td>
<td>~3.5 kB</td>
<td>~2.3 kB</td>
</tr>
<tr>
<td><em>crt cluster</em></td>
<td><em>BglII</em> (truncated <em>crtN1</em>): 5'-AGATCTAATGT GCAGGCAGC GGCAGC GACG-3' (SEQ ID NO:27)</td>
<td><em>NotI</em>(promoter deletion) 5'-GGGCGCGCGCTG TCATTGGAATTACAT CAGC-3' (SEQ ID NO:30)</td>
<td>~4.3 kB</td>
<td>~2.1 kB</td>
</tr>
<tr>
<td><em>crt cluster</em></td>
<td><em>BglII</em>(promoter deletion): 5'-AGATCTGGGTCGGC TTGATGAAATG CTGGC-3' (SEQ ID NO:29)</td>
<td><em>SphII/crtN2</em> (deletion) #2: 5'-GCATGCTGCG TTTGATAGCAGTAGE CG-3' (SEQ ID NO:22)</td>
<td>~7.2 kB</td>
<td>~2.1 kB</td>
</tr>
</tbody>
</table>

**Underlined sequences represent restriction endonuclease recognition sites.**
ΔcrtN1, Δald(ΔcrtN1), Δcrt Cluster Promoter, and Δcrt Cluster Mutants

The *Methylomonas* strains that had “white” phenotypes, designated herein as (MWM1400 (ΔcrtN1), MWM1000 (Δald(ΔcrtN1)), MWM1100 (Δcrt cluster promoter), and MWM1600 (Δcrt cluster) were easily distinguished from the wild-type cells. However, the construction of these strains was still verified via PCR amplification using PCR primers *BglII/ald* (deletion) #1 (SEQ ID NO:13) and *SphI/ald* (deletion) #2 (SEQ ID NO:16), which permitted cells that contained an intact *ald* gene (wherein the expected size of the PCR product was ~3.7 kB, see Table 4) to be distinguished from those that contained a deletion of the *ald* gene (wherein the expected size of the PCR product was ~2.1 kB). Two of the three cultures produced a PCR product that was ~2.1 kB, while the third culture appeared to have the wildtype gene present. Thus, these two PCR reactions confirmed the deletion of the *crtN1/ald* genes in the genome of *Methylomonas* MWM1000. To verify the construction of the *crtN1* deletion mutant, PCR amplification using PCR primers *BglII/crtN1* (deletion) (SEQ ID NO: 11) and *SphI/crtN1* (deletion) (SEQ ID NO: 12) was employed. Cells having an intact *crtN1* gene produced a DNA fragment that was ~3.5 kB, whereas cells comprised of a disrupted *crtN1* gene produced a DNA fragment of ~2.2 kB. An additional pair of primers was used to verify the construction of the *crtN1* deletion. PCR primers *NcoI/crtN1* (SEQ ID NO: 52) and *ScaI/NdeI/crtN1* (SEQ ID NO: 53) were used as an additional set of primers to verify the construction of the *crtN1* deletion. Cells having an intact *crtN1* gene produced a DNA fragment that was ~1.7 kB, whereas cell possessing a disrupted *crtN1* gene resulted in a fragment of ~0.3 kB. Similarly, the construction of *Methylomonas* MWM1100 (Δcrt cluster promoter) was confirmed using PCR methodology and PCR primers *BglII* (truncated *crtN1*) (SEQ ID NO:27) and *NsiI* (promoter deletion) (SEQ ID NO:30). Cells that contained an intact promoter region for the *crt* cluster had the expected PCR product size of ~4.3 kB. In contrast, cells in which the promoter region of the *crt* cluster had been deleted, gave rise to PCR products that were ~2.1 kB (Table 4). Likewise, the construction of *Methylomonas* MWM1600 (Δcrt cluster) was established using PCR methodology and PCR primers *BglII* (promoter deletion) (SEQ ID NO:29) and *SphI/crtN2* (deletion) #2 (SEQ ID NO:22). Cells that remained wild-type for the *crt* cluster were expected to produce a PCR product that was ~7.2 kB. This long PCR fragment could not be detected under our PCR
reaction conditions, thus we used the absence of a PCR product to indicate that the wild-type genes were the product of the double-crossover event. In contrast, the PCR fragment was ~2.1 kB for the cells in which the entire *crt* gene cluster had been deleted.

**ΔcrtN2, ΔcrtN3, and Δald Mutants**

The identification of the “pink” mutants required more effort since these cells could not be easily distinguished from the wild-type cells. Nevertheless, generally less than five cultures were evaluated for the verification of the *Methylomonas* “pink” mutants. Cells containing a deletion of the *crtN2* gene, designated herein as MPM1200, were easily distinguished from wild-type cells using PCR methodology and PCR primers *BglII*/crtN2 (deletion) #1 (SEQ ID NO:19) and *SphI*/crtN2 (deletion) #2 (SEQ ID NO:22). The wild-type cells gave a PCR fragment that was ~3.4 kB, while the cells containing the *crtN2* deletion gave a fragment that was ~2.1 kB (Table 4). The verification of the *Methylomonas* *crtN3* deletion mutant, designated herein as MPM1000, was demonstrated using PCR amplification with PCR primers *BglII*/crtN3 (deletion) #1 (SEQ ID NO:23) and 3′*NotI*/crtN3 (deletion) #2 (SEQ ID NO:26). An ~2.3 kB PCR fragment was detected for the ΔcrtN3 mutant as compared to an ~3.5 kB PCR fragment that was produced by the wild-type *Methylomonas* cells (Table 4). The affirmation of the *Methylomonas ald* deletion mutant, designated herein as MPM1300, was revealed using PCR amplification with PCR primers *BglII*/ald (deletion) #1 (SEQ ID NO:13) and *SphI*/ald (deletion) #2 (SEQ ID NO:16). The wild-type cells gave a PCR fragment that was ~3.7 kB, while cells containing the *ald* deletion gave a fragment that was ~2.2 kB (Table 4).

**EXAMPLE 6**

**ANALYSIS OF C30 CAROTENOIDs PRODUCED BY METHYLOMONAS SP. 16a DELETION MUTANTS USING HPLC-PHOTODIODE ARRAY**

To investigate the effect of deleting genes involved in the biosynthesis of the C30 carotenoid from the *Methylomonas* sp. 16a genome, the carotenoids were extracted from each mutant strain, using two different methods, and evaluated using high performance liquid chromatography with photodiode array detection (HPLC-photodiode array). As a control, the carotenoid from the wild-type *Methylomonas* strain was also extracted and analyzed using similar methods.
Methanol Extraction Method

Each *Methylobacterium* culture (wild-type, Δ.crtN1, Δ.ald(crtN1), Δ.crtN2, Δ.crtN3, Δ.crt promoter deletion, and Δ.crt cluster deletion mutant) was grown with aeration in two 500 mL bottles containing 100 mL of BTZ and 25% methane until they reached saturation (~24 hr). The cells were harvested by centrifugation for 15 min at 4000 rpm. The cell pellet was extracted twice with 10 mL of methanol for 15 min at room temperature with agitation. This step was followed by two extractions with 10 mL of a methanol/acetone (1:1) mixture for 15 min at room temperature also with agitation. The extracted C₃₀ carotenoids were dried with nitrogen and were subsequently re-dissolved in 1 mL of methanol.

THF/Methanol Extraction Method

As in the methanol extraction method, each *Methylobacterium* culture to be analyzed was grown in two 500 mL bottles containing 100 mL of BTZ and 25% methane until saturation (~24 hr). The cells were split into two aliquots and harvested by centrifugation. For one aliquot, the cell pellets were dried in an oven at ~100°C for ~24 hr and the dry cell mass was determined. The other aliquot was used for C₃₀ carotenoid extractions. The cells were lysed using glass beads (0.5 mL/sample). Also added to each sample was 150 μL of the internal standard, ethyl-β-apo-8'-carotene (trans) (100 mg/L stock solution) and 5 mL of a THF (tetrahydrofuran)/methanol (1:1) solution. This mixture was vortexed for ~2 min, followed by a 15 min centrifugation at 8,000 rpm. The supernatant was collected and the sample was vortexed again for ~2 min and centrifuged at 8,000 rpm for another 15 min. The supernatants were combined and were dried using nitrogen. The carotenoid samples were stored at ~80°C until analyzed using HPLC-photodiode array.

HPLC-Photodiode Array Analysis

A Beckman System Gold® HPLC with Beckman Gold® Nouveau Software (Beckman Coulter, Inc. Fullerton, CA) was used for the study. A 0.1 mL aliquot of the crude acetone extraction was loaded onto a 150 mm X 4.6 mm ZORBAX C18 (3.5 μm particles) column (Agilent Technologies Inc., Wilmington, DE). The flow rate was 1 mL/min using a solvent gradient consisting of Buffer A and Buffer B. Buffer A was 95% acetonitrile and 5% dH₂O, Buffer B was 100% tetrahydrofuran. The running program was: 0 – 1st min, 95% Buffer A and 5% Buffer B; 1 - 11th min, linear gradient from 95% Buffer A and 5% Buffer B to 60% Buffer A and 40% Buffer B; 11 – 22nd min, linear gradient from 60% Buffer A and 40% Buffer
B to 50% Buffer A and 50% Buffer B; 22 - 25th min, 95% Buffer A and 5% Buffer B. The spectrum data was collected by a Beckman photodiode array detector (Model 168).

In the HPLC-diode array analysis, two peaks were detected for the wild-type strain, a large peak that eluted at ~5.5 min and a smaller peak that eluted at ~7 min. Although these two peaks had different elution profiles, their UV spectrums were very similar, suggesting that they were similar forms of the same C₃₀ carotenoid (data not shown). When the extracts from Methyloccoccus sp. strain MWM1000 (Δald/crtN1) were evaluated via HPLC-photodiode array, neither of the two peaks seen for wild-type Methyloccoccus sp. 16a were observed. This result suggests that the deletion of the aldehyde dehydrogenase (ald) gene and the crtN1 gene impaired Methyloccoccus' ability to make the C₃₀ carotenoid. This was a surprising result, because it was expected that the ald gene encoded an enzyme that modified a C₃₀ molecule further down the biosynthesis chain. Interestingly, when only the crtN1 gene was deleted, it was also failed to produce the two peaks detected for the wild-type Methyloccoccus. This suggested that perhaps it was the partial deletion of the crtN1 gene and not the deletion of the ald gene in the Δald/crtN1 strain that was responsible of the loss of the pink pigment in these strains. As expected, when the promoter region of the crt cluster (MWM1100) was deleted or the entire crt cluster (MWM1600) was deleted, a “white” mutant was generated (i.e. the C₃₀ carotenoid was no longer produced).

Evaluation of the crtN2 and crtN3 deletion strains yielded very different results when analyzed using HPLC-photodiode array. It was found that Methyloccoccus MPM1200 (ΔcrtN2) had an elution profile that was very similar to that of the wild-type Methyloccoccus sp. 16a, i.e., two peaks were observed. However, the ratios of the two peaks were distinct. The levels of Peak 2 were similar between the two strains. Interestingly, the level of Peak 1 was approximately 75% lower in MPM1200 as compared to Methyloccoccus sp. 16a, i.e., 15 mAU versus 60 mAU, respectively. However, the exact structure of the C₃₀ carotenoid present in each fraction is not known. In contrast to MPM1200, an increase (~15%) in the biosynthesis of the C₃₀ carotenoid was observed in MPM1000 (ΔcrtN3).

As described above, “Clustal W” analysis, conducted to show the relationship between crtN1, crtN2, crtN3, and sqs, revealed that crtN3 is not closely linked to crtN1 and crtN2 (Figure 3). When crtN3 (which
contains sequences that are homologous to domains of other FAD-dependent oxidoreductases) was viewed in context of its surrounding ORFs, it was observed that crtN3 is located at the end of a cluster of ORFs that have high homology to proteins that play a role in fatty acid metabolism (Figure 2). It is hypothesized that the crtN3 may have an indirect role in carotenoid biosynthesis and perhaps the deletion of crtN3 from the *Methylomonas* genome relieved a drain on common intermediates. It is possible that crtN3 may also be involved in fatty acid metabolism; however, additional experiments are necessary to confirm this hypothesis.

**EXAMPLE 7**

**CONSTRUCTION OF THE CAROTENOID DOUBLE MUTANTS:**

MWM1300 (Δ*ald(crtN1) + ΔcrtN3*), MWM1200 (Δcrt cluster promoter + ΔcrtN3, MWM1800 (ΔcrtN1 + ΔcrtN3), and MWM1900 (Δcrt cluster + ΔcrtN3).

Since an increase in the production of the C₃₀ carotenoid was observed when the crtN3 gene was deleted from *Methylomonas* sp. 16a background, it was hypothesized that the crtN3 deletion would also have a positive impact on the synthesis of the C₄₀ carotenoids (i.e. canthaxanthin and astaxanthin). Thus, deletion of crtN3 was made in each of the *Methylomonas* "white" mutants (MWM1000, MWM1100, MWM1400, and MWM1600).

The pGP704::sacB::ΔcrtN3 integration plasmid was transferred into MWM1000, MWM1100, MWM1400, and MWM1600 via conjugation using the same procedures described above in Example 4. Once inside the *Methylomonas*, the crtN3 gene was deleted via homologous recombination using the same two step strategy described in Example 5. Once again, the deletion of the crtN3 gene was confirmed using PCR methodology and PCR primers *BglII/crtN3* (deletion) #1 (SEQ ID NO:23) and 3' *NotI/crtN3* (deletion) #2 (SEQ ID NO:26). If the "white" mutants still contained the intact crtN3 gene, a PCR fragment that was ~3.5 kB was produced. In contrast, cells in which the crtN3 gene was deleted produced an ~2.3 kB PCR fragment (Table 4). The new *Methylomonas* strains that were produced are referred to herein as MWM1200 (Δcrt cluster promoter + ΔcrtN3), MWM1300 (Δ*ald(crtN1) + ΔcrtN3), MWM1800 (ΔcrtN1 + ΔcrtN3), and MWM1900 (Δcrt cluster + ΔcrtN3).
EXAMPLE 8
SYNTHESIS OF CODON-OPTIMIZED GENES FOR EXPRESSION
IN METHYLOMONAS SP. 16A

This Example describes the design and synthesis of a codon-optimized β-carotene ketolase gene **crtW** and a β-carotene hydroxylase gene **crtZ** for production of canthaxanthin, astaxanthin and intermediates in *Methyloomonas sp.* 16a based on the **crtW** and **crtZ** sequences from *Agrobacterium aurantiacum*.

**Source of the Genes**

*Agrobacterium aurantiacum* is a marine bacterium that naturally produces astaxanthin. The relative percentages of astaxanthin and 4-ketozeaxanthin (adenixanthin) to the total carotenoids produced in this strain are reported to be 19.5% and 70.3%, respectively (Yokoyama et al., *Biosci. Biotech. Biochem.*, 58:1842-1844 (1994)). The carotenoid biosynthesis gene cluster in *A. aurantiacum* contains the β-carotene ketolase gene **crtW** and the β-carotene hydroxylase gene **crtZ** (Misawa et al., *J. Bacteriol.*, 177:6575-6584 (1995)). The ketolase and the hydroxylase have low substrate specificity, resulting in the production of many presumed intermediates of astaxanthin, i.e., adonixanthin, adonirubin, canthaxanthin, 3'-hydroxyechinenone, 3-hydroxyechinenone.

To ensure optimal expression in *Methyloomonas sp.* 16a to produce more desired products such as canthaxanthin or astaxanthin, the codons of the *A. aurantiacum** **crtW** and **crtZ** genes were codon optimized based on the preferred codon usage table for *Methyloomonas sp.* 16a (Table 5).

**Determination of Preferred Codon Usage Table for Methyloomonas sp. 16a**

The genome of *Methyloomonas sp.* 16a was sequenced and 201 highly expressed *Methyloomonas* genes, according to microarray analysis, were used to determine the preferred codon usage profile in *Methyloomonas sp.*16a, which is shown in Table 5. The coding regions of these genes, comprising 164,751 bp, were translated by the Editseq program of DNASTAR to the corresponding 54,917 amino acids. The column titled "Number" refers to the number of times a given codon encodes a particular amino acid in the sample of 54,917 amino acids. The column titled "Fraction" refers to the frequency that a given codon encodes a particular amino acid. The stop codons were not included in the coding regions for tabulation.
<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>Codon</th>
<th>Number</th>
<th>Fraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly</td>
<td>GGG</td>
<td>288.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Gly</td>
<td>GGA</td>
<td>300.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Gly</td>
<td>GGU</td>
<td>1168.00</td>
<td>0.27</td>
</tr>
<tr>
<td>Gly</td>
<td>GGC</td>
<td>2541.00</td>
<td>0.59</td>
</tr>
<tr>
<td>Glu</td>
<td>GAG</td>
<td>966.00</td>
<td>0.28</td>
</tr>
<tr>
<td>Glu</td>
<td>GAA</td>
<td>2514.00</td>
<td>0.72</td>
</tr>
<tr>
<td>Asp</td>
<td>GAU</td>
<td>1435.00</td>
<td>0.46</td>
</tr>
<tr>
<td>Asp</td>
<td>GAC</td>
<td>1712.00</td>
<td>0.54</td>
</tr>
<tr>
<td>Val</td>
<td>GUG</td>
<td>1287.00</td>
<td>0.32</td>
</tr>
<tr>
<td>Val</td>
<td>GUA</td>
<td>508.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Val</td>
<td>GUU</td>
<td>717.00</td>
<td>0.18</td>
</tr>
<tr>
<td>Val</td>
<td>GUC</td>
<td>1450.00</td>
<td>0.37</td>
</tr>
<tr>
<td>Ala</td>
<td>GCG</td>
<td>1576.00</td>
<td>0.31</td>
</tr>
<tr>
<td>Ala</td>
<td>GCA</td>
<td>607.00</td>
<td>0.12</td>
</tr>
<tr>
<td>Ala</td>
<td>GCU</td>
<td>658.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Ala</td>
<td>GCC</td>
<td>2279.00</td>
<td>0.45</td>
</tr>
<tr>
<td>Lys</td>
<td>AAG</td>
<td>1055.00</td>
<td>0.35</td>
</tr>
<tr>
<td>Lys</td>
<td>AAA</td>
<td>1988.00</td>
<td>0.65</td>
</tr>
<tr>
<td>Asn</td>
<td>AAU</td>
<td>877.00</td>
<td>0.40</td>
</tr>
<tr>
<td>Asn</td>
<td>AAC</td>
<td>1317.00</td>
<td>0.60</td>
</tr>
<tr>
<td>Met</td>
<td>AUG</td>
<td>1443.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Ile</td>
<td>AUA</td>
<td>301.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Ile</td>
<td>AUU</td>
<td>933.00</td>
<td>0.28</td>
</tr>
<tr>
<td>Amino Acid</td>
<td>Codon</td>
<td>Number</td>
<td>Fraction</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>Ile</td>
<td>AUC</td>
<td>2122.00</td>
<td>0.63</td>
</tr>
<tr>
<td>Thr</td>
<td>ACG</td>
<td>544.00</td>
<td>0.19</td>
</tr>
<tr>
<td>Thr</td>
<td>ACA</td>
<td>263.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Thr</td>
<td>ACU</td>
<td>380.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Thr</td>
<td>ACC</td>
<td>1738.00</td>
<td>0.59</td>
</tr>
<tr>
<td>Trp</td>
<td>UGG</td>
<td>600.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Cys</td>
<td>UGU</td>
<td>151.00</td>
<td>0.24</td>
</tr>
<tr>
<td>Cys</td>
<td>UGC</td>
<td>474.00</td>
<td>0.76</td>
</tr>
<tr>
<td>Tyr</td>
<td>UAU</td>
<td>779.00</td>
<td>0.52</td>
</tr>
<tr>
<td>Tyr</td>
<td>UAC</td>
<td>723.00</td>
<td>0.48</td>
</tr>
<tr>
<td>Phe</td>
<td>UUU</td>
<td>793.00</td>
<td>0.38</td>
</tr>
<tr>
<td>Phe</td>
<td>UUC</td>
<td>1308.00</td>
<td>0.62</td>
</tr>
<tr>
<td>Ser</td>
<td>AGU</td>
<td>317.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Ser</td>
<td>AGC</td>
<td>868.00</td>
<td>0.27</td>
</tr>
<tr>
<td>Ser</td>
<td>UCG</td>
<td>733.00</td>
<td>0.23</td>
</tr>
<tr>
<td>Ser</td>
<td>UCA</td>
<td>318.00</td>
<td>0.10</td>
</tr>
<tr>
<td>Ser</td>
<td>UCU</td>
<td>291.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Ser</td>
<td>UCC</td>
<td>701.00</td>
<td>0.22</td>
</tr>
<tr>
<td>Arg</td>
<td>AGG</td>
<td>186.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Arg</td>
<td>AGA</td>
<td>287.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Arg</td>
<td>CGG</td>
<td>411.00</td>
<td>0.13</td>
</tr>
<tr>
<td>Arg</td>
<td>CGA</td>
<td>250.00</td>
<td>0.08</td>
</tr>
<tr>
<td>Arg</td>
<td>CGU</td>
<td>693.00</td>
<td>0.22</td>
</tr>
<tr>
<td>Arg</td>
<td>CGC</td>
<td>1292.00</td>
<td>0.41</td>
</tr>
<tr>
<td>Gln</td>
<td>CAG</td>
<td>893.00</td>
<td>0.40</td>
</tr>
<tr>
<td>Gln</td>
<td>CAA</td>
<td>1345.00</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Amino Acid  | Codon | Number  | Fraction  
------------|-------|---------|-----------
His         | CAU   | 635.00  | 0.51      
His         | CAC   | 600.00  | 0.49      
Leu         | UUG   | 1825.00 | 0.38      
Leu         | UUA   | 319.00  | 0.07      
Leu         | CUG   | 1980.00 | 0.41      
Leu         | CUA   | 172.00  | 0.04      
Leu         | CUU   | 221.00  | 0.05      
Leu         | CUC   | 277.00  | 0.06      
Pro         | CCG   | 1104.00 | 0.44      
Pro         | CCA   | 443.00  | 0.18      
Pro         | CCU   | 441.00  | 0.18      
Pro         | CCC   | 520.00  | 0.21      

**Design and Synthesis of the Synthetic Genes for β-Carotene Ketolase and β-Carotene Hydroxylase**

The β-carotene ketolase gene (SEQ ID NO:31) and the β-carotene hydroxylase gene (SEQ ID NO:32) from *Agrobacterium aurantiacum* are 729 bp and 489 bp in length, respectively (US 5,972,690; GenBank® D58420). The codon-optimized β-carotene ketolase gene and the β-carotene hydroxylase gene were first designed by back translating the amino acid sequences of the *A. aurantiacum* CrtW and the CrtZ, according to the preferred codon usage table for *Methylomonas sp.* 16a. To further optimize the genes, most strong hairpin structures were disrupted by replacing with alternative sub-optimal codons. The AT-rich mRNA instability region (Guhaniyogi, G. and J. Brewer, *Gene*, 265(1-2):11-23 (2001)) and the long runs of the same nucleotide were also eliminated. In the case of a string of more than 3 or 4 of the same amino acids, a sub-optimal codon was also introduced to prevent shortage of the most preferred codon pool for this amino acid. The ribosomal binding site (RBS) was engineered upstream of the start codon as the RBS sequence from pTrcHis2-TOPO vector (Invitrogen, Carlsbad, CA). Several
restriction sites were also engineered at the 5' and 3' ends of the genes to facilitate cloning.

A comparison between the synthetic crtW gene (SEQ ID NO:7) and the synthetic crtZ gene (SEQ ID NO:8) with the full-length wildtype crtW and crtZ sequences from A. aurantiacum is shown in Figure 6 and Figure 7, respectively, wherein nucleotides in shaded boxes correspond to nucleotides that are identical between the native gene and the synthetic gene. For both crtW and crtZ, there is 84% nucleotide identity between the native gene and the synthetic gene. For the codon-optimized crtW gene, 99 codons were optimized, while for the codon-optimized crtZ gene 73 codons were optimized. None of the modifications in the codon-optimized gene changed the amino acid sequence of the encoded ketolase protein (SEQ ID NO:33) or hydroxylase protein (SEQ ID NO:34). The synthetic crtW gene was synthesized and cloned onto the pCRScript vector by Aptagen Inc. (Herndon, VA) to form pCRScript-Dup1. The synthetic crtZ gene was synthesized and cloned onto the pUC18 vector by GenScript Corp. (Scotch Plains, NJ) to form pUC-SynCrtZ.

EXAMPLE 9

CONSTRUCTION OF CANTHAXANTHIN EXPRESSION PLASMID

pDCQ307

The purpose of this Example was to prepare a canthaxanthin expression plasmid, referred to herein as pDCQ307. This canthaxanthin-producing plasmid was prepared by coupling the codon-optimized crtW gene (SEQ ID NO:7), described in Example 8, to a β-carotene synthesis gene cluster. The crtW gene was cloned downstream of crtE in the reorganized crtEYIB cluster from P. stewartii ATCC8199 to form the operon crtEWYIB (see Figure 8).

Pantoea stewartii ATCC #8199 (WO 03/016503) contains the natural gene cluster crtEXYIYZ. The genes required for β-carotene synthesis (i.e., crtE and crtYIB) were joined together by PCR. Specifically, the crtE gene (SEQ ID NO:35) and crtYIB genes (SEQ ID NO:36) were each amplified using chromosomal DNA as template and the primers given in Table 6.
Table 6
Primers Used for Creation of the *crtEYIB* Reporter Construct

<table>
<thead>
<tr>
<th>Gene(s)</th>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>crtE</em></td>
<td>pBHRcrt_1F: 5'-GAATTCCCGCTTACGGTCT-3' (SEQ ID NO:37)</td>
<td>pBHRcrt_1R: 5'-CGTTGATACTGGCCCACTTTTTTT-3' (SEQ ID NO:38)</td>
</tr>
<tr>
<td><em>crtYIB</em></td>
<td>pBHRcrt_2F: 5'-AAAACATCGGCTGCACTGAC-3' (SEQ ID NO:39)</td>
<td>pBHRcrt_2R: 5'-GGTACCTAGATCGGCGCCTGAC-3' (SEQ ID NO:40)</td>
</tr>
</tbody>
</table>

* Note: Underlined portions within each primer correspond to restriction sites for *EcoR*1, *Mfe*1.

The PCR reactions were performed with *Pfu* DNA polymerase in buffer supplied by the manufacturer containing dNTPs (200 μM of each). Parameters for the thermocycling reactions were: 92°C (5 min), followed by 30 cycles of: 95°C (30 sec), 55°C (30 sec), and 72°C (5 min). The reaction concluded with 1 cycle at 72°C for 10 min. The two PCR products were gel purified and joined together by a subsequent PCR reaction using the primers pBHRcrt_1F (SEQ ID NO:37) and pBHRcrt_2R (SEQ ID NO:40). Parameters for the thermocycling reaction were: 95°C (5 min), followed by 20 cycles of: 95°C (30 sec), 55°C (1 min) and 72°C (8 min). A final elongation step at 72°C for 10 min completed the reaction. The final 4511 bp PCR product was cloned into the pTrcHis2-Toho vector (Invitrogen, Carlsbad, CA) in the forward orientation, resulting in plasmid pDCQ300. The ~4.5 kb *EcoR*1 fragment of pDCQ300 containing the *crtEYIB* gene cluster was ligated into the unique *EcoR*1 site of vector pBHR1 (MobTec GmbH, Goettingen, Germany), to create construct pDCQ301. In pDCQ301, a unique *Mfe*1 site was engineered in the intergenic region of *crtE* and *crtY*.

The ~0.8 kb *EcoR*1 fragment of pCRScript-Dup1, prepared as described in Example 8, containing the synthetic codon-optimized *crtW* gene was ligated to the unique *Mfe*1 site in pDCQ301. In the resulting
construct pDCQ307, the _crtEWYIB_ genes were under the control of the chloramphenicol resistant gene promoter of the vector.

**EXAMPLE 10**

**EVALUATION OF CANTHAXANTHIN PRODUCTION IN FOUR METHYLOMONAS SP. "WHITE" MUTANTS**

The production of canthaxanthin was evaluated in all four of the _Methylomonas sp._ "white" mutant backgrounds, i.e., _Methylomonas_ strains MWM1200 (_Δcrt_ cluster promoter + _ΔcrtN3_), MWM1300 (_Δald(crtN1) + _ΔcrtN3_), MWM1100 (_Δcrt_ cluster promoter), and MWM1000 (_Δald(crtN1)_).

The canthaxanthin expression plasmid pDCQ307, prepared as described in Example 9, was transferred into each "white" mutant via conjugation using the same method described in Example 4. The resulting strains were designated as MCS001 (MWM1000 + pDCQ307), MCS002 (MWM1100 + pDCQ307), MCS003 (MWM1200 + pDCQ307), and MCS004 (MWM1300 + pDCQ307).

After purifying the strains away from the contaminating _E. coli_ cells that were introduced during the mating, each strain was evaluated for the effect of canthaxanthin production on growth rate (data not shown). In addition, the canthaxanthin titers were determined for each strain, as described below.

**THF/Methanol Extraction of C₄₀ Carotenoid**

As described in Example 6 for the extraction of the C₃₀ carotenoid, a THF/methanol (1:1) solution was also employed for the extraction of the C₄₀ carotenoid from the _Methylomonas sp._ carotenoid-producing strains.

The procedures were virtually identical to the C₃₀ carotenoid extraction method with the exception of the culture volumes used. Two 500-mL bottles containing ~60 mL of BTZ and 25% methane were grown until saturation (~24 hr) for each of the _Methylomonas_ canthaxanthin-producing cultures (MCS001, MCS002, MCS003, and MCS004) to be analyzed. The cultures used in the growth rate experiments were used in this study. The cells were split into two fractions harvested via centrifugation. For one fraction, the cell pellets were dried at ~100°C in an oven for ~24 hr and the dry cell mass was determined. The other fraction was used for C₄₀ carotenoid extractions. The cells were lysed using glass beads (0.5 mL/sample). Also added to each sample was 150 μL of the internal standard, ethyl-β-apo-8'-carotene (trans) (100 mg/L stock solution) and 5 mL of a THF/methanol (1:1) solution. This mixture was vortexed for ~2 min, followed by a 15 min centrifugation at 8,000 rpm. The supernatant
was collected and the sample was vortexed again for ~2 min and centrifuged at 8,000 rpm for another 15 min. The supernatants were combined and were dried using nitrogen. The carotenoid samples were stored at -80°C until analyzed using HPLC-photodiode array.

5  **HPLC Diode Array Analysis of C_{40} Carotenoid**

The carotenoid samples were analyzed using HPLC diode array analysis, as described in Example 6. The measured canthaxanthin titers are given in Table 7. The highest titer was seen for MCS003 (~1400 ppm). This reflected an ~80% increase in titer over its parent strain MCS002 (~750 ppm). The only difference between these two strains is the deletion of *crtN3* in MCS003. Thus, the deletion of *crtN3* resulted in an increase in the production of the C_{40} carotenoid that was greater than the increase that was observed in the production of the C_{30} carotenoid, i.e., ~80% increase versus ~15% increase, respectively. However, the effect of the *crtN3* deletion was not as pronounced in the MCS004 strain. The canthaxanthin titer of MCS004 (~1100 ppm) was only slightly higher (~11%) than its parent strain MCS001 (~1000 ppm). This result suggests that there are other contributing factors that influence the much higher titer that is seen in MCS003. One hypothesis is that it is the combination of the *crtN2* deletion and the *crtN3* deletion may have an additive effect on canthaxanthin production. Note that *crtN2* is still present in MCS001 and MCS004.

**Table 7**

**Canthaxanthin Titers for *Methylomonas* Mutants**

<table>
<thead>
<tr>
<th>Strain</th>
<th>C_{30} Carotenoid Genes Expressed</th>
<th>C_{30} Carotenoid Genes Not Expressed</th>
<th>Canthaxanthin Titer (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCS001</td>
<td><em>crtN2, crtN3</em></td>
<td><em>crtN1, ald</em></td>
<td>~1000</td>
</tr>
<tr>
<td>MCS002</td>
<td><em>ald, crtN2, crtN3</em></td>
<td><em>crtN1</em></td>
<td>~750</td>
</tr>
<tr>
<td>MCS003</td>
<td><em>ald, crtN2</em></td>
<td><em>crtN1, crtN3</em></td>
<td>~1400</td>
</tr>
<tr>
<td>MCS004</td>
<td><em>crtN2</em></td>
<td><em>crtN1, crtN3, ald</em></td>
<td>~1100</td>
</tr>
</tbody>
</table>
EXAMPLE 11
CONSTRUCTION OF A CANTHAXANTHIN-PRODUCING
METHYLOMONAS SP. "WHITE" MUTANT

This Example describes the construction of another canthaxanthin-produc-
ing Methylomonas sp. 16a white mutant. The plasmid pDCQ333
was constructed by coupling the codon-optimized crtW gene to the β-
carotene synthesis gene cluster crtEidiYIB from Pantoea agglomerans
DC404 (US 10/808807). The crtW gene was cloned upstream of the crtE
in the native cluster to form the operon crtWEidiYIB (see Figure 8). This
operon was introduced into Methylomonas sp. 16a MWM1200 using tri-
parental conjugation and the production of canthaxanthin by this
transformed strain was analyzed using HPLC.

Codon-Optimized Ketolase Gene Coupled to β-Carotene Synthesis Genes
from P. agglomerans DC404

P. agglomerans DC404 was an environmental isolate that
contained the carotenoid synthesis gene cluster crtEidiYIBZ (SEQ ID
NO:41), as described by Cheng et al. in copending US 10/808807.
The soil from a residential vegetable garden in Wilmington, Delaware was
collected and resuspended in LB medium. A 10 µL loopful of
resuspension was streaked onto LB plates and the plates were incubated
at 30°C. Pigmented bacteria with diverse colony appearances were
picked and streaked twice to homogeneity on LB plates and incubated at
30°C. From these colonies, one which formed pale yellow smooth
translucent colonies was designated as "strain DC404".

P. agglomerans strain DC404 was grown in 25 mL of LB medium at
30°C overnight with aeration. Bacterial cells were centrifuged at 4,000 x g
for 10 min. The cell pellet was gently resuspended in 5 mL of 50 mM
Tris-10 mM EDTA (pH 8.0) and lysozyme was added to a final
concentration of 2 mg/mL. The suspension was incubated at 37°C for
1 hr. Sodium dodecyl sulfate was then added to a final concentration of
1% and proteinase K was added at 100 µg/mL. The suspension was
incubated at 55°C for 2 hr. The suspension became clear and the clear
lysate was extracted twice with an equal volume of
phenol:chloroform:isoamyl alcohol (25:24:1) and once with
chloroform:isoamyl alcohol (24:1). After centrifuging at 4,000 rpm for
20 min, the aqueous phase was carefully removed and transferred to a
new tube. Two volumes of ethanol were added and the DNA was gently
spooled with a sealed glass Pasteur pipette. The DNA was dipped into a
tube containing 70% ethanol. After air drying, the DNA was resuspended in 400 μL of TE (10 mM Tris-1 mM EDTA, pH 8.0) with RNaseA (100 μg/mL) and stored at 4°C. The concentration and purity of DNA was determined spectrophotometrically by OD\textsubscript{260}/OD\textsubscript{280}.

A cosmid library of DC404 was constructed using the pWEB cosmid cloning kit from Epicentre (Madison, WI) following the manufacturer’s instructions. Genomic DNA was sheared by passing it through a syringe needle. The sheared DNA was end-repaired and size-selected on low-melting-point agarose by comparison with a 40 kb standard. DNA fragments approximately 40 kb in size were purified and ligated into the blunt-ended cloning-ready pWEB cosmid vector. The library was packaged using ultra-high efficiency MaxPlax Lambda Packaging Extracts, and plated on EPN100 E.coli cells. Two yellow colonies were identified from the cosmid library clones. The cosmid DNA from the two clones had similar restriction digestion patterns. This cosmid DNA, referred to herein as pWEB-404, contained the \textit{crtEidiYIBZ} gene cluster, given as SEQ ID NO:41.

Primers pWEB404F: 5’-GAATTCACTAGTCGAGACGCGGATACCAACCAT-3’ (SEQ ID NO:42) and pWEB404R: 5’-GAATTCTAGCAGCGCCGCTGCCAGA-3’ (SEQ ID NO:43) were used to amplify a fragment from DC404 containing the \textit{crtEidiYIB} genes (SEQ ID NO:6) by PCR. Cosmid DNA pWEB-404 was used as the template with PfuTurbo™ polymerase (Stratagene, La Jolla, CA), and the following thermocycler conditions: 92°C (5 min); 94°C (1 min), 60°C (1 min), 72°C (9 min) for 25 cycles; and 72°C (10 min). A single product of approximately 5.6 kb was observed following gel electrophoresis. Taq polymerase (Roche Applied Science, Indianapolis, IN) was used in a ten minute 72°C reaction to add additional 3’ adenosine nucleotides to the fragment for TOPO\textsuperscript{®} cloning into pTrcHis2-TOPO (Invitrogen, Carlsbad, CA). Following transformation to \textit{E. coli} TOP10 cells, several colonies appeared bright yellow in color, indicating that they were producing a carotenoid compound. The gene cluster was then subcloned into the broad host range vector pBHR1 (MoBiTec, LLC, Marco Island, FL), and electroporated into \textit{E. coli} 10G cells (Lucigen, Middletown, WI). The transformants containing the resulting plasmid pDCQ330 were selected on LB medium containing 50 μg/mL kanamycin. In pDCQ330, a unique \textit{SpeI} site was engineered upstream of \textit{crtE}.
The ~0.8 kb *EcoRI* fragment of pCRScript-Dup1, prepared as described in Example 8, containing the synthetic, codon-optimized *crtW* gene was first blunt-ended and then ligated to pDCQ330, which was digested by *SpeI* and blunt-ended. In the resulting construct pDCQ333, the *crtW* gene was cloned upstream of the *crtE*/*diYIB* cluster and the *crtW*/*E*/*di*/*YIB* genes were under the control of the chloramphenicol resistant gene promoter of the vector.

Tri-Parental Conjugation of the Plasmid pDCQ333 into *Methylomonas sp.* 16a MWM1200

Plasmid pDCQ333 was transferred into *Methylomonas* 16a MWM1200 by tri-parental conjugal mating, which is described in Example 4. The *E. coli* helper strain containing pRK2013 and the *E. coli* 10G donor strain containing pDCQ333 were grown overnight in LB medium containing kanamycin (50 µg/mL), washed three times in LB, and resuspended in a volume of LB representing approximately a 60-fold concentration of the original culture volume.

The *Methylomonas* 16a MWM1200 strain, described in Example 7, was grown as the recipient using the general conditions described in Example 4. Briefly, *Methylomonas* 16a MWM1200 strain was grown in serum-stoppered Wheaton bottles (Wheaton Scientific, Wheaton IL) using a gas/liquid ratio of at least 8:1 (i.e., 20 mL of liquid BTZ (nitrate) medium in 160 mL total volume) at 30°C with constant shaking. The standard gas phase for cultivation contained 25% methane in air. The recipient cells were cultured under these conditions for 48 h in BTZ (nitrate) medium, washed three times in BTZ (nitrate), and resuspended in a volume of BTZ (nitrate) representing a 150-fold concentration of the original culture volume.

The donor, helper, and recipient cell pastes were then combined in ratios of 1:1:2, respectively, on the surface of BTZ (nitrate) agar plates containing 0.5% (w/v) yeast extract. Plates were maintained at 30°C in 25% methane for 16-72 hr to allow conjugation to occur, after which the cell pastes were collected and resuspended in BTZ (nitrate). Dilutions were plated on BTZ (nitrate) agar containing kanamycin (50 µg/mL) and incubated at 30°C in 25% methane for up to 1 week. Orange-red transconjugants were streaked onto BTZ (nitrate) agar containing kanamycin (50 µg/mL).
Analysis of Canthaxanthin Production

For analysis of carotenoid composition, transconjugants were cultured in 25 mL of BTZ (nitrate) medium containing kanamycin (50 μg/mL) and incubated at 30°C in 25% methane as the sole carbon source for 3-4 days. The cells were harvested by centrifugation and frozen at -20°C. After thawing, the pellets were extracted with 10 mL of acetone, dried under nitrogen and redissolved in 1-2 mL of acetone. The extract was filtered using an Acrodisc CR25 mm syringe filter (Pall Corporation, Ann Arbor, MI). The extract was then concentrated in 0.1 mL of 10% acetone/90% acetonitrile for HPLC analysis using an Agilent Series 1100 LC/MSD SL (Agilent, Foster City, CA).

A cell extract sample (20 μL) was loaded onto a 150 mm X 4.6 mm ZORBAX C18 (3.5 μm particles) column (Agilent Technologies, Inc.). The column temperature was kept at 40°C. The flow rate was 1 mL/min, while the solvent running program used was:
- 0 - 2 min: 95% Buffer A and 5% Buffer B;
- 2 - 10 min: linear gradient from 95% Buffer A and 5% Buffer B to 60% Buffer A and 40% Buffer B;
- 10 - 12 min: linear gradient from 60% Buffer A and 40% Buffer B to 50% Buffer A and 50% Buffer B;
- 12 - 18 min: 50% Buffer A and 50% Buffer B; and,
- 18 - 20 min: 95% Buffer A and 5% Buffer B.
Buffer A was 95% acetonitrile and 5% dH₂O; Buffer B was 100% tetrahydrofuran.

The resulting chromatogram showed that in Methyloclas MWM1200 containing pDCQ333, the major carotenoid produced was canthaxanthin, which eluted at 8.2 min. The echinenone intermediate was also present at a low level.

EXAMPLE 12

CONSTRUCTION OF ASTAXANTHIN-PRODUCING METHYLOMONAS SP. "WHITE" MUTANTS

This Example describes the construction of two astaxanthin-producing Methyloclas sp. 16a white mutants. Two plasmids were constructed by coupling the codon-optimized ctw and ctwZ genes, described in Example 8, to two different β-carotene synthesis gene clusters. In the first construct pDCQ324, the ctwZ genes were cloned downstream of ctwE in the reorganized ctwEYIB cluster from P. stewartii ATCC8199 to form the operon ctwEWZYIB (see Figure 8). In the second
construct pDCQ334, the crtWZ genes were cloned upstream of crtE in the
native crtEidiYIB cluster from P. agglomerans DC404 to form the operon
crtWZEidiYIB (see Figure 8). These operons were introduced into
Methylomonas sp. 16a MWM1200 using tri-parental conjugation and the
production of C40 carotenoids by the two resulting transformed strains
were analyzed using HPLC.
Codon-Optimized Ketolase and Hydroxylase Genes Coupled to β-
Carotene Synthesis Genes from P. stewartii ATCC8199

In order to create convenient restriction sites for incorporating the
codon-optimized hydroxylase gene crtZ (SEQ ID NO:8) into the cluster,
unique XbaI and BglII sites were introduced downstream of the crtW gene.
The codon-optimized crtW gene was PCR amplified using pDCQ307 DNA
as template with primers crtW_F_Agro(Syn) 5’-
GCCAATTGAAGGAGGAATAAACCATG −3’ (SEQ ID NO:44) and
crtW_R_Agro(Syn) 5’-
GCGAATTCAGATCTTGCTTAGATCAGCGGCTGTCGCTTTTG-3 (SEQ
ID NO:45). The underlined portions of these primer sequences are
restriction sites for EcoRI, MfeI, BglII, or XbaI. The ~0.8 kB PCR product
containing the crtW gene was digested with MfeI/EcoRI and ligated into
the unique MfeI site in pDCQ301. The ~0.5 kB XbaI/BglII fragment of
pUC-synCrtZ, described in Example 8, containing the codon-optimized
crtZ gene was then cloned to the XbaI/BglII sites downstream of crtW. In
the resulting construct pDCQ324, the crtEWZYIB genes were organized in
an operon and were under the control of the chloramphenicol resistant
gene promoter of the vector.
Codon-Optimized Ketolase and Hydroxylase Genes Coupled to β-
Carotene Synthesis Genes from P. agglomerans DC404

The codon-optimized crtWZ genes were PCR amplified using
pDCQ324 DNA as template, with primers crtWZ_F(syn agro): 5’-
ACTAGTAAGGAGGATAAACCATGAGCGCC-3’ (SEQ ID NO:46) and
crtWZ_R(syn agro): 5’-
GCTAGCTGTACATTAGGTGCTTCTTGCGCTTC-3’ (SEQ ID NO:47).
The underlined portions of these primer sequences were engineered
restriction enzyme sites for SpeI and Nhel. The ~1.3 kb PCR product was
gel-purified and cloned into the pTrcHis2-topo vector. The ~1.3 kb SpeI/
Nhel fragment containing crtWZ genes was then subcloned into the SpeI
site of pDCQ330. In the resulting plasmid pDCQ334, the crtWZEidiYIB
genes were organized in an operon and were under the control of the chloramphenicol resistant gene promoter of the vector.

Tri-Parental Conjugation of the Plasmids pDCQ324 or pDCQ334 into Methylomonas sp. 16a MWM1200

Plasmid pDCQ324 or pDCQ334 was transferred into Methylomonas 16a MWM1200 using tri-parental conjugation, as described in Example 11. The transconjugants were grown and analyzed as described in Example 11. The HPLC results showed that in Methylomonas MWM1200 containing pDCQ324 or pDCQ334, the major carotenoid produced was astaxanthin, which eluted at about 5.5 min. The intermediates such as adonixanthin, adonirubin and canthaxanthin were also present. Identification of astaxanthin and canthaxanthin were based on comparison of elution time, absorption spectrum and the molecular weight with those of the authentic standards obtained from CaroteNature (Lupsingen, Switzerland). Identification of adonixanthin and adonirubin were based on the comparison of the absorption spectrum and the molecular weight with those reported in the literature.
CLAIMS

What is claimed is:

1. A high growth methanotrophic bacterial strain which:
   a) grows on a C1 carbon substrate selected from the group consisting of methanol and methane;
   b) comprises a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate dependent phosphofructokinase enzyme; and
   c) possesses a disruption such that at least one endogenous gene selected from the group consisting of \textit{crtN1}, \textit{ald}, \textit{crtN2}, and \textit{crtN3} is down-regulated.


3. The bacterial strain of Claim 1 comprising a 16s rRNA sequence as set forth in SEQ ID NO:1.

4. The bacterial strain of Claim 3 wherein the strain is a \textit{Methylomonas} \textit{sp}.

5. The \textit{Methylomonas} \textit{sp}. strain of Claim 4 further comprising:
   (a) a \textit{crtE}-\textit{idi}-\textit{crtY}-\textit{crtI}-\textit{crtB} gene cluster having the sequence as set forth in SEQ ID NO:6; and
   (b) a codon-optimized \(\beta\)-carotene ketolase gene having the sequence as set forth in SEQ ID NO:7.

6. The \textit{Methylomonas} \textit{sp}. strain of Claim 4 further comprising:
   (a) a \textit{crtE}-\textit{idi}-\textit{crtY}-\textit{crtI}-\textit{crtB} gene cluster having the sequence as set forth in SEQ ID NO:6;
   (b) a codon-optimized \(\beta\)-carotene ketolase gene having the sequence as set forth in SEQ ID NO:7; and
   (c) a codon-optimized \(\beta\)-carotene hydroxylase gene having the sequence as set forth in SEQ ID NO:8.

7. A \textit{Methylomonas} \textit{sp}. strain optimized for the production of C40 carotenoids selected from the group consisting of: MWM1000 (\textit{\Delta}\textit{ald(crtN1)}), MWM1100 (\textit{\Delta}\textit{crt} cluster promoter), MWM1400 (\textit{\Delta}\textit{crtN1}), MWM1600 (\textit{\Delta}\textit{crt} cluster), MWM1200 (\textit{\Delta}\textit{crt} cluster promoter + \textit{\Delta}\textit{crtN3}), MWM1300 (\textit{\Delta}\textit{ald(crtN1)} + \textit{\Delta}\textit{crtN3}), MWM1800 (\textit{\Delta}\textit{crtN1} + \textit{\Delta}\textit{crtN3}), and MWM1900 (\textit{\Delta}\textit{crt} cluster + \textit{\Delta}\textit{crtN3}).
8. The *Methylocomas* sp. strain of any of Claims 4-7 wherein the down regulated genes have the nucleic acid sequences (SEQ ID NO:2), (SEQ ID NO:3), (SEQ ID NO:4), (SEQ ID NO:5) corresponding to the genes *crtN1*, *ald*, *crtN2*, and *crtN3* respectively.

9. A method for the production of C_{40} carotenoids, comprising:
   a) providing the high growth methanotrophic bacterial strain of Claim 1 comprising at least one gene encoding an enzyme of the C_{40} carotenoid biosynthesis pathway;
   b) growing the bacterial strain of step (a) under conditions wherein the gene encoding an enzyme of the C_{40} carotenoid biosynthesis pathway is expressed producing a C_{40} carotenoid; and
   c) optionally recovering the C_{40} carotenoid.

10. The method of Claim 9, wherein the at least one gene encoding an enzyme of the C_{40} carotenoid biosynthesis pathway is selected from the group consisting of: *crtE*, *crtX*, *crtY*, *crtl*, *crtB*, *crtR*, *crtZ*, *crtW*, *crtO*, *crtA*, *crtC*, *crtD*, *crtF*, and *crtU*.

11. The method according to Claim 10, wherein the C_{40} carotenoid is selected from the group consisting of: antheraxanthin, adonirubin, adonixanthin, astaxanthin, canthaxanthin, capsorubin, β-cryptoxanthin, α-carotene, β-carotene, epsilon-carotene, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone, γ-carotene, 4-keto-γ-carotene, ζ-carotene, α-cryptoxanthin, deoxyflexigraphan, diatoxanthin, 7,8-didehydroastaxanthin, fucoxanthin, fucoxanthinol, isorenieratene, lactucalexanthin, lutein, lycopene, myxobactone, neoaxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, 4-keto-rubixanthin, siphoneanxanthin, spheroidene, spheroidenone, spirilloxanthin, 4-ketotorulene, 3-hydroxy-4-keto-torulene, urilide, urilide acetate, violaxanthin, zeaxanthin-β-diglucoside, and zeaxanthin.

12. A method for the production of canthaxanthin comprising:
   a) providing a high growth methanotrophic bacterial strain comprising:
      i) a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate dependent phosphofructokinase enzyme; and
      ii) a disruption in at least one endogenous gene selected from the group consisting of *crtN1*, *ald*, *crtN2*, and *crtN3*;
iii) a *crtE-idi-crtY-crtl-crtB* gene cluster having the sequence as set forth in SEQ ID NO:6; and

iv) a codon-optimized β-carotene ketolase gene having the sequence as set forth in SEQ ID NO:7;

wherein said high growth methanotrophic bacterial strain grows on a C1 carbon substrate selected from the group consisting of methanol and methane;

(b) growing the high growth methanotrophic bacterial strain of step (a) under conditions wherein the *crtE-idi-crtY-crtl-crtB* gene cluster genes and the codon-optimized β-carotene ketolase gene are expressed producing canthaxanthin; and

(c) optionally recovering the canthaxanthin.

13. A method for the production of astaxanthin comprising:

(a) providing a high growth methanotrophic bacterial strain comprising:

i) a functional Embden-Meyerhof carbon pathway, said pathway comprising a gene encoding a pyrophosphate dependent phosphofructokinase enzyme; and

ii) a disruption in at least one endogenous gene selected from the group consisting of *crtN1, ald, crtN2*, and *crtN3*;

iii) a *crtE-idi-crtY-crtl-crtB* gene cluster having the sequence as set forth in SEQ ID NO:6;

iv) a codon-optimized β-carotene ketolase gene having the sequence as set forth in SEQ ID NO:7; and

(v) a codon-optimized β-carotene hydroxylase gene having the sequence as set forth in SEQ ID NO:8;

wherein said high growth methanotrophic bacterial strain grows on a C1 carbon substrate selected from the group consisting of methanol and methane;

(b) growing the high growth methanotrophic bacterial strain of step (a) under conditions wherein the *crtE-idi-crtY-crtl-crtB* gene cluster genes, the codon-optimized β-carotene ketolase gene, and the codon-optimized β-carotene hydroxylase gene are expressed producing astaxanthin; and

(c) optionally recovering the astaxanthin.
14. An isolated nucleic acid molecule selected from the group consisting of:
   (a) an isolated nucleic acid molecule as set forth in SEQ ID NO:7 which encodes a β-carotene ketolase enzyme; and
   (b) an isolated nucleic acid molecule that is completely complementary to (a).

15. An isolated nucleic acid molecule which encodes a β-carotene ketolase enzyme as set forth in SEQ ID NO:33 wherein at least 99 codons are codon-optimized for expression in *Methylophaga sp.*

16. An isolated nucleic acid molecule selected from the group consisting of:
   (a) an isolated nucleic acid molecule as set forth in SEQ ID NO:8 which encodes a β-carotene hydroxylase enzyme; and
   (b) an isolated nucleic acid molecule that is completely complementary to (a).

17. An isolated nucleic acid molecule which encodes a β-carotene hydroxylase enzyme as set forth in SEQ ID NO:34 wherein at least 73 codons are codon-optimized for expression in *Methylophaga sp.*
Figure 1

Upper Isoprenoid Pathway

C₄₀ Synthesis

Lower Carotenoid Biosynthetic Pathway

Lycopene,
β-carotene,
Zeaxanthin,
Zeaxanthin-β–glucosides,
Astaxanthin, etc.
Figure 2

Dehydrogenases (short-chain alcohol dehydro)

3-oxoacyl (ACP) synthase
dehydrogenase

Acyl-CoA synthetase
crtN3 (FAD-dependent oxidoreductase)

SAM-dependent methyl transferase

hypothesized protein

3-oxoacyl (ACP) synthase

10,050 bp
Figure 3

- CrtN1
  - Phytoene dehydrogenase
    - (Oceanabaccillus)
  - Orf 3
    - (Staphylococcus aureus)

- CrtN2
  - Phytoene dehydrogenase
    - (Listeria monocytogenes)
  - Phytoene dehydrogenase
    - (Listeria innocua)
  - Hypothetical protein
    - (Prochlorococcus marinus)
  - Unknown Conserved Protein
    - (Bacillus halodurans)

- CrtN3
  - Hypothetical protein
    - (Nitrosomonas europaea)
  - Farnesyldiphosphate farnesyltransferase
    - (Methylococcus capsulatus)
  - Sqs
Figure 7
CL2230 PCT Seq Lst.
SEQUENCE LISTING

<110>  E. I. du Pont de Nemours & Company
<120>  Optimized Bacterial Host Strains of Methylomonas sp. 16A
<130>  CL-2230
<150>  US 60/527,083
<151>  2003-12-03
<150>  US 60/527,877
<151>  2003-12-08
<160>  53
<170>  PatentIn version 3.3

<210>  1
<211>  1429
<212>  DNA
<213>  Methylomonas sp. 16A

<400> 1
cggtagctt aacacatgca agtcgaacgc tgaaaggtctc ttgcacctgg atgagtggcg 60
  gacggtgtag taatgcatag gaaatgcttc attaagtttg ggataaactgg ggaacttcac 120
gtaatcgcg ctaacgctct acacagggac gccggggacc ttgcggcctg gcgcctttag 180
atgagcctct gtccgattag ctaggttggtg ggttaaaggg ctacccaagcg cagcagctcg 240
agcctggtctg aagagagttg cagccacact gggactgaga cagcgcccgag atctctacgg 300
gagcacgtag tgagccatgag gcgcgaagct gatcagacact taccgctgtg 360
gtgccacag cctgaagggatt ttaaatgcatt ccaattgtgacag aagggactac atcgtttag 420
accgcgtgag cttcacattt gacggttgctc ccccagttct ggaacttcctg ggaacttcctg 480
ctgtaatacgg agggtgcaag ctgttaatcgg aattactggg cgtgaagcgct gcgtgagggc 540
ttttttaagc cagatggtaac agccgctgtgg ttaaaccttgag aatcgctattt gatgcattgg 600
aactagagg ttagttaggg gagttgagaatt ctcaatttga cggggaaagc cgccttgagc 660
tgaaggaaaca ccagtgccga aggcggttct cttggactcag actgacgcctg aggtacgaaa 720
gcgtggttag cgaagcggtg tatttcacct ggttgtcact gcgcgtaaacg gcgcctttctta 780
accgcccttg gctggaggttt gctatattcag ctcgccctgc cggagctcagc gacgctgg 840
agttcctgta ctaatttgag aagagctcag ttccttcctag gcccttgcgg gcccttgcgg 900
atgtggttaa atctgcagtaa cagccgctagt cgcggctctct cctgatcaca cctggacggttg 960
gtcagagatct cccttggtttg gtctggagtg gctggtgtaat ccctctttct cttatggcctt 1020
ggtcctctgtt gtagttgttac gtgccttac ccctccccag ccccttttta cccctctttcag 1080
cagcgcgtta cggctgactt gcttctgagtg cgcctctggta aacccgcttc ggggggggg 1140
agctgtgctc gtcattgag cgcggctttg gccggcgctg tctgcctggct aatggccttg 1200
acaggggggtt gggccagttgcc ggcacagcag cttgccctga atcagagatc gacgagccgc 1260
tgcgtctttc aacctcgactt ccgagaagcg cagagcgcagc cactgctatc aatggtctgtg 1320
WO 2005/072086

CL2230 PCT Seq Lst.

cgcgggtgaat aegttccccgg gcccttgatc cacgccgccc gt cacaacatgg gagtggggttg 1380
caaaaagaagt aggtgtttta accttccccgga ggccgttac caaccttggg 1429

<210> 2
<211> 1536
<212> DNA
<213> Methylomonas sp. 16A

<400> 2
atggccaaaca ccaacacat catcatcgtc ggccggggtc ccgggctgacg ttgcgccggc 60
atgttgctgca gcagcggccgt cccttcaagta tcgatctctcg acacatcagtc cagaaatacgcc 120
ggcgcgcaacc gcggctacaagcatgctggcgc gacatctggc tcttcgaccgcagcagcagc 180
ttggtgaggag gcggctggtggc gagatgcttgac agcgcggagt gaggattatatctgcagcagc 240
tcggatccacagcgatcagcgc gacgtgttcgta cgcgtgtctgcagcgtatcgc 300
gcgtctcatcgg aacgctcggg caaagcctctc gcgcgacgctg gccgggctgacg 360
agcgtgctacag gc gagatccgata ccggccggtg gggaaatctcg cggcggggtg 420
atccggcgtgc caccctccag gccggcatggtg gccggctggtggc gagatgcttgacagcagcagc 480
cggccggggcga ttcgagtttgg gccgggctgacg aagatctcgc ggggaggatctcgcagcagc 540
aaatctgcttc gcccctctccg ctcacatcgtcagcgtatcgc ggggaggatctcgcagcagc 600
cggccgtgtc gcgcgacgctg gccgggctgacg aagatctcgc ggggaggatctcgcagcagc 660
gcgcgcctcgg ctcgacgctg gccgggctgacg aagatctcgc ggggaggatctcgcagcagc 720
attcagtttg ccgcgccgctg gccgggctgacg aagatctcgc ggggaggatctcgcagcagc 780
aaatctgcttc gcccctctccg ctcacatcgtcagcgtatcgc ggggaggatctcgcagcagc 840
cgcggcgtgc ctcgacgctg gccgggctgacg aagatctcgc ggggaggatctcgcagcagc 900
aaatctgcttc gcccctctccg ctcacatcgtcagcgtatcgc ggggaggatctcgcagcagc 960
attcagtttg ccgcgccgctg gccgggctgacg aagatctcgc ggggaggatctcgcagcagc 1020
ttcgacacacaa aacccctgag cggccggtc agcttggtttccc gcggaaacgg gcacgctcggcagc 1080
gacgacacgc tagccggcagc gcgccaaacgc cggctgtctcg cggctggttgcc gtcgggccacac 1140
aacgcacgac gccggctgtctcg cggccggtc acgatcgcgactgcggccgctg gccgggctgacg 1200
agcgtggggc gcggctcctgg cggccggtc acgatcgcgactgcggccgctg gccgggctgacg 1260
atccgcgctc gcgccaaacgc cggccggtc acgatcgcgactgcggccgctg gccgggctgacg 1320
tgcgtgctctc gcgccaaacgc cggccggtc acgatcgcgactgcggccgctg gccgggctgacg 1380
aatctgctcgc gggcgcctgtgc gcgcggcgcc gccgggctgacg gtcgggccacac 1440
tgcgtgctctc gcgccaaacgc cggccggtc acgatcgcgactgcggccgctg gccgggctgacg 1500
gcgcgcagcg cggccggtc acgatcgcgactgcggccgctg gccgggctgacg 1536

<210> 3
<211> 1593
<212> DNA
<213> Methylomonas sp. 16A

Page 2
CL2230 PCT Seq Lst.

<400>  3
atgacagaca tagcagccgct ctccccactgt gatggccgct tgctgggaca ttttccagtc 60
agcaagcgcgg cgctcattca gcaacagctg acaaaatccc gccgcgccgc cctgcttttg 120
cgcagctgccc gcggctacgc aagcggtctata cgcctgtgcg ccttgaaaca acagctgctg 180
gataacccgg agacacttcg cgaaacaccg ccgctctcagc ccgcaaggtg tccgaccagg 240
gctttgtcgg ggggaatta ttcgccgtcg gatttaactgg gtatattacca aagccgaggcc 300
cgcgggattc tacgcacgcg ccgcgtgtcgc acctgctgcct tcgcgttttc aggccgaccgc 360
gccgcagctcg aacggcccct tttggcttctg gtgcgccgtgct tctgccatag gaaactcccctg 420
tttcacgtaga ggctgcggccct cgcgtctgcgct cttctgctgttg cgcacagcgg 480
aaacctccgg aacticgtcct gcggctgtctg cgtgttcttc ggcgagctctg 540
agtttgccgg aagccgtttgc gacatggtgct cctggattcgg ggcgctacgc 600
attagcgcgcc gcgcgcagcccc ggtgtttttcg accgccgcggc cgcgcggcggctt cgcgcggccc 660
attgcaacggc ccccgccgca caggattcgcg cgcacgctgctc agttgagcggc ttacgccacc 720
attgcgtgttgc tggcgcagcgc gcacaattcag ggccgcacgcg ctggccgctgct gtcgcgacgcg 780
ttttgcaata ttcgccagcat cggctgttctct cggccacagtt cggccagcgcg 840
tttgacgagat tccctgtccat cgcctgtgagct ggtctggccac ggtctccagcg 900
ccggcgcacgc ggcttgccgct ccgcctgctgct gcggccacagc gggccgccagc ggcgcggccct 960
tacggagcgc cccatcgccag gcgcgcgcgc gcgcgtgctgcgc gcagccatcagct ggcggcgcgcct 1020
gtcgctcag gcctggctgct ctgggcgctgc cagccacgcgc tgaaggtgcgc gcggcgcgcct 1080
aacctcggct gcgtgctgcc ggtcgctccgc ttcgacgacgc aagccgcggg cctgaacgct 1140
gccgacggac gcgtgctggct tccaagggc cgagtcatgct ggcgcggcct cccgatctgg 1200
aggccgctct tgcgacagaact atggcgtgctgc aactggcgcgct ctaacgcagct atggccagccc 1260
gtggggccatt cccggcgcct ccctgggccc ctcagccaaa gcgggtttgg tccgctagct 1320
ggcgcggacag ccggctgctgca tccaatggctc ggcgctacgc ggcgcgtgctgc gcggcgcct 1380
ttcggcttaag aaccccaactgt ctgcccttcg cgggctatgg gctatggaaaa ttctagaaggt 1440
ttcctcagtt tttactccag gcagagctgc tggctcgcag gcggcgcgcgc catcagccaa 1500
ggcgctgcaag cctttcgccgga ttttttcatt ttcgattgga cacaacggct gcacaaccctg 1560
aaactgtcgtg ttcttggagg ccgagatgact taa

<210>  4
<211>  1494
<212> DNA
<213> Methylocomas sp. 16A

<400>  4
atgacactcaag acgcgtgatgc gtcgacgccg ccgctctcg gcggctgtcct 60
ccggctatgtg cggcgcctttg tccgtcagcct gcacgacggc aaacacccaag 120
gtcgacggcg cgcgtcactca gacgctggctt cttccagctct gggcgcggccct 180
CL2230 PCT Seq Lst.

atattgacga tgcggcacat ctgggagcct tgttccagct ggcgcggcaca aacatggccg
240

gattacgtgc aatcccgaaa agtcgagccac tcgggcagcct tgttccagct ggcgcggcaca aacatggccg
300

gtcatcgcact tgtgcaagga ggcgcggcaca aacatggccg 240

gcgttcgga tggcagctgct ttgggagcct tgttccagct ggcgcggcaca aacatggccg
360

gccatcctg caatcctctg gacattcgtg aaaaaacacc aacatggccg 420

gagcgcgggtt atccggcgaa gggcgcggcct tgttccagct ggcgcggcaca aacatggccg
480

cgcttcggcca gcgttcggtgct ttgggagcct tgttccagct ggcgcggcaca aacatggccg
540

cttgtttagct acgacgccgct ctcgaatgatc acaatatcata atccagggc
600

tcttcgtagct acgacgccgct ctcgaatgatc acaatatcata atccagggc
660

tgatctgtgcgct ggcagttcgtgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
720

gacggcggaagcttgctgtgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
780

gacggcgcgctagctgctgtgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
840

aacatgagctacgatcaggtcgtgtgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
900

aaaatcgacgcgtttagctgctc ggcagttcgtgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
960

tcgtctgcggca acggacgggtgctgctgtgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1020

gatgtctgcggcatatggtt actvggacgc ttcgagtttgct cagcagctggatggtggatcct ggcgcggcaca aacatggccg
1080

tgacgcggcgtctcgtgtgctgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1140

acggacgggtgctgctgtgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1200

acggacgggtgctgctgtgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1260

agacgatgtgctgctgtgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1320

tcggcgctggactgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1380

atcgtcgactgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1440

gtggacgctgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
1494

<210> 5
<211> 1542
<212> DNA
<213> Methylomonas sp. 16A
<400> 5
atgctgctgca ttcacaccag aacatggtgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
60

ggtacgtgctgctgcgct ttcacaccag aacatggtgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
120

gtgcagctgcgct ttcacaccag aacatggtgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
180

ttcgactgctgctgcgct ttcacaccag aacatggtgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
240

gaagcgctgcgct ttcacaccag aacatggtgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
300

cggcgctgcgct ttcacaccag aacatggtgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
360

gtggacgctgctgcgct ttcacaccag aacatggtgctgctgcgct tgggagcct tgttccagct ggcgcggcaca aacatggccg
420

Page 4
CL2230 PCT Seq Lst.

agcaaaatca ctcgcgtacc acccaacgca gcggctgcca atttgttctc ttacgcgcgc 540
gcacccttgg cggagagct gatgaatatt ttgctcgcacc cacatctttaa aagtgccgctg 600
gccgcttcat ggcctttatttt ggccctaccc gcctctaaac tgcctttttt atataaggcc 660
agcattagtt cggggtacac tcatcagagtt gctgttatttt gcggcggcag ttttcaaccc 720
tatgcaccca gacctgcggca aggctattgg aagccggcgcc gcgcaggttgc cattggaccc 780
agctgctgagc ggaatttggtt gggaaaacgc gcggcagagct gcacctggtc ggaataatgg 840
cacaatatac gcgcaaaagcg cggtgttctc cgtattggcgc cggcgaccac gcgcaggaatta 900
cgcagttggtc cgccagatttgc gattgcgcac actatactatttc gcgcaccaca 960
tgcgctgagtt ttctcggcag ctatcatgcaga accgatggatt ccatacgagac gcggcgtttcat 1020
agccacgagct ggtttttttttt ccaaaacttcc gatcagaaag cgggggtttggc atccacgcac 1080
aagggcgcag ccaatgttgg gtgcgcaccc ctgtgcagct gtagcgagtc tctgcgtgga 1140
cgcggcggttc cgcacatccct gcctcgctgac accttatgct gcgttgcaaat agggccaaagc 1200
tgcgcacagg cccaaactttga ctttgcagca cgctttatgg gcagcaagcgc acaacaataaaa 1260
cgggcctgtaa aagacatttt gtttgctgata gaacctgggt gcggcggcgc gctggaacgc 1320
tacaacctca accacggccta ggcgcgttcg gggcttggcc ctacccccca tcaaatgctgg 1380
ccaaaactgtc cggacggcttc cgaggcttgt tcggggtttgt tccacacgg gcaacctggagc 1440
cgtcggggcgc gcggcgcctgc gcggcgctcgtat atttcggcgtc aactgggcgcgc acaacgcaattt 1500
ttgaaacctgc cccataacgc gcatttttgtgg aacagccctgg at 1542

<210> 6
<211> 5632
<212> DNA
<213> Pantoaea agglomerans

<400> 6
cattgaccggc gcggccggfggc gcggccgaga caatcggcga caatcggccc aggggagccga 60
aaccaattgg ccagtcgggttg gcggccttt gcgggtggagc gcggggcggcgc 120
gacggcgttg gatcgcgcgt(gc cccggcgggac gttgacggggt tcacttcaggt 180
tggcgccggcg gcggcgttgc gtttgacgtgc gcgcgggggc gcgggagcgt 240
gatgcgacgc gcggggctttaaac gttgtgctat gaaacggttc caggggctgc gcacggcgtg 300
cggggagtgc gcctggggtc gcattctca gcgtatttgc caggttgttc cgacgccccgt 360
gttacgggtg tgtgttaaac ggtggtggcc gcggatctcc cgggagcgttg cctgccgact 420
gcgccggcggc gcgggtgtgg gatgccgata ggcgcgggtc gcgggacacc gcgtggcgtgc 480
ggcgcgggt gcagcggggtt ggacgcggcc ggacatctgg gcgggaaact gcgagggggcc 540
tgcgggcaattt cattgcctgg aatgcggggtc tgcggttccat cttgctgcctc tctgatctcc 600
gccatggctgc cttgccggccg ctcgggacca gcttgatggc gttgggttttg gcctttggcg 660
tgctagtccgt gtggggcaggg cggccgggttg gtcggcctgc gacaacgcct cggcggacgcg 720
gcgaatttcag cctttcgcgg cccatgattgc gcagctgcgg gatcgcgggt tctgtacccgc 780
CL2230 PCT Seq Lst.

ggtaacgggg cagcggcgctgg aggctgcggcc ggtggtcctcg aacgcggcgacg tggtaataac 2880
catatgccca atctgcgctgc atacccgca cggcggcgcct acggcgcacat agctgaaacg 2940
cagcgcatag gaacaactgcg ttgctgtgtcct atatattggtct ctgagatcacc atacacccca 3000
gctggcgcacc cacaatgctct gtcttgcgccg tctatattcag agtattatgtct 3060
cagcgccgac acccttctggg aaagatatttc gctctatctg cattgcgccttg cgtcgaaccg 3120
ccgctgcctgg gccgccggccg ggtgcggccag ctctactatgc cctgcgcgccc gtcgcaacct 3180
cagtgtaaccgc ccggctgcact gcagcgtgtaga aagggcgcgctgt ctgcgagggtc gcatttttga 3240
ttagtctgaa gcgcggctata tgcgcgcccgt gcgccttcgac cggtgtacgc accgcatcttt 3300
cacgggacgag gattttccgc atacgctcag tgcctcgccag ggtctcgctgag ttttcactgga 3360
gcggatctcc aaccagacgc cttgcctgcgc gcgcgccacac cgccagacgcg tgtggtgataa 3420
ccttcatcctg gtcgctgcccgc gcggctcgcag ggtgccggcctgt agatgcgctgcgcgc 3480
cgcccacgac ccgctcgccag ttataaaa gagatattgc taataatctca gcgccttcccct 3540
gaacaagcga gcggcaccat gacgcggcgt tctaaatgttc tgccaccgcg ctcgaagctcgt 3600
tttgacacac gcacccggcgc caggcgctgcgt atgctctata cctggctgccgc ctctagctgcag 3660
egatggtacgc gcggagacacg gttggtgttctg cggcccccgcg ccggagcgacgc gcgcagcgc 3720
agggcgggctg tcaacagcaccgc gcctcaagatgc acgcgcggccg cttacgcagcg ggaacacagtg 3780
cagagagcgc gcgttcgccgc cttctcaggat gcgtctcctgg gcccatgcacac tccgcctact 3840
cagggcttcgc agcaccctgcc agtggagctgac gcaacagacgc gcgctgccgccgc 3900
cctgcagcgtact ccgccccgact gttgcgcccggc tgtcggctgctg cttgactggtgctgaccg 3960
aggggtgattg gatgcggcagga ggaagccacg tggctcgtcgc gcctggcagcgc ggttgctgcgctg 4020
ctccagctca caatcatgcgc cagggatattgc tgtgacagatgc gcgcagtggcg acgcagctac 4080
ctgcccggccgc agttggctgcgg gcacagtgcgag cctcaggttgcgc cttcagacgcgctg 4140
aaccgctgggg cgtgcggcgccgc ttgctgcgttaga ccggctgcttgc gcgccctcatatc 4200
cagtcagcgc gtcggcggcgc ggggtgatcgc ccggctcgccgc ggcgccggccgc cattgtgccgctgc 4260
gcgcagcgggg tgtgctcgcgtg aaggggcttcg tggccgggtcg aagcgcagcctg 4320
tgatccccccgc ccggcggggcg aggcgcggcttg cgctcttcgct ccgcgcgtgcgg 4380
caggcgatcgg cttccggctgg ggccgagcgggg ccgcgcggcgct ctcgccactcctgcgcgcgc 4440
ccgcgcgtagak atcggacttc actagctgcag acggcgcggta ccaaccccctga caagaccccct 4500
tgaacacatc cccgccgtcgcc gcgcggataact gcagttcgctgc gcctgcgctgctgc 4560
cctggagtgaa ctcggctcggc gttggcgatct gcggcagctcgc atgctgcctgg gcggcgcgcac 4620
agggctactgc gcagcgggga aacgcattctcg cccgcgtgcctgcagtcctgcgcg cccgcgcgcag 4680
ctgctgctcgc gatgcgcgacc cccgccgctgc gctggtgatatg gcctgctgctgg tgcaatagtgctg 4740
gcaacgcttcgc tgcgtcatact gcctgtgtctg cttcgcgcggc gcgttcgcccgg 4800
cggtcgccctc accattcact gcgcgctgtaggc taagcagctga gcaatttcctgct gcgcgctgtagc 4860
CL2230 PCT Seq Lst.

gttgctcagc agcgccctttt gcgtgatgtg gcggcgcgac ggatgtcttc cccgagtgcgc 4920
cagccaggcg tgtgccagagc tgctcagtgcc ggtctgctgcc cagggtcttg ggacggtgca 4980
gtataaggct cttgccttcttg ccaccgcggg cgcacagcgg gaggagatcg ccaccacccaa 5040
cgaacttgaaa accagcaagc gtgtttgttc cacagtctcc aatcgcggcag tgtggcgccg 5100
cgctctgccg gggccgcgcgcc agaaaaatccg ctgctgggctg caggatttag gccagccgct 5160
cgcagcgtcttg gacgcatctgg cggacgggcca tgcgcgggacc ggcacagaga tcaataagga 5220
cgcggttaag tccacgctgg tgcgagatct ggcagacgac gcctgctggc agcggcctga 5280
cacccatctcg cgccgcgcag acgcctcttt ttgcacgcgcgc ttggagaaaaa accagggcacc 5340
gcagcgcgtt atgcaggcct gggtttcaaa acagcctgggc gccttttagct gacggaacgg 5400
tacaccccggg taattattgt ggagatcaca tgaagagcac gcctcctggtt cagcgttaaa 5460
atgaacactct ttgctgttgctt acccgccgat ggatccactc gcagcagcgtt 5520
ttgcagccttg gcgtttgtaa cactggcggc tccggcagct ggatctcagc ggtatctgatc 5580
ttcacccacct cctgtttttccc gcgccgcgtga aagccgccggt gcgtgatcagc tc 5632

<210> 7
<211> 756
<212> DNA
<213> Artificial Sequence

<220> Codon-optimized crtW gene from Agrobacterium aurantiacum

<400> 7
caatggaagg aggaaaaac catgagcggc catgccctgac gcaggccga cctgcagccg 60
accagccctga tcgctagcgg tgtgcatcatt gcggcctggc tgccgctgcc tggctctgca 120
cggctgtttcc tggacgcggc gcggccatcga tccctgagcgc ccggccacgta ctcggccctg 180
acctggtgctcg gcggctggct gttctcatcgc gcgcatgacg ccaggtgctgc gcagcggctgc 240
cgcgggtggct cgcgctgcacag gcgcgcatggc ggccacgctag cccagctgctt tcagcgcctgc 300
ttgcagcctgac gcagagatgtgc ctgttaacat gttgcctcgtc atgcgcgcgta tggccgcgctg 360
gacgatgcggc acctgcgaccg tgtggccctgg gcggctgtgcc atgcgcgcctg tcggcgcaccg 420
tttttgcctg gccgctgagag ccctgctgctgc ggctgtacgc tcagctgctag tcgcgcgtgatc 480
cggccgctgc gcggctgtgct gcagcagcctg ccaccgccgc ccggccggtcc gcggccggtcc 540
caatggttgc ttcggctgcc catgcgctgag ccggcagcgc cttcgcgggc gcggccggtcc 600
cggccatacgc ccgcgacgcc gcgcatgcggc gggccgggca gcggtctgctg tcgttctctg 660
ttcggcggct gcacacagcgc gcagcagcgc gcgccgccgc gcggccggctgc gcggtctggctg 720
agccacccga ccaagccgca caccgctgta caatgg

<210> 8
<211> 533
<212> DNA
<213> Artificial Sequence
Codon-optimized crtZ gene from Agrobacterium aurantiacum

caatgtgtgc ctgaagagga ggaataacc atgaccacct tcctgatggt ggtcgcccc  60
gttgccgtca tggaaactac ccggttatagc gttccatctg tgggtcccttgg  120
ggtggggggct gcacaagag cacatctgaa gaacatgacc atgccccctgg aaaaagatgc  180
cctgctgtgtgct gtacgctgcg cgcatacgcc accgtcctgt tcaccgctgg ctggatctgg  240
gctccagttct tgtggtggtat gcgcctggtg gacggtgctct atgggtttgat ctaactttgtc  300
cgctcaggtg gcctggtcgct tcaacgcttgg ccgttccctct acatcaccgt ccaaggtctat  360
gctcgttcgt cttatatcagc ccatcgcttgg cactatgccc tcgaagttgc tgaattttgc  420
gttcgtctgg gcttcatccta tggcccacccgt ctcgacaaac tgaacaagatg cctggaagatg  480
agcgggctct tgtcgtgccga agcccacaaag cgcaccaaat atgcttggaa ttc  533

9
34
DNA
Artificial Sequence

Primer

9
gacatcgata gcgatgtc gacgtgctgg gat  34

10
32
DNA
Artificial Sequence

Primer

10
32
gcctcctcct gcttatttag ttcactgtca gc  32

11
27
DNA
Artificial Sequence

Primer

11
27
agatctcggga cgctcact cccacat  27

12
26
DNA
Artificial Sequence

Primer
<400>  12
gcatgcgccg ggtagacgtc gaacac

26

<210>  13
<211>  29
<212>  DNA
<213>  Artificial Sequence

<220>  
<223>  Primer

<400>  13
agatcttttcg aaggggtatt cgacgaagg

29

<210>  14
<211>  37
<212>  DNA
<213>  Artificial Sequence

<220>  
<223>  Primer

<400>  14
gcatgcgcgg ccgccgcat gtttctggga aatcagc

37

<210>  15
<211>  28
<212>  DNA
<213>  Artificial Sequence

<220>  
<223>  Primer

<400>  15
gcggccgcaa tcagcaagcg tgcagc

28

<210>  16
<211>  26
<212>  DNA
<213>  Artificial Sequence

<220>  
<223>  Primer

<400>  16
gcatgcgttg cggaagc ctgctcc

26

<210>  17
<211>  39
<212>  DNA
<213>  Artificial Sequence

<220>  
<223>  Primer

<400>  17
gcatgcctcg agttcgtatcg tcgtcatact caggctttg

39

<210>  18
<211>  26
<212>  DNA
<213>  Artificial Sequence
<220> Primer
<400> 18
cctgagaatc agcaagcgct gcaagc 26

<210> 19
<211> 28
<212> DNA
<213> Artificial Sequence

<220> Primer
<400> 19
agatctatcc gattccggtc atgctgga 28

<210> 20
<211> 25
<212> DNA
<213> Artificial Sequence

<220> Primer
<400> 20
gcatgccttt ggtcatgatg tgagc 25

<210> 21
<211> 26
<212> DNA
<213> Artificial Sequence

<220> Primer
<400> 21
acgcgtggtg agggcaaga ttgtgg 26

<210> 22
<211> 26
<212> DNA
<213> Artificial Sequence

<220> Primer
<400> 22
gcatgcgct ttggttagcg atagcg 26

<210> 23
<211> 26
<212> DNA
<213> Artificial Sequence

<220> Primer
<400> 23
agatctccggt tctgtacact gatccg 26
<210> 24
<211> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 24
tagctgccc cgcccatttg ttgctgatat aatccggc

<210> 25
<211> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 25
gcggccgccgc aagcgcgca acagggattc c

<210> 26
<211> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 26
gcggccgccgc aatacctcga cattcaagc

<210> 27
<211> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 27
tagcttaact gtgcggcgc cgtagc

<210> 28
<211> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 28
gcatgccgac atctagttgt ccagc

<210> 29
<211> DNA
<213> Artificial Sequence

<220>
<223> Primer
CL2230 PCT Seq Lst.

<400>  29
agatctttgga gcttgatcga aatcgtcg

<210>  30
<211>  28
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400>  30
gccgctcgctg tctgtgcaat gcacacgac

<210>  31
<211>  729
<212> DNA
<213> Agrobacterium aurantiacum

<400>  31
atgagcgcac atgcctggtcc caaggagcag atgcggcctgca ccagcctgctgc tgcctggtggc  60
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  120
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  180
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  240
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  300
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  360
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  420
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  480
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  540
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  600
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  660
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  720
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  789

<210>  32
<211>  489
<212> DNA
<213> Agrobacterium aurantiacum

<400>  32
atgaccaatt tctgtatcctg tctgcgcacc gcgggtcctt gggagttgcgc gcgccttattc  60
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  120
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  180
ggccttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  240
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  300
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  360
ttcgcttcgctc gcgcctgctg ggcgcctgctgc ggtgcctgctg gcgcctgctg gcgcctgctg  420

caccacggcgc gcgcctgctgc gcgcctgctg gcgcctgctg gcgcctgctg gcgcctgctg  480
CL2230 PCT Seq Lst.

gtcgacaagc tgaagcagga cctgaagatg tcggcgctg tcggcggcga ggcgcaggg

cgcacggtga

<210> 33
<211> 242
<212> PRT
<213> Agrobacterium aurantiacum

<400> 33

Met Ser Ala His Ala Leu Pro Lys Ala Asp Leu Thr Ala Thr Ser Leu 1 5 10 15

Ile Val Ser Gly Gly Ile Ile Ala Ala Trp Leu Ala Leu His Val His 20 25 30

Ala Leu Trp Phe Leu Asp Ala Ala Ala His Pro Ile Leu Ala Ile Ala 35 40 45

Asn Phe Leu Gly Leu Thr Trp Leu Ser Val Gly Leu Phe Ile Ile Ala 50 55 60

His Asp Ala Met His Gly Ser Val Val Pro Gly Arg Pro Arg Ala Asn 65 70 75 80

Ala Ala Met Gly Glu Leu Val Leu Trp Leu Tyr Ala Gly Phe Ser Trp 85 90 95

Arg Lys Met Ile Val Lys His Met Ala His His Arg His Ala Gly Thr 100 105 110

Asp Asp Asp Pro Asp Phe Asp His Gly Gly Pro Val Arg Trp Tyr Ala 115 120 125

Arg Phe Ile Gly Thr Tyr Phe Gly Trp Arg Glu Gly Leu Leu Leu Pro 130 135 140

Val Ile Val Thr Val Tyr Ala Leu Ile Leu Gly Asp Arg Trp Met Tyr 145 150 155 160

Val Val Phe Trp Pro Leu Pro Ser Ile Leu Ala Ser Ile Gln Leu Phe 165 170 175

Val Phe Gly Thr Trp Leu Pro His Arg Pro Gly His Asp Ala Phe Pro 180 185 190

Asp Arg His Asn Ala Arg Ser Ser Arg Ile Ser Asp Pro Val Ser Leu 195 200 205

Leu Thr Cys Phe His Phe Gly Gly Tyr His His Glu His His Leu His 210 215 220

Page 14
Thr Ala

<210> 34
<211> 162
<212> PRT
<213> Agrobacterium aurantiacum
<400> 34

Met Thr Asn Phe Leu Ile Val Val Ala Thr Val Leu Val Met Glu Leu
1 5 10 15

Thr Ala Tyr Ser Val His Arg Trp Ile Met His Gly Pro Leu Gly Trp
20 25 30

Gly Trp His Lys Ser His His Glu Glu His Asp His Ala Leu Glu Lys
35 40 45

Asn Asp Leu Tyr Gly Leu Val Phe Ala Val Ile Ala Thr Val Leu Phe
50 55 60

Thr Val Gly Trp Ile Trp Ala Pro Val Leu Trp Trp Ile Ala Leu Gly
65 70 75 80

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

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

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

His Cys Val Ser Phe Gly Phe Ile Tyr Ala Pro Pro Val Asp Lys Leu
130 135 140

Lys Gln Asp Leu Lys Met Ser Gly Val Leu Arg Ala Glu Ala Gln Glu
145 150 155 160

Arg Thr

<210> 35
<211> 950
<212> DNA
<213> Pantoea stewartii

<400> 35
gaatgcggccc ttgacgtcct ggcgaaaaa acagttcac cttactggca tttcggctga
CL2230 PCT Seq Lst.
gcagttgctg gctgtatagt atagccgccct tgtatcagtt ctcgccggttc aggggtgacgc 120
 ggattgttgtgc gggtggcggc cagcgtgaaggtc cagcgtggcc cccgggaacaac gttatcgtgcc 180
gatgctgtcgtg ttatattacag ccggcgatactct tgtggtgtgcg atccattcag ggggtcattact 240
 ggatttaccc tgtcggcgtgtg aaattgggtca tgtgcgctgtcctg tgtgattcag atgatatgctcc 300
tctgctggagct gcctcggggc gctgcccacc attcacgacgcc atgcacggtga 360
 actgtggggct acttgctgcctt actcaagcaaa gcggtgtggag tggatccgaga 420
 ggcgtgaaggt ctcagcgcctaca tagcccaaaaact ctcgccggttc tccgagctcgtg ccacctgcatg 480
tgccatcgag tgcctgggttc aggccagcctt taaggacccct ctcgggaagccc ataataaccccg 540
cagccggcag gccaatactcg tcaaccaatactgc gtttaaaacagc agcagctgtccttttgccgtgc 600
 aacgcaaatgtg cgcgctcattgc cggcccaacgc gtctgcgtcag cgcacgtgaga acctgcactcg 660
tttctcgtcctc gatctggcag aggcccccttca gttgtgttcgat cttcgatcagctgct 720
gcataccggag aagacatcgctc atcgaggtgacctc aggtaaatcag acgcgtgagtc attatattagg 780
cgcggcggagtgc gcgaaacgctgc gcttggcagca agcattggcctggctggcagtttacccctcc 840
cgcggcctcg ccaaaacccggc atcgcctcactc ccaactttttatcaggtgcctgc gtttgacaaac 900
 aaaacttcgct cgcgtcagttt gaaatattgag tgggccaggtact gtagcagagc 950

<210> 36
<211> 3611
<212> DNA
<213> Pantoea stewartii

<400> 36
aaaaactcgct gcgcgctcggtt aacaattgagt gggtgcaggt catgtcagcc 60
tcattctggtc cggctgcgggt cttgcagttctg gctttactgc gctcgcgggtt cagccagcagc 120
atccggatag gcttgcttctgg cgggtctctgag ggcggagagg gaccataacct 180
gttcccttcaga cagaggtgata ccaacgctgctgctgcagctgc gttgcttctgg 240
ctacctacag gacctactgat ggcctgttcgag cccggtatac cccgggcaacg ctttgagcc 300
atctatggtgc gctttccgcag ctcgcgttcgttt ccccccaacg cggctggcaaat gttgcaggtg 360
attatatggtg ccatacgtgc tgcgcgcttggt tctgtatgca atcggctggtttagc 420
ggcggatatc ctctgcttctgc ctcagagtcag cggaggaggg cttacagcctg gatcgtgctc 480
taccgcgttggc ctttcagcctgc aggtgtgctac acttgagcggc cgcagattggtctc 540
atatcgcacct gattatcagtc gatcagcagcgc ctcgtcagaca aatggtcatac gctttgagtc 600
atatcctgcag cttccgcagag ccgcactggtg gcatgcaagag cccacactag atccgagac 660
ctaatctccgcc gcggcaagcgc ggcggcctaga accattgcagtctgc gctttgagtc 720
ggccgggtaca gacttggtgcg cgggaagacg agggtgcatt gccctattcgc ctcagctggttctg 780
atactgctca gttgctggcc cagcagcgcag gacgcggttag cgggtactgc gcggctggccgc gcggcggctgt 840
ttcactgcggc aacggcggttc accttcagcag ctcgggctagc gcttcagccgctgcagtgcgcgc 900
cgctgaggtatttacacgcttc ttcctgcttgc accacagagacatgtcactgt ggccccacggcagc 960
gttgccagca acaaggggttt ttccgcagtc gaatatgccat ttttttttttt gcccggaccgg 1020
ccgagtcagc tggcgcgtgtg atgacaggttt tcttttacctt acccaggagat ttgattgcc 1080
gcttttatgac gggaaaactc accgtgacgc atcggctaca cattctgcag gccaagccgc 1140
cgcagggcgt tttcgcggca ttcgagccaa ttatgacgac ttcactgtgga agagcgcata 1200
catgaaacca actacggtaa tttggtgcggg ttttttacgtg ctgggcacggt caattctgtt 1260
acagggcgcag ggtatttccgt tttttgctgt ctggcagcgc gcacagcccg ggtgcggcggc 1320
tatgttttat caggagcagg gcttttacttt tggatgcggc cctacgctta tcacggatcc 1380
cagcgcgatt gaagaacctgt ttgctgtcgc gctgtaaacc ccttaaggatt agctgcaagct 1440
gtggccgctg acgccagtttt atcgcctgtg tctggagcttc gcagaagctc tcaattacga 1500
taaagccagc ggcagcttag aagcgcagat acagcagttt aatcgcgcggc atggctgggg 1560
ttacctagcg ttcctcgtac actcgcggttc ctggctcata aagggctcac tggagctg 1620
cactgtgctc tttttatcgt tcaaaagacat gcttccgcggc gcgcgacagtt tgggaaacag 1680
gcagccgcatg cgccagctttt aagcagctaat tggccgctaac atggcagagt acgaatcttg 1740
gcagccgctt tctttccact ctgctttttat ggagggggaat ccccttgcaac ctcctgtcct 1800
atatgcgctg atccacgcgtg tagaagccggt atggggcgttc ttgcttttca ggcggggtgac 1860
cgtgtgcgtgc tgaatggcagc gggcgtggagt ctcgggctgc aagtcgctctg 1920
taagcgcggc gtcagctcata tgggaacccct tgggaaacag attcagccgc gtcgttgaga 1980
agacgcgcag cggtgctttg ccgggtgcaac gggcgtgcttg ttcagaatctca 2040
tgcggatctg tgggctgacg atccggccagc cgctaaggct gcgaaagacg tcgaaccttc 2100
gcggatagt aacctcactgt ttgtagactta aacacttccct aagcagcttc acgtcacttc 2160
cgccccatcat accgtctgtgt ttggggcacc ctcccctgaaa tctgatcccg ccaatctcct 2220
ccatgaggtg tttggtgcgctttt attttttacc gacacccggttg tcacgggatcc 2280
gtcactggaga cgggacgagtt ggcggctgctg tatgttgctgt gcggctgcttt cacacattgc 2340
cagccgcaac ctcgatcggg ctgtagaagg aacccgcgctg gcgcgctgtga ttttgacta 2400
cttggagcaat cattccatcg gcggctctcg gcggacgacc ctgggtactggtg tccgggttcctt 2460
gcagggcgtt tacggcgacg agctccaatgc gcggacgagtt tccgcgacttc ctgggtggatc 2520	tatctgacc cagagcgcct ggtttccgacc acataacgcgc gataacgccat tggataatcgtt 2580
atatgtgtgtt gcggaggcaa cccatctggt ccgcyggcac aaccgcgcgtaa tcgctgccctg 2640
gagggcgcag gcaggtcatat ttgcctggga cctgatcggta cgaatagctc accaaccct 2700
catgcgcgtgc aaccctaggg ggtgggcctgc aaagcttcttg gcagcgtcctc gacgccttcttc 2760
gacgcacacac cccgtgcagcg tggctgcgta tctttacgcatt gcggcgcgca tgcgaacgcc 2820
gtcttgacgc atccaacact gcgcgctgtc gcgggacgcgc cccctgctga gatgccgctag 2880
cagccgcgtgc aagcagcttc aatgaaacag ctcgagccct acgcgggttc gcgaaatgccac 2940
agccgcgttt ttgccgcggtt tcaggagtt gcggctgcctgc atgatatgctg tcccgcctcct 3000
CL2230 PCT Seq Lst.

gcgttgaacc atctggaagg ttttgcattg gatgtgtgcccc aaacgcgcctcta cctgaacactg 3060
gacgatacgcc tgcgttaattt gctacacgctc gccggtgtggt tgggccctgat gatggcgcaaa 3120
attatgggagg tcgccgataaa cgccacgcctc gatcgccgctc gcctctcctgct gcctggttttcc 3180
caagtaaccac aacttgcggcc tcatttttgct gcagatcttcct cctgtggttctgctctgct 3240
ccttaaagct ggtgggaaga ggagagactgt acagaaacgca attatgcttc gcagaaacaaac 3300
cggcgaccct tgatggtatcg cggcgaggcag gcgtgtcaggg aagcgggaacc cctattgctgat 3360
tcatcaatgg ccgtgctggcc acaatttcaccc ttacgtccgg cctgggcccccg cgcagacgacg 3420
aagcatgctgt acgttaaatg tggcgtgaaat ggtaacaggg cggtaagcg cggctgggatt 3480
catgcgcacgt ccacgttccac gcggcggaaaaa ttacgctttt tgctgacgggc atccgggtcag 3540
gcagggattc cccggtgaga cacgtatccta cccctgtcttg cttctctctg gcagcggcccg 3600
actctagttt ac 3611

<210> 37
<211> 20
<212> DNA
<213> Artificial Sequence

<220> Primer

<400> 37
gaattgcgcc ttgacgttct

<210> 38
<211> 50
<212> DNA
<213> Artificial Sequence

<220> Primer

<400> 38
cgggccagata atccgctcaca ctcattaagttg aactgacggcc aagcgagttt 50

<210> 39
<211> 50
<212> DNA
<213> Artificial Sequence

<220> Primer

<400> 39
aaaactctgct gcgcgtagttt aacatgggag tgggcaggta atcagcaccg 50

<210> 40
<211> 25
<212> DNA
<213> Artificial Sequence

<220> Primer

<223> Primer

Page 18
CL2230 PCT Seq Lst.

<400> 40
ggtaccttga tcgggccgctg ccaga

<210> 41
<211> 8814
<212> DNA
<213> Pantoea agglomerans

<400> 41
accgccaaga cagcaacgtg ctggagaccc gctttgagac ggaaccgggt cgggtgctgat  60
caccgagtcc ctgtaacagca gcgttctggc ccggctgcgg tggacgacag tgggcccgcgg  120
catcgacggt attgaggcc ggctgagcgc ctgctgctgct gacgtcgagc ctgcgctgctgc  180
cgagacgcgc tccctcgtgga gggccagacac ccatttgccg gatggtgttgg aacttgccga  240
tcgtgatgcag atgcgtcgcg caagccgaga gagcttgatt accacactgc acgcagaaac  300
aattaccggc cagctgagta ccctccaggg gccgacgcgg ctgggtgctggt ctggggctgca  360
cgagaaaaag cccgctgctgg tgcgggatct tccggcgcac ctggagccga tcggagacag  420
ccacctgccc cggggtgtcg actggtgaatt ttcctggtacct cccggagacgg ccgggctgccc  480
cgctaaaaagct cccgctgctgg tgcgggatct ttcctggtacct cccggagacgg ccgggctgccc  540
gccggcgccacc accacctgcg tgcggagaggg cattggcccg gagaagac gacgactaccc  600
ctatgcctgtt cttgctgat gcctgctggat catcaacggc cttggtttgc cttggtgctct  660
ggagacgtgc aaacgccgctt ccctctgggtct gcggagaggg cattggcccg gagaagac gacgactaccc  720
gtgcgccgct gctgtaagct gcggagaggg cattggcccg gagaagac gacgactaccc  780
gccgggtatc cggcagttcc cccgctgctgg tgcgggatct ttcctggtacct cccggagacgg ccgggctgccc  840
ggctcgacag tgcgacacac gtcggcgtcct tttatcggag cggagcacgt  900
actggtatcgc gccacctgcct gcgtgctgcttg cgaactgcgg cattgcgtcg ccggacagct  960
gcgcgagac gatccgacgca tctggggaatt ccggagcagc cagcaacttt cccgctgctgg 1020
gatggcctgc tggctgctgac tggagcggagt ccgagacac gacgactaccc 1080
accggacctgg gtcgggcgcct gcggagacac gcggagctgt ccctggctgg agtcagagcaac 1140
ccacctgctgg ctcggagagtg cagccgtccatt gtccgggttc gatccgggaat cttttcggag 1200
gatggcctgc tggctgctgac tggagcggagt ccgagacac gacgactaccc 1260
gccccctatc cgcgccatcat aagcagagct gcggagacac gcggccctgc tttacccgcc 1320
cagcgaggtg gagaagaaaag aagcgcatttt gccgctgctgc cggagcggagt ccggagccgg 1380
cccgcggcgc ctcgctgctgaa cccgagcggc ccggcggcgcct ctcgctgctgc 1440
gctcgtgagac cggggccatag ttgaaacttt taacgagctg tttgatgctg gatccgggaat 1500
gtgccccgccct aaacccctcgc agggttcgag ccctctggcct gcggagctgt ccggagccgg 1560
gctttggaa cccgacgcgc ctcgctgctgaa cccgagcggc ccggcggcgcct ctcgctgctgc 1620
tgacacatcag gatctgtgaa catatggtta ttacccgtgc cggccctccgc gggggcgagc 1680
ggggggggcttc cggggcgtctg gatccgagtt ccgccgctgg ctcgctgctgc 1740
cggacaatgc catcagcggc gaccagatga ggcgcgtcata cttgctttccc cgggagatgc 1800
ggttacccgg ccctggccatg ccctggcttcc gcagccgcttg caaatgtaggc ctgtttgaga 1860
tcgatcgggc ggttggccgac gcggccgaaat atatcgagcga ggccggcccttg aaccacctgt 1920
cggtttacgt ggcagacatt cagcagccgc gcggcagccgta aagaacacag ggccgccacgc 1980
tgttggaggg cgcggagcggag tcggctttgcttc aggagagagg ccggcgacgac cagacgtcggt 2040
tctgagcggcact gctttttgcg ccctgtgtgtag aactcacttct cttctctctcgg cctctctcgt 2100
aagatgcggccag ccctggcagcaacctagctgg gctggccagctgc tgccgactgtc ctcttcagcgc 2160
ggccgcatgcc ccggtggtctca taccggctgct acctgctgctg aaaaaatattt caataaaattt 2220
tctattttta gcggactatttta aagcagacatg cggcgccgtag acatttatttt caattttttctc 2280
cctacagtta aaaaacattgag tggataacttt tcggatacctgc tgcacagcggg ccctggccactg 2340
gacccagcgtgg ccgcttgctcc gctttctcc gtaaactgcgc gtcagcagcgc gaacactggca 2400
tgcagcagc ccctgcgtgctt gcaactcgtgc cggactggtcg ccctggctggag cggactggtcg 2460
cctcgcgagcc ccctggcgtgctt gctggagctgg cggcagagca gcggtcagctcc cggctgagcag 2520
gccgctcccgc ccctggcgtgctt gctggagctgg cggcagagca gcggtcagctcc cggctgagcag 2580
tgccgctggcg ccctggcgtgctt gctggagctgg cggcagagca gcggtcagctcc cggctgagcag 2640
cggctccggcct gtcggccggtgc acatgtctggg cggggtggctg gctggagctgg cggcagagca 2700
cgggcgtcggc ggctggggcgc aacacggggg actctgcatcga ctgctgctgctt aagcagacatg 2760
tccctcgctgact gcagcattcct tggctgatgact gcagcagcgc gcgtgagctgg cggcagagca 2820
tccctcgctgact gcagcattcct tggctgatgact gcagcagcgc gcgtgagctgg cggcagagca 2880
tcctggagtgc ccgctgcagcgg ccggttggcag ctggctgctgctt aagcagacatg cggcagagca 2940
tccctcgctgact gcagcattcct tggctgatgact gcagcagcgc gcgtgagctgg cggcagagca 3000
ctcctggagtgc ccgctgcagcgg ccggttggcag ctggctgctgctt aagcagacatg cggcagagca 3060
tcctggagtgc ccgctgcagcgg ccggttggcag ctggctgctgctt aagcagacatg cggcagagca 3120
acacatataa ggagacgcggtt aagctcagcgc tggattggcag tcctggcagcgc gcggtcagctgg 3180
ggcgagcggtct gcacacccagtgc ccctggcgtgctt aagcagacatg cggcagagca 3240
ccctggcgtgctt aagcagacatg cggcagagca gcggtcagctcc cggcagagca 3300
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3360
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3420
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3480
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3540
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3600
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3660
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3720
gccctcgtgact gctggagctgg cggcagagca gcggtcagctcc cggcagagca 3780
gacatgatcgc acgcgacgacg attaattttg cattctgaacc cgcgctccag 3840
ggcgccgccc atgcgacgct gccgccccatc ctcaacgcac tcggcgcgcct ccctccgacgagcttt accatcggcg 3900
tcccggggttc ccggagtattt ccgtgttcgct ccccctccggg ctggggatccg 3960
cagctcactcc cgcgcgagtc aaccagcacc aacgctggattt ccggctgggag 4020
gccgggtgtta ggctgcgcggc tggggcgtgct tggggcgtgct tggggcgtgct 4080
gccgactcgtt gctcgctgct ctcgctcggc cccgctcggc tggggctggg 4140
atccgctcctt ctcgctcggc cggccatcgc cccgctcggc tggggctggg 4200
ttgtaaatgg atgatgcgatc cgcggccggc tggggctggg 4260
gacccgcggaa cggcccttgg ctccggctcgc cggccgctcgc cccgctcggc tggggctggg 4320
acggccagtg caacatgtgca cagcgcgctcgc ctcgctcggc cccgctcggc tggggctggg 4380
gacccgcggaa cggcccttgg ctccggctcgc cggccgctcgc cccgctcggc tggggctggg 4440
aggttgatgt gcggccccggc tggggctggg 4500
tccgccgccat cactgctgct ctcgctcggc cccgctcggc tggggctggg 4560
gctctttccca cacaagcata ctcgctcggc cccgctcggc tggggctggg 4620
cgacctgggg gacggcctcgc ccggtccctg tggggcctcgc cggccgctcgc cccgctcggc tggggctggg 4680
cctacccgacg atgctgctgct ctcgctcggc cccgctcggc tggggctggg 4740
tccgccgccat cactgctgct ctcgctcggc cccgctcggc tggggctggg 4800
acggccagtg caacatgtgca cagcgcgctcgc ctcgctcggc cccgctcggc tggggctggg 4860
aggttgatgt gcggccccggc tggggctggg 4920
tccgccgccat cactgctgct ctcgctcggc cccgctcggc tggggctggg 4980
tccgccgccat cactgctgct ctcgctcggc cccgctcggc tggggctggg 5040
gacggccagtg caacatgtgca cagcgcgctcgc ctcgctcggc cccgctcggc tggggctggg 5100
acggccagtg caacatgtgca cagcgcgctcgc ctcgctcggc cccgctcggc tggggctggg 5160
tccgccgccat cactgctgct ctcgctcggc cccgctcggc tggggctggg 5220
tccgccgccat cactgctgct ctcgctcggc cccgctcggc tggggctggg 5280
cacgcgaagc gacgctgctgc cccgctcggc tggggctggg 5340
cacgccct cccgctcggc tggggctggg 5400
cacgccct cccgctcggc tggggctggg 5460
ccggcccagcg ctgatcgcgct cccgctcggc tggggctggg 5520
gccgcttcg ctgatcgcgct cccgctcggc tggggctggg 5580
cccgcttcctt ctcgctcggc cccgctcggc tggggctggg 5640
cccgcttcctt ctcgctcggc cccgctcggc tggggctggg 5700
ccgcttcctt ctcgctcggc cccgctcggc tggggctggg 5760
ccgcttcctt ctcgctcggc cccgctcggc tggggctggg 5820
ggtgagctgta tcgggtgatc gcggctcttc tgccttctgtt ggaggaagctg caggttttc 5880
gactacggca acgtacaggg gcggctgtag tcggatacgg cggcttttaa cccgaacgagc 5940
gtgccgggct atccagccgtt ccaggcaatcc tccgggcggg tggcttggcag aggtatcttg 6000
aagcctcgggg cggggcgggtt tcctctgtgg tgcggcatgc tgcgctccgg ctcctcaactg 6060
ggccccctgc agggcttgcag cagcggtgta acaacagtta gcggctgcttg gaaagcttg 6120
cactctggta acggcagcgc gctggagagt gcggccggtg cctcgaacgc gcggcttgta 6240
aatctcattag ccggcactgc ccgccacgcgc ggtgcacgcc cccagttactg caaaaaagctg 6480
aaccggaacg gcgtacgaaa cggctctgcttc gttggcttatt ttgctgccga ccacctcaac 6540
accgcctgg ctcgccacata cgctctctgg ggccccgggtg ataaagcgct aatagcatgaa 6600
atccgccgag ccggcagcgc gctggtgagat ttttcctctct atggctcctc gcggctgctgta 6660
accgccagcg cgctgctggccgc cggccggctgt gcgacgactc atctgctgctgc gcggcctggc 6720
ccctccggtg aacgccggtg acgtgcgagtc gtaaggggag gcgtggctgc gtagtcgatt 6780
tttgattatct tccgagcccc gcgtctgctgg gcgggtcctg tggcctgcgt gcgagccgctg 6840
attctcgagc cgagggatct tcggcatacg tctgagcctgg gcggcctggc gcgaggggaga 6900
ctgggaggga tccctcaccata gcgccgctgtgt ttggggcgcgc acaaccccgaga ccagcgtgttt 6960
gataaactc acttggtcggt ccggggacgct catccgggcc gcgtgctgcc gggcggctact 7020
ggatcgcgca aggccagcgcc ccaaggattg ttaaaggtgata tcggcagcgc gcggccgagc 7080
ctttcggcaac gcgcaggccg ccattaactg gcttttcgcc aatgggtgac accgctcctcgg 7140
agcttgtgga caaaccgacc gcggcgaggtg cttggccagct ctatattctgt gcgctgctact 7200
gccagctgtg tattcgaggg caggtggttg gtttgctgct gcggccggcag cagcgccaga 7260
ggcgcgggag gcggcttcgaa cggcccgagc agatgaccgg ccgaggctgca gcggggaaga 7320
ccatgcaaga gcgcggcttgc gcggcttcctc aggggtgtgc ctcgctgccatt gcgcctacgcg 7380
ctatctcgag ctcgccacac ctggaggattc atgcgagctgc gcgggcacgc gacgctgatt 7440
acacgctcaga tgaactcgtc gcgtactctgt atcaaggtgc gcggctcggc gcggctgtcga 7500
tggccagggt gatgggagta gcggcagccgc ccaggtgtgga tgcgctcgctg gattggtcca 7560
tggcttttca gcctcagcag atcggtgcttg gcgtcggcatgc gttggagctt gcggctgctgc 7620
gctactggcc gcagctgctgt gcggcagggg tggacaccag gcggctcagc gcggccggggc 7680
gggccacgcctg ctcggccgctc gcggctcccgt ggtgcagctt cgtgcagctcgt ctgaacaccc 7740
attcgcctgc agcgctgtgc gcggctggggg atctgcgtccgg ctcggcctgcc gtcggatttg 7800
cacccgcgca gcgggttgtat ctgagctgga gcggagagtt gcgtgagcctg gcgggaaaga 7860
CL2230 PCT Seq Lst.
catgggatac ccgcgacggc aggacgctgct gcggagaagct ggcgcgtggttt atttccggc g 7920
cgaagcaggc gatggctccc gggagggcag gctggccgac gcggatccg cacctctggc 7980
agccggccgct ctagggcgtct tcggcgtaac ggctgagctag acctggcagtac accggctg 8040
cgggtgctgc tagataaac gcagagagac gcacccctgct gcggccggcag cgcgtgatg 8100
cagccgcgtgt ggcctgtaga gcggcgcgctg atagccgagc gggagcggcag 8160
ccagcgctgg ttgactaaac cagctgtaac gtaaaagtag atacacggctg agccggctcat 8220
tccggcgcca atccactgaa gcggccagta cccctcggtg cccggtgaac ttcagcgcaat 8280
ggccagtgac gcgaagacca ccgcatagag atggttacgc tcaacgcgct cttgcgcggg 8340
ggtatggtgc gataagtgcc agcccccttcc ccagccggtgc atagatgtact gtgtgctgaa 8400
cgttgccacc ccttccatga tgtgataagt cagtagcagc atccccggtat tccacaacgc 8460
aagcataggt tcctctgtga ggttacgcct gctaaagcgt agcttggaat gcagaagaaac 8520
ataaggctaa cctgggggat aatgcgccttt tcaggggctaa aagcattttat gacaattatt 8580
catcgccgca cttcagccag gtcgcgcgcct gttccacggcg cgccagagcg cgcatccgct 8640
gataacgcgg cccgctttct gcgcgctcct gcggcgtgtg cgccggctgca agcgaatataa 8700
acttaacttcc aaaaaagccgg cccgaaaggg gccgggcttct tattctgctc atatactcga 8760
tctaaacctga attatgtggc cagctatccg cttccctttgtt ggtagaaccg ggc 8814

<210> 42
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 42
gaattcacta gtagagacgc gcggatatc tcat 34

<210> 43
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 43
gaattctagc gcgagcgtc gcta 25

<210> 44
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 44
gccaattgaa ggaggaataa accatg 26
Artificial Sequence

Primer

DNA

Primer

DNA

Primer

DNA

Primer

DNA

Primer
CL2230 PCT Seq Lst.

<400>  50
acgcgtggtg agggacaaga ttgtgg  26

<210>  51
<211>  27
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400>  51
gcatgccttc taaagtccag gccttta  27

<210>  52
<211>  28
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400>  52
ccatggccaa caccacacac atcatcat  28

<210>  53
<211>  48
<212> DNA
<213> artificial sequence

<220>
<223> primer

<400>  53
agtaacctccgc atatgtcagg ctttggccttt ggctttttttc agccaggc  48