



- (51) **International Patent Classification:**  
C12N 9/02 (2006.01) C12P 23/00 (2006.01)
- (21) **International Application Number:**  
PCT/IB2013/056583
- (22) **International Filing Date:**  
27 August 2013 (27.08.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
12198352.2 20 December 2012 (20.12.2012) EP  
61/830,273 3 June 2013 (03.06.2013) US
- (71) **Applicant:** DSM IP ASSETS B.V. [—/NL]; Het Overloot 1, NL-6411 The Heerlen (NL).
- (72) **Inventors:** FARRELL, Christopher; 6134 Lexington Ridge Drive, Lexington, 02421 (US). MAYORGA, Maria; 22 Bradford Avenue, Medford, 02155 (US). CHEVREUX, Bastien; DSM Nutritional Products, c/o Patent Department, P.O. Box 2676, CH-4002 Basel (CH).
- (74) **Agent:** SCHWANDER, Kuno Josef; DSM Nutritional Products Ltd, Wurmisweg 576, 241/636, CH-4303 Kaiseraugst (CH).

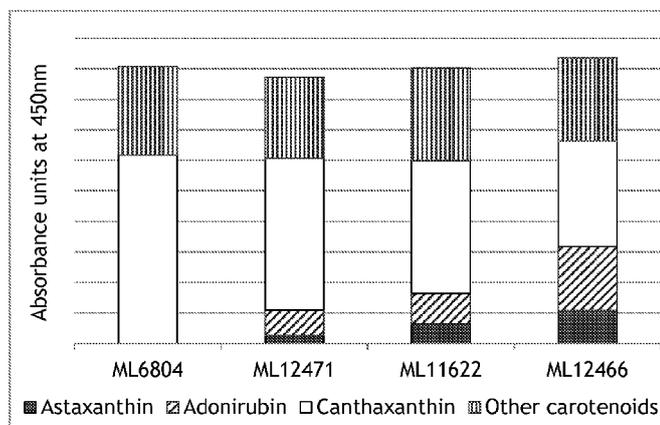
(81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**Published:**  
— with international search report (Art. 21(3))

(54) **Title:** CAROTENE HYDROXYLASE AND ITS USE FOR PRODUCING CAROTENOIDS

Fig. 1



(57) **Abstract:** The present invention relates to a novel hydroxylase, nucleic acid sequence coding therefore, expression constructs and vectors comprising this sequence, microorganisms transformed therewith, processes for the microbiological hydroxylation of isoprenoids, for example processes for converting beta-carotene into beta-cryptoxanthin and zeaxanthin, canthaxanthin into astaxanthin.

WO 2014/096990 A1

## CAROTENE HYDROXYLASE AND ITS USE FOR PRODUCING CAROTENOIDS

### Background of the Invention

5 The present invention relates to a novel hydroxylase, nucleic acid sequences coding therefore, expression constructs and vectors comprising this sequence, microorganisms transformed therewith, processes for the microbiological hydroxylation of isoprenoids, for example processes for converting beta-carotene into zeaxanthin or beta-cryptoxanthin and canthaxanthin into astaxanthin.

10 Carotenoids are organic pigments ranging in color from yellow to red that are naturally produced by certain organisms, including photosynthetic organisms (e.g., plants, algae, cyanobacteria), and some fungi.

Carotenoids such as lutein, zeaxanthin or astaxanthin are important additives in the human and livestock diet as pigmenting substances and precursors of vitamin A derivatives. In  
15 addition, carotenoids have a health-promoting action such as enhancing the immune response and, by reason of their antioxidant properties, a cancer-preventing action, which makes their use as nutraceuticals of interest. An economic process for preparing carotenoids and foodstuffs with an increased carotenoid content is therefore of great importance. Particularly economic processes for preparing carotenoids are biotechnological processes  
20 which make use of proteins and biosynthesis genes of carotenoid biosynthesis from carotenoid-producing organisms.

Prokaryotic [beta]-carotene hydroxylases which catalyze the enzymatic conversion of [beta]-carotene into zeaxanthin via [beta]-cryptoxanthin, and the genes which encode these proteins are known from the bacteria *Erwinia uredovora* (Misawa et al., *J. Bacteriol.* 1990,  
25 6704-6712; EP 393690 B1), *Erwinia herbicola* (WO 9113078), *Agrobacterium aurantiacum* (Misawa et al., *J. Bacteriol.* 1995, 6575-6584; EP 735 137 A1), *Alcaligenes* sp. PC-1 (EP

735 137 A1), *Flavobacterium* sp. strain R1534 (Pasamontes et al., *Gene*. 1997, 185: 35-41; EP 747483 A2) and from the cyanobacterium *Synechocystis* sp. PCC6803 (Masamoto et al., *Plant Cell Physiol*. 1998, 39(5): 560-564).

It is also known that the prokaryotic [beta]-carotene hydroxylases from *Agrobacterium*  
5 *aurantiacum*, *Alcaligenes* and *Erwinia uredovora* are additionally able to convert  
canthaxanthin via adonirubin in astaxanthin (Misawa et al., *J. Bacteriol*. 1995, 6575-6584;  
Fraser et al., *J. Biol. Chem*. 1997, 272: 6128-6135).

From eukaryotic sources, three plant [beta]-carotene hydroxylases are known to catalyze the  
enzymatic conversion of [beta]-carotene into zeaxanthin via [beta]-cryptoxanthin. The  
10 corresponding cDNAs have been isolated from *Arabidopsis thaliana* (Cunningham et al, *J.*  
*Biol. Chem*. 1996, 271: 24349-24352, WO 9736998), and from *Capsicum annuum* L.  
(Bouvier et al., *Biochim. Biophys. Acta*. 1998, 1391: 320-328).

Genes of eukaryotic origin have the advantage over prokaryotic genes that they are  
expressed better in higher transgenic organisms such as plants. Nevertheless, there is still a  
15 need to improve and increase the carotenoid productivity for an economic process for  
preparing carotenoid derivatives or foodstuffs with an increased carotenoid content by  
incorporating eukaryotic nucleic acids into organisms.

In addition, the appropriate eukaryotic [beta]-carotene hydroxylases in the prior art have the  
disadvantage that they have only a narrow substrate range so that there is a build-up of  
20 metabolic products which cannot be converted by the hydroxylases and may exert an  
inhibiting effect on the hydroxylases.

### Summary of the Invention

It is an object of the present invention to remedy the described deficiencies of the prior art  
and to provide a eukaryotic carotene hydroxylase with improved properties.

25 The inventors have found that the above object is surprisingly achieved by a protein which  
has an enzymatic activity for hydroxylating isoprenoids, for example [beta]-carotene or  
canthaxanthin, comprising the amino acid sequence SEQ ID NO:2.

Therefore, the present invention is related to a protein or polypeptide, comprising the amino  
acid sequence SEQ ID NO:2 or a sequence derived from this sequence by substitution,

insertion or deletion of amino acids and having a homology of at least 50% at the amino acid level with the sequence SEQ ID NO:2.

In particular the hydroxylase according to the invention if expressed in a strain selected or able to produce a specific carotenoid as for example zeaxanthin or astaxanthin increases the level of these carotenoids compared to strains not transformed with the gene encoding the protein or polypeptide according to the invention.

The present invention also relates to the isolated polynucleotide encoding the polypeptide of the present invention, nucleic acid constructs, recombinant expression vectors, and recombinant host cells comprising the polynucleotide, and to methods of producing the polypeptide.

The present invention also provides improved systems for the biological production of carotenoids. In one preferred example, the invention provides oleaginous fungi (including, for example, yeast) that produce one or more carotenoids. The present invention also provides methods of constructing such yeast and fungi, methods of using such yeast and fungi to produce carotenoids, and methods of preparing carotenoid-containing compositions, such as food or feed additives, or nutritional supplements, using carotenoids produced in such oleaginous yeast or fungi. In particular, the present invention provides systems and methods for generating yeast and fungi containing polynucleotides encoding the polypeptides of the present invention.

#### Overview of Sequence Listing

SEQ ID NO:1 is the non-optimized DNA sequence encoding the low frequency carotene hydroxylase from *Haematococcus pluvialis*.

SEQ ID NO:2 is the amino acid sequence as deduced from SEQ ID NO:1.

SEQ ID NO:3 is the DNA sequence encoding the low frequency carotene hydroxylase from *Haematococcus pluvialis* as optimized for expression in *Yarrowia lipolytica*.

SEQ ID NO:4 is the DNA sequence encoding the carotene hydroxylase from *Cronobacter pulveris* (formerly *Enterobacter pulveris*) (*Ep*) as optimized for expression in *Yarrowia lipolytica*.

SEQ ID NO:5 is the DNA sequence encoding the carotene hydroxylase from *Enterobacteriaceae bacterium* DC404 (*Dc*) as optimized for expression in *Yarrowia lipolytica*.

### Definitions

**Isolated polypeptide:** The term “isolated polypeptide” means a polypeptide that is modified by the hand of man relative to that polypeptide as found in nature. In one aspect, the polypeptide is at least 1% pure, *e.g.*, at least 5% pure, at least 10% pure, at least 20% pure, at least 40% pure, at least 60% pure, at least 80% pure, and at least 90% pure, as determined by SDS-PAGE.

**Substantially pure polypeptide:** The term “substantially pure polypeptide” means a preparation that contains at most 10%, at most 8%, at most 6%, at most 5%, at most 4%, at most 3%, at most 2%, at most 1%, and at most 0.5% by weight of other polypeptide material with which it is natively or recombinantly associated. Preferably, the polypeptide is at least 92% pure, *e.g.*, at least 94% pure, at least 95% pure, at least 96% pure, at least 97% pure, at least 98% pure, at least 99% pure, at least 99.5% pure, and 100% pure by weight of the total polypeptide material present in the preparation. The polypeptides of the present invention are preferably in a substantially pure form. This can be accomplished, for example, by preparing the polypeptide by well known recombinant methods or by classical purification methods.

**Sequence Identity:** The relatedness between two amino acid sequences or between two nucleotide sequences is described by the parameter “sequence identity”.

For purposes of the present invention, the degree of sequence identity between two amino acid sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, *J. Mol. Biol.* 1970, 48: 443-453) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice et al., *Trends Genet.* 2000, 16: 276-277), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labeled “longest identity” (obtained using the `-nobrief` option) is used as the percent identity and is calculated as follows:

$$\text{Percent Identity} = (\text{Identical Residues} \times 100) / (\text{Length of Alignment} - \text{Total Number of Gaps in Alignment})$$

For purposes of the present invention, the degree of sequence identity between two deoxyribonucleotide sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *supra*) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice  
5 *et al.*, 2000, *supra*), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EDNAFULL (EMBOSS version of NCBI NUC4.4) substitution matrix. The output of Needle labeled "longest identity" (obtained using the `-nobrief` option) is used as the percent identity and is calculated as follows:

(Identical Deoxyribonucleotides x 100)/(Length of Alignment – Total Number of Gaps in  
10 Alignment)

**Fragment:** The term "fragment" means a polypeptide having one or more (several) amino acids deleted from the amino and/or carboxyl terminus of a mature polypeptide; wherein the fragment has hydroxylase activity.

**Allelic variant:** The term "allelic variant" means any of two or more alternative forms of a  
15 gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

**Isolated polynucleotide:** The term "isolated polynucleotide" means a polynucleotide that is  
20 modified by the hand of man relative to that polynucleotide as found in nature. In one aspect, the isolated polynucleotide is at least 1% pure, *e.g.*, at least 5% pure, at least 10% pure, at least 20% pure, at least 40% pure, at least 60% pure, at least 80% pure, at least 90% pure, and at least 95% pure, as determined by agarose electrophoresis. The polynucleotides may  
25 be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

**Substantially pure polynucleotide:** The term "substantially pure polynucleotide" means a polynucleotide preparation free of other extraneous or unwanted nucleotides and in a form suitable for use within genetically engineered polypeptide production systems. Thus, a  
30 substantially pure polynucleotide contains at most 10%, at most 8%, at most 6%, at most 5%, at most 4%, at most 3%, at most 2%, at most 1%, and at most 0.5% by weight of other polynucleotide material with which it is natively or recombinantly associated. A substantially

pure polynucleotide may, however, include naturally occurring 5' and 3' untranslated regions, such as promoters and terminators. Preferably, the polynucleotide is at least 90% pure, *e.g.*, at least 92% pure, at least 94% pure, at least 95% pure, at least 96% pure, at least 97% pure, at least 98% pure, at least 99% pure, and at least 99.5% pure by weight. The  
5 polynucleotides of the present invention are preferably in a substantially pure form.

**Coding sequence:** The term "coding sequence" means a polynucleotide, which directly specifies the amino acid sequence of a polypeptide. The boundaries of the coding sequence are generally determined by an open reading frame, which usually begins with the ATG start codon or alternative start codons such as GTG and TTG and ends with a stop codon such as  
10 TAA, TAG, and TGA. The coding sequence may be a DNA, cDNA, synthetic, or recombinant polynucleotide.

**cDNA:** The term "cDNA" means a DNA molecule that can be prepared by reverse transcription from a mature, spliced, mRNA molecule obtained from a eukaryotic cell. cDNA lacks intron sequences that may be present in the corresponding genomic DNA. The initial,  
15 primary RNA transcript is a precursor to mRNA that is processed through a series of steps, including splicing, before appearing as mature spliced mRNA.

**Nucleic acid construct:** The term "nucleic acid construct" means a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or is modified to contain segments of nucleic acids in a manner that would not otherwise exist in  
20 nature or which is synthetic. The term nucleic acid construct is synonymous with the term "expression cassette" when the nucleic acid construct contains the control sequences required for expression of a coding sequence of the present invention.

**Control sequences:** The term "control sequences" means all components necessary for the  
25 expression of a polynucleotide encoding a polypeptide of the present invention. Each control sequence may be native or foreign to the polynucleotide encoding the polypeptide or native or foreign to each other. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and  
30 transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the polynucleotide encoding a polypeptide.

**Operably linked:** The term “operably linked” means a configuration in which a control sequence is placed at an appropriate position relative to the coding sequence of a polynucleotide such that the control sequence directs the expression of the coding sequence.

5

**Expression:** The term “expression” includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

10 **Expression vector:** The term “expression vector” means a linear or circular DNA molecule that comprises a polynucleotide encoding a polypeptide and is operably linked to additional nucleotides that provide for its expression.

15 **Host cell:** The term “host cell” means any cell type that is susceptible to transformation, transfection, transduction, and the like with a nucleic acid construct or expression vector comprising a polynucleotide of the present invention. The term “host cell” encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication.

20 **Variant:** The term “variant” means a polypeptide having hydroxylase activity comprising an alteration, *i.e.*, a substitution, insertion, and/or deletion of one or more (several) amino acid residues at one or more (several) positions. A substitution means a replacement of an amino acid occupying a position with a different amino acid; a deletion means removal of an amino acid occupying a position; and an insertion means adding 1-3 amino acids adjacent to an amino acid occupying a position.

25

### **Detailed Description of the Invention**

Carotene hydroxylase hereinafter means a protein or polypeptide according to the invention, *i.e.* a protein which has for example an enzymatic activity for converting [beta]-carotene into zeaxanthin or canthaxanthin into astaxanthin, comprising the amino acid sequence SEQ ID NO:2 or a sequence derived from this sequence by substitution, insertion or deletion of  
30 amino acids and having a homology of at least 50% at the amino acid level with the sequence SEQ ID NO:2.

The amino acid sequence depicted in SEQ ID NO:2 is derived from translation of the cDNA sequence depicted in SEQ ID NO:1. The optimized sequence of SEQ ID NO: 1 is shown in SEQ ID NO: 3..

Substitution means replacement of one or more amino acids by one or more amino acids.

- 5 The replacements are preferably those called conservative, in which the replaced amino acid has a similar property to the original amino acid, for example replacement of Glu by Asp, Gln by Asn, Val by Ile, Leu by Ile, Ser by Thr.

- 10 Deletion is the replacement of an amino acid by a direct linkage. Preferred positions for deletions are the termini of the polypeptide and the linkages between the individual protein domains.

Insertions are introductions of amino acids into the polypeptide chain, there formally being replacement of a direct linkage by one or more amino acids.

- 15 The homology between two proteins means identity of the amino acids over the entire length of each protein, which is calculated by comparison with the aid of the computer program GAP (UWGCG, University of Wisconsin, Genetic Computer Group, program algorithm of Needleman and Wunsch, *J. Mol. Biol.* 1970, 48: 443-453), setting the following parameters:

Gap Weight:	12
Length Weight:	4
Average Match:	2.912
20 Average Mismatch:	-2.003

A protein which has a homology of at least 50% at the amino acid level with the sequence SEQ ID NO:2 means a protein which, in comparison of its sequence with the sequence SEQ ID NO:2 using the above program algorithm with the above set of parameters, has an identity of at least 50%, preferably 60%, particularly preferably 70%.

- 25 The proteins according to the invention are able to catalyze the conversion of a [beta]-ionone structural element into a 3-hydroxy-[beta]-ionone structural element, such as the conversion of [beta]-carotene into zeaxanthin, [beta]-carotene into [beta]-cryptoxanthin, [beta]-cryptoxanthin into zeaxanthin, echinenone into 3'-hydroxyechinenone, 3-hydroxyechinenone into adonixanthin (4-ketozeaxanthin), [alpha]-carotene into [alpha]-cryptoxanthin or other  
30 chemical compounds which have up to 40 C atoms and contain a [beta]-ionone ring into the

corresponding 3-hydroxy-[beta]-ionone compounds or the conversion of a 4-keto-[beta]-ionone structural element into a 3-hydroxy-4-keto-[beta]-ionone structural element, such as the conversion of canthaxanthin into astaxanthin, canthaxanthin into phoenicoxanthin (adonirubin), phoenicoxanthin (adonirubin) into astaxanthin, echinenone into 3-  
5 hydroxyechinenone, 3'-hydroxyechinenone into adonixanthin (4-ketozeaxanthin) or other chemical compounds which have up to 40 C atoms and contain a 4-keto-[beta]-ionone ring into the corresponding 3-hydroxy-4-keto-[beta]-ionone compounds.

The carotene hydroxylase can be prepared, as described hereinafter, by gene expression of the appropriate nucleic acids which encode these proteins from natural or genetically  
10 manipulated organisms.

The invention further relates to a process for the preparation of carotenoids and carotenoid derivatives, which comprises converting a [beta]-ionone structural element into a 3-hydroxy-[beta]-ionone structural element and/or a 4-keto-[beta]-ionone structural element into a 3-hydroxy-4-keto-[beta]-ionone structural element in the presence of the protein according to  
15 the invention.

Carotenoids and carotenoid derivatives are for example zeaxanthin, [beta]-cryptoxanthin, 3'-hydroxyechinenone, 3-hydroxyechinenone, adonixanthin (4-ketozeaxanthin), astaxanthin, phoenicoxanthin (adonirubin), [alpha]-cryptoxanthin, or lutein or derivatives thereof having up to 40 C atoms and containing at least one 3-hydroxy-[beta]-ionone or at least one 3-hydroxy-  
20 4-keto-[beta]-ionone structural element in the molecule, such as, for example, 3-hydroxy-6-vinyl-[beta]-ionone, 3-hydroxy-4-keto-6-vinyl-[beta]-ionone, 3-hydroxyretinol, 3-hydroxy-4-ketoretinol, 3-hydroxyretinal, 3-hydroxy-4-ketoretinal, 3-hydroxyretinoic acid or 3-hydroxy-4-ketoretinoic acid.

In the process according to the invention there is conversion in the presence of the proteins  
25 according to the invention of a [beta]-ionone structural element into a 3-hydroxy-[beta]-ionone structural element, such as [beta]-carotene into zeaxanthin, via  $\beta$ -cryptoxanthin, [beta]-carotene into [beta]-cryptoxanthin, [beta]-cryptoxanthin into zeaxanthin, echinenone into 3'-hydroxyechinenone or 3-hydroxyechinenone, 3-hydroxyechinenone or 3'-hydroxyechinenone into adonixanthin (4-ketozeaxanthin), [alpha]-carotene into zeinoxanthin  
30 or [alpha]-cryptoxanthin or a chemical compound having up to 40 C atoms and containing a [beta]-ionone ring into the corresponding 3-hydroxy-[beta]-ionone compound or a 4-keto-[beta]-ionone structural element into a 3-hydroxy-4-keto-[beta]-ionone structural element,

such as canthaxanthin into astaxanthin, via adonirubin, canthaxanthin into phoenicoxanthin (adonirubin), phoenicoxanthin (adonirubin) into astaxanthin, echinenone into 3-hydroxyechinenone, 3'-hydroxyechinenone into adonixanthin (4-ketozeaxanthin) or a chemical compound having up to 40 C atoms and containing a 4-keto-[beta]-ionone ring into  
5 the corresponding 3-hydroxy-4-keto-[beta]-ionone compound.

The invention also relates to nucleic acid sequence coding for the hydroxylase according to the invention. A preferred nucleic acid has the sequence of SEQ ID NO:1 and the sequence of SEQ ID NO: 3.

The invention moreover relates to functional analogs of the nucleic acid according to SEQ ID  
10 NO:1 and SEQ ID NO: 3, obtained by addition, substitution, insertion and/or deletion of individual or multiple nucleotides, which furthermore code for a hydroxylase having the desired specificity.

The invention also encompasses those nucleic acid sequences which comprise so-called silent mutations or which are modified in comparison with a specifically mentioned sequence  
15 in accordance with the codon usage of a specific origin or host organism, and naturally occurring variants of such nucleic acid sequences.

The invention also encompasses modifications of the nucleic acid sequences obtained by degeneration of the genetic code (i.e. without any changes in the corresponding amino acid sequence) or conservative nucleotide substitution (i.e. the corresponding amino acid is  
20 replaced by another amino acid of the same charge, size, polarity and/or solubility), and sequences modified by nucleotide addition, insertion, inversion or deletion, which sequences encode a hydroxylase according to the invention having a "modified substrate profile", and the corresponding complementary sequences.

The invention furthermore relates to expression constructs comprising a nucleic acid  
25 sequence according to the invention under the genetic control of regulatory nucleic acid sequences; and vectors comprising at least one of these expression constructs.

Preferably, the constructs according to the invention encompass a promoter 5'-upstream of the encoding sequence in question and a terminator sequence 3'-downstream, and, optionally, further customary regulatory elements, and, in each case operatively linked with  
30 the encoding sequence. Operative linkage is to be understood as meaning the sequential

arrangement of promoter, encoding sequence, terminator and, if appropriate, other regulatory elements in such a manner that each of the regulatory elements can fulfill its intended function on expression of the encoding sequence. Examples of operatively linkable sequences are targeting sequences, or else translation enhancers, enhancers,  
5 polyadenylation signals and the like. Further regulatory elements encompass selectable markers, amplification signals, replication origins and the like.

In addition to the artificial regulatory sequences, the natural regulatory sequence can still be present upstream of the actual structural gene. If desired, this natural regulation may be switched off by genetic modification, and the expression of the genes may be enhanced or  
10 lowered. However, the gene construct may also be simpler in construction, i.e. no additional regulatory signals are inserted upstream of the structural gene and the natural promoter with its regulation is not removed. Instead, the natural regulatory sequence is mutated in such a way that regulation no longer takes place and the gene expression is increased or reduced. One or more copies of the nucleic acid sequences may be present in the gene construct.

15 Examples of suitable promoters are: cos, tac, trp, tet, trp-tet, lpp, lac, lpp-lac, lacIq, T7, T5, T3, gal, trc, ara, SP6, I-PR or I-PL promoter, which are advantageously employed in Gram-negative bacteria; and Gram-positive promoters amy and SPO2, the yeast promoters ADC1, MFa, Ac, P-60, CYC1, GAPDH, TEF1 or the plant promoters CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or the ubiquitin or phaseolin promoter. Particular preference is given to  
20 using inducible promoters, for example light- and in particular temperature-inducible promoters, such as the PrP1 promoter.

In principle, all natural promoters with their regulatory sequences can be used. In addition, synthetic promoters may also be used in an advantageous fashion.

The above mentioned regulatory sequences are intended to allow the targeted expression of  
25 the nucleic acid sequences and of protein expression. Depending on the host organism, this may mean, for example, that the gene is expressed or over expressed only after induction has taken place, or that it is expressed and/or over expressed immediately and/or constitutively.

The regulatory sequences or factors can preferably have a positive effect on expression and  
30 in this manner increase or reduce the latter. Thus, an enhancement of the regulatory elements may advantageously take place at the transcriptional level by using strong

transcription signals such as promoters and/or "enhancers". In addition, translation may also be enhanced by improving, for example, mRNA stability.

An expression cassette is generated by fusing a suitable promoter with a suitable hydroxylase nucleotide sequence and a terminator signal or polyadenylation signal. To this  
5 end, customary recombination and cloning techniques are used as they are described, for example, in T. Maniatis, E. F. Fritsch and J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989) and in T. J. Silhavy, M. L. Berman and L. W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) and in Ausubel, F. M. et al., *Current Protocols in*  
10 *Molecular Biology*, Greene Publishing Assoc. and Wiley Interscience (1987).

For expression in a suitable host organism, the recombinant nucleic acid construct or gene construct is advantageously inserted into a host-specific vector which allows optimal gene expression in the host. Vectors are well known to the skilled worker and can be found, for example, in "Cloning Vectors" (Pouwels P. H. et al., Ed., Elsevier, Amsterdam-New York-  
15 Oxford, 1985). Vectors are to be understood as meaning not only plasmids, but all other vectors known to the skilled worker such as, for example, phages, viruses, such as SV40, CMV, baculovirus and adenovirus, transposons, IS elements, plasmids, cosmids, and linear or circular DNA. These vectors can be replicated autonomously in the host organism or chromosomally.

20 The vectors according to the invention allow the generation of recombinant microorganisms which are transformed, for example, with at least one vector according to the invention and which can be employed for producing the mutants. The above-described recombinant constructs according to the invention are advantageously introduced into a suitable host organism and expressed. It is preferred to use usual cloning and transfection methods known  
25 to the skilled worker in order to bring about expression of the abovementioned nucleic acids in the expression system in question. Suitable systems are described, for example, in current protocols in molecular biology, F. Ausubel et al., eds., Wiley Interscience, New York 1997.

Suitable host organisms are, in principle, all organisms which allow expression of the nucleic acids according to the invention, their allelic variants, and their functional equivalents or  
30 derivatives. Preferred initial organisms are those naturally able to synthesize carotenoids. However, initial organisms able to synthesize carotenoids because of the introduction of carotenoid biosynthesis genes are also suitable. Initial organisms mean prokaryotic or

eukaryotic organisms such as, for example, microorganisms or plants. Preferred microorganisms are bacteria, yeasts, algae or fungi.

Therefore, the invention further relates to a process for preparing the genetically modified organisms described below, wherein the carotene hydroxylase gene according to the invention is introduced into the genome of the initial organism. By initial organisms are meant  
5 the organisms before the genetic modification according to the invention.

The carotene hydroxylase gene according to the invention can in principle be introduced by all methods known to the skilled worker into the initial organisms described below, which are genetically modified thereby.

10 They are advantageously introduced into the initial organisms or cells thereof by transformation, transfection, electroporation, using the so-called particle gun, or by microinjection.

The skilled worker can find appropriate methods for microorganisms in the textbooks by Sambrook, J. et al. (1989) *Molecular cloning: A laboratory manual*, Cold Spring Harbor  
15 Laboratory Press, by F. M. Ausubel et al. (1994) *Current protocols in molecular biology*, John Wiley and Sons, by D. M. Glover et al., *DNA Cloning Vol. 1*, (1995), IRL Press (ISBN 019-963476-9), by Kaiser et al. (1994) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory Press or Guthrie et al. *Guide to Yeast Genetics and Molecular Biology*, *Methods in Enzymology*, 1994, Academic Press.

20 Examples of advantageous methods which may be mentioned are those such as the introduction of the DNA by homologous or heterologous recombination, for example using the *URA3* gene, specifically the *URA3* gene from *Ashbya*, as described in the German Application DE 19801120.2, and/or by the REMI method ("restriction enzyme mediated integration") which is described below.

25 The REMI technique is based on the cotransformation of a linear DNA construct which has been cut at both ends with the same restriction endonuclease, together with the restriction endonuclease which was used for this restriction of the DNA construct, into an organism. The restriction endonuclease then cuts the genomic DNA of the organism into which the DNA construct has been introduced together with the restriction enzyme. This leads to an  
30 activation of the cell's own repair mechanisms. These repair mechanisms repair the strand

breaks in the genomic DNA which have been caused by endonuclease, and during this also incorporate with a certain frequency the cotransformed DNA construct into the genome. Ordinarily, the restriction cleavage sites are retained at both ends of the DNA during this.

This technique was described by Bölker et al. (*Mol. Gen. Genet.*, 1995, 248: 547-552) for the  
5 insertion mutagenesis of fungi. The method was used by Von Schiestl and Petes (*Proc. Natl. Acad. Sci. USA*, 1991, 88: 7585-7589) to find out whether there is heterologous recombination in *Saccharomyces*. The method has been described by Brown et al. (*Mol. Gen. Genet.* 1996, 251: 75-80) for the stable transformation and regulated expression of an inducible reporter gene.

10 It is possible using the REMI method to position the nucleic acid fragments according to the invention or the aforementioned carotene hydroxylase gene according to the invention at transcriptionally active sites in the genome.

It is possible and advantageous to clone the nucleic acids together with at least one reporter gene into a DNA construct, which is introduced into the genome. This reporter gene ought to  
15 make detectability easy by a growth, fluorescence, chemo- or bioluminescence assay or by a photometric measurement. Examples which may be mentioned of reporter genes are antibiotic resistance genes, hydrolase genes, fluorescent protein genes, bioluminescence genes, glucosidase genes, the luciferase gene, [beta]-galactosidase gene, gfp gene, lipase gene, esterase gene, peroxidase gene, [beta]-lactamase gene, acetyl-, phospho- or  
20 adenylyltransferase gene. These genes make it possible easily to measure and quantify the transcription activity and thus the expression of the genes. This means that it is possible to identify sites in the genome which have a productivity differing by up to a factor of 2.

If it is intended to introduce a plurality of genes, such as, for example, further genes of carotenoid biosynthesis, into the organism, they can all be introduced together with a  
25 reporter gene in a single vector, or each individual gene with a reporter gene can be introduced in one vector in each case, into the organism, it being possible to introduce the various vectors at the same time or successively. It is also possible to insert gene fragments coding for the respective activities using the REMI techniques.

Restriction enzymes suitable in principle for integrating the carotene hydroxylase gene or  
30 nucleic acid constructs according to the invention into the genome of initial organisms are all known to the skilled worker. Restriction enzymes which recognize only 4 base pairs as

restriction cleavage site are less preferred because they cut too often in the genome or in the vector to be integrated, and preferred enzymes recognize 6, 7, 8 or more base pairs as cleavage site, such as *Bam*HI, *Eco*RI, *Bgl*II, *Sph*I, *Spe*I, *Xba*I, *Xho*I, *Nco*I, *Sal*I, *Cla*I, *Kpn*I, *Hind*III, *Sac*I, *Pst*I, *Bpn*I, *Not*I, *Srf*I or *Sfi*I, to mention only a few of the possible enzymes. It is advantageous if the enzymes used no longer have cleavage sites in the DNA to be introduced; this increases the efficiency of integration. Ordinarily, 5 to 500 U, preferably 10 to 250, particularly preferably 10 to 100 U of the enzymes are used in the REMI mixture. The enzymes are advantageously employed in an aqueous solution which contains substances for osmotic stabilization, such as sugars such as sucrose, trehalose or glucose, polyols such as glycerol or polyethylene glycol, a buffer with an advantageous buffering in the range of pH 5 to 9, preferably 6 to 8, particularly preferably 7 to 8, such as tris, MOPS, HEPES, MES or PIPES and/or substances to stabilize the nucleic acids, such as inorganic or organic salts of Mg, Cu, Co, Fe, Mn or Mo. It is also possible where appropriate for other substances to be present, such as EDTA, EDDA, DTT, [beta]-mercaptoethanol or nuclease inhibitors. However, it is also possible to carry out the REMI technique without these additions.

The process is carried out at a temperature in the range from 5 to 80°C, preferably from 10 to 60°C, particularly preferably from 20 to 40°C. Other known methods for destabilizing cell membranes are suitable for the process, such as, for example, electroporation, fusion with loaded vesicles or destabilization with various alkali metal or alkaline earth metal salts such as lithium, rubidium or calcium salts, with lithium salts being preferred.

The invention further relates to a correspondingly genetically modified organism, with the expression of the carotene hydroxylase gene according to the invention being increased by comparison with a wild type organism in the case where the initial organism contains a carotene hydroxylase gene, or being caused in the case where the initial organism does not contain a carotene hydroxylase gene, by the genetic modification.

A genetically modified organism means an organism in which the carotene hydroxylase gene or nucleic acid construct according to the invention has been inserted, preferably by one of the methods described above.

The genetically modified organism contains at least one carotene hydroxylase gene according to the invention or at least one nucleic acid construct according to the invention. Depending on the initial organism, the nucleic acid may be present inside or outside the chromosome.

Carotenoid metabolism in the genetically modified organisms is preferably altered by comparison with the wild type.

Preferred organisms are recombinant fungi and yeast. In a particular embodiment, the recombinant fungus is oleaginous in that it can accumulate lipid to at least about 20% of its dry cell weight; and produces at least one carotenoid selected from the group consisting of

5 antheraxanthin, adonirubin, adonixanthin, astaxanthin, canthaxanthin, capsorubin,  $\beta$ -cryptoxanthin,  $\alpha$ -carotene,  $\beta,\psi$ -carotene,  $\delta$ -carotene,  $\epsilon$ -carotene, echinenone, 3-hydroxyechinenone, 3'-hydroxyechinenone,  $\gamma$ -carotene, 4-keto- $\gamma$ -carotene,  $\zeta$ -carotene,  $\alpha$ -cryptoxanthin, deoxyflexixanthin, diatoxanthin, 7,8-didehydroastaxanthin, didehydrolycopene,

10 fucoxanthin, fucoxanthinol, isorenieratene,  $\beta$ -isorenieratene, lactucaxanthin, lutein, lycopene, myxobactone, neoxanthin, neurosporene, hydroxyneurosporene, peridinin, phytoene, rhodopin, rhodopin glucoside, 4-keto-rubixanthin, siphonaxanthin, spheroidene, spheroidenone, spirilloxanthin, torulene, 4-keto-torulene, 3-hydroxy-4-keto-torulene, uriolide, uriolide acetate, violaxanthin, zeaxanthin- $\beta$ -diglucoside, zeaxanthin, a C30 carotenoid, and

15 combinations thereof, and can accumulate the produced carotenoid to at least about 1% of its dry cell weight. Preferably, the recombinant fungus is a member of a genus selected from the group consisting of: *Aspergillus*, *Blakeslea*, *Botrytis*, *Candida*, *Cercospora*, *Cryptococcus*, *Cunninghamella*, *Fusarium* (*Gibberella*), *Kluyveromyces*, *Lipomyces*, *Mortierella*, *Mucor*, *Neurospora*, *Penicillium*, *Phycomyces*, *Pichia* (*Hansenula*), *Puccinia*, *Pythium*,

20 *Rhodospiridium*, *Rhodotorula*, *Saccharomyces*, *Sclerotium*, *Trichoderma*, *Trichosporon*, *Xanthophyllomyces* (*Phaffia*), and *Yarrowia*, or is of a species selected from the group consisting of: *Aspergillus terreus*, *Aspergillus nidulans*, *Aspergillus niger*, *Blakeslea trispora*, *Botrytis cinerea*, *Candida japonica*, *Candida pulcherrima*, *Candida revkaufi*, *Candida tropicalis*, *Candida utilis*, *Cercospora nicotianae*, *Cryptococcus curvatus*, *Cunninghamella echinulata*, *Cunninghamella elegans*, *Fusarium fujikuroi* (*Gibberella zeae*), *Kluyveromyces lactis*, *Lipomyces starkeyi*, *Lipomyces lipoferus*, *Mortierella alpina*, *Mortierella ramanniana*, *Mortierella isabellina*, *Mortierella vinacea*, *Mucor circinelloides*, *Neurospora crassa*,

25 *Phycomyces blakesleanus*, *Pichia pastoris*, *Puccinia distincta*, *Pythium irregulare*, *Rhodospiridium toruloides*, *Rhodotorula glutinis*, *Rhodotorula graminis*, *Rhodotorula mucilaginoso*, *Rhodotorula pinicola*, *Rhodotorula gracilis*, *Saccharomyces cerevisiae*, *Sclerotium rolfii*, *Trichoderma reesei*, *Trichosporon cutaneum*, *Trichosporon pullulans*,

30 *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*), and *Yarrowia lipolytica*.

Of these naturally oleaginous strains, some also naturally produce carotenoids and some do not; these strains may be additionally utilized as a host cell by introduction of carotenoid biosynthesis genes as disclosed in US Patent 7 851 199.

In other embodiments, the present invention provides a method of producing a carotenoid, the method comprising steps of cultivating a fungus under conditions that allow production of the carotenoid; and isolating the produced carotenoid.

Cultivation of the genetically modified organism according to the invention takes place in a manner known per se, such as cultivation of the appropriate wild type, for example in the case of microorganisms in a suitable medium such as, for example, on agar plates or in suspension culture, or in the case of plants in soil or appropriately suitable nutrient media. By harvesting is meant in the case of microorganisms the isolation of the microorganisms, and in the case of plants the cutting off of the plant or, where appropriate, particular plant parts containing the carotenoids. The carotenoids are isolated in a manner known per se, for example by disruption of the organism cells, extraction of the carotenoids and subsequent purification of the carotenoids by chemical or physical separation methods such as extraction or chromatography.

The following examples illustrate the invention.

### Examples

Table 1 below describes certain *Yarrowia lipolytica* strains used in the following exemplification:

5

Table 1: *Yarrowia lipolytica* strains.

ML5252	<i>MATA erg9-4789::ura3 tef1P-{HMG-tr GGS carB carRP}</i> prototrophic	Classical and standard molecular genetic techniques
ML6804	<i>MATB erg9-4789::ura3 tef1P-{HMG-tr GGS carB carRP crtW}</i> prototrophic	Classical and standard molecular genetic techniques
ML9335	<i>MATA erg9-4789::URA3 tef1P-{HMG-tr GGS carB carRP crtW Dc-crtZ}</i> prototrophic	Classical and standard molecular genetic techniques
ML9863	<i>MATB erg9-4789::URA3 tef1P-{HMG-tr GGS carB carRP crtW Xa-crtZ Dc-crtZ}</i> prototrophic	Classical and standard molecular genetic techniques
ML11218	<i>ML9863 crtW-Δ6180</i>	Targetted disruption with Hyg <sup>R</sup> cassette ; subsequent marker removal using <i>cre-lox</i> system
ML11453	<i>ML9335 tef1P-carRP/Hyg<sup>R</sup></i>	Untargeted transformation; additional unmarked copies of pre-existing activities may also have been incorporated.
ML11584	<i>ML9863 tef1P-Ep-crtZ/Nat<sup>R</sup></i>	Untargeted transformation; additional unmarked copies of pre-existing activities may also have been incorporated.
ML11956	<i>MATB erg9-4789::ura3 tef1P-{HMG-tr GGS carB carRP crtW Dc-crtZ}</i>	ML11453 x ML11584
ML12526	<i>ML9335 with tef1P-HMG-tr tef1P-carB 3X-tef1P-carRP</i>	Untargetted transformations followed by removal of Hyg <sup>R</sup> and Nat <sup>R</sup> using <i>cre-lox</i> system

*Yarrowia* strains ML5252, ML6804, ML9863, and ML9335 were constructed by the introduction of heterologous genes under the control of the endogenous *TEF1* promoter,

coupled with several generations of crossbreeding, starting with ML350 and ATCC201249 as described in US Patent 7,851,199. The GGS gene and the truncated *HMG* gene ("*HMG-tr*") were derived from *Yarrowia* sequences corresponding to native geranylgeranyl pyrophosphate synthase and hydroxymethylglutaryl-CoA reductase genes, respectively. The *carRP* and *carB* genes were derived from *Mucor circinelloides*, and they encode a bifunctional phytoene synthase/lycopene cyclase and a phytoene dehydrogenase, respectively. The *crtW* gene was synthesized to encode the carotene ketolase of *Parvularcula bermudensis* (US 2012/0156718).

The *crtZ* gene was amplified from *Xanthobacter autotrophicus* (*Xa*) or synthesized to encode the carotene hydroxylase of *Cronobacter pulveris* (formerly *Enterobacter pulveris*) (*Ep*) (SEQ ID NO:4), *Enterobacteriaceae* bacterium DC404 (*Dc*) (SEQ ID NO:5), or *Flavobacterium* sp. R1534 (*Fb*) (US Patent 6,087,152). These genes are sometimes but not always associated with auxotrophic markers (*URA3*, *LEU2*, *URA2*, *LYS1*, *ADE1*) or a *loxP* site, remnant of a Hyg<sup>R</sup> (hygromycin resistance) or Nat<sup>R</sup> (nourseothricin resistance) marker.

15 Table 2: Plasmids

Plasmid	Backbone	Insert	Oligos or source
pMB6486	pMB6157 (Hyg <sup>R</sup> <i>tef1P-xprT</i> )	<i>Hp<sup>LF</sup>-crtZ</i>	Synthesized <i>NheI-MluI</i> fragment
pMB6487	pMB6157 (Hyg <sup>R</sup> <i>tef1P-xprT</i> )	<i>Fb-crtZ</i>	Synthesized <i>NheI-MluI</i> fragment
pMB6056	pMB6157 (Hyg <sup>R</sup> <i>tef1P-xprT</i> )	<i>Ep-crtZ</i>	Synthesized <i>NheI-MluI</i> fragment

All basic molecular biology and DNA manipulation procedures described herein are generally performed according to Sambrook *et al.* or Ausubel *et al.* (J. Sambrook, E.F. Fritsch, T. Maniatis (eds). 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press: New York; F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl (eds.). 1998. *Current Protocols in Molecular Biology*. Wiley: New York).

**Example 1a: Production of pMB6486 (Hyg<sup>R</sup> *tef*- *Hp<sup>LF</sup>-crtZ*), encoding *H. pluvialis* LF carotene hydroxylase.**

As a result of an mRNA sequence analysis using *Haematococcus pluvialis* (UTEX No. 2505) inventors surprisingly found the new  $\beta$ -carotene hydroxylase, which was named Hp<sup>LF</sup>-CrtZ (*H. pluvialis* LF (low frequency) hydroxylase).

A codon optimized carotene hydroxylase (Hp<sup>LF</sup>-CrtZ) ORF sequence was synthesized *de novo* based on the deduced protein sequence of the *H. pluvialis* gene, using *Y. lipolytica* codon bias as specified in SEQ ID NO:3. During the *de novo* synthesis the sequence 5'-TGCTAGCCACAAAA, containing an *NheI* restriction site and a typical Kozak sequence for enabling efficient translation, was added immediately upstream of the ATG. The sequence ACGCGT-3', comprising an *MluI* restriction site, was added immediately downstream of the stop codon. This sequence was cleaved using *NheI* and *MluI* and ligated to pMB6157 cut with *NheI* and *MluI* to produce pMB6486. The resulting encoded Hp<sup>LF</sup>-CrtZ protein of pMB6486 is specified in SEQ ID NO:2.

**Example 1b: Production of pMB6056 (Hyg<sup>R</sup> tef-crtZ-Ep), encoding *C. pulveris* (formerly *E. pulveris*) carotene hydroxylase.**

A codon optimized carotene hydroxylase (Ep-CrtZ) ORF sequence was synthesized *de novo* based on the deduced protein sequence of the *Cronobacter pulveris* gene, formerly known as *Enterobacter pulveris*, (EMBL Accession No. CAZ90621.1), using *Y. lipolytica* codon bias as specified in SEQ ID NO:4. During the *de novo* synthesis the sequence 5'-TGCTAGCCACAAAA, containing an *NheI* restriction site and a typical Kozak sequence for enabling efficient translation, was added immediately upstream of the ATG. The sequence ACGCGT-3', comprising an *MluI* restriction site, was added immediately downstream of the stop codon. This sequence was cleaved using *NheI* and *MluI* and ligated into pMB6157 cut with *NheI* and *MluI* to produce pMB6056.

**Example 1c: Production of pMB6487 (Hyg<sup>R</sup> tef-crtZ-Fb), encoding *Flavobacterium sp.* R1534 carotene hydroxylase.**

A codon optimized carotene hydroxylase (Fb-CrtZ) was synthesized *de novo* based on the protein sequence of *Flavobacterium sp.* R1534 as specified in US Patent 6,087,152, using *Y. lipolytica* codon bias. During the *de novo* synthesis the sequence 5'-TGCTAGCCACAAAA, containing an *NheI* restriction site and a typical Kozak sequence for enabling efficient translation, was added immediately upstream of the ATG. The sequence ACGCGT-3', comprising an *MluI* restriction site, was added immediately downstream of the stop codon.

The sequence was cleaved using *NheI* and *MluI* and ligated to pMB6157 cut with *NheI* and *MluI* to produce pMB6487.

**Example 2A: Introduction of three carotene hydroxylase genes into the *Y. lipolytica* canthaxanthin producing strain ML6804 to produce astaxanthin**

To test the hydroxylating potential of *Hp*<sup>LF</sup> *CrtZ* and compare it to *Fb* *CrtZ* and *Ep* *CrtZ*, strain ML6804 was transformed with three different constructs:

- 1) a *HindIII* – *XbaI* fragment of pMB6487 that harbors the *Flavobacterium crtZ* gene under the control of *tef1p* and the hygromycin resistance marker *Hyg*<sup>R</sup>;
- 10 2) a *PvuII* fragment of pMB6056 harboring the *Cronobacter pulveris* (formerly *Enterobacter pulveris*) *crtZ* gene under the control of *tef1p* and the hygromycin resistance marker *Hyg*<sup>R</sup>;
- 3) a *HindIII* – *XbaI* fragment of pMB6486 harboring the *H. pluvialis* “low frequency” *crtZ* gene under the control of *tef1p* and the hygromycin resistance marker *Hyg*<sup>R</sup>.

Transformants were selected on YPD with 100 mg/L hygromycin after 3-4 days of growth at  
15 30°C. Ten transformants from each construct were grown in shake flasks for 4 days in YPD. All transformants produced astaxanthin. A representative transformant from each transformation is shown in Fig. 1 alongside the parent strain ML6804. Strain ML12471, containing the *Flavobacterium* sp. R1534 *crtZ* gene, produced 3% astaxanthin and 9% adonirubin (as a percentage of the total carotenoid). Strain ML11622, containing the  
20 *Cronobacter pulveris* (formerly *Enterobacter pulveris*) *crtZ* gene, produced 7% astaxanthin and 11% adonirubin. And strain ML12466, containing the *H. pluvialis*<sup>LF</sup> *crtZ* gene, produced 12% astaxanthin and 22% adonirubin.

**Example 2B. Introduction of *H. pluvialis* LF carotene hydroxylase into *Y. lipolytica* astaxanthin producing strains to increase astaxanthin purity**

25 Strains ML9863 and ML11956 were transformed with a *HindIII* – *XbaI* fragment of pMB6486 that harbors the *tef1* promoter, the *Hp*<sup>LF</sup> *crtZ* gene and the selectable marker for hygromycin resistance, *Hyg*<sup>R</sup>. Twenty *Hyg*<sup>R</sup> transformants from each strain were chosen that appeared darker than the parents on the transformation plates (YPD with 100 mg/L hygromycin) after 3-4 days of growth at 30°C. Transformants were grown in shake flasks for 4 days in YPD.  
30 Two representative transformants are shown in Fig. 2 compared to their parent strains.

ML12562 is derived from ML9863 and ML12566 from ML11956. As a percentage of total carotenoid, ML12562 and ML12566 produced 3-fold (40% vs. 13%) and 2-fold (27% vs. 13%) more astaxanthin, respectively, than their parents (Fig. 2).

5 **Example 2C. Introduction of *H. pluvialis* LF hydroxylase into a *Y. lipolytica*  $\beta$ -carotene producing strain to produce  $\beta$ -cryptoxanthin:**

Strain ML5252 was transformed with three different constructs:

- 1) a *Hind*III – *Xba*I fragment of pMB6487 that harbors the *Flavobacterium* sp. R1534 *crtZ* gene under the control of *tef1p* and the hygromycin resistance marker Hyg<sup>R</sup>;
- 10 2) a *Pvu*II fragment of pMB6056 harboring the *Cronobacter pulveris* (formerly *Enterobacter pulveris*) *crtZ* gene under the control of *tef1p* and the hygromycin resistance marker Hyg<sup>R</sup>;
- 3) a *Hind*III – *Xba*I fragment of pMB6486 harboring the *H. pluvialis* LF *crtZ* gene under the control of *tef1p* and the hygromycin resistance marker Hyg<sup>R</sup>.

Transformants were selected on YPD with 100 mg/L hygromycin after 3-4 days of growth at  
15 30°C. Ten transformants from each construct were grown in shake flasks for 4 days in YPD. All transformants produced zeaxanthin and  $\beta$ -cryptoxanthin. Representative transformants are shown in Fig. 3 compared to the parent strain ML5252. Strain ML12458, containing the *Flavobacterium crtZ* gene, produced 6%  $\beta$ -cryptoxanthin and 6% zeaxanthin (as a percentage of the total carotenoid). Strain ML10341, containing the *Cronobacter pulveris*  
20 (formerly *Enterobacter pulveris*) *crtZ* gene, produced 3%  $\beta$ -cryptoxanthin and 39% zeaxanthin (as a percentage of total carotenoid). And strain ML12453 containing the *H. pluvialis* LF *crtZ* gene produced 21%  $\beta$ -cryptoxanthin and 4% zeaxanthin (as a percentage of total carotenoid).

25 **Extraction and quantification of carotenoid production by HPLC from *Yarrowia lipolytica* cells**

Shake flask testing and carotenoid analysis of generated strains were performed according to the methods described previously in US patent No. 7,851,199 B2.



### Claims

1. A polypeptide having hydroxylase activity, selected from the group consisting of:
  - (a) a polypeptide of SEQ ID NO:2;
  - (b) a polypeptide derived from SEQ ID NO:2 by substitution, insertion or deletion of amino acids and having a homology of at least 50% at the amino acid level with the sequence of SEQ ID NO:2.for the microbiological hydroxylation of isoprenoids.
2. A polypeptide of claim 1, wherein the protein has an enzymatic activity for converting [beta]-carotene into [beta]-cryptoxanthin and/or zeaxanthin, and/or an enzymatic activity for converting canthaxanthin into andirubin and/or astaxanthin.
3. An isolated nucleic acid encoding the polypeptide of claim 1.
4. The isolated nucleic acid of claim 3, which consists of the sequence depicted in SEQ ID NO:1. and SEQ ID NO: 3.
5. A nucleic acid construct or expression vector comprising the polynucleotide of claim 3 or 4 operably linked to one or more (several) control sequences that direct the production of the polypeptide in an expression host cell.
6. A transformed microorganism (host cell) wherein the nucleic acid of claim 3 or 4 is expressed.
7. The transformed microorganism of claim 6, whose carotenoid metabolism is different from that of a wild type.
8. The transformed microorganism of claim 7, wherein the said microorganism is an oleaginous strain.
9. The transformed microorganism of claim 8, wherein the said oleaginous strain is a strain of *Yarrowia lipolytica*.
10. A process for producing the transformed microorganism of claim 6, which comprises introducing a nucleic acid or nucleic acid construct consisting of the sequence depicted in SEQ ID NO:1 which is functionally linked to one or more regulation signals.

11. A process for the preparation of carotenoids or carotenoid derivatives, which comprises converting a [beta]-ionone into a 3-hydroxy-[beta]-ionone and/or a 4-keto-[beta]-ionone into a 3-hydroxy-4-keto-[beta]-ionone structural element in the presence of the polypeptide of claim 1.
- 5 12. A process of claim 11, comprising the steps of:
- (a) cultivating a transformed microorganism according to any of claims 6 to 9, under conditions conducive for production of carotenoids; and
  - (b) recovering the carotenoid or carotenoid derivative.
13. A composition comprising the polypeptide of any of claims 1 or 2.

Fig. 1

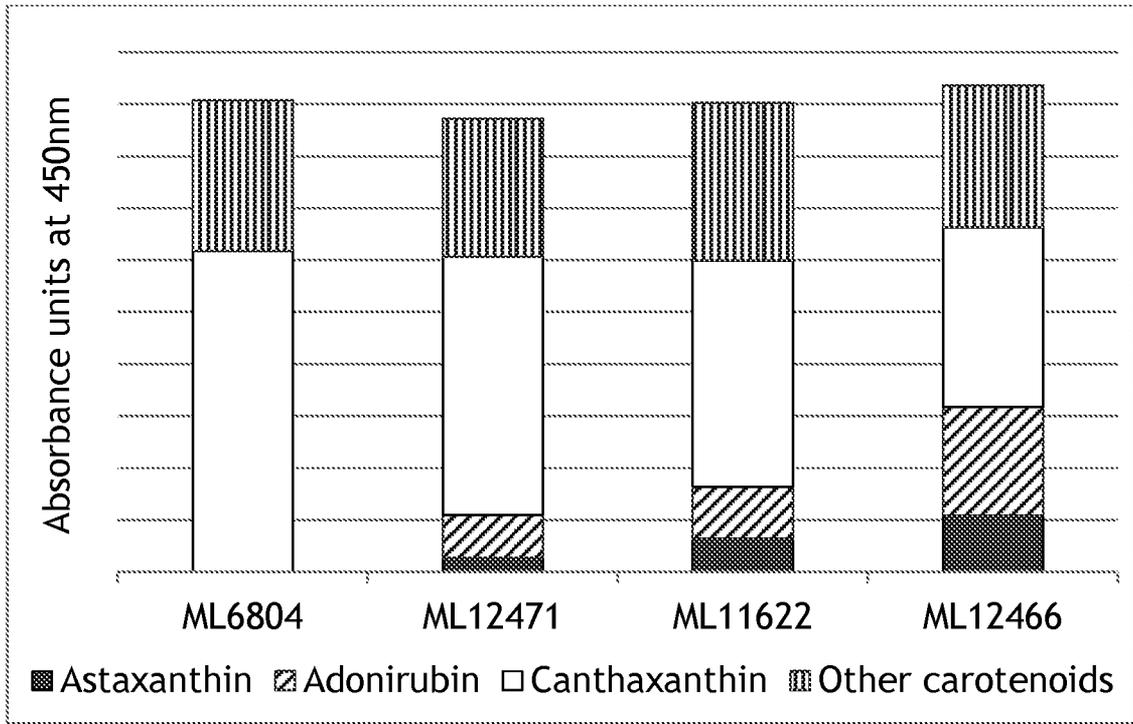


Fig. 2

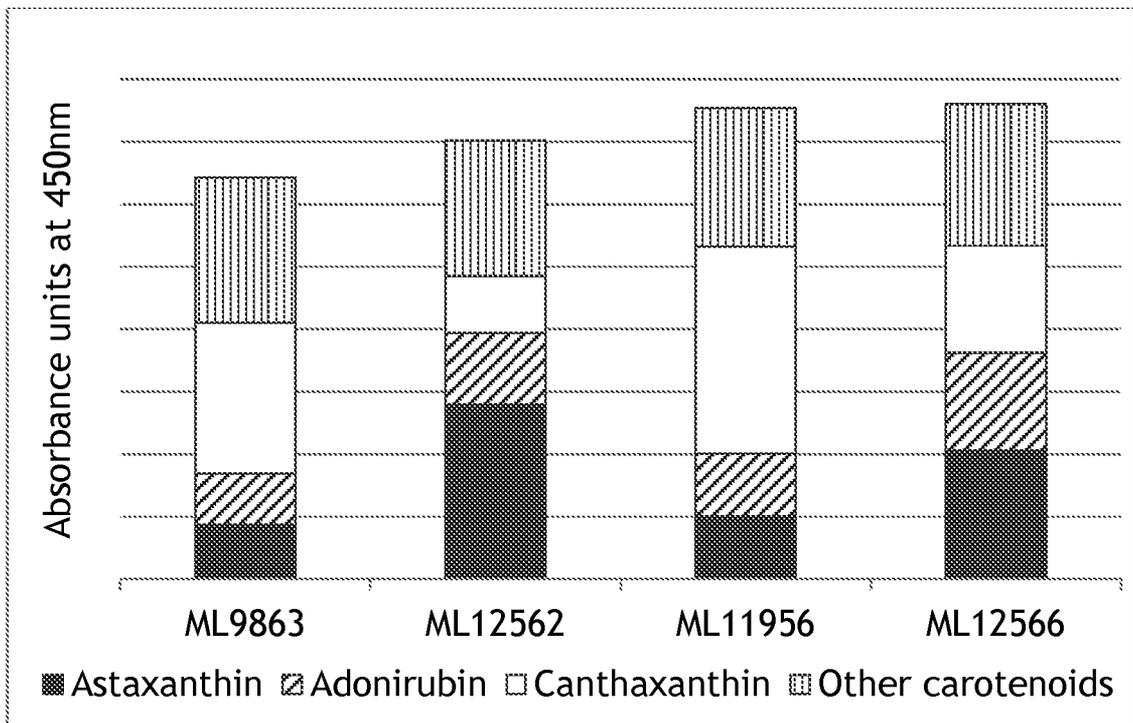
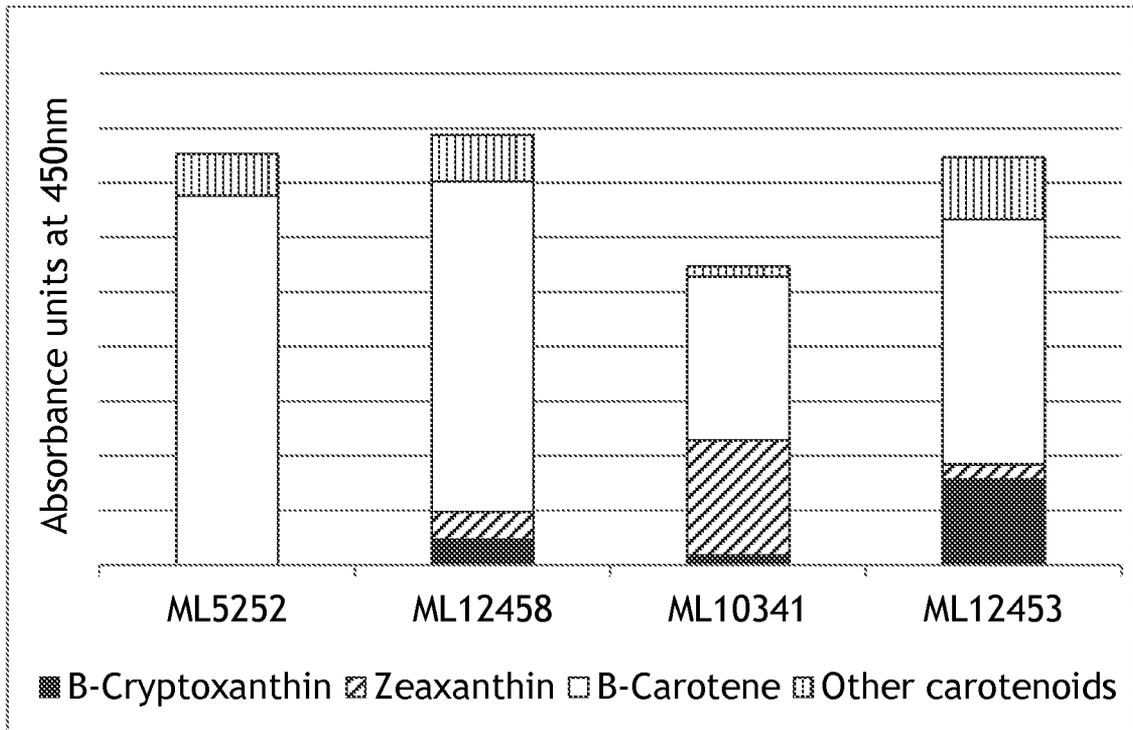


Fig. 3



# INTERNATIONAL SEARCH REPORT

International application No <b>PCT/IB2013/056583</b>
--

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. C12N9/02                      C12P23/00 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) C12N C12P		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, Sequence Search, BIOSIS, EMBASE, WPI Data		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE Geneseq [Online]  19 May 2005 (2005-05-19), "Haematococcus pluvialis ketocarotenoid-related hydroxylase protein SEQ ID NO: 16 from W02005019461", XP002718519, retrieved from EBI accession no. GSP:ADY52920 Database accession no. ADY52920 sequence  <div style="text-align: center;">-----</div> <div style="text-align: center;">-/--</div>	1-13
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents :		
"A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search	Date of mailing of the international search report	
10 January 2014	22/01/2014	
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  <b>Chakravarty, Ashok</b>	

## INTERNATIONAL SEARCH REPORT

International application No

PCT/IB2013/056583

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE Geneseq [Online]</p> <p>19 May 2005 (2005-05-19),  "Novel ketocarotenoid preparation  method-related protein SeqID78.",  XP002718520,  retrieved from EBI accession no.  GSP:ADY52444  Database accession no. ADY52444  sequence</p> <p style="text-align: center;">-----</p>	1-13
X	<p>DATABASE Geneseq [Online]</p> <p>19 May 2005 (2005-05-19),  "Haematococcus pluvialis  ketocarotenoid-related hydroxylase DNA.  SEQ ID NO 15 from W02005019461",  XP002718521,  retrieved from EBI accession no.  GSN:ADY52919  Database accession no. ADY52919  sequence</p> <p style="text-align: center;">-----</p>	1-13
X	<p>DATABASE EMBL [Online]</p> <p>1 November 2004 (2004-11-01),  "Sequence 31 from Patent W02004063359.",  XP002718522,  retrieved from EBI accession no.  EM PAT:CQ890446  Database accession no. CQ890446  sequence</p> <p style="text-align: center;">-----</p>	1-13