

NUCLEIC ACID MOLECULE ENCODING TRITERPENOID SYNTHASE

FIELD

The present invention relates to nucleic acid molecules encoding a polypeptide capable of producing a triterpenoid hydrocarbon. Additionally, the invention relates to polypeptides encoded by such nucleic acid molecules and use of such nucleic acid molecules or their encoded polypeptides in triterpenoid hydrocarbon production.

BACKGROUND

Fossil fuel is widely recognised as an unsustainable source of petroleum because of depleting supplies and the net contribution of these fuels to the carbon dioxide in the atmosphere. Renewable, carbon neutral fuels are necessary for environmental and economic sustainability. Biofuel derived from oil crops is a potential renewable and carbon neutral alternative to petroleum fuels. Currently, biofuels are produced mainly from soybeans, canola oil, animal fat, palm oil, corn oil and waste cooking oil.

Biofuel from the above sources cannot realistically satisfy even a small fraction of the existing demand for fuels. While researchers are seeking alternative feedstock for biofuel, algae have emerged as one of the most promising sources for biofuel production for three main reasons: (1) the yields of oil from algae are orders of magnitude higher than those for traditional oilseeds; (2) algae can grow in places away from the farmlands and forests, thus minimising the damages caused to ecosystems and food chain systems; and (3) algae can be grown in sewers utilising sewage and next to power-plant smokestacks where they digest pollutants and produce oil.

Algae are phototrophic cell factories, capable of deriving energy from sunlight and carbon from carbon dioxide. Algae convert carbon dioxide to potential biofuels, foods, feeds and high-value bioactives. Algae can provide several different types of renewable biofuels and valuable by-products such as antibiotics.

Not all algae are satisfactory for producing biofuel because of their low oil content or slow growth. Some species of the genus *Botryococcus* are characterised by an ability to produce high levels of hydrocarbons. For example, *Botryococcus braunii* is a unique colonial green alga that synthesises and accumulates an unusually high level of hydrocarbons up to 76% dry weight. This alga is a potentially good renewable source of useful lipids, hydrocarbons, polysaccharides, and other specialty chemicals.

The hydrocarbons produced by *Botryococcus* include (1) *n*-alkadienes and trienes (Race A), (2) triterpenoid botryococcenes and methylated squalenes (Race B), or (3) a tetraterpenoid, lycopadiene (Race L). Triterpenoid hydrocarbons can be used as feedstock for hydrocracking in an oil refinery to produce octane (gasoline, petrol), kerosene, and diesel, for example. Botryococcenes are preferred over alkadienes and alkatrienes for hydrocracking because botryococcenes will likely be transformed into a biofuel with a higher octane rating. It follows that use of algal lipids or hydrocarbons can greatly reduce the environmental impact associated with using coal and petroleum.

However, the production of photosynthetic fuel oils from *B. braunii* is not competitive with petroleum derived fuels. One major reason for this is the relatively slow growth rate of *B. braunii*. Furthermore, the gene(s) that causes the algae to produce botryococcene has not been identified or isolated in the art.

Disclosed in the prior art, however, is a squalene synthase (SEQ ID NO: 10) and an alleged botryococcene synthase (SEQ ID NO: 42), each derived from the Race B *B. braunii* Berkeley (Showa) strain. The alleged botryococcene synthase was not shown to produce botryococcene by molecular identification of botryococcene. Furthermore, the supporting data are consistent with the alleged botryococcene synthase having squalene synthase activity rather than botryococcene synthase activity as alleged.

Therefore, a need exists for synthetic or recombinant tools to facilitate triterpenoid hydrocarbon, particularly botryococcene, production.

SUMMARY

A first aspect provides an isolated nucleic acid molecule encoding a polypeptide capable of producing a triterpenoid hydrocarbon, wherein the polypeptide comprises: at least 15 contiguous amino acid residues of any one of SEQ ID NOs: 1 to 5 or 20 to 29; at least 70% sequence identity to any one of SEQ ID NOs: 1 to 5 or 20 to 29; or a functional fragment of SEQ ID NO: 29.

The nucleic acid molecule provides a tool for triterpenoid hydrocarbon production.

In one embodiment of the first aspect, the nucleic acid molecule comprises any one of SEQ ID NOs: 6 to 9 or 30. The nucleic acid molecule may be recombinant or synthetic.

A second aspect provides an isolated polypeptide capable of producing a triterpenoid hydrocarbon, wherein the polypeptide comprises: at least 15 contiguous amino acid residues of any one of SEQ ID NOs: 1 to 5 or 20 to 29; at least 70% sequence identity to any one of SEQ ID NOs: 1 to 5 or 20 to 29; or a functional fragment of SEQ ID NO: 29.

Particular examples of triterpenoid hydrocarbon molecules are botryococcene and squalene, which are isomers of each other and are produced from farnesyl pyrophosphate (FPP) substrate via presqualene pyrophosphate (PSPP) intermediate in *B. braunii*. Botryococcene occurs in high concentrations in Race B strains of *B. braunii*, but does not occur in Races A and L of *B. braunii*. Squalene occurs at much lower levels than botryococcene in Race B strains and occurs at low levels also in Races A and L of *B. braunii*.

The polypeptide of the second aspect may be used to synthesise a triterpenoid hydrocarbon in a cell-free system. Alternatively, the polypeptide of the second aspect may be expressed in a recombinant organism for production of a triterpenoid hydrocarbon or a botryococcene, respectively.

In one embodiment of the second aspect, the polypeptide comprises any one of SEQ ID NOs: 1 to 5 or 20 to 29. The polypeptide may be recombinant or synthetic.

The nucleic acid molecule encodes, or the polypeptide is, a botryococcene synthase, which is capable of converting FPP and/or PSPP to botryococcene. Alternatively, the nucleic acid molecule may encode, or the polypeptide may be, a squalene synthase, which is capable of converting FPP and/or PSPP to squalene.

A third aspect provides a vector, comprising the nucleic acid molecule of the first aspect.

A fourth aspect provides a recombinant non-human organism capable of producing a triterpenoid hydrocarbon, the organism comprising the nucleic acid molecule of the first aspect or the vector of the third aspect.

5 In another embodiment of the fourth aspect, the organism is a microorganism. The microorganism may be an alga or a bacterium. In another embodiment, the microorganism is selected from the group consisting of: *Escherichia coli*; *Chlamydomonas reinhardtii*; *Saccharomyces cerevisiae*; and *Pichia* sp.

10 A fifth aspect provides use of the nucleic acid molecule of the first aspect, the polypeptide of the second aspect, the vector of the third aspect, or the recombinant organism of the fourth aspect, to produce a triterpenoid hydrocarbon.

A sixth aspect provides a method for producing a triterpenoid hydrocarbon, comprising the step of growing the recombinant organism of the fourth aspect under conditions sufficient for the organism to produce a triterpenoid hydrocarbon.

15 A seventh aspect provides a triterpenoid hydrocarbon produced by the method of the sixth aspect. In various embodiments, the triterpenoid hydrocarbon is squalene, botryococcene, dehydrosqualene or dehydrobotryococcene.

20 An eighth aspect provides use of the nucleic acid molecule of the first aspect, the polypeptide of the second aspect, the vector of the third aspect, the recombinant organism of the fourth aspect, the triterpenoid hydrocarbon produced by the method of the sixth aspect, or the triterpenoid hydrocarbon of the seventh aspect, in the production of an intermediate of a biofuel, a bioplastic, a pharmaceutical, a food additive or an industrial chemical production.

25 A ninth aspect provides a method for producing an intermediate of a biofuel, a bioplastic, a pharmaceutical, a food additive or an industrial chemical production, comprising the steps of growing the recombinant organism of the fourth aspect under conditions sufficient for the organism to produce a triterpenoid hydrocarbon, harvesting the triterpenoid hydrocarbon from the organism, and producing the intermediate from the triterpenoid hydrocarbon.

A tenth aspect provides an intermediate produced by the method of the ninth aspect of a biofuel, a bioplastic, a pharmaceutical, a food additive or an industrial chemical, production.

30 In an embodiment of any one of the first to ninth aspects, the triterpenoid hydrocarbon is produced from FPP and/or PSPP.

In another embodiment of any one of the first to ninth aspects, the triterpenoid hydrocarbon is a botryococcene or a squalene.

35 To overcome the slow growth rate of some organisms, such as *B. braunii*, a triterpenoid hydrocarbon nucleic acid molecule, e.g. a nucleic acid molecule encoding a botryococcene synthase or a squalene synthase, may be transformed into other organisms that are faster growing, such as *C. reinhardtii*, *Escherichia coli*, *S. cerevisiae*, *Pichia* sp., or any other organism that is amenable to transformation and faster growth than *B. braunii*. Such synthetic or recombinant tools and recombinant organisms will facilitate the commercialisation of triterpenoid hydrocarbon production, 40 such as for biofuel or bioplastic production. Such a recombinant organism may be used commercially

in an enclosed mass growth or culture system, e.g., to provide an intermediate in or a feedstock for biofuel or bioplastic production, or to provide a source of triterpenoid hydrocarbons, such as squalene or botryococcene, for use in other processes such as chemical synthesis or cosmetic manufacture.

5 BRIEF DESCRIPTION OF THE FIGURES

Figures 1 to 5 and 20 to 28 depict the amino acid sequences encoded by exons of squalene synthase/botryococcene synthase homologues of the Ayamé 1 strain of *B. braunii*, Race B. The amino acid sequences of Figures 1 to 5 and 20 to 28 correspond with SEQ ID NOs: 1 to 5 and 20 to 28, respectively, and are deduced from the nucleic acid sequences provided in Figures 6 to 9 and 30.

10 An asterisk represents a stop codon.

Figures 6 to 9 depict the nucleic acid sequences of exons of squalene synthase/botryococcene synthase homologues of the Ayamé 1 strain of *B. braunii*, Race B. Some of the figures include nucleic acid sequences of introns intervening between adjacent exons. The nucleic acid sequences of Figures 6 to 9 and 30 correspond with SEQ ID NOs: 6 to 9 and 30, respectively.

15 Figure 10 depicts the amino acid sequence of squalene synthase of *B. braunii* Berkeley strain, Race B (accession identifications: gi|6636500; gb|AAF20201.1; AF205791_1) and corresponds with SEQ ID NO: 10.

20 Figure 11 depicts the nucleic acid sequence (mRNA, complete cds; accession identifications: gi|6636499; gb|AF205791.1; AF205791) encoding the amino acid sequence of Figure 10 and corresponds with SEQ ID NO: 11.

Figure 12 depicts the partial amino acid sequence of squalene synthase of *B. braunii* Berkeley strain, Race B (accession identifications: gi|7532841; gb|AAF63255.1) and corresponds with SEQ ID NO: 12.

25 Figure 13 depicts the nucleic acid sequence (partial cds; accession identifications: gi|7532838; gb|AH009227.1; SEG_AF205789S) encoding the amino acid sequence of Figure 12 and corresponds with SEQ ID NO: 13.

Figure 14 depicts the amino acid sequence of Figure 10 against which the consensus deduced amino acid sequences of Figures 1 to 5 have been positioned. An asterisk represents a stop codon. The consensus sequence corresponds with SEQ ID NO: 14.

30 Figure 15 illustrates the nucleic acid sequence (SEQ ID NO: 15) of the squalene synthase pSP124S construct transformed into *C. reinhardtii* in Example 11.

Figures 16 and 17 depict the nucleic acid sequences (corresponding with SEQ ID NO: 16 and SEQ ID NO: 17, respectively) of primer pair 1 (PP1) used to detect the presence of the squalene synthase pSP124S construct in transformed *C. reinhardtii* in Example 11.

35 Figures 18 and 19 depict the nucleic acid sequences (corresponding with SEQ ID NO: 18 and SEQ ID NO: 19, respectively) of primer pair 2 (PP2) used to detect the presence of the squalene synthase pSP124S construct in transformed *C. reinhardtii* in Example 11.

Figure 29 depicts the amino acid sequence of ATB1, a triterpenoid hydrocarbon synthetic polypeptide of *B. braunii* Race B, Ayamé 1 strain, and corresponds with SEQ ID NO: 29. The asterisk

represents the stop codon. The polypeptide comprises exons that correspond with SEQ ID NOs: 1 to 5 and 20 to 28. ATB1 does not comprise an exon 8.

Figure 30 depicts the inferred cDNA nucleic acid sequence encoding the amino acid sequence of Figure 29 and corresponds with SEQ ID NO: 30. The gene comprises 8 exons (labelled 1 to 7 and 9), of which exons 2, 4, 6 and 9 are underlined. The cDNA comprises exons that correspond with SEQ ID NOs: 6 to 9.

Figure 31 depicts a genomic nucleic acid sequence encoding the amino acid sequence of Figure 29 and corresponds with SEQ ID NO: 31.

Figure 32 depicts a nucleic acid sequence used for cloning the *ATB1* triterpenoid hydrocarbon synthetic nucleic acid sequence. The full sequence corresponds to SEQ NO: 32. The sequence, other than immediately flanking the start and stop codons, corresponds to the sequence of Figure 30, i.e. SEQ ID NO: 30. The sequence was used to generate pUC57/ATB1, pET11a/ATB1, pET302/NT-His-ATB1, and CT-His-ATB1/pET302, referred to herein as constructs 1, 2, 3, and 4, respectively.

Figure 33 is a photograph of an agarose gel for molecular detection of the squalene synthase pSP124S construct in transformed *C. reinhardtii* of Example 11 using the PCR primers of Figures 16 to 19. PCR was performed on genomic DNA extracted from individual colonies positive for ZeocinTM resistance. Lanes: 1 untransformed *C. reinhardtii* cells, PP1 negative control; 2 & 24 no cells, squalene synthase pSP124S construct, PP1 positive control; 3 plasmid pSP124S, PP1 negative control; 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, transformed colonies 12, 13, 18, 24, 28, 76, 82, 87, 88, and 89 respectively, PP1; 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, transformed colonies 12, 13, 18, 24, 28, 76, 82, 87, 88, and 89 respectively, PP2.

Figure 34 is a photograph of agarose gel electrophoresis separating restriction digests of pET11a/ATB1, pET302/NT-His-ATB1, and CT-His-ATB1/pET302 (constructs 2, 3 and 4, respectively) and confirming that the expected fragment sizes are present. Figure 34A shows that construct 2 was cut with *NdeI* and *BstXI*. Figure 34B shows that constructs 3 and 4 were each cut with *NdeI* and *XhoI*. Figure 34A and B 100 bp ladder fragment sizes: 1500; 1000; 900; 800; 700; 600; 500; 400; 300; 200; 100 bp. Figure 34A 1 kb ladder fragment sizes: 10; 8; 6; 5; 4; 3; 2.5; 2; 1.5; 1; 0.75; 0.5; 0.25 kb. Figure 34B 1 kb ladder fragment sizes: 10; 8; 6; 5; 4; 3; 2; 1.5; 1.0; 0.5.

Figure 35 is a photograph of an SDS-PAGE minigel of proteins of whole untransformed (*E. coli*) and whole *E. coli* transformed (T) with construct 2 (Lanes 1 to 6) and supernatant of a cell lysate of *E. coli* transformed (T) with construct 2 (Lanes 8 and 9). *E. coli* were induced with IPTG (1 mM) for 0 h (Lanes 1 and 2), 2 h (Lanes 3 and 4) or 4 h (Lanes 5, 6, 8 and 9). Proteins were visualised with Coomassie blue dye. Arrows lane 7: molecular weight markers upper arrow 46 kDa, lower arrow 30 kDa. Arrow lane 9: soluble ATB1 protein.

Figure 36 depicts the chemical conversion of two farnesyl pyrophosphate C₁₅ units to the squalene C₃₀ unit by the enzyme squalene synthase with NADPH cofactor (A and B) and the chemical conversion of two farnesyl pyrophosphate C₁₅ units to the botryococcene C₃₀ unit by the enzyme botryococcene synthase with NADPH cofactor (B).

Figure 37 is a gas chromatography-mass spectrometry (GC-MS) analysis of authentic squalene.

Figure 38 is a GC-MS analysis of authentic C₃₀ botryococcene.

Figure 39 is a GC-MS analysis of heptane extracts of untransformed *E. coli* derived from Example 15.

Figure 40 is a GC-MS analysis of heptane extracts of *E. coli* transformed with pET11a/ATB1 (construct 2) and induced with IPTG for 16 h, showing mass spectrum of generated dehydrosqualene.

Figure 41 is a GC-MS analysis of heptane extracts of *E. coli* transformed with pET11a/ATB1 (construct 2) and induced with IPTG for 16 h, showing mass spectrum of generated squalene.

Figure 42 is a GC-MS analysis of heptane extracts of lysates, derived from *E. coli* transformed with CT-His-ATB1/pET302 (construct 4), induced with IPTG for 2 h, then lysed and the lysate incubated for 30 min with an excess both of FPP and NADPH.

Figure 43 depicts a nucleic acid sequence (SEQ ID NO: 33) of nucleotides inserted immediately upstream of the native start codon of the nucleotide sequence provided in Figure 30.

Figure 44 depicts the nucleic acid sequence of the *Nsi*I restriction site (SEQ ID NO: 34).

Figure 45 depicts the nucleic acid sequence of the *Nde*I restriction site (SEQ ID NO: 35).

Figure 46 depicts the amino acid sequence (SEQ ID NO: 36) of the ATB1-MH polypeptide encoded by pET11a/ATB1 (construct 2), i.e. the ATB1 polypeptide of Figure 29 (SEQ ID NO: 29) augmented consecutively with one methionine and one histidine residue at the C-terminus of ATB1.

The asterisk represents the stop codon.

Figure 47 depicts a nucleic acid sequence (SEQ ID NO: 37) of nucleotides, comprising *Xho*I and *Sma*I restriction sites, inserted immediately downstream of the native stop codon of the nucleotide sequence provided in Figure 30.

Figure 48 depicts the amino acid sequence (SEQ ID NO: 38) of the MHHHHHHHHMH-ATB1-MH polypeptide encoded by pET302/NT-His-ATB1 (construct 3), i.e. the ATB1 polypeptide of Figure 29 (SEQ ID NO: 29) augmented consecutively with one methionine, eight histidine, one methionine, and one histidine residues at the N-terminus of ATB1 and augmented consecutively with one methionine and one histidine residue at the C-terminus of ATB1 (Figure 29, SEQ ID NO: 29). The asterisk represents the stop codon.

Figure 49 depicts the amino acid sequence (SEQ ID NO: 39) of the MH-ATB1-MH-HHHHHVNSLEIDDIRA polypeptide encoded by CT-His-ATB1/pET302 (construct 4), i.e. the ATB1 polypeptide of Figure 29 (SEQ ID NO: 29) augmented consecutively with at its N-terminus with one histidine and one methionine residue and augmented consecutively at its N-terminus with five histidine residues and residues VNSLEIDDIRA at the C-terminus of ATB1 (Figure 29, SEQ ID NO: 29). The asterisk represents the stop codon.

Figure 50 depicts the nucleic acid sequence of the *Bst*XI restriction site (SEQ ID NO: 40).

Figure 51 depicts the nucleic acid sequence of the *Xho*I restriction site (SEQ ID NO: 41).

Figure 52 depicts the amino acid sequence (SEQ ID NO: 42) provided as SEQ ID NO: 2 of US20100041120.

Figure 53 illustrates a polypeptide sequence alignment of SEQ ID NO: 29 disclosed herein, and SEQ ID NOs: 10 and 42 disclosed previously. The boxed regions indicate Domains I, II, III, IV and V, which have been observed previously to be highly conserved amongst diverse squalene synthases. Domains III and IV have been correlated with conversion of FPP to PSPP, whereas Domains III and IV have been correlated with conversion of PSPP to squalene. SEQ ID NO: 10 comprises a hydrophobic C-terminus, which is not present in SEQ ID NOs: 29 or 42. SEQ ID NO: 29 is 87% identical to SEQ ID NO: 42 and 38% identical to SEQ ID NO: 10. SEQ ID NO: 42 is 39% identical to SEQ ID NO: 10.

Figure 54 depicts the nucleic acid sequence of plasmid for *Chlamydomonas* Expression 3-ATB1 (pCE3-ATB1) of Example 19.

Figures 55 to 62 depict nucleic acid sequences of oligonucleotide primers used in Example 19.

DETAILED DESCRIPTION

Recombinant DNA technology has provided the main impetus for the rise in biotechnology since the 1980s. Currently, almost every major industrial biotechnological process is based around the use of genetically modified organisms.

Disclosed herein are synthetic and recombinant tools useful for triterpenoid hydrocarbon production. Such tools include nucleic acid molecules, polypeptides, vectors, recombinant organisms, uses and methods. Triterpenoid hydrocarbons can be used for production of biofuel, bioplastic, pharmaceuticals, food additives, industrial chemicals or specialty chemicals, for example.

Examples of use of triterpenoid hydrocarbons for manufacture of intermediates and products such as biofuel, bioplastic, pharmaceuticals (e.g. vaccine adjuvant), food additives (e.g. for cancer prevention), industrial chemicals (e.g. lubricant), specialty chemicals or cosmetics (e.g. light protection) is disclosed in Ab Gapor *et al.* *Palm Oil Developments* 32: 36-40; Newmark, *Cancer Epidemiology, Biomarkers & Prevention*, 1997, 6, 1101-1103; Huang *et al.*, *Molecules* 2009, 14, 540-554; Auffray, *International Journal of Cosmetic Science*, 2007, 29, 23-29; Cox and Coulter, *Vaccine*, 1997, 15, 248-256; Fox, *Molecules* 2009, 14, 3286-3312; Schroepfer, *Ann. Rev. Biochem.*, 1981, 50, 585-621; Tran *et al.*, *Fuel*, 2010, 89, 265-274; Catchpole *et al.* *Ind. Eng. Chem. Res.* 1997, 36, 4318-4324; He *et al.*, *J. Agric. Food Chem.* 2002, 50, 368-372.

Exemplary triterpenoid hydrocarbon synthetic nucleic acid molecules encode a botryococcene synthase or a squalene synthase.

Whilst not wishing to be bound to any particular hypothesis, a botryococcene synthase may be advantageous when compared with a squalene synthase, since squalene synthase is subject to biosynthetic feedback inhibition, whereas it is proposed that botryococcene synthase is not subject to biosynthetic feedback inhibition. Alternatively, it is proposed that botryococcene synthase is subject to minimal or reduced biosynthetic feedback inhibition relative to biosynthetic feedback inhibition exerted upon squalene synthase. This hypothesis may account for the unusually high level of triterpenoid hydrocarbons synthesised and accumulated by *B. braunii*. It follows that although a squalene synthase may be a useful synthetic or recombinant tool for producing a triterpenoid hydrocarbon, a

botryococcene synthase may be a superior synthetic or recombinant tool for producing a triterpenoid hydrocarbon.

In one embodiment of the first aspect, a nucleic acid molecule encoding a botryococcene synthase polypeptide may be isolated from Race B *B. braunii*, specifically strain Ayamé 1 (Ivory Coast). Alternatively, the Race B *B. braunii* may be selected from the group consisting of Kossou (Ivory Coast), Overuyo 3 (Bolivia), Paquemar (Martinique), La Manzo (Martinique), CCAC 0121, strain CH 28, strain CH 86, or strain 1284.

Whilst not wishing to be bound by any particular hypothesis, strains Ayamé 1 and Paquemar were considered to be high hydrocarbon-producing strains, a consideration that appeared to be confirmed in respect of Ayamé 1 by photographs showing substantial accumulation of oil deposits. Qualitatively, Ayamé 1 has been confirmed by gas chromatography to contain greater amounts of hydrocarbon than other strains. Accordingly, Ayamé 1 was selected for isolation of triterpenoid hydrocarbon nucleic acid molecules.

In another embodiment of the first aspect, a nucleic acid molecule encoding a squalene synthase polypeptide may be isolated from a Race A, a Race B or a Race L *B. braunii*. The race A *B. braunii* may be selected from the Lingoult strain (France), Overjuyo 7 strain (Bolivia), or Jillamatong strain (Australia). The race B strains may be selected from the group of strains Ayamé 1 (Ivory Coast), Kossou (Ivory Coast), Overuyo 3 (Bolivia), Paquemar (Martinique), La Manzo (Martinique), CCAC 0121, strain CH 28, strain CH 86, or strain 1284. The Race L *B. braunii* may be selected from the Madras 3 strain (India) or the Yamoussoukro 4 strain (Ivory Coast).

Alternatively, a nucleic acid molecule encoding a triterpenoid hydrocarbon synthetic polypeptide may be synthesised.

The terms “nucleic acid molecule” and “polynucleotide” are used synonymously and refer to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. A nucleic acid of the present disclosure will generally contain phosphodiester bonds, although in some cases, nucleic acid analogues may be used that may have alternate backbones, comprising, e.g., phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphosphoroamidite linkages, and peptide nucleic acid backbones and linkages. Other analogue nucleic acids include those with positive backbones; non-ionic backbones, and non-ribose backbones. Thus, nucleic acid molecules or polynucleotides may also include modified nucleotides that permit correct read-through by a polymerase.

“Nucleic acid sequence” or “polynucleotide sequence” includes both the sense and antisense strands of a nucleic acid molecule as either individual single strands or in a duplex. As will be appreciated by those skilled in the art, the depiction of a single strand also defines the sequence of the complementary strand; thus the sequences described herein also provide the complement of the sequence. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses variants thereof, and complementary sequences, as well as the sequence explicitly indicated. The nucleic acid molecule may be DNA, both genomic and cDNA, RNA or a hybrid.

The phrase “a nucleic acid molecule encoding” refers to a nucleic acid molecule which contains sequence information for a structural RNA such as rRNA, a tRNA, or the primary amino acid

sequence of a specific protein or polypeptide, or a binding site for a *trans*-acting regulatory agent. This phrase specifically encompasses degenerate codons (i.e., different codons which encode a single amino acid) of the native sequence or sequences that may be introduced to conform to codon preference in a specific host cell.

Letters other than A, T, C, and G when present in a nucleic acid sequence disclosed herein represent ambiguity. That is, of all the molecules sampled, there is more than one kind of nucleotide at that position. The symbols (letters) are defined as follows: A = adenine; C = cytosine; G = guanine; T = thymine; R = G A (purine); Y = T C (pyrimidine); K = G T (keto); M = A C (amino); S = G C (strong bonds); W = A T (weak bonds); B = G T C (all but A); D = G A T (all but C); H = A C T (all but G); V = G C A (all but T).

The term "gene" may be used interconvertibly with the term "nucleic acid molecule" or "nucleic acid sequence".

Nucleic acid molecules disclosed herein, e.g. encoding botryococcene synthase or squalene synthase can be expressed recombinantly in organisms, e.g., algae, cyanobacteria, photosynthetic or non-photosynthetic bacteria, fungi, including yeasts, or plant cells. As appreciated by one of skill in the art, expression constructs can be designed taking into account such properties as codon usage frequencies of the organism in which the nucleic acid molecule is to be expressed. Codon usage frequencies can be tabulated using known methods. Codon usage frequency tables, including those for algae and cyanobacteria, are also available in the art.

The term "complementary" is used herein to mean that the sequence is complementary to all or a portion of a reference polynucleotide sequence.

A "polypeptide" refers to a polymer formed from the linking, in a defined order, of amino acids by means of an amide bond. Usually, a polypeptide performs some specific function in an organism. As used herein, the term "polypeptide" includes a peptide.

As used herein, "enzyme" refers to a polypeptide that catalyses the specific conversion of a molecular substrate to a molecular product. The term "enzyme" is used herein interconvertibly with the term "polypeptide".

The term "isolated", when applied to a nucleic acid molecule or polypeptide, denotes that the nucleic acid or polypeptide is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state and may be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high pressure liquid chromatography. A polypeptide that is the predominant species present in a preparation is substantially purified. In particular, an isolated nucleic acid molecule is separated from open reading frames, which encode polypeptides other than the polypeptide of interest, that flank the gene from which the nucleic acid molecule is derived.

As used herein, an "isolated" nucleic acid molecule or polypeptide may be purified to at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 12.5%, 15%, 17.5%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 82.5%, 85%, 87.5%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more purity by weight.

As used herein, "variant" refers to a nucleic acid molecule that comprises degenerate codon substitutions or combinations of deoxyribo- and ribo-nucleotides, and combinations of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine hypoxanthine, isocytosine, isoguanine, etc. A "variant" nucleic acid molecule may also comprise a deletion or an insertion of a nucleotide. As used herein, a "variant" includes a homologue of a nucleic acid molecule. A "variant" further includes a nucleic acid molecule that under stringent conditions hybridises to a nucleic acid molecule encoding a polypeptide provided as any one of SEQ ID NOs: 1 to 41, 86, 87, 89 to 97, 115 to 117 or 124 to 179, or a functional fragment thereof, or to a nucleic acid molecule provided any one of SEQ ID NOs: 42 to 75, 88, 98, 118 to 120 or 180 to 229, or a functional fragment thereof. A "variant" also refers to the complement of any such nucleic acid sequence described above. All variants of a nucleic acid molecule encode a polypeptide that has triterpenoid hydrocarbon synthetic activity.

A "variant" polypeptide includes a polypeptide comprising a deletion, an insertion, or a substitution of an amino acid. As used herein, a "variant" includes a homologue of polypeptide. All variants of a polypeptide have triterpenoid hydrocarbon synthetic activity.

As used herein, "homologue", "homology" and "homologous" refer to nucleotide sequence similarity and amino acid sequence similarity among nucleic acid molecules and polypeptides, respectively. "Homology" may also refer to functional similarity. Homology is an indicator of ancestry or evolution.

An "orthologue" and a "paralogue" are two types of homologous sequences and are encompassed by the term "homologue". "Orthology" describes nucleotide sequences or polypeptides in different species that derive from a common ancestor. "Orthologous" nucleic acid molecules or polypeptides may or may not have the same function. "Paralogy" describes homologous nucleotide sequences or polypeptides within a single species that diverged by gene duplication.

As used herein, a "homologue" of a first nucleic acid molecule or sequence (e.g., any one of SEQ ID NOs: 6 to 9 or 30), or first polypeptide (e.g., any one of SEQ ID NOs: 1 to 5 or 20 to 29), is a second nucleic acid molecule or sequence, or second polypeptide, in the same organism type or in a different organism type, that is substantially similar to the first nucleic acid sequence or first polypeptide, in terms of its sequence. Otherwise, the "homologue" is substantially similar to the first nucleic acid molecule or first polypeptide in terms of function.

Accordingly, a triterpenoid hydrocarbon synthetic nucleic acid molecule, or the triterpenoid hydrocarbon synthetic polypeptide, includes a homologue, an orthologue or a paralogue of a nucleic acid molecule or polypeptide disclosed herein. The homologue, orthologue or paralogue corresponds with a nucleic acid molecule or polypeptide disclosed herein. The homologue, orthologue or paralogue may be a homologue, orthologue or paralogue of SEQ ID NO: 29 or SEQ ID NO: 30, but not of SEQ ID NOs: 10 to 13 or 42.

As used herein, "fragment" refers to a nucleic acid molecule that comprises a portion of a nucleic acid molecule disclosed herein, provided that the encoded polypeptide has triterpenoid hydrocarbon synthetic activity. For example, a fragment may comprise deletion of a nucleotide from the 5' terminus, the 3' terminus, both the 5' terminus and the 3' terminus, or from between the 5'

terminus and the 3' terminus of the nucleic acid molecule, provided that the encoded polypeptide has triterpenoid hydrocarbon synthetic activity.

Similarly, a "fragment" of a polypeptide may comprise deletion of an amino acid from the N-terminus, the C-terminus, both the N-terminus and the C-terminus, or from between the N-terminus and the C-terminus of the polypeptide, provided that the polypeptide fragment has triterpenoid hydrocarbon synthetic activity.

As used herein, a "triterpenoid hydrocarbon" is a compound that may be characterised by the presence of six 5-carbon isoprene (2-methyl-1,3-butadiene) units, or three 10-carbon terpene units. A "triterpenoid hydrocarbon" includes triterpenoids, methylated triterpenoids and partially cyclic triterpenoids. Particular examples of triterpenoid hydrocarbons include those synthesised by *B. braunii*. Triterpenoid hydrocarbons include botryococcenes and squalenes.

As used herein, "botryococcene" refers to a hydrocarbon compound that is produced by a living organism, and is sparingly soluble in water, but is readily soluble in organic solvents. More particularly, "botryococcene" refers to a hydrocarbon comprising isoprene units. More specifically, "botryococcene" refers to a triterpenoid hydrocarbon comprising the formula C_nH_{2n-10} , where $n=30$ to 37 (C_{30} to C_{37}). Whilst not wishing to be bound to any particular hypothesis, triterpenoid synthesis, hence botryococcene synthesis, is dependent upon isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) precursors. IPP and DMAPP combine to produce geranyl pyrophosphate (GPP), which then combines with IPP to produce FPP. Two molecules of FPP combine to form the intermediate molecule PSPP within the active site of the botryococcene synthase enzyme, which then forms botryococcene. Botryococcene, in particular C_{30} botryococcene, which is the precursor of all higher botryococcenes (C_{31} to C_{37}), is a triterpene resulting from head-to-head 1'-3 linkage of two FPP molecules (Figure 36). Ultimately, botryococcene refers to any one of C_{30} to C_{37} hydrocarbons produced by the initial activity of a botryococcene synthase, including cyclic and acyclic botryococcenes and dehydrobotryococcenes.

As used herein, "squalene" refers to a hydrocarbon compound that is produced by a living organism, and is sparingly soluble in water, but is readily soluble in organic solvents. More particularly, "squalene" refers to a hydrocarbon comprising isoprene units. More specifically, squalene refers to a triterpenoid hydrocarbon comprising 30 carbon atoms (C_{30}). Triterpenoid synthesis, hence squalene synthesis, is dependent upon IPP and DMAPP precursors. IPP and DMAPP combine to produce GPP, which then combines with IPP to produce FPP. Two molecules of FPP combine to form the intermediate molecule PSPP within the active site of the squalene synthase enzyme, which then forms squalene. Squalene is a triterpene resulting from head-to-head 1'-1 linkage of two FPP molecules (Figure 36). Ultimately, squalene refers to a C_{30} hydrocarbon produced by a squalene synthase.

As used herein, a "substrate" is any compound that may be acted upon by a triterpenoid hydrocarbon synthetic polypeptide to produce a triterpenoid hydrocarbon. The substrate may itself be a terpenoid hydrocarbon. More particularly, the substrate may be FPP and/or PSPP or a precursor of FPP. The substrate, FPP and/or PSPP or precursor of FPP may be synthetic, recombinant or may be produced by an organism via one or more enzymes. The substrate, FPP and/or PSPP or precursor of

FPP may be produced by a recombinant organism of the fourth aspect harbouring a nucleic acid molecule of the first aspect, a polypeptide of the second aspect, or a vector of the third aspect. The substrate, FPP and/or PSPP or precursor of FPP may be provided from an exogenous source to a polypeptide of the second aspect or a recombinant organism of the fourth aspect.

As stated above, triterpenoid hydrocarbon synthetic polypeptide has an enzymatic activity of catalysing the condensation of two C_{15} molecules of FPP to a long-chain C_{30} triterpenoid hydrocarbon, e.g. squalene or botryococcene. In addition, a triterpenoid hydrocarbon synthetic polypeptide may catalyse the condensation of two C_{20} geranylgeranyl pyrophosphate (GGPP) molecules to a long-chain C_{40} lycopadiene hydrocarbon.

A "triterpenoid hydrocarbon synthetic nucleic acid molecule" is a nucleic acid molecule encoding a polypeptide capable of converting a substrate to a triterpenoid hydrocarbon. In one embodiment, a "triterpenoid hydrocarbon synthetic nucleic acid molecule" comprises a nucleic acid molecule that encodes a polypeptide that comprises, or is derived from, any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment or variant thereof. In another embodiment, the triterpenoid hydrocarbon synthetic nucleic acid molecule comprises, or is derived from, any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment or variant thereof. In another embodiment, the triterpenoid hydrocarbon synthetic nucleic acid molecule encodes a polypeptide that is substantially similar to any one of SEQ ID NOs: 10, 12 or 42, or a fragment or variant thereof. In yet another embodiment, the triterpenoid hydrocarbon synthetic nucleic acid molecule is substantially similar to SEQ ID NO: 11 or SEQ ID NO: 13. However, the triterpenoid hydrocarbon synthetic nucleic acid molecule does not encode a squalene synthase identical to SEQ ID NOs: 10, 12 or 42 and is not identical to SEQ ID NO: 11 or SEQ ID NO: 13.

A "triterpenoid hydrocarbon synthetic nucleic acid molecule" may:

1) comprise a region of about 100, 150, 200, 300, 400, 500, 1000, 1500, 2000, or 5000, or more nucleotides encoding a polypeptide comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment or variant thereof;

2) comprise a region of about 100, 150, 200, 300, 400, 500, 1000, 1500, 2000, or 5000, or more nucleotides comprising any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment or variant thereof;

3) under stringent conditions hybridise to a nucleic acid molecule encoding a polypeptide comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment thereof, or the complement thereof;

4) under stringent conditions hybridise to a nucleic acid molecule comprising any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment thereof, or the complement thereof;

5) encode a polypeptide comprising at least 15 contiguous amino acids, or at least 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, or 550, or more contiguous residues of a polypeptide comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment or variant thereof;

6) encode a polypeptide that has at least 55%, at least 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater amino acid identity to a polypeptide comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment thereof, over a comparison window of at least 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, or 550 amino acid residues;

7) have a nucleic acid sequence that has greater than about 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher nucleotide sequence identity to a nucleic acid molecule encoding a polypeptide comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment thereof, over a comparison window of at least about 20, 50, 100, 200, 300, 400, 500, 1000, or more nucleotides;

8) have a nucleic acid sequence that has greater than about 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher nucleotide sequence identity to a nucleic acid molecule comprising any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment thereof, over a comparison window of at least about 20, 50, 100, 200, 300, 400, 500, 1000, or more nucleotides;

9) be amplified by primers to a nucleic acid molecule encoding a polypeptide comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29;

7) is amplified by primers to a nucleic acid molecule comprising any one of SEQ ID NOs: 6 to 9 or 30.

Exons 2.1 (SEQ ID NO: 6, Figure 6), 3.11 and 4.1 (SEQ ID NO: 7, Figure 7), 5.3 and 6.1 (SEQ ID NO: 8, Figure 8) and 9.2 (SEQ ID NO: 9, Figure 9) correspond with SEQ ID NO: 30 (Figure 30) and encode amino acids that correspond with SEQ ID NO: 29 (Figure 29).

Exons 2.1, 4.1, 5.3, 6.1 and 9.2 encode amino acids (SEQ ID NOs: 1 to 5, Figures 1 to 5) that correspond with SEQ ID NO: 29 (Figure 29).

As used herein, an "exon" refers to a nucleic acid sequence (either DNA or RNA) that is represented in the mature form of an RNA molecule after portions (introns) of a precursor RNA have been removed by splicing. The mature RNA molecule can be a messenger RNA or a functional form of a non-coding RNA such as rRNA or tRNA.

The person skilled in the art will appreciate that the amino acid translation of a nucleotide sequence encoding an exon may not provide all amino acid residues due to the presence of incomplete codons arising from splice junctions.

In all cases, a "triterpenoid hydrocarbon synthetic nucleic acid molecule" encodes a polypeptide capable of producing a triterpenoid hydrocarbon.

The term "triterpenoid hydrocarbon synthetic nucleic acid molecule" refers to a double stranded or single stranded nucleic acid molecule. The triterpenoid hydrocarbon synthetic nucleic acid molecule encodes an active or functional triterpenoid hydrocarbon synthetic polypeptide that catalyses the conversion of a substrate to a hydrocarbon.

A "triterpenoid hydrocarbon synthetic polypeptide" is an active or functional polypeptide that is capable of converting a substrate to a triterpenoid hydrocarbon. The term includes polypeptide

fragments or variants that retain function sufficient to convert a substrate to a triterpenoid hydrocarbon.

In one embodiment, a "triterpenoid hydrocarbon synthetic polypeptide" comprises the amino acid sequence of any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a fragment or variant of SEQ ID NO: 29, or a sequence substantially similar to any one of SEQ ID NOs: 10, 12 or 42, or a fragment or variant thereof.

Thus, a triterpenoid hydrocarbon synthetic polypeptide can:

1) have at least 55% identity, or least 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or greater amino acid identity to any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a fragment of SEQ ID NO: 97, over a comparison window of at least 15, 20, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, or 550 or more amino acids;

2) comprise at least 15 contiguous amino acids, or at least 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, or more contiguous amino acids of any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a fragment or variant of SEQ ID NO: 29; or

3) bind to an antibody raised against an immunogen comprising a partial or full amino acid sequence of any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a fragment of SEQ ID NO: 29, and conservatively modified variants thereof.

In all cases, a "triterpenoid hydrocarbon synthetic polypeptide" is a polypeptide capable of producing a triterpenoid hydrocarbon.

A triterpenoid hydrocarbon synthetic polypeptide may be used to produce a triterpenoid hydrocarbon in a host organism. Alternatively, a triterpenoid hydrocarbon synthetic polypeptide may be used to synthesise a triterpenoid hydrocarbon, i.e. in the absence of a host cell.

Any reference herein to "botryococcene synthase" or a "botryococcene synthase polypeptide" is to a polypeptide or amino acid sequence having activity sufficient to synthesise a botryococcene, including a dehydrobotryococcene. In a particular embodiment, botryococcene synthase is capable of converting FPP and/or PSPP to botryococcene. In another embodiment, botryococcene synthase is capable of converting FPP and/or PSPP to dehydrobotryococcene, by limiting, decreasing, minimising or omitting reducing equivalents such as NADPH or NADH from the reaction system. A "botryococcene synthase" includes polypeptide fragments or variants that retain function sufficient to convert FPP and/or PSPP to botryococcene.

"Botryococcene synthase" refers to a protein that belongs to a family of prenyl transferases. Accordingly, specific motifs that are characteristic of prenyl transferases are also present in botryococcene synthase, such as, for example, an aspartate-rich region and/or a substrate-Mg²⁺ binding site.

Methods for identifying amino acid residues necessary for the biochemical activity of a triterpenoid synthetic polypeptide are described in Pandit *et al.*, *J. Biol. Chem.*, 2000, 275, 30610-30617. Methods for altering the amino acid residues of a triterpenoid synthetic polypeptide for structure-function analyses are described in Gu *et al.*, *J. Biol. Chem.*, 1998, 273,

12515–12525. Numerous methods for conducting structure-function analyses of a triterpenoid synthetic polypeptide nucleic acid molecule or polypeptide are referenced in US 2010/0041120. These methods include site directed (oligonucleotide directed) mutagenesis, kits for which are commercially available. Furthermore, methods for structure-functional analyses of a triterpenoid synthetic polypeptide nucleic acid molecule or polypeptide will be known to the person skilled in the art.

If the polypeptide of the second aspect is a botryococcene synthase and the polypeptide is able to produce a squalene and a botryococcene, more botryococcene will be produced than squalene. For example, the ratio of botryococcene to squalene will be at least greater than about 1, or at least greater than about 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10 000, 20 000, 50 000, 100 000, 200 000, 500 000, 1 000 000, 2 000 000, 5 000 000, 10 000 000 or more.

As used herein, “botryococcene synthase gene” refers to a nucleic acid molecule that encodes botryococcene synthase polypeptide, or fragment or variant thereof. Thus, such a gene is often a cDNA sequence that encodes botryococcene synthase. In other embodiments, a botryococcene synthase gene may include sequences, such as introns, that are not present in a cDNA.

The term “botryococcene synthase coding region” refers to the region of the nucleic acid molecule that encodes a mature polypeptide.

A “botryococcene synthase nucleic acid molecule” is a nucleic acid molecule encoding a polypeptide capable of converting FPP and/or PSPP to botryococcene.

The term “botryococcene synthase nucleic acid molecule” refers to a double stranded or single stranded nucleic acid molecule.

Appropriate primers for identifying a botryococcene synthase gene from an organism can be generated from comparisons of the sequences provided herein.

The term “primers” refers to oligonucleotides of 10 to 100 nucleotides, which are used to amplify a gene by the polymerase chain reaction (PCR) as known by one skilled in the art.

Any reference herein to “squalene synthase” or “squalene synthase polypeptide” is to an active or functional polypeptide having activity sufficient to synthesise a squalene, including a dehydrosqualene. In a particular embodiment, squalene synthase is capable of converting FPP and/or PSPP to squalene. In another embodiment, squalene synthase is capable of converting FPP and/or PSPP to a dehydrosqualene. The term includes polypeptide fragments or variants that retain function sufficient to convert FPP and/or PSPP to squalene.

A squalene synthase of the present disclosure does not encompass any one of SEQ ID NOs: 10, 12 or 42.

If a squalene synthase polypeptide is able to produce a squalene and a botryococcene, more squalene will be produced than botryococcene. For example, the ratio of squalene to botryococcene will be at least greater than about 1, or at least greater than about 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10 000, 20 000, 50 000, 100 000, 200 000, 500 000, 1 000 000, 2 000 000, 5 000 000, 10 000 000 or more.

As used herein, "squalene synthase gene" refers to a nucleic acid molecule that encodes a squalene synthase polypeptide, or fragment or variant thereof. Thus, such a gene is often a cDNA sequence that encodes squalene synthase. In other embodiments, a squalene synthase gene may include sequences, such as introns that are not present in a cDNA.

5 The term "squalene synthase coding region" refers to the region of the nucleic acid molecule that encodes the polypeptide.

A "squalene synthase nucleic acid molecule" is a nucleic acid molecule encoding a polypeptide capable of converting FPP and/or PSPP to squalene.

10 The term "squalene synthase nucleic acid molecule" refers to a double stranded or single stranded nucleic acid molecule.

Appropriate primers for identifying and characterising a squalene synthase gene from an organism can be generated from the sequences provided herein.

15 Because botryococcene synthase and squalene synthase are exemplary triterpenoid hydrocarbon synthetic polypeptides, a botryococcene synthase or squalene synthase nucleic acid molecule may have any of the features of a triterpenoid hydrocarbon synthetic nucleic acid molecule. Similarly, a botryococcene synthase or squalene synthase polypeptide may have any of the features of a triterpenoid hydrocarbon synthetic polypeptide.

20 The person skilled in the art will understand that optimal alignment of sequences for comparison may be conducted by the local homology algorithm, by the homology alignment algorithm, by the search for similarity method, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, TFASTA, and DASH), or by inspection.

25 Two nucleic acid molecules or polypeptides are said to be "identical" if the sequence of nucleotides or amino acid residues, respectively, in the two sequences is the same when aligned for maximum correspondence as described below. Optionally, the percentage of sequence identity between a reference sequence and a query sequence is only considered over the length of the query sequence.

30 "Percentage of sequence identity" is determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the nucleic acid or amino acid sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

35 A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions, e.g., 15 or 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

40 In the case of expression of synthetic or recombinant nucleic acid molecules, one of skill in the art will recognise that the inserted nucleic acid molecule need not be identical and may be

“substantially identical” to a sequence of the nucleic acid molecule from which it was derived. As explained, these variants are specifically covered by this term.

The term “substantial identity” in the context of nucleic acid or amino acid sequences means that a nucleic acid molecule or polypeptide comprises a sequence that has at least 55% sequence identity to a reference sequence. Alternatively, percent identity can be any integer from 55% to 100%. Exemplary embodiments include at least: 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or more identity compared to a reference sequence using the programs described herein; preferably BLAST using standard parameters, as described below.

Accordingly, a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or a squalene synthase nucleic acid molecule, includes a nucleic acid molecule that has substantial identity to a nucleic acid molecule encoding a polypeptide comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment thereof, or a nucleic acid molecule that has substantial identity to any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment thereof. As noted above, a triterpenoid hydrocarbon synthetic polypeptide, e.g. a botryococcene synthase or a squalene synthase, includes a polypeptide having substantial identity to any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment thereof.

Polypeptides that are “substantially similar” share sequences as noted above except that residue positions that are not identical may differ by conservative amino acid changes. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains is cysteine and methionine. Exemplary conservative amino acids substitution groups are: valine-leucine-isoleucine; phenylalanine-tyrosine; lysine-arginine; alanine-valine; aspartic acid-glutamic acid; and asparagine-glutamine.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridise to each other, or to a third nucleic acid, under stringent hybridisation conditions.

The triterpenoid hydrocarbon synthetic nucleic acid molecule or polypeptide, e.g. botryococcene synthase or squalene synthase nucleic acid molecule or polypeptide, disclosed herein includes nucleic acid molecules and polypeptides identified and characterised by techniques such as hybridisation and/or sequence analysis using exemplary nucleic acid sequences including any one of SEQ ID NOs: 6 to 9 or 30, or functional fragment thereof, and exemplary polypeptide sequences including any one of SEQ ID NOs: 1 to 5 or 20 to 29, or a functional fragment thereof.

The phrase “stringent hybridisation conditions” refers to conditions under which a probe will hybridise to its target subsequence, typically in a complex mixture of nucleic acid molecules, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridise specifically at higher temperatures. Generally, stringent

conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic acid molecule concentration) at which 50% of the probes complementary to the target hybridise to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent hybridisation conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 M to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). Stringent hybridisation conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridisation, a positive signal is at least two times background, optionally 10 times background hybridisation. Exemplary stringent hybridisation conditions can be as following: 50% formamide, 5X SSC, and 1% SDS, incubating at 42°C, or 5X SSC, 1% SDS, incubating at 65°C, with wash in 0.2X SSC, and 0.1% SDS at 55°C, 60°C, or 65°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

Nucleic acids that do not hybridise to each other under stringent hybridisation conditions are still substantially identical if the polypeptides that they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid molecule is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acid molecule typically hybridises under moderately stringent hybridisation conditions. For example, a botryococcene synthase or squalene synthase nucleic acid molecule can also be identified by its ability to hybridise under moderately stringent hybridisation conditions (e.g., T_m ~40°C) to nucleic acid probes having the sequence of any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment thereof. Such a botryococcene synthase nucleic acid sequence can have, e.g., about 25-30% base pair mismatches or fewer relative to the selected nucleic acid probe. Any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment or variant thereof, comprises an exemplary botryococcene synthase or squalene synthase nucleic acid sequence. Exemplary "moderately stringent hybridisation conditions" include hybridisation in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridisation is at least twice background.

Those of ordinary skill in the art will readily recognise that alternative hybridisation and wash conditions can be utilized to provide conditions of similar stringency.

Appropriate probes for identifying a triterpenoid hydrocarbon synthetic gene, such as a botryococcene synthase or squalene synthase gene, from an organism can be generated from comparisons of the sequences provided herein.

Consequently, a "percentage of sequence identity" and/or hybridisation can aid identification of a homologue, an orthologue or a paralogue of a triterpenoid hydrocarbon synthetic gene or polypeptide disclosed herein.

Isolation or generation of a homologous triterpenoid hydrocarbon synthetic nucleic acid molecule, such as a botryococcene synthase or squalene synthase nucleic acid molecule, can be accomplished by a number of techniques. Cloning and expression of such technique will be

addressed in the context of botryococcene synthase or squalene synthase nucleic acid molecules. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired nucleic acid molecule in a cDNA or genomic DNA library from a desired species. Such a cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned
5 botryococcene synthase or squalene synthase gene, e.g., any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment or variant thereof. Probes may be used to hybridise with genomic DNA or cDNA sequences to isolate homologous genes in the same or different species.

Alternatively, the nucleic acid molecule of interest can be amplified from nucleic acid samples using amplification techniques. For instance, PCR may be used to amplify the sequences of
10 the genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid molecules that code for polypeptides to be expressed, to make nucleic acid molecules to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

As used herein, "expression" of a triterpenoid hydrocarbon synthetic nucleic acid molecule,
15 such as a botryococcene synthase or squalene synthase nucleic acid molecule, refers to introducing a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or squalene synthase nucleic acid molecule, into a cell, e.g., algae, such as green algae, cyanobacteria, photosynthetic or non-photosynthetic bacteria, fungi, including yeasts, or a plant, in which it is not normally expressed. Accordingly, an "increase" in triterpenoid hydrocarbon synthetic activity or
20 expression is generally determined relative to wild type cells, e.g., algae, cyanobacteria, photosynthetic or non-photosynthetic bacteria, fungi, including yeasts, or plants that have no, or low, triterpenoid hydrocarbon synthetic activity.

"Increased" or "enhanced" activity or expression of a triterpenoid hydrocarbon synthetic nucleic acid molecule or polypeptide, e.g. a botryococcene synthase or squalene synthase nucleic
25 acid molecule or polypeptide, refers to a change in triterpenoid hydrocarbon synthetic activity or expression, e.g. botryococcene synthase or squalene synthase activity or expression. Examples of such increased activity or expression include the following: botryococcene synthase expression or activity or squalene synthase expression or activity of a botryococcene synthase nucleic acid molecule or polypeptide or squalene synthase nucleic acid molecule or polypeptide is increased
30 above the level of that in the wild-type, non-transgenic control organism (i.e., the quantity of botryococcene synthase or squalene synthase activity, or expression of botryococcene synthase or squalene synthase nucleic acid molecule, is increased); botryococcene synthase or squalene synthase activity, or expression of a botryococcene synthase or squalene synthase nucleic acid molecule, is in a cell where it is not easily detected in wild-type, non-transgenic cells (i.e., expression
35 of the botryococcene synthase or squalene synthase protein is increased, as is its activity); botryococcene synthase or squalene synthase activity, or botryococcene synthase or squalene synthase nucleic acid molecule expression, is also increased when botryococcene synthase or squalene synthase activity or expression of the botryococcene synthase or squalene synthase nucleic acid molecule is present in a cell for a longer period than in a wild-type, non-recombinant control (i.e.,

duration of botryococcene synthase or squalene synthase activity, or expression of the botryococcene synthase or squalene synthase protein is increased).

The activity or expression of a triterpenoid hydrocarbon synthetic nucleic acid molecule or polypeptide, or the production of a triterpenoid hydrocarbon, is considered to be "increased" or "enhanced" if the activity, expression or production is at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 12.5%, 15%, 17.5%, 20%, 25%, 30%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 82.5%, 85%, 87.5%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 100% or more, for example at least about 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, 10-, 20-, 50-, 100-fold or more, compared with an organism lacking a triterpenoid hydrocarbon synthetic gene or polypeptide, or compared with an organism comprising a wild-type or endogenous triterpenoid hydrocarbon synthetic gene or polypeptide, or in a cell-free system compared to another putative or homologous triterpenoid hydrocarbon synthetic polypeptide.

A "vector" as used herein refers to any vehicle used to transfer genetic material, particularly to transfer exogenous genetic material into a cell.

An "expression vector" refers to a nucleic acid construct, which when introduced into a host cell, results in transcription and/or translation of a RNA or polypeptide, respectively.

The introduction of polynucleotide sequences can be either temporary, e.g. by use of vectors, or permanent, e.g. by integration of the entire vector or a fragment thereof into either the nuclear genome, the plastid genome, or the mitochondrial genome of the host alga.

Suitable vectors for increasing expression of the polynucleotide sequence that encodes the desired polypeptide sequence in an algal cell are known in the art, such as the expression vectors described in US Patent 7,232,679. Such vectors for increasing expression of the polynucleotide sequence are incorporated herein by reference. For example, suitable vectors include pBBR-K-mev-opl6-l, pBBR-K-mev-opl6-2, pDS-mvaA, pDS-idi, pDS-hcs, pDS-mvk, pDS-pmk, pDS-mvd, pDS-His-mvaA, pDS-His-idi, pDS-His-hcs, pDS-His-mvk, pDS-His-pmk, pDS-His-mvd, pBBR-K-Zea4, pBBR-K-Zea4-up, pBBR-K-Zea4-down, pBBR-K-PcrtE-crtE-3, pBBR-tK-PcrtE-mvaA, pBBR-tK-PcrtE-idi, pBBR-tK-PcrtE-hcs, pBBR-tK-PcrtE-mvk, pBBR-tK-PcrtE-pmk, pBBR-tK-PcrtE-mvd, pBBR-K-PcrtE-mvaA-crtE-3, pDS-His-phaA, pBBR-K-PcrtE-crtW, pBBR-K-PcrtE-crtWZ, pBBR-K-PcrtE-crtZW, and combinations thereof.

A vector comprising a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. botryococcene synthase nucleic acid molecule or squalene synthase nucleic acid molecule, may comprise a marker gene that confers a selectable phenotype on algae, plant or bacterial cells. For example, the marker may encode antibiotic resistance, such as resistance to kanamycin, G418, phleomycin (ZeocinTM), bleomycin, hygromycin, and the like. In one embodiment, a selectable marker for use in *Chlamydomonas* can be a marker that provides spectinomycin resistance, kanamycin and amikacin resistance, zeomycin, bleomycin or phleomycin (ZeocinTM) resistance, paramomycin, or neomycin resistance.

Selectable markers for host organisms are well known in the art.

A "regulatory element" includes a promoter, which may be either constitutive or inducible. In one embodiment, a promoter can be used to direct expression of botryococcene synthase nucleic

acid molecules under the influence of changing environmental conditions. Examples of environmental conditions that may effect transcription via an inducible promoter include anaerobic conditions, elevated temperature, or the presence of light. Promoters that are inducible upon exposure to chemical reagents may also be used to express a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or a squalene synthase nucleic acid molecule.

Other useful inducible regulatory elements include copper-inducible regulatory elements, tetracycline and chlor-tetracycline-inducible regulatory elements, ecdysone inducible regulatory elements, heat shock inducible regulatory elements, and *lac* operon elements, which are used in combination with a constitutively expressed *lac* repressor to confer, for example, IPTG-inducible expression. An inducible regulatory element also can be, for example, a nitrate-inducible promoter, e.g., derived from the spinach nitrite reductase gene, or a light-inducible promoter, such as that associated with the small subunit of RuBP carboxylase or the LHCP gene families, or a light.

The term "promoter" or "regulatory element" refers to a region or sequence determinants located upstream or downstream from the start of transcription that direct transcription. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal elements, which can be located as much as several thousand base pairs from the start site of transcription.

A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions.

An "inducible" promoter is a promoter that is active under environmental or developmental regulation.

The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter) and a second nucleic acid sequence, such as a botryococcene synthase or squalene synthase gene, wherein the expression control sequence directs transcription of the nucleic acid molecule corresponding to the second sequence.

An "algae promoter", "bacterial promoter", "fungal promoter", "yeast promoter" or "plant promoter" is a promoter capable of initiating transcription in algae, bacteria, fungi, including yeasts, and/or plant cells, respectively. Such a promoter is therefore active in algae, cyanobacteria, bacterial, or plant cells, but need not originate from that organism. It is understood that limited modifications can be made without destroying the biological function of a regulatory element and that such limited modifications can result in regulatory elements that have substantially equivalent or enhanced function as compared to a wild type regulatory element. These modifications can be deliberate, as through site-directed mutagenesis, or can be accidental such as through mutation in hosts harbouring the regulatory element. All such modified nucleotide sequences are included in the definition of a regulatory element as long as the ability to confer expression is substantially retained.

In one example, a promoter sequence that is responsive to light may be used to drive expression of a botryococcene synthase nucleic acid construct that is introduced into *Chlamydomonas* that is exposed to light. Other light-inducible promoter systems may also be used, such as the phytochrome/ PIF3 system. Further, a promoter can be used that is also responsive to

heat can be employed to drive expression in algae such as *Chlamydomonas*. Additional promoters, e.g., for expression in algae such as green algae, include the RbcS2 and Psad promoters.

In one embodiment, the promoter may be from a gene associated with photosynthesis in the species to be transformed or another species. For example such a promoter from one species may be used to direct expression of a protein in transformed algal cells or cells of another photosynthetic marine organism. Suitable promoters may be isolated from or synthesized based on known sequences from other photosynthetic organisms. Preferred promoters are those for genes from other photosynthetic species that are homologous to the photosynthetic genes of the algal host to be transformed. For example, a series of light harvesting promoters from the fucoxanthin chlorophyll binding protein have been identified in *Phaeodactylum tricornutum*. In other embodiments, a carotenoid chlorophyll binding protein promoter, such as that of peridinin chlorophyll binding protein, can be used.

In one embodiment, a promoter used to drive expression of a heterologous triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or squalene synthase nucleic acid molecule, is a constitutive promoter. Examples of constitutive strong promoters for use in algae include, e.g., the promoters of the *atpA*, *atpB*, and *rbcL* genes. Various promoters that are active in cyanobacteria are also known. These include promoters such as the (constitutive) promoter of the *psbA3* gene in cyanobacteria. Other promoters that are operative in plants, e.g., promoters derived from plant viruses, such as the CaMV35S promoters, can also be employed in algae.

Promoter sequences for an algal cell are preferably isolated from an algal species or a closely related organism. Promoters that are functional in higher plants are less preferred except for groups of algae closely related to higher plants. For example, the 35S CaMV promoter, which is active in many plant species, is completely inactive in *Chlamydomonas* (Day et al. (1990) *Physiol. Plantarum* 78:254-260).

Examples of suitable promoters include hydrogenase promoters, Cytochrome C6 (*Cyc6*) promoter, *Nia1* promoter, *CabII-1* promoter, *Ca1* promoter, *Ca2* promoter, coproxy oxidase promoter, algal ribulose biphosphate carboxylase small subunit gene (*SSU*) promoter, and algal pyruvate kinase promoter. Additional suitable promoters include the arylsulfatase promoter, and the aminoglycoside 3'-phosphotransferase gene (*aph VIII*) promoter from the multicellular green alga *Volvox*, and *RbcS2* promoter which has been widely used to drive gene expression in the nucleus of *C. reinhardtii*.

Another suitable promoter for use in *C. reinhardtii* is the Hsp70A-RbcS2 hybrid promoter (e.g. in pBC1 and other plasmids) as disclosed in Schroda et al. *The Plant Journal*, 2000, 21, 121-131 and Schroda et al., *The Plant Journal*, 2002, 31, 445-455.

In one embodiment, promoters are identified by analysing the 5' sequences of a genomic clone corresponding to a triterpenoid hydrocarbon synthetic gene, e.g. a botryococcene synthase or squalene synthase gene. Sequences characteristic of promoter sequences can be used to identify the promoter.

A promoter can be evaluated, e.g., by testing the ability of the promoter to drive expression in plant cells, e.g., green algae, in which it is desirable to introduce a triterpenoid hydrocarbon synthetic construct, such as a botryococcene synthase or squalene synthase expression construct.

Examples of suitable regulatory elements, e.g. enhancer elements, include: EE-1 and EE-2 described in Kucho *et al.* (*Plant Physiol* 2003, 133(2):783-93); the GCC-box enhancer element described in Wu *et al.* (*Mol Genet Genomics* 2001, 265(5):763-70); and those described in Fischer, *et al.* (*Mol Genet Genomics* 2001, 265(5):888-94) regarding flanking regions of PsdD.

Enhancer elements suitable for transgene expression in *C. reinhardtii* are also described in Lumbreras *et al.*, *The Plant Journal*, 1998, 14, 441-447 and Colombo *et al.*, *Funct. Plant Biol.*, 2002, 29, 231-241.

"*Botryococcus braunii*" or "*B. braunii*" refers to a colonial green alga of the order Chlorococcales (class Chlorophyceae). Certain strains of *B. braunii* synthesise and accumulate an unusually high level of lipids or triterpenoid hydrocarbons that are typically around 30-40 percent, or more, of their dry weight.

As used herein, "strain" refers to an organism belonging to the same species, but displaying at least one different characteristic compared with a second organism of the same species. The different characteristic can be caused by differences in the genomes of the strains under comparison. Thus, a recombinant organism is also a "strain" of the host organism. Specific types of strains are referred to herein as "races". "Race A" refers to strains of *B. braunii* whose major lipid hydrocarbon is "alkadiene" and "alkatriene" (derivatives of fatty acids) having the formula C_nH_{2n-10} , wherein n is an odd number 23 to 31. "Race B" refers to strains of *B. braunii* whose major lipid hydrocarbon is "botryococcene". "Race L" refers to strains of *B. braunii* whose major lipid hydrocarbon is "lycopadiene".

As used herein, "organism" refers to an individual living system.

A "host organism" refers to a first organism that comprises exogenous DNA or RNA, or a polypeptide, which is derived from a second organism other than the first organism.

The host organism is a "non-human" organism, which excludes vertebrate organisms.

The host organism may be a a single cell organism, a filamentous organism, a higher plant or a macroalgae such as a seaweed, for example Pacific giant kelp. Pacific giant kelp is a species of marine alga found on the Pacific coast of North and Central America. It begins life as a microscopic spore, but may grow up to 60 m with the upper fronds forming a dense canopy at the surface. Kelp fronds may grow at 300 mm to 600 mm per day which makes it ideal to produce large amounts of triterpenoid hydrocarbons in a short period of time. See, for example, Copping, *et al.*, 2008, Techno-Economic Feasibility Analysis of Offshore Seaweed Farming for Bioenergy and Biobased products, Independent Research and Development Report, Battelle Pacific Northwest Division, incorporated herein by reference.

In a particular embodiment, the host organism is a microorganism.

As used herein, "microorganism" refers to microalgae, bacteria, fungi, including yeast, cyanobacteria, diatoms, photosynthetic bacteria or non-photosynthetic bacteria, or Euglenids.

In one embodiment, an alga is used. "Alga", "algae", "microalga", "microalgae" or the like, refer to plants belonging to the subphylum Algae of the phylum *Thallophyta*. The algae are unicellular, photosynthetic, oxygenic algae and are non-parasitic plants without roots, stems or leaves; they contain chlorophyll and have a great variety in size, from microscopic to large seaweeds. Green algae, which are single cell eukaryotic organisms of oxygenic photosynthesis, endowed with chlorophyll a and chlorophyll b, belonging to Eukaryota—Viridiplantae—Chlorophyta—Chlorophyceae, may be used. However, algae may also be blue-green, red, or brown, so long as the algae are able to host a nucleic acid molecule of the first aspect or a vector of the third aspect to produce a triterpenoid hydrocarbon such as botryococcene or squalene.

In one embodiment, a green alga, cyanobacterium (blue-green algae), or a diatom is used. Examples of green algae include: *Botryococcus* sp. other than that from which botryococcene synthase was isolated, such as *B. braunii* Race A or Race L, or *Botryococcus sudeticus*; *Chlamydomonas* sp., e.g. *C. reinhardtii*, which is classified as Volvocales—Chlamydomonadaceae; *Chlorella* sp., e.g. *Chlorella vulgaris* or NC64A; *Scenedesmus obliquus*; *Dunaliella salina*; or *Haematococcus pluvialis*. Other examples include *Choricystis* sp., e.g. *Ch. minor* and *Ch. chodatii*. Examples of cyanobacteria include: *Microcystis* sp.; *Synechocystis* sp.; or *Spirulina platensis*. *Navicula tennerima* is an example of a diatom species that may be used.

Alternatively, the alga may be *Volvox carteri*, *Ostreococcus tauri*, *Ostreococcus lucimarinus*, *Thalassiosira pseudonana*, *Pheodactylum tricornutum*, or *Aureococcus anophagefferens*.

In another embodiment, a euglenid may be used, e.g. *Euglena* sp.

In yet another embodiment, a bacterium may be used. The bacterium may be photosynthetic, e.g. *Rhodospirillum rubrum*, or non- photosynthetic, e.g. *E. coli*.

In an alternative embodiment, another photosynthetic organism may be used, e.g. a plant cell. Plant cells that may be suitable include those of *Nicotiana tabacum* or *Arabidopsis thaliana*.

In one embodiment, a fungus may be used. For example, a yeast may be used, e.g. *S. cerevisiae* or *Pichia* sp.

As used herein, "recombinant organism" or "transgenic organism" refers to a second organism, or "host organism", that harbours genetic material derived from a first, or donor, source. "Recombinant organism" refers to any organism, such as an alga, bacterium, fungus, including a yeast, or plant, in which an exogenous nucleic acid molecule is expressed. More specifically, the term refers to any transformed alga, bacterium, fungus, including a yeast, or plant that displays the desired characteristics of both the donor and host organisms. The desired characteristic of the donor organism is a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or squalene synthase nucleic acid molecule, or the desired characteristic is a triterpenoid hydrocarbon synthetic polypeptide, e.g. a botryococcene synthase or a squalene synthase polypeptide, either of which contributes in the host organism to high triterpenoid hydrocarbon content similar to the donor organism. The desired characteristics of the host organism include stability and/or fast growth. Therefore, a "recombinant organism" will synthesise levels of triterpenoid hydrocarbon in excess of the host organism and more similar to those of the donor strain. A desired "recombinant organism" will also be more stable and/or grow faster than the donor organism.

As used herein, "recombinant" describes a second, or host species of genetic material that harbours genetic material derived from a first, or donor, source. Thus, the host harbours exogenous genetic material, e.g. DNA or RNA. "Recombinant" is used interchangeably with "genetically manipulated" or "transgenic".

5 As used herein, "genetic manipulation" or "molecular biotechnology" refers to any human-exerted artificial process that is performed on DNA or RNA, including, but not limited to, disruption, isolation, replication, modification and introduction into a different organism. An organism that has undergone genetic manipulation is also referred to herein as "recombinant".

10 As used herein, "exogenous" refers to a molecule, e.g. DNA, RNA or compound, originating externally to an organism. With respect to DNA or RNA, the term can be used interchangeably with "heterologous".

Conversely, as used herein, "endogenous" refers to a molecule, e.g. DNA, RNA or compound, originating within an organism.

15 The term "heterologous" refers to a host, e.g. a strain, species, or genus, not normally possessing a gene that is transformed into the host, or refers to a gene not normally present in a host, e.g. a strain, species, or genus, artificially endowed with the gene. The term "heterologous" is used interconvertibly with the term "exogenous", in the case of an exogenous gene.

20 A nucleic acid molecule is "heterologous to" a second nucleic acid molecule if it originates from a foreign species, or, if from the same species, is modified by human action from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

25 This disclosure contemplates a recombinant organism or a host organism further transformed with a nucleic acid molecule or vector encoding, in addition to the triterpenoid hydrocarbon synthetic polypeptide, a polypeptide useful in synthesis of a triterpenoid hydrocarbon such as a botryococcene. For example, the recombinant organism may also comprise an FPP synthase or a methyltransferase such as a botryococcene methyltransferase or a squalene methyltransferase.

30 As used herein, "transform", and variations thereof, refers to the process of introducing exogenous DNA or RNA from a donor organism into a host organism. The term is not used in a limiting sense and includes any mechanism by which foreign DNA can be introduced into a host organism. Thus, "transform" includes, but is not limited to, electroporation, transfection, gene guns, viral methods, and glass beads methods for example. The term "transformed" essentially unites a gene and organism not previously containing that gene, and refers to the artificial conjunction of the
35 organism and the gene by genetic modification of the organism by a human being, using processes other than sexual conjugation of organisms. After transformation, function of the gene may or may not occur in the new organism.

40 In algae, the nuclear, mitochondrial, and chloroplast genomes are transformed through a variety of known methods, including by microparticle bombardment (e.g. DNA-coated gold particles), or using a glass bead method.

In one embodiment, a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or squalene synthase nucleic acid molecule, is introduced into and/or targeted to the chloroplast of an alga. In another embodiment, a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or squalene synthase nucleic acid molecule, is introduced into and/or targeted to the nucleus.

To use isolated sequences in the above techniques, synthetic or recombinant DNA vectors suitable for transformation of green algae, other eukaryotic algae, cyanobacteria, photosynthetic or non-photosynthetic bacteria, fungi, including yeasts, or plant cells are prepared.

Techniques for transformation are well known in the art and described in the technical and scientific literature. For example, a DNA sequence encoding a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. a botryococcene synthase or squalene synthase nucleic acid molecule, can be combined with transcriptional and other regulatory sequences which will direct the transcription of the nucleic acid molecule in the intended cells of the transformed alga, cyanobacterium, bacterium, fungus, including a yeast, or plant. In one embodiment, an expression vector that comprises the botryococcene synthase or squalene synthase nucleic acid molecule further comprises a promoter operably linked to the botryococcene synthase or squalene synthase nucleic acid molecule. In other embodiments, a promoter and/ or other regulatory elements that direct transcription of the botryococcene synthase or squalene synthase nucleic acid molecule are endogenous to the organism and the vector comprising the botryococcene synthase or squalene synthase nucleic acid molecule is introduced, e.g., by homologous recombination, such that the heterologous botryococcene synthase or squalene synthase nucleic acid molecule is operably linked to an endogenous promoter and expression is driven by the endogenous promoter.

The present disclosure employs various routine synthetic or recombinant nucleic acid techniques. Generally, the nomenclature and the laboratory procedures in synthetic or recombinant DNA technology described are those well known and commonly employed in the art.

As used herein, "biofuel" refers to fuel derived from carbon sources that can be replenished and make reduced, minimal or no net contribution to atmospheric carbon dioxide levels. More specifically, biofuel refers to carbon sources derived from an organism that expresses a triterpenoid hydrocarbon synthetic gene, and the lipid or hydrocarbon can be used as fuel or to produce fuel. To produce biofuel, it may be necessary for hydrolytic cracking or hydrocracking of the triterpenoid hydrocarbon. Biofuel is distinct from traditional "fossil" fuels.

A triterpenoid hydrocarbon produced according to this disclosure may be combusted directly as a fuel. However for performance in internal combustion engines, the triterpenoid hydrocarbon should be modified by processes such as pyrolysis and catalytic cracking. Crude triterpenoid hydrocarbons can be converted to petrol (gasoline), light cycle oil, heavy cycle oil, and coke by catalytic cracking as has been taught by Kitazato *et al.*, Sekiyu Gakkaishi 32:28. The yields of petrol (gasoline) obtained by catalytic cracking of algal hydrocarbons are comparable to yields obtained from petroleum. Also the gasoline produced has sufficiently high octane numbers for direct use in transportation vehicles. Aviation fuel may also be obtained using processes known in the art.

Similarly, "bioplastic" refers to plastic derived from carbon sources that can be replenished and make minimal or no net contribution to atmospheric carbon dioxide levels. More specifically, bioplastic refers to carbon sources derived from an organism that expresses a triterpenoid hydrocarbon synthetic gene, and the triterpenoid hydrocarbon can be used to produce plastic.

5 The disclosure contemplates application to any product traditionally derived from petrochemical sources of fossil fuels to which the triterpenoid hydrocarbon is amenable. For example, in addition to biofuel and bioplastic, it is envisaged that the disclosure will be useful for, but not limited to, production of pharmaceuticals, food additives, industrial chemicals, specialty chemicals, detergents, fertilizers, medicines, paints, plastics, synthetic fibres, or synthetic rubber.

10 In an embodiment of the method of the sixth aspect, the triterpenoid hydrocarbon is squalene or botryococcene and the conditions comprise supplementing the recombinant organism with NADPH or NADH. In another embodiment, the triterpenoid hydrocarbon is dehydrosqualene or dehydrobotryococcene and the conditions comprise not supplementing the recombinant organism with, or minimising the concentration of, NADPH or NADH. By limiting, decreasing, minimising or
15 omitting reducing equivalents such as NADPH or NADH in the reaction system, the method provides for the production of a dehydrosqualene or a dehydrobotryococcene, which each lack at least two hydrogen atoms when compared with squalene or botryococcene, respectively. Dehydrosqualene and/or dehydrobotryococcene can be identified using GC-MS.

A triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. botryococcene synthase or
20 squalene synthase nucleic acid molecule, can be expressed in any number of algae, e.g., green algae, other eukaryotic algae, cyanobacteria, or photosynthetic or non-photosynthetic bacteria, fungi, including yeasts, or plant cells where it is desirable to produce a triterpenoid hydrocarbon, e.g. botryococcene or squalene. Transformed algae, cyanobacteria, bacteria (photosynthetic bacteria or non-photosynthetic bacteria), or fungi, including yeasts, that express a heterologous triterpenoid
25 hydrocarbon synthetic nucleic acid molecule, e.g. botryococcene synthase or squalene synthase nucleic acid molecule, may be grown under mass culture conditions for the production of the triterpenoid hydrocarbon, e.g. botryococcene or squalene. The transformed or recombinant organisms are grown in bioreactors or fermentors that provide an enclosed environment to contain the triterpenoid hydrocarbon, e.g. botryococcene or squalene. In certain embodiments for mass
30 culture, the algae, cyanobacteria, bacteria, or fungi, including yeasts, are grown in enclosed reactors in quantities of at least about 100 litres, 200 litres, 500 litres, at least about 1000 litres, 2000 litres, 5000 litres, 10,000 litres, 20,000 litres, 50,000 litres, 100,000 litres, 200,000 litres, 500,000 litres or greater, and in one embodiment, in quantities of about 1,000,000 litres or more, even 1,500,000 litres, or more.

35 As used herein, "mass-culturing" refers to growing large quantities of algae, cyanobacteria, photosynthetic or non-photosynthetic bacteria, or fungi, including yeasts, that have been modified to express a triterpenoid hydrocarbon synthetic nucleic acid molecule, e.g. botryococcene synthase or squalene synthase. A "large quantity" is generally in the range of about 10 litres or more, 20 litres, 50 litres, or about 100 litres to about 1,500,000 litres, for example, 200 litres, 500 litres, 1000 litres, 2000
40 litres, 5000 litres, 10,000 litres, 20,000 litres, 50,000 litres, 100,000 litres, 200,000 litres, 500,000

litres, 1,000,000 litres, even 1,500,000 litres, or more. In one embodiment, the organism may be cultured in large quantities in modular bioreactors, each having a capacity of about 1,000 to about 1,000,000 litres, for example, 200 litres, 500 litres, 1000 litres, 2000 litres, 5000 litres, 10,000 litres, 20,000 litres, 50,000 litres, 100,000 litres, 200,000 litres, or 500,000 litres. In one embodiment, a recombinant plant may be amenable to mass culture.

A "bioreactor" refers to any enclosed large-capacity vessel in which microalgae, cyanobacteria or photosynthetic or non-photosynthetic bacteria, or fungi, including yeasts, are grown. A "large-capacity vessel" can hold about 10 litres, or about 500 litres to about 1,000,000 litres, or more, for example, 20 litres, 50 litres, 100 litres, 200 litres, 500 litres, 1000 litres, 2000 litres, 5000 litres, 10,000 litres, 20,000 litres, 50,000 litres, 100,000 litres, 200,000 litres, or 500,000 litres, or more.

Methods of mass-culturing algae are known. For example, algae can be grown in high density photobioreactors, bioreactors such as those for sewage and waste water treatments, mass-cultured for the elimination of heavy metals from contaminated water, mass-cultured for the production of β -carotene, hydrogen, and pharmaceutical compounds, as well as nutritional supplements for both humans and animals and for the production of other compounds of nutritional value.

Conditions for growing triterpenoid hydrocarbon-expressing algae, cyanobacteria, photosynthetic or non-photosynthetic bacteria, or fungi, including yeasts, e.g. botryococcene synthase-expressing or squalene synthase-expressing algae or bacteria, are known in the art. For example, methods for growing *Botryococcus* are disclosed in WO 2006/121950.

Alternatively, a triterpenoid hydrocarbon may be produced in a cell-free system. Examples of cell-free systems are disclosed herein at Example 10 and Example 16, and in US 2010/0041120, Okada *et al. Arch Biochem Biophys* 2000, 373(2): 307-317, and Okada *et al. Arch Biochem Biophys* 2004, 422: 110-118, which are incorporated herein by reference.

Triterpenoid hydrocarbon, e.g. botryococcene or squalene, produced by the genetically modified organisms disclosed herein is not currently known in the art, but could be harvested using known techniques. As used herein, to "harvest" a triterpenoid hydrocarbon, e.g. botryococcene or squalene, refers to capturing and sequestering the triterpenoid hydrocarbon, e.g. botryococcene or squalene, in a closed or contained environment.

Harvesting may include flocculating the microorganism, e.g. alga, by adjusting the pH of the medium in which the organism resides. Botryococcenes and squalene are not miscible in water and they may rise to and float at the surface of the growth medium if released by the organism. If so, they may be siphoned off from the surface and sequestered in suitable containers.

Alternatively, destructive harvesting of the triterpenoid hydrocarbon, e.g. botryococcene or squalene, from the recombinant organism, such as algae (including seaweed) or bacteria, may be required, for example pressing, solvent extraction or supercritical fluid extraction. Under suitable conditions, 70% of hydrocarbons can be released by contact with hexane for 30 minutes. Hydrocarbons can be continuously extracted with proper agitation and recovery methodologies known

in the art. Supercritical fluids such as carbon dioxide (CO₂) allow rapid extraction of hydrocarbons, and are nontoxic, inexpensive, easily separated from the extracted hydrocarbons and are reusable.

Also disclosed herein is a business method comprising providing a carbon credit to a party growing a genetically modified organism, comprising a nucleic acid molecule according to the first aspect, adapted to produce a biofuel, a pharmaceutical, a food additive, an industrial chemical, a specialty chemical, a detergent, a fertilizer, a medicine, a paint, a plastic, a synthetic fibre, or a synthetic rubber. As such, the methods and compositions described herein may be used in a business method in exchange for carbon credits.

Carbon credits may be an allowance, permit, credit, or the like which are or have been allowed, authorised, or recognised by some relevant sovereign entity (such as, but not limited to, a city (including municipalities of all sizes and types including both incorporated and unincorporated municipalities), a county, a state or province, or a nation, as well as related governmental entities such regional, multi-national, or other international bodies such as the United Nations, the European Union, or the United Arab Emirates).

The carbon credit may be substantially received directly from a regulatory agency or administrative entity. In other instances, they may be received indirectly, for example, an entity using the methods or compositions herein may receive the carbon credits directly from a regulatory agency, and may then transfer the carbon credits to another entity. Transfer of the carbon credit may be in association with a given process or product using the genetically modified organism.

For example, a first entity may be identified that provides a consumable product that is distributed for consumption in an end-user mobile platform, wherein the consumption and/or production of the consumable product includes a corresponding resultant emission. For example, combustion of diesel fuel often results in the environmental release of corresponding nitrogen oxides and combustion of gasoline often results in the environmental release of corresponding sulfur oxide.

The first party may adopt a method of producing its product using the genetically modified organism, or use the product generated by the genetically modified organisms in their compositions, resulting in less harmful effects on the environment than conventional methods of generating, for example, diesel fuel, thus off-setting the environmental effects of the end product. The first party may then receive a carbon, or emission, credit as a result of a reduction of the total emission. The carbon credit may be received from a regulatory or administrative agency, or may be transferred to the first party from a second party, wherein the second party may have sold the genetically modified organism or the product of the genetically modified organism to the first party.

The carbon credit may be exchanged for a substantially liquid monetary instrument. For example, the carbon credit may be exchanged for a cash equivalent, such as cash, cheque, and the like. The carbon credit may also be exchanged for a legal grant regarding an intellectual property right, for example, but not limited to, an assignment or a licence. The carbon credit may also be exchanged for a government tax subsidy or access to purchasers of a given market. The carbon credit may also be exchanged for use of another carbon emission process, such as one not comprising growing the organism. For example, a party may have a limited number of emissions it may release in a time period, for example, a month or a year, and going over the limit may incur fines

and penalties. However, with carbon credits, the party going over the limit may exchange carbon credits to offset the fines or penalties or may be taken into account when determining the amount of emissions generated by the party.

The business method can also involve the production of a product other than a biofuel, for example a pharmaceutical, a food additive, an industrial chemical, a specialty chemical, a detergent, a fertilizer, a medicine, a paint, a plastic, a synthetic fibre, or a synthetic rubber. A business method associated with a biofuel, including those involving the use of carbon credits, are also relevant to the production of other types of useful products and materials.

It will be understood to persons skilled in the art of the invention that many modifications may be made without departing from the spirit and scope of the invention.

In the claims which follow and in the preceding description of the invention, except where the context requires otherwise due to express language or necessary implication, the word "comprise" or variations such as "comprises" or "comprising" is used in an inclusive sense, i.e. to specify the presence of the stated features but not to preclude the presence or addition of further features in various embodiments of the invention.

It is to be understood that, if any prior art publication is referred to herein, such reference does not constitute an admission that the publication forms a part of the common general knowledge in the art, in Australia or any other country.

Pursuant to the Budapest Treaty, plasmids pUC57/ATB1 (construct 1, Example 12), pET11/ATB1 (construct 2, Example 12), pSP124S (Example 11), and pBC1 were deposited on 14 September 2010 with the National Measurement Institute (NMI), Australia in its capacity as an International Depositary Authority. Deposited plasmid pUC57/ATB1 was assigned the accession number V10/022706. Deposited plasmid pET11/ATB1 was assigned the accession number V10/022707. Deposited plasmid pSP124S was assigned the accession number V10/022708. Deposited plasmid was assigned the accession number V10/022709.

EXAMPLES

Example 1 – Culture Conditions

The *B. braunii* strain Ayamé 1 (Race B) was cultured under an atmosphere of air enriched by 1% CO₂, 2% CO₂ or up to 10% CO₂ or more CO₂, using an air-lift system, with continuous illumination, and temperature around 25°C. The growth medium, was a modified CHU 13 (KNO₃ 200 mg/L, MgSO₄•7H₂O 100 mg/L, K₂HPO₄ 39 mg/L, CaCl₂•2H₂O 54 mg/L, FeNaEDTA 15 mg/L, H₃BO₃ 2.68 mg/L, MnSO₄•H₂O 1.54 mg/L, ZnSO₄•7H₂O 0.22 mg/L, CuSO₄•5H₂O 0.08 mg/L, Na₂MoO₄•2H₂O 0.06 mg/L, CoSO₄•7H₂O 0.09 mg/L) medium at pH 7.5. The tubes for illumination were from Mazda (SF 36w AZU/965; length: ~120 cm). The illumination intensity began at ~40 microeinsteins, and increased progressively up to ~120-160 microeinsteins (in the stationary phase), or more when culturing in a larger container (10 litres or greater).

Example 2 – Genomic DNA Preparation

DNA was extracted from the Ayamé 1 strain of Example 1 as follows. First, approximately 2 x 200 milligrams of cell pellet was obtained by flocculation of cells in 50% ethanol (final concentration). Each cell pellet was resuspended in 2 mL of high salt DNA extraction buffer (200 mM Tris.HCl pH 7.5, 100 mM EDTA pH 8.0, 0.5 M NaCl) and subjected to 10 rounds of freeze-thaw using liquid nitrogen and a 45°C water bath. Proteinase K was then added to a final concentration of 2 mg/mL and the mix was incubated at 55°C for one hour, then 60°C overnight. The resulting lysate was used whole in a Qiagen DNeasy Plant Maxi Kit (cat. no. 68163), to extract the DNA, according to the manufacturer's instructions.

Example 3 - DNA Sequencing

The genome of *B. braunii* has not been sequenced and therefore is not known. Hence, it was necessary for the genome of *B. braunii* to be determined.

Twenty micrograms of the genomic DNA from *B. braunii* Ayamé 1 strain (Example 2) was used for genomic sequencing on an Illumina GenomeAnalyser machine (Geneworks, South Australia), to a total of 3 giga bases of data (comprising reads of ~35 bases long each). Another 20 µg of this DNA was used for genomic sequencing on a Roche FLX machine (Australian Genome Research Facility (AGRF), Queensland) to a total of 3 giga bases of data (comprising reads of ~250 bases long each).

Another twenty micrograms of the genomic DNA from *B. braunii* Ayamé 1 strain (Example 1) was used for genomic sequencing on an Illumina GenomeAnalyser machine (Illumina, Hayward California USA), to a total of 5 giga bases of data (comprising reads of ~75 bases long each).

Example 4 - Genome Assembly and Gene Identification

Preparatory assembly of DNA contigs by merging of overlapping raw sequence data (Example 3) was performed at the AGRF. Subsequently, data from the Illumina platform was integrated at Flinders University using a variety of algorithms such as EULER, VELVET, CLC Genomics Workbench, DASH, to extend contigs and establish additional contigs.

Using BLAST or DASH algorithms, putative genes were identified and the genome will be assembled.

Exons encoded by amino acid sequences provided as SEQ ID NOs: 1 to 5 or 20 to 28 were identified according to this example. The full polypeptide sequence (*ATB1*) provided as SEQ ID NO: 29 was identified according to this example.

Example 5 – RNA Preparation

Cells of *B. braunii* Ayamé 1 strain were grown to stationary phase in CHU 13 medium (Example 1), then subcultured into an equal volume of fresh CHU 13 medium, and grown further for three days. On the third day, at the same time of day as the inoculation occurred three days beforehand, 1-10 mg cells were harvested using a 0.45 µm pore-size filter and scraping the cells from the filter using a scalpel blade. RNA was extracted from these cells using a Qiagen RNeasy Plant Mini Kit (cat. no. 74903). Extraction was begun by resuspending the cells in "buffer RLT" of that kit,

and subjecting the suspension to at least 5 cycles of freeze-thaw using liquid nitrogen and 45°C water bath, respectively. Subsequently, RNA was isolated according to the manufacturer's instructions for plant cell RNA extraction.

5 **Example 6 – Host Cell Species (Gene Recipient)**

A. Algae

The disclosure is not limited to algae, but in one embodiment the host cell is an alga. The desirable criteria for host algae include ease and rapidity of growth, and flagellate morphology, because flagellate algae have positive phototaxis and promote ease of harvest. Given these criteria, the non-limiting example of species listed below may be gene recipients. Alternatively, the alga may be a seaweed, for example Pacific giant kelp.

1. Algae

i. ***Botryococcus*** sp. other than that from which botryococcene synthase was isolated (Race B, Ayamé 1) and which satisfy at least one of the above criteria may be used according to the disclosure. Some examples include *B. braunii* Race A or Race L, or *Botryococcus sudeticus*.

ii. ***Chlamydomonas*** sp. is a genus of green alga. They are unicellular flagellates and used as a model organism for molecular biology, especially studies of flagellar motility and chloroplast dynamics, biogenesis, and genetics. One example is *C. reinhardtii*.

iii. ***Chlorella*** sp. is a genus of single-celled green algae. It is spherical in shape, about 2 to 10 µm in diameter, and lacks flagella. Through photosynthesis it multiplies rapidly requiring only carbon dioxide, water, sunlight, and small amounts of minerals to reproduce. One example is *Chlorella vulgaris*.

iv. Other green algal species that may be used according to the disclosure include *Scenedesmus obliquus*, *Dunaliella salina*, or *Haematococcus pluvialis*.

v. ***Navicula tennerima*** is a diatom species that has the potential for heterotrophic culture.

vi. Any alga of the dinophyceae, haptophyceae, stramenopiles, alveolates, cryptophyceae, euglenophyceae, or rhodophyceae may be used as hosts.

vii. ***Choricystis*** sp. is a genus of freshwater microalga with typical characteristic features comprising: small size of the cells; lack of mucilage; lack of pyrenoids and autospore formation with prevalent formation of two autospores. Two examples are *Ch. minor* and *Ch. chodatii*.

2. Cyanobacteria

i. ***Microcystis*** sp. is a type of blue-green algae, also known as cyanobacteria, found in fresh water. Cyanobacteria are prokaryotic (meaning they possess no nucleus) and can be found in freshwater, marine water, or brackish water. *Microcystis* grows well in nutrient-rich water.

ii. Other cyanobacteria that may be used according to the disclosure include *Spirulina platensis* or *Synechocystis* sp.

B. Euglenids

Euglena sp. is typical of the euglenids, and is commonly found in nutrient-rich fresh water. The cells vary in length from around 20 to 300 µm, and are typically cylindrical, oval, or spindle-

shaped with a single emergent flagellum used for movement. If sunlight is not available, it can absorb nutrients from decayed organic material. *Euglena* is also found in sewage systems. *Euglena* reproduces through binary fission, a process in which one cell is divided into two. *Euglena* has a contractile vacuole which prevents the cell from bursting.

C. Bacteria

1. Photosynthetic Bacteria

Rhodospirillum rubrum is a photosynthetic bacterium that may be a host cell according to the disclosure.

2. Non-Photosynthetic Bacteria

E. coli is a non- photosynthetic bacterium that is commonly found in the lower intestine of warm-blooded animals and may be a host cell according to the disclosure. Some strains possess flagella and are motile, whereas other strains lack flagella.

D. Other Photosynthetic Organisms

Nicotiana tobaccum, an example of a tobacco plant, is one example of a photosynthetic organism that may be used according to the disclosure. Another example may be *Arabidopsis thaliana*.

E. Yeast

Fungi may also be a suitable host cell, for example a yeast such as *S. cerevisiae* or *Pichia* sp.

Example 7 – Expression Vector

A nucleic acid construct suitable for use in an expression vector for expressing botryococcene synthase or squalene synthase in *E. coli* may be designed. This may be based on a pET vector or similar, and the protein produced by the expression of the construct in *E. coli* may carry an N-terminal or C-terminal 6-histidine tag for purification of the protein. The expression construct may be synthesized.

A second nucleic acid construct suitable for use in an expression vector for expressing botryococcene synthase in *Chlamydomonas* may be designed. The expression construct may be synthesized. The *B. braunii* botryococcene synthase or squalene synthase construct or vector may include a selection marker comprising an antibiotic resistance gene, for example, the bleomycin resistance gene endowing resistance to phleomycin (ZeocinTM, Invitrogen). The construct or vector may also include a promoter suitable for expression in *Chlamydomonas*. One suitable *Chlamydomonas* promoter may be derived from the RuBisCO gene). The construct or vector may also include at least one enhancer sequence that contributes to strong RNA expression in *Chlamydomonas*. The enhancer sequence may be positioned in at least one intron, and the intron may be deliberately positioned close to the transcription start site of the botryococcene synthase or squalene synthase coding region. The enhancer element, or the intron comprising the enhancer element, may be substituted into a native intron present in the botryococcene synthase or squalene synthase gene. Because an intron is spliced out during RNA processing, the native intron may be first identified by comparing the genomic botryococcene synthase or squalene synthase sequence with its

corresponding RNA (cDNA) coding sequence. Furthermore, the recombinant intron may also be spliced out during RNA processing. The construct or vector may also include a transcription terminator (poly adenylation region) suitable for *Chlamydomonas*.

The expression construct may be ligated into a vector, for example a plasmid.

Nucleic acid sequences may be manipulated using computer software such as Sequencher (GeneCodes USA) or Bioedit.

Example 8 – Transformation

Electroporation and the glass beads method may be used for the transformation of algal species. The transformed cells may be incubated for two weeks in 24 hour light at 25°C. Transformed cells may be subjected to antibiotic selection (phleomycin (ZeocinTM) resistance). Colonies of resistant *Chlamydomonas* may become visible on plates from both kinds of transformation method after ten days. These *Chlamydomonas* transformants may require seven days further growth to produce enough material for reliable detection of the squalene synthase or botryococcene synthase gene.

Example 9 – Functional Tests of Botryococcene Synthase or Squalene Synthase In Vivo

A botryococcene synthase gene or squalene synthase gene may be functionally tested in *E. coli*, *S. cerevisiae*, *C. reinhardtii*, or other living organisms. A gene may be expressed as its native open reading frame, or as the codon-optimised version optimised according to the codon-usage of highly expressed genes of the host organism. A botryococcene synthase or squalene synthase gene may be expressed with N-terminal or C-terminal tags for purification (for example, hexa-histidine, or any other peptide-based purification tag). Botryococcene synthase or squalene synthase may be expressed fused to Green Fluorescent Protein (GFP) or other fluorescent polypeptide. Botryococcene synthase or squalene synthase may be expressed as a chimera of a *B. braunii* polypeptide and a host (e.g. *C. reinhardtii*) polypeptide.

Identification and quantification of botryococcene or squalene production in the transgenic organism may be conducted by lysing the transgenic host cells, then detecting botryococcene or squalene directly.

Botryococcene is not present at all in *E. coli*, *S. cerevisiae*, or *C. reinhardtii*, so any botryococcene detected must be due to the presence of the transgene.

Squalene is present in *E. coli*, *S. cerevisiae*, or *C. reinhardtii*, so any squalene detected must be above the background level of squalene in cells of the untransformed control strain in order to be concluded that its production was due to the presence of the transgene.

Alternatively, the activity of the transgenic botryococcene synthase or squalene synthase enzyme in the transgenic host may be assayed by lysing the transgenic host cells, then detecting the botryococcene or squalene using a biochemical assay for conversion of FPP to botryococcene or squalene, for example incorporation of tritium from tritium-labelled FPP into tritium-labelled botryococcene or tritium-labelled squalene.

The botryococcene or squalene can be detected by chromatography, for example gas chromatography (GC), liquid chromatography (LC), or thin layer chromatography (TLC). The TLC

procedure is useful for detecting tritium-labelled botryococcene or tritium-labelled squalene using photographic film or other detectors.

In representative experiments, a host organism such as *C. reinhardtii* or *E. coli* may be transformed with a nucleic acid molecule encoding botryococcene synthase or squalene synthase derived from *B. braunii* Race B strain Ayamé 1 (e.g. a nucleic acid molecule comprising any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment or variant thereof), and botryococcene or squalene production may be assayed.

Example 10 – Functional Tests of Botryococcene Synthase or Squalene Synthase In Vitro

Botryococcene synthase or squalene synthase enzymes expressed in *E. coli*, *S. cerevisiae*, *C. reinhardtii* or other living organisms, and with or without N-terminal or C-terminal tags for purification, with or without GFP or other fluorescent polypeptide fusion, and with or without fused polypeptide of a homologous or heterologous *C. reinhardtii* polypeptide, can be assayed for specific enzyme activity *in vitro*.

Botryococcene or squalene production may be achieved by purification of transgenic botryococcene synthase or squalene synthase enzyme from the *in vivo* host, until it exists in a high state of purity substantially free of other enzymes of that host. Identification and quantification of botryococcene or squalene may be conducted using a biochemical assay for conversion of FPP to botryococcene or squalene. The botryococcene or squalene thus produced can be detected by chromatography, e.g. GC, LC or TLC. The TLC procedure may rely on incorporation of tritium-labelled FPP into tritium-labelled botryococcene or tritium-labelled squalene, and may be detected using photographic film or other detectors to detect tritium labelled botryococcene or squalene.

Example 11 - Transformation of *C. reinhardtii* with the Squalene Synthase Gene from *B. braunii* Berkeley strain

The host strain of *C. reinhardtii* (CC400) possesses a cell wall weakness that allows DNA to be transformed into the cells by shaking the exogenous DNA and alga with glass beads to penetrate the weak cell wall. The glass beads technique is an established protocol. The cell wall deficient *Chlamydomonas* strain may also be transformed using electroporation.

Squalene Synthase Construct

To insert the *B. braunii* squalene synthase gene into *C. reinhardtii* CC400, a construct was made which has the capacity to express in *Chlamydomonas*. The artificial gene construct was based on pSP124S and was synthesised. The antibiotic resistance was used as a marker for transformation (bleomycin resistance gene, ZeocinTM antibiotic, a very active form of bleomycin antibiotic). To the *B. braunii* squalene synthase gene sequence was added a *Chlamydomonas* promoter from the *Chlamydomonas* RuBisCO gene), two enhancer/intron sequences that also promote strong RNA expression, and a *Chlamydomonas* transcription terminator.

Insertion into Chlamydomonas Genome

The squalene synthase construct was cut with the restriction enzyme *SacI* (Promega, WI, USA) to linearise it, and was then transformed into *Chlamydomonas* cells by a glass beads method and in parallel by electroporation.

Bioinformatic Methods

The bleomycin resistance plasmid was obtained from Dr Britta Forster (Australian National University) and Dr Saul Purton (University College London). The gene construct was designed by downloading from Genbank the genomic DNA sequence, and RNA sequence for *B. braunii* squalene synthase, checking its completeness and locating introns by means of the program Sequencher (GeneCodes USA), then adding *Chlamydomonas* elements cut and pasted from the bleomycin resistance gene, using the freeware program Bioedit. Introns were positioned close to the start of the squalene synthase gene, where the enhancer element in them has easiest access to assist in promotion of transcription. The open reading frame for squalene synthase was not modified.

Detection of Transgenic Chlamydomonas

Primer pair 1 (5'-CCTTGATGGGGTATTTGAGCAC-3', SEQ ID NO: 16 and 5'-GGAATTCGGTTAGGCGCTGAGTGTGGGTCTAGG-3', SEQ ID NO: 17) and primer pair 2 (5'-TGAGCACTTGCAACCCTTATCCG-3', SEQ ID NO: 18 and 5'-GGAATTCGGTTAGGCGCTGAGTGTGGGTCTAGG-3', SEQ ID NO: 19) were used to amplify the *B. braunii* squalene synthase gene from genomic DNA that had been extracted from individual *C. reinhardtii* colonies positive for ZeocinTM resistance.

Results

Transformants

After the electroporation and glass bead protocols had been carried out using the squalene synthase pSP124S construct, and the transformed cells incubated for two weeks at 25°C in 24 hour light on Tris Acetate Phosphate (TAP) agar with ZeocinTM, colonies of ZeocinTM resistant *Chlamydomonas* became visible on plates from both kinds of transformation method, electroporation and glass beads. Individual colonies of these *Chlamydomonas* transformants were picked and grown in liquid TAP media to produce enough material for reliable detection of the *B. braunii* squalene synthase gene. TAP medium for *C. reinhardtii* is described in Gorman and Levine, *Proc. Natl. Acad. Sci. USA*, 1965, 54, 1665-1669.

Detection of Recombinant Chlamydomonas

Each transformed cell grew to a colony on the plate. Each colony represents a transformed cell carrying the bleomycin resistance gene. Thus, the *B. braunii* squalene synthase was detected in transformed cells of *C. reinhardtii* CC400. Primer pair 1 and primer pair 2 amplified the artificial gene from some colonies of transformed CC400 colonies (Figure 33).

Other CC400 colonies that were bleomycin resistant but were not PCR-positive for the artificial gene carried only the antibiotic resistance parts of the squalene synthase pSP124S construct that was transformed into the cells, but did not also carry a full squalene synthase cassette. In such colonies the other parts of the DNA (i.e. some section of the squalene synthase gene, or its entirety) had been excluded by the organism during transformation and recombination.

Test of Function

Colonies that were both bleomycin resistant and positive for the squalene synthase pSP124S construct (i.e. using primer pairs 1 and 2) were assayed for squalene content. The transformed cells did not contain significantly more squalene than untransformed control cells.

EXAMPLE 12 - Cloning the *ATB1* Triterpenoid Hydrocarbon Synthetic Nucleic Acid Sequence

Gene *ATB1* was assembled from the *Botryococcus braunii* Ayamé genome and is 1212 nucleotides in length (Figure 30, SEQ ID NO: 30). To facilitate cloning the *ATB1* gene, as shown in Figure 32, the nucleotides CACCACATGCAT (Figure 43, SEQ ID NO: 33) comprising an *NsiI* restriction site (ATGCAT; Figure 44, SEQ ID NO: 34) were inserted immediately upstream of the native start codon of the nucleotide sequence provided in Figure 30 (SEQ ID NO: 30). By joining these 33 additional nucleotides to the start codon, an *NdeI* restriction site (CATATG; Figure 45, SEQ ID NO: 35) was also generated at the start codon. The addition of the 33 nucleotides resulted, in some constructs, in addition of one methionine and one histidine residue upstream of the native start codon of the polypeptide sequence provided in Figure 29 (SEQ ID NO: 29), but not in the construct pET11a/*ATB1* which utilises the *NdeI* site of the publicly available pET-11a vector to join to the *NdeI* site of the synthesised *ATB1* gene (SEQ ID NO: 32).

Similarly, the nucleotides ATGCAT (SEQ ID NO: 34) comprising an *NsiI* restriction site were inserted immediately upstream of the native stop codon of the nucleotide sequence of the *ATB1* gene provided in Figure 30 (SEQ ID NO: 30). In all constructs this resulted in addition to the polypeptide sequence provided in Figure 29 (SEQ ID NO: 29), of one methionine and one histidine residue, upstream of the native stop codon (giving ATB1-MH; Figure 46, SEQ ID NO: 36). Furthermore, the nucleotides TCACTCGAGCCCGGG (Figure 47, SEQ ID NO: 37) comprising *XhoI* and *SmaI* restriction sites were inserted immediately downstream of the native stop codon to facilitate cloning.

The resulting *ATB1* sequence with these 33 additional nucleotides was synthesised (by Genscript, NJ USA).

Initially, *ATB1* was cloned into standard vector pUC57 (construct 1) for verification by sequencing.

Subsequently, *ATB1* was subcloned into vector pET-11a for overexpression of ATB1 protein including one methionine and one histidine residue at the C-terminus of ATB1 (ATB1-MH; SEQ ID NO: 36) via the inducible T7-lac promoter in that vector.

The synthesised *ATB1* gene was cut with restriction enzymes *NdeI* and *BclI* and cloned into the *NdeI* and *BamHI* sites of pET-11a to produce construct pET11a/*ATB1* (construct 2).

ATB1 was also subcloned into vectors that supply an inducible T7-lac promoter and either a N-terminal or C-terminal poly-histidine tag that can be used for affinity purification of the ATB1 protein. The synthesised *ATB1* gene (SEQ ID NO: 32) was cut with restriction enzyme *XhoI* and cloned into the *PmlI* and *XhoI* sites of pET-302, to produce pET302/NT-His-ATB1 (construct 3) that encodes protein ATB1-MH augmented consecutively with one methionine, eight histidine, one methionine, and one histidine residues at its N-terminus (to produce MHHHHHHHMH-ATB1-MH; Figure 48, SEQ ID NO: 38). Alternatively, the synthesised *ATB1* gene was cut with restriction enzyme *NsiI* and cloned

directionally into the *Nsi*I site of pET-302 (such that the T7-lac promoter is upstream) to produce CT-His-ATB1/pET302 (construct 4) that encodes protein ATB1-MH augmented consecutively at its N-terminus with one histidine and one methionine residue and augmented consecutively at its N-terminus with five histidine residues and residues VNSLEIDDIRA (to produce MH-ATB1-MH-HHHHHVNSLEIDDIRA; Figure 49, SEQ ID NO: 39).

EXAMPLE 13 - Transformation of *E. coli* with the ATB1 Triterpenoid Hydrocarbon Synthetic Nucleic Acid Sequence

Transformation

The inducible promoter in *ATB1* constructs 2, 3, and 4 is the viral T7 promoter. Constructs 2, 3, and 4 were transformed into *E. coli* strain BL21 StarTM (DE3), which is itself a transgenic *E. coli* containing a viral (T7) RNA polymerase that allows transcription from the T7 promoter.

Three vials of *E. coli* BL21 StarTM (DE3) were thawed on ice, to be used for transformation of the three constructs. To an aliquot of these cells 100 ng of DNA was added, mixed by stirring with the pipette tip, and incubated on ice for 30 minutes. The cells were heat shocked at 42°C for 30 seconds, and immediately transferred back onto ice. To the transformed cells, 250 µl of SOC medium (20g Bacto tryptone, 5g Bacto yeast extract, 2 ml of 5 M NaCl, and 2.5 ml of 1 M KCl per litre of water, autoclaved before use, 20 ml of sterilized 1 M glucose added thereafter) was added and the tube containing the cells + SOC mix was shaken at 37°C for 1 hour to allow cells to recover from transformation, and to allow the cells to express the ampicillin resistance protein from the *b/a* gene locus that is present on all constructs. Cells were spread onto LB agar (10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl, 15 g bacteriological agar per litre of water, autoclaved before use).

Glycerol stocks of *E. coli* transformants were prepared (20% glycerol + 80% rich LB growth medium (10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl per litre of water, autoclaved before use)) by scraping transformants from LB agar containing ampicillin, and transferring several loopfuls of these cells into the glycerol + LB mix, resuspending the cells with a pipette tip, waiting an hour for cells to absorb glycerol, freezing in liquid nitrogen, and then placing and storing in a -80°C freezer.

Restriction Digests

Plasmid DNA from the transformants was verified by extraction and restriction digestion. To prepare for DNA extraction, a small aliquot of each transformant was grown overnight at 37°C in rich LB growth medium. DNA was extracted from transformed *E. coli* using a commercial miniprep kit (Wizard® Promega, WI, USA).

All constructs were cut with restriction enzyme *Nde*I, which cuts at the specific DNA sequence CATATG at the start of the synthesised *ATB1* gene. Each construct was also cut with a second restriction enzyme as follows: construct 2 – *Bst*XI (cuts at sequence CCANNNNNNNTGG; Figure 50, SEQ ID NO: 40); constructs 3 and 4 – *Xho*I (cuts at sequence CTCGAG; Figure 51, SEQ ID NO: 41). This was achieved by combining 0.5 µl of *Nde*I and 0.5 µl of the appropriate second restriction enzyme, together with 2 µl of 10x "Promega restriction buffer 4", 0.5 µl of bovine serum albumin (stock 10 mg/ml, final concentration 250 ng per µl), and 480 ng of DNA (about 6 to 12 µl), making a total reaction volume of 20 µl per reaction (by adding nuclease free water), and then

incubating at 37°C for 1 hour. Samples were electrophoresed on a 1% agarose gel in TAE buffer (40mM Tris acetate, 1 mM EDTA) at 120 volts for 30 minutes, and DNA bands were visualised with SYBR-Safe fluorescent dye, on a UV illuminator, and digitally captured to disc using an 8 second exposure, with camera aperture fully open.

The pattern observed when construct 2 was cut with *Nde*I and *Bst*XI included a band of 1070 base pairs (Figure 34A) as predicted, confirming the presence of *ATB1*. The pattern observed when constructs 3 and 4 were cut with *Nde*I and *Xho*I included a band of 1230 base pairs (Figure 34B) as predicted, confirming the presence of *ATB1*.

EXAMPLE 14 - Expression and Folding of ATB1 by Transformed E. coli

Untransformed *E. coli* and *E. coli* transformed with pET11a/ATB1 (construct 2) were each grown overnight at 37°C in 5 ml LB medium containing 1% glucose. Cultures of transformed *E. coli* contained 100 µg/ml (final concentration) of ampicillin, whereas cultures of untransformed *E. coli* contained no ampicillin. Next day, cultures were centrifuged and the pellets resuspended in 1 ml of LB medium, which was added to 49 ml of fresh LB medium. Again, cultures of transformed *E. coli* contained 100 µg/ml (final concentration) of ampicillin, whereas cultures of untransformed *E. coli* contained no ampicillin. The 50 ml cultures were incubated at 37°C with shaking and were monitored by spectrophotometer at wavelength 600 nm. Sterile technique was maintained throughout. The OD600 had reached 0.65 by 1.5 hours, and accordingly cells were ready to be induced.

The inducing agent isopropylthiogalactoside (IPTG) was added to each culture at a final concentration of 1 mM. At 0, 2 and 4 hours post IPTG induction, 1 ml of culture was removed and centrifuged. The resulting pellet was boiled for 10 minutes in 200 µl of protein assay buffer (50 mM Tris-HCl pH 6.8, 10 mM Dithiothreitol, 1% sodium dodecyl sulfate), and then protein concentration was measured on a Nanodrop spectrophotometer set at a wavelength of 280 nm. A sample containing 60 µg of protein was added to 3x strength Laemmli buffer (6% SDS, 30% glycerol, 15% β-mercaptoethanol, 0.006% bromophenol blue, 0.1875 M Tris-HCl) and boiled for 2 min.

Also at 4 hours post IPTG induction, 5 ml of culture of *E. coli* transformed with pET11a/ATB1 was removed and centrifuged. The resulting pellet was resuspended in 500 µl of lysis buffer B (50 mM Tris-HCl pH 7.5, 1 mM dithiothreitol, 2 mM PMSF, 0.5% Triton X-100, 100 µg DNase) then repeatedly freeze-thawed (10 times). The lysate was centrifuged for 15 min at 13,000 rpm in a microfuge at 4°C. The supernatant was removed, its protein content was measured with the Nanodrop spectrophotometer, and the supernatant sample was diluted using 3x Laemmli buffer. Lanes 8 and 9 of Figure 35 contain 15 µl and 20 µl, respectively, of this supernatant.

The samples were loaded on a SDS-PAGE minigel comprising a stacking gel (4% acrylamide) and a resolving gel (10% acrylamide). The minigel was run at 170 volts. All lanes of the minigel contained 1x Laemmli buffer final concentration. The proteins on the minigel were visualised using Coomassie dye (Figure 35).

ATB1 protein was not present in untransformed *E. coli* cells at any time (Figure 35, Lanes 1, 3, 5), nor was it present in *E. coli* cells transformed with pET11a/ATB1 (construct 2) and sampled at 0 h post IPTG-induction (Figure 35, Lane 2). However, ATB1 protein (running between 46 kDa and 30

kDa, i.e. between arrows compared to lane 7: molecular weight markers upper arrow 46 kDa, lower arrow 30 kDa) was expressed by transformed *E. coli* (transformed with construct 2) at 2 h and 4 h post IPTG induction (Figure 35, Lanes 4, 6, 8 and 9). ATB1 protein (arrowed in lane 9, Figure 35: soluble ATB1 protein) was present in the soluble supernatant of the cell lysate of *E. coli* transformed with construct 2 (Figure 35, Lanes 8 and 9), after lysing and centrifugation at 13000 rpm in a microfuge for 15 minutes. This indicates that ATB1 was correctly folded, since incorrectly folded ATB1 would be in the insoluble pellet.

EXAMPLE 15 - Generation of a Triterpenoid Hydrocarbon by Transformed *E. coli*

A few microlitres of each of untransformed *E. coli*, and *E. coli* transformed with either pET11a/ATB1 (construct 2) or pET302/NT-His-ATB1 (construct 3) were taken from glycerol stocks and added to 200 ml of LB medium. LB medium in which transformed *E. coli* were grown contained ampicillin (100 µg/ml), whereas LB broth in which untransformed *E. coli* were grown contained no ampicillin. Each culture also contained sterile glucose (final concentration 1% glucose w/v) to prevent premature expression of *ATB1*. The cultures were shaken overnight at 37°C to obtain a sufficient number of cells in stationary phase to be used as inoculates.

Each 200 ml culture was centrifuged and the pellet resuspended in 10 ml of LB medium, then added to 990 ml of fresh LB medium containing the inducer IPTG (final concentration 1 mM), and grown for 1 hour at 37°C to allow cells to recover from lag phase and begin to divide. The 1 litre cultures were then incubated overnight at 28°C to 33°C with shaking. The temperature 28°C to 33°C may increase the yield of correctly folded ATB1 protein relative to the yield at 37°C.

Following the overnight incubation, the cultures were centrifuged and the pellets resuspended in 10 ml of lysis buffer A (50 mM MOPS (4-Morpholinepropanesulfonic acid) buffer pH 7.3; 25 mM β-mercaptoethanol; 20 mM MgCl₂; 0.5 mM PMSF (Phenylmethylsulfonyl fluoride)). First the sample of untransformed *E. coli* placed into a cell crusher at 4°C, and then the sample of *E. coli* transformed with pET11a/ATB1 (construct 2) was placed into a cell crusher at 4°C. The cell crusher was extensively washed with 75% ethanol 25% water, and then fresh distilled water, between samples. After crushing, these two samples were extracted with 10 ml heptane at 60°C for two minutes with vigorous shaking. A third sample of *E. coli* transformed with pET302/NT-His-ATB1 (construct 3) was extracted directly with 10 ml heptane, by heating over a bunsen burner until boiling, without cell crushing.

Heptane was separated from aqueous material and cell debris by centrifugation in PTFE/Teflon® tubes. The heptane fraction was removed using a single-use heptane-rinsed glass pipette. Samples were evaporated in the order untransformed *E. coli*, *E. coli* transformed with pET11a/ATB1 then *E. coli* transformed with pET302/NT-His-ATB1 (construct 3). The rotary evaporator was rinsed twice, using 10 ml of clean heptane per rinse, between samples. Each sample was concentrated to 2 ml.

Heptane-soluble polar compounds, consisting of mostly fatty acids, their esters and lipids, were removed by: 1) Adding sufficient heptane to the sample to increase its volume to 10 ml; 2) Passing the entire 10 ml of heptane through a column(LCTech (Germany), Elufix Column, Part

number 9370) consisting of 2 g of anhydrous sodium sulfate and 2 g of synthetic magnesia-silica gel (Florisil®) and collecting the eluent; 3) Reducing the volume of heptane eluent collected to approximately 0.8 ml under a stream of nitrogen.

5 **EXAMPLE 16 - Generation of a Triterpenoid Hydrocarbon by Lysate of Transformed *E. coli***

A few microlitres of each of *E. coli* transformed with either pET11a/ATB1 (construct 2) or pET302/NT-His-ATB1 (construct 3) were taken from glycerol stocks and added to 20 ml of LB medium (also known as nutrient broth NB). LB medium in which transformed *E. coli* were grown contained ampicillin (100 µg/ml), whereas LB broth in which untransformed *E. coli* were grown
10 contained no ampicillin. Each culture also contained sterile glucose (1% final concentration). The cultures were incubated overnight at 37°C with shaking to obtain a sufficient number of cells in stationary phase to be used as inoculates. A glycerol stock of untransformed *E. coli* BL21 Star™ (DE3) was used directly as an inoculate. After overnight culture, cells were centrifuged to remove glucose, and the pellets each resuspended in LB medium without glucose, to allow transgene
15 expression.

Untransformed *E. coli* were incubated at 37°C only, whereas each of the two transformed *E. coli* strains was incubated both at 18°C and 37°C (i.e. making a total of 5 samples).

Each of the samples to be induced at 37°C was resuspended in 75 ml of LB (without glucose), grown for 2 hours at 37°C to approximately attain log phase, and then induced with 0.5 mM
20 IPTG. These were then further incubated for 2 hours at 37°C, to induce transgene expression.

Each of the samples to be induced at 18°C were passaged directly from the pellets obtained after removing glucose from overnight cultures. The pellets were resuspended in 4 ml of overnight culture, then transferred to 500 ml of LB containing 0.5 mM IPTG (final concentration) and incubated overnight at 18°C.

25 After induction, cultures were centrifuged. The pellet was resuspended in 5 ml of lysis buffer A. Lysis was achieved by sonication for 1 minute, on ice (ice-water slurry) in a glass beaker. Samples were clarified by aliquotting into microfuge tubes, and centrifuging at 13000 rpm, at 4°C for 15 minutes, then combining aliquots of identical samples to make 3.6 ml in total of each sample. To each clarified lysate, 35 µl of FPP stock (stock 10 mM in ethanol) and 15 mg of NADPH powder was
30 added and the mix was incubated with gentle shaking at 37°C for 30 minutes (Figure 36).

After incubation with FPP and NADPH, 10 ml of heptane was added to each lysate, and each was vigorously shaken by hand for about 20 seconds. Then the lysates were allowed to separate into hydrophobic and hydrophilic fractions overnight at room temperature with gentle shaking. Next day, the samples were centrifuged in PTFE/ Teflon® tubes then evaporated as
35 described in Example 14.

EXAMPLE 17 - Detection of a Triterpenoid Hydrocarbon by Gas Chromatography - Mass Spectrometry (GC-MS)

Samples derived from Examples 15 and 16 were analysed by GC-MS.

GC-MS was performed on an Agilent Technologies 5975C mass spectrometer equipped with an Agilent technologies 7890A gas chromatograph and a 7683B autosampler.

The capillary column was an Alltech EC-5 (15 m x 0.25 mm i.d.). Splitless injection was used with a split time of 60 seconds. Injection port temperature was 300°C. The carrier gas was helium in constant flow mode at 1.8 mL/min. Injection volume was 1 µL.

The GC temperature program was: initial temperature of 40°C for 4 min, ramp rate of 20°C/min, final temperature of 350°C for 4 min. The GC/MS interface temperature was set to 280°C. The FID temperature was 325°C.

For full scan mass spectra, the mass spectrometer was scanned from m/z 35 to m/z 550.

Electron ionisation (EI) was used with the electron energy set to 70 eV. The source temperature was 230°C. The MS quadrupole was set to 150°C.

Authentic squalene and authentic C₃₀ botryococcene were analysed by GC-MS (Figures 37 and 38, respectively).

Squalene (elution time 16.5 min, mass 410) and botryococcene (elution time 15.8 min, mass 410) were undetectable in untransformed *E. coli* when analysed by GC-MS (Figure 39). In contrast, Compound X (elution time 15.7 min, mass 408) and Compound Y (elution time 16.5 min) with a similar elution time to authentic squalene were detected in *E. coli* transformed with construct 2 and induced with IPTG (Figures 40 and 41).

When supplemented with FPP and NADPH, a cell lysate of *E. coli* transformed with construct 2 and induced with IPTG produced Compound Y (elution time 16.5 min) as determined by GC-MS, whereas Compound X (elution time 15.7 min) was undetectable (Figure 42).

The mass spectrum of Compound X after electron cracking (Figure 39) is similar to that of both squalene (Figure 37) and botryococcene (Figure 38), except that a majority of peaks are smaller by two mass units than those of squalene and botryococcene (Table 1). The absence of two hydrogens, demonstrated both by the overall molecular weight and the electron cracked mass spectrum, indicates that the triterpenoid Compound X is a “dehydro” triterpenoid. Indeed, the mass spectrum is consistent with a dehydrotriterpenoid, such as 12,13-dehydrosqualene, also known as 4,4'-diapophytoene.

On the basis of GC-MS, Compound X and Compound Y are identified as triterpenoid hydrocarbons. Specifically, Compound X is dehydrosqualene and Compound Y is squalene.

Table 1

Mass spectra of triterpenoids squalene, botryococcene and Compound X after electron cracking.

Major peaks (companion peaks in brackets) are reported.

Compound X	Squalene	C₃₀ Botryococcene
408	410	410
393	395	395
378 (377)		
365	367	367
356 (352)		354
339 (340)	341	341
327 (326)	325	325 (327, 328)
311		313
297	299	299 (297)
285 (282)	285 (287)	285
271	273	271 (273)
257 (255, 258)	257 (259)	257 (259)
243 (241)	245 (244)	245 (243)
229	231	231 (229)
215	217 (218)	217 (215)
203 (201)	203 203	
187 (189)	191 (189, 192)	191 (189)
173 (175)	175 (177)	175 (177)
159 (161)	161 (163)	161 (163)
147 (145)	149	149 (147)
133	137 (136)	135
119 (121)	121 (123)	121 (123)
109 (107, 105)	109 (107)	109 (107)
93 (95, 91)	95 (93)	95 (93)
81 (79)	81	81
69	69	69
55	55	55
41	41	41

EXAMPLE 18 – Algae Culture

Algae, for example *Chlamydomonas reinhardtii*, is grown in Aquasol[®] soluble fertiliser with trace elements at 150 mg/L. Aquasol[®] comprises nitrogen as monoammonium phosphate 1.8%, nitrogen as potassium nitrate 2.6%, nitrogen as urea 18.6% (total nitrogen 23%), phosphate as monoammonium phosphate 4% (total phosphate 4%), potassium as potassium nitrate 7.8%, potassium as potassium chloride 10.2% (total potassium 18%), zinc as sulphate 0.05%, copper as sulphate 0.06%, molybdenum as sodium molybdate 0.0013, manganese as sulphate 0.15%, iron as sodium ferric EDTA 0.06%, boron as sodium borate 0.011%.

Alternatively, algae is grown in TAP medium, which is described in Gorman and Levine, *Proc. Natl. Acad. Sci. USA*, 1965, 54, 1665-1669.

Algae are cultured in raceway ponds made of opaque plastic with dimensions 1.52 m wide, 2.44 m long and 0.61 m deep with a capacity of 2000 L. The raceway ponds are powered by a paddle wheel and operate at 20–30 cm depth. The paddle wheel rotation is adjusted to generate a flow rate of $21 \pm 3 \text{ cm s}^{-1}$. Algae are also cultured in vertical tank reactors (VTR) with dimensions 0.45 m diameter and 1.52 m height with 100 L working volume and made of transparent acrylic sheets. A roll of low density polyethylene (LDPE-Uline-6-Mil heavy duty polytubing with 50.8 cm circumference) material is used to fabricate hanging polybags with dimensions 95 cm deep, 15 cm diameter, 20 L working volume. All the reactors are attached with delivery tubings and air stones for bubbling 5–6% CO₂ and air mixture.

Supplemental CO₂ is derived from a liquid CO₂ storage tank blended with air using a Concoa BlendMaster Model 1000 mixer and passed through a Whatman HEPA-Vent filter at 5–6% CO₂ concentration in air. Cole Parmer rotameters are used to regulate air flow rates among the raceways, VTR and polybags. Raceways are stirred continuously with paddle wheels powered by Dayton Model 4Z129B motors (90 V, 1.27 Amps). For the VTR, mixing is accomplished by bubbling CO₂ and air mixture through rectangular air stones (15 × 4 × 4 cm), whereas for the polybags, the mixture is simply bubbled into a port disk (0.72 cm opening) placed at the bottom of the polybag. To keep the VTR and polybags stirred after terminating the supply of supplemental CO₂, each evening ambient air is pumped into these cultures at the same flow rate as that of the supplemental CO₂ gas mix during the day. Culture temperature and pH for the raceways, VTR and polybags are measured daily.

A MasterFlex Model 77250-62 peristaltic pump is used to fill polybags at a flow rate of 3–4 L min⁻¹. Inoculum is drained from the bottom of VTR through a gate valve fitted with a 1.3 cm internal diameter garden hose that is either directly sent into the raceways or into pre-autoclaved glass carboys/Erlenmeyer flasks for subsequent delivery to the intended culture system.

Algae are grown batch-wise in a green house. The raceways, VTR and polybags are arranged in a row parallel to the east–west direction.

Biomass is harvested with a Lavin process centrifuge (Model: 12-413 V) operated at 2250g, manually removed from the interior drum with a spoon and dried at 40°C in a hot air oven for 72 h. It is stored at 4–5°C.

Raceway ponds are filled with 800 L of growth medium. These are inoculated with 50 L each of exponentially growing cultures of algae. After inoculation, the raceways operate at 30 cm depth.

Alternatively, raceways with 496 L growth medium are inoculated with 18 L of algal culture and operate at 20 cm depth with a total volume of 550 L. VTR have 88 L growth medium and 4 L of algal culture is added as inoculum. VTR operate with 100 L working volume and the depth of water column is maintained at 61 cm.

Alternatively, algae are grown in raceways (working volume: 500 L; Depth: 18 cm), VTR (working volume: 100 L; diameter: 45 cm; depth: 61 cm) or polybags (working volume: 20 L; diameter: 16 cm; depth: 95 cm). To each of the raceways, VTR and polybags, 450, 90 and 18 L of growth medium is filled and 25, 5 and 1 L of the inoculum is added, respectively. Final volumes are, 500 L (18 cm deep) in raceways, 100 L (61 cm deep) in VTR and 20 L (95 cm deep) in polybags.

EXAMPLE 19 - Generating an ATB1 Construct Containing *Chlamydomonas* Control Elements

Disclosed is a method for generating an expression vector pCE2 (plasmid for *Chlamydomonas* Expression 2). Also disclosed is a method for putting the *ATB1* gene under the control of elements present in this vector pCE2, thus creating construct pCE3-ATB1 (Figure 54, SEQ ID NO: 43). Control elements used are a *C. reinhardtii* Rubisco small subunit (*RbcS2*) promoter and a *C. reinhardtii* *RbcS2* terminator.

An interim vector pCE1 is first constructed. The *C. reinhardtii* *RbcS2* terminator is amplified from pSP124S using PCR, and is ligated back into pSP124S at the polylinker to generate the interim vector pCE1. The *C. reinhardtii* *RbcS2* terminator is amplified from pSP124S using PCR, and is ligated into pCE1 at the polylinker upstream of the terminator to generate expression vector pCE2.

Stepwise, firstly the *RbcS2* terminator is amplified from pSP124S using forward primer *RbcS2-termforw* (5'-GAGTACGCGGCCGCCCCGCTCCGTGTAAAT-3'; Figure 55, SEQ ID NO: 44), which contains a *NotI* restriction site, and reverse primer *RbcS2-termrev* (5'-GCACTCGAATTCGCTTCAAATACGCCCAGC-3'; Figure 56, SEQ ID NO: 45), which contains an *EcoRI* restriction site. The polymerase used is Promega (Madison, Wisconsin, USA) GoTaq. All conditions used of GoTaq enzyme activity are within the manufacturer's instructions (dNTPs concentration 0.1 mM, MgCl₂ concentration 1 mM, primer concentration 0.4 pmol/μl, template concentration 1 pg/μl, 1x manufacturer's buffer) and the PCR cycle conditions used are 94°C for 3 min, then 40 cycles of: 94°C for 30 s, 55°C for 30 s, 72°C for 1 min, and a final extension of 72°C for 7 min. The amplicon generated is 260 bp long and is purified in a Promega Wizard PCR purification column, then cut with *NotI* and *EcoRI* for 1 hr at 37°C, and subsequently heat treated for 10 minutes at 65°C to denature enzymes, before purifying again through a Wizard PCR purification column.

The amplified and restricted terminator is then ligated into vector pSP124S that has been prepared as follows: the vector is cut with *NotI* and *EcoRI* by incubating with these enzymes for 1 hr at 37°C, and the vector is then purified through a Wizard PCR purification column; the cut and purified vector is treated with Promega Calf Intestinal Alkaline Phosphatase according to the manufacturer's instructions, in order to prevent self ligation; the phosphatase-treated vector is purified through a

Wizard PCR purification column. Ligation of the restriction-cut terminator amplicon and phosphatase treated pSP124S is accomplished with Promega T4 ligase according to the manufacturer's instructions. The ligation mix is transformed into DH10 α electrocompetent *E. coli* cells, using a Biorad (Hercules, California USA) Gene Pulser II electroporator set at 2.5 kV, together with a Molecular Bioproducts (San Diego, California, USA) electroporation cuvette with a 2 mm gap width.

Transformants are screened by plating the transformation mix onto Oxoid (Basingstoke, Hampshire, UK) nutrient agar (1 g/L Lab-Lemco powder, 2 g/L yeast extract, 5 g/L peptone, 5 g/L NaCl, 15 g/L agar) plates, containing 100 μ g/ml (final concentration in agar) of sterile-filtered ampicillin. Colonies are picked and grown overnight in LB medium (nutrient broth), and their plasmid DNA extracted the following day. These plasmids are screened by PCR using forward primer pSP124S-preSacl (5'-CACTAAAGGGAACAAAAGCTG-3'; Figure 57, SEQ ID NO: 46) and reverse primer *RbcS2-termrev* (5'-GCACTCGAATTCGCTTCAAATACGCCCAGC-3'; Figure 56, SEQ ID NO: 45), and those plasmids which yield a PCR amplicon of 291 bp are considered positive. Positive plasmids are sequenced by the Australian Genomic Research Facility (AGRF) using a T3 primer (5'-GCAATTAACCCTCACTAAAGGGA-3'; Figure 58, SEQ ID NO: 47) and dye-terminator chemistry to confirm the *RbcS2* terminator sequence. This interim vector that contains a *C. reinhardtii* *RbcS2* terminator in the polylinker is named pCE1.

Similarly, in making pCE2, the *RbcS2* promoter is first amplified from pSP124S using forward primer *RbcS2-promforw* (5'-GAGTGCAGCTCTAAATGCCAGAAGGAGCG-3'; Figure 59, SEQ ID NO: 48), which contains a *SacI* restriction site, and reverse primer *RbcS2-promrev* (5'-GTCAGTGCAGCGCTTTAAGATGTTGA-3'; Figure 60, SEQ ID NO: 49), which contains a *NotI* restriction site. The polymerase used is Promega GoTaq. All conditions used of GoTaq enzyme activity are within the manufacturer's instructions (dNTPs concentration 0.1 mM, MgCl₂ concentration 1 mM, primer concentration 0.4 μ M, template concentration 1 μ g/ μ l, 1x manufacturer's buffer) and the PCR cycle conditions used are 94°C for 3 min, then 40 cycles of: 94°C for 30 s, 52°C for 30 s, 72°C for 1 min, and a final extension of 72°C for 7 min. The amplicon generated is 239 bp long and is purified in a Wizard PCR purification column, then cut with *SacI* and *NotI* for 1 hr at 37°C, and subsequently heat treated for 10 minutes at 65°C to denature enzymes, before purifying again through a Wizard PCR purification column. The amplified and restricted promoter is then ligated into vector pCE1 that has been prepared as follows: the vector is cut with *SacI* and *NotI*, by incubating with these enzymes for 1 hr at 37°C, and the vector is then purified through a Wizard PCR purification column. The cut and purified vector is treated with Promega Calf Intestinal Alkaline Phosphatase according to the manufacturer's instructions in order to prevent self ligation. The phosphatase-treated vector is purified through a Wizard PCR purification column.

Ligation of the restriction-cut promoter amplicon and phosphatase treated pCE1 is accomplished with Promega T4 ligase according to the manufacturer's instructions. The ligation mix is transformed into DH10 α electrocompetent *E. coli* cells, using a Biorad Gene Pulser II electroporator, together with a Molecular Bioproducts electroporation cuvette with a 2 mm gap width. Transformants are screened by plating the transformation mix onto Oxoid nutrient agar plates, containing 100 μ g/ml (final concentration in agar) sterile-filtered ampicillin. Colonies are picked and grown overnight in LB

medium (nutrient broth), and their plasmid DNA extracted the following day. These plasmids are screened by PCR using forward primer pSP124S-preSacl (5'-CACTAAAGGGAACAAAAGCTG-3'; Figure 57, SEQ ID NO: 46) and reverse primer *RbcS2-termrev* (5'-GCACTCGAATTCGCTTCAAATACGCCAGC-3'; Figure 56, SEQ ID NO: 45), and those plasmids which yield a PCR amplicon of 494 bp are considered positive. Positive plasmids are sequenced by the AGRF using a T3 primer (5'-GCAATTAACCCTCACTAAAGGGA-3'; Figure 58, SEQ ID NO: 47) and dye-terminator chemistry, to confirm the *RbcS2* promoter sequence.

This novel expression vector that contains a *C. reinhardtii* *RbcS2* terminator and a *C. reinhardtii* *RbcS2* promoter, both present in the polylinker, is named pCE2. It contains a *NotI* site in between the *RbcS2* promoter and the *RbcS2* terminator. For greater expression, a hybrid *C. reinhardtii* Hsp70A/*RbcS2* promoter may be used instead of the *RbcS2* promoter (Schroda *et al.* *The Plant Journal*, 2000, 21, 121-131 and Schroda *et al.*, *The Plant Journal*, 2002, 31, 445-455).

As a further example, the *ATB1* gene is ligated into vector pCE2 at the *NotI* site between the *RbcS2* promoter and terminator, to create an expression construct pCE3-*ATB1*. Specifically, in generating the expression construct pCE3-*ATB1*, the *ATB1* gene is amplified from construct pUC57/*ATB1* that was manufactured at Genscript (Piscataway, New Jersey, USA). Construct pUC57/*ATB1* contains the inferred cDNA of the *Botryococcus braunii* Ayame *ATB1* gene, that was shown to function in *E.coli* in the form of subcloned version pET11/*ATB1* (Example 15). Amplification of the *ATB1* gene from construct pUC57/*ATB1* is accomplished using forward primer *ATB1forw* (5'-GTCACTGCGGCCGCGTCACTTTGTAGGGCT-3'; Figure 61, SEQ ID NO: 50) that contains a *NotI* restriction site, and reverse primer *ATB1rev* (5'-GAGTACGCGGCCGCAAAATGAGTATGCACC-3'; Figure 62, SEQ ID NO: 51) that also contains a *NotI* restriction site. The polymerase used is Phusion (Finnzymes, Espoo, Finland). All conditions used for Phusion enzyme activity are within the manufacturer's instructions (dNTPs concentration 0.7 mM, primer concentration 0.4 pmol/μl, template concentration 1 pg/μl, 1x manufacturer's buffer) and the PCR cycle conditions used are 94°C for 2 min, then 40 cycles of: 94°C for 10 s, 53°C for 30 s, 72°C for 1 min, and a final extension of 72°C for 7 min. The amplicon generated is 1244 bp long and is purified in a Wizard PCR purification column, then cut with *NotI* for 1 hr at 37°C, and subsequently heat treated for 10 minutes at 65°C to denature the *NotI* enzyme, before purifying again through a Wizard PCR purification column.

The amplified and restricted *ATB1* gene is then ligated into vector pCE2 that has been prepared as follows: the vector is cut with *NotI*, by incubating with these enzymes for 1 hr at 37°C, and the vector is then purified through a Wizard PCR purification column; the cut and purified vector is treated with Promega Calf Intestinal Alkaline Phosphatase according to the manufacturer's instructions to prevent self ligation; the phosphatase-treated vector is purified through a Wizard PCR purification column. Ligation of the restriction-cut *ATB1* gene and phosphatase treated pCE2 is accomplished with Promega T4 ligase according to the manufacturer's instructions. The ligation mix is transformed into DH10α electrocompetent *E. coli* cells, using a Biorad Gene Pulser II set at 2.5 kV, using a Molecular Bioproducts electroporation cuvette with a 2 mm gap width. Transformants are screened by plating the transformation mix onto Oxoid nutrient agar plates, containing 100 μg/ml (final concentration in agar) of sterile-filtered ampicillin. Colonies are picked and grown overnight in LB

medium (nutrient broth), and their plasmid DNA extracted the following day. These plasmids are screened by PCR using forward primer pSP124S-preSacl (5'-CACTAAAGGGAACAAAAGCTG-3'; Figure 57, SEQ ID NO: 46) and reverse primer *RbcS2-termrev* (5'-GCACTCGAATTCGCTTCAAATACGCCCAGC-3'; Figure 56, SEQ ID NO: 45), and those plasmids which yield a PCR amplicon of 1718 bp are considered positive. Positive plasmids are sequenced by the AGRF using a T3 primer (5'-GCAATTAACCCTCACTAAAGGGA-3'; Figure 58, SEQ ID NO: 47) and dye-terminator chemistry, to confirm the *ATB1* gene sequence, as well as the *RbcS2* promoter sequence and *RbcS2* terminator sequence. This novel construct that contains the *B. braunii* *ATB1* gene under the control of the *C. reinhardtii* *RbcS2* promoter and the *C. reinhardtii* *RbcS2* terminator, is named pCE3-ATB1.

EXAMPLE 20 - Transforming pCE3-ATB1 into *C. reinhardtii* CC503 to Generate Transgenic Organism

Construct pCE3-ATB1 is transformed into *C. reinhardtii* cells (cell-wall deficient strain CC503) by the method of vortexing the DNA and cells together in the presence of glass beads. To prepare the *C. reinhardtii* cells, CC503 culture is grown to log phase in Tris Acetate Phosphate (TAP) medium (Gorman and Levine 1965), then 50 ml are centrifuged and the pellet aseptically resuspended in 0.5 ml of TAP medium. To a sterile glass tube containing 300 mg of autoclaved Biospec (Bartlesville, Oklahoma, USA) glass beads (0.5 mm diameter), 500 µl of this thick cell suspension is added together with 2 µg of DNA that has been linearised by cutting with *SacI* site at 37°C for 1 hr. Prior to adding to the cells and beads, the DNA is heat treated at 65°C for 10 mins, cooled and purified with a Wizard PCR purification column, and resuspended in 45 µl of water. The cells, DNA, and beads mix is vortexed on a Ratek Instruments (Boronia, Victoria, Australia) vortexer at top speed for 30 seconds, thus transforming the *C. reinhardtii* cells with the pCE3-ATB1 construct. Following vortexing the transformed cells are spread on a TAP-agar plate that contains 3 µg per ml (final concentration in agar) of Invitrogen Zeocin (Carlsbad, California, USA). After two weeks growth at 25°C in a constant light, or a light-dark cycle of 16 hr light, 8 hours dark, the TAP agar plate containing transformed cells shows colonies of transformed *C. reinhardtii*, that have been positively selected on the basis of Zeocin resistance. Zeocin is a form of bleomycin. A bleomycin resistance gene is present in pSP124S and in pCE3-ATB1. The bleomycin gene also functions to confer resistance to zeocin.

Zeocin resistant colonies of *C. reinhardtii* are picked individually into individual 1 ml aliquots of fresh TAP liquid medium using a sterile wire loop to transfer them. The isolates are then each grown at 25°C in a constant light, or a light-dark cycle of 16 hr light, 8 hours dark, for up to 1 week, and then DNA is extracted from each by the procedure of Example 2. Isolates are screened by PCR to ascertain presence absence of the *ATB1* gene. Primers used for PCR screening to identify transgenic *C. reinhardtii* are the forward primer *RbcS2-promforw* (5'-GAGTGCGAGCTCTAAATGCCAGAAGGAGCG-3'; Figure 59, SEQ ID NO: 48) and reverse primer *RbcS2-termrev* (5'-GCACTCGAATTCGCTTCAAATACGCCCAGC-3'; Figure 60, SEQ ID NO: 56, SEQ ID NO: 45). Zeomycin resistant isolates from which an amplicon of 1703 bp is obtained are

considered positive, i.e. contain the ATB1 gene. The 1703 bp amplicon from five of these isolates is purified from an agarose electrophoresis gel, using a Wizard PCR and Gel purification kit, sequenced directly at AGRF, using either one of the PCR primers that were used to generate the amplicon. The sequence of the 1703 bp amplicon is compared to the ATB1 sequence as well as to its flanking *RbcS2* promoter and *RbcS2* terminator. Isolates which have identical sequence as the ATB1 gene source DNA and as the *RbcS2* control element source DNAs, are named as specific cases of transgenic organism CC503-ATB1, as follows: CC503-ATB1A, CC503-ATB1B, CC503-ATB1C, CC503-ATB1D, CC503-ATB1E.

EXAMPLE 21 - Assaying Triterpenoids in Transgenic Alga CC503

C. reinhardtii cells are assayed for levels of triterpenoids, and identities of triterpenoids, using GC-MS. Transgenic *C. reinhardtii* grown using standard media, temperatures, and light regimes, is centrifuged at 1,000 to 20,000 rpm to pellet the cells, the supernatant is decanted and discarded, and cells are resuspended in lysis buffer (50 mM 3-morpholinopropane-1-sulfonic acid pH7.3, 25 mM β -mercaptoethanol, 20 mM $MgCl_2$, 1 mM Phenylmethylsulfonylfluoride) then subjected to a French Press or similar, multiple times, to burst cells. The lysate is then extracted with heptane or hexane, for example as described in Example 15, and analysed by GC-MS as described in Example 17. Both the untransformed CC503, and the transformed CC503 are assayed for hydrocarbon content by identical methods, and the GC-MS data are compared to detect any hydrocarbons that are present in the transformed cells but not present in the untransformed cells. Any such novel hydrocarbon peaks that are indicated by mass spectrometry to be triterpenoids are concluded to have been derived from the action of the transgene. Five separate transformants are assayed for triterpenoid content, to allow for differences in the chromosomal location of integration events. Integration in euchromatin is expected to facilitate greater expression of the *ATB1* gene than integration in heterochromatin, however it is not necessary to know which type of chromatin the gene integrated in, rather it is simply necessary to screen multiple isolates. The isolate which produces the greatest level of triterpenoid (such as squalene, dehydrosqualene, botryococcene, dehydrobotryococcene, or other triterpenoid) compared with non-transgenic control is chosen over the other isolates that were assayed.

EXAMPLE 22 – Structure-Function Analysis

Functional units (motifs) in triterpenoid synthetic polypeptides have been identified. Prosite accession number PDOC00802 discloses two squalene and phytoene synthase signatures (PS01044 signature 1 and PS01045 signature 2). PS01044 signature 1 provides the consensus pattern:

Y-[CSAM]-x(2)-[VSG]-A-[GSA]-[LIVAT]-[IV]-G-x(2)-[LMSC]-x(2)-[LIV].

PS01045 signature 2 provides the consensus pattern:

[LIVM]-G-x(3)-Q-x(2,3)-[ND]-[IFL]-x-[RE]-D-[LIVMFY]-x(2)-[DE]-x(4,7)-R-x-[FY]-x-P.

These consensus signatures are used to find homologous signatures in the triterpenoid hydrocarbon synthetic polypeptide, thereby identifying one or more amino acids, and the

corresponding codon(s), to be mutated. Alternatively, the consensus sequence of Figure 53 is used to identify one or more amino acids, and the corresponding codon(s), to be mutated.

5 Oligonucleotide primers for site directed mutagenesis are designed to mutate the one or more codons selected for mutation in a plasmid encoding *ATB1*. The mutated triterpenoid hydrocarbon synthetic polypeptide is then expressed and its function assayed according to Examples 15 to 17 or Examples 20 and 21 for its triterpenoid hydrocarbon synthetic activity. The function of the mutated polypeptide is then correlated with the structure of the mutated polypeptide to ascertain amino acid residues that contribute to triterpenoid hydrocarbon synthetic activity.

CLAIMS

1. An isolated nucleic acid molecule encoding a polypeptide capable of producing a triterpenoid hydrocarbon, wherein the polypeptide comprises: at least 15 contiguous amino acid residues of any one of SEQ ID NOs: 1 to 5 or 20 to 29; at least 70% sequence identity to any one of SEQ ID NOs: 1 to 5 or 20 to 29; or a functional fragment of SEQ ID NO: 29.
2. The nucleic acid molecule of claim 1, comprising any one of SEQ ID NOs: 6 to 9 or 30, or a functional fragment thereof.
3. An isolated polypeptide capable of producing a triterpenoid hydrocarbon, wherein the polypeptide comprises: at least 15 contiguous amino acid residues of any one of SEQ ID NOs: 1 to 5 or 20 to 29; at least 70% sequence identity to any one of SEQ ID NOs: 1 to 5 or 20 to 29; or a functional fragment of SEQ ID NO: 29.
4. The polypeptide of claim 3, comprising any one of SEQ ID NOs: 1 to 5 or 20 to 29.
5. A vector, comprising the nucleic acid molecule of claim 1 or claim 2.
6. A recombinant non-human organism capable of producing a triterpenoid hydrocarbon, the organism comprising the nucleic acid molecule of claim 1 or claim 2, or the vector of claim 5.
7. The recombinant organism of claim 6, wherein the organism is a microorganism.
8. The recombinant organism of claim 7, wherein the microorganism is an alga or a bacterium.
9. The recombinant organism of claim 7, wherein the microorganism is selected from the group consisting of: *Escherichia coli*; *Chlamydomonas reinhardtii*; *Saccharomyces cerevisiae*; and *Pichia sp.*
10. Use of the nucleic acid molecule of claim 1 or claim 2, the polypeptide of claim 3 or claim 4, the vector of claim 5, or the recombinant organism of any one of claims 6 to 9, to produce a triterpenoid hydrocarbon.
11. A method for producing a triterpenoid hydrocarbon, comprising the step of growing the recombinant organism of any one of claims 6 to 9 under conditions sufficient for the organism to produce the triterpenoid hydrocarbon.
12. The method of claim 11, wherein the triterpenoid hydrocarbon is squalene or botryococcene and the conditions comprise supplementing the recombinant organism with NADPH or NADH.

13. The method of claim 11, wherein the triterpenoid hydrocarbon is dehydrosqualene or dehydrobotryococcene and the conditions comprise not supplementing the recombinant organism with, or minimising the concentration of, NADPH or NADH.
- 5 14. A triterpenoid hydrocarbon produced by the method of any one of claims 11 to 13.
15. Use of the nucleic acid molecule of claim 1 or claim 2, the polypeptide of claim 3 or claim 4, the vector of claim 5, or the recombinant organism of any one of claims 6 to 9, the triterpenoid hydrocarbon produced by the method of any one of claims 11 to 13, or the triterpenoid hydrocarbon of claim 14, in the production of an intermediate of a biofuel, a bioplastic, a pharmaceutical, a food additive, an industrial chemical or a specialty chemical production.
- 10
16. A method for producing an intermediate of a biofuel, a bioplastic, a pharmaceutical, a food additive, an industrial chemical or a specialty chemical production, comprising the steps of growing the recombinant organism of any one of claims 6 to 9 under conditions sufficient for the organism to produce a triterpenoid hydrocarbon, harvesting the triterpenoid hydrocarbon from the organism, and producing the intermediate from the triterpenoid hydrocarbon.
- 15
17. An intermediate produced by the method of claim 16.
- 20
18. The nucleic acid molecule of claim 1 or claim 2, the polypeptide of claim 3 or claim 4, the recombinant organism of any one of claims 6 to 9, the use of claim 10 or claim 15, the method of any one of claims 11 to 13 or 16, or the triterpenoid hydrocarbon of claim 14, wherein the triterpenoid hydrocarbon is produced from FPP and/or PSPP.
- 25
19. The nucleic acid molecule of claim 1 or claim 2, the polypeptide of claim 3 or claim 4, the recombinant organism of any one of claims 6 to 9, the use of claim 10 or claim 15, the method of any one of claims 11 to 13 or 16, or the triterpenoid hydrocarbon of claim 14, wherein the triterpenoid hydrocarbon is botryococcene or squalene.
- 30
20. A vector deposited as V10/022706 or V10/022707.