



## (51) International Patent Classification:

C12N 15/52 (2006.01) C12N 9/02 (2006.01)  
C12N 1/19 (2006.01) C07K 14/37 (2006.01)  
C12N 9/90 (2006.01) C12P 7/44 (2006.01)  
C12N 9/88 (2006.01)

## (21) International Application Number:

PCT/EP2015/055976

## (22) International Filing Date:

20 March 2015 (20.03.2015)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

92409 21 March 2014 (21.03.2014) LU

(71) Applicants: **PHILIPPS-UNIVERSITÄT MARBURG** [DE/DE]; Biegenstrasse 10, 35032 Marburg (DE).  
**RHEINISCH-WESTFÄLISCHE TECHNISCHE HOCHSCHULE (RWTH) AACHEN** [DE/DE]; Templergraben 55, 52062 Aachen (DE).

(72) Inventors: **BÖLKER, Michael**; Calvinstraße 8, 35037 Marburg (DE). **BUCKEL, Wolfgang**; Am Köpper 8, 35043 Marburg (DE). **PRZYBILLA, Sandra**; Am alten Feld 10, 51645 Gummersbach (DE). **WIERCKX, Nick**; Sileneweerd 4, NL-229TX Maastricht (NL). **BLANK, Lars M.**; Gibbenhey 10, 44227 Dortmund (DE).

(74) Agents: **WEINZIERL, Gerhard** et al.; SCHIWECK WEINZIERL KOCH, Landsbergerstraße 98, 80339 München (DE).

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

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

## Published:

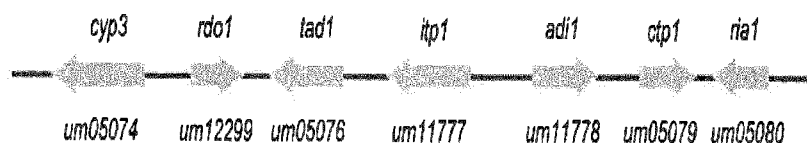
- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

## (54) Title: MEANS AND METHODS FOR ITACONIC ACID PRODUCTION

Figure 2

Itaconic acid production in deletion mutants and overexpression strains

| Gene    | Deletion | Overexpression |
|---------|----------|----------------|
| Um05074 | 0        | -              |
| Um12299 | 0        | 0              |
| Um05076 | --       | 0              |
| Um11777 | -        | 0              |
| Um11778 | --       | 0              |
| Um05079 | -        | ++             |
| Um05080 | --       | ++             |



(57) Abstract: The present invention relates to a method of producing itaconic acid. Further the present invention relates to nucleic acids encoding an aconitate-delta-isomerase (ADI) and trans- aconitate decarboxylase (TAD) and uses of such nucleic acids. Provided is additionally a recombinant host cell engineered to overexpress nucleic acids of the present invention. Furthermore an expression cassette and a vector are provided which include the respective nucleic acid.

---

**MEANS AND METHODS FOR ITACONIC ACID PRODUCTION**

---

**CROSS-REFERENCE TO RELATED APPLICATIONS**

5  
[001] The present application claims the benefit of and the priority to an application for  
“Means and Methods for Itaconic Acid Production” filed on 21 March 2014 with the Office  
de la propriété intellectuelle of the Ministère de l’Économie of Luxemburg, and there duly  
assigned serial number LU 92 409. The content of said applications filed on 21 March 2014  
10 is incorporated herein by reference for all purposes in their entirety including all tables,  
figures, and claims - as well as including an incorporation of any element or part of the  
description, claims or drawings not contained herein and referred to in Rule 20.5(a) of the  
PCT, pursuant to Rule 4.18 of the PCT.

**FIELD OF THE INVENTION**

15  
[002] The present invention relates to a method of producing itaconic acid. Further the  
present invention relates to nucleic acids encoding an aconitate-delta-isomerase (ADI) and a  
trans-aconitate decarboxylase (TAD) and uses of such nucleic acids. Provided is additionally  
a recombinant host cell engineered to overexpress nucleic acids of the present invention.  
20 Furthermore an expression cassette and a vector are provided which include the respective  
nucleic acid.

**BACKGROUND OF THE INVENTION**

25  
[003] Methylsuccinic acid (CAS number: 97-65-4), known as itaconic acid (ITA) and  
also called butanedioic acid, methylene- is a promising new platform compound for  
production of biofuels, chemical compounds, including e.g. detergents, and polymers,  
including plastics and artificial glass. ITA has been identified by the US Department of  
Energy as one of the top 12 bio-based chemical building blocks (Werpy, T, and Petersen, G,  
Top Value Added Chemicals from Biomass Volume I, Results of Screening for Potential  
Candidates from Sugars and Synthesis Gas. Technical report, U.S. Department of Energy,  
30 2004, available at <http://www.osti.gov/bridge/>). Over 80000 tons per year are produced,  
mostly for the production of polymers (Okabe et al., Appl Microbiol Biotechnol (2009) 84,  
597-606; Dwiarti, L, et al., Bioresour Technol. (2007) 98, 17, 3329-3337). ITA can for  
example be used in the manufacture of acrylic latexes, plasticizer, additives, adhesives,

emulsifying agents, paint chemicals, textile and paper.

[004] Currently, most ITA is produced by fermentation of *Aspergillus* strains (Okabe 2009, supra; Willke T, and Vorlop K.D., Appl Microbiol Biotechnol. (2001) 56, 3-4, 289-295; US2385283A). The parameters of ITA production using this fungus have been investigated (Willke, 2001, supra; Okabe 2009, supra), and the biochemical production route has been extensively studied (Bonnarme, P, et al., Journal of Bacteriology (1995) 177, 12, 3573-3578; Bentley, R, and Thiessen, C.P., J Biol Chem. (1957) 226, 2, 673-687; Bentley, R, and Thiessen, C.P., J Biol Chem. (1957) 226, 2, 689-701; Bentley, R, Thiessen, C.P., J Biol Chem. (1957) 226, 2, 703-720; Kanamasa, S, et al., Appl Microbiol Biotechnol. (2008) 80, 2, 223-229). In short, in *Aspergillus terreus*, the substrate (i.e. glucose) is converted to the primary metabolite cis-aconitate through glycolysis and the TCA cycle (tricarboxylic acid cycle). Cis-aconitate is subsequently decarboxylated by the enzyme cis-aconitate decarboxylase (CAD), to yield ITA and CO<sub>2</sub>. The identification of the CAD enzyme and especially its corresponding gene sequence has led to a wide range of metabolic engineering opportunities, which is rapidly expanding (e.g. Dwiarti, L, et al., J Biosci Bioeng. (2002) 94, 1, 29-33; Kanamasa et al., 2008, supra; WO 2009/014437 A1, US 2010/330632 A1, US 2010/285546 A1, US 2011/053232 A1). Alternative hosts for the production of ITA are available, such as strains of *Rhodotorula*, *Candida* and *Ustilago* (Willke, 2001, supra). Nevertheless, up to date virtually nothing is known about how these alternative hosts produce ITA.

[005] ITA production with *Aspergillus* strains has been well characterized (WO 2009/014437 A1, WO 2009/104958 A1, WO 2009/110796 A1). Further, *Ustilago* strains have been used for ITA production (JP55034017A, JP3035785A, Willke, 2001, supra).

[006] There are several disadvantages related to fermentation of filamentous fungi in general, and with *Aspergillus terreus* specifically (Willke, 2001, supra). Filamentous fungi are particularly problematic in that their morphology can be difficult to be controlled in fermentation systems.

[007] *Ustilago* strains that are presently applied for the production of itaconic acid are uncharacterized and thus are not amenable without undue efforts to modifications and improvements. Moreover, these strains produce itaconic acid under natural conditions, i.e. regulatory elements of genes involved in the biosynthesis of itaconic acid may be subject to feedback inhibition or end product inhibition, thereby limiting the spectrum of such strains for permanent itaconic acid production.

[008] It is therefore an object of the present invention to offer alternative means and

methods of obtaining ITA. It would be desirable if such alternative means and methods could provide a way of obtaining ITA in high yield.

[009] This object is solved by providing the claimed subject matter and the embodiments and aspects which follow.

## SUMMARY OF THE INVENTION

[010] The present invention is based on the elucidation of the biosynthesis pathway of itaconic acid production in *Ustilago maydis*. Though at first glance, the elucidation of said biosynthetic pathway may appear to be a straightforward matter, since *Ustilago maydis* is known to produce itaconic acid. However, while a knock-out of the key enzyme of itaconic acid, i.e., cis-aconitate decarboxylase (CAD) which converts cis-aconitate into itaconic acid, results in a complete abolishment of itaconic acid in presently applied host cells such as *Aspergillus* sp., a knock-out of the gene encoding a protein with sequence similarity to CAD in *Ustilago maydis* does not abolish itaconic acid production. Thus, it is apparent that *Ustilago maydis* must have a different pathway. This different pathway was elucidated by the present inventors. In particular, rather than using CAD as the key enzyme in itaconic acid production, *Ustilago maydis* uses an aconitate-delta-isomerase (ADI), which converts cis-aconitate into trans-aconitate, which is further converted into itaconic acid by the action of a trans-aconitate decarboxylase (TAD) (see Figure 14). Thus, in *Ustilago maydis*, two enzymes are required for itaconic acid production. These two enzymes have, to the best of the inventors' knowledge, no known homolog among the usual itaconic acid production hosts. This is an unusual and thus surprising finding and shows that inventive efforts were required to figure out that two enzymes are necessary, since neither, e.g., a standard complementation assay of mutants nor protein identification based on the available genome from *Ustilago maydis* would have led to success.

[011] Moreover, in accordance with data retrieved from knock-out and over-expression strains, the present inventors concluded that in the biosynthesis of itaconic acid, cis-aconitate, or a precursor thereof, is exported from the mitochondria into the cytosol by the mitochondrial transporter CTP1. In the cytosol, cis-aconitate is the substrate of ADI, which catalyzes isomerization to trans-aconitate. TAD uses trans-aconitate as a substrate for decarboxylation to itaconic acid. Finally, itaconic acid is exported by a transporter of the major facilitator superfamily (MFS). Hence, in accordance with this model the itaconic acid biosynthesis route in *Ustilago maydis* deviates fundamentally from that in, e.g., *Aspergillus*

*terreus* – a well-known and used working horse for the production of itaconic acid. In *Aspergillus terreus* cis-aconitate decarboxylase (CAD) is solely responsible for itaconic acid biosynthesis by catalyzing a decarboxylation of cis-aconitate, while *Ustilago maydis* has two key enzymes, i.e., ADI and TAD.

[012] The above being said, the present invention is broadly applicable to host cells and not limited to *Ustilago maydis*, since the present inventors showed that the expression of TAD and ADI in the baker's yeast *Saccharomyces cerevisiae* results in itaconic acid production, while *S. cerevisiae* is otherwise not capable of producing itaconic acid. This result demonstrates that ADI and TAD are functional in a heterologous host cell and are indeed the key players in the synthesis of itaconic acid. Moreover, given the fact that baker's yeast does not at all have homologs to TAD and ADI from *Ustilago maydis*, it is fully reasonable that TAD and ADI, when expressed in an microorganism otherwise incapable of producing itaconic acid, are functional and can thus convey to such an organism the capability to produce itaconic acid.

[013] Accordingly, in a first aspect the present invention relates to a method of producing itaconic acid, comprising

- (a) culturing a recombinant host cell which is engineered to overexpress
  - (i) a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity; and/or
  - (ii) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity,under conditions to allow said host cell to overexpress said polynucleotide sequence (i) and (ii), thereby said host cell is capable of converting cis-aconitate via trans-aconitate to itaconic acid; and
- (b) obtaining itaconic acid.

TAD and ADI activity can preferably be determined as described in the appended Examples.

[014] As mentioned before, the present invention is, inter alia, based on the identification of an aconitate-delta-isomerase (ADI) enzyme and a trans-aconitate decarboxylase (TAD) enzyme, which together compose the biochemical pathway for ITA production in *Ustilago maydis*. Although *U. maydis* possesses a gene encoding an enzyme with low similarity to the *A. terreus* CAD (UM06344 on the genome of *U. maydis* 521, 23% similarity to *A. terreus*

CAD), disruption of this gene in *U. maydis* does not affect ITA production. Instead, two genes encoded by the loci UM11778 and UM05076 on the genome of *U. maydis* 521 were found to be essential for ITA production. Disruption of either of these genes led to a drastic decrease of ITA production. Without being bound by any particular theory, UM11778 is  
5 believed to encode the enzyme aconitate-delta-isomerase (ADI). UM05076 is believed to encode the enzyme trans-aconitate decarboxylase (TAD). The amino acid sequence of neither of these enzymes has any apparent similarity to the CAD enzyme of *A. terreus*. An alternative pathway to produce itaconic acid was therefore discovered by the present inventors. In a first step, cis-aconitate will be converted into trans-aconitate by the ADI. The  
10 so-produced trans-aconitate is then converted into itaconate (itaconic acid) by the TAD. This principle can also be seen in the examples. Of course, this principle can also be expanded to other fungi and yeasts, as shown in the examples.

[015] Itaconic acid is a top-value biobased chemical building block used for the production of polymers, pharmaceuticals and fuels. Itaconic acid is currently produced via  
15 fermentation of the filamentous ascomycete *Aspergillus terreus*. Biosynthesis of itaconic acid in *A. terreus* occurs by decarboxylation of cis-aconitate, a common intermediate of the citric acid cycle. Itaconic acid production has been observed also in other fungal species but its biosynthetic routes have not been elucidated. The present inventors show that the basidiomycetous yeast *Ustilago maydis* uses an alternative biosynthesis pathway to produce  
20 itaconic acid. In this fungus, itaconic acid is generated by decarboxylation of trans-aconitate. This unusual substrate is generated in the cytoplasm by isomerization of cis-aconitate. All genes required for itaconic acid production and secretion are arranged in a single gene cluster in the *U. maydis* genome. The present inventors have characterized in vitro both trans-aconitate decarboxylase (TAD) and aconitate-delta-isomerase (ADI). They were also  
25 able to reconstitute itaconic acid production in the yeast *Saccharomyces cerevisiae* by expression of these two enzymes. This shows that at least TAD and ADI, when expressed in a heterologous host cell, are sufficient to produce itaconic acid. Hence, it is plausible and reasonable that a host cell equipped with genes encoding TAD and/or ADI is able to produce itaconic acid. The identification of an alternative route for itaconic acid production will help  
30 to improve both yield and application potential of this interesting biobased chemical building block, since the elucidation of the alternative route will allow a targeted manipulation of the enzymes that are correlated in this pathway. For example, any feedback inhibition or the like could be prevented and proteins could be massively overexpressed. In addition, the yeast-like growing fungus *U. maydis* might be a more suitable production strain for e.g. large-scale

submerged fermentation.

[016] In some embodiments, in the method of the present invention said host cell further overexpresses a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a mitochondrial citrate transporter, preferably of a mitochondrial cis-aconitate transporter.

[017] In further embodiments, in the method of the present invention said host cell further overexpresses a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a multidrug transporter of the major facilitator superfamily, preferably of a transporter which exports itaconic acid or itaconate, respectively, out of the host cell.

[018] In some embodiments, in the method of the present invention said host cell further overexpresses a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a transcription factor for at least one of polynucleotide sequences of any one of SEQ ID NOs: 1, 3, 4, and/or 6.

[019] In further embodiments, in the method of the present invention at least one of the polynucleotide sequences of any of the SEQ ID NO: 1, 3, 4, 6, and/or 8 is overexpressed in comparison to a host cell prior to engineering.

[020] In other embodiments, in the method of the present invention overexpression is achieved by using a recombinant promoter, which drives expression of said polynucleotide(s) in said host cell. The overexpression can be achieved by expressing 2, 3, 4 or more copies of said polynucleotide(s) in said host cell.

[021] In some embodiments, in the method of the present invention said polynucleotide is integrated in the genome of said host cell. The integration can be ectopically or in the natural locus. The overexpression can be achieved by using an enhancer to express the polynucleotide.

[022] In further embodiments, in the method of the present invention said polynucleotide, when overexpressed in said host cell results in production of itaconic acid of 2 g/L or more after 48 h culture in comparison to a host cell prior to engineering.

[023] In some embodiments, in the method of the present invention said host cell is

engineered to underexpress a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 9, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a cytochrome P450 monooxygenase. In some embodiments, the underexpression in said host cell results in production of itaconic acid of 2 g/L or more after 48 h culture in comparison to a host cell prior to engineering.

[024] In other embodiments, in the method of the present invention said polynucleotide sequence is a heterologous polynucleotide sequence.

[025] In some embodiments, in the method of the present invention said host cell is a cell of a fungus or a yeast, preferably said cell of a fungus is a cell from *Ustilago maydis*, *Aspergillus terreus* or *Saccharomyces cerevisiae*.

[026] In further embodiments, the method of the present invention further comprises providing the host cell with a suitable carbon source. Preferably, the carbon source comprises at least one of a monosaccharide, a polysaccharide, a lipid and a fatty acid. Preferably, the monosaccharide is one of mannose, glucose, arabinose and xylose. Preferably, the polysaccharide is one of starch, a mannan and cellulose. Preferably, the lipid is triacylglycerol.

[027] In some embodiments, the method of the present invention further comprises isolating itaconic acid from the medium in which the host cell is cultured.

[028] In a further aspect, the present invention relates to a recombinant host cell which is engineered to overexpress

- (i) a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity; and
- (ii) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity.

[029] In some embodiments, the recombinant host cell of the present invention comprises a host cell further overexpressing at least one polynucleotide sequence having at least 50% sequence identity with the SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a mitochondrial citrate transporter, preferably of a mitochondrial cis-aconitate transporter and/or a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said

polynucleotide sequence encoding a protein or fragment thereof having the activity of a multidrug transporter of the major facilitator superfamily, preferably of a transporter which exports itaconic acid out of the host cell, and/or a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a transcription factor for at least one of polynucleotide sequences of any one of SEQ ID NOs: 1, 3, 4, and/or 6.

[030] In some embodiments, the recombinant of the present invention comprises a host cell further underexpressing a polynucleotide sequences having at least 50% sequence identity with the sequence of SEQ ID NO: 9, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a cytochrome P450 monooxygenase.

[031] In another aspect the present invention relates to a nucleic acid molecule comprising a polynucleotide sequence selected from

- (a) a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity;
- (b) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a multidrug transporter of the major facilitator superfamily, preferably of a transporter which exports itaconic acid out of the host cell;
- (c) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a mitochondrial citrate transporter, preferably of a mitochondrial cis-aconitate transporter;
- (d) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity;
- (e) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a transcription factor for at least one of polynucleotide sequences of any one of SEQ ID NOs: 1, 3, 4, and/or 6.

[032] In some embodiments the nucleic acid of the present invention is a nucleic acid molecule, wherein

- (i) the polynucleotide sequence of (a) as described above is not the polynucleotide sequence having accession number UM11778 in MUMDB
- (ii) the polynucleotide sequence of (b) as described above is not the polynucleotide sequence having accession number UM11777 in MUMDB
- (iii) the polynucleotide sequence of (c) as described above is not the polynucleotide sequence having accession number UM05079 in MUMDB
- (iv) the polynucleotide sequence of (d) as described above is not the polynucleotide sequence having accession number UM05076 in MUMDB; and/or
- (v) the polynucleotide sequence of (e) as described above is not the polynucleotide sequence having accession number UM05080 in MUMDB.

[033] In a further aspect, the present invention relates to a polypeptide encoded by the nucleic acid of the present invention.

[034] In another aspect, the present invention relates to an expression cassette comprising one or more nucleic acids of the present invention.

[035] In yet another aspect the present invention relates to a vector comprising the nucleic acid sequence of the present invention or the expression cassette of the present invention.

[036] In a further aspect the present invention relates to a host cell overexpressing at least one of the nucleic acid molecules of the present invention, such as one, two, three, four or five nucleic acid molecules of the present invention.

[037] In another aspect the present invention relates to a host cell underexpressing a polynucleotide having at least 50% sequence identity with the sequence of SEQ ID NO: 9, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a cytochrome P450 monooxygenase.

[038] In another aspect the present invention relates to a host cell overexpressing at least one of nucleic acid molecules of the present invention and underexpressing a polynucleotide having at least 50% sequence identity with the sequence of SEQ ID NO: 9, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a cytochrome P450 monooxygenase.

[039] In another aspect the present invention relates to a method of producing itaconic acid, comprising

- (a) culturing a recombinant host cell of the present invention, and
- (b) obtaining itaconic acid

[040] In yet another aspect the present invention relates to a use of a nucleic acid of the present invention, an expression cassette of the present invention, a vector of the present invention, or a host cell of the present invention for the production of itaconic acid in a host cell.

[041] In another aspect the present invention relates to use of a nucleic acid of the present invention, an expression cassette of the present invention, or a vector of the present invention, for the production of a host cell, which is capable of producing itaconic acid.

[042] In a further aspect the present invention relates to a method for producing a host cell, which is capable of producing itaconic acid, comprising genetically engineering of a host cell to overexpress a nucleic acid of the present invention or transforming said host cell with an expression cassette of the present invention, or with a vector of the present invention.

[043] In some embodiments in the use of the present invention or in the method for producing a host cell of the present invention the host cell is a cell of a fungus or a yeast.

## BRIEF DESCRIPTION OF THE DRAWINGS

[044] **Figure 1** is a graphical outline of the itaconic acid cluster in *U. maydis* including a comparison to the *Pseudozyma hubeiensis* cluster. Sequence similarities at the protein level of the *Pseudozyma hubeiensis* genes to the corresponding *U. maydis* genes were determined using ClustalW and are indicated in brackets.

[045] **Figure 2** lists the genes of the itaconic acid cluster in *U. maydis* and the effects of their deletion/overexpression. -- strongly decreased production; - decreased production; 0 no effect; + increased production; ++ strongly increased production.

[046] **Figure 3** illustrates relative itaconic acid production by *Ustilago maydis* MB215 strains carrying either mutations or overexpression constructs of itaconic acid biosynthesis genes. Itaconic acid production of wild type cells was set to 100 %. (A) MB215 carrying deletions for indicated genes. (B) MB215 carrying overexpression constructs for indicated genes.

[047] **Figure 4** depicts nucleotide and derived amino acid sequences as disclosed in the instant document. Intron sequences are indicated by lowercase letters. **Fig 4A:** SEQ ID NO:

1, nucleic acid sequence of UM11778(MB215), **Fig 4B**: SEQ ID NO: 2, amino acid sequence of UM11778(MB215), **Fig 4C**: SEQ ID NO: 3, nucleic acid sequence of UM11777, **Fig 4D**: SEQ ID NO: 4, nucleic acid sequence of UM05079, **Fig 4E**: SEQ ID NO: 5, nucleic acid sequence of UM12299, **Fig 4F**: SEQ ID NO: 6, nucleic acid sequence of UM05076, **Fig 4G**: SEQ ID NO: 7, nucleic acid sequence of UM11778, **Fig 4H**: SEQ ID NO: 8, nucleic acid sequence of UM05080, **Fig 4I**: SEQ ID NO: 9, nucleic acid sequence of UM05074, **Fig 4J**: SEQ ID NO: 10, amino acid sequence of UM11778, **Fig 4K**: SEQ ID NO: 11, amino acid sequence of UM05074, **Fig 4L**: SEQ ID NO: 12, amino acid sequence of UM12299 **Fig 4M**: SEQ ID NO: 13, amino acid sequence of UM05076, **Fig 4N**: SEQ ID NO: 14, amino acid sequence of UM11777, **Fig 4O**: SEQ ID NO: 15, amino acid sequence of UM05079, **Fig 4P**: SEQ ID NO: 16, amino acid sequence of UM05080, **Figure 4Q**: SEQ ID NO: 17, nucleic acid sequence of UM02807, **Figure 4R**: SEQ ID NO: 18, polypeptide sequence of UM02807, **Figure 4S**: SEQ ID NO: 19, nucleic acid sequence of UM02806, **Figure 4T**: SEQ ID NO: 20, polypeptide sequence of UM02806, **Figure 4U**: SEQ ID NO: 21, nucleic acid sequence UM02808, **Figure 4V**: SEQ ID NO: 22, Polypeptide sequence UM02808.

**[048] Figure 5** Generation of UM05076 deletion mutant (A) Plasmid map of pRS426-um05076-hyg. (B) Outline of the region around the UM05076 gene in the *U. maydis* genome and the deletion construct integrated into the *U. maydis* genome by homologous recombination.

**[049] Figure 6** Generation of UM11778 deletion mutant (A) Plasmid map of pRS426-um11778-hyg. (B) Outline of the region around the UM11778 gene in the *U. maydis* genome and the deletion construct integrated into the *U. maydis* genome by homologous recombination.

**[050] Figure 7** shows the production kinetics of ITA from *U. maydis* MB215 wildtype and deletion mutants  $\Delta$ UM11778 and  $\Delta$ UM05076 in 50 ml MTM containing 50 g l<sup>-1</sup> glucose, 0.8 g l<sup>-1</sup> NH<sub>4</sub>Cl and 19.5 g L<sup>-1</sup> 2-(N-morpholino)ethanesulfonic acid (MES) (n = 2; the values are mean of two analytical determinations)

**[051] Figure 8** Plasmid map of pOTEF-UM05076 overexpression plasmid used in the Examples. Indicated are the Potef promoter, the cbx resistance gene and the open reading frame of UM05076.

**[052] Figure 9** Plasmid map of pOTEF-UM11778 overexpression plasmid used in the Examples. Indicated are the Potef promoter, the cbx resistance gene and the open reading frame of UM11778.

**[053] Figure 10** *In vivo* reconstitution of itaconic acid production in *S. cerevisiae*.

*S. cerevisiae* ESM356 (■), *S. cerevisiae* ESM356 + pGREG503-um05076 (▲), *S. cerevisiae* ESM356 + pGREG506-UM11778 (▼) and *S. cerevisiae* ESM356 + pGREG503-UM05076 + pGREG506-UM11778 (◆). Error bars indicate deviation from the mean (n=2). Coexpression of *um05076* and *um11778* resulted in itaconic acid formation (◆).

**[054] Figure 11** Plasmid map of pGREG503-UM05076 overexpression plasmid used in the Examples. Indicated are the GAL1 promoter, the HIS3 marker and the open reading frame of UM05076.

**[055] Figure 12** Plasmid map of pGREG506-UM11778 overexpression plasmid used in the Examples. Indicated are the GAL1 promoter, the URA3 marker and the open reading frame of UM11778.

**[056] Figure 13** Characterization of enzyme activities. (A) Aconitate-delta-isomerase (ADI) encoded by UM11778 catalyzes isomerization of cis-aconitate into trans-aconitate in both directions. (B) Trans-aconitate decarboxylase (TAD) encoded by UM05076 catalyzes decarboxylation of trans-aconitate to itaconic acid.

**[057] Figure 14** Incubation of cis-aconitate with aconitate-delta-isomerase (ADI) and trans-aconitate decarboxylase (TAD) leads to formation of itaconic acid.

**[058] Figure 15:** Model for intracellular localization of itaconic acid biosynthesis in *U. maydis*.

**[059] Figure 16:** The *um02807* cluster and its influence on the itaconate biosynthesis of *U. maydis* MB215. Figure 16A: *U. maydis* genes present in the cluster of *um02807* encode a putative CTP1-mitochondrial citrate transporter (*ctp1*), a putative aconitate- $\Delta$ -isomerase protein (*PrpF*), a putative transcriptional regulator (*regulator*), and genes present in the itaconate cluster encode a putative Cytochrome P<sub>450</sub> monooxygenase (*cyp3*), a trans-aconitate decarboxylase (*tad1*), a Major Facilitator Superfamily transporter (*itp1*), an aconitate- $\Delta$ -isomerase (*adi1*), a putative mitochondrial tricarboxylate transporter (*mtt1*), and a putative transcriptional regulator (*ria1*). Figure 16B: Relative itaconate concentration of different deletion, overexpression, and complementary mutants of *U. maydis* compared to the wildtype after 96 h in the screening medium specified in the Examples section is shown. Values are the arithmetic mean of two biological determinations. Error bars indicate deviation from the mean.

**[060] Figure 17:** Pairwise sequence alignment between the polynucleotide

sequences of UM11778 (SEQ ID NO: 1) and UM02807 (SEQ ID NO: 17).

### DETAILED DESCRIPTION OF THE INVENTION

[061] As indicated above, the present inventors have identified two genes, UM05076 and UM11778 encoding the key enzymes for itaconic acid via cis- and trans-aconitate. The gene UM05076 encodes a trans-aconitate decarboxylase (TAD) and UM11778 encodes an aconitate-delta-isomerase (ADI). However, even though both genes are not existent in *A. terreus* or any other itaconic acid producing host known thus far they are found in the genome sequence of *P. hubeiensis*, which is a close relative to *U. maydis* (Fig. 1). In *A. terreus* other genes that are implicated in ITA production surround the *cadA* gene. ATEG\_09970 encodes a putative mitochondrial tricarboxylic acid transporter (WO 2009/104958 A1) and ATEG\_09972 encodes an ITA transporting major facilitator superfamily (MFS) transporter (WO 2009/ 110796 A1). The similarity in the organization of genes surrounding the genes encoding ADI and TAD in *U. maydis* and the *cadA* gene in *A. terreus* makes it likely that the surrounding *U. maydis* genes are also involved in ITA production. However, there is only little similarity between the amino acid sequences encoded by the genes in *A. terreus* and *U. maydis*. In fact, both knock-out and overexpression experiments confirmed that further genes, which surround the genes encoding ADI and TAD, are involved in the biosynthesis pathway for the production of itaconic acid. In particular a transporter of the major facilitator superfamily (MFS) with the accession number UM11777 in MUMDB is involved in the biosynthesis pathway for the production of itaconic acid. In fact, knock-out of UM11777 led to a decrease in ITA production.

[062] Also the mitochondrial transporter CTP1 (CTP1) with the accession number UM05079 in MUMDB is particularly involved in the biosynthesis pathway for the production of itaconic acid. In fact, overexpression led to an increase in ITA production and knock-out led to a decrease in ITA production.

[063] In addition, the functional transcription factor (transcription factor) with the accession number UM05080 in MUMDB is particularly involved in the biosynthesis pathway for the production of itaconic acid. In fact, overexpression led to an increase in ITA production and knock-out led to a decrease in ITA production.

[064] Further, the cytochrome P450 monooxygenase (P450 monooxygenase) with the accession number UM05074 in MUMDB has an influence on the biosynthesis pathway for the production of itaconic acid. In fact, its overexpression led to a decrease in the production

of ITA.

[065] Based on the identification of the *U. maydis* enzymes and their corresponding genetic sequences new opportunities are provided for metabolic engineering of ITA production hosts with improved production characteristics. The identification of an ADI and a TAD enzyme from the phylum of Basidiomycota allows selection of the optimal enzyme in a heterologous host. Furthermore, in particular *U. maydis* has several characteristics that make it superior to *A. terreus* in an industrial setting, such as a yeast-like growth pattern and low sensitivity to medium impurities. The identification of the genes involved in ITA production in *U. maydis* will aid in optimizing this strain but also other strains of species different from *U. maydis* for biotechnological ITA production.

### Definitions

[066] Unless otherwise stated, the following terms used in this document, including the description and claims, have the definitions given below.

[067] The word “about” as used herein refers to a value being within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, “about” can mean within 1 or more than 1 standard deviation, per the practice in the art. The term “about” is also used to indicate that the amount or value in question may be the value designated or some other value that is approximately the same. The phrase is intended to convey that similar values promote equivalent results or effects according to the invention. In this context “about” may refer to a range above and/or below of up to 10%. The word “about” refers in some embodiments to a range above and below a certain value that is up to 5%, such as up to up to 2%, up to 1%, or up to 0.5 % above or below that value. In one embodiment “about” refers to a range up to 0.1 % above and below a given value.

[068] The use of the terms “5’ “ and “3’ “ is a convention used to describe features of a nucleotide sequence related to either the position of genetic elements and/or the direction of events (5’ to 3’), such as e.g. transcription by RNA polymerase or translation by the ribosome which proceeds in 5' to 3' direction. Synonyms are upstream (5') and downstream (3'). Conventionally, nucleotide sequences, gene maps, vector cards and RNA sequences are drawn with 5' to 3' from left to right or the 5' to 3' direction is indicated with arrows, wherein

the arrowhead points in the 3' direction. Accordingly, 5' (upstream) indicates genetic elements positioned towards the left hand side, and 3' (downstream) indicates genetic elements positioned towards the right hand side, when following this convention.

[069] The term “antibody” generally refers to an immunoglobulin, a fragment thereof or a proteinaceous binding molecule with immunoglobulin-like functions. Examples of (recombinant) immunoglobulin fragments are Fab fragments, Fv fragments, single-chain Fv fragments (scFv), diabodies, triabodies (Iliades, P., et al., *FEBS Lett* (1997) 409, 437-441), decabodies (Stone, E., et al., *Journal of Immunological Methods* (2007) 318, 88-94) and other domain antibodies (Holt, L.J., et al., *Trends Biotechnol.* (2003), 21, 11, 484-490). An example of a proteinaceous binding molecule with immunoglobulin-like functions is a mutein based on a polypeptide of the lipocalin family (WO 2003/029462; WO 2005/019254; WO 2005/019255; WO 2005/019256; Beste et al., *Proc. Natl. Acad. Sci. USA* (1999) 96, 1898-1903). Lipocalins, such as the bilin binding protein, the human neutrophil gelatinase-associated lipocalin, human Apolipoprotein D, human tear lipocalin, or glycodefin, possess natural ligand-binding sites that can be modified so that they bind to selected small protein regions known as haptens. Other non-limiting examples of further proteinaceous binding molecules so-called glubodies (see WO 96/23879), proteins based on the ankyrin scaffold (Mosavi, L.K., et al., *Protein Science* (2004) 13, 6, 1435-1448) or the crystalline scaffold (WO 2001/04144), the proteins described by Skerra (*J. Mol. Recognit.* (2000) 13, 167-187), AdNectins, tetranectins, avimers and peptoids. Avimers contain so called A-domains that occur as strings of multiple domains in several cell surface receptors (Silverman, J, et al., *Nature Biotechnology* (2005) 23, 1556-1561). Adnectins, derived from a domain of human fibronectin, contain three loops that can be engineered for immunoglobulin-like binding to targets (Gill, D.S. & Damle, N.K., *Current Opinion in Biotechnology* (2006) 17, 653-658). Tetranectins, derived from the respective human homotrimeric protein, likewise contain loop regions in a C-type lectin domain that can be engineered for desired binding (ibid.). Peptoids, which can act as protein ligands, are oligo(N-alkyl) glycines that differ from peptides in that the side chain is connected to the amide nitrogen rather than the  $\alpha$  carbon atom. Peptoids are typically resistant to proteases and other modifying enzymes and can have a much higher cell permeability than peptides (see e.g. Kwon, Y.-U., and Kodadek, T., *J. Am. Chem. Soc.* (2007) 129, 1508-1509). Where desired, a modifying agent may be used that further increases the affinity of the respective moiety for any or a certain form, class etc. of target matter.

[070] The polynucleotide may be integrated in its natural locus. “Natural locus” means the location on a specific chromosome, where the polynucleotide is located. However,

in another embodiment, the polynucleotide is present in the genome of the host cell not at their natural locus, but integrated ectopically. The term "ectopic integration" means the insertion of a nucleic acid into the genome of a microorganism at a site other than its usual chromosomal locus, i.e., predetermined or random integration.

5        [071] By the use of the term "enriched" in reference to a polypeptide, a nucleic acid or a cell is meant that the specific amino acid/nucleotide sequence or cell, including cell population, constitutes a significantly higher fraction (2 - 5 fold) of the total amino acid sequences or nucleic acid sequence present in the sample of interest than in the natural source from which the sample was obtained. The polypeptide, a nucleic acid or a cell may  
10 also constitute a significantly higher fraction than in a normal or diseased organism or than in normal or diseased cells or in the cells from which the sequence was taken. This could be caused by preferential reduction in the amount of other amino acid/nucleotide sequences or cells present, or by a preferential increase in the amount of the specific amino acid/nucleotide sequence or cell of interest, or by a combination of the two. However, it should  
15 be noted that enriched does not imply that there are no other amino acid sequences, nucleotide sequences or cells present. The term merely defines that the relative amount of the sequence of interest has been significantly increased. The term significant here is used to indicate that the level of increase is useful to the person achieving such an increase, and generally means an increase relative to other amino acid or nucleic acid sequences of about  
20 at least 2-fold, for example at least about 5- to 10-fold or even more. The term is meant to cover only those situations in which man has intervened to increase the proportion of the desired amino acid sequence, nucleotide sequence or cell.

25        [072] The term "essentially consists of" is understood to allow the presence of additional components in a sample or a composition that do not affect the properties of the sample or a composition. As an illustrative example, a pharmaceutical composition may include excipients if it essentially consists of an active ingredient.

30        [073] The terms "expressing" and "expression" in reference to a nucleic acid as described herein are intended to be understood in the ordinary meaning as used in the art. A nucleic acid is expressed by a cell via transcription of a nucleic acid into mRNA, followed by translation into a polypeptide, which is folded and possibly further processed.

      [074] With regard to the respective biological process itself, the terms "expression", "gene expression" or "expressing" refer to the entirety of regulatory pathways converting the information encoded in the nucleic acid sequence of a gene first into messenger RNA (mRNA) and then to a protein. Accordingly, the expression of a gene includes its

transcription into a primary hnRNA, the processing of this hnRNA into a mature RNA and the translation of the mRNA sequence into the corresponding amino acid sequence of the protein. In this context, it is also noted that the term "gene product" refers not only to a protein, including e.g. a final protein (including a splice variant thereof) encoded by that gene and a respective precursor protein where applicable, but also to the respective mRNA, which may be regarded as the "first gene product" during the course of gene expression.

[075] By the term "expression cassette", also referred to as an expression system, is meant a nucleic acid molecule capable of directing expression of a particular nucleotide sequence in an appropriate host cell. An expression cassette includes a promoter operatively linked to the nucleotide sequence of interest, which is operatively linked to one or more termination signals. It may also include sequences required for proper translation of the nucleotide sequence. The coding region can encode a polypeptide of interest and can also encode a functional RNA of interest, including but not limited to, antisense RNA or a non-translated RNA, in the sense or antisense direction. The expression cassette comprising the nucleotide sequence of interest can be chimeric, meaning that at least one of its components is heterologous with respect to at least one of its other components. The expression cassette can also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression. In some embodiments, however, the expression cassette is heterologous with respect to the host; i.e., the particular nucleic acid sequence of the expression cassette does not occur naturally in the host cell and was introduced into the host cell or an ancestor of the host cell by a transformation event. The expression of the nucleotide sequence in the expression cassette can be under the control of a constitutive promoter or of an inducible promoter that initiates transcription only when the host cell is exposed to some particular external stimulus. In the case of a multicellular organism such as a plant or an animal, the promoter can also be specific to a particular tissue, organ, or stage of development.

[076] By "fragment" in reference to a polypeptide as described herein is meant any amino acid sequence present in a polypeptide as described herein, as long as it is shorter than the full length sequence and as long as it is capable of performing the function of a protein involved in the biosynthesis of itaconic acid as described herein. Preferred fragments have at least 20, 40, 60, 80, 100, 120, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600 amino acids. Such preferred fragments have the function of a protein involved in the biosynthesis of itaconic acid as described herein.

[077] By "gene" is meant a unit of inheritance that occupies a specific locus on a

chromosome and that is a segment of nucleic acid associated with a biological function. A gene encompasses transcriptional and/or translational regulatory sequences as well as a coding region. Besides a coding sequence a gene may include a promoter region, a cis-regulatory sequence, a non-expressed DNA segment that is a specific recognition sequence for regulatory proteins, a non-expressed DNA segment that contributes to gene expression, a DNA segment designed to have desired parameters, or combinations thereof. A gene can be obtained by a variety of methods, including cloning from a biological sample, synthesis based on known or predicted sequence information, and recombinant derivation of an existing sequence.

[078] A protein has “identity”, or is “identical” to a second protein if the nucleic acid sequence that encodes the protein has a similar or identical sequence to the nucleic acid sequence that encodes the second protein. Also, a protein has identity to a second protein if the two proteins have “similar” amino acid sequences. Thus, the term “identical proteins” is defined to mean that the two proteins have similar or identical amino acid sequences. . In a preferred embodiment, an identical protein is one that exhibits at least 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90% or more such as 95%, 97%, 98% or 99% sequence identity to the wild type protein. In another preferred embodiment, an identical protein is one that exhibits at least 60% sequence identity to the wild type protein, more preferred is at least 70% sequence identity. Even more preferred are identical proteins that exhibit at least 80%, 85% or 90% sequence identity to the wild type protein. In a yet more preferred embodiment, an identical protein exhibits at least 95%, 97%, 98% or 99% sequence identity. As used herein, identity between two regions of amino acid sequence (especially with respect to predicted structural similarities) is interpreted as implying similarity in function.

[079] Sequence identity for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. See, e.g., the Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using measure of identity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as “Gap” and “Bestfit” which can be used with default parameters to determine sequence identity or sequence identity between closely related polypeptides, such as identical polypeptides from different species of organisms or between a wild type protein and a mutin thereof. See, e. g., GCG Version 6.1. Alignment for purposes of determining percent amino acid sequence identity can be

achieved in various ways that are within the skill in the art, for instance, using publically available computer software such as BLAST, ALIGN, or Megalign (DNASTAR) software. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximum alignment over the full length of the sequences being compared. The same is true for nucleotide sequences disclosed herein.

[080] The term “isolated” indicates that the cell or cells, or the peptide(s) or nucleic acid molecule(s) has/have been removed from its/their normal physiological environment, e.g. a natural source, or that a peptide or nucleic acid is synthesized. Use of the term “isolated” indicates that a naturally occurring sequence has been removed from its normal cellular (i.e., chromosomal) environment. Thus, the sequence may be in a cell-free solution or placed in a different cellular environment. An isolated cell or isolated cells may for instance be included in a different medium such as an aqueous solution than provided originally, or placed in a different physiological environment. Typically isolated cells, peptides or nucleic acid molecule(s) constitute a higher fraction of the total cells, peptides or nucleic acid molecule(s) present in their environment, e.g. solution/suspension as applicable, than in the environment from which they were taken. By “isolated” in reference to a polypeptide or nucleic acid molecule is meant a polymer of amino acids (2 or more amino acids) or nucleotides coupled to each other, including a polypeptide or nucleic acid molecule that is isolated from a natural source or that is synthesized. The term “isolated” does not imply that the sequence is the only amino acid chain or nucleotide chain present, but that it is essentially free, e.g. about 90 - 95% pure or more, of e.g. non-amino acid material and/or non-nucleic acid material, respectively, naturally associated with it.

[081] Isolation of a desired population of cells may in some embodiments include general cell enrichment techniques such as centrifugation, filtration or cell chromatography. Generally, isolating or enriching a desired population of cells may be carried out according to any desired technique known in the art. In some embodiments isolation of a desired population of cells may include the use of a commercially available cell isolation kit. “Isolating/isolation of itaconic acid” may in some embodiments include isolation of itaconic acid from the medium in which the host cell is cultured. This can be done e.g. after the cells were lysed or without lysis of the cells. Also it is possible, that the host cells are removed from the medium before isolation of itaconic acid takes place. Methods to isolate molecules such as itaconic acid are known to the person skilled in the art. One possibility is to isolate itaconic acid e.g. from the medium in which the host cell is cultured is via high pressure liquid chromatography (HPLC).

[082] The term “MUMDB” as used herein refers to the MIPS *Ustilago maydis* DataBase provided by the Helmholtz Zentrum Muenchen. Access to that page can be obtained via the webpage <http://mips.helmholtz-muenchen.de/genre/proj/ustilago>.

[083] The term “nucleic acid molecule” as used herein refers to any nucleic acid in any possible configuration, such as single stranded, double stranded or a combination thereof. Examples of nucleic acids include for instance DNA molecules, RNA molecules, analogues of the DNA or RNA generated using nucleotide analogues or using nucleic acid chemistry, locked nucleic acid molecules (LNA), protein nucleic acids molecules (PNA), alkylphosphonate and alkylphosphotriester nucleic acid molecules and tecto-RNA molecules (e.g. Liu, B., et al., J. Am. Chem. Soc. (2004) 126, 4076-4077). LNA has a modified RNA backbone with a methylene bridge between C4' and O2', providing the respective molecule with a higher duplex stability and nuclease resistance. Alkylphosphonate and alkylphosphotriester nucleic acid molecules can be viewed as a DNA or an RNA molecule, in which phosphate groups of the nucleic acid backbone are neutralized by exchanging the P-OH groups of the phosphate groups in the nucleic acid backbone to an alkyl and to an alkoxy group, respectively. DNA or RNA may be of genomic or synthetic origin and may be single or double stranded. Such nucleic acid can be e.g. mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, a copolymer of DNA and RNA, oligonucleotides, etc. A respective nucleic acid may furthermore contain non-natural nucleotide analogues and/or be linked to an affinity tag or a label. A “nucleic acid” as used herein comprises a polynucleotide or, as also referred to herein, a nucleotide sequence. The terms “polynucleotide” and “nucleotide sequence” can thus be interchangeably used.

[084] Many nucleotide analogues are known and can be used in nucleic acids used in the methods of the invention. A nucleotide analogue is a nucleotide containing a modification at for instance the base, sugar, or phosphate moieties. As an illustrative example, a substitution of 2'-OH residues of siRNA with 2'F, 2'O-Me or 2'H residues is known to improve the in vivo stability of the respective RNA. Modifications at the base moiety may be a natural or a synthetic modification of A, C, G, and T/U, a different purine or pyrimidine base, such as uracil-5-yl, hypoxanthin-9-yl, and 2-aminoadenin-9-yl, as well as a non-purine or a non-pyrimidine nucleotide base. Other nucleotide analogues serve as universal bases. Examples of universal bases include 3-nitropyrrole and 5-nitroindole. Universal bases are able to form a base pair with any other base. Base modifications often can be combined with for example a sugar modification, such as for instance 2'-O-methoxyethyl, e.g. to achieve unique properties such as increased duplex stability.

[085] Those skilled in the art will be familiar with the fact that corresponding sequences need to be compared. The use of a corresponding sequence includes that a position is not only determined by the number of the preceding nucleotides/amino acids. Accordingly, the position of a given amino acid in accordance with the disclosure which may be substituted may vary due to deletion or addition of amino acids elsewhere in a (mutant or wild-type) protein such as NS1. Thus, by a "corresponding position" in accordance with the disclosure it is to be understood that amino acids may differ in the indicated number – for instance when comparing data base entries – but may still have similar neighbouring amino acids.

[086] The terms "overexpress," "overexpressing," "overexpressed" and "overexpression" in the present invention refer an expression of a gene or gene product or a polypeptide at a level greater than the expression of the same gene or gene product or polypeptide prior to a genetic alteration of the host cell or in a comparable host, which has not been genetically altered. If a host cell does not comprise a given gene or gene product, it is possible to introduce the gene or gene product into the host cell for expression; in this case, any detectable expression is encompassed by the term "overexpression." In one embodiment, overexpression is achieved by expressing 2, 3, 4 or more copies of said polynucleotide(s) in said host cell.

[087] Overexpression can be achieved in a number of ways. In general, it can be achieved by increasing transcription/translation of the gene, e.g. by increasing the copy number of the gene or altering or modifying regulatory sequences or sites associated with expression of a gene. For example, the gene can be operably linked to a strong constitutive promoters and/or strong ubiquitous promoters in order to reach high expression levels. Alternatively, it is possible to remove regulatory sequences such that expression becomes constitutive. One can substitute a promoter with a heterologous promoter which increases expression of the gene or leads to constitutive expression of the gene. Using inducible promoters additionally make it possible to increase the expression in the course of the translation of the gene by fermentation. Furthermore, overexpression can also be achieved by, for example, modifying the chromosomal location of a particular gene, altering nucleic acid sequences adjacent to a particular gene such as a ribosome binding site or transcription terminator, modifying proteins (e.g., regulatory proteins, suppressors, enhancers, transcriptional activators and the like) involved in transcription of the gene and/or translation of the gene product, or any other conventional means of deregulating expression of a particular gene routine in the art (including but not limited to use of antisense nucleic acid molecules, for example, to block expression of repressor proteins or deleting or mutating the

gene for a transcriptional factor which normally represses expression of the gene desired to be overexpressed. Prolonging the life of the mRNA may also improve the level of expression. If multiple copies of genes are included, the genes can either be located in plasmids of variable copy number or integrated and amplified in the chromosome.

5       **[088]** Those skilled in the art will find relevant instructions in Martin et al. (Bio/Technology 5, 137-146 (1987)), Guerrero et al. (Gene 138, 35-41 (1994)), Tsuchiya and Morinaga (Bio/Technology 6, 428-430 (1988)), Eikmanns et al. (Gene 102, 93-98 (1991)), EP 0 472 869, US 4,601,893, Schwarzer and Pühler (Bio/Technology 9, 84-87 (1991)), Reinscheid et al. (Applied and Environmental Microbiology 60, 126-132 (1994)),  
10       LaBarre et al. (Journal of Bacteriology 175, 1001- 1007 (1993)), WO 96/15246, Malumbres et al. (Gene 134, 15- 24 (1993)), JP-A-10-229891, Jensen and Hammer (Biotechnology and Bioengineering 58, 191-195 (1998)) and Makrides (Microbiological Reviews 60, 512-538 (1996)), inter alia, and in well-known textbooks on genetics and molecular biology.

15       **[089]** The terms "underexpress," "underexpressing," "underexpressed" and "underexpression" in the present invention refer an expression of a gene or gene product or a polypeptide at a level lower than the expression of the same gene or gene product or polypeptide prior to a genetic alteration of the host cell or in a comparable host which has not been genetically altered. If a host cell does not comprise a given gene or gene product, it is possible to decrease or delete the gene or gene product from the host cell. In one  
20       embodiment, underexpression is achieved by deletion of said polynucleotide(s) in said host cell.

25       **[090]** Underexpression can be achieved in a number of ways. In general, it can be achieved by decreasing transcription/translation of the gene, e.g. by altering or modifying regulatory sequences or sites associated with expression of a gene. Other methods that can be used to achieve underexpression are e.g. siRNA methods or gene knock-out strategies.

**[091]** The terms "polypeptide" and "protein" refer to a polymer of amino acid residues and are not limited to a certain minimum length of the product. Where both terms are used concurrently, this twofold naming accounts for the use of both terms side by side in the art.

30       **[092]** The term "promoter" as used throughout this document, refers to a nucleic acid sequence needed for gene sequence expression. Promoter regions vary from organism to organism, but are well known to those skilled in the art for different organisms. For example, in prokaryotes, the promoter region contains both the promoter (which directs the initiation of RNA transcription) as well as the DNA sequences which, when transcribed into RNA, will signal synthesis initiation. Such regions will normally include those 5'-non-coding

sequences involved with initiation of transcription and translation, such as the TATA box, capping sequence, CAAT sequence, and the like. Both constitutive and inducible promoters can be used in the present invention, in accordance with the needs of a particular embodiment. A large number of promoters recognized by a variety of potential host cells are well known. The selected promoter can be operably linked to cistron DNA encoding a polypeptide described herein by removing the promoter from the source DNA via restriction enzyme digestion and inserting the isolated promoter sequence into the vector of choice. Both the native promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of a selected nucleic acid sequence. The promoter can also be a recombinant promoter.

[093] In case of the host cell being *Ustilago maydis*, a suitable promoter is the constitutive tef, otef promoter (Spellig et al. (1996), Mol Gen Genet 252, 503-509), hsp70 promoter (Holden et al., EMBO J. 8, 1927-1934). A preferred inducible promoter is the nar1 promoter (Brachmann et al., (2001), Mol Microbiol. 42, 1047-63) or the crg1 promoter (Bottin et al. (1996), Mol Gen Genet 253, 342-352).

[094] The term "purified" is understood to be a relative indication in comparison to the original environment of the cell, thereby representing an indication that the cell is relatively purer than in the natural environment. It therefore includes, but does not only refer to, an absolute value in the sense of absolute purity from other cells (such as a homogeneous cell population). Compared to the natural level, the level after purifying the cell will generally be at least 2-5 fold greater (e.g., in terms of cells/ml). Purification of at least one order of magnitude, such as about two or three orders, including for example about four or five orders of magnitude is expressly contemplated. It may be desired to obtain the cell at least essentially free of contamination, in particular free of other cells, at a functionally significant level, for example about 90%, about 95%, or 99% pure. With regard to a nucleic acid, a peptide, a protein or a peptidomimetic, the above applies mutatis mutandis. In this case purifying the nucleic acid, peptide or protein will for instance generally be at least 2-5 fold greater (e.g., in terms of mg/ml).

[095] The word "recombinant" is used in this document to describe a nucleic acid molecule that, by virtue of its origin, manipulation, or both is not associated with all or a portion of the nucleic acid molecule with which it is associated in nature. Generally a recombinant nucleic acid molecule includes a sequence which does not naturally occur in the respective wildtype organism or cell. Typically a recombinant nucleic acid molecule is obtained by genetic engineering, usually constructed outside of a cell. Generally a

recombinant nucleic acid molecule is at least substantially identical and/or substantial complementary to at least a portion of the corresponding nucleic acid molecule occurring in nature. A recombinant nucleic acid molecule may be of any origin, such as genomic, cDNA, mammalian, bacterial, viral, semisynthetic or synthetic origin. The term "recombinant" as used with respect to a protein / polypeptide means a polypeptide produced by expression of a recombinant polynucleotide. A recombinant cell, polypeptide, or nucleic acid can be typically described with reference to how it differs from a naturally occurring counterpart (the "wild-type"). A "recombinant cell" or "recombinant host cell" refers to a cell or host cell that has been genetically altered to comprise a nucleic acid sequence which was not native to said cell.

[096] The recombinant host cell within the present invention does not necessarily contain the nucleic acid sequences encoding a protein of interest. It is appreciated by a skilled person in the art that the host cells can be provided for inserting desired nucleotide sequences into the host cell, for example, in a kit.

[097] As used herein, "engineered" host cells are host cells which have been manipulated using genetic engineering, i.e. by human intervention. When a host cell is "engineered to overexpress" a given protein, the host cell is manipulated such that the expression of the given protein is increased compared to the host cell under the same condition prior to manipulation (or "prior to engineering"). The degree of overexpression may be 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500% or more compared to the host cell under the same condition prior to manipulation (or "prior to engineering"). When a host cell is "engineered to underexpress" a given protein, the host cell is manipulated such that the expression of the given protein is decreased compared to the host cell under the same condition prior to manipulation (or "prior to engineering"). The degree of underexpression may be 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100% compared to the host cell under the same condition prior to manipulation (or "prior to engineering").

[098] The term "vector", sometimes also referred to as gene delivery system or gene transfer vehicle, relates to a macromolecule or complex of molecules that include(s) a polynucleotide to be delivered to a host cell, whether *in vitro*, *ex vivo* or *in vivo*. Typically a vector is a single or double-stranded circular nucleic acid molecule that allows or facilitates the transfer of a nucleic acid sequence into a cell. A vector can generally be transfected into cells and replicated within or independently of a cell genome. A circular double-stranded nucleic acid molecule can be cut and thereby linearized upon treatment with restriction enzymes. An assortment of nucleic acid vectors, restriction enzymes, and the knowledge of

the nucleotide sequences cut by restriction enzymes are readily available to those skilled in the art. A nucleic acid molecule encoding a peptide, such as a sequence that includes a sequence of the present invention can be inserted into a vector by cutting the vector with restriction enzymes and ligating the two pieces together. A vector may for instance be a viral vector, such as a retroviral vector, a Lentiviral vector, a herpes virus based vector or an adenoviral vector. A vector may also be a plasmid vector, which is also a typical example of a prokaryotic vector. A respective plasmid may in some embodiments be a plasmid capable of replication in *E. coli*, such as, for example, pBR322, ColEI, pSC101, pACYC 184 or  $\pi$ VX. Bacillus plasmids include pC194, pC221 or pT127. Suitable *Streptomyces* plasmids include p1J101, and *Streptomyces* bacteriophages such as  $\phi$ C31. A vector may also be a liposome-based extrachromosomal vector, also called episomal vector. Two illustrative examples of an episomal vector are an oriP-based vector and a vector encoding a derivative of EBNA-1. Lymphotropic herpes virus is a herpes virus which replicates in a lymphoblast and becomes a plasmid for a part of its natural life-cycle. A vector may also be based on an organically modified silicate. In some embodiments a vector may be a transposon-based system, i.e. a transposon/transposase system, such as the so called Sleeping Beauty, the Frog Prince transposon - transposase system or the TTAA-specific transposon piggyBac system. Transposons are mobile genetic elements in that they are sequences of DNA that can move around to different positions within the genome of a single cell, a process called transposition. In the process, a transposon can cause mutations and change the amount of DNA in the genome.

[0099] The terms “comprising”, “including,” “containing”, “having” etc. shall be read expansively or open-ended and without limitation. Singular forms such as “a”, “an” or “the” include plural references unless the context clearly indicates otherwise. Thus, for example, reference to a “vector” includes a single vector as well as a plurality of vectors, either the same - e.g. the same operon - or different. Likewise reference to “cell” includes a single cell as well as a plurality of cells. Unless otherwise indicated, the term “at least” preceding a series of elements is to be understood to refer to every element in the series. The terms “at least one” and “at least one of” include for example, one, two, three, four, or five or more elements. It is furthermore understood that slight variations above and below a stated range can be used to achieve substantially the same results as a value within the range. Also, unless indicated otherwise, the disclosure of ranges is intended as a continuous range including every value between the minimum and maximum values.

[0100] The scope and meaning of any use of a term will be apparent from the

specific context in which the term is used. Certain further definitions for selected terms used throughout this document are given in the appropriate context of the detailed description, as applicable. Unless otherwise defined, all other scientific and technical terms used in the description, figures and claims have their ordinary meaning as commonly understood by one of ordinary skill in the art.

*ADI/TAD and further genes and proteins*

**[0101]** The nucleic acids and proteins of the present invention as well as their abbreviation are disclosed below.

| Abbreviation         | UM Number<br>in MUMDB | Protein SEQ ID                   | Nucleic acid SEQ ID             |
|----------------------|-----------------------|----------------------------------|---------------------------------|
| ADI                  | UM11778               | SEQ ID NO: 2 or<br>SEQ ID NO: 10 | SEQ ID NO: 1 or<br>SEQ ID NO: 7 |
| TAD                  | UM05076               | SEQ ID NO: 13                    | SEQ ID NO: 6                    |
| MFS                  | UM11777               | SEQ ID NO: 14                    | SEQ ID NO: 3                    |
| CTP1                 | UM05079               | SEQ ID NO: 15                    | SEQ ID NO: 4                    |
| Transcription factor | UM05080               | SEQ ID NO: 16                    | SEQ ID NO: 8                    |
| P450-Monooxygenase   | UM05074               | SEQ ID NO: 11                    | SEQ ID NO: 9                    |
| ADI2                 | UM02807               | SEQ ID NO: 18                    | SEQ ID NO: 17                   |
| CFP1                 | UM02806               | SEQ ID NO: 20                    | SEQ ID NO: 19                   |
| regulator            | UM02808               | SEQ ID NO: 22                    | SEQ ID NO: 21                   |

**[0102]** A suitable polynucleotide sequence of an aconitate-delta-isomerase (ADI) and/or a trans-aconitate decarboxylase (TAD) according to the invention, or a complement of such a polynucleotides sequence, may be included in any nucleic acid molecule. In some embodiments such a polynucleotide sequence is operably linked to a sequence that is capable of regulating gene expression. The polynucleotide sequence may for instance be operably linked to a promoter. As a further example, the polynucleotide sequence may be operably linked to an enhancer, including a silencer. In some embodiments the sequence capable of regulating gene expression is located on the same nucleic acid molecule as the polynucleotide sequence according to the invention. In some embodiments the sequence capable of regulating gene expression is located on a nucleic acid molecule that is different from the nucleic acid molecule on which the polynucleotide sequence according to the

invention is located.

[0103] A polynucleotide sequence of an ADI (UM11778) included in a nucleic acid molecule according to the invention in some embodiments is at least essentially identical to the sequence of SEQ ID NO: 1. The polynucleotide sequence of an ADI according to the invention may have at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1.

[0104] In some embodiments a nucleic acid molecule according to the invention includes a sequence that encodes a functional fragment of an ADI. Such a polynucleotide sequence may for example include a sequence of about 1000 contiguous bases of SEQ ID NO: 1. In some embodiments a respective polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 1.

[0105] In some embodiments the polynucleotide sequence of an ADI (UM11778) is at least essentially identical to the sequence of SEQ ID NO: 7. The polynucleotide sequence of an ADI according to the invention may have at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 7.

[0106] Another polynucleotide sequence of an aconitate-delta-isomerase is a polynucleotide sequence on an ADI2 (UM02807). The polynucleotide sequence of an ADI2 (UM02807) is at least essentially identical to the sequence of SEQ ID NO: 17. The polynucleotide sequence of an ADI2 according to the invention may have at least 30% sequence identity with the sequence of SEQ ID NO: 17. It is envisioned by the invention that an ADI2 can substitute an ADI. Hence, embodiments comprising ADI also envision embodiments in which ADI is substituted by ADI2 as well as embodiments, where ADI and ADI2 are present. The terms "ADI" and "ADI2" as used in this paragraph refer to the polynucleotide sequence encoding for aconitate-delta-isomerase as well as to a polypeptide encoded by said polynucleotide sequences.

[0107] In some embodiments a nucleic acid molecule according to the invention includes a functional fragment of the sequence of SEQ ID NO: 7. The functional fragment may be a polynucleotide sequence that includes a sequence of about 1000 contiguous bases of SEQ ID NO: 7. In some embodiments such a polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 7.

[0108] A polynucleotide sequence of a TAD (UM05076) included in a nucleic acid molecule according to the invention in some embodiments is at least essentially identical to the sequence of SEQ ID NO: 6. The polynucleotide sequence of a TAD according to the invention may have at least 50% sequence identity with the sequence of SEQ ID NO: 6.

[0109] In some embodiments a nucleic acid molecule according to the invention includes a sequence that encodes a functional fragment of a TAD. Such a polynucleotide sequence may for example include a sequence of about 1000 contiguous bases of SEQ ID NO: 6 or an identical sequence of the sequence of this length. In some embodiments a  
5 respective polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 6 or an identical sequence of the sequence of this length.

[0110] A polynucleotide sequences of an ADI and a TAD included in a nucleic acid molecule according to the invention in some embodiments is at least essentially identical to the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and SEQ ID NO: 6. The  
10 polynucleotide sequence of an ADI and a TAD according to the invention may have at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and SEQ ID NO: 6.

[0111] In some embodiments a nucleic acid molecule according to the invention includes a sequence that encodes a functional fragment of an ADI and a TAD. Such a  
15 polynucleotide sequence may for example include a sequence of about 1000 contiguous bases of SEQ ID NO: 1 and/or SEQ ID NO: 7 and SEQ ID NO: 6 or an identical sequence of the sequence of this length. In some embodiments a respective polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 1 and/or SEQ ID NO: 7 and SEQ ID NO: 6 or an identical sequence of the sequence of this length.

[0112] Similarly, an isolated nucleic acid molecule of the invention can  
20 comprise polynucleotide sequences wherein the polynucleotide having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and SEQ ID NO: 6 encoding a polypeptide having activity of an aconitate-delta-isomerase (ADI) or encoding a polypeptide having activity of a trans-aconitate decarboxylase (TAD).

[0113] A polynucleotide sequence of a CTP1 (UM05079) included in a nucleic acid molecule according to the invention in some embodiments is at least essentially identical to the sequence of SEQ ID NO: 4. The polynucleotide sequence of a CTP1  
25 according to the invention may have at least 50% sequence identity with the sequence of SEQ ID NO: 4.

[0114] In some embodiments a nucleic acid molecule according to the invention includes a sequence that encodes a functional fragment of a CTP1. Such a polynucleotide sequence may for example include a sequence of about 1000 contiguous bases of SEQ ID  
30 NO: 4 or an identical sequence of the sequence of this length. In some embodiments a

respective polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 4 or an identical sequence of the sequence of this length.

[0115] A polynucleotide sequence of a MFS (UM11777) included in a nucleic acid molecule according to the invention in some embodiments is at least essentially identical to the sequence of SEQ ID NO: 3. The polynucleotide sequence of a MFS according to the invention may have at least 50% sequence identity with the sequence of SEQ ID NO: 3.

[0116] In some embodiments a nucleic acid molecule according to the invention includes a sequence that encodes a functional fragment of a MFS. Such a polynucleotide sequence may for example include a sequence of about 1500 contiguous bases of SEQ ID NO: 3 or an identical sequence of the sequence of this length. In some embodiments a respective polynucleotide sequence may include a sequence of about 950 contiguous bases of SEQ ID NO: 3 or an identical sequence of the sequence of this length.

[0117] A polynucleotide sequence of a transcription factor (UM05080) included in a nucleic acid molecule according to the invention in some embodiments is at least essentially identical to the sequence of SEQ ID NO: 8. The polynucleotide sequence of a transcription factor according to the invention may have at least 50% sequence identity with the sequence of SEQ ID NO: 8.

[0118] In some embodiments a nucleic acid molecule according to the invention includes a sequence that encodes a functional fragment of the transcription factor. Such a polynucleotide sequence may for example include a sequence of about 1000 contiguous bases of SEQ ID NO: 8 or an identical sequence of the sequence of this length. In some embodiments a respective polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 8 or an identical sequence of the sequence of this length.

[0119] A polynucleotide sequence of a cytochrome P450-Monooxygenase (UM05074) included in a nucleic acid molecule according to the invention in some embodiments is at least essentially identical to the sequence of SEQ ID NO: 9. The polynucleotide sequence of a cytochrome P450-Monooxygenase according to the invention may have at least 50% sequence identity with the sequence of SEQ ID NO: 9.

[0120] In some embodiments a nucleic acid molecule according to the invention includes a sequence that encodes a functional fragment of the cytochrome P450 monooxygenase. Such a polynucleotide sequence may for example include a sequence of about 1500 contiguous bases of SEQ ID NO: 9 or an identical sequence of the sequence of this length. In some embodiments a respective polynucleotide sequence may include a sequence of about 950 contiguous bases of SEQ ID NO: 9 or an identical sequence of the

sequence of this length.

[0121] In addition, the polynucleotide sequence may include a nucleotide sequence, which results from the addition, deletion or substitution of at least one nucleotide to the 5'-end and/or the 3'-end of the nucleic acid formula shown in a given sequence. Any nucleotide or polynucleotide may be used in this regard, provided that its addition, deletion or substitution does not alter the amino acid sequence, which is encoded by the nucleotide sequence. For example, the present invention is intended to include any nucleic acid sequence resulting from the addition of ATG as an initiation codon at the 5'-end of the inventive nucleic acid sequence or its derivative, or from the addition of TTA, TAG or TGA as a termination codon at the 3'-end of the inventive nucleotide sequence or its derivative. Moreover, a nucleic acid molecule according to the present invention may, as necessary, have restriction endonuclease recognition sites added to its 5'-end and/or its 3'-end. Such functional alterations of a given nucleic acid sequence afford an opportunity to promote secretion and/or processing of heterologous proteins encoded by foreign nucleic acid sequences fused thereto.

[0122] Hence, in some embodiments a nucleic acid of the inventions has a polynucleotide sequence that lacks a portion of a SEQ ID NO defined in this document. Typically an encoded amino acid stretch is lacking so that the remaining amino acids are in the form of a shortened polypeptide, i.e., the respective amino acid stretch is deleted "in frame" and accordingly the encoding polynucleotide sequence lacks a base triplet or a plurality of consecutive base triplets.

[0123] Further, it is possible to delete codons or to substitute one or more codons with codons other than degenerate codons to produce a structurally modified polypeptide, but one which has substantially the same utility or activity as the polypeptide produced by the unmodified nucleic acid molecule. As recognized in the art, the two polypeptides are functionally equivalent, as are the two nucleic acid molecules that give rise to their production, even though the differences between the nucleic acid molecules are not related to the degeneracy of the genetic code.

[0124] A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional ADI. The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 2. In some embodiments the polypeptide sequence of an ADI according to the invention has at least 30%, 40%, or 50% sequence identity with the

sequence of SEQ ID NO: 2.

[0125] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 2. The functional fragment may be a polynucleotide sequence that includes a sequence of about 400 contiguous amino acids of SEQ ID NO: 2. In some embodiments such an amino acid sequence may include a sequence of about 350 contiguous bases of SEQ ID NO: 2.

[0126] As indicated above, instead of or in addition to a peptide, a peptidomimetic may likewise be used in the context of the present invention. The term “peptidomimetic” as used herein refers to a compound that has the same general structure as a corresponding polypeptide, but which includes modifications that increase its stability or biological function. In some embodiments a peptidomimetic may include one or more D-amino acids, essentially consist of D-amino acids or consist of D-amino acids. D-amino acids are the optical isomer of a naturally occurring L-amino acid. A D-amino acid can be taken to be a mirror image of a L-amino acid. Stretches of D-amino acids are less prone to be degraded in a host organism via proteolysis. In some embodiments a peptidomimetic may be an inverso analog, which is an analog of the same sequence that consists only of D-amino acids. In some embodiments a peptidomimetic may be a “reverse” analogue of a given peptide, which means that the peptidomimetic includes the reverse sequence of the peptide. In some embodiments a peptidomimetic may be a “D-retro-enantiomer peptide”, which is an analog that consists of D-amino acids, with the sequence arranged in the reversed order. A peptidomimetic may also include, essentially consist of or consist of a peptoid. A peptoid differs from peptides in that the side chain is connected to the amide nitrogen rather than the carbon atom. A peptoid can thus be taken to be an oligo(N-alkyl) glycine, which nevertheless has the same or substantially the same amino acid sequence as the corresponding polypeptide. Peptoids are typically resistant to proteases and other modifying enzymes and can have a much higher cell permeability than peptides (see e.g. Kwon, Y.-U., and Kodadek, T., J. Am. Chem. Soc. (2007) 129, 1508-1509).

[0127] The peptide or peptidomimetic may be prepared by any method, such as by synthesizing the peptide or peptidomimetic, or by expressing a nucleic acid encoding an appropriate amino acid sequence in a cell and harvesting the peptide from the cell. A combination of such methods may likewise be used. Methods of *de novo* synthesizing peptides and peptidomimetics, and methods of recombinantly producing peptides and peptidomimetics are well known in the art.

[0128] As mentioned above, in some embodiments a sequence of the invention

corresponds to one of the SEQ ID NOs of this document, such as a sequence corresponding to SEQ ID NO: 1 or SEQ ID NO: 7, and contains a conservative substitution. Conservative substitutions are generally the following substitutions, listed according to the amino acid to be mutated, each followed by one or more replacement(s) that can be taken to be conservative: Ala → Gly, Ser, Val; Arg → Lys; Asn → Gln, His; Asp → Glu; Cys → Ser; Gln → Asn; Glu → Asp; Gly → Ala; His → Arg, Asn, Gln; Ile → Leu, Val; Leu → Ile, Val; Lys → Arg, Gln, Glu; Met → Leu, Tyr, Ile; Phe → Met, Leu, Tyr; Ser → Thr; Thr → Ser; Trp → Tyr; Tyr → Trp, Phe; Val → Ile, Leu. Other substitutions are also permissible and can be determined empirically or in accord with other known conservative or non-conservative substitutions. As a further orientation, the following eight groups each contain amino acids that can typically be taken to define conservative substitutions for one another:

- 1) Alanine (Ala), Glycine (Gly);
- 2) Aspartic acid (Asp), Glutamic acid (Glu);
- 3) Asparagine (Asn), Glutamine (Gln);
- 4) Arginine (Arg), Lysine (Lys);
- 5) Isoleucine (Ile), Leucine (Leu), Methionine (Met), Valine (Val);
- 6) Phenylalanine (Phe), Tyrosine (Tyr), Tryptophan (Trp);
- 7) Serine (Ser), Threonine (Thr); and
- 8) Cysteine (Cys), Methionine (Met)

**[0129]** A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional ADI. The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 10. In some embodiments the polypeptide sequence of a ADI according to the invention has at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 10.

**[0130]** polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional ADI2. The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 18. In some embodiments the polypeptide sequence of an ADI according to the invention has at least 30% sequence identity with the sequence of SEQ ID NO: 18.

[0131] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 10. The functional fragment may be a polynucleotide sequence that includes a sequence of about 400 contiguous amino acids of SEQ ID NO: 10. In some embodiments such an amino acid sequence may include a sequence of about 350 contiguous bases of SEQ ID NO: 10.

[0132] A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional TAD (trans-aconitate decarboxylase). The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 13. In some embodiments the polypeptide sequence of a TAD according to the invention has at least 50% sequence identity with the sequence of SEQ ID NO: 13.

[0133] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 13. The functional fragment may be a polynucleotide sequence that includes a sequence of about 400 contiguous amino acids of SEQ ID NO: 13. In some embodiments such an amino acid sequence may include a sequence of about 450 contiguous bases of SEQ ID NO: 13.

[0134] A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional CTP1 (mitochondrial transporter CTP1). The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 15. In some embodiments the polypeptide sequence of a CTP1 according to the invention has at least 50% sequence identity with the sequence of SEQ ID NO: 15.

[0135] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 15. The functional fragment may be a polynucleotide sequence that includes a sequence of about 200 contiguous amino acids of SEQ ID NO: 15. In some embodiments such an amino acid sequence may include a sequence of about 250 contiguous bases of SEQ ID NO: 15.

[0136] A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional MFS (transporter of the major facilitator superfamily). The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 14. In some

embodiments the polypeptide sequence of a MFS according to the invention has at least 50% sequence identity with the sequence of SEQ ID NO: 14.

[0137] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 14. The functional  
5 fragment may be a polynucleotide sequence that includes a sequence of about 400 contiguous amino acids of SEQ ID NO: 14. In some embodiments such an amino acid sequence may include a sequence of about 450 contiguous bases of SEQ ID NO: 14.

[0138] A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence  
10 of a functional transcription factor (UM05080). The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 16. In some embodiments the polypeptide sequence of a transcription factor according to the invention has at least 50% sequence identity with the sequence of SEQ ID NO: 16.

[0139] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 16. The functional  
15 fragment may be a polynucleotide sequence that includes a sequence of about 300 contiguous amino acids of SEQ ID NO: 16. In some embodiments such an amino acid sequence may include a sequence of about 350 contiguous bases of SEQ ID NO: 16.

[0140] A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence  
20 of a functional ADI and a sequence of a functional TAD. The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 10 and SEQ ID NO: 13. In some embodiments the polypeptide sequence of an ADI and a TAD according to the  
25 invention has at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 10 and at least 50% sequence identity with the sequence of SEQ ID NO: 13.

[0141] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 10 and SEQ ID  
30 NO: 13. The functional fragment may be a polynucleotide sequence that includes a sequence of about 400 contiguous amino acids of SEQ ID NO: 10 and SEQ ID NO: 13. In some embodiments such an amino acid sequence may include a sequence of about 350 contiguous bases of SEQ ID NO: 10 and about 450 contiguous bases of SEQ ID NO: 13.

[0142] A polypeptide/protein/peptidomimetic according to the invention may be a

polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional ADI and a sequence of a functional TAD. The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 2 and SEQ ID NO: 13. In some embodiments the polypeptide sequence of an ADI and a TAD according to the invention has at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2 and at least 50% sequence identity with the sequence of SEQ ID NO: 13.

[0143] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 2 and SEQ ID NO: 13. The functional fragment may be a polynucleotide sequence that includes a sequence of about 400 contiguous amino acids of SEQ ID NO: 2 and SEQ ID NO: 13. In some embodiments such an amino acid sequence may include a sequence of about 350 contiguous bases of SEQ ID NO: 2 and about 450 contiguous bases of SEQ ID NO: 13.

[0144] A polypeptide/protein/peptidomimetic according to the invention may be a polypeptide molecule that has, including consists of and essentially consists of, a sequence of a functional ADI. The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 2. In some embodiments the polypeptide sequence of an ADI according to the invention has at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2.

[0145] In some embodiments a polypeptide or peptidomimetic according to the invention includes a functional fragment of the sequence of SEQ ID NO: 2. The functional fragment may be a polynucleotide sequence that includes a sequence of about 400 contiguous amino acids of SEQ ID NO: 2. In some embodiments such an amino acid sequence may include a sequence of about 350 contiguous bases of SEQ ID NO: 2.

[0146] Similarly, the present invention relates to an isolated polypeptide molecule comprising an amino acid sequences having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and having at least 50% sequence identity with the sequence of SEQ ID NO: 13 and having ADI activity or TAD activity.

[0147] The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and/or SEQ ID NO: 13 and SEQ ID NO: 14. In some

embodiments the polypeptide sequence according to the invention has at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and/or at least 50% sequence identity with the sequence of SEQ ID NO: 13 and SEQ ID NO: 14.

[0148] The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and/or SEQ ID NO: 13 and SEQ ID NO: 15. In some embodiments the polypeptide sequence according to the invention has at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and/or at least 50% sequence identity with the sequence of SEQ ID NO: 13 and SEQ ID NO: 15.

[0149] The polypeptide in some embodiments includes, including consist of and at least essentially consists of, a sequence that is at least essentially identical to the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and/or SEQ ID NO: 13 and SEQ ID NO: 16. In some embodiments the polypeptide sequence according to the invention has at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and/or at least 50% sequence identity with the sequence of SEQ ID NO: 13 and SEQ ID NO: 16.

[0150] In the absence of any identifiable protein with a sequence of 70 % or more amino acid identity, a protein of *Pseudogulbenkiania* sp. NH8B with GenBank GENE ID: 11156442 NH8B\_2749 (SEQ ID NO: 23), having 53 % amino acid identity to the protein of SEQ ID NO: 10 may serve as an illustrative example. This protein has at amino acid positions 12 to 17 the sequence GGTSKG, whereas the protein of SEQ ID NO: 10 has at amino acid positions 12 to 17 the sequence AGTSRG. The exchange of R to K at position 16 and the exchange of A to G at position 12 can both be taken to define a conservative substitution. Further, the protein of *Pseudogulbenkiania* sp. NH8B has at amino acid positions 19 to 29 the sequence FFLAD DLPADP, while at amino acid positions 19 to 29 the protein of SEQ ID NO: 10 has the sequence YFLASDLPAEP. The exchange of Y to F at position 19 and the exchange of E to D at position 28 can both be taken to define a conservative substitution. In contrast thereto, the exchange of S to D at position 23 can be taken to define a non-conservative substitution.

[0151] As already implied above, more substantial changes, such as the following, do not represent conservative substitutions: Ala → Leu, Ile; Arg → Gln; Asn → Asp, Lys, Arg, His; Asp → Asn; Cys → Ala; Gln → Glu; Glu → Gln; His → Lys; Ile → Met, Ala, Phe; Leu → Ala, Met, Norleucine; Lys → Asn; Met → Phe; Phe → Val, Ile, Ala; Trp → Phe; Tyr → Thr, Ser; Val → Met, Phe, Ala.

[0152] A polynucleotide sequence according to the invention may also be a

sequence that is capable of hybridizing to a sequence encoding a ADI, TAD, CTP1, MFS, transcription factor having activity for one of nucleic acids of SEQ ID NO: 1, 7, 3, 4, 6, 8 and/or 9, (e.g. UM05080; SEQ ID NO: 8) or a cytochrome P450 monooxygenase (e.g. UM05074; SEQ ID NO: 9). The nucleic acid molecule may for instance be capable of hybridizing to a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity.

[0153] The nucleic acid molecule may for instance be capable of hybridizing to a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 7, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity.

[0154] The nucleic acid molecule may for instance be capable of hybridizing to a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity.

[0155] The nucleic acid molecule may for instance be capable of hybridizing to a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a mitochondrial citrate transporter, preferably of a mitochondrial cis-aconitate transporter.

[0156] The nucleic acid molecule may for instance be capable of hybridizing to a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a multidrug transporter of the major facilitator superfamily, preferably of a transporter which exports itaconic acid out of the host cell.

[0157] The nucleic acid molecule may for instance be capable of hybridizing to a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a transcription factor for at least one of polynucleotide sequences of any one of SEQ ID NOs: 1, 3, 4, and/or 6.

[0158] In some embodiments a nucleic acid molecule of the invention is a sequence that is complementary to the sequence of an ADI (SEQ ID NO: 1 and/or 7), an

ADI2 (SEQ ID NO: 17), a TAD (SEQ ID NO: 6), a CTP1 (SEQ ID NO: 4), a MFS (SEQ ID NO: 3), a transcription factor having activity for one of nucleic acids of SEQ ID NO: 1, 7, 3, 4, 6, 8 and/or 9, (e.g. UM05080; SEQ ID NO: 8) or a cytochrome P450-monooxygenase (e.g. UM05074; SEQ ID NO: 9). A nucleotide sequence is complementary to or the complement of another nucleotide sequence if all of the nucleotides of the first sequence are complementary to all of the nucleotides of the second sequence. Accordingly, the respective nucleotide sequence will specifically hybridise to, or undergo duplex formation with, the respective portion of the target nucleic acid molecule under suitable hybridisation assay conditions, in particular of ionic strength and temperature.

**[0159]** As an illustrative example, the respective nucleic acid sequence may be included in a single-stranded nucleic acid molecule. Such a single-stranded nucleic acid molecule may have a nucleic acid sequence that is at least partially complementary to at least a portion of a strand of the sequence of an ADI (SEQ ID NO: 1 and/or 7), an ADI2 (SEQ ID NO: 17), a TAD (SEQ ID NO: 6), a CTP1 (SEQ ID NO: 4), a MFS (SEQ ID NO: 3), a transcription factor having activity for one of nucleic acids of SEQ ID NO: 1, 7, 3, 4, 6, 8 and/or 9 (e.g. UM05080; SEQ ID NO: 8) or a cytochrome P450 monooxygenase (e.g. UM05074; SEQ ID NO: 9). The respective nucleotide sequence may for example be 50, 60, 70, for example 80 or 85, including 100 % identical to another nucleic acid sequence. The higher the percentage to which the two sequences are complementary to each other (i.e. the lower the number of mismatches), the easier will they form a complex by hybridization. In typical embodiments the respective nucleotide sequence is substantially complementary to at least a portion of the sequence of an ADI (SEQ ID NO: 1 and/or 7), an ADI2 (SEQ ID NO: 17), a TAD (SEQ ID NO: 6), a CTP1 (SEQ ID NO: 4), a MFS (SEQ ID NO: 3), transcription factor having activity for one of nucleic acids of SEQ ID NO: 1, 7, 3, 4, 6, 8 and/or 9 (e.g. UM05080; SEQ ID NO: 8) or a cytochrome P450-monooxygenase (e.g. UM05074; SEQ ID NO: 9). “Substantially complementary” as used in this document refers to the fact that a given nucleic acid sequence is at least 90 % identical to another nucleic acid sequence. A substantially complementary nucleic acid sequence is in some embodiments about 95 % or more identical to another nucleic acid sequence. The term “complementary” or “complement” refers to two nucleotides that can form multiple favourable interactions with one another. Such favourable interactions are specific association between opposing or adjacent pairs of nucleic acid (including nucleic acid analogue) strands via matched bases, and include Watson-Crick base pairing. As an illustrative example, in two given nucleic acid molecules (e.g. DNA molecules) the base adenosine is complementary to thymine or uracil,

while the base cytosine is complementary to guanine.

[0160] Interactions between two or more nucleic acid molecules are generally sequence driven interactions referred to as hybridization. Sequence driven interaction is an interaction that occurs between two nucleotides or nucleotide analogs or nucleotide derivatives in a nucleotide specific manner (supra). Typically sequence driven interactions occur on the Watson-Crick face or Hoogsteen face of the respective nucleotide. The hybridization of two nucleic acid molecules is affected by a number of conditions and parameters known to those skilled in the art. For example, the concentrations of salts, pH, and temperature of the reaction all affect whether two nucleic acid molecules will hybridize. In some embodiments selective hybridization conditions can be defined as stringent hybridization conditions. For example, stringency of hybridization is controlled by both temperature and salt concentration of either or both of the hybridization and washing steps. For example, conditions of hybridization that achieve selective interactions between complementary sequences may involve hybridization in high ionic strength solution ( $6 \times$  SSC or  $6 \times$  SSPE) at a temperature that is in the range from about 12 to about 25 °C below the  $T_m$ , the melting temperature at which half of the molecules of a sequence dissociate from hybridization partners that are a perfectly matched probe, followed by washing at a combination of temperature and salt concentration chosen so that the washing temperature is in the range from about 5 °C to about 20 °C below the  $T_m$ . The temperature and salt conditions are readily determined empirically in preliminary experiments in which samples of reference DNA immobilized on filters are hybridized to a labeled nucleic acid of interest and then washed under conditions of different stringencies. Hybridization temperatures are typically higher for DNA-RNA and RNA-RNA hybridizations than for DNA-DNA hybridizations.

[0161] A polynucleotide sequence of the invention may for instance be complementary to the sequence of SEQ ID NO: 7 or to the sequence of SEQ ID NO: 1. A polynucleotide sequence of the invention may be fully complementary to one of these sequences, i.e. have a sequence of nucleotides, where in the form of this sequence each of the nucleotides can undergo specific association such as Watson-Crick base pairing with a nucleotide in the sequence of SEQ ID NO: 7 or of SEQ ID NO: 1.

[0162] A nucleic acid molecule according to the present invention may also include a sequence that is capable of modulating transcription, also called transcriptional control elements suitable such as a promoter or an enhancer (supra). A nucleic acid molecule according to the present invention may also include a polyadenylation signal, a transcription

pausing signal or a transcription termination signal. A sequence capable of modulating transcription, including a transcription termination site, may be operably linked to a polynucleotide sequence as described above. For proper expression of a respective polypeptide, a suitable translational control element may be operably linked to the encoding

5 polynucleotide sequence, such as e.g. a suitable leader sequence, a non-translated region of an mRNA that is important for translation by the cell. The leader sequence is operably linked to the 5'-terminus of the nucleic acid sequence encoding the ADI. Any leader sequence, which is functional in the selected host cell, may be used in the context of the present invention. A further illustrative example of a suitable translational control element is a 5'

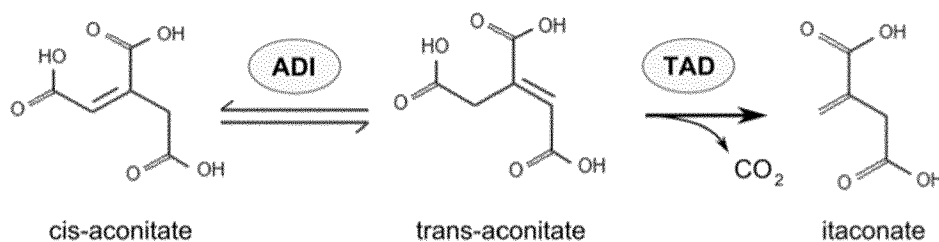
10 untranslated region leading to a 5' cap structure suitable for recruiting ribosomes, and a stop codon to terminate the translation process. In some embodiments a nucleic acid molecule of the invention includes an expression cassette (cf. above). Such an expression cassette may for example have a sequence capable of modulating transcription, such as a promoter and/or a transcription termination site, operably linked to the sequence as described above. A

15 nucleic acid molecule according to the present invention may be an isolated molecule, for example isolated from one or more cells or from tissue. A nucleic acid molecule of the invention may be a recombinant molecule. In some embodiments a nucleic acid molecule of the invention is included in a vector (supra).

The sequence of an ADI according to the invention encodes a biologically active

20 polypeptide/protein in that such a polypeptide has activity of an aconitate-delta-isomerase. This cytoplasmic enzyme (EC 5.3.3.7) is involved in ITA formation by converting the TCA cycle intermediate cis-aconitic acid into trans-aconitic acid. The sequence of a TAD according to the invention encodes a biologically active polypeptide/protein in that such polypeptide has activity of a trans-aconitate decarboxylase (TAD). This cytoplasmic enzyme

25 (no EC number attributed yet) catalyzes the second step of ITA formation by converting trans-aconitic acid into itaconic acid:



[0163] The conversion of cis-aconitate to ITA is in *A. terreus* the last step of a

biosynthesis that includes the Embden-Meyerhof pathway followed by the TCA cycle. In *A. terreus* the enzyme CAD has the sequence of SwissProt/Uniprot accession No B3IUN8 (version 11 of 3 October 2012), encoded by the gene of GenBank accession No DM010748 (version DM010748.1, GI: 224430794). Since ITA is not essential for the viability of *U. maydis*, it is regarded a secondary metabolite. Secondary metabolism genes are often arranged in clusters and are coregulated (cf. Fig. 1). Although the general enzyme activities implicated in ITA production are similar in *A. terreus* and *U. maydis*, the sequences of the genes or enzymes show little to no relation to each other. The present inventors found in cultivation experiments with the *U. maydis* strain deleted for the UM11778 gene that the enzyme encoded by this gene plays a crucial role in the ITA synthesis pathway. Its deletion almost completely disrupted the ITA production. In addition, deletion of UM05076 resulted in a total loss of ITA production, which indicates that both genes are essential in the ITA synthesis pathway of *U. maydis*. The deletion of the two other genes, UM06058 and UM02807, encoding proteins, which have a high sequence similarity to UM11778, did not result in any detectable decrease in ITA production.

[0164] The enzyme methylitaconate-Delta-isomerase (Mii), which shows 33% similarity to the *U. maydis* ADI, has previously been identified in *Eubacterium barkeri* (Velarde 2009). Although this Mii enzyme can catalyze the conversion of citraconate to itaconate, this activity is likely a side-activity of the main reaction, which is an isomerisation of methylitaconate into dimethylmaleate. To our knowledge, the Mii enzyme cannot convert cis-aconitate to ITA, and *Eubacterium* strains have not been implicated in ITA production.

[0165] With regard to nucleic acid sequences, the degeneracy of the genetic code permits substitution of certain codons by other codons that specify the same amino acid and hence would give rise to the same protein. The nucleic acid sequence can vary substantially since, with the exception of methionine and tryptophan, the known amino acids can be coded for by more than one codon. Thus, portions or all of the nucleic acid sequences described herein could be synthesized to give a nucleic acid sequence significantly different from that shown in their indicated sequence. The encoded amino acid sequence thereof would, however, be preserved.

[0166] A nucleic acid molecule according to the present invention may also include a polynucleotide sequence that is at least essentially identical to the sequence of SEQ ID NO: 3. The respective polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 3. The sequence of SEQ ID NO: 3 is believed to encode a protein that is involved in the export of methylenesuccinic acid from the cytosol of a cell

expressing the protein (MFS transporter or MFS). In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7 and the sequence of SEQ ID NO: 3. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and the sequence of SEQ ID NO: 3. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 6 and the sequence of SEQ ID NO: 3.

**[0167]** In some embodiments a nucleic acid molecule according to the present invention includes a sequence that is at least essentially identical to the sequence of SEQ ID NO: 4. Such a polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 4. The sequence of SEQ ID NO: 4 is believed to encode a protein that is involved in the export of citrate or cis-aconitate from the mitochondria to the cytosol of a cell expressing the protein (CTP1). In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and the sequence of SEQ ID NO: 4. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7 and the sequence of SEQ ID NO: 4. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 6 and the sequence of SEQ ID NO: 4.

**[0168]** In some embodiments a nucleic acid molecule according to the invention includes a functional fragment of the sequence of SEQ ID NO: 3 or of the sequence of SEQ ID NO: 4. The functional fragment may be a polynucleotide sequence that includes a sequence of about 1000 contiguous bases of SEQ ID NO: 3 and of SEQ ID NO: 4, respectively. In some embodiments such a polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 3 and of SEQ ID NO: 4, respectively.

**[0169]** A nucleic acid molecule according to the present invention may also include a sequence that is at least essentially identical to the sequence of SEQ ID NO: 5. Such a polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 5. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and the sequence of SEQ ID NO: 5. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7 and the sequence of SEQ

ID NO: 5. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 6 and the sequence of SEQ ID NO: 5.

**[0170]** A nucleic acid molecule according to the present invention may also include a sequence that is at least essentially identical to the sequence of SEQ ID NO: 6. Such a polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 6. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and the sequence of SEQ ID NO: 6. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7 and the sequence of SEQ ID NO: 6. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 6.

**[0171]** A nucleic acid molecule according to the present invention may also include a sequence that is at least essentially identical to the sequence of SEQ ID NO: 9. Such a polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 9.

**[0172]** A nucleic acid molecule according to the present invention may also include a sequence that is at least essentially identical to the sequence of SEQ ID NO: 8. Such a polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 8. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and the sequence of SEQ ID NO: 8. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7 and the sequence of SEQ ID NO: 8. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 8. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 6 and SEQ ID NO: 8. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 6 and/or SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 8.

**[0173]** A nucleic acid molecule according to the present invention may also include a sequence that is at least essentially identical to the sequence of SEQ ID NO: 1. Such a

polynucleotide sequence may have at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1.

[0174] A nucleic acid molecule according to the present invention may also include a sequence that is at least essentially identical to the sequence of SEQ ID NO: 7. Such a polynucleotide sequence may have at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 7.

[0175] Similarly, the present invention relates to a recombinant nucleic acid molecule comprising an expression cassette, the expression cassette comprising (i) a transcription regulating nucleotide sequence and (ii) a polynucleotide sequence

- having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 7 and encoding a polypeptide having ADI activity or
  - capable of hybridizing under stringent conditions to the sequence of SEQ ID NO: 1
- SEQ ID NO: 7 and

(iii) a polynucleotide sequence

- having at least 50% sequence identity with the sequence of SEQ ID NO: 6 and encoding a polypeptide having TAD activity or
- capable of hybridizing under stringent conditions to the sequence of SEQ ID NO: 6.

[0176] The present invention also relates to a recombinant nucleic acid molecule comprising an expression cassette, the expression cassette comprising (i) a transcription regulating nucleotide sequence and (ii) a polynucleotide sequence having at least 30%, 40% or 50% sequence identity with any of the sequences of SEQ ID NO: 1, 3, 4, 6-9 and encoding a protein having ADI, MFS, CTP1, TAD, transcription factor having activity for one of nucleic acids of SEQ ID NO: 1, 7, 3, 4, 6, 8 and/or 9 (e.g. UM05080) or cytochrome P450-Monooxygenase activity, respectively or capable of hybridizing under stringent conditions to the sequence any of SEQ ID NO: 1, 3, 4, 6-9, respectively.

[0177] In some embodiments the polynucleotide sequence of a nucleic acid molecule of the invention is included in a nucleic acid molecule that further contains a polynucleotide sequence that is at least essentially identical to the sequence of SEQ ID NO: 6. The respective polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 6. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 6 and the sequence of SEQ ID NO: 7. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 6 and the

sequence of SEQ ID NO: 1. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 6 and the sequence of SEQ ID NO: 4.

[0178] In some embodiments a nucleic acid molecule according to the invention includes a functional fragment of the sequence of SEQ ID NO: 5 or of the sequence of SEQ ID NO: 6. The functional fragment may be a polynucleotide sequence that includes a sequence of about 1000 contiguous bases of SEQ ID NO: 5 and of SEQ ID NO: 6, respectively. In some embodiments such a polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 5 and of SEQ ID NO: 6, respectively.

[0179] In some embodiments a nucleic acid molecule according to the present invention includes a sequence that is at least essentially identical to the sequence of SEQ ID NO: 8 and/or a sequence that is at least essentially identical to the sequence of SEQ ID NO: 9. Such a polynucleotide sequence may have at least 50% sequence identity with the sequence of SEQ ID NO: 8, and with the sequence of SEQ ID NO: 9, respectively. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and the sequence of SEQ ID NO: 8. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7 and the sequence of SEQ ID NO: 8. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7 and the sequence of SEQ ID NO: 9. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 7, the sequence of SEQ ID NO: 8 and the sequence of SEQ ID NO: 9. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and the sequence of SEQ ID NO: 9. In some embodiments a respective nucleic acid molecule is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1, the sequence of SEQ ID NO: 8 and the sequence of SEQ ID NO: 9. In some embodiments the nucleic acid molecule of the invention is a recombinant nucleic acid molecule that includes the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and/or SEQ ID NO: 6 and SEQ ID NO: 8.

[0180] In some embodiments a nucleic acid molecule according to the invention includes a functional fragment of the sequence of SEQ ID NO: 8 or of the sequence of SEQ ID NO: 9. The functional fragment may be a polynucleotide sequence that includes a sequence of about 1000 contiguous bases of SEQ ID NO: 8 and of SEQ ID NO: 9,

respectively. In some embodiments such a polynucleotide sequence may include a sequence of about 850 contiguous bases of SEQ ID NO: 8 and of SEQ ID NO: 9, respectively.

[0181] A polynucleotide sequence of the invention may in some embodiments be complementary, including at least essentially complementary, to the sequence of SEQ ID NO: 7 or to the sequence of SEQ ID NO: 1. A polynucleotide sequence of the invention may be fully complementary to one of these sequences, i.e. have a sequence of nucleotides, where in the form of this sequence each of the nucleotides can undergo specific association such as Watson-Crick base pairing with a nucleotide in the sequence of SEQ ID NO: 7 or of SEQ ID NO: 1.

[0182] A nucleic acid molecule of the invention may include a polynucleotide sequence that is complementary, including at least essentially complementary, to the sequence of SEQ ID NO: 3. In some embodiments a polynucleotide sequence of the invention is fully complementary to the sequence of SEQ ID NO: 3. In some embodiments a polynucleotide nucleic acid molecule of the invention includes a polynucleotide sequence that is complementary, including at least essentially complementary or fully complementary, to the sequence of SEQ ID NO: 4.

[0183] A polynucleotide sequence of the invention may in some embodiments be complementary, including at least essentially complementary or fully complementary, to the sequence of SEQ ID NO: 5. In some embodiments a polynucleotide nucleic acid molecule of the invention includes a polynucleotide sequence that is complementary, including at least essentially complementary or fully complementary, to the sequence of SEQ ID NO: 6. A nucleic acid molecule of the invention may also include a sequence that is complementary, including at least essentially complementary or fully complementary, to the sequence of SEQ ID NO: 5.

[0184] A polynucleotide sequence of the invention may in some embodiments be complementary, including at least essentially complementary or fully complementary, to the sequence of SEQ ID NO: 8. In some embodiments a polynucleotide nucleic acid molecule of the invention includes a polynucleotide sequence that is complementary, including at least essentially complementary or fully complementary, to the sequence of SEQ ID NO: 9.

[0185] A tag may be included into a polynucleotide sequence to allow identification and/or purification of the encoded protein. Examples of affinity tags that may be used in accordance with the invention include, but are not limited to, a streptavidin binding tag such as the STREP-TAGS® described in US patent application US 2003/0083474, US patent 5,506,121 or 6,103,493, an immunoglobulin domain, maltose-binding protein, glutathione-S-

transferase (GST), calmodulin binding peptide (CBP), FLAG-peptide (e.g. of the sequence Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys-Gly, SEQ ID NO: 24), the T7 epitope (Ala-Ser-Met-Thr-Gly-Gly-Gln-Gln-Met-Gly, SEQ ID NO:25), maltose binding protein (MBP), the HSV epitope of the sequence Gln-Pro-Glu-Leu-Ala-Pro-Glu-Asp-Pro-Glu-Asp (SEQ ID NO: 26) of herpes simplex virus glycoprotein D, the Vesicular Stomatitis Virus Glycoprotein (VSV-G) epitope of the sequence Tyr-Thr-Asp-Ile-Glu-Met-Asn-Arg-Leu-Gly-Lys (SEQ ID NO: 27), the hemagglutinin (HA) epitope of the sequence Tyr-Pro-Tyr-Asp-Val-Pro-Asp-Tyr-Ala (SEQ ID NO: 28) and the “myc” epitope of the transcription factor c-myc of the sequence Glu-Gln-Lys-Leu-Ile-Ser-Glu-Glu-Asp-Leu (SEQ ID NO: 29). Further examples of a suitable tag include, but are not limited to, a HAT tag, c-myc, His (e.g., 6xHis, SEQ ID NO: 30) tags, a TAP-tag, chitin binding domain, immunoglobulin A (IgA), intein and streptavidine binding protein (SBP) tag.

[0186] A nucleic acid molecule of the invention may further include a nucleotide sequence that encodes a marker protein. Such a marker protein may for instance confer resistance against an antibiotic or anti-metabolite.

[0187] A marker protein, in accordance with the invention, means a protein, which provides the transformed cells with a selection advantage (e.g. growth advantage, resistance against an antibiotic) by expressing the corresponding gene product. Marker genes code, for example, for enzymes causing a resistance to particular antibiotics. As used herein, the term “marker gene” refers to a gene the product of which confers a characteristic to the cell expressing the marker gene that allows it to be distinguished from cells that do not express the marker gene. In some embodiments, the marker gene allows screening and/or selection of cells. In some such embodiments, the marker gene is a “screenable marker” or a “selectable marker”. Screening and/or selection may be accomplished based on the presence or absence of the marker. In some embodiments, the screenable or selectable marker confers resistance to an agent such as an antibiotic. In some embodiments, the screenable or selectable marker confers an ability that provides an advantage in a particular set of growth conditions over cells that do not express the screenable or selectable marker.

[0188] As described above, the selectable marker can be the expression product of a gene encoding a protein restoring prototrophy for an organic compound, also referred to as a prototrophy restoring gene. In this case, the selectable marker introduced enables the cell to synthesize the compound by itself so that it is no longer or less dependent on the external supply of said compound with the medium. Accordingly, a prototrophy restoring gene is a gene encoding an expression product, i.e. the selectable marker, which reduces or preferably

abolishes the dependency of the host cell on external supply of an organic compound by facilitating its synthesis in the cell.

[0189] Selection for cells expressing said prototrophy restoring gene is carried out by culturing the cells on/in medium not containing the compound. Only cells expressing the prototrophy restoring gene will grow. The expression product of the gene may be a constituent of a synthesis pathway and the product produced by the constituent may have to be further processed in order to obtain the organic compound otherwise externally supplied. Prototrophy restoring genes commonly applied to plant or fungal cells are e.g. those expressing proteins conferring arginine prototrophy, tryptophan prototrophy, uridine prototrophy or genes enabling for nitrate or sulphate utilization. If the selectable marker is the expression product of a prototrophy restoring gene, the selecting agent is the medium in which the cell is cultivated and which does not contain the respective organic compound. Responsiveness in that case is expressed e.g. in growth rates of the cell. Thus, the higher the expression of the selectable marker, the higher the growth rate of the cell in the absence of the respective compound.

[0190] For some prototrophy restoring genes, the amount of expression product sufficient to result in prototrophy is very low. Accordingly, it is more laborious to distinguish cells expressing said prototrophy restoring selectable marker at a low level from those that express it at a high level. In order to facilitate said distinction, such a prototrophy restoring gene can be co-introduced together with a nucleic acid encoding a reporter gene the detectability of which is proportional to its expression level. Accordingly, in this embodiment, the selectable marker according to the invention is composed of the auxotrophy gene and the reporter gene.

#### Expression, Methods and Uses

[0191] A host cell used in the context of the present invention is typically capable of expressing a protein encoded by e.g. SEQ ID NO: 1 or any of SEQ ID NO: 3-9 such as a polypeptide molecule of SEQ ID NO: 2 or any of SEQ ID NO: 10-16, in that it includes a respective nucleic acid sequence, for example in the form of a functional gene of an ADI, TAD, MFS, CTP1 or a transcription factor (whether homologous or heterologous). In some embodiments, the host cell is capable of expressing a protein encoded by any combination of nucleic acids of the present invention. In other embodiments the host cell is able to express the proteins or protein combinations of the present invention. In some embodiments e.g. a nucleic acid molecule encoding a biologically active ADI has a nucleic acid sequence that

has at least 70 % sequence identity with the sequence of SEQ ID NO: 1 or the sequence of SEQ ID NO: 7. In some embodiments a respective, for instance identical, gene encoding a protein of e.g. SEQ ID NO: 1 or SEQ ID NO: 7 is functionally active and expressing a protein of e.g. SEQ ID NO: 2. In some embodiments an expression cassette, for instance heterologous, encoding an ADI protein is functionally active and expressing the respective protein. In some embodiments an endogenous nucleic acid sequence encoding a protein of e.g. SEQ ID NO: 2 is functionally inactive. In some of these embodiments a respective protein is nevertheless expressed – generally from a heterologous expression cassette. A heterologous gene or expression cassette encoding a corresponding protein may be introduced by means of recombinant technology, for instance by means of a vector carrying a gene encoding a protein of e.g. SEQ ID NO: 2 (cf. also above). It may in this regard be advantageous to further use a vector that contains a promoter effective to initiate transcription in the respective host cell (whether of endogenous or exogenous origin).

**[0192]** In case the host cell is a fungal host cell that is an *Ustilago maydis* cell, suitable marker genes encode a resistance gene against hygromycin, G418, phleomycin, nourseothricin and carboxin.

**[0193]** A nucleic acid molecule of the invention may be included in a vector such as an expression vector. When included in a vector, typically a polynucleotide sequence of the invention is included in an expression cassette. An expression cassette as used in the context of the invention is typically driven by an expression control sequence, i.e. its expression is controlled by an expression control sequence which may be a constitutively active or inducible expression control sequence (such as a promoter) that is operatively linked with the expression cassette. A respective expression cassette is further designed such that it allows the expression of the incorporated nucleic acid sequence in selected host cells. For this purpose a respective expression cassette usually has the necessary regulatory sequences, such as a promoter and/or a transcription termination sequence such as a poly A site. A suitable host cell may for instance be a fungal host cell, such as a fungal host cell that is capable of filamentous growth in liquid culture or a cell of basidiomycetales, i.e. basidiomycetes. Typically the host cell does not secrete proteases, which take action on a protein of interest as described herein. In this regard the host cell may be a cell that has been manipulated so that any such proteases are inactivated, e.g., by knockout or pull-down by, e.g. iRNA or siRNA. As an illustrative example, where a fungal host cell or a yeast host cell is used, the cell may have an inactivated Kex2 protease, i.e. a Kex2-negative fungal or yeast host cell may be used. In some embodiments the fungi host cell is a cell of the class

Saccharomycetes. Preferred are e.g. *Saccharomyces cerevisiae*, and *Pichia pastoris*. In other embodiments the host cell is a cell of the class *Ustilaginomycetes* or *Eurotiomycetes*, e.g. *Ustilago maydis* or *Aspergillus terreus*. In some embodiments where the host cell is *Ustilago maydis*, the protease that may be inactivated is Kex 2, encoded by the gene um02843 of SwissProt/ Uniprot accession No Q4PAM0 (version 55 as of 03 October 2012). The skilled person is aware of means and methods for inactivating any such protease. In case of *Ustilago maydis*, Kex2 can, for example, be knocked-out, either fully or partially, e.g., by homologous recombination. Other proteases that may be inactivated, either additionally or alternatively to Kex2, in *Ustilago maydis* are a secreted aspartic protease Um04926, designated Pep4; a lysosomal serine protease Um04400, designated Prb1 and/or a lysosomal tripeptidyl peptidase Um06118 of SwissProt/ Uniprot accession No Q4P195 (version 36 as of 05 September 2012).

[0194] Any desired restriction endonuclease site may be incorporated into an expression cassette and/or vector according to the invention. Typically the expression cassette includes at least one restriction enzyme recognition site at about the 3'-end and at least one restriction enzyme recognition site at about the 5'-end. As an example, a sequence cassette may be created enzymatically (e.g., by using a type I or type II restriction endonuclease or exonuclease), by mechanical means such as shearing, by chemical synthesis, or by a recombinant method such as PCR. An expression cassette usually includes the following elements (presented in the 5'-3' direction of transcription): a transcriptional and translational initiation region, a coding sequence for a gene of interest, and a transcriptional and translational termination region functional in the organism where it is desired to express the gene of interest.

[0195] Any desired protein expression system may be employed to generate a polypeptide according to the invention. In some embodiments a bacterial, a fungal or a mammalian cell expression system may be used. In one embodiment the cell is a cell of a fungus of the order *Ustilaginales*. As explained above, a host cell may in some embodiments be a fungus or a yeast that includes a nucleic acid molecule as described above, including an expression cassette or a vector described herein. A large number of suitable methods exist in the art to produce polypeptides (or fusion proteins) in host cells. Conveniently, the produced protein is harvested from the culture medium (medium), lysates of the cultured host cell or from isolated (biological) membranes by established techniques. For example, the expression cassettes as described herein comprising, inter alia, the nucleotide sequence encoding the protein of interest can be synthesized by PCR and inserted into an expression vector.

Subsequently a cell produced with the method of the present invention may be transformed with the expression vector. Thereafter, the cell is cultured to produce/express the desired protein(s), which is/are isolated and purified. For example, the product may be recovered from the host cell and/or culture medium by conventional procedures including, but not limited to, cell lysis, breaking up host cells, centrifugation, filtration, ultra-filtration, extraction or precipitation. Purification may be performed by a variety of procedures known in the art including, but not limited to, chromatography (e.g. ion exchange, affinity, hydrophobic, chromatofocussing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing), differential solubility (e.g. ammonium sulphate precipitation) or extraction.

**[0196]** In embodiments where a nucleic acid sequence of the invention is included in a eukaryotic expression cassette the expression cassette usually includes a polyadenylation site. Most eukaryotic nascent mRNAs possess a polyA-tail at their 3' end which is added during a complex process that involves cleavage of the primary transcript and a coupled polyadenylation reaction. The polyA-tail is advantageous for mRNA stability and transferability.

**[0197]** When a polypeptide of interest or of the present invention is expressed in a selected host cell, it may be necessary to modify the nucleotide sequence encoding the polypeptide by adapting the codon usage of the nucleotide sequence to meet the frequency of the preferred codon usage of the respective host cell. In this regard, "frequency of preferred codon usage" refers to the preference exhibited by the host cell of the invention in usage of nucleotide codons to specify a given amino acid. To determine the frequency of usage of a particular codon in a gene, the number of occurrences of that codon in the gene is divided by the total number of occurrences of all codons specifying the same amino acid in the gene. Similarly, the frequency of preferred codon usage exhibited by a host cell can be calculated by averaging frequency of preferred codon usage in a large number of genes expressed by the host cell. It is preferable that this analysis be limited to genes that are highly expressed by the host cell. The percent deviation of the frequency of preferred codon usage for a synthetic gene from that employed by a host cell is calculated first by determining the percent deviation of the frequency of usage of a single codon from that of the host cell followed by obtaining the average deviation over all codons. As defined herein, this calculation includes unique codons (i.e., ATG and TGG). In general terms, the overall average deviation of the codon usage of an optimized gene from that of a host cell is calculated using the equation  $1A = n-1 \sum Z X_n - Y_n X_n \text{ times } 100 Z$  where  $X_n$  = frequency of usage

for codon  $n$  in the host cell;  $Y_n$ =frequency of usage for codon  $n$  in the synthetic gene;  $n$  represents an individual codon that specifies an amino acid; and the total number of codons is  $Z$ . The overall deviation of the frequency of codon usage,  $A$ , for all amino acids should preferably be less than about 25%, such as less than about 10%.

5       **[0198]**       Allowing a host cell to express a protein encoded by a sequence as defined above or a protein of the present invention may include fermenting a medium that includes a carbon source such as a carbohydrate-based carbon source, e.g. starch, a molasse such as sugar cane molasse, a hydrolysate of e.g. corn syrup or wood, (cf. also above) with a transformed cell as defined above. The cell may be allowed to ferment the carbon source to  
10       itaconic acid. An example of a suitable fermentation process is an aerobic fermentation process. An aerobic fermentation process according to an embodiment of the invention may be run under aerobic oxygen-limited conditions. As an example, in an aerobic process under oxygen-limited conditions, the rate of oxygen consumption is at least 5.5, including at least 6 or at least 7 mmol/L/h. The fermentation process may be a submerged or a solid state  
15       fermentation process. Itaconic acid may be produced via submerged fermentation starting from a carbohydrate raw material such as for instance cassava and/or corn, which may be milled and mixed with water. A seed fermentation may be prepared in a separate fermenter. The liquefaction of the starch may be performed in the presence of an amylolytic enzyme such as for instance an amylase, a cellulase, a lactase or a maltase and one or more additives  
20       and nutrients such as antifoam may be added before or during fermentation.

**[0199]**       Culturing a cell that includes a nucleic acid molecule having a polynucleotide sequence that is at least essentially identical to a polynucleotide sequence as described above, e.g. the sequence of SEQ ID NO: 1 or SEQ ID NO: 7 and/or SEQ ID NO: 6, may be carried out in any suitable device such as a conventional stirred tank reactor, a bubble column, a  
25       tubular reactor, or an air-lift reactor.

**[0200]**       Once expressed by a host cell, the itaconic acid may be recovered. It may be isolated by lysis or any other measure applied to break the cell membrane and/or cell wall of a host cell and subsequently be purified. However, a convenient way to harvest a protein of interest is its isolation from the culture medium. In that case, the protein of interest must be  
30       secreted by the host. Secretion of a protein is usually achieved by the use of signal sequences. Specifically, proteins equipped with a signal sequence are secreted through the conventional endoplasmic reticulum (ER)-Golgi secretory pathway, i.e., the conventional secretion pathway. However, some proteins, for example, cytoplasmic, nuclear and signal-peptide-containing proteins have been shown to reach the cell surface by non-conventional transport

pathways. The mechanisms and molecular components of unconventional protein secretion are beginning to emerge. Unconventional protein secretion may have some advantages vis-à-vis conventional protein secretion, since proteins subject to unconventional secretion are not processed by ER or Golgi-dependent post-translational modifications.

5       **[0201]**     In a method or use of the invention a nucleic acid may be introduced into a host cell by any suitable technique of nucleic acid delivery for transformation of a cell available in the art. Examples of suitable techniques include, but are not limited to, direct delivery of DNA, e.g. via transfection, injection, including microinjection, electroporation, calcium phosphate precipitation, by using DEAE-dextran followed by polyethylene glycol, 10 direct sonic loading, liposome mediated transfection, receptor-mediated transfection, microprojectile bombardment, agitation with silicon carbide fibers, Agrobacterium-mediated transformation, desiccation/ inhibition-mediated DNA uptake or any combination thereof. As indicated above, any desired host cell may be selected, for example a cell of a fungus of the order *Ustilaginales*.

15       **[0202]**     A polypeptide of the invention may consist of the twenty conventional amino acids. In some embodiments a polypeptide of the invention may include stereoisomers (e.g., D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as  $\alpha,\alpha$ -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for a polypeptide according to the present invention. Examples 20 of unconventional amino acids include: 4-hydroxyproline,  $\gamma$ -carboxyglutamate, N,N,N-trimethyllysine, E-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids such as 4-hydroxyproline. In the polypeptide notation used in this document, the left-hand direction is the amino terminal direction and the right hand direction is the 25 carboxy-terminal direction, in accordance with standard usage and convention.

30       **[0203]**     The present invention provides a method of producing itaconic acid, the method comprising culturing a cell that comprises polynucleotide sequences having at least 50% sequence identity with the sequence of SEQ ID NO: 1 and/or SEQ ID NO: 7 and SEQ ID NO: 6 allowing the host cell to express or overexpress the polynucleotide sequences, forming a polypeptides that have activity of an ADI and a TAD, thereby producing itaconic acid. In some embodiments, the method further comprises a cell further comprising a polynucleotide sequence of SEQ ID NO: 4. In some embodiments, the method further comprises a cell further comprising a polynucleotide sequence of SEQ ID NO: 3. In some embodiments, the method further comprises a cell further comprising a polynucleotide

sequence of SEQ ID NO: 8. In some embodiments, the method includes a cell overexpressing polynucleotides of any of the SEQ ID NO: 1, 3-9 in comparison to a wildtype cell. In some embodiments, the method includes a cell with a decreased expression or deletion of the polynucleotide sequence of SEQ ID NO: 9 in comparison to a wildtype cell.

**[0204]** The recombinant host cell of the present invention is engineered to overexpress

- (i) a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity; and
- (ii) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity.

**[0205]** The recombinant host cell of the present invention can also be engineered to overexpress

- (i) a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 7 or SEQ ID NO: 17, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity; and
- (ii) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity.

**[0206]** The recombinant host cell of the present invention also includes a recombinant host cell that is engineered to overexpress a polynucleotide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity.

**[0207]** The recombinant host cell of the present invention also includes a recombinant host cell that is engineered to overexpress a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity.

[0208] The recombinant host cell of the present invention also includes a recombinant host cell that is engineered to overexpress polynucleotide sequences having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 71 or SEQ ID NO: 17, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity.

[0209] The recombinant host cell of the present invention also includes a recombinant host cell that is engineered to overexpress polynucleotide sequences having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said polynucleotide sequence encoding a protein or fragment thereof having MFS activity.

[0210] The recombinant host cell of the present invention also includes a recombinant host cell that is engineered to overexpress polynucleotide sequences having at least 50% sequence identity with the sequence of SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having CTP1 activity.

[0211] The recombinant host cell of the present invention also includes a recombinant host cell that is engineered to overexpress polynucleotide sequences having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or fragment thereof having transcription factor activity. Preferably, the transcription factor has activity for one of nucleic acids of SEQ ID NO: 1, 7, 3, 4, 6, 8 and/or 9, as for example, UM05080.

[0212] The recombinant host cell of the present invention also includes a recombinant host cell that is engineered to under-express polynucleotide sequences having at least 50% sequence identity with the sequence of SEQ ID NO: 9 said polynucleotide sequence encoding a protein or fragment thereof having P450-Monooxygenase (UM05074) activity.

[0213] The recombinant host cell of the present invention can be engineered to overexpress

(i) a polypeptide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2 and/or SEQ ID NO: 10 and/or SEQ ID NO: 18, wherein the polypeptide has ADI activity; and

(ii) a polypeptide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 13, wherein the polypeptide has TAD activity.

[0214] The recombinant host cell of the present invention can be engineered to overexpress a polypeptide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 2 or SEQ ID NO: 18, wherein the polypeptide has ADI activity.

[0215] The recombinant host cell of the present invention can be engineered to overexpress a polypeptide sequence having at least 30%, 40%, or 50% sequence identity with the sequence of SEQ ID NO: 10 or SEQ ID NO: 18, wherein the polypeptide has ADI activity.

[0216] The recombinant host cell of the present invention can be engineered to overexpress a polypeptide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 13, wherein the polypeptide has TAD activity.

[0217] The recombinant host cell of the present invention can be engineered to underexpress a polypeptide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 11, wherein the polypeptide has P450-Monooxygenase (UM05074) activity.

[0218] The recombinant host cell of the present invention can be engineered to overexpress a polypeptide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 14, wherein the polypeptide has MFS activity.

[0219] The recombinant host cell of the present invention can be engineered to overexpress a polypeptide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 15, wherein the polypeptide has CTP1 activity.

[0220] The recombinant host cell of the present invention can be engineered to overexpress a polypeptide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 11, wherein the polypeptide has transcription factor (UM05080) activity.

[0221] In a method or use of generating itaconic acid (methylenesuccinic acid) according to the invention a recombinant nucleic acid molecule as defined above is employed, for example including an expression cassette. As an example and as explained

above, in some embodiments the recombinant nucleic acid molecule includes a polynucleotide sequence, e.g. a heterologous polynucleotide sequence, that has at least 30%, 40%, or 50 % identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17 or at least 50% identity with the sequence of any of SEQ ID NO: 3-9. In some embodiments the recombinant nucleic acid molecule includes a polynucleotide sequence that differs from the sequence of SEQ ID NO: 1 or SEQ ID NO: 17 and/or from any of the sequence of SEQ ID NO: 3-9 in at least one nucleotide position. In some embodiments the recombinant nucleic acid molecule includes a polynucleotide sequence that has at least 70 % identity with any of the sequence of SEQ ID NO: 1, 3-9, 17. The recombinant nucleic acid molecule may also include a polynucleotide sequence that has at least 80 % identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17 or any of sequence of SEQ ID NO: 3-9. The recombinant nucleic acid molecule may also include a polynucleotide sequence that has at least 85 % identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17 or any of sequence of SEQ ID NO: 3-9. The recombinant nucleic acid molecule may also include a polynucleotide sequence that has at least 90 % identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17 or any of sequence of SEQ ID NO: 3-9. The recombinant nucleic acid molecule may also include a polynucleotide sequence that has at least 95 % identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17 or any of sequence of SEQ ID NO: 3-9. The recombinant nucleic acid molecule may also include a polynucleotide sequence that has at least 99 % identity with the sequence of SEQ ID NO: 1 or SEQ ID NO: 17 or any of sequence of SEQ ID NO: 3-9.

**[0222]** The recombinant nucleic acid molecule may include a further polynucleotide sequence as described above. In some embodiments a plurality such as two, three or more recombinant nucleic acid molecules as defined above are used. Each recombinant nucleic acid molecule of such a plurality may include one or more polynucleotide sequences as described above. For example, a sequence included in a nucleic acid molecule used may have at least 70 % identity with the sequence of SEQ ID NO: 3 or with the sequence of SEQ ID NO: 4. In one embodiment a recombinant nucleic acid molecule is used that includes a sequence that is at least essentially identical to the sequence of SEQ ID NO: 5 or to the sequence of SEQ ID NO: 6. As a further example a sequence included in a recombinant nucleic acid molecule used may have at least 70 % identity with the sequence of SEQ ID NO: 8 or with the sequence of SEQ ID NO: 9. In some embodiments the recombinant nucleic acid molecule includes a polynucleotide sequence that differs from the sequence of SEQ ID NO: 3, the sequence of SEQ ID NO: 4, the sequence of SEQ ID NO: 5, the sequence

of SEQ ID NO: 6 and/or from the sequence of SEQ ID NO: 7 in at least one nucleotide position. In some embodiments the recombinant nucleic acid molecule includes a polynucleotide sequence that differs from the sequence of SEQ ID NO: 8 and/or from the sequence of SEQ ID NO: 9 in at least one nucleotide position.

5       **[0223]**     The nucleic acid molecule may be included in a host cell as described above. Where a host cell contains the nucleic acid molecule, the host cell is cultured. The respective conditions for culturing the host cell depend on the selected type of cell and are within the knowledge of the skilled person. The host cell is allowed to express the polynucleotide sequence, so that a polypeptide is forming that is encoded by the sequence included in the  
10 recombinant nucleic acid molecule. The polypeptide may for instance have activity of a CAD. Typically the methylenesuccinic acid formed is isolated from the medium in which the cell is cultured (supra).

**[0224]**     A method or use of producing itaconic acid may be a method of producing nitrilon. A method or use of producing itaconic acid may also be a method of manufacturing  
15 paper. In some embodiments a method or use of producing itaconic acid is a method of biofuel production or a method of wastewater treatment. In a method or use of producing itaconic acid typically the recombinant nucleic acid molecule is typically included in a host cell (supra). In typical embodiments the cell is provided with a suitable carbon source such as a monosaccharide, a polysaccharide, a polyol such as sorbitol, xylitol or mannitol, a  
20 lipid such as triacylglycerol or a fatty acid. Examples of a suitable monosaccharide include, but are not limited to, mannose, glucose, arabinose and xylose. Examples of a suitable polysaccharide include, but are not limited to, starch, a mannan and cellulose. Illustrative examples of a suitable fatty acid are omega-3 and/or omega-6 polyunsaturated fatty acids including decosahexaenoic acid or eicosapentaenoic acid, as well as palmitoleic acid, oleic  
25 acid or citric acid.

### Kit

**[0225]**     Reagents needed or useful in the context of the present invention may be provided in the form of a kit. Such a kit may in particular include means for expression of one or more genes as described in the present document. Means for expressing a biomarker  
30 are known in the art, and may include, for example, the use of suitable host cell and a nucleic acid molecule, for example included in an expression vector, carrying a sequence of a gene described above. In some embodiments a respective nucleic acid molecule according to the invention has a nucleic acid sequence that has at least 90 % sequence identity with the sequence of SEQ ID NO: 3. A nucleic acid molecule according to the invention may also

have a nucleic acid sequence that has at least 90 % sequence identity with the sequence of SEQ ID NO: 4 or the sequence of SEQ ID NO: 5. The kit may include a first container that has a host cell such as a fungal host cell as described above. The kit may include a second container that has a nucleic acid molecule, such as a nucleic acid molecule described above.

5 A nucleic acid molecule according to the invention may also have a nucleic acid sequence that has at least 90 % sequence identity with the sequence of SEQ ID NO: 6 or the sequence of SEQ ID NO: 7. Typically the second container includes a nucleic acid molecule with a sequence that is at least essentially identical to the sequence of SEQ ID NO: 1 or the sequence of SEQ ID NO: 7 or the sequence of SEQ ID NO: 17. One or more further  
10 containers may be included in the kit that contain(s) a nucleic acid molecule with a sequence that is at least essentially identical to the sequence of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 17 or the sequence of SEQ ID NO: 9. In some embodiments a nucleic acid molecule according to the invention has a nucleic acid sequence that has at least 90 % sequence identity with the  
15 sequence of SEQ ID NO: 8 or with the sequence of SEQ ID NO: 9.

[0226] In some embodiments the kit may also include a reagent that allows the detection of a detectable label, which may be expressed in the form of a tag, attached to/fused to a protein encoded by a nucleic acid molecule of the invention. As an illustrative example, the detectable label may be an enzyme and the reagent may be a substrate of the  
20 enzyme. The substrate may for instance be converted by such enzyme into a product that emits a signal such as a fluorescent or a colour signal. In some embodiments the kit may include a binding partner directed to a protein encoded by a sequence as defined above. A respective binding partner may be an antibody, such as an immunoglobulin, a fragment of an immunoglobulin or a proteinaceous binding molecule with immunoglobulin-like functions.  
25 In one embodiment the kit includes components for setting up a method of detecting the expression of a nucleic acid sequence as defined above.

[0227] The kit may further include instructions and/or imprint indicating on how to express a nucleic acid molecule contained in the kit, as well as – where applicable – how to detect the expressed product. The kit may also include positive and/or negative controls,  
30 which allow a comparison to the control.

[0228] The listing or discussion of a previously published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

[0229] The invention illustratively described herein may suitably be practiced in the

absence of any element or elements, limitation or limitations, not specifically disclosed herein. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by exemplary embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

[0230] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0231] Other embodiments are within the appending claims. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0232] The invention is further illustrated by the following non-limiting examples. As one of ordinary skill in the art will readily appreciate from the disclosure of the present invention, other compositions of matter, means, uses, methods, or steps, presently existing or later to be developed that perform substantially the same function or achieve substantially the same result as the corresponding exemplary embodiments described herein may likewise be utilized according to the present invention.

## Examples

### Generation of UM05076 and UM11778 deletion mutants in *Ustilago maydis*

[0233] The deletion constructs were constructed by using the yeast drag and drop method [Jansen et al. (2005) Gene 344:43-51]. For the deletion of UM05076, the shuttle vector pRS426 was combined with the left flank (5'UTR) and the right flank (3'UTR) of UM05076, respectively, as well as a hygromycin cassette derived from pMF1-h [Brachmann, A, et al. (2004) Mol Gen Genomics 272: 216-226]. For the deletion of

UM11778, the shuttle vector pRS426 was combined with the left flank (5'UTR) and the right flank (3'UTR) of UM11778, respectively, as well as a hygromycin cassette derived from pMF1-h. Plasmid maps are shown in Figure 5 and Figure 6.

[0234] Vector pRS426-UM05076-Hyg was restricted with *SspI* to obtain the deletion-construct. Afterwards *U. maydis* strain MB215 was transformed with this construct, leading to genomic integration via homologous recombination. Disruption of UM05076 was checked by Southern blot analysis using *HindIII* as restriction enzyme for the genomic DNA and the right flank of UM05076 as a probe for radioactive labeling.

[0235] The resulting mutants were assessed for itaconate production in System Duetz® 24 well plates (Duetz et al. Applied and Environmental Microbiology, 2000, 66(6):2641–2646) with a filling volume of 1.5 mL (d = 50 mm, n = 300 rpm, T = 30°C and  $\Phi$  = 80 %) or in 500 ml Erlenmeyer flasks with a filling volume of 50 ml (d = 25 mm, n = 250 rpm, T = 30°C and  $\Phi$  = 80 %). The screening medium contained 50 g L<sup>-1</sup> glucose, 0.8 g L<sup>-1</sup> NH<sub>4</sub>Cl, 0.2 g L<sup>-1</sup> MgSO<sub>4</sub>•7H<sub>2</sub>O, 0.01 g L<sup>-1</sup> FeSO<sub>4</sub>•7H<sub>2</sub>O, 0.5 g L<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub>, 1 mL L<sup>-1</sup> vitamin solution, 10 mL L<sup>-1</sup> trace element solution and as buffer 19.5 g L<sup>-1</sup> 2-(N-morpholino)ethanesulfonic acid (MES). The pH of the MES stock solution was adjusted to 6.5 with NaOH. The vitamin solution contained (per liter) 0.05 g D-biotin, 1 g D-calcium pantothenate, 1 g nicotinic acid, 25 g myo-inositol, 1 g thiamine hydrochloride, 1 g pyridoxol hydrochloride and 0.2 g para-aminobenzoic acid. The trace element solution contained (per liter) 1.5 g EDTA, 0.45 g ZnSO<sub>4</sub>•7H<sub>2</sub>O, 0.10 g MnCl<sub>2</sub>•4H<sub>2</sub>O, 0.03 g CoCl<sub>2</sub>•6H<sub>2</sub>O, 0.03 g CuSO<sub>4</sub>•5H<sub>2</sub>O, 0.04 g Na<sub>2</sub>MoO<sub>4</sub>•2H<sub>2</sub>O, 0.45 g CaCl<sub>2</sub>•2H<sub>2</sub>O, 0.3 g FeSO<sub>4</sub>•7H<sub>2</sub>O, 0.10 g H<sub>3</sub>BO<sub>3</sub> and 0.01 g KI. Itaconate in the supernatant was analyzed in a Beckman Coulter System Gold High Performance Liquid Chromatography (Beckman Coulter GmbH, Germany) with an Organic Acid Resin 300 × 8 mm column (CS-Chromatography, Germany) and a differential refractometer LCD 201 (MELZ, Germany). As solvent, 5 mM H<sub>2</sub>SO<sub>4</sub>, with a flow rate of 0.6 mL min<sup>-1</sup> and a temperature of 30°C, was used. All samples were filtered with Rotilabo® syringe filters (CA, 0.20 µm, Ø 15 mm) and afterwards 1:5 diluted with 5 mM H<sub>2</sub>SO<sub>4</sub>.

[0236] A comparison of ITA production between *U. maydis* wildtype and the deletion mutants ΔUM05076, ΔUM11778, ΔUM05074, ΔUM12299, ΔUM11777, ΔUM05079 or ΔUM05080 is shown in Figure 2 and Figure 3 A. A strong decrease of the itaconic acid titer was observed in ΔUM05076, ΔUM11778, ΔUM11777, ΔUM05079 and ΔUM05080 mutant strains if compared to the wild type strain. These data indicate that these genes encode enzymes important for itaconic acid production in *U. maydis*.

### Overexpression of itaconic acid biosynthesis genes in *Ustilago maydis*

[0237] The ORFs for genes UM05076, UM11778, UM05074, UM12299, UM11777, UM05079 and UM05080 were amplified via PCR, restricted with suitable enzymes and ligated into vector Potef-GFP (Spellig et al., 1996). The final vectors contain an ampicillin resistance marker, a carboxin resistance marker, the Potef promoter and the corresponding ORFs. The corresponding plasmid map for UM05076 is shown in Figure 8. The corresponding plasmid map for UM11778 is shown in Figure 9. Constructs were tested by restriction analysis and confirmed by DNA sequencing.

[0238] Prior to transformation, the vector was linearized using *SspI*. Approximately 1 µg DNA were used for transformation of *U. maydis* MB215 strains, following integration of the construct into the *ip*-locus (Cbx) via homologous recombination. Transformants were checked by PCR analysis. *U. maydis* MB215 was transformed with overexpression constructs for either UM05076, UM11778, UM05074, UM12299, UM11777, UM05079 or UM05080.

[0239] The resulting overexpression mutants were assessed for itaconic acid production as described above. A comparison of itaconic acid production between *U. maydis* wildtype and the overexpression mutants is shown in Figure 3 B. A strong increase of the itaconic acid titre was observed in strains containing UM05079 and UM05080 overexpression constructs if compared to the wild type strain. These data indicate that these genes encode enzymes that are important for itaconic acid production in *U. maydis*.

### In vivo reconstitution of itaconic acid production in yeast

[0240] We were able to reconstitute itaconate biosynthesis in the yeast *Saccharomyces cerevisiae* by expression of um11778 and um05076. um11778 and um05076 were expressed in yeast either singly or in combination. Plasmid maps of yeast expression constructs are shown in Figure 11 and Figure 12. The *S. cerevisiae* strains were cultivated in 50 mL YNB+gal media (1.7 g L<sup>-1</sup> yeast nitrogen base without amino acids (Sigma-Aldrich, Germany), 5.0 g L<sup>-1</sup> NH<sub>4</sub>Cl, 40 g L<sup>-1</sup> galactose, 1.4 g L<sup>-1</sup> yeast synthetic drop out medium supplements without histidine, leucine, tryptophan and uracil (Sigma-Aldrich, Germany), 0.075 g L<sup>-1</sup> tryptophan and 0.5 g L<sup>-1</sup> leucine) in 500 mL Erlenmeyer flasks. According to the yeast selection markers 0.15 g L<sup>-1</sup> uracil and/or 0.5 g L<sup>-1</sup> histidine were added. Cells were grown at 30°C and 250 rpm. The supernatant of the culture was analyzed for itaconic acid production via HPLC as described above. Single expression of either gene did not result in

detectable itaconic acid production (Figure 10). However, when both genes were coexpressed, we detected significant levels of itaconic acid production (Figure 10). This indicates that both enzymes are sufficient to produce itaconic acid in a heterologous host system.

#### Characterization of enzyme activities of Um05076 and Um11778

**[0237]** We performed *in vitro* enzyme assays for both UM05076 and UM11778. Protein expression of UM05076 and UM11778 was performed in strain Rosetta 2(DE3) (Novagen, Madison, WI) using plasmids derived from pGEX4T-1 (GE Healthcare, Waukesha, WI). Cells were grown overnight in inducing dYT medium at 22°C. Cells were harvested by centrifugation, resuspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT, 1 mM PMSF) and lysed by French cell pressure. After ultra-centrifugation, the supernatant was loaded on GSH Sepharose beads (GE Healthcare) and incubated at 4°C for 1 h. Beads were washed five times with lysis buffer and then eluted in elution buffer (100 mM Tris-HCl, pH 9.0, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT and 50 mM reduced glutathione). Protein concentration was determined either by Bradford or by Nanodrop.

**[0238]** As potential substrates we used cis-aconitate, trans-aconitate, citrate and isocitrate. 10 µg of the corresponding protein was mixed with assay-buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM MgCl<sub>2</sub>, 1 mM DTT and 5mM substrate) to a final volume of 100 µl in a 96-well microtiterplate (Greiner Bio-One). The samples were incubated shaking for 15 min at RT. Reaction products were analyzed in an Agilent Technologies 1260 Series system with an organic acid resin (250 x 8 mm; CS Chromatographie). For detection at 210 nm a variable wavelength detector (1260 VWD; Agilent Technologies) was used. The solvent was 0.25 mM sulfuric acid with a flow rate of 1,0 mL min<sup>-1</sup> at RT.

HPLC analysis clearly demonstrated that UM11778 is able to catalyze cis-trans-isomerization of aconitate (Figure 13A). Therefore, we called this enzyme aconitate-delta-isomerase (ADI). UM05076 was able to convert trans-aconitate nearly completely into itaconic acid (Figure 13B). We called this second enzyme trans-aconitate decarboxylase (TAD). To our knowledge, such enzyme activity has never been described before. In a further experiment we incubated both ADI and TAD with cis-aconitate as substrate to mimic the putative biosynthesis pathway for itaconic acid. HPLC analysis demonstrated that cis-

aconitate was completely converted to itaconic acid by the sequential activity of the two enzymes (Fig. 14).

[0239] These data allowed us to propose a potential biosynthesis pathway of itaconic acid production in *U. maydis* (Fig. 15). We assume that cis-aconitate is exported from the mitochondria into the cytosol by the mitochondrial transporter CTP1. In the cytosol cis-aconitate is the substrate of ADI, which catalyzes isomerization to trans-aconitate. TAD uses trans-aconitate as a substrate for decarboxylation to itaconic acid. Finally, itaconic acid is exported by a transporter of the major facilitator superfamily (MFS). According to this model the itaconic acid biosynthesis route in *U. maydis* deviates from that in *A. terreus*. In *A. terreus* cis-aconitate decarboxylase (CAD) is solely responsible for itaconic acid biosynthesis by catalyzing a decarboxylation of cis-aconitate. In *U. maydis*, two enzymes are required for itaconic acid production, starting also with cis-aconitate as a substrate and ending with itaconic acid as a product.

#### Characterization of the ADI2 protein encoded by *um02807*

[0240] *U. maydis* possesses besides *um11778* (*adi1*) another gene encoding a putative aconitate- $\Delta$ -isomerase (ADI) protein, *um02807*. Deletion of this gene did not influence itaconate biosynthesis (Figure 16B). However, *um02807* is located in a small cluster, surrounded by a probable CTP1-mitochondrial citrate transporter, member of the mitochondrial carrier family (MCF), encoded by *um02806* and a putative transcriptional regulator related to the nitrogen assimilation transcription factor nit-4 encoded by *um02808* (Figure 16B). These three genes showed functional similarities to the genes in the itaconate cluster of *U. maydis* and are arranged in the same way. BLAST analysis on protein level 200 indicated 37 % sequence identity for *um02806* and *um05079* (*mtt1*) and 32 % for *um02807* and *um11778* (*adi1*) (Figure 16A and Figure 17). However, the putative transcriptional regulator *um02808* did not show any significant sequence similarity to *um05080* (*ria1*).

[0241] The deletion of the gene *um11778* (*adi1*) encoding the aconitate- $\Delta$ -isomerase completely abolished itaconate biosynthesis of *U. maydis*. Since both proteins (UM11778 and UM02807) are aconitate- $\Delta$ -isomerase proteins, UM02807 could possibly substitute the function of UM11778. To prove this assumption, the deletion mutant for *um11778* (ADI) was complemented with the gene *um02807* under the strong constitutive promoter  $P_{etef}$  (SEQ ID NO: 31). Interestingly, the itaconate concentration of this complementation mutant *U. maydis* MB215  $\Delta um11778 + P_{etef} um02807$  after 96 h was approximately threefold higher ( $p = 4.9 \times 10^{-4}$ ) compared to wildtype indicating that UM02807 has a similar activity to

UM11778 (ADI) (Figure 16B). Also the single overexpression of *um02807* with the help of the strong constitutive promoter  $P_{\text{etef}}$  could increase the itaconate concentration after 96 h to approximately 150 % ( $p=1.6 \times 10^{-3}$ ) compared to the wildtype. The same was done for the gene *um02808* encoding a putative transcriptional regulator. However, the gene could  
5 neither complement the function of the transcriptional regulator of the known itaconate cluster UM05080 (RiaA) nor influenced its overexpression the itaconate biosynthesis positively ( $p=0.111$ ) (Figure 16B).

[0242] These results show that the protein encoded by UM02807, also annotated as putative methylitaconate  $\Delta$ -isomerase, is probably an additional  
10 aconitate- $\Delta$ -isomerase and can take over the function of the aconitate- $\Delta$ -isomerase UM11778 (ADI). However, it was shown, that in the *um11778 (adi1)* deletion mutant no itaconate is formed, indicating that *um02807* is not expressed under the itaconate producing conditions used in this study. Thus, *um02807* is not able to compensate the deletion of *um11778 (adi1)* under control of its native regulator, supporting the  
15 theory of independently regulated clusters. Furthermore, this led to the assumptions that also other, redundant genes might also be involved in itaconate biosynthesis of *U. maydis* in other conditions, as for instance *in planta*.

[0243] The listing or discussion of a previously published document in this  
20 specification should not be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

[0244] The inventions illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including," "containing", etc.  
25 shall be read expansively and without limitation. Thus, the word "comprise" or variations such as "comprises" or "comprising" will be understood to imply the inclusion of a stated integer or groups of integers but not the exclusion of any other integer or group of integers. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and  
30 expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by exemplary embodiments and optional features, modification

and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

5       **[0245]**       The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

10       **[0246]**       Other embodiments are within the following claims. In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

## CLAIMS

1. A method of producing itaconic acid, comprising
  - (a) culturing a recombinant host cell which is genetically engineered to overexpress
    - (i) a polynucleotide sequence having at least 30% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity; and
    - (ii) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity,in comparison to a comparable host cell which has not been genetically altered under conditions to allow said host cell to overexpress said polynucleotide sequence (i) and (ii), thereby said host cell is capable of converting cis-aconitate via trans-aconitate to itaconic acid; and
  - (b) obtaining itaconic acid.
2. The method of claim 1, wherein said host cell further overexpresses a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a mitochondrial citrate transporter, preferably of a mitochondrial cis-aconitate transporter.
3. The method of claim 1 or 2, wherein said host cell further overexpresses a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a multidrug transporter of the major facilitator superfamily, preferably of a transporter which exports itaconic acid out of the host cell.
4. The method of any one of the preceding claims, wherein said host cell further overexpresses a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or

fragment thereof having the activity of a transcription factor for at least one of polynucleotide sequences of any one of SEQ ID NOs: 1, 3, 4, and/or 6.

5. The method of any one of the preceding claims, wherein at least one of the polynucleotide sequences of any of the SEQ ID NO: 1, 3, 4, 6, and/or 8 is overexpressed in comparison to a host cell prior to genetically engineering.
6. The method of any one of the preceding claims, wherein overexpression is achieved by using a recombinant promoter, which drives expression of said polynucleotide(s) in said host cell.
7. The method of any one of the preceding claims, wherein overexpression is achieved by expressing 2, 3, 4 or more copies of said polynucleotide(s) in said host cell.
8. The method of any one of the preceding claims, wherein said polynucleotide is integrated in the genome of said host cell.
9. The method cell of claim 8, wherein the integration is ectopically or in the natural locus.
10. The method of any one of the preceding claims, wherein the overexpression is achieved by using an enhancer to express the polynucleotide.
11. The method of any one of the preceding claims, wherein said polynucleotide, when overexpressed in said host cell results in production of itaconic acid of 2 g/L or more after 48 h culture in comparison to a host cell prior to genetically engineering.
12. The method of any one of the preceding claims, wherein said host cell is genetically engineered to underexpress a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 9 in comparison to a comparable host cell which has not been genetically altered, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a cytochrome P450-Monooxygenase.
13. The method of claim 12, wherein said polynucleotide, when underexpressed in said host cell results in production of itaconic acid of 2 g/L or more after 48 h culture in comparison to a host cell prior to engineering.

14. The method of any one of the preceding claims, wherein said polynucleotide sequence is a heterologous polynucleotide sequence.
15. The method of any one of the preceding claims, wherein said host cell is a cell of a fungus or yeast.
16. The method of claim 15, wherein said cell of a fungus is a cell from *Ustilago maydis*, *Aspergillus terreus* or *Saccharomyces cerevisiae*.
17. The method of any one of the preceding claims, further comprising providing the host cell with a suitable carbon source.
18. The method of any one of the preceding claims, further comprising isolating itaconic acid from the medium in which the host cell is cultured.
19. The method of claim 17, wherein the carbon source comprises at least one of a monosaccharide, a polysaccharide, a lipid and a fatty acid.
20. The method of claim 19, wherein the monosaccharide is one of mannose, glucose, arabinose and xylose.
21. The method of claim 19, wherein the polysaccharide is one of starch, a mannan and cellulose.
22. The method of claim 19, wherein the lipid is triacylglycerol.
23. A recombinant host cell which is genetically engineered to overexpress
  - (i) a polynucleotide sequence having at least 30% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity; and
  - (ii) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity;in comparison to a comparable host cell which has not been genetically altered.
24. The host cell of claim 23, further overexpressing at least one polynucleotide sequence as defined in any one of claims 2-4.

25. The host cell of claim 23 or 24, further underexpressing a polynucleotide sequences as defined in claim 12.
26. A host cell genetically engineered to overexpress at least one of the nucleic acid molecules having a polynucleotide sequence selected from
- (a) a polynucleotide sequence having at least 30% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity;
  - (b) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a multidrug transporter of the major facilitator superfamily, preferably of a transporter which exports itaconic acid out of the host cell;
  - (c) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a mitochondrial citrate transporter, preferably of a mitochondrial cis-aconitate transporter,
  - (d) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity;
  - (e) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a transcription factor for at least one of polynucleotide sequences of any one of SEQ ID NOs: 1, 3, 4, and/or 6
- in comparison to a comparable host cell which has not been genetically altered.
27. A host cell underexpressing a polynucleotide as defined in claim 12 in comparison to a comparable host cell which has not been genetically altered.
28. A host cell overexpressing at least one of the nucleic acid molecules as defined in claim 26 and underexpressing a polynucleotide as defined in claim 12 in comparison to a comparable host cell which has not been genetically altered.
29. A method of producing itaconic acid, comprising
- (a) culturing a recombinant host cell of claim 26, 27, or 28, and

- (b) obtaining itaconic acid.
- 30.** Use of a nucleic acid having a polynucleotide sequence selected from
- (a) a polynucleotide sequence having at least 30% sequence identity with the sequence of SEQ ID NO: 1, said polynucleotide sequence encoding a protein or fragment thereof having aconitate-delta-isomerase (ADI) activity;
  - (b) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 3, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a multidrug transporter of the major facilitator superfamily, preferably of a transporter which exports itaconic acid out of the host cell;
  - (c) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 4, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a mitochondrial citrate transporter, preferably of a mitochondrial cis-aconitate transporter,
  - (d) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 6, said polynucleotide sequence encoding a protein or fragment thereof having trans-aconitate decarboxylase (TAD) activity;
  - (e) a polynucleotide sequence having at least 50% sequence identity with the sequence of SEQ ID NO: 8, said polynucleotide sequence encoding a protein or fragment thereof having the activity of a transcription factor for at least one of polynucleotide sequences of any one of SEQ ID NOs: 1, 3, 4, and/or 6,
- an expression cassette, a vector, or a host cell comprising said nucleic acid for the production of itaconic acid in a host cell.
- 31.** Use of a nucleic acid as defined in claim 30, an expression cassette as defined in claim 30 in comparison to a comparable host cell which has not been genetically altered, or a vector as defined in claim 30, for the production of a host cell which is capable of producing itaconic acid.
- 32.** A method for producing a host cell which is capable of producing itaconic acid, comprising genetically engineering a host cell to overexpress a nucleic acid as defined in claim 30 or transforming said host cell with an expression cassette as defined in claim 30, or with a vector as defined in claim 30.

- 33.** The use of claim 30 or 31, or the method of claim 32, wherein the host cell is a cell of a fungus or a yeast.

Figure 1

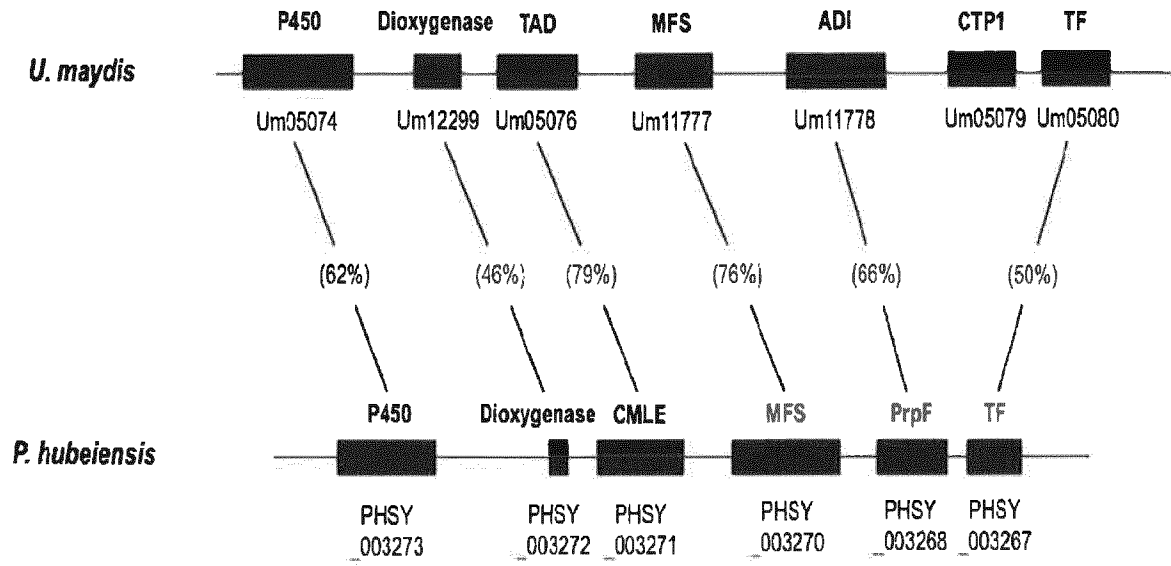
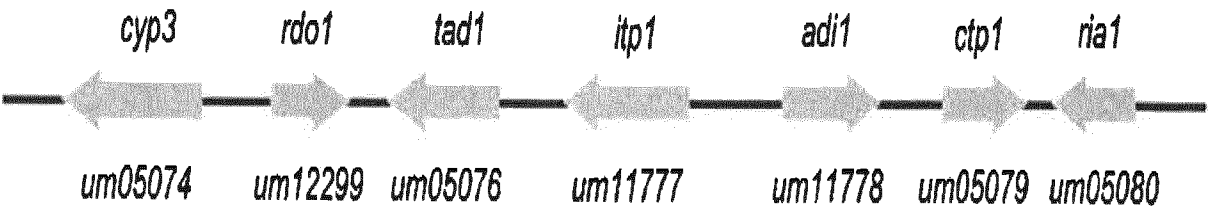


Figure 2

Itaconic acid production in deletion mutants and overexpression strains

| Gene    | Deletion | Overexpression |
|---------|----------|----------------|
| Um05074 | 0        | -              |
| Um12299 | 0        | 0              |
| Um05076 | --       | 0              |
| Um11777 | -        | 0              |
| Um11778 | --       | 0              |
| Um05079 | -        | ++             |
| Um05080 | --       | ++             |



3/27

Figure 3A

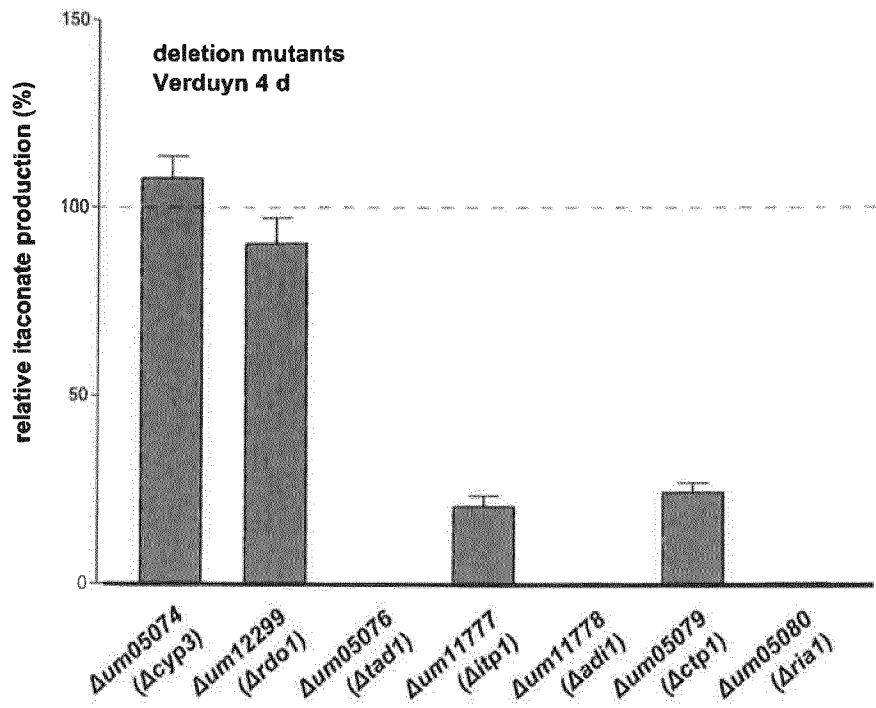
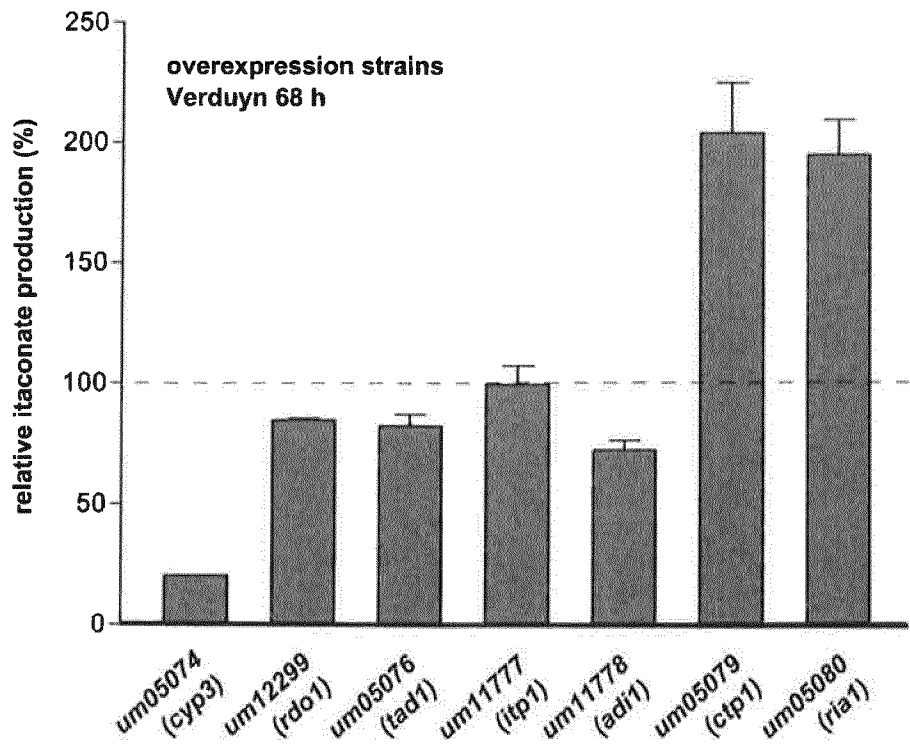


Figure 3B



4/27

**Figure 4A**

SEQ ID NO: 1 - Nucleic acid sequence adi1 UM11778 (ADI)

ATGTTGCATCCGATCGATACCACCATCTATCGTGCCGGCACCAGCCGCGGGCTCTACTTT  
CTCGCCAGCGATCTGCCCCTGAACCAAGCGAGCGAGATGCGGCGCTCATCTCGATCATG  
GGCTCTGGTCATCCATTGCAGATCGATGGCATGGGCGGTGGCAATTCGTTGACCTCCAAG  
GTGGCCATCGTCTCTGCTAGCACACAGCGCAGCGAGTTTGACGTCGACTATCTCTTCTGT  
CAGGTCGGCATCACTGAGCGCTTTGTGATACCGCCCCCACTGCGGCAATCTCATGTGCG  
GGCGTCGCTGCATTTGCCATTGAACGAGGTCTGGTGCAACCGCATCCGTCGGACACGACT  
TGTCTGGTTCGCATCTTTAACCTCAACTCCAGACAGGCTTCCGAGCTCGTCATCCCGGTC  
TACAACGGTCGCGTTCACTACGATGACATTGATGATATGCACATGCAGCGTCTTTCGGCG  
CGCGTCGGGCTGCGCTTTCTCGACACGGTGGGCTCGTGCACTGGGAAGCTTCTGCCACC  
GGAAATGCGAGCGACTGGATCGACGGCCTCAAAGTGTCCATCATCGACTCGGCAGTCCCC  
GTGGTGTTCATTTCGTCAGCACGATGTTGGTATCACCGGTAGCGAGGCGCCCGCCACGCTC  
AACGCCAACACTGCTCTCCTTGATCGGCTCGAGCGCGTTTCGGCTCGAGGCGGGCCGACGC  
ATGGGTCTCGGCGACGTCTCTGGTAGCGTAGTCCCCAAGCTTTTCGCTCATCGGTCCCGGT  
ACAGAGACGACCACGTTTACCGCACGCTgtaagtcgagtatattgttcgagatcgcatagc  
gttgatacctagtggtgacgaatacgaattcggtgttgagatggtcgacagATTTTACGC  
CAAAGGCTTGTCACAACGCACATGCGGTGACGGGTGCCATCTGTACGGCCGGGGCGGCGT  
ATATCGACGGAAGCGTGGTGTGCGAGATTCTTTTCGTCGCGTGCTTCGGCGTGTAGCGCGT  
CTCAGCGTCGCATTTTCGATCGAGCATCCGAGCGGCGTGCTCGAGGTGGGTCTCGTACCGC  
CTGAAAATGCGGCGCAGTCGCTCGTGGATGTGGCAGTGGTAGAGCGGTCCGTCGCGCTGA  
TCGCGCATGCTCGCGTCTACTACACCACCCAGATAGGCGGCGCTCGTACGACTCACCGC  
TCACTTCGCCCTCCACGCCCGCCGACACGCACAACCTGTTTCGATGCAGCGTACCGTCCCG  
TGATACAGCCTAGTGACACTGACGTAGAGGCTCCACACATGCTTGCGCTCGAAAACAAGG  
AGCAATGCGTGTCTCGGTGCGATACCGCGCTCCACCACATCGTAGCCAGCTACGGCGCTA  
GCGATGCACACGCATCCGACCGCAGCCTCTCTTAG  
(variant, source MB215, sequenced, introns indicated by  
lowercase letters)

**Figure 4B:**

SEQ ID NO: 2 - Polypeptide sequence UM11778 (ADI)

MLHPIDTTIYRAGTSRGLYFLASDLPAEPSEERDAALISIMSGHPLQIDGMGGGNSLTSK  
VAIVSASTQRSEFDVDYLCQVGITERFVDTAPNCGNLMSGVAAFAIERGLVQPHPSDTT  
CLVRIFNLNSRQASELVIPVYNGRVHYDDIDDMHMQRPSARVGLRFLDTVGSCTGKLLPT  
GNASDWIDGLKVSIIIDSAVPVVFIRQHDVGITGSEAPATLNANTALLDRLERVRLEAGRR  
MGLGDVSGSVVPKLSLIGPGTETTTFTARYFTPKACHNAHAVTGAICTAGAAAYIDGSVVC  
EILSSRASACSASQRRISIEHPSGVLEVGLVPPENAAQSLVDVAVVERSVALIAHARVYY  
TTPDRRRSYDSPLTSPSTPADTHNLFDAAYRPVIQPSD TDVEAPHMLALENKEQCVSRCD  
TALHHIVASYGASDAHASDRSLS\*  
(variant, derived from MB215 sequence, deviation at position  
350 underlined)

5/27

Figure 4C

SEQ ID NO: 3 - Nucleic acid sequence UM11777

ATGGACCAAGCCGATCATTCGGCGTCCCAGACGACGCTGCGCTCGAGGAGGCGCCCAAT  
ACAGTACCGATTCAAGAAAAGAGCGCACAGCCTCACGACACGCAGCCGTACTGTGCATTC  
ACGAAGCGGTCCAAGCTGTTTATCGTCCTCACCGTCTCTCTTGCCGGCTTTTTTTCGCCG  
TTTGCCATCAACATTTACATCCCTGCCCTGCCCCAGATCGCCGGTATGCTGCATACCTCT  
GAAGgtgagtgaaacctatcggctgaggcgcgaaacgtgcagagcaaccgattcgggagc  
Cgtcaacttggtttctcattcacctctgtcctcctttccgcttcttctctttccccctt  
acaccttgccagCCGCTACCAACGTCAACGTGACCGTCTACATGATTGCACAGGGACTTT  
CACCGTTATCTGGGCTCCTCTCTCGGATgtgagtagccaaaattgcgctcgggtggctt  
Tagcttggcggtgtctctctctgaccacatatgcaagcgtttcttgcgttttttgggcttg  
atcttggtctcgatcttggtgatgcatcacagGTGTTTGGGCGCCGACCGATCTACATCTT  
GACGTTTTTTCATCTTTTTCATTGCCAACCTCGGGCTGTCGTTACCAACGTCTACTGGCT  
CTTGGTGGTGCTGCGTATGGTGCAGGCTGCAGGCGCGTGCAGTGCATTGCGATTGGCGC  
TGGTACGATTGGCGACGTTACTGAGCGCAAGGAACGAGGAAGTTACATGGGCTACTATGC  
GCTTGCAGCAATACAGGGCCCCGCCATCGGACCGGTTGTGCGTGGCGCGCTTTCGCAGAG  
ATGGGACTACCATGCTACGTTTTTCTTCTTGACGGCGATCTCGGGTCCGTTTTTGTGTT  
CATGCTTCTCTTCTCGTTGAAACGCTTCGAGTCATTGTTGGCAACGGCAGTGCAAAAAC  
GTCGGGCATCTATCGCCCCCTTGGTGGAGCCCAAGCTGCAACGCTCGATCGCCAATGCGCC  
TCGGCCTGGCATCAAGAACCCTGTCATGGCACGCTCGATTTTCGGCTTCCACCGTCCGTT  
TTTGGTGTGTTGCGCGCCCCGAGACCAGCCTAGCCATCCTGGCTTTTTCGATGGTATATGC  
GAGCTACTACTTGTCTATCTGGTTCGCTGCCGTATCTGTTCAAGCAGGTGTACGGTCTGGA  
CGAGCTCCTGATCGGCGTATGCTTTGTTCCGAGCGGTGTGGGGTGTGCGGTGGGCACAGT  
GCTCGCCGGCAAGATCCTCGACTGGGACTATCGTCGTGCGTTGGACAAGAGCAAGCTTGG  
TGTCAGGTGACGCGCGCAAGGTTGCAGTCGGCGTGGATCTACCTGCCGTGCTACTGCGC  
TTCGCTTCTGGCGTACGGATGGTGTGTTTCGTGCGCATACGCATATCGCCGCTCCGATCGT  
GTTTCAGTTTACACgtaggttgcagaccctcgctcggctagatactgtactgcgattg  
ctgaccatgcttggtgtgctgttcgcatttggtgatttctacagTGGGCATGTTTTTCGACC  
ATGTACTTTACCAATGTCAATACGCTCATTGTGACCTGTATCCTGGCAAAGCGGCCAGC  
GCAACCGCAGCGGTCAACGTGCGACGGTGCTTGCTCGGCGCAGTAGCAGTCGCCGTGGTC  
CAGCCCATGATCGACGCAATGGGCGCCGGCTGGACTTTTACGCTCGGCGCACTACTCACA  
CTGATCGTCGGTCTCATCTGCCAAGTTCTCATCTACCTCTACGGCGAAATGTGGGCAGCT  
CGCAAACACTCGTGA

(U.maydis 521 data base entry Um11777, introns indicated by lowercase letters)

6/27

## Figure 4D

SEQ ID NO: 4 - Nucleic acid sequence UM05079 (CTP1)

ATGCCGCCGTCTGGCCGTAAAGTGTCGCCTAGCGTCTCAGTAGTAGCAGGAGCCACAGCC  
GGCGCTGTTGAAGGAGgtgagtttacttggcgccaacacgaaatgacgaatcaccacgat  
ctgaccaagagccggaaccgaatgcttgtgttgctcgcttggcttgttggtggatcagTCG  
CTACGTTTCCGATCGAGTATCTCAAGACAGTTTCGCAGTTTGCACCGCGCGATGTGCACG  
GGAACCAGCAGCGGCTATCTCCGATCGAGGTGGTCCGATCGACGCTCCAGAAGGAAGGTC  
CCAAGGGGCTCTTTTCGAGGCTGTACCGCTATGGTTGTAGGCAATGCCGGTAAGGCTGGTG  
TACGATTTTTTCGCTTTTGAAAACCTCCGCAGTATGCTCAAGAACAAGTCTACAgtaagtc  
Ggcgcacatcttctgctctacagtgcgagctagagcgtggtttgccattccacgaacgctga  
cgtttgccccggttttttggcggttgacctttagGGAAAATTGTCCAACCTCGAGCAACTAC  
CTTGCCGGTATGGGCGCCGGGACGTTAGAGGCCATCTTTGCCGTCCTCCGAGCGAGACG  
ATCAgtgagtggttccagcggtagatcaaaccacgcaattgcgcgcacccgaacgccaac  
Tgaccggctggtgcggatgtttgtcttgttttgtttgatcatttttcttcttatggctat  
tgtgcggtacgcaacagAGACCAAGCTGATCGATGATAGCAAGCGAGCCAAGCCACGCTA  
CGAGCAAGGGCTTGTGCGTGGTACGGCGTCGATTGTACGACAGGAAGGGTTGGCAGGCAT  
CTACCAGGGGGTTGTACCGGTAGTGATGCGGCAGGGGTCTGCATCTGCGATTCTGGCTGGG  
GACGTACTCTGCGCTGCGAGATTGGCTTCCGAAAGCGCACGGTAGTGGATCGTCATTGAT  
CAACTGGCTGGCTACGTTTTTCGATCGGCGCGGCATCTGGCGTAGTCGCGGTGTATGGAAC  
GATGCCATTTCGACGTGCTGAAAACGCGCATGCAGGCCATAGACGCTGCACGCTACCGCTC  
CACCTGGCACTGTCTACCAACACCCTGAAAACAGAAGGCGCAGCTGCCCTGTGGCGTGG  
CTCGGTATCACGCAGTATGCGTCTCATCGTCAGCGGAGGCGTCATCTTCTCGGTCTACGA  
ACAGGTCGTCTGGCTCCTAGCAGGTCCCGAGTCGTAG  
(U.maydis 521 data base entry Um05079, introns indicated by  
lowercase letters)

7/27

**Figure 4E**

SEQ ID NO: 5 - Nucleic acid sequence UM12299

ATGTTGCGTTCTAGCCAGGCCAGAAGCGTGTTACGCTCCAGTCAGTGGGCCACCACCGCG  
CGAGTTCACCAACTGGAGCTGCCAAGCGGGTGGAAGCCGAGCGCACTCGGAGTAGCACCA  
TGGCAGCAGCGCCAGCAGCGCCAGCAGCGCCAGCTGTCGGTCAGGTCGCTCGATCACTTG  
GTGATCACCTGCCACGATATGGACAAGACCATTGACTTTTACACGCGCCTCGGCATGGGC  
GTGGTGCAGTTTCGGACAAGGGCGCAAAGCGCTCGAGTTTGGCTCCCAGAAGATCAACCTG  
CACCAGAAAGGCAAAGAGTTgtaagtacagtgtggtgtgtggtcagcaagcagctcaaca  
aggacccaggccactgacaccgactttgtttgtctgcttggtctcggctcagCGAACCGA  
gtaagttgcaaggcggcgcatgtggtgcggcgaatgtaggtggttgcggtggctgcgaatg  
catgtactgacggttggtgtgtcttcagAGGCGTTGGTCCCGCAACCGGGATCCCAGGA  
CTTGTGTTTCGTCATTACGACAGCATCGCCGACGCCCAGgtaagaagcgtgtactagct  
gcacccgtgcatccgtgcaaggcgcccttaagctgacgcctgagctctcgacagAAACAC  
CTGCAAGAGCACGGCATCCAAGTTGTCGAGGGCCCCGTCAAGCGCACCGGTGCTGTGCGGA  
CCGATCCTCAGCATCTACGTGCGCGATCCCGACAATAACCTGATCGAGgtgagtgccaaag  
ccccgcgattttcagtgcaatacaaaacaagcctcgctaacacacacgctctgctctccagC  
TCTCGTCGTACCAAGACGCAAAGTGA

(U.maydis 521 data base entry Um012299, introns indicated by lowercase letters, manually corrected for wrong intron prediction)

8/27

Figure 4F

SEQ ID NO: 6 - Nucleic acid sequence UM05076 (TAD)

ATGGCACCTGCACTCAACGCAAACCCTACCACGAAACGCGACGAGCTGAGCGCTCCGTCTG  
GCATCGCACAAAGCTCGGCATGTCTGAGCATGGCGAGCAGGGCGGCAGGCGGCGGTCTCAAG  
CTCACTGGTCTGCCCCGATCTCTCGGACTCGGCAGGAACGCTGAGCGACATTTTCGGTACG  
CCGCAGATGCGCGAGATCTGGTCTGGACAGAACCGTGTGGCGTGCTATCTCGAGATCGAA  
GCGGCGCTCGCGATTGTGCAGGCGGATCTCGGGATCATCCCCAAGAATGCGGCACACGAG  
ATCGTCGAACACTGCCGCGTTCAAGAGATCGACTGGGCTCTGTACAAGCAAAAGACCGAG  
CTGATCGGCTACCCCGTGCTGGGCATCGTGCAGCAGCTCGTCGCCAACTGCAAAGATGGT  
CTCGGCGAGTACTGTCACTGGGGCGCCACAACGCAGGATATCACTGACACCGCCACCGTC  
ATGCAGATCCGCCAGTCGCTCACGCTCGTCAAGCAGAGACTCGACAGCATCGTCTCGAGC  
CTCGAGCATCTCGCCGAGCAGCATCGCAACGTGCCCATGGCGGCTCGTTCCAACCTCAAG  
CAGGCGGTACCGATCACGTTTGGCTTCAAGATGGCGCGCTTCCTCGCCACGTTCCGCCGA  
CACCAGCAGCGTCTCGTCGAGCTCGAAAAGCGCGTCTACACGCTCGAGTTTGGCGGTGCA  
GCGGGCAACTTGTCGTCTGGGTGACCAGGGCATTGCGACGCACGATGCGCTTGCCAAG  
ATGCTCGACCTGGCGCCCGCCGAGATTGCGTGGCACACGGAACACGACCGCTTCGCCGAG  
GTAGGTACCTTCCTCGGCCTGCTCACTGGAACGCTTGCCAAACTCGCCACCGACATCAAG  
CTCATGTCTGCAGACCGAGGTGGGCGAGGTGGGCGAGCCGTTTATCTCGAACCGCGGCTCG  
TCGTCTGACGATGCCACAGAAGAACAATCCGATCTCGTGTGTCTACATTCACGCGTGTGCG  
GCGAATGTGCGTCAGGGCGCTGCAGCGCTACTCGATGCCATGCAGTCTGATCACGAACGT  
GGCACGGGTCCCTGGGAGATCATCTGGGTCCAGCTGCCACTCATGATGAACTGGACCTCG  
GCCGCTCTCAACAACGCCGACTTTGTCTGCGCGGCCTCCAGGTGTTCCAGACGCAATG  
CAACACAACCTGGACCTCTCGAAAGGGCTCATCGTCTCGGAAGCCGTATGATGGGTCTC  
GGTAACACGCTCGGCCGTGAGTACGCACACGACGCCGTCTACGAATGCTGTGCAACCGCG  
TTCGTCCAAGACAGACCGCTCCTCGACGTCCTCCTCGAAAATCACGAGATCGCCTCCAAA  
CTAGACCGCACCGAGCTTGAAAACTCTGTGATCCCGCCAACTACCTCGGCCAGTGTTTCG  
CAGTGGATCGATCGCGTGCTGTCTCGCCCATCGTCTGGCCTGA  
(U.maydis 521 data base entry Um005076, no intron)

9/27

## Figure 4G

SEQ ID NO: 7 - Nucleic acid sequence UM11778 (ADI)

ATGTTGCATCCGATCGATAACCACCATCTATCGTGCCGGCACCAGCCGCGGGCTCTACTTT  
CTCGCCAGCGATCTGCCCCGCTGAACCAAGCGAGCGAGATGCGGCGCTCATCTCGATCATG  
GGCTCTGGTCATCCATTGCAGATCGATGGCATGGGCGGTGGCAATTCGTTGACCTCCAAG  
GTGGCCATCGTCTCTGCTAGCACACAGCGCAGCGAGTTTGACGTGACTATCTCTTCTGT  
CAGGTCGGCATCACTGAGCGCTTTGTGATAACGCCCCCAACTGCGGCAATCTCATGTGCG  
GGCGTCGCTGCATTTGCCATTGAACGAGGTCTGGTGCAACCGCATCCGTCGGACACGACT  
TGTCTGGTTTCGCATCTTTAACCTCAACTCCAGACAGGCTTCCGAGCTCGTCATCCCGGTC  
TACAACGGTCGCGTTCACTACGATGACATTGATGATATGCACATGCAGCGTCCTTCGGCG  
CGCGTCGGGCTGCGCTTTCTCGACACGGTGGGCTCGTGCACTGGGAAGCTTCTGCCACC  
GGAAATGCGAGCGACTGGATCGACGGCCTCAAAGTGTCCATCATCGACTCGGCAGTCCCC  
GTGGTGTTCATTTCGTCAGCACGATGTTGGTATCACCGGTAGCGAGGCGCCCGCCACGCTC  
AACGCCAACACTGCTCTCCTTGATCGGCTCGAGCGCGTTCGGCTCGAGGCGGGCCGACGC  
ATGGGTCTCGGCGACGTCTCTGGTAGCGTAGTCCCCAAGCTTTCGCTCATCGGTCCCGGT  
ACAGAGACGACCACGTTTACCGCACGCTgtaagtcgagtatttgttcgagatcgcatagc  
gttgatacctagtggctgacgaatacgattcgtgttgagatggtcgacagATTTTACGC  
CAAAGGCTTGTCAACCGCACATGCGGTGACGGGTGCCATCTGTACGGCCGGGGCGGCGT  
ATATCGACGGAAGCGTGGTGTGCGAGATTCTTTTCGTCGCGTGCTTCGGCGTGTAGCGCGT  
CTCAGCGTCGCATTTTCGATCGAGCATCCGAGCGGCGTGCTCGAGGTGGGTCTCGTACCGC  
CTGAAAATGCGGCGCAGTCGCTCGTGGATGTGGCAGTGGTAGAGCGGTCCATCGCGCTGA  
TCGCGCATGCTCGCGTCTACTACACCACCCAGATAGGCGGCGCTCGTACGACTCACCGC  
TCACTTCGCCCTCCACGCCCGCCGACACGCACAACCTGTTTCGATGCAGCGTACCGTCCCG  
TGATACAGCCTAGTGACACTGACGTAGAGGCTCCACACATGCTTGCGCTCGAAAACAAGG  
AGCAATGCGTGTCTCGGTGCGATAACGCGCTCCACCACATCGTAGCCAGCTACGGCGCTA  
GCGATGCACACGCATCCGACCGCAGCCTCTCTTAG

(U.maydis 521 data base entry Um11778, introns indicated by lowercase letters)

10/27

**Figure 4H**

SEQ ID NO: 8 - Nucleic acid sequence UM05080 (Txf)

ATGCGCTTTGCAGGGATGAGTTGCGACGACGAGCGTCCTGCCAACATGTTTGATCTGATGGC  
GCCACAGTTGGCCTCGACATCGTGCAACGAGCACTACTTTTCCACTGCGGACTTGGGAGCTT  
CGACGTTGTACGCCACGACGACGATGCACCAGCCACGATAGCTGGCATACTTACGCCACAG  
CCAGCGACGTTGGCTCCGATGTACAGCACGTGTCCGGTGCGTTTTGACGACCAGCGCTCTGC  
ACCGGCATCCACATCCGTGTGCGGGCAAACGCAAACACTCGGACGTGGAAAAGGACCGTCGAA  
GAAGCATTTCAAATGGGTTTGCGGtaagtaggctacttggcctcgttggatcacgcgccagc  
cgaagctcaagctgaccgaatgcgttgaatggtgcatgaacctgtgcgtcgacttgatcagG  
TGCTGCAGAAATGTGCTTCACAACGAGTCCAACGCTAAACCGATCTCCAAGTCGATTCTGCTC  
CAGCAGGCGTGCAGACGAGATCCGCGAACTTCGCAAGAAACTCGACGCTAGTACTACCATCAT  
CTCGCGTTTTGGTCTGGAAAACCTGTTCTGTACCCACACCCTCATCCACGCATGCGTCGCCGC  
CGAATGCGTCTAGTGAATCTACTCGCCCATCAACCAGGCTTCCGATGTACTGGCTGATACA  
CGTCGTGCTTCCATATCGACGAGTGCTACGCCGATCCTGTATAGCGAAGAGAAGCGCAAGGC  
GAATGCGAAACGCAGACACTCGTACGATGGGTGCTGGCAAGCATCGGATCGTGGCTCGATCG  
ACGACGAAGCCAGTGCCAGTGCCAGTGCCAGTAGCAGCGCTAGCTGTAGCAGTAGTAGCCAC  
ACGCATTCCGACGATAACCGACTGCGACGACACCGACACCGACATCCC  
AGCCGAATCCGCACTCAAGGAGAGAACCAAACGCCACAAGGCCAGATCCAAGAAAGAACGAG  
ATCGCACCAAACCGCGTTACAGACCGAAACCATCAACCAACCGTTCACCAACTCCGTCCTCC  
TGCGCTTCCAGTACGCCCAACTCGCCGCCACATCATCAACCGCAACCGCGACCTCCAACA  
GGCCATCCTCTCTCTCTCTCGAACTCCCCTCGCATCTCGAAGACGTCAAAAACAAAAGAC  
GTGCTTCGCAACCAACCGAACTCGCCGATCCTTCCAGCGTAAAGTCCAGGTCCAAGAAACGT  
CACCGATAA

(U.maydis 521 data base entry Um05080, introns indicated by lowercase letters)

**Figure 4I**

SEQ ID NO: 9 - Nucleic acid sequence UM05074 (P450)

ATGAACACGACCAAACACTACTCGGTACCGGGGCGCTCAGCCCGTCATTTGTGTTTTGACCAC  
GATTCTGGCAACGCCATCTTTGGACTTTTCGAGCTCGACGCTGGTTGTTCTTGTAGCCATG  
ATCGCGGTCTCGACGCTCACGCTCAAGAGCGTACTGCCCGGAGATCGCAGCATCAATCTT  
CCGGGACCTCGAGGCTGGCCGATCGTCGGCTCTTGGTTCGATCTCGGCAACAACTGGGGC  
GAGTACTTCCGTCAAGCTGCCAAAGAATACGGTGATgtgcgctctctcgtggtttccgcag  
Tgcatccatgtctcgtgtggttggcatgggtcgagctgacgtgctgacgtgctgacgttc  
tgacgttctgacgtgtatcttgccctctgtatgtacgccagGTATTCAAAGTACACATCGG  
GAATCGCACCGTGGTGGTCGTCAACTCACCAAGGCGGCGCACATTCTGTTCAACGAACA  
CGGCTCGTCGCTCATCTCACGACCATGGTTCATACGTTCCACGGCGTGCTGTCCAAGTC  
GTCTGCCTTCACTATCGGTACGTGCGCCTGGAGCGATAGCACAAAGAACAAGCGCAAGGC  
AGCTGCCACTGCGCTCAATCGTCCGGCGGTCCAGTCGTACATGCCCATCATTGTCTGAAGA  
GTCGTTGGACGCCGTTTCGTGCGATCCTCAACGACGGCAACGCGGGCAAGAACGGCATTTGT  
TCCGTACAGCTACTTTCAGCGTCTTGCGCTCAACACGAGCTTTCAGGTCAACTACGGCTT  
CCGCATGGGCGAGCGGACGACGCTGTGTTTGATGAAATCTCCGAGGTGATTGCTAAAGT  
GGCCTCgtacgtatcagtcgactcgactcttgccacctacctgcttgacttgcatcgcga  
tttgctaacttggcgctctgcttctcgtacagTGTTTCGCGCTGTGACCGGTTCGTTGCAG  
GACTACGTTCCGCTCATGCGCTACCTGCCCGCTAATGCAAAGAGCAAGGCGGCTGCTTCG  
TACGGTCTGCGTCGTAAAAAGTTTCATGTGCAAACTGTACGAAGAACTCGAGCAACGCGTC  
AACCAAGGCAAGGATGAGAGTTGCATCAgtacgttgatcacatccaaccaagcgtgccag  
Tcgtgccagtcgtgccaatcgtaccagacaggctgctcaccacactcaccactcttggtg  
cttttcttggttcgcagCCGGCAACATCCTTAAAGACACTGAATCGCGCAAGAAGCTGTCT  
CGCTTGGAGATCGACTCGATCTGTCTCTCGATGGTCTCGGCAGGTCTGGATACTTTTGCC  
AACAgtagttagtcaactcagagatacaacacaagtgcaccgatcccgtgtctgaccag  
ttcttttctgtgtattggcggtgttggcagCAATGATCTGGACGATCGGTTTCCTGGCCAA  
GCATCCAGAGATCCAGCGCAAGGCACAGGCCGAACCTGTTGGCTCACTATCCGAACCGGGA  
GCTTCCGCATGTTGACTCGGAGGACTTGGTGTATATCCATGCCATGGCCAAGGAGGCATC

**11/27**

ACGACTGTTCAACGTGTTCCGCATCTGTCTGCCGCGTACAAACGTGAGCGACGTACCTA  
CAACAACGCCGTGATTCAGCGGGGACGACGTTTTTCCTCAACTCATGGGCGTGCAACGT  
GGACGCTGAAAAGTTGCGCGACCCGTTTCGAGTTCAAGCCGAGCGCTTCATGGACAAGAG  
CGCCAGTAATGCACACGTCGAGAACAAAGATGGGCGGCGTCGAGACGTACGCTTTCGGCAT  
GGGAAGACGCATGTGTCCCGGCGTGTTCTTGGCGCTTCGCGAAATCTACACGACGCTCGT  
CTTCCTCACCCACTTTTTTCGATATTGCTCCCGACGGAGAGTATGACATCGACCCTCTCAC  
AGCCGTAGAGGACGGTCGCGCGTTTCAGCGTACGTCCGAAACCGTTCAAGGTGCGCTGCAC  
GCCTCGACCCGGCGTCGACCTTTCCCCCGTGCTCGACAAACAATAG

(U.maydis 521 data base entry Um05074, introns indicated by  
lowercase letters)

**Figure 4J**

SEQ ID NO: 10 - Polypeptide sequence UM11778 (ADI)

MLHPIDTTIYRAGTSRGLYFLASDLPAEPSEDAALISIMSGHPLQIDGMGGGNSLTSK  
VAIVSASTQRSEFDVDYLCQVGITERFVDTAPNCGNLMSGVAAFAIERGLVQPHPSDTT  
CLVRIFNLNSRQASELVIPVYNGRVHYDDIDDMHMQRPSARVGLRFLDTVGSCTGKLLPT  
GNASDWIDGLKVSIIIDSAVPVVFIRQHDVGITGSEAPATLNANTALLDRLERVRLEAGRR  
MGLGDVSGSVVPKLSLIGPGTETTTFTARYFTPKACHNAHAVTGAICTAGAAYIDGSVVC  
EILSSRASACSASQRRISIEHPSGVLEVGLVPPENAAQSLVDVAVVERSIALIAHARVYY  
TTPDRRRSYDSPLTSPSTPADTHNLFDAAYRPVIQPSD TDVEAPHMLALENKEQCVRCD  
TALHHIVASYGASDAHASDRSLS\*

(U.maydis 521 data base entry Um11778)

12/27

**Figure 4K**

SEQ ID NO: 11 - Polypeptide sequence UM05074 (P450)

MNTTKLLGTGALSPSFVFDHDSGNAIFGLSSSTLVVLVAMIAVSTLTLSVLPGDRSINL  
PGPRGWPIVGSWFDLGNNWAEYFRQAAKEYGDVFKVHIGNRTVVVVVNSPKAAHILFNEHG  
SSLISRPWFYTFHGVLSKSSAFTIGTSAWSDSTKNKRKAAATALNRPAVQSYMPIIVEES  
LDAVRRILNDGNAGKNGIVPYSYFQRLALNTSFQVNYGFRMGERDDGLFDEISEVIAKVA  
SVRAVTGSLQDYVPLMRYLPANAKSKAAASYGLRRKKFMSKLYEELEQRVNQKDESCIT  
GNILKDTESRKKLSRLEIDSICLSMVSAGLDTFANTMIWTIGFLAKHPEIQRKAQAELLA  
HYPNRELPHVDSEDLVYIHAMAKEASRLFNVFRI CLPRTNVSDVTYNNAPIAGTTFFLN  
SWACNVDAEKFADPFEFKPERFMDKSASNAHVENKMGGVETYAFGMGRRMCPGVFLALRE  
IYTTLVFLTHFFDIAPDGEYDIDPLTAVEDGRAFSVRPKPFKVRCTPRPGVDLSPVLDKQ  
(U.maydis 521 data base entry Um05074)

**Figure 4L**

SEQ ID NO: 12 - Polypeptide sequence UM12299 (Dioxygenase)

MLRSSQARSVVRSSQWATTARVHQLELPSGWKPSALGVAPWQQRQQRQQRQLSVRSLDHL  
VITCHDMDKTIDFYTRLGMGVVQFGQGRKALEFGSQKINLHQKGKEFEPKALVPQPGSQD  
LCFVIHDSIADAQKHLQEHGIQVVEGPVKRTGAVGPILSIYVRDPDNNLIELSSYQDAK  
(U.maydis 521 data base entry Um12299, manually corrected for  
wrong intron prediction)

13/27

**Figure 4M**

SEQ ID NO: 13 - Polypeptide sequence UM05076 (TAD)

MAPALNANPTTKRDELSAPSASHKLGSMSSMASRAAGGGLKLTGLPDLSDSAGTSLSDIFGT  
PQMREIWSQNRVACYLEIEAALAIVQADLGIIPKNAHEIVEHCRVQEIDWALYKQKTE  
LIGYPVLGIVQQLVANCKDGLGEYCHWGATTQDITDTATVMQIRQSLTLVKQRLDSIVSS  
LEHLAEQHRNVPMARSNLKQAVPITFGFKMARFLATFRRHQORLVELEKRVYTLEFGGA  
AGNLSSLGDQGIATHDALAKMLDLAPAEIAWHTEDRFAEVGTFLGLLTGLAKLATDIK  
LMSQTEVGEVGEFFISNRGSSSTMPQKNNPISCVYIHACAANVRQGAAALLDAMQSDHER  
GTGPWEIIWVQLPLMMNWTSAALNNADFLRGLQVFPDAMQHNLDLSKGLIVSEAVMMGL  
GNTLGRQYAHDAVYECCRTAFVQDRPLLDVLLNHEIASKLDRTELEKLCDPANYLGQCS  
QWIDRVLSRPSSA\*

(U.maydis 521 data base entry Um05076)

**Figure 4N**

SEQ ID NO: 14 - Polypeptide sequence UM11777

MDQADHSGVPDDAALEEAPNTVPIQEKSAPPHDTQPYCAFTKRSKLFIVLTVSLAGFFSP  
FAINIYIPALPQIAGMLHTSEATNVTVTVMIAQGLSPVIWAPLSDVFGRRPPIYILTFF  
IFFIANLGLSFTNVYWLLVVLRMVQAAGACSAIAIGAGTIGDVTERKERGSYMGYYALAQ  
YTGPAIGPVVGGALSQRWDYHATFFFLTAISGPFLFLFLLVETLRVIVGNGSAKTSGI  
YRPLVEPKLQRSIANAPRPGIKNPLHGTLDGFGHRPFLVFARPETSLAILAFSMVYASY  
LSSGSLPYLFKQVYGLDELLIGVCFVPSGVGCAVGTVLAGKILDWDYRRALDKSKLGVKV  
TRARLQSAWIYLPYCASLLAYGWCVRATHIAAPIVFQFTLGMFSTMYFTNVNTLIVDL  
YPGKAASATAAVNVGRCLLGAVAVAVVQPMIDAMGAGWTFTLGALLTLIVGLICQVLIYL  
YGEMWAARKHS

(U.maydis 521 data base entry Um11777)

**Figure 4O**

SEQ ID NO: 15 - Polypeptide sequence UM05079 (CTP1)

MPPSGRKVSPSVSVVAGATAGAVEGVATFPIEYLKTVSQFAPRDVHGNQORLSPIEVVRS  
TLQKEGPKGLFRGCTAMVVGNAGKAGVRFFAFENFRSMLKNKSTGKLSNSSNYLAGMGAG  
TLEAIFAVTPSETIKTKLIDDSKRAKPRYEQGLVRGTASIVRQEGLAGIYQGVVPVVMRQ  
GSASAIRLGTYSALRDWLPKAHSGSSLINWLATFSIGAASGVVAVYGTMPFDVLKTRMQ  
AIDAARYRSTWHCLTNTLKTEGAAALWRGVSRSMLIVSGGVIFSVEQVWVLLAGPES

(U.maydis 521 data base entry Um05079)

14/27

**Figure 4P**

SEQ ID NO: 16 - Polypeptide sequence UM05080 (Txf)

MRFAGMSCDDDERPANMFDLMAPQLASTSCNEHYFSTADLGASTLYATTTDAPATIAGILT  
PQPATLAPMYSTCPVRFDDQRSAPASTSVSGKRKHS DVEKDRRRSISNGFAVLQNVLHNE  
SNAKPISK SILLQQACDEIRELRKKLDASTTIIISRFGLENLFVPTPSSTHASPPNASSRI  
YSPINQASDVLADTRRASISTSATPILYSEEKRRKANAKRRHSYDGSWQASDRGSIDDEAS  
ASASASSSASCSSSSSHTHSDDTDCDDTDTDIPAESALKERTKRHKARSKKERDRTKPRYR  
PKPSTNRSPTPSSCASSTPNSPPTSSNRNRDLQQAILSLLELPSHLEDVKNKRASQPT  
ELADPSSVKSRSKKRHR

(U.maydis 521 data base entry Um05080)

**Figure 4Q**

SEQ ID NO: 17 - Nucleic acid sequence UM02807 (ADI2)

ATGTTGCTCCAGCCTACCTCGAGCCGTCACCTTTGAAGACTTTCGCTATGTCGAGCCCAAG  
ATCTTGACCAAAAGCGCTCTGCCCATCGCTCCGGCGCGACTACGGCAGGTGCGTCTTCT  
AGCAAGTCGCCATCGTTTGAACGTCGCTCCCTTCGCACGGTAATCATGCGAGCAGGAACC  
TCAAAGGGACTTTTCATCAAGGCTTCCGATCTGCCAGCTCTCGAGCCGAATGGCAAAC  
ATCTTGCCCTTCGATCATGGGCAGTCCTGATCCGTTTGGCCGTCAGCTCAATGGATTGGGT  
GGTGGCACTTCGACAACCTTCCAAGATTGCTGTTGTCTCGCAGAGCTCTGCTCCTCACATC  
GCCGACGTCGATTACCTGTTTCATCCAGTGCCCCATCGAAGGCGACAAGCTCGACTTTACC  
GGCAACTGCGGCAACATTCTTAGTGGTGGTGGCCCTTTTGCAATTGAAGAGGGTCTCATC  
CCGGCGTCAGTCCTGGCACCCTTGCTGCATTTCGCCAAGACCAACGGCAAAGGCCGACAA  
GACAAGGTCGCACTCACTCTTCGCTGTCTCAACAACAACCAGCTCATCCGATCTACATTC  
CTTGTTTCGCAATGGAAAGCCGGTCGAATTTGGCGACGTGGTCATCGACGGTGTTCAGGT  
ACAGGCTCGCCGATCCAGTTGGATTTCTTGACCCAGCGGGCAGCATGTGTACCTCGCTC  
TGCCCCACTGGCAAACCTTTGGATTTGCTGCACATACACGGCGAAGACTTGCCGATCGAA  
GTATCATGCGTCGATGCAGCCAATCCGTTCTGTTGTTGCGCTCTCCGACATTGATGAG  
ACCCTGCGCGGAAACGAGCCAGCTTCCGTACTGGAACGCCACTCGGCTCGCGTAGAGCTG  
ATCCGTCAAGCAGCCGCCGTCGTAATGGGTCTTGACCCGACACGGCCACTGCAGCAAAG  
ACCAAAGGCACACCCAAGATCTGTCTGGTTTCCGACGCGCTCCCCGGCTCCGACGCACAC  
GTTCTCTCGCGAAGCTTCAGCATGGGTGACACATCCTGCATTACAGCTCTCCGGTGGT  
GTCTGCCTCGCTGCTGCCTGCAGCATTCCAGGGTCGATCCCGAACCAGATTCTGCTCAAG  
CAAACCAAGATGATGCCGGAAGGCTCAAGTTTCGCACACGCATGCGGTGCCATTGAAGCT  
ACCGCGGATGTGAGATGGATAGCAAGAGCCCGGTTGGCGTCACTGTGAGGAGCACGAGT  
CTTTTCAGGACTGCAAGGAGACTGGCTTCTGCTGAAGCGTATTACCTCTCGCCTAGCCAG  
TAG

(U.maydis 521 data base entry Um02807)

15/27

Figure 4R

SEQ ID NO: 18 - Polypeptide sequence UM02807 (ADI2)

MLLQPTSSRHFEFRYVEPKILTKSALPIAPGATTAGASSSKSPSFERRSLRTVIMRAGT  
 SKGLFIKASDLPSRAEWQNILPSIMGSPDPFGRQLNGLGGGTSTTSKIAVVSQSSAPHI  
 ADVDYLFIQCPIEGDKLDFTGNCGNILSGVGPFAFEEGLIPASVLAPLAAFAKTNGKGRQ  
 DKVALTLRCLNNNQLIRSTFLVRNGKPVEFGDVVIDGVAGTGSPIQLDFLDPA GSMCTSL  
 CPTGKPLDLLHIHGEDLPIEVSCVDAANPFVVRVLSIDIDETLRGNEPASVLERHSARVEL  
 IRQAAAVVMGLAPDTATAAKTKGTPKICLVSDALPGSDAHVLSRSFSMGRPHPALQLSGG  
 VCLAAACSI PG SIPNQILLKQTKMMPERLKF AHACGAIEATADVEMDSKSPVGVTVRSTS  
 LFR TARRLASAEAYYLSPSQ

(U.maydis 521 data base entry Um02807)

Figure 4S

SEQ ID NO: 19 - Nucleic acid sequence UM02806 (CFP1)

ATGTCGACAACAGAGAAACGAATCTCGCCTTCGCCCATGCACTCGCTGCTCGCCGGAACC  
 ATCGCCGGTGCGGTCTGAAGGGTTTCTTACTTACCCGACCGAGTTTGTCAAGACGCAGGCT  
 CAGCTTGCCCTCGAATGCAGCCGCAAGTGCCAGCAAGACATGTTCCGGTCCCCAAGAGTGTT  
 CCGGGGGCGGTTTCGACACATCTCGTATGGTGCAATTGCCAGCGCACAAAGCTCAACATCGCC  
 GCCGCCGCCGCCGCCACATCAGCACCATCAACCACTGTACCGATGCCAAAAGCAGGCGCC  
 TCGGCGATGCAGATTGTCCGGGATACTTGGAAGACTCGCGGGATCACTGGCTTCTTCAGC  
 GGCGCAGGTGCCATGGTTACGGGCAACTCTGCTAAAGCGGGTGTTTCGCTTCCTCACGTAC  
 GATACGATTCAAAACCTTCTGCGACCAAAGTCGATCGATGCAAACCAGAAGCTCGGAATG  
 GGCAGATCGATCCTTGCCGGCTTCTTGCTGGATCGGCCGAGGCGATGCTCGCGGTACAG  
 CCTTCCGAGGCGATCAAGACACGGATGATCCAGGATTCGCTGCAGCCTGCTCACATGCGC  
 AAGTACAAAGGCGCCATCGATGCGGTTCAAAAGATTGTGGGCGCTGAGGGCTTGGCTGGT  
 CTGTACAGAGGGTTGGGTGCCACGGTGCTGCGACAGGGAGCAAACCTCATCGGTGAGGTTG  
 ACATCCTACTCCATCTTGAAATCCGTACAAACGCAAGCAGGCTATGCCAAGTCCACAGCA  
 GCGACATTTCGCATCGGGAGCAGGCGCTGGCTTGATCACCGTCTACCTCACAATGCCGTTT  
 GACGTGGTCAAGACGCGAATGCAACAGTCACCCTCGACCACAGGAGCTACACAGAGCAAG  
 CCGAGCATCGTCTCGTGCGGTCTCGACATTGTCAAGAGGGAAGGAGTAAAGAGCCTCTGG  
 AAGGGTACCACCCACGATTGACCAGGTTGATTTTCAGCGGTGGCATCGCGTTTACTGCT  
 TACGAGACGGTTATTGGCTGGCTCAACCCGACCACAGTTGCTTGA

(U.maydis 521 data base entry Um02806)

Figure 4T

SEQ ID NO: 20 - Polypeptide sequence UM02806 (CFP1)

MSTTEKRISPSPMHSLLAGTIAGAVEGFLTYPTFEVKTQAQLASNAASASKTCSVPKSV  
 PGAVRHISYGALPAHKLNIAAAAAATSAPSTTVMPKAGASAMQIVRDTWKTRGITGFFS  
 GAGAMVTGNSAKAGVRFLTYDTIQNLLRPKSIDANQKLGMR SILAGFLAGSAEAMLAVT  
 PSEAIKTRMIQDSLQPAHMRKYKGAIDAVQKIVGAEGLAGLYRGLGATVLRQGANSSVRL  
 TSY SILKSVQTQAGYAKSTAATFASGAGAGLITVYLTMPFDVVKTRMQQSPSTTGATQSK  
 PSIVSCGLDIVKREGVKSLWKGTTPRLRLIFSGGIAFTAYETVIGWLNPTTVA

(U.maydis 521 data base entry Um02806)

Figure 4U

SEQ ID NO: 21 - Nucleic acid sequence UM02808 (regulator)

ATGAGGTGGAAACAAGTACACGCTGTACCTGGCAGTAACACCTCGGACACCTCGCGTCCG  
 TCGCAAGGCTCATCGGCTTCTCTACCGCTTGCGATTTTTGCCGTCACCGAAAGATTTCGT

16/27

TGCAATCGAGAACAGCCATGCGACAAGTGCCAGAAGCACGGTAGGGCCAACACGTGTCAC  
 TTTGAAGATGGTTCGAACCAAAAAGCGCCAACGCCCAACAATGCGGCTTGTGCGGCAGTA  
 CATCAAGACGAAAGCCGCGTTGGTTCAGCTCGGGGACGAGCTCGAGCTCGAGCTCGAGATC  
 GCCCGTGTGGTTCAGCTTTCCGCGACGCCAGGTTCGACACGGTAATGTACGCTTCAACCGC  
 GCTTCATCAGAGGCTGCTGGTCAAGTCAGCGTGGATGAAGCAAGTCGTGTTGAAGCGCAC  
 AATACGTCGGGCCAGGCTAAGGACGGCTCACACGCGTTTGCCAATCCACGGCTCAACGAG  
 ACTGCGTTCAGACATACCGCAGAGACGCACCATGCTGCTCCAGTAGCCACACGGACGTG  
 CGTGTCATCGTCTGTATCATGCTCGTCTGCTGAATGCTGCAGTGCAGGCTCACAGCTCGTCA  
 CCTCTGCCGACTTCGCCAACTTCGTGCGCGGTGCTAGCATCATTGCGGATCTGGGCGAT  
 CCGAGGCATTCAATCGGACATGCTGAACGGACGCGCGCCTCTGCGTCACCCAGCCAGCAA  
 GCTACACGTGGAGTCATGTCAAATCTGCGCGTCAACAATCGAGGCGAATCGCGATACCAC  
 GGTCCGACGAGTGCTTCGTTCAACGCGCCTGAAGCATTGAGCGAGGATTCTCGAGCTGGA  
 ATTGCGTCTGCGCAGGGCCGTCTCAGTGTGGATGGTGGCGGTGGCGCCAATCCAGCAGGT  
 GGTAGTCAAGAGCAGCGTGTCTCGCTGGATGCAGCCGTGCTCACCGACAGATTGACATCG  
 CTGGCTGCTCGCGAAAGACAGGAAGAGGTGATGAACGTTGTGCTGGCAGACTCGACTTT  
 GACGGTGTAGAGCCTGAGCTGGCGATCCACCTGCTCAACTTGCACTGGAATCGGCAGCAT  
 CACGCCTACCTGGTTCAGTACCGACCGCTGTTTCATGCGGGATATGGCGGCTCCTCCAGGC  
 AAAGCAAAGTACTTTTCCAAGCTACTCTTCCACGCTATCCTGTTGCGTGCCGCCAAGTTT  
 TCAGACCGCATCAATCAGGTCAGGTCAGACGTCAATGATCCCTCGACGGCTGGCCAACAG  
 TTTCTCGAACGCGTCCAGCAACTGTTACCAGAAGCGTTGATCAAGAGCCGCATCACGACG  
 GTGCAAGCGTTGCTGTTACTCGCATGCACGCACTATGCCCGTGGTAGCGAGAGTGCAGCT  
 TGGCTACATTCCGGATTGGCTTTCCGCATGGCGATCGATCTGGGGATGAACGAAGATAGT  
 ATCGAGCTGGTTGCATCGGGCAACATGGCACTCGAGGAGCTCGAGCTTCGTGACGTGTA  
 GTATGGGCGCGTTTTGTGATTGACAAGATCCACTCGCTCTACCAAGGTGCGCAGGCTTCG  
 ATTGAACGGCGCAACCTGTATGTGCCATCGAGTTCACGACCATTTCGAAGAGACCGAG  
 TTCTGGACGCCTGTTGCGTTTACCGACCTCTGAGCGAGGCAGAGGCGCAAAGCAAAGCT  
 ACAGCCAGAGGCACCAACGCCGCCACTCCGAGTGGCTACACGGGACCGATCTACAGTGTT  
 TCTACGTTTGCCGAGCTGTGCGAGGCTGACGGTGATTATGGAAAAGATCATCCAGATCTTC  
 TATTTCGATCGATAGCGGTGCGCGATCCGAAAGAGCACAAACGGAACACTTGATTGAAATG  
 CGCTCCGAGCTCGCCGCTTGGCGCACGTGCTTGCCCTGCACATCTGCGTATCGACTCGAGC  
 TCGCCAGCAGGTAGAAGTGGCGGCCAACGCTCTCGCACTTGCCCCGCCAATCTGATCTCG  
 TTGCACGCACTCTATCATGCGTTGACGATCCTGCTGAATCGACCATTCCTACCACATGGA  
 CATCTTCGAGCGACGATGCAGCTTGCGGCAGATCGTCATGGAAAGCATGCGTGGATGCA  
 GCGTCGAGCATCTCGAGCCTCATGAATCTGTATCGTCAGACGGTGAGCATGAAGGGAGCG  
 CCGCAGCTCATCTCATACATCAACTTTTGTGCTGCAGGAATTCACGTTGCTGTAGCAGCA  
 CAGCTGCAACAGTCCAACACGTCGAGCGAAGGTGCTAATCTGGCAGCCGCATCCAGTCGC  
 AGATACGAGCTGAAAGCGCTCCGACAATGCCTGCAGGATTTGAGGAGAATCAAGATCCC  
 AACCCGGGCGTCGCTAAAGCCAAGAGTGTGATCCGCAACATGGCCGAGAGAGCTGGCATT  
 CTCAACCTGCTTCAAGACGACTCGGGAGAAGGTGGATTGTGCGCGATCAACCAAGATGAC  
 AGTCCACCGCTGATGAATCTCGTCGGCCACAGTGGTAGCCAAGCGAATCGAGGGGGTGCT  
 GCAAATCGCGTCGAGCTTGTGGAGCAAGCGCACTTGACGCGCATGCGAGCGCGCTGTCC  
 ATGACAACACCGTCTGCTCGACTTCGATCGCCGAGCAGGAGGAGCAACAGCCGCATCA  
 GCAGCGTTGAATCGAGTAGACGCAGCGCAGAACTCAACTTCTGCACTCACACCGCCCTAC  
 GAGATCGACTTTTGAAAGCATTCTCGCCTCGTTTGATCACTTTGAGTCCCGCAGCCGGTCTC  
 GGTCGAGACGGCCTATTTCGCAACAGGCACAGCGGATCCCAACACCAACTCGGACATTCTC  
 TTTGGCTTTCTTCGTGACTATGATGCACCCCTCTGCCCTCTCGCTATGA  
 (U.maydis 521 data base entry Um02808)

Figure 4V

SEQ ID NO: 22 - Polypeptide sequence UM02808 (regulator)

MRWKQVHAVPGSNTSDTSRPSQGSSAFSTACDFCRHRKIRCNRQPCDKCQKHGRANTCH  
 FEDGRTKKRQRPNNAAAVHQDESRVQLGDELELELEIARVQLSATPGRHGNVRFNR  
 ASSEAAGQVSVDASRVEAHNTSGQAKDGSHAFANPRLNETAFRHTAETHHAAPVAHTDV  
 RASSVSCSSLNAAVRPHSSSPLPTSPTSSAVASIIADLGDPRH SIGHAERTRASASPSQQ  
 ATRGVMSNLRVNNRGESRYHGPTSASFNAPEALSEDSRAGIASAQGRLSVDGGGGANPAG

17/27

GSQEQRVSLDAAVLTDRLTSLAARERQEEVMNVVAGRLDFDGVEPELAIHLLNLHWNQRH  
HAYLVTYRPLFMRDMAAPPGKAKYFSKLLFHAILFGAAKFSDRINQVRSDVNDPSTAGQQ  
FLERVQQLLPEALIKSRITTVQALLLLACTHYARGSESAAWLHSGLAFRMAIDLGMNEDS  
IELVASGNMALEEELELRRRVVWAAFVIDKIHSLYQGRQASIERRNLYVP IEFHDHFEETE  
FWTPVAFHRPLSEAEAQSKATARGTNAATPSGYTGPIYSVSTFAELCRLTVIMEKIIQIF  
YSIDSGARSERAQTEHLIEMRSELAAWRTSLPAHLRIDSSSPAGRSGGQRSRTCPPNLIS  
LHALYHALTILLNRPFLPHGHLRSDDAACGRSSWKACVDAASSISSLMNLYRQTVSMKGA  
PQLISYINFCAAGIHVRVAAQLQQSNTSSEGANLAAASSRRYELKALRQCLQDFEENQDP  
NPGVAKAKSVIRNMAERAGILNLLQDDSGEGGLSPINQDDSPPLMNLVGHSGSQANRGGA  
ANRVELVGASALARHASALSMTTPSSSTSIAAAGGATAASAALNRVDAAQNSTSAITPPY  
EIDFESILASFDHFESAAGPGRDGLFATGTADPNTNSDILFGFLRDYDAPSALSL  
(U.maydis 521 data base entry Um02808)

18/27

Figure 5A

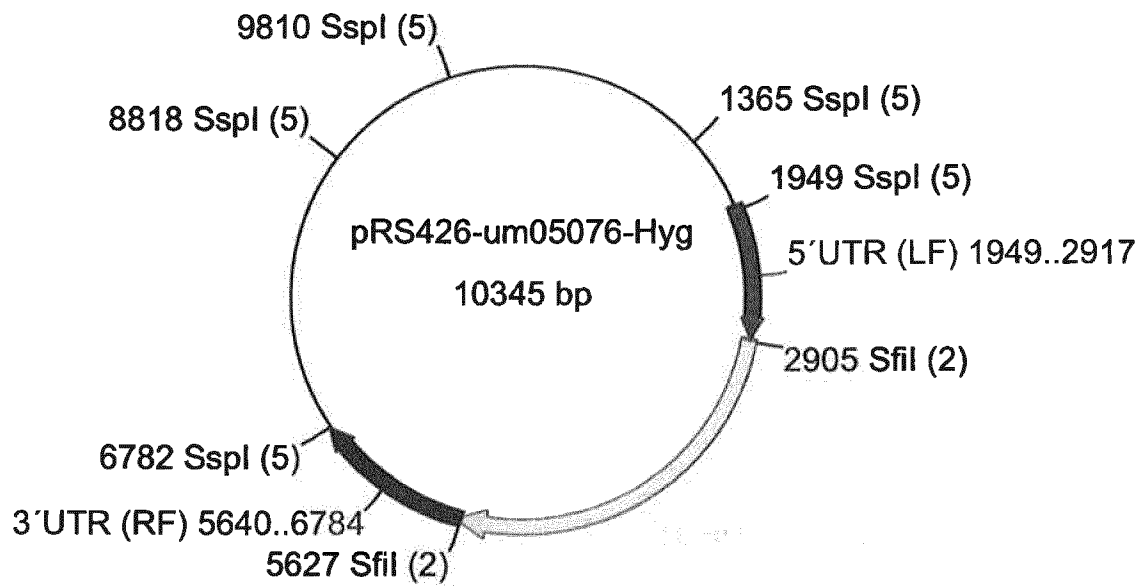
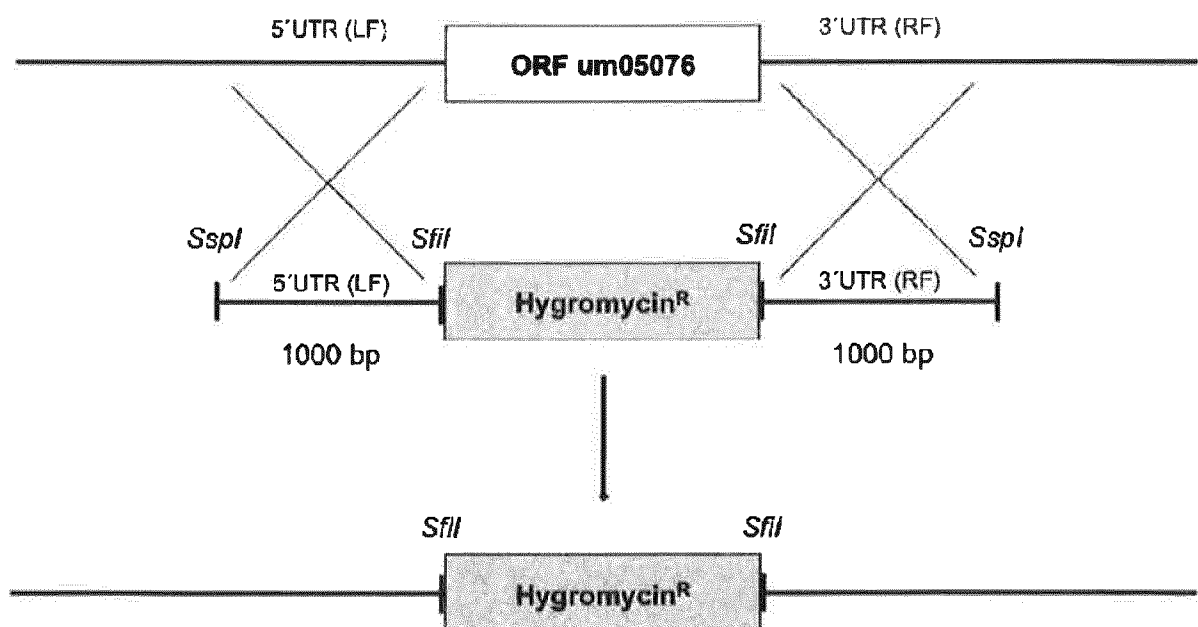


Figure 5B



19/27

Figure 6A

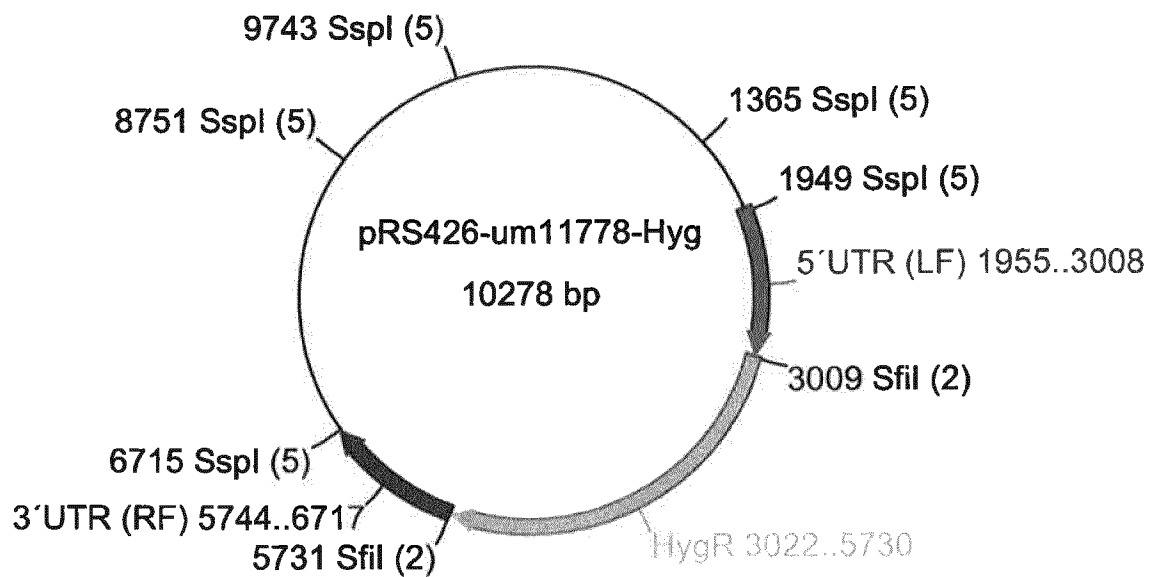
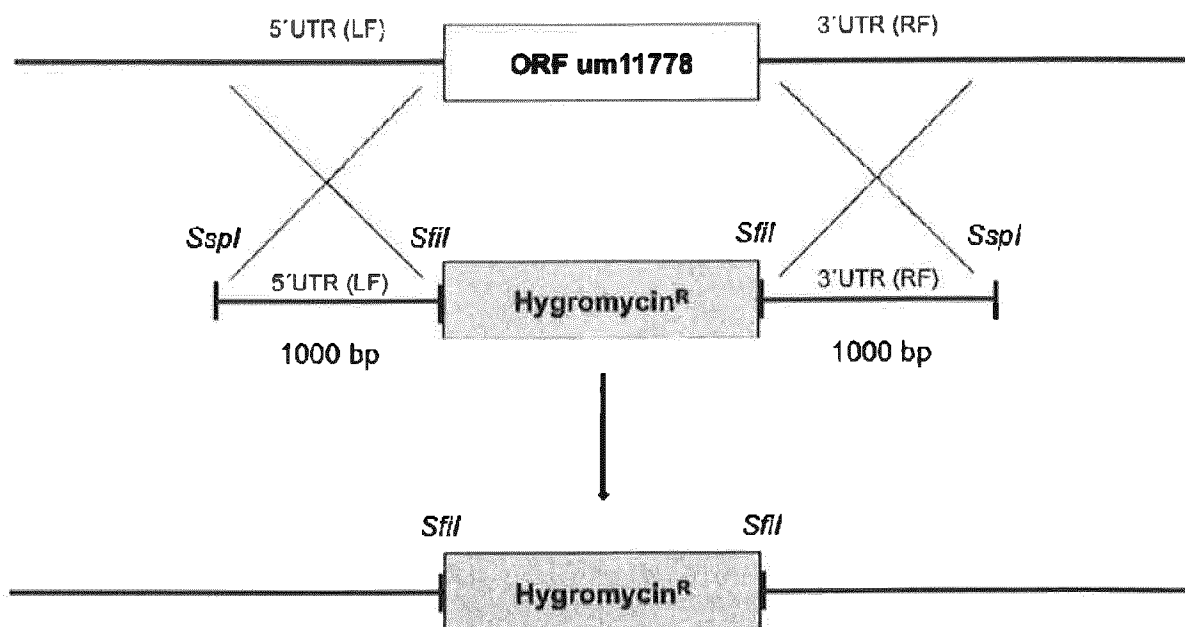
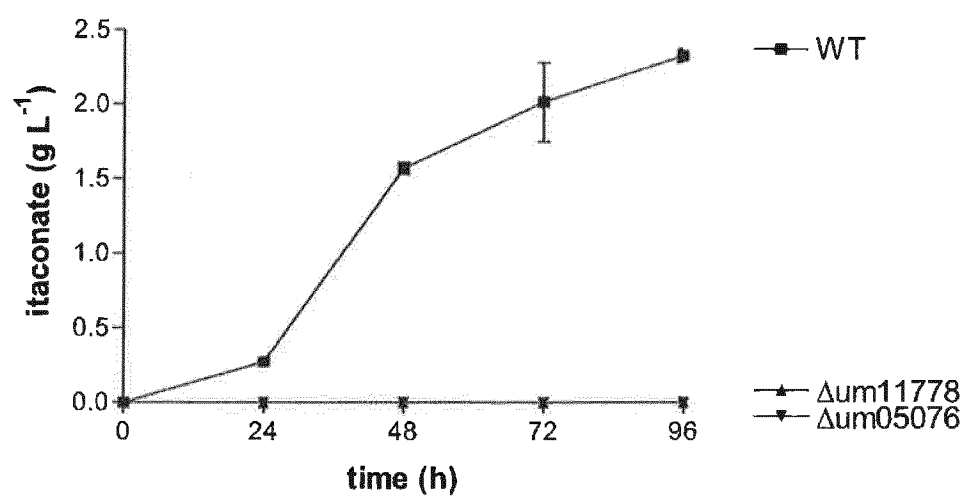


Figure 6B



20/27

Figure 7



21/27

Figure 8

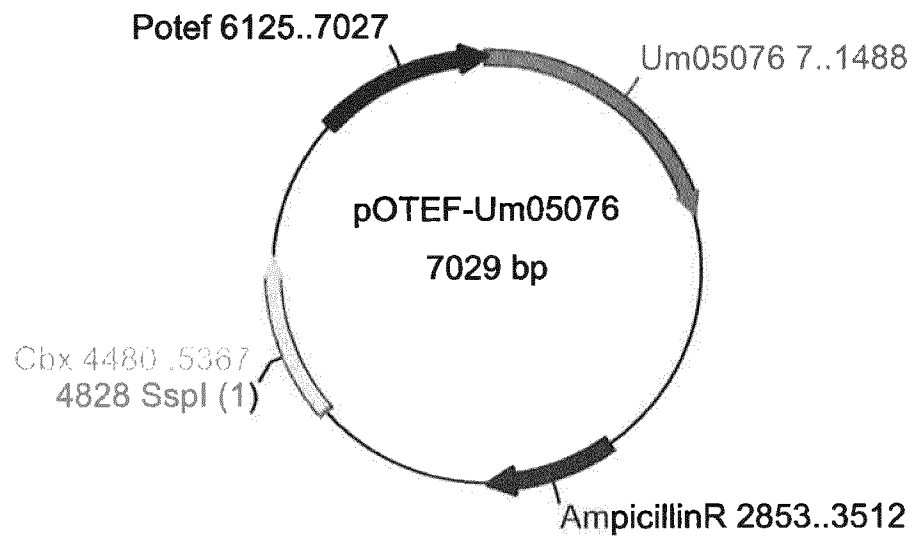
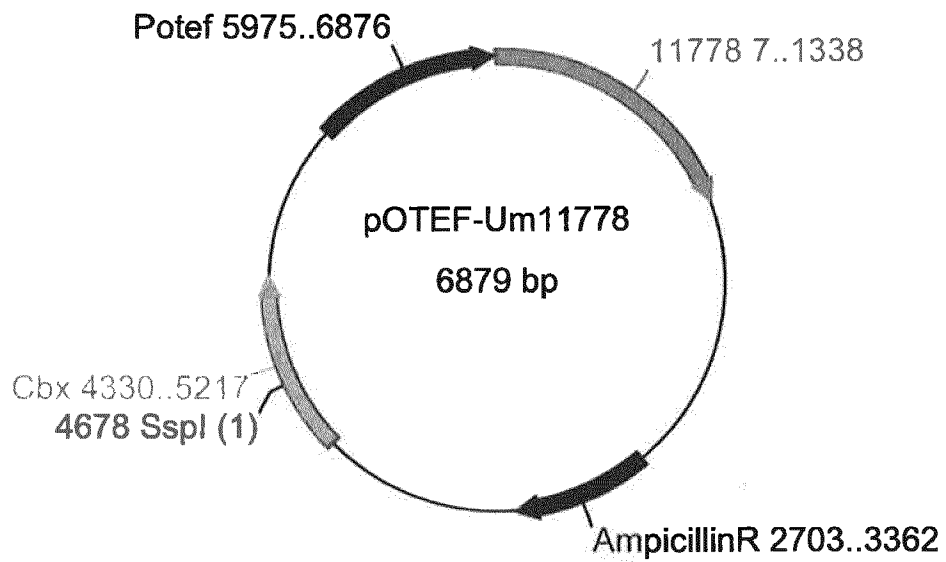
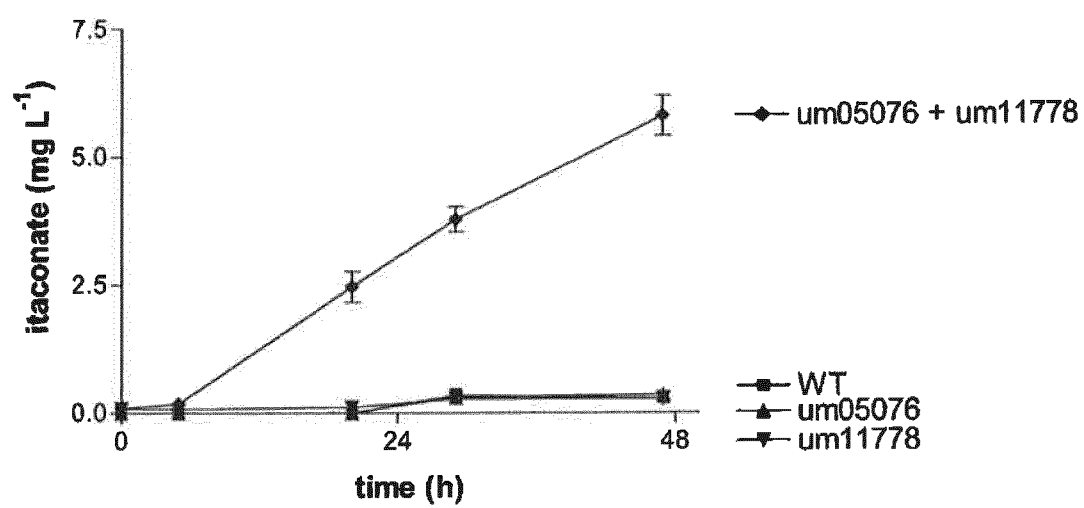


Figure 9



22/27

Figure 10



23/27

Figure 11

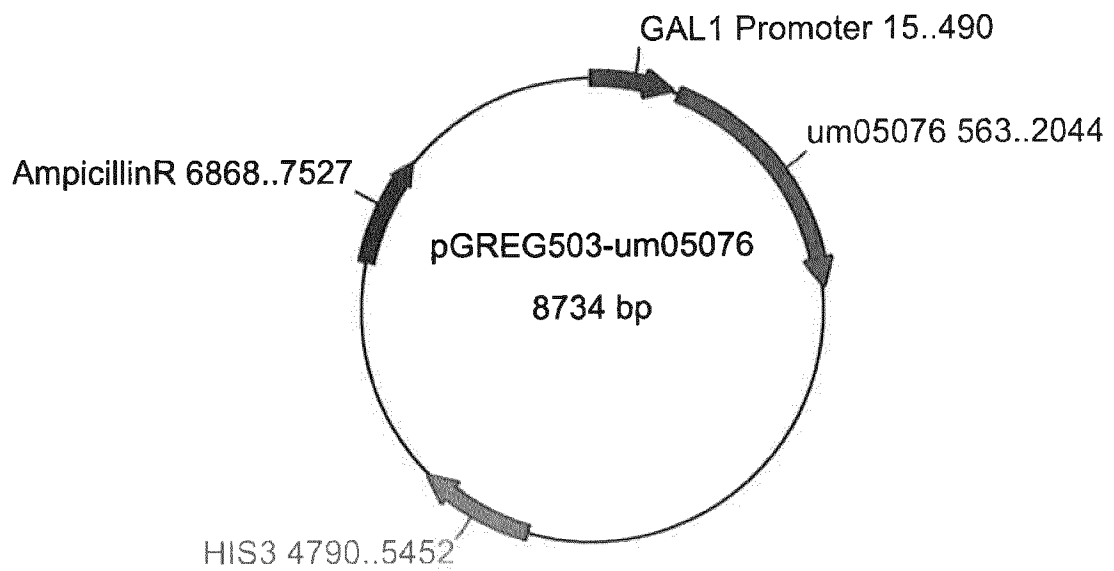
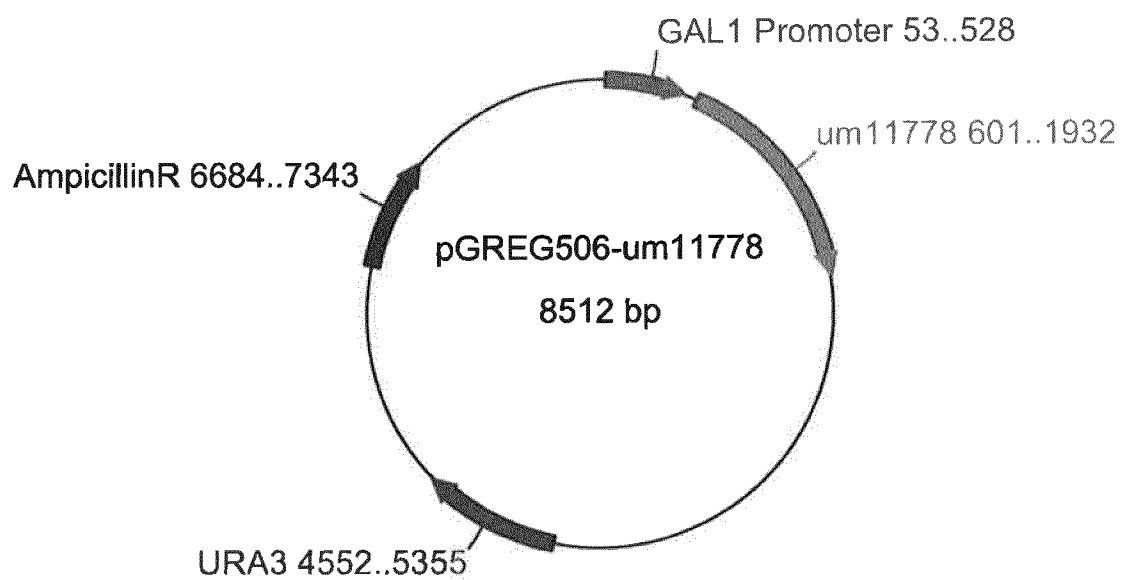


Figure 12



24/27

Figure 13A

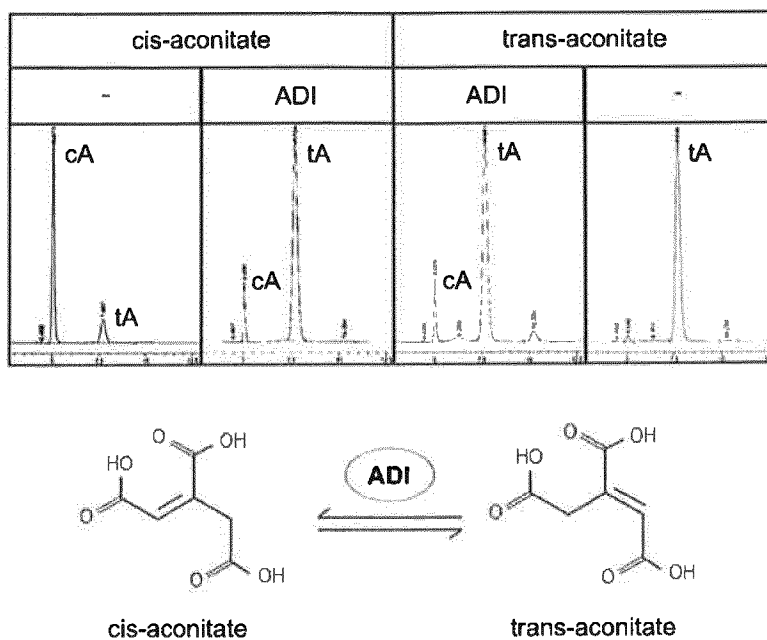
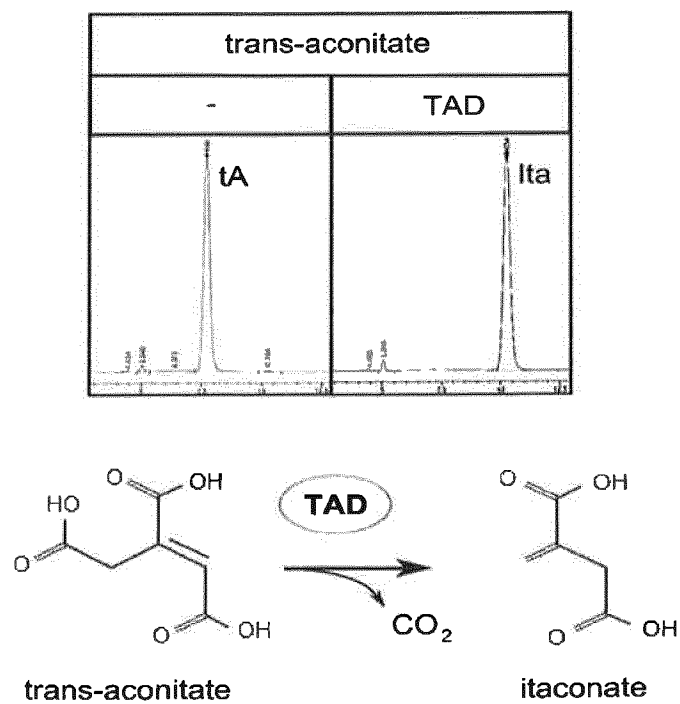


Figure 13B



25/27

Figure 14

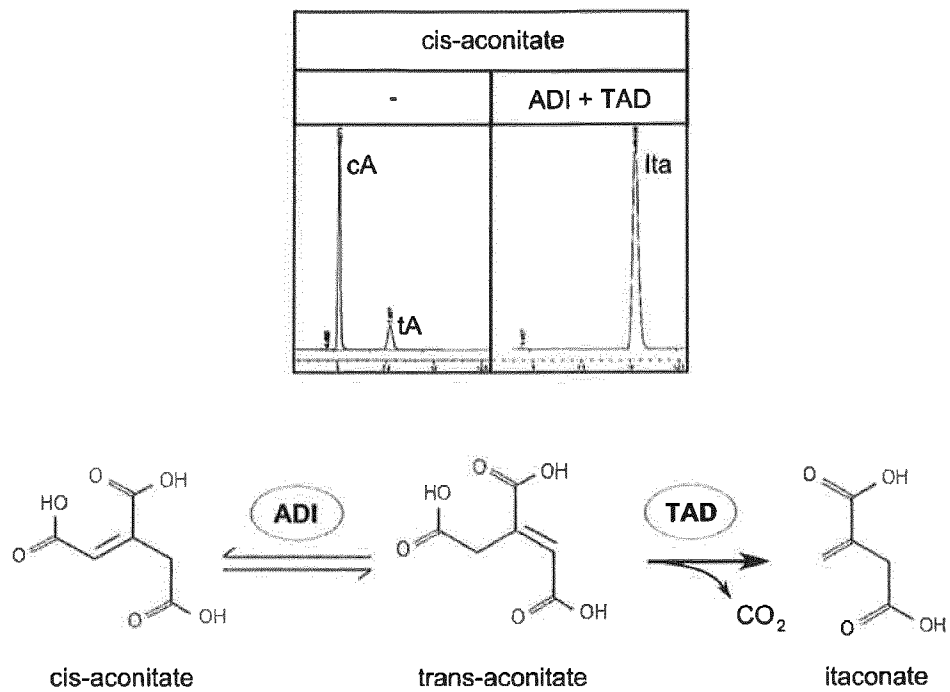
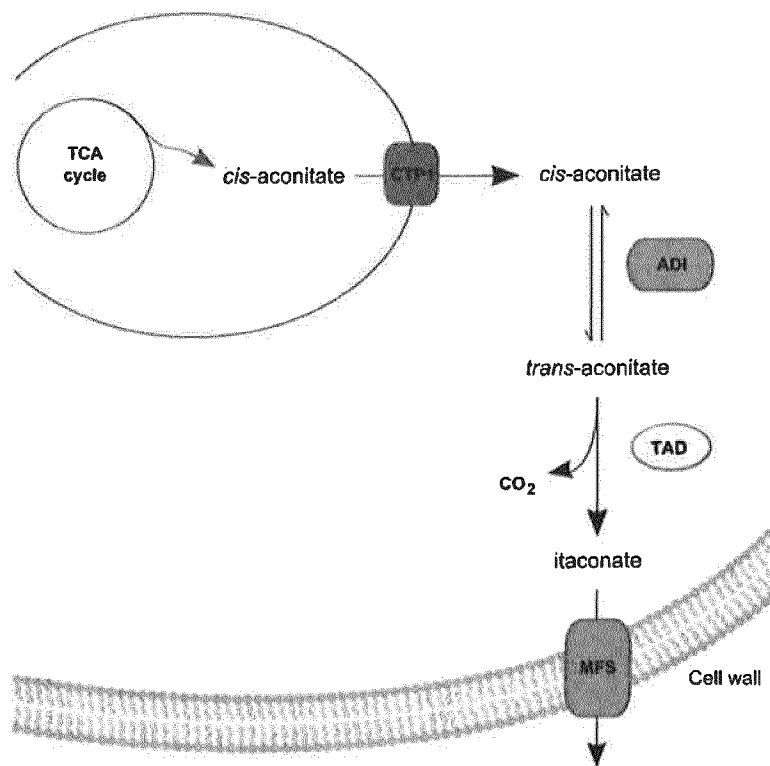


Figure 15



26/27  
Figure 16

Figure 16 A

A

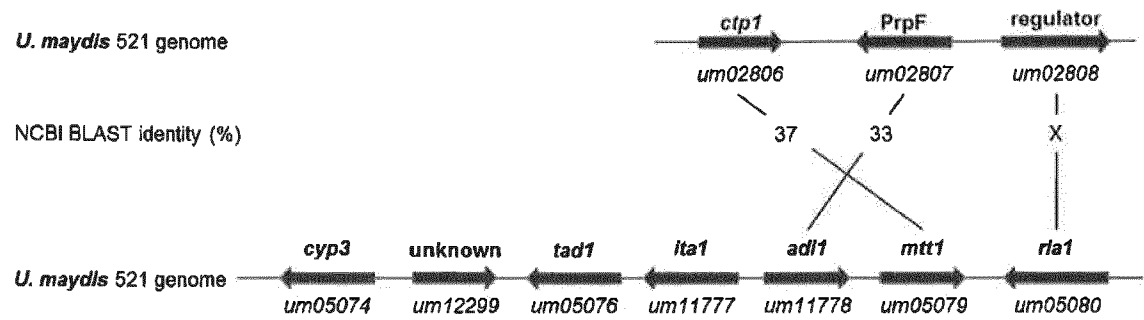
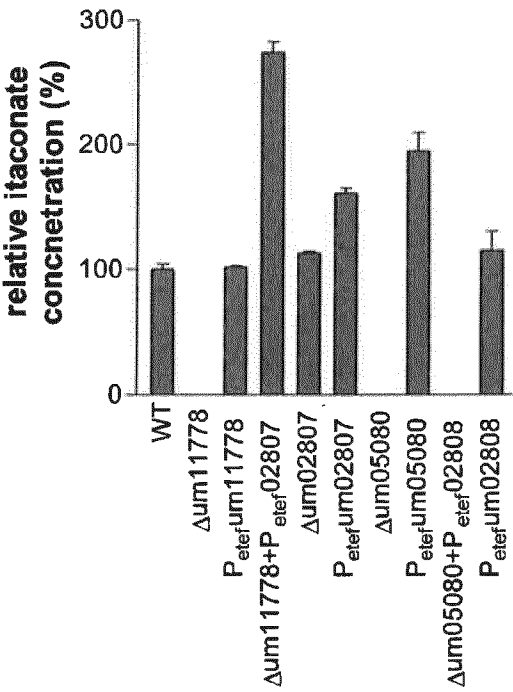


Figure 16B

B



27/27

Figure 17

Pairwise sequence alignment between the polynucleotide sequences of UM11778 (SEQ ID NO: 1) and UM02807 (SEQ ID NO: 17).

```

um028075 1  LRTVIMRAGTSKGLFIKASDLPPSSRAEWQNILPSIMGSPDPFGRQLNGLGGGTSTTSKIA 110
+ T I RAGTS+GL+ ASDLP+ +E L SIMGS P Q++G+GGG S TSK+A
um11778 5  IDTTIYRAGTSRGLYFLASDLPAEFSEDAALISIMSGHPL--QIDGMGGGNSLTSKVA 62

um02807 111 VVSQSSAPHIADVDYLFIQCPIEGDKLDFTGNCNLSGVGPFPAFEEGLI--PASVLAP 167
+VS S+ DVDYLF Q I +D NCGN++SGV FA E GL+ P+
um11778 63  IVSASTQRSEFDVDYLFQCVGITERFVDTPNCGNLMSGVAAFAIERGLVQPHPSDTTCL 122

um02807 168 LAAFAKTNGKGRQDKVALTLRLCLNNQLIRSTFLVRNGKPVFEGDVVIDGVAGTGSPIQL 227
+ F LN+ Q V NG+ V + D+ + + + L
um11778 123 VRIFN-----LNSRQASELVIPVYNGR-VHYDDIDDMHMQRPSARVGL 164

um02807 228 DFLDPAGSMCTSLCPTGKPLDLLHIHGEDLPIEVSCVDAANPFVVFVRLSDIDETLRGNEP 287
FLD GS L PTG D + ++VS +D+A P VF+R D+ T G+E
um11778 165 RFLDTVGSCCTGKLLPTGNASDWID-----GLKVSIIIDSAVPVVFIRQHDVGIT--GSEA 216

um02807 288 ASVLERHSA--RVELIRQAAAVVMGLAPDTATAAKTKGTPKICLVSDALPGSDAHVLS- 343
+ L ++A R+E +R A MGI+ + + PK+ L+ PG++ +
um11778 217 PATLNANTALLDRLEVRLEAGRRMGLGDVSGSV-----VPKLSLIG---PGTETTTFTA 268

um02807 344 RSFSMGRPHPALQLSGGVCLAAACSI PGSIPNQIL---LKQTKMMPERLKFACGAIE- 399
R F+ H A ++G +C A A I GS+ +IL R+ H G +E
um11778 269 RYFTPKACHNAHAVTGAICTAGAAIDGSVVCEILSSRASACSASQRRISIEHPSGVLEV 328

um02807 400 ATADVEMDSKSPVGVTVRSTSLFRTARRLASAEAYLSPSQ 440
E ++S V V V + R+ +A A YY +P +
um11778 329 GLVPPENAAQSLVDVAV----VERSIALIAHARVYYTTPDR 365

```

Select for  
downloading or  
viewing reports



Select seq

| Description | Max score | Total score | Query cover | E value | Ident | Accession |
|-------------|-----------|-------------|-------------|---------|-------|-----------|
|-------------|-----------|-------------|-------------|---------|-------|-----------|

um11778

155

155

88%

3e-46

33%

Query\_29917

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments

| Description | Max score | Total score | Query cover | E value | Ident | Accession   |
|-------------|-----------|-------------|-------------|---------|-------|-------------|
| um11778     | 155       | 155         | 88%         | 3e-46   | 33%   | Query_29917 |

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2015/055976

| <b>A. CLASSIFICATION OF SUBJECT MATTER</b><br>INV. C12N15/52 C12N1/19 C12N9/90 C12N9/88 C12N9/02<br>C07K14/37 C12P7/44<br>ADD.<br>According to International Patent Classification (IPC) or to both national classification and IPC                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                                     |                       |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|-----------|------------------------------------------------------------------------------------|-----------------------|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| <b>B. FIELDS SEARCHED</b><br>Minimum documentation searched (classification system followed by classification symbols)<br>C12P C12N C07K<br>Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched<br>Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br>EPO-Internal, BIOSIS, Sequence Search, FSTA, WPI Data                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                                                                                                                                                                                                                                                                                                                     |                       |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 10%; padding: 5px;">Category*</th> <th style="width: 70%; padding: 5px;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="width: 20%; padding: 5px;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td style="text-align: center; vertical-align: top; padding: 10px;">X</td> <td style="padding: 10px;">           KAEMPER JOERG ET AL: "Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis", NATURE, vol. 444, no. 7115, 2 November 2006 (2006-11-02), pages 97-101, XP002579269, ISSN: 0028-0836, DOI: 10.1038/NATURE05248 the whole document<br/> <div style="text-align: center;">-----<br/>-/--</div> </td> <td style="text-align: center; vertical-align: top; padding: 10px;">26,27, 31-33</td> </tr> </tbody> </table>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |                                                                                                                                                                                                                                                                                                                     |                       | Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. | X | KAEMPER JOERG ET AL: "Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis", NATURE, vol. 444, no. 7115, 2 November 2006 (2006-11-02), pages 97-101, XP002579269, ISSN: 0028-0836, DOI: 10.1038/NATURE05248 the whole document<br><div style="text-align: center;">-----<br/>-/--</div> | 26,27, 31-33 |
| Category*                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                  | Relevant to claim No. |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |
| X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | KAEMPER JOERG ET AL: "Insights from the genome of the biotrophic fungal plant pathogen Ustilago maydis", NATURE, vol. 444, no. 7115, 2 November 2006 (2006-11-02), pages 97-101, XP002579269, ISSN: 0028-0836, DOI: 10.1038/NATURE05248 the whole document<br><div style="text-align: center;">-----<br/>-/--</div> | 26,27, 31-33          |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |
| <div style="display: flex; justify-content: space-between;"> <div> <input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C.         </div> <div> <input checked="" type="checkbox"/> See patent family annex.         </div> </div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |                                                                                                                                                                                                                                                                                                                     |                       |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |
| <div style="display: flex;"> <div style="flex: 1;"> <p>* Special categories of cited documents :</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="flex: 1;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</p> <p>"&amp;" document member of the same patent family</p> </div> </div> |                                                                                                                                                                                                                                                                                                                     |                       |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |
| Date of the actual completion of the international search<br><br><div style="text-align: center; font-size: 1.2em;">30 June 2015</div>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Date of mailing of the international search report<br><br><div style="text-align: center; font-size: 1.2em;">10/07/2015</div>                                                                                                                                                                                       |                       |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |
| Name and mailing address of the ISA/<br>European Patent Office, P.B. 5818 Patentlaan 2<br>NL - 2280 HV Rijswijk<br>Tel. (+31-70) 340-2040,<br>Fax: (+31-70) 340-3016                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               | Authorized officer<br><br><div style="text-align: center; font-size: 1.2em;">van de Kamp, Mart</div>                                                                                                                                                                                                                |                       |           |                                                                                    |                       |   |                                                                                                                                                                                                                                                                                                                     |              |

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/055976

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                                                         |                       |
|------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                      | Relevant to claim No. |
| X                                                    | <p>DATABASE UniProt [Online]</p> <p>19 July 2005 (2005-07-19),<br/> "SubName: Full=Uncharacterized protein<br/> {ECO:0000313 EMBL:EAK82869.1}";",<br/> XP055158332,<br/> retrieved from EBI accession no.<br/> UNIPROT:Q4P485<br/> Database accession no. Q4P485<br/> sequence</p> <p>-----</p>                                         | 26,31-33              |
| X                                                    | <p>DATABASE UniProt [Online]</p> <p>19 July 2005 (2005-07-19),<br/> "SubName: Full=Uncharacterized protein<br/> {ECO:0000313 EMBL:EAK82868.1}";",<br/> XP055158337,<br/> retrieved from EBI accession no.<br/> UNIPROT:Q4P486<br/> Database accession no. Q4P486<br/> sequence</p> <p>-----</p>                                         | 26,31-33              |
| X                                                    | <p>DATABASE UniProt [Online]</p> <p>19 July 2005 (2005-07-19),<br/> "SubName: Full=Uncharacterized protein<br/> {ECO:0000313 EMBL:EAK82870.1}";",<br/> XP055158338,<br/> retrieved from EBI accession no.<br/> UNIPROT:Q4P484<br/> Database accession no. Q4P484<br/> sequence</p> <p>-----</p>                                         | 26,31-33              |
| X                                                    | <p>DATABASE UniProt [Online]</p> <p>19 July 2005 (2005-07-19),<br/> "SubName: Full=Uncharacterized protein<br/> {ECO:0000313 EMBL:EAK82867.1}";",<br/> XP055158334,<br/> retrieved from EBI accession no.<br/> UNIPROT:Q4P487<br/> Database accession no. Q4P487<br/> sequence</p> <p>-----</p>                                         | 26,31-33              |
| X                                                    | <p>DATABASE UniProt [Online]</p> <p>19 July 2005 (2005-07-19),<br/> "SubName: Full=Uncharacterized protein<br/> {ECO:0000313 EMBL:EAK82871.1}";",<br/> XP055158340,<br/> retrieved from EBI accession no.<br/> UNIPROT:Q4P483<br/> Database accession no. Q4P483<br/> sequence</p> <p>-----</p> <p style="text-align: center;">-/--</p> | 26,31-33              |

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/055976

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                                                   | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | <p>DATABASE UniProt [Online]</p> <p>19 July 2005 (2005-07-19),<br/> "SubName: Full=Uncharacterized protein<br/> {ECO:0000313 EMBL:EAK82865.1}";",<br/> XP055158418,<br/> retrieved from EBI accession no.<br/> UNIPROT:Q4P489<br/> Database accession no. Q4P489<br/> sequence</p> <p>-----</p>                                                                                                      | 27                    |
| A         | <p>GRAEME S. GARVEY ET AL: "The<br/> three-dimensional crystal structure of the<br/> PrpF protein of Shewanella oneidensis<br/> complexed with trans -aconitate: Insights<br/> into its biological function",<br/> PROTEIN SCIENCE,<br/> vol. 16, no. 7, 1 July 2007 (2007-07-01),<br/> pages 1274-1284, XP055158309,<br/> ISSN: 0961-8368, DOI: 10.1110/ps.072801907<br/> abstract</p> <p>-----</p> | 26,31-33              |
| A         | <p>KLEMENT TOBIAS ET AL: "Itaconic acid - A<br/> biotechnological process in change",<br/> BIORESOURCE TECHNOLOGY,<br/> vol. 135, 8 December 2012 (2012-12-08),<br/> pages 422-431, XP028579661,<br/> ISSN: 0960-8524, DOI:<br/> 10.1016/J.BIORTECH.2012.11.141<br/> the whole document<br/> 3. Biosynthesis of IA<br/> table 2<br/> 5.4.4. New organisms</p> <p>-----</p>                           | 1,23,<br>26-32        |
| A         | <p>WO 2009/110796 A1 (TNO [NL]; VAN DER WERF<br/> MARIA JOHANNA [NL]; PUNT PETER JAN [NL])<br/> 11 September 2009 (2009-09-11)<br/> cited in the application<br/> the whole document</p> <p>-----</p>                                                                                                                                                                                                | 1,23,<br>26-32        |
| A         | <p>WO 2009/104958 A1 (TNO [NL]; JORE JOHANNES<br/> PETRUS MARIA [NL]; PUNT PETER JAN [NL];<br/> VAN DE) 27 August 2009 (2009-08-27)<br/> cited in the application<br/> the whole document</p> <p>-----</p>                                                                                                                                                                                           | 1,23,<br>26-32        |
|           | -----<br>-/--                                                                                                                                                                                                                                                                                                                                                                                        |                       |

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/055976

| C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                            |                       |
|------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category*                                            | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                         | Relevant to claim No. |
| A                                                    | <p>NICOLE MAASSEN ET AL: "Influence of carbon and nitrogen concentration on itaconic acid production by the smut fungus Ustilago maydis", ENGINEERING IN LIFE SCIENCES, vol. 14, no. 2, 6 January 2014 (2014-01-06), pages 129-134, XP055158158, ISSN: 1618-0240, DOI: 10.1002/elsc.201300043 abstract</p> | 1,23, 26-32           |
| A                                                    | <p>PANAKOVA M ET AL: "High-level formation of itaconic acid by the fungus Ustilago maydis", NEW BIOTECHNOLOGY, vol. 25, 1 September 2009 (2009-09-01), page S270, XP026461617, ISSN: 1871-6784, DOI: 10.1016/J.NBT.2009.06.605 the whole document</p>                                                      | 1,23, 26-32           |
| A                                                    | <p>JEONG WOOK LEE ET AL: "Microbial production of building block chemicals and polymers", CURRENT OPINION IN BIOTECHNOLOGY, vol. 22, no. 6, 17 March 2011 (2011-03-17), pages 758-767, XP028397462, ISSN: 0958-1669, DOI: 10.1016/J.COPBIO.2011.02.011 the whole document</p>                              | 1,29-33               |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2015/055976

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|-------------------------------------------|---------------------|----------------------------|---------------------|
| WO 2009110796 A1                          | 11-09-2009          | EP 2262826 A1              | 22-12-2010          |
|                                           |                     | US 2011070616 A1           | 24-03-2011          |
|                                           |                     | US 2014199754 A1           | 17-07-2014          |
|                                           |                     | WO 2009110796 A1           | 11-09-2009          |
| -----                                     |                     |                            |                     |
| WO 2009104958 A1                          | 27-08-2009          | EP 2262827 A1              | 22-12-2010          |
|                                           |                     | US 2011124066 A1           | 26-05-2011          |
|                                           |                     | US 2014193885 A1           | 10-07-2014          |
|                                           |                     | WO 2009104958 A1           | 27-08-2009          |
| -----                                     |                     |                            |                     |