



(86) Date de dépôt PCT/PCT Filing Date: 2011/11/18
(87) Date publication PCT/PCT Publication Date: 2012/05/24
(85) Entrée phase nationale/National Entry: 2013/10/22
(86) N° demande PCT/PCT Application No.: NL 2011/050787
(87) N° publication PCT/PCT Publication No.: 2012/067510
(30) Priorités/Priorities: 2010/11/18 (US61/415,054);
2010/11/18 (EP10191736.7); 2011/04/05 (US61/471,836)

(51) Cl.Int./Int.Cl. *C12P 7/06* (2006.01),
C12N 1/32 (2006.01), *C12N 9/02* (2006.01),
C12N 9/04 (2006.01), *C12N 9/10* (2006.01),
C12N 9/12 (2006.01), *C12P 7/40* (2006.01)
(71) Demandeur/Applicant:
DSM IP ASSETS B.V., NL
(72) Inventeurs/Inventors:
DE BONT, JOHANNES ADRIANUS MARIA, NL;
TEUNISSEN, ALOYSIUS WILHELMUS RUDOLPHUS
HUBERTUS, NL
(74) Agent: SMART & BIGGAR

(54) Titre : SOUCHES DE LEVURES MODIFIEES POUR PRODUIRE DE L'ETHANOL A PARTIR DE GLYCEROL
(54) Title: YEAST STRAINS ENGINEERED TO PRODUCE ETHANOL FROM GLYCEROL

(57) **Abrégé/Abstract:**

The present invention relates to genetically modified yeast cells comprising exogenous genes coding for pyruvate formate lyase and acetaldehyde dehydrogenase activities. The yeast cells further comprises genetic modifications that improve glycerol utilization such as modifications that increases NAD^+ -linked glycerol dehydrogenase activity, and preferably one or more of dihydroxyacetone kinase activity and transport of glycerol into the cell. The yeast cell further preferably comprises a functional exogenous xylose isomerase gene and/or functional exogenous genes which confer to the cell the ability to convert L-arabinose into D-xylulose 5-phosphate and they may comprise a genetic modification that increase acetyl-CoA synthetase activity. The process is further characterised in that glycerol is present in or fed into the culture medium, whereby the modified yeast cell ferments the hexoses, pentoses, acetic acid and glycerol to ethanol. The invention further relates to processes for producing a fermentation product and formate from carbon sources comprising glycerol and at least one of a hexose and a pentose, wherein the genetically modified yeast cells are used to ferment the carbon sources to the fermentation product and formate. A preferred fermentation product is ethanol.

- (51) **International Patent Classification:**

<i>C12P 7/06</i> (2006.01)	<i>C12N 9/04</i> (2006.01)
<i>C12N 1/32</i> (2006.01)	<i>C12N 9/12</i> (2006.01)
<i>C12N 9/10</i> (2006.01)	<i>C12P 7/40</i> (2006.01)
<i>C12N 9/02</i> (2006.01)	

(21) **International Application Number:** PCT/NL2011/050787

(22) **International Filing Date:** 18 November 2011 (18.11.2011)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**

10191736.7	18 November 2010 (18.11.2010)	EP
61/415,054	18 November 2010 (18.11.2010)	US
61/471,836	5 April 2011 (05.04.2011)	US

(71) **Applicant** (*for all designated States except US*): **C5 YEAST COMPANY B.V.** [NL/NL]; Lelyweg 29, NL-4612 PS Bergen op Zoom (NL).

(72) **Inventors; and**

(75) **Inventors/Applicants** (*for US only*): **DE BONT, Johannes Adrianus Maria** [NL/NL]; Lawickse Allee 68, NL-6707 AK Wageningen (NL). **TEUNISSEN, Aloysius Wilhelmus Rudolphus Hubertus** [NL/NL]; Tjaskerlaan 28, NL-3052 HP Rotterdam (NL).

(74) **Agent:** **HUTTER, Hans**; J.W. Frisolaan 13, NL-2517 JS The Hague (NL).
- (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, 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:** Yeast strains engineered to produce ethanol from glycerol
- (57) **Abstract:** The present invention relates to genetically modified yeast cells comprising exogenous genes coding for pyruvate formate lyase and acetaldehyde dehydrogenase activities. The yeast cells further comprises genetic modifications that improve glycerol utilization such as modifications that increases NAD⁺ -linked glycerol dehydrogenase activity, and preferably one or more of dihydroxyacetone kinase activity and transport of glycerol into the cell. The yeast cell further preferably comprises a functional exogenous xylose isomerase gene and/or functional exogenous genes which confer to the cell the ability to convert L-arabinose into D-xylulose 5-phosphate and they may comprise a genetic modification that increase acetyl-CoA synthetase activity. The process is further characterised in that glycerol is present in or fed into the culture medium, whereby the modified yeast cell ferments the hexoses, pentoses, acetic acid and glycerol to ethanol. The invention further relates to processes for producing a fermentation product and formate from carbon sources comprising glycerol and at least one of a hexose and a pentose, wherein the genetically modified yeast cells are used to ferment the carbon sources to the fermentation product and formate. A preferred fermentation product is ethanol.

Yeast strains engineered to produce ethanol from glycerolField of the invention

The present invention relates to metabolic engineering in microorganisms such as yeast. In particular the invention relates to yeast strains that have been engineered to produce ethanol from glycerol. These strains have retained their natural ability to produce ethanol from hexoses (glucose, fructose, galactose, etc) and comprise an engineered ability to produce ethanol from pentoses like xylose. The invention further relates to the processes wherein the engineered strains of the invention produce ethanol from glycerol, either as main fermentation feedstock, or concomitantly with one or more of hexoses and pentoses.

Background of the invention

Both the formation and the degradation of glycerol in the yeast *Saccharomyces cerevisiae* are important processes that have been studied for more than a century. During the last decades, the emphasis has been on (1) suppression of the formation of glycerol during ethanol production. In the beginning of the previous century, (2) the optimization of glycerol production from sugars has been studied. Both of these aspects have profound economic impacts. More recently, a fully new aspect has been introduced in this glycerol field. It deals with (3) the anaerobic fermentation of glycerol into ethanol.

(1) Suppression of glycerol formation during ethanol production.

Bioethanol is produced by *Saccharomyces cerevisiae* from a range of substrates including lignocellulosic hydrolysates of non-food feedstocks (e.g. energy crops and agricultural residues). One problem of yeast-based ethanol production is that during anaerobic ethanolic fermentation of sugar feedstocks, substantial amounts of glycerol are invariably formed as a by-product. Sugar dissimilation during anaerobic growth of *S. cerevisiae* occurs via alcoholic fermentation. In this process, the NADH formed in glycolysis is reoxidised via a NAD^+ dependent alcohol dehydrogenase which converts acetaldehyde (formed by decarboxylation of pyruvate) to ethanol. This dissimilatory pathway is redox-neutral and can therefore not compensate a net reduction of NAD^+ to NADH occurring elsewhere in metabolism. Such net reduction of NAD^+ to NADH occurs in assimilation when yeast biomass is synthesized under anaerobic conditions

from sugars and e.g. ammonia. Under anaerobic conditions, NADH reoxidation in *S. cerevisiae* to compensate for the assimilatory-associated NADH-formation is mainly dependent on reduction of at least part of the sugar carbon source to glycerol, resulting in a lower ethanol yield. To address this problem, several metabolic engineering
5 approaches have been explored to reduce or eliminate glycerol production in anaerobically grown *S. cerevisiae*. Two examples are:

Nissen et al (2000) disclose an approach in which the enzyme NADPH-dependent glutamate dehydrogenase was deleted. Glutamine synthetase and glutamate synthase were overexpressed. The resulting strain no longer synthesized glutamate
10 from ammonium and 2-oxoglutarate via a NADPH-requiring route, but rather via a NADH- and ATP-requiring pathway. The resulting strain had a 10% higher yield in ethanol and a 38% lower glycerol yield than the wild type strain.

Guadalupe Medina et al. (2009, Appl. Environ. Microbiol., 76: 190–195) disclose a *S. cerevisiae* strain wherein production of the by-product glycerol is eliminated by
15 disruption of the endogenous NAD-dependent glycerol 3-phosphate dehydrogenase genes (*GPD1* and *GPD2*). Expression of the *E. coli mhpF* gene, encoding the acetylating NAD-dependent acetaldehyde dehydrogenase restored the ability of the GPD-disrupted strain to grow anaerobically. However, the GPD-disrupted strain could only grow anaerobically if the medium is supplemented with acetic acid.

20 (2) Optimization of glycerol production from sugars.

Production of glycerol from sugars, as opposed to minimizing its production during ethanol formation, has been an economically important process during World War 1 in Germany. In recent years, a renewed interest in this process is seen, as witnessed for instance by the work of Overkamp et al (2002, Appl. Environ. Microbiol.
25 68:2814-21).

(3) Fermentation of glycerol into ethanol.

The fermentation of glycerol into ethanol is not feasible in normal *S. cerevisiae* and under fully anoxic cultivation conditions due to an unbalanced balance for NADH. Glycerol is more reduced than ethanol and hence the organism cannot dispose of its
30 excess NADH by reoxidation under normal anoxic conditions. Some initial work has been done in the field of fermentation of glycerol into ethanol: Yu et al. (2010, Bioresour. Technol. 101(11):4157-61. Epub 2010 Feb 9) disclose *S. cerevisiae* strains metabolically engineered for improved ethanol production from glycerol by

simultaneous overexpression of glycerol dehydrogenase (*GCH*), dihydroxyacetone kinase (*DAK*) and the glycerol uptake protein (*GUP1*). In a later communication, Yu et al (2010, J. Biotechnol. doi:10.1016/j.jbiotec.2010.09.932) disclose an optimisation of their strain for ethanol production from glycerol by deleting two glycerol production genes, *FPS1* and *GPD2*. It was shown that ethanol production from glycerol is possible in their strains. However, the increase in yield was dependent on micro aerobic conditions, which can be explained by the requirement for NADH oxidation. The excess of NADH produced in the path from glycerol to ethanol apparently could only be reoxidised via an oxygen-dependent reaction.

Waks and Silver (2009, Appl. Environ. Microbiol., 75:1867–1875) disclose a synthetic dual-organism system for biohydrogen production. In a first step formate is produced by an engineered *S. cerevisiae* strain wherein a formate-overproducing pathway has been implemented. In a second step the formate produced by the engineered yeast is processed into hydrogen by *Escherichia coli*. The *S. cerevisiae* strain was engineered to produce formate by expressing the anaerobic enzyme pyruvate formate lyase (PFL) from *E. coli*. Formate production was further increased by also introducing expression of a downstream enzyme, the AdhE of *E. coli*, the bifunctional enzyme that reduces acetyl-CoA generated by PFL into ethanol.

It is an object of the present invention to provide for yeasts that are capable of producing ethanol from glycerol while retaining their abilities of fermenting hexoses (glucose, fructose, galactose, etc) as well as pentoses like xylose, as well as processes wherein these strains are used for the production of ethanol and/or other fermentation products.

Description of the invention

Definitions

Sequence identity is herein defined as a relationship between two or more amino acid (polypeptide or protein) sequences or two or more nucleic acid (polynucleotide) sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between amino acid or nucleic acid sequences, as the case may be, as determined by the match between strings of such sequences. "Similarity" between two amino acid sequences is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the

sequence of a second polypeptide. "Identity" and "similarity" can be readily calculated by known methods. The terms "sequence identity" or "sequence similarity" means that two (poly)peptide or two nucleotide sequences, when optimally aligned, preferably over the entire length (of at least the shortest sequence in the comparison) and maximizing the number of matches and minimizes the number of gaps such as by the programs ClustalW (1.83), GAP or BESTFIT using default parameters, share at least a certain percentage of sequence identity as defined elsewhere herein. GAP uses the Needleman and Wunsch global alignment algorithm to align two sequences over their entire length, maximizing the number of matches and minimizes the number of gaps. Generally, the GAP default parameters are used, with a gap creation penalty = 50 (nucleotides) / 8 (proteins) and gap extension penalty = 3 (nucleotides) / 2 (proteins). For nucleotides the default scoring matrix used is nwsgapdna and for proteins the default scoring matrix is Blosum62 (Henikoff & Henikoff, 1992, PNAS 89, 915-919). A preferred multiple alignment program for aligning protein sequences of the invention is ClustalW (1.83) using a blosum matrix and default settings (Gap opening penalty:10; Gap extension penalty: 0.05). It is clear than when RNA sequences are said to be essentially similar or have a certain degree of sequence identity with DNA sequences, thymine (T) in the DNA sequence is considered equal to uracil (U) in the RNA sequence. Sequence alignments and scores for percentage sequence identity may be determined using computer programs, such as the GCG Wisconsin Package, Version 10.3, available from Accelrys Inc., 9685 Scranton Road, San Diego, CA 92121-3752 USA or the open-source software Emboss for Windows (current version 2.10.0-0.8). Alternatively percent similarity or identity may be determined by searching against databases such as FASTA, BLAST, etc.

A variant of a nucleotide or amino acid sequence disclosed herein may also be defined as a nucleotide or amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the nucleotide or amino acid sequence specifically disclosed herein (e.g. in de the sequence listing).

Optionally, in determining the degree of amino acid similarity, the skilled person may also take into account so-called "conservative" amino acid substitutions, as will be clear to the skilled person. Conservative amino acid substitutions refer to the interchangeability of residues having similar side chains. For example, a group of amino acids having aliphatic side chains is glycine, alanine, valine, leucine, and

isoleucine; a group of amino acids having aliphatic-hydroxyl side chains is serine and threonine; a group of amino acids having amide-containing side chains is asparagine and glutamine; a group of amino acids having aromatic side chains is phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains is lysine, arginine, and histidine; and a group of amino acids having sulphur-containing side chains is cysteine and methionine. Preferred conservative amino acids substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine. Substitutional variants of the amino acid sequence disclosed herein are those in which at least one residue in the disclosed sequences has been removed and a different residue inserted in its place. Preferably, the amino acid change is conservative. Preferred conservative substitutions for each of the naturally occurring amino acids are as follows: Ala to ser; Arg to lys; Asn to gln or his; Asp to glu; Cys to ser or ala; Gln to asn; Glu to asp; Gly to pro; His to asn or gln; Ile to leu or val; Leu to ile or val; Lys to arg; gln or glu; Met to leu or ile; Phe to met, leu or tyr; Ser to thr; Thr to ser; Trp to tyr; Tyr to trp or phe; and, Val to ile or leu.

Nucleotide sequences of the invention may also be defined by their capability to hybridise with parts of specific nucleotide sequences disclosed herein, respectively, under moderate, or preferably under stringent hybridisation conditions. Stringent hybridisation conditions are herein defined as conditions that allow a nucleic acid sequence of at least about 25, preferably about 50 nucleotides, 75 or 100 and most preferably of about 200 or more nucleotides, to hybridise at a temperature of about 65°C in a solution comprising about 1 M salt, preferably 6 x SSC or any other solution having a comparable ionic strength, and washing at 65°C in a solution comprising about 0.1 M salt, or less, preferably 0.2 x SSC or any other solution having a comparable ionic strength. Preferably, the hybridisation is performed overnight, i.e. at least for 10 hours and preferably washing is performed for at least one hour with at least two changes of the washing solution. These conditions will usually allow the specific hybridisation of sequences having about 90% or more sequence identity.

Moderate conditions are herein defined as conditions that allow a nucleic acid sequences of at least 50 nucleotides, preferably of about 200 or more nucleotides, to hybridise at a temperature of about 45°C in a solution comprising about 1 M salt, preferably 6 x SSC or any other solution having a comparable ionic strength, and washing at room temperature in a solution comprising about 1 M salt, preferably 6 x

SSC or any other solution having a comparable ionic strength. Preferably, the hybridisation is performed overnight, i.e. at least for 10 hours, and preferably washing is performed for at least one hour with at least two changes of the washing solution. These conditions will usually allow the specific hybridisation of sequences having up to 50% sequence identity. The person skilled in the art will be able to modify these hybridisation conditions in order to specifically identify sequences varying in identity between 50% and 90%.

A “nucleic acid construct” or “nucleic acid vector” is herein understood to mean a man-made nucleic acid molecule resulting from the use of recombinant DNA technology. The term “nucleic acid construct” therefore does not include naturally occurring nucleic acid molecules although a nucleic acid construct may comprise (parts of) naturally occurring nucleic acid molecules. The terms “expression vector” or “expression construct” refer to nucleotide sequences that are capable of affecting expression of a gene in host cells or host organisms compatible with such sequences. These expression vectors typically include at least suitable transcription regulatory sequences and optionally, 3’ transcription termination signals. Additional factors necessary or helpful in effecting expression may also be present, such as expression enhancer elements. The expression vector will be introduced into a suitable host cell and be able to effect expression of the coding sequence in an in vitro cell culture of the host cell. The expression vector will be suitable for replication in the host cell or organism of the invention.

As used herein, the term “promoter” or “transcription regulatory sequence” refers to a nucleic acid fragment that functions to control the transcription of one or more coding sequences, and is located upstream with respect to the direction of transcription of the transcription initiation site of the coding sequence, and is structurally identified by the presence of a binding site for DNA-dependent RNA polymerase, transcription initiation sites and any other DNA sequences, including, but not limited to transcription factor binding sites, repressor and activator protein binding sites, and any other sequences of nucleotides known to one of skill in the art to act directly or indirectly to regulate the amount of transcription from the promoter. A “constitutive” promoter is a promoter that is active in most tissues under most physiological and developmental conditions. An “inducible” promoter is a promoter that is physiologically or developmentally regulated, e.g. by the application of a chemical inducer.

The term “selectable marker” is a term familiar to one of ordinary skill in the art and is used herein to describe any genetic entity which, when expressed, can be used to select for a cell or cells containing the selectable marker. The term “reporter” may be used interchangeably with marker, although it is mainly used to refer to visible markers, such as green fluorescent protein (GFP). Selectable markers may be dominant or recessive or bidirectional.

As used herein, the term “operably linked” refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is “operably linked” when it is placed into a functional relationship with another nucleic acid sequence. For instance, a transcription regulatory sequence is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein encoding regions, contiguous and in reading frame.

The terms “protein” or “polypeptide” are used interchangeably and refer to molecules consisting of a chain of amino acids, without reference to a specific mode of action, size, 3-dimensional structure or origin.

“Fungi” (singular fungus) are herein understood as heterotrophic eukaryotic microorganism that digest their food externally, absorbing nutrient molecules into their cells. Fungi are a separate kingdom of eukaryotic organisms and include yeasts, molds, and mushrooms. The terms fungi, fungus and fungal as used herein thus expressly includes yeasts as well as filamentous fungi.

The term “gene” means a DNA fragment comprising a region (transcribed region), which is transcribed into an RNA molecule (e.g. an mRNA) in a cell, operably linked to suitable regulatory regions (e.g. a promoter). A gene will usually comprise several operably linked fragments, such as a promoter, a 5’ leader sequence, a coding region and a 3’ nontranslated sequence (3’ end) comprising a polyadenylation site. “Expression of a gene” refers to the process wherein a DNA region which is operably linked to appropriate regulatory regions, particularly a promoter, is transcribed into an RNA, which is biologically active, i.e. which is capable of being translated into a biologically active protein or peptide.

The term “homologous” when used to indicate the relation between a given (recombinant) nucleic acid or polypeptide molecule and a given host organism or host cell, is understood to mean that in nature the nucleic acid or polypeptide molecule is

produced by a host cell or organisms of the same species, preferably of the same variety or strain. If homologous to a host cell, a nucleic acid sequence encoding a polypeptide will typically (but not necessarily) be operably linked to another (heterologous) promoter sequence and, if applicable, another (heterologous) secretory
5 signal sequence and/or terminator sequence than in its natural environment. It is understood that the regulatory sequences, signal sequences, terminator sequences, etc. may also be homologous to the host cell. In this context, the use of only “homologous” sequence elements allows the construction of “self-cloned” genetically modified organisms (GMO’s) (self-cloning is defined herein as in European Directive 98/81/EC
10 Annex II). When used to indicate the relatedness of two nucleic acid sequences the term “homologous” means that one single-stranded nucleic acid sequence may hybridize to a complementary single-stranded nucleic acid sequence. The degree of hybridization may depend on a number of factors including the amount of identity between the sequences and the hybridization conditions such as temperature and salt
15 concentration as discussed later.

The terms “heterologous” and “exogenous” when used with respect to a nucleic acid (DNA or RNA) or protein refers to a nucleic acid or protein that does not occur naturally as part of the organism, cell, genome or DNA or RNA sequence in which it is present, or that is found in a cell or location or locations in the genome or DNA or
20 RNA sequence that differ from that in which it is found in nature. Heterologous and exogenous nucleic acids or proteins are not endogenous to the cell into which it is introduced, but have been obtained from another cell or synthetically or recombinantly produced. Generally, though not necessarily, such nucleic acids encode proteins, i.e. exogenous proteins, that are not normally produced by the cell in which the DNA is
25 transcribed or expressed. Similarly exogenous RNA encodes for proteins not normally expressed in the cell in which the exogenous RNA is present. Heterologous/exogenous nucleic acids and proteins may also be referred to as foreign nucleic acids or proteins. Any nucleic acid or protein that one of skill in the art would recognize as foreign to the cell in which it is expressed is herein encompassed by the term heterologous or
30 exogenous nucleic acid or protein. The terms heterologous and exogenous also apply to non-natural combinations of nucleic acid or amino acid sequences, i.e. combinations where at least two of the combined sequences are foreign with respect to each other.

The “specific activity” of an enzyme is herein understood to mean the amount of activity of a particular enzyme per amount of total host cell protein, usually expressed in units of enzyme activity per mg total host cell protein. In the context of the present invention, the specific activity of a particular enzyme may be increased or decreased as compared to the specific activity of that enzyme in an (otherwise identical) wild type host cell.

“Anaerobic conditions” or an anaerobic fermentation process is herein defined as conditions or a fermentation process run in the absence of oxygen or in which substantially no oxygen is consumed, preferably less than 5, 2.5 or 1 mmol/L/h, more preferably 0 mmol/L/h is consumed (i.e. oxygen consumption is not detectable), and wherein organic molecules serve as both electron donor and electron acceptors.

Detailed description of the invention

Biodiesel is having and will continue to have a substantial impact in the area of biofuels. It can be produced from a variety of sources including algae, rape and palm. A comprehensive list of potential feedstock sources is available from “Feedstock and Biodiesel Characteristics Report, “Renewable Energy Group, Inc., www.regfuel.com (2009)”. The vegetable oils from which the biodiesel is produced by interesterification inevitably result in large byproduct streams of glycerol. Furthermore, the plant material from which the oil is extracted will give rise to lignocellulosic byproduct. The latter by-product will contain cellulose and hemicelluloses in varying relative concentrations. These polymeric sugars can be hydrolyzed into glucose and pentoses such as xylose, respectively, by known processes that are still optimized. Hence, two by-product streams may result from the production of biodiesel. The present invention describes a method for the concomitant fermentation of both glycerol and the sugars into ethanol. This process can be applied for various feedstocks. As an example, the situation in the palm oil industry is presented in some detail below.

The palm oil industry produces huge amounts of by-products. The oil consists of only 10% of the total biomass produced in the plantation. The remainder consists of lignocellulosic materials such as oil palm fronds, trunks and empty fruit bunches (EFB) (See e.g. http://umpir.ump.edu.my/697/1/Kamarul_Azlan_Abd._Hamid.pdf). These lignocellulosic materials can be used for the production of value added products such as ethanol that may be used as biofuel. The lignocellulosic materials may be hydrolysed

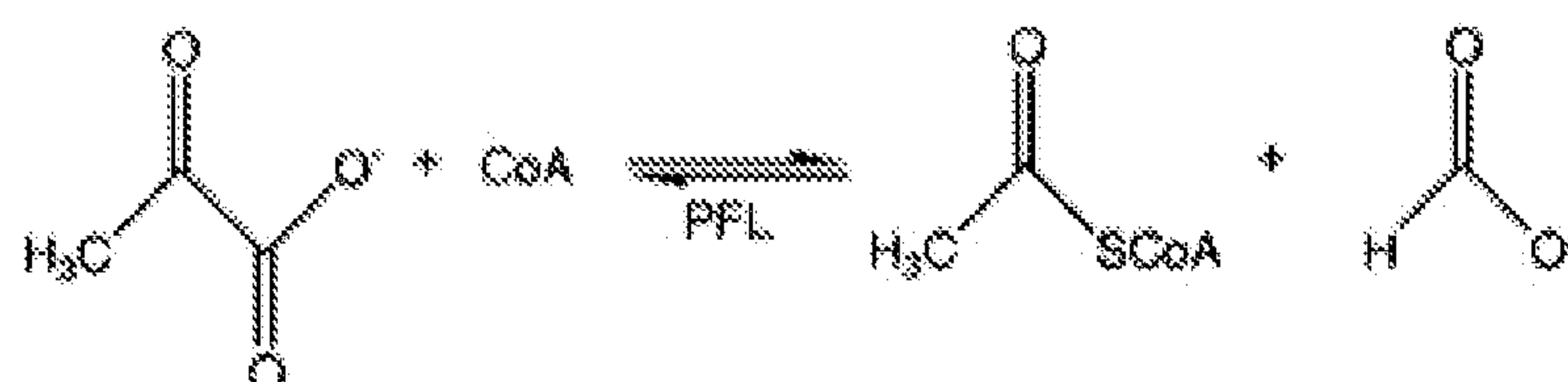
by known methods into streams containing hexoses and pentose such as glucose and xylose, which can then be fermented to ethanol by engineered yeast strains as e.g. described by Kuypers et al (2005, FEMS Yeast Res. 5:399-409). However, hexose- and pentose-containing streams obtained from hydrolysis of this type of lignocellulosic materials are relatively dilute, and will therefore yield relatively low ethanol concentrations of no more than about 5% (v/v).

At the same time, large amounts of concentrated glycerol are generated as a by-product in biodiesel production from transesterification reactions using vegetable oils such as palm oil and alcohols. The availability of crude glycerol is predicted to increase over the next years as a result of the growth in biodiesel production worldwide. Consequently large amounts of concentrated glycerol will be available at low cost near palm oil plantations. This offers the possibility to increase the carbon concentration of diluted streams obtained in the above-described hydrolysis of lignocellulosic materials by mixing in the available highly concentrated glycerol. Utilisation of such mixed streams would require engineered yeast strains which can ferment not only hexoses and pentose but also glycerol to ethanol. However, anaerobic consumption of large amount of glycerol would produce a redox imbalance in yeast. The present invention addresses this problem by engineering the yeast to produce formic acid, in addition to ethanol. The present invention therefore provides yeast strains engineered to produce ethanol and formic acid from carbon sources containing one or more of glycerol, hexose and pentose, as well as processes wherein these strains are used to produce ethanol and formic acid from these carbon sources.

In a first aspect the invention relates to a fungal host cell comprising an exogenous gene coding for a enzyme with the ability to convert pyruvate and coenzyme-A into formate and acetyl-CoA. An enzyme with the ability to convert pyruvate and coenzyme-A into formate and acetyl-CoA katalyses the reaction (EC 2.3.1.54):



11



Such an enzyme is herein understood as an enzyme having pyruvate formate lyase activity and is referred to as a pyruvate formate lyase (PFL) or formate C-acetyltransferase.

5 A suitable exogenous gene coding for an enzyme with pyruvate formate lyase activity is e.g. a prokaryotic pyruvate formate lyase, such as the pyruvate formate lyase from *E. coli*. The *E. coli* pyruvate formate lyase is a dimer of PflB (encoded by *pflB*), whose maturation requires the activating enzyme PflAE (encoded by *pflA*), radical S-adenosylmethionine, and a single electron donor, which in the case of *E. coli* is
10 flavodoxin (Buis and Broderick, 2005, Arch. Biochem. Biophys. 433:288–296; Sawers and Watson, 1998, Mol. Microbiol. 29:945–954). However, Waks and Silver (*supra*) have shown that for activation of the pyruvate formate lyase in yeast, only co-expression of an activating enzyme is required but expression of flavodoxin is not necessary.

15 The exogenous gene coding for an enzyme with pyruvate formate lyase activity preferably comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 1. Suitable examples of organisms comprising an enzyme with pyruvate formate lyase activity are provided in Table 1. Further examples of such organisms are
20 listed by Lehtiö and Goldman (2004, Prot. Engin. Design & Selection, 17:545-552). The amino acid sequences of these enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding enzyme with pyruvate formate lyase activity (see e.g. SEQ ID NO: 2). The exogenous gene coding for an enzyme with pyruvate formate lyase activity
25 may also comprises a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence of SEQ ID NO: 1. Preferably the amino acid sequence has no more than 420, 380, 300, 250, 200, 150, 100, 75, 50, 40, 30, 20, 10 or 5 amino acid substitutions, insertions and/or deletions as compared to SEQ ID NO: 1.

Table 1: Enzymes with pyruvate formate lyase activity related to *E.coli* pflB

Organism	Amino acid identity (%)
<i>Escherichia coli</i> str. K12 substr. MG1655	100%
<i>Shigella boydii</i>	100%
<i>Escherichia albertii</i> TW07627	99%
<i>Salmonella enterica</i>	97%
<i>Citrobacter rodentium</i> ICC168	97%
<i>Klebsiella pneumoniae</i> NTUH-K2044	96%
<i>Yersinia aldovae</i> ATCC 35236	91%
<i>Proteus mirabilis</i> HI4320	87%
<i>Haemophilus influenzae</i> Rd KW20	86%
<i>Actinobacillus succinogenes</i> 130Z	83%
<i>Piromyces</i> sp. E2	57%

The host cell of the invention further preferably comprises an exogenous gene coding for the PflAE activating enzyme for activation of the pyruvate formate lyase.

- 5 The pyruvate formate lyase activating enzyme is herein understood as an enzyme that katalyses the reaction:

S-adenosyl-L-methionine + dihydroflavodoxin + [pyruvate formate lyase]-glycine



10 5'-deoxyadenosine + L-methionine + flavodoxin semiquinone +
[formate C-acetyltransferase]-glycin-2-yl radical

The exogenous gene coding for the pyruvate formate lyase activating enzyme preferably comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 3. Suitable examples of organisms comprising an enzyme with pyruvate formate lyase activity are provided in Table 2. The amino acid sequences of these enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding enzyme with pyruvate formate lyase activity (see e.g. SEQ ID NO: 4). The exogenous gene

15

20

coding for the pyruvate formate lyase activating enzyme may also comprises a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence of SEQ ID NO: 3. Preferably the amino acid sequence has no more than 135, 125, 100, 75, 50, 25, 20, 15, 10, 8, 5 or 2 amino acid substitutions, insertions and/or deletions as compared to SEQ ID NO: 3.

Table 2: Pyruvate formate lyase activating enzymes related to *E.coli* pflA

Organism	Amino acid identity (%)
<i>Escherichia coli</i> str. K12 substr. MG1655	100%
<i>Shigella boydii</i>	100%
<i>Escherichia albertii</i> TW07627	99%
<i>Salmonella enterica</i>	98%
<i>Citrobacter rodentium</i> ICC168	98%
<i>Klebsiella pneumoniae</i> NTUH-K2044	97%
<i>Yersinia rohdei</i> ATCC 43380	89%
<i>Proteus penneri</i> ATCC 35198	85%
<i>Haemophilus parasuis</i> 29755	70%

In a preferred host cell of the invention, the exogenous genes coding for the enzyme with pyruvate formate lyase activity and the pyruvate formate lyase activating enzyme are from the same donor organism, i.e. be homologous to each other. However, the exogenous genes coding for the enzyme with pyruvate formate lyase activity and the pyruvate formate lyase activating enzyme may also be from different donor organisms, i.e. be heterologous to each other.

In one aspect the invention relates to methods for preparing or constructing the yeast cells of the invention. For this purpose standard genetic and molecular biology techniques are used that are generally known in the art and have e.g. been described by Sambrook and Russell (2001, "Molecular cloning: a laboratory manual" (3rd edition), Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press) and Ausubel et al. (1987, eds., "Current protocols in molecular biology", Green Publishing and Wiley

Interscience, New York). Furthermore, the construction of mutated host yeast strains is carried out by genetic crosses, sporulation of the resulting diploids, tetrad dissection of the haploid spores containing the desired auxotrophic markers, and colony purification of such haploid host yeasts in the appropriate selection medium. All of these methods
5 are standard yeast genetic methods known to those in the art. See, for example, Sherman et al., *Methods Yeast Genetics*, Cold Spring Harbor Laboratory, NY (1978) and Guthrie et al. (Eds.) *Guide To Yeast Genetics and Molecular Biology* Vol. 194, Academic Press, San Diego (1991).

The exogenous genes coding for the enzyme having pyruvate formate lyase
10 activity and the pyruvate formate lyase activating enzyme, preferably are expression constructs comprising the nucleotide sequence coding for the enzymes operably linked to suitable expression regulatory regions/sequences to ensure expression of the enzymes upon transformation of the expression constructs into the host cell of the invention. Thus, the gene or expression construct will at least comprise a promoter that
15 is functional in the host cell operably linked to the coding sequence. The gene or construct may further comprise a 5' leader sequence upstream of the coding region and a 3'-nontranslated sequence (3'end) comprising a polyadenylation site and a transcription termination site downstream of the coding sequence. It is understood that the nucleotide sequences coding for the enzyme having pyruvate formate lyase activity
20 and the pyruvate formate lyase activating enzyme may be present together on a single expression construct, or each enzyme may be present on a separate expression construct.

Suitable promoters for expression of the nucleotide sequences coding for the enzyme having pyruvate formate lyase activity and the pyruvate formate lyase
25 activating enzyme (as well as other enzymes of the invention; see below) include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Promoters having these characteristics are widely available and known to the skilled person. Suitable examples of such promoters include e.g.
30 promoters from glycolytic genes such as the phosphofructokinase (*PPK*), triose phosphate isomerase (*TPI*), glyceraldehyde-3-phosphate dehydrogenase (*GPD*, *TDH3* or *GAPDH*), pyruvate kinase (*PYK*), phosphoglycerate kinase (*PGK*), glucose-6-phosphate isomerase promoter (*PGII*) promoters from yeasts. More details about such

promoters from yeast may be found in (WO 93/03159). Other useful promoters are ribosomal protein encoding gene promoters (*TEF1*), the lactase gene promoter (*LAC4*), alcohol dehydrogenase promoters (*ADH1*, *ADH4*, and the like), the enolase promoter (*ENO*) and the hexose(glucose) transporter promoter (*HXT7*). Alternatively, the
5 nucleotide sequences encoding the enzyme having pyruvate formate lyase activity and the PflAE activating enzyme are expressed under anaerobic conditions by using an anoxic promoter such as e.g. the *S. cerevisiae ANB1* promoter (SEQ ID NO: 24). Other promoters, both constitutive and inducible, and enhancers or upstream activating sequences will be known to those of skill in the art. Preferably the promoter that is
10 operably linked to nucleotide sequence as defined above is homologous to the host cell. Suitable terminator sequences are e.g. obtainable from the cytochrome c1 (*CYC1*) gene or an alcohol dehydrogenase gene (e.g. *ADH1*).

To increase the likelihood that the enzyme having pyruvate formate lyase activity is expressed at sufficient levels and in active form in the transformed host cells of the
15 invention, the nucleotide sequence encoding these enzymes, as well as the pyruvate formate lyase activating enzyme and other enzymes of the invention (see below), are preferably adapted to optimise their codon usage to that of the host cell in question. The adaptiveness of a nucleotide sequence encoding an enzyme to the codon usage of a host cell may be expressed as codon adaptation index (CAI). The codon adaptation index is
20 herein defined as a measurement of the relative adaptiveness of the codon usage of a gene towards the codon usage of highly expressed genes in a particular host cell or organism. The relative adaptiveness (w) of each codon is the ratio of the usage of each codon, to that of the most abundant codon for the same amino acid. The CAI index is defined as the geometric mean of these relative adaptiveness values. Non-synonymous
25 codons and termination codons (dependent on genetic code) are excluded. CAI values range from 0 to 1, with higher values indicating a higher proportion of the most abundant codons (see Sharp and Li, 1987, Nucleic Acids Research 15: 1281-1295; also see: Jansen et al., 2003, Nucleic Acids Res. 31(8):2242-51). An adapted nucleotide sequence preferably has a CAI of at least 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 or 0.9. Most
30 preferred are the sequences which have been codon optimised for expression in the fungal host cell in question such as e.g. *S. cerevisiae* cells.

The fungal host cell to be transformed with a nucleic acid construct comprising a nucleotide sequence encoding an enzyme with pyruvate formate lyase activity

preferably is a yeast host cell. Preferably the host cell is a cultured cell. The host cell of the invention, preferably is a host capable of active or passive pentose (xylose and preferably also arabinose) transport into the cell. The host cell preferably contains active glycolysis. The host cell may further preferably contains an endogenous pentose phosphate pathway and may contain endogenous xylulose kinase activity so that xylulose isomerised from xylose may be metabolised to pyruvate. The host further preferably contains enzymes for conversion of a pentose (preferably through pyruvate) to a desired fermentation product such as ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propane-diol, butanols (1-butanol, 2-butanol, isobutanol) and isoprenoid-derived products. A particularly preferred host cell is a yeast cell that is naturally capable of alcoholic fermentation, preferably, anaerobic alcoholic fermentation. The yeast host cell further preferably has a high tolerance to ethanol, a high tolerance to low pH (i.e. capable of growth at a pH lower than 5, 4, or 3) and towards organic acids like lactic acid, acetic acid or formic acid and sugar degradation products such as furfural and hydroxy-methylfurfural, and a high tolerance to elevated temperatures. Any of these characteristics or activities of the host cell may be naturally present in the host cell or may be introduced or modified by genetic modification, preferably by self cloning or by the methods of the invention described below. A suitable cell is a cultured cell, a cell that may be cultured in fermentation process e.g. in submerged or solid state fermentation. Particularly suitable host cells are eukaryotic microorganism like e.g. fungi, however, most suitable for use in the present inventions are yeasts.

Yeasts are herein defined as eukaryotic microorganisms and include all species of the subdivision Eumycotina (Yeasts: characteristics and identification, J.A. Barnett, R.W. Payne, D. Yarrow, 2000, 3rd ed., Cambridge University Press, Cambridge UK; and, The yeasts, a taxonomic study, C.P. Kurtzman and J.W. Fell (eds) 1998, 4th ed., Elsevier Science Publ. B.V., Amsterdam, The Netherlands) that predominantly grow in unicellular form. Yeasts may either grow by budding of a unicellular thallus or may grow by fission of the organism. Preferred yeasts cells for use in the present invention belong to the genera *Saccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*. Preferably the yeast is capable of anaerobic fermentation, more preferably anaerobic alcoholic fermentation. Over the years suggestions have been made for the introduction

of various organisms for the production of bio-ethanol from crop sugars. In practice, however, all major bio-ethanol production processes have continued to use the yeasts of the genus *Saccharomyces* as ethanol producer. This is due to the many attractive features of *Saccharomyces* species for industrial processes, i.e., a high acid-, ethanol- and osmo-tolerance, capability of anaerobic growth, and of course its high alcoholic fermentative capacity. Preferred yeast species as host cells include *S. cerevisiae*, *S. exiguus*, *S. bayanus*, *K. lactis*, *K. marxianus* and *Schizosaccharomyces pombe*.

In a further embodiment, the host cell of the invention further comprises a genetic modification that reduces specific NAD^+ -dependent formate dehydrogenase activity in the cell. NAD^+ -dependent formate dehydrogenases (FDH; EC 1.2.1.2) catalyse the oxidation of formate to bicarbonate, donating the electrons to NAD^+ . In the cells of the invention, the specific formate dehydrogenase activity is preferably reduced by at least a factor 0.8, 0.5, 0.3, 0.1, 0.05 or 0.01 as compared to a strain which is genetically identical except for the genetic modification causing the reduction in specific activity, preferably under anaerobic conditions. Formate dehydrogenase activity may be determined as described by Overkamp et al. (2002, Yeast 19:509-520).

Preferably, formate dehydrogenase activity is reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an formate dehydrogenase. Preferably, the genetic modifications reduce or inactivate the expression of each endogenous copy of the gene encoding a specific formate dehydrogenase in the cell's genome. A given cell may comprise multiple copies of the gene encoding a specific formate dehydrogenase with one and the same amino acid sequence as a result of di-, poly- or aneu-ploidy. In such instances preferably the expression of each copy of the specific gene that encodes the formate dehydrogenase is reduced or inactivated. Alternatively, a cell may contain several different (iso)enzymes with formate dehydrogenase activity that differ in amino acid sequence and that are each encoded by a different gene. In such instances, in some embodiments of the invention it may be preferred that only certain types of the isoenzymes are reduced or inactivated while other types remain unaffected. Preferably, however, expression of all copies of genes encoding (iso)enzymes with formate dehydrogenase activity is reduced or inactivated.

Preferably, a gene encoding formate dehydrogenase activity is inactivated by deletion of at least part of the gene or by disruption of the gene, whereby in this context

the term gene also includes any non-coding sequence up- or down-stream of the coding sequence, the (partial) deletion or inactivation of which results in a reduction of expression of formate dehydrogenase activity in the host cell.

A preferred gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is the *S. cerevisiae FDH1* as
5 described by van den Berg and Steensma (1997, Yeast 13:551-559), encoding the amino acid sequence of SEQ ID NO: 5 and orthologues thereof in other species. Therefore a gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention preferably is a gene encoding a formate
10 dehydrogenase having an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 5 or a gene encoding a formate dehydrogenase having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence of SEQ ID NO: 5.

However, in some strains of *S. cerevisiae* a second gene encoding a formate
15 dehydrogenase is active, i.e. the *FDH2*, see e.g. Overkamp et al. (2002, *supra*). Another preferred gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention therefore is an *S. cerevisiae FDH2* as described by Overkamp et al. (2002, *supra*), encoding the amino acid sequence of SEQ ID NO: 6 and orthologues thereof in other species. Therefore a gene encoding a formate
20 dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention preferably is a gene encoding a formate dehydrogenase having an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 6 or a gene encoding a formate dehydrogenase having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the
25 amino acid sequence of SEQ ID NO: 6.

In a further embodiment, the activity of all the genes in the host cell encoding a formate dehydrogenase is reduced or inactivated. In such cells preferably all copies of endogenous genes encoding a formate dehydrogenase having an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ
30 ID NO: 5 or 6 (or having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to the amino acid sequences of SEQ ID NO: 5 or 6) are inactivated or at least reduced in expression.

In a preferred embodiment, the host cell of the invention further comprises an exogenous gene coding for an enzyme with the ability to reduce acetylCoA into acetaldehyde, which gene confers to the cell the ability to convert acetylCoA (and/or acetic acid) into ethanol. An enzyme with the ability to reduce acetylCoA into acetaldehyde is herein understood as an enzyme which catalyzes the reaction (ACDH; EC 1.2.1.10):



Thus, the enzyme catalyzes the conversion of acetylCoA into acetaldehyde (and *vice versa*) and is also referred to as an (acetylating NAD-dependent) acetaldehyde dehydrogenase or an acetyl-CoA reductase. The enzyme may be a bifunctional enzyme which further catalyzes the conversion of acetaldehyde into ethanol (and *vice versa*; see below). For convenience we shall refer herein to an enzyme having at least the ability to reduce acetylCoA into either acetaldehyde or ethanol as an “acetaldehyde dehydrogenase”. It is further understood herein that the host cell has endogenous alcohol dehydrogenase activities which allow the cell, being provided with acetaldehyde dehydrogenase activity, to complete the conversion of acetyl-CoA into ethanol. It is further also preferred that the host cell has endogenous acetyl-CoA synthetase which allow the cell, being provided with acetaldehyde dehydrogenase activity, to complete the conversion of acetic acid (via acetyl-CoA) into ethanol.

The exogenous gene may encode for a monofunctional enzyme having only acetaldehyde dehydrogenase activity (i.e. an enzyme only having the ability to reduce acetylCoA into acetaldehyde) such as e.g. the acetaldehyde dehydrogenase encoded by the *E.coli* mhpF gene. A suitable exogenous gene coding for an enzyme with acetaldehyde dehydrogenase activity comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 7. Suitable examples of prokaryotes comprising monofunctional enzymes with acetaldehyde dehydrogenase activity are provided in Table 3. The amino acid sequences of these monofunctional enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding monofunctional enzyme (see e.g. SEQ ID NO: 8). The exogenous gene coding for the monofunctional enzyme

having only acetaldehyde dehydrogenase activity may also comprises a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 7.

5 Table 3: Enzymes with acetaldehyde dehydrogenase activity related to *E.coli* mhpF

Organism	Amino acid identity (%)
<i>Escherichia coli</i> str. K12 substr. MG1655	100%
<i>Shigella sonnei</i>	100%
<i>Escherichia coli</i> IAI39	99%
<i>Citrobacter youngae</i> ATCC 29220	93%
<i>Citrobacter</i> sp. 30_2	92%
<i>Klebsiella pneumoniae</i> 342)	87%
<i>Klebsiella variicola</i>	87%
<i>Pseudomonas putida</i>	81%
<i>Ralstonia eutropha</i> JMP134	82%
<i>Burkholderia</i> sp. H160	81%
<i>Azotobacter vinelandii</i> DJ	79%
<i>Ralstonia metallidurans</i> CH34	70%
<i>Xanthobacter autotrophicus</i> Py2	67%
<i>Burkholderia cenocepacia</i> J2315	68%
<i>Frankia</i> sp. EAN1pec	67%
<i>Polaromonas</i> sp. JS666	68%
<i>Burkholderia phytofirmans</i> PsJN	70%
<i>Rhodococcus opacus</i> B4	64%

Preferably, the host cell comprises an exogenous gene coding for a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activity, which gene confers to the cell the ability to convert acetylCoA into ethanol. The advantage of using a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activities as opposed to separate enzymes for each of the acetaldehyde dehydrogenase and alcohol dehydrogenase activities, is that it allows for direct channelling of the intermediate between enzymes that catalyze consecutive reactions in

a pathway offers the possibility of an efficient, exclusive, and protected means of metabolite delivery. Substrate channelling thus decreases transit time of intermediates, prevents loss of intermediates by diffusion, protects labile intermediates from solvent, and forestalls entrance of intermediates into competing metabolic pathways. The bifunctional enzyme therefore allows for a more efficient conversion of acetylCoA into ethanol as compared to the separate acetaldehyde dehydrogenase and alcohol dehydrogenase enzymes. A further advantage of using the bifunctional enzyme is that it may also be used in host cells having little or no alcohol dehydrogenase activity under the condition used, such as e.g. anaerobic conditions and/or conditions of catabolite repression.

Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity are known in the art prokaryotes and protozoans, including e.g. the bifunctional enzymes encoded by the *Escherichia coli* adhE and *Entamoeba histolytica* ADH2 genes (see e.g. Bruchaus and Tannich, 1994, J. Biochem. 303: 743-748; Burdette and Zeikus, 1994, J. Biochem. 302: 163-170; Koo et al., 2005, Biotechnol. Lett. 27: 505-510; Yong et al., 1996, Proc Natl Acad Sci USA, 93: 6464-6469). Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity are larger proteins consisting of around 900 amino acids and they are bifunctional in that they exhibit both acetaldehyde dehydrogenase (ACDH; EC 1.2.1.10) and alcohol dehydrogenase activity (ADH; EC 1.1.1.1). The *E. coli* adhE and *Entamoeba histolytica* ADH2 show 45% amino acid identity. Therefore, in one embodiment of the invention, a suitable exogenous gene coding for a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activity comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO: 9 and 11. Suitable examples of prokaryotes comprising bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity are provided in Tables 4 and 5. The amino acid sequences of these bifunctional enzymes are available in public databases and can be used by the skilled person to design codon-optimised nucleotide sequences coding for the corresponding bifunctional enzyme (see e.g. SEQ ID NO: 10 or 12). The exogenous gene coding for the a bifunctional enzyme with acetaldehyde dehydrogenase and alcohol dehydrogenase activity may also comprises a nucleotide sequence coding for an amino

acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO: 9 and 11.

5 Table 4: Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity related to *E.coli* adhE

Organism	Amino acid identity (%)
<i>Escherichia coli</i> O157:H7 str. Sakai	100%
<i>Shigella sonnei</i>	100%
<i>Shigella dysenteriae</i> 1012	99%
<i>Klebsiella pneumoniae</i> 342	97%
<i>Enterobacter</i> sp. 638	94%
<i>Yersinia pestis</i> biovar Microtus str. 91001	90%
<i>Serratia proteamaculans</i> 568	90%
<i>Pectobacterium carotovorum</i> WPP14	90%
<i>Sodalis glossinidius</i> str. 'morsitans'	87%
<i>Erwinia tasmaniensis</i> Et1/99	86%
<i>Aeromonas hydrophila</i> ATCC 7966	81%
<i>Vibrio vulnificus</i> YJ016]	76%

Table 5: Bifunctional enzymes with acetaldehyde dehydrogenase and alcohol dehydrogenase activity related to *Entamoeba histolytica* ADH2

Organism	Amino acid identity (%)
<i>Entamoeba histolytica</i> HM-1:IMSS	99%
<i>Entamoeba dispar</i> SAW760	98%
<i>Mollicutes bacterium</i> D7	65%
<i>Fusobacterium mortiferum</i> ATCC 9817	64%
<i>Actinobacillus succinogenes</i> 130Z	63%
<i>Pasteurella multocida</i> Pm70	62%
<i>Mannheimia succiniciproducens</i> MBEL55E	61%
<i>Streptococcus</i> sp. 2_1_36FAA]	61%

For expression of the nucleotide sequence encoding the bifunctional enzyme having acetaldehyde dehydrogenase and alcohol dehydrogenase activities, or the enzyme having acetaldehyde dehydrogenase activity, the nucleotide sequence (to be
5 expressed) is placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure expression of the enzyme upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters for expression of the nucleotide sequence coding for the enzyme having the bifunctional enzyme having acetaldehyde dehydrogenase and
10 alcohol dehydrogenase activities, or the enzyme having acetaldehyde dehydrogenase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given above.

Preferably, the nucleotide sequence encoding the bifunctional enzyme having
15 acetaldehyde dehydrogenase and alcohol dehydrogenase activities, or the enzyme having acetaldehyde dehydrogenase activity is adapted to optimise its codon usage to that of the host cell in question (as described above).

The enzyme having acetaldehyde dehydrogenase and optionally alcohol dehydrogenase activities preferably is expressed in active form in the transformed host
20 cell. Thus, expression of the nucleotide sequence in the host cell produces an acetaldehyde dehydrogenase with a specific activity of at least 0.005, 0.010, 0.020, 0.050 or 0.10 $\mu\text{mol min}^{-1} (\text{mg protein})^{-1}$, determined as acetyl-CoA dependent rate of NADH reduction in cell extracts of the transformed host cell at 30 °C as described in the Examples herein.

25 In a further embodiment, the host cell of the invention further comprises a genetic modification that increases at least one of: i) the specific activity of glycerol dehydrogenase; ii) the specific activity of dihydroxyacetone kinase; and, iii) transport of glycerol into the cell.

Preferably, the genetic modification that increases the specific activity of at least
30 one of glycerol dehydrogenase and dihydroxyacetone kinase is overexpression of a nucleotide sequence encoding at least one of a glycerol dehydrogenase and dihydroxyacetone kinase. However, alternatively, the specific activity of the glycerol dehydrogenase and/or dihydroxyacetone kinase may be increased by expressing an

enzyme having increased activity as compared to the endogenous wild type enzyme of the host cell, in addition to, or as a replacement for the wild type enzyme.

A glycerol dehydrogenase is herein understood as an enzyme that catalyzes the chemical reaction (EC 1.1.1.6):

5



Other names in common use include glycerin dehydrogenase, NAD^+ -linked glycerol dehydrogenase and glycerol: NAD^+ 2-oxidoreductase. Preferably the genetic
10 modification causes overexpression of a glycerol dehydrogenase, e.g. by overexpression of a nucleotide sequence encoding a glycerol dehydrogenase. The nucleotide sequence encoding the glycerol dehydrogenase may be endogenous to the cell or may be a glycerol dehydrogenase that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of glycerol dehydrogenase in the cells
15 of the invention are e.g. the glycerol dehydrogenase gene from *S. cerevisiae* (*GCY1*) as e.g. described by Oechsner et al. (1988, FEBS Lett. 238: 123-128) or Voss et al. (1997, Yeast 13: 655-672). Preferably, the nucleotide sequence encoding the glycerol dehydrogenase comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence
20 identity with SEQ ID NO: 13 or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 13. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the glycerol dehydrogenase is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the glycerol dehydrogenase of SEQ ID NO:
25 13.

The glycerol dehydrogenase encoded by the yeast *GCY1* gene however appears to be specific for the cofactor NADP^+ (EC 1.1.1.72) as opposed to NAD^+ (EC 1.1.1.6). The yeasts such *S. cerevisiae* appear to lack NAD^+ -dependent glycerol dehydrogenase activity (EC 1.1.1.6) (see e.g. KEGG pathway 00561). More preferred nucleotide
30 sequences for overexpression of a heterologous glycerol dehydrogenase in the cells of the invention are therefore e.g. sequences encoding bacterial glycerol dehydrogenases which use NAD^+ as cofactor (EC 1.1.1.6), such as e.g. the *gldA* gene from *E. coli* described by Truniger and Boos (1994, J Bacteriol. 176(6):1796-1800), the expression

of which in yeast has already been reported (Lee and Dasilva, 2006, Metab Eng. 8(1):58-65). Preferably, the nucleotide sequence encoding a heterologous glycerol dehydrogenase comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 49 or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 49. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the heterologous glycerol dehydrogenase is overexpressed, such as e.g. a codon-optimised nucleotide sequence encoding the amino acid sequence of the glycerol dehydrogenase of SEQ ID NO: 49. Such a codon-optimised nucleotide sequence is e.g. provided in SEQ ID NO: 50 (positions 10 - 1113; CAI = 0.976).

For overexpression of the nucleotide sequence encoding the glycerol dehydrogenase, the nucleotide sequence (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure overexpression of the glycerol dehydrogenase enzyme upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters for (over)expression of the nucleotide sequence coding for the enzyme having glycerol dehydrogenase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given above. In the cells of the invention, a glycerol dehydrogenase to be overexpressed is preferably overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the glycerol dehydrogenase is overexpressed under anaerobic conditions by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity in the cell), the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme in the cell. Overexpression of the nucleotide sequence in the host cell produces a specific glycerol dehydrogenase activity of at least 0.2, 0.5, 1.0, 2.0, or 5.0 U min⁻¹ (mg protein)⁻¹, determined in cell extracts of the transformed host cells at 30 °C as described in the Examples herein.

A dihydroxyacetone kinase is herein understood as an enzyme that catalyzes the chemical reaction ((EC 2.7.1.29):



5

Other names in common use include glycerone kinase, ATP:glycerone phosphotransferase and (phosphorylating) acetol kinase. It is understood that glycerone and dihydroxyacetone are the same molecule. Preferably the genetic modification causes overexpression of a dihydroxyacetone kinase, e.g. by overexpression of a nucleotide sequence encoding a dihydroxyacetone kinase. The nucleotide sequence encoding the dihydroxyacetone kinase may be endogenous to the cell or may be a dihydroxyacetone kinase that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of dihydroxyacetone kinase in the cells of the invention are e.g. the dihydroxyacetone kinase genes from *S. cerevisiae* (*DAK1*) and (*DAK2*) as e.g. described by Molin et al. (2003, J. Biol. Chem. 278:1415-1423). Preferably, the nucleotide sequence encoding the dihydroxyacetone kinase comprises an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO's: 14 and 15 or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 14 and 15. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the dihydroxyacetone kinase is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the dihydroxyacetone kinase of SEQ ID NO: 14 or a codon optimised nucleotide sequence encoding the dihydroxyacetone kinase of SEQ ID NO: 15. A preferred nucleotide sequence for overexpression of a dihydroxyacetone kinase is a nucleotide sequence encoding a dihydroxyacetone kinase comprises an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO's: 14 (*S. cerevisiae* (*DAK1*), having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 14.

30 Nucleotide sequences that may be used for overexpression of a heterologous dihydroxyacetone kinase in the cells of the invention are e.g. sequences encoding bacterial dihydroxyacetone kinases such as the *dhaK* gene from *Citrobacter freundii* e.g. described by Daniel et al. (1995, J. Bacteriol. 177:4392-4401). Preferably, the

nucleotide sequence encoding a heterologous dihydroxyacetone kinase comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with SEQ ID NO: 52 or a nucleotide sequence coding for an amino acid sequence having one or several
5 substitutions, insertions and/or deletions as compared to SEQ ID NO: 52. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the heterologous dihydroxyacetone kinase is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the amino acid sequence of the dihydroxyacetone kinase of SEQ ID NO: 52. Such a codon-optimised nucleotide sequence is e.g. provided in
10 SEQ ID NO: 53 (positions 10 - 1668).

For overexpression of the nucleotide sequence encoding the dihydroxyacetone kinase, the nucleotide sequence (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure overexpression of the dihydroxyacetone kinase enzyme
15 upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters for (over)expression of the nucleotide sequence coding for the enzyme having dihydroxyacetone kinase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction.
20 Examples of such promoters are given above. In the cells of the invention, a dihydroxyacetone kinase to be overexpressed is preferably overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the dihydroxyacetone kinase is overexpressed under anaerobic conditions by at least a
25 factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity in the cell), the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme in the cell.
30 Overexpression of the nucleotide sequence in the host cell produces a specific dihydroxyacetone kinase activity of at least 0.002, 0.005, 0.01, 0.02 or 0.05 U min⁻¹ (mg protein)⁻¹, determined in cell extracts of the transformed host cells at 30 °C as described in the Examples herein.

Preferably, the genetic modification that increases transport of glycerol into the cell preferably is a genetic modification that causes overexpression of a nucleotide sequence encoding at least one of a glycerol uptake protein and a glycerol channel.

A glycerol uptake protein is herein understood as a multimembrane-spanning protein that belongs to the included in the membrane bound O-acyltransferases (MBOAT) superfamily including e.g. the *S. cerevisiae* glycerol uptake proteins encoded by the *GUP1* and *GUP2* genes. Preferably the genetic modification causes overexpression of a glycerol uptake protein, e.g. by overexpression of a nucleotide sequence encoding a glycerol uptake protein. The nucleotide sequence encoding the glycerol uptake protein may be endogenous to the cell or may be a glycerol uptake protein that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of glycerol uptake protein in the cells of the invention are e.g. the glycerol uptake protein genes from *S. cerevisiae* (*GUP1*) and (*GUP2*) and orthologues thereof as e.g. described by Neves et al. (2004, FEMS Yeast Res. 5:51-62). Preferably, the nucleotide sequence encoding the glycerol uptake protein comprises a nucleotide sequence coding for an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO's: 16 (Gup1p) and 17 (Gup2p) or a nucleotide sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 16 and 17. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the glycerol uptake protein is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the glycerol uptake protein SEQ ID NO: 16 or a codon optimised nucleotide sequence encoding the glycerol uptake protein of SEQ ID NO: 17. Although the exact nature of the influence of GUP1 on glycerol transport is not yet clear, Yu et al. (2010, *supra*) have shown that overexpression of GUP1 in *S. cerevisiae* improves ethanol production on glycerol grown cells. A preferred nucleotide sequence for overexpression of a glycerol uptake protein is therefore a nucleotide sequence encoding a glycerol uptake protein that is capable of rescuing salt stress-associated phenotype of a *S. cerevisiae* *gup1Δ* mutant by complementation as described by Neves et al. (2004, *supra*). Such complementing orthologues of *S. cerevisiae* *GUP1* include nucleotide sequences encoding amino acid sequences having at least 60, 68, 72, 75, 80, 85, 90, 95, 98, 99% identity with the amino acid sequence of SEQ ID NO: 16 and may be obtained from yeast species

belonging to the genera of *Saccharomyces*, *Zygosaccharomyces*, *Kluyveromyces*, *Candida*, *Pichia*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*.

A glycerol channel is herein understood as a member of the MIP family of channel proteins reviewed by Reizer et al. (1993, CRC Crit. Rev. Biochem. Mol. Biol.,
5 28: 235-257), which channel proteins comprise a 250 -280 amino acid transmembrane domain consisting of six membrane-spanning domains and have at least 30, 35, 40, 45, 50, 60, 70, 80, 90, 95, 98 or 99 % amino acid identity, or at least 55, 60, 65, 70, 80, 90, 95, 98 or 99% amino acid similarity with the amino acid sequence between amino acids 250 and 530 of SEQ ID NO: 18, the *S. cerevisiae* FPS1 aquaglyceroporin. Alternatively
10 the channel protein comprises a 250 -280 amino acid transmembrane domain consisting of six membrane-spanning domains, and having one or several substitutions, insertions and/or deletions as compared to the amino acid sequence between amino acids 250 and 530 of SEQ ID NO: 18.

Nucleotide sequences that may be used for overexpression of a glycerol channel
15 in the cells of the invention include nucleotide sequences encoding the yeast aquaglyceroporin FPS1 gene from e.g. *S. cerevisiae* (Van Aelst et al., 1991, EMBO J. 10:2095-2104) and orthologues thereof from other yeasts including *Kluyveromyces lactis*, *Kluyveromyces marxianus* and *Zygosaccharomyces rouxii* as e.g. described by Neves et al. (2004, *supra*). However, the use of bacterial or plant glycerol channels is
20 not excluded as e.g. Luyten et al. (1995, EMBO J. 14:1360-1371) have shown that the *E.coli* glycerol facilitator, having only 30% sequence identity with the amino acid sequence between amino acids 250 and 530 of the *S. cerevisiae* FPS1 aquaglyceroporin, can complement glycerol uptake in a *S. cerevisiae fps1Δ* mutant. The nucleotide sequence encoding the glycerol channel may be endogenous to the cell or
25 may be a glycerol channel that is heterologous to the cell. In a preferred embodiment a codon-optimised (see above) nucleotide sequence encoding the glycerol channel is overexpressed, such as e.g. a codon optimised nucleotide sequence encoding the aquaglyceroporin of SEQ ID NO: 18.

For overexpression of the nucleotide sequence encoding the glycerol uptake
30 protein and/or the glycerol channel protein, the nucleotide sequence (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure overexpression of the glycerol uptake protein and/or the glycerol channel protein upon transformation of the

expression construct into the host cell of the invention (see above). Suitable promoters for (over)expression of the nucleotide sequence coding for the glycerol uptake protein and/or the glycerol channel protein include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given above. In the cells of the invention, a glycerol uptake protein and/or a glycerol channel protein to be overexpressed are preferably overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the glycerol uptake protein and/or the glycerol channel protein are overexpressed under anaerobic conditions by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity in the cell), the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme in the cell.

In a further embodiment, the host cell of the invention further comprises a genetic modification that increases the specific acetyl-CoA synthetase activity in the cell, preferably under anaerobic conditions as this activity is rate-limiting under these conditions. Acetyl-CoA synthetase or acetate-CoA ligase (EC 6.2.1.1) is herein understood as an enzyme that catalyzes the formation of a new chemical bond between acetate and coenzyme A (CoA). Preferably the genetic modification causes overexpression of a acetyl-CoA synthetase, e.g. by overexpression of a nucleotide sequence encoding a acetyl-CoA synthetase. The nucleotide sequence encoding the acetyl-CoA synthetase may be endogenous to the cell or may be a acetyl-CoA synthetase that is heterologous to the cell. Nucleotide sequences that may be used for overexpression of acetyl-CoA synthetase in the cells of the invention are e.g. the acetyl-CoA synthetase genes from *S. cerevisiae* (*ACS1* and *ACS2*) as e.g. described by de Jong-Gubbels et al. (1998, FEMS Microbiol Lett. 165: 15-20). Preferably, the nucleotide sequence encoding the acetyl-CoA synthetase comprises an amino acid sequence with at least 45, 50, 60, 65, 70, 75, 80, 85, 90, 95, 98, 99% amino acid sequence identity with at least one of SEQ ID NO's: 19 and 20, or a nucleotide

sequence coding for an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 19 and 20.

In one embodiment, the nucleotide sequence that is overexpressed encodes an acetyl-CoA synthetase with a high affinity for acetate. Use of an acetyl-CoA synthetase with a high affinity for acetate is preferred for conditions under which there is a relatively low concentration of acetic acid in the culture medium, e.g. no more than 2 g acetic acid/L culture medium. An acetyl-CoA synthetase with a high affinity for acetate is herein defined as an acetyl-CoA synthetase with a higher affinity for acetate than the acetyl-CoA synthetase encoded by the *S. cerevisiae* *ACS2* (SEQ ID NO: 20). Preferably, an acetyl-CoA synthetase with a high affinity for acetate has a K_m for acetate of no more than 10, 5, 2, 1, 0.5, 0.2 or 0.1 mM, such e.g. the acetyl-CoA synthetase encoded by the *S. cerevisiae* *ACS1* gene. More preferably a codon-optimised (see above) nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 19 is overexpressed.

In another embodiment, the nucleotide sequence that is overexpressed encodes an acetyl-CoA synthetase with a high maximum rate (v_{max}). Use of an acetyl-CoA synthetase with a high maximum rate is preferred for condition under which there is a relatively high concentration of acetic acid in the culture medium, e.g. at least 2, 3, 4 or 5 g acetic acid/L culture medium. An acetyl-CoA synthetase with a high maximum rate is herein defined as an acetyl-CoA synthetase with a higher maximum rate than the acetyl-CoA synthetase encoded by the *S. cerevisiae* *ACS1*. Preferably, the acetyl-CoA synthetase with a high maximum rate is the acetyl-CoA synthetase encoded by the *S. cerevisiae* *ACS2* gene. More preferably a codon-optimised (see above) nucleotide sequence encoding the amino acid sequence of SEQ ID NO: 20 is overexpressed.

For overexpression of the nucleotide sequence encoding the acetyl-CoA synthetase (to be overexpressed) can be placed in an expression construct wherein it is operably linked to suitable expression regulatory regions/sequences to ensure overexpression of the acetyl-CoA synthetase enzyme upon transformation of the expression construct into the host cell of the invention (see above). Suitable promoters for (over)expression of the nucleotide sequence coding for the enzyme having acetyl-CoA synthetase activity include promoters that are preferably insensitive to catabolite (glucose) repression, that are active under anaerobic conditions and/or that preferably do not require xylose or arabinose for induction. Examples of such promoters are given

above. In the cells of the invention, an acetyl-CoA synthetase to be overexpressed is overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. Preferably, the acetyl-CoA synthetase is overexpressed under
5 anaerobic conditions by at least a factor 2, 5, 10, 20, 50, or 100 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity (specific activity), the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the
10 enzyme.

In a further embodiment, the host cell of the invention further comprises a genetic modification that reduces specific NAD^+ -dependent glycerol 3-phosphate dehydrogenase activity in the cell. Glycerol 3-phosphate dehydrogenase or glycerolphosphate dehydrogenase (EC 1.1.1.8) katalyses the reduction of
15 dihydroxyacetone phosphate to sn-glycerol 3-phosphate while oxidising NADH to NAD^+ . In the cells of the invention, the specific glycerolphosphate dehydrogenase activity is preferably reduced by at least a factor 0.8, 0.5, 0.3, 0.1, 0.05 or 0.01 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression, preferably under anaerobic conditions.

20 Preferably, glycerolphosphate dehydrogenase activity is reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an glycerolphosphate dehydrogenase. Preferably, the genetic modifications reduce or inactivate the expression of each endogenous copy of the gene encoding a specific glycerolphosphate dehydrogenase in the cell's genome. A given cell may
25 comprise multiple copies of the gene encoding a specific glycerolphosphate dehydrogenase with one and the same amino acid sequence as a result of di-, poly- or aneu-ploidy. In such instances preferably the expression of each copy of the specific gene that encodes the glycerolphosphate dehydrogenase is reduced or inactivated. Alternatively, a cell may contain several different (iso)enzymes with glycerolphosphate
30 dehydrogenase activity that differ in amino acid sequence and that are each encoded by a different gene. In such instances, in some embodiments of the invention it is preferred that only certain types of the isoenzymes are reduced or inactivated while other types remain unaffected (see below). Preferably, the gene is inactivated by deletion of at least

part of the gene or by disruption of the gene, whereby in this context the term gene also includes any non-coding sequence up- or down-stream of the coding sequence, the (partial) deletion or inactivation of which results in a reduction of expression of glycerolphosphate dehydrogenase activity in the host cell.

5 A preferred gene encoding a glycerolphosphate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is the *S. cerevisiae GPD2* gene as described by Eriksson et al. (1995, Mol. Microbiol. 17: 95-107), encoding the amino acid sequence of SEQ ID NO: 21 and orthologues thereof in other species. Therefore a gene encoding a glycerolphosphate dehydrogenase whose activity is to be reduced or
10 inactivated in the cell of the invention preferably is a gene encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 21 or an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 21.

15 In a preferred embodiment of the invention, the host cell of the invention comprises a functional high-osmolarity glycerol response pathway. Preferably therefore, only the activity of the gene(s) encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 21 are reduced or inactivated, while at least one endogenous gene encoding a
20 glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 22 is functional. SEQ ID NO: 22 depicts the amino acid sequence encoded by the *S. cerevisiae GPD1* gene as described by Albertyn et al. (1994, Mol. Cell. Biol. 14: 4135-4144), which has 69% amino acid identity with the *S. cerevisiae GPD2* glycerolphosphate dehydrogenase.
25 The *S. cerevisiae GPD1* gene is the stress-induced glycerolphosphate dehydrogenase of *S. cerevisiae*, which is important for growth under osmotic stress as may occur under industrial fermentations conditions. Its expression is *inter alia* regulated by the high-osmolarity glycerol response pathway. It is therefore advantageous that a host cell of the invention has at least one functional copy of a endogenous gene encoding a
30 glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to SEQ ID NO: 22 or an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to SEQ ID NO: 22.

In a further embodiment, the activity of all the genes in the host cell encoding a glycerolphosphate dehydrogenase is reduced or inactivated. In such cells preferably all copies of endogenous genes encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70, 75, 80, 85, 90, 95, 98 or 99% sequence identity to
5 SEQ ID NO: 21 or 22 (or having an amino acid sequence having one or several substitutions, insertions and/or deletions as compared to at least one of SEQ ID NO's: 21 and 22) are inactivated or at least reduced in expression.

In a further preferred embodiment, the host cell of the invention has at least one of: a) the ability of isomerising xylose to xylulose; and, b) the ability to convert L-arabinose into D-xylulose 5-phosphate. For a) the cell preferably has a functional
10 exogenous xylose isomerase gene, which gene confers to the cell the ability to isomerise xylose into xylulose. For b) the cell preferably has functional exogenous genes coding for a L-arabinose isomerase, a L-ribulokinase and a L-ribulose-5-phosphate 4-epimerase, which genes together confers to the cell the ability to isomerise
15 convert L-arabinose into D-xylulose 5-phosphate.

Fungal host cells having the ability of isomerising xylose to xylulose as e.g. described in WO 03/0624430 and in WO 06/009434. The ability of isomerising xylose to xylulose is preferably conferred to the cell by transformation with a nucleic acid construct comprising a nucleotide sequence encoding a xylose isomerase. Preferably
20 the cell thus acquires the ability to directly isomerise xylose into xylulose. More preferably the cell thus acquires the ability to grow aerobically and/or anaerobically on xylose as sole energy and/or carbon source though direct isomerisation of xylose into xylulose (and further metabolism of xylulose). It is herein understood that the direct isomerisation of xylose into xylulose occurs in a single reaction catalysed by a xylose
25 isomerase, as opposed to the two step conversion of xylose into xylulose via a xylitol intermediate as catalysed by xylose reductase and xylitol dehydrogenase, respectively.

Several xylose isomerases (and their amino acid and coding nucleotide sequences) that may be successfully used to confer to the cell of the invention the ability to directly isomerise xylose into xylulose have been described in the art. These
30 include the xylose isomerases of *Piromyces* sp. and of other anaerobic fungi that belongs to the families *Neocallimastix*, *Caecomyces*, *Piromyces* or *Ruminomyces* (WO 03/0624430), *Cyllamyces aberensis* (US 20060234364), *Orpinomyces* (Madhavan et al., 2008, DOI 10.1007/s00253-008-1794-6), the xylose isomerase of the bacterial

genus *Bacteroides*, including e.g. *B. thetaiotaomicron* (WO 06/009434), *B. fragilis*, and *B. uniformis* (WO 09/109633), the xylose isomerase of the anaerobic bacterium *Clostridium phytofermentans* (Brat et al., 2009, Appl. Environ. Microbiol. 75: 2304–2311), and the xylose isomerases of *Clostridium difficile*, *Ciona intestinales* and
5 *Fusobacterium mortiferum* (WO 10/074577).

Fungal host cells having the ability to convert L-arabinose into D-xylulose 5-phosphate as e.g. described in Wisselink et al. (2007, AEM Accepts, published online ahead of print on 1 June 2007; Appl. Environ. Microbiol. doi:10.1128/AEM.00177-07) and in EP 1 499 708. The ability of to converting L-arabinose into D-xylulose 5-phosphate is preferably conferred to the cell by transformation with a nucleic acid
10 construct(s) comprising nucleotide sequences encoding a) an arabinose isomerase; b) a ribulokinase, preferably a L-ribulokinase a xylose isomerase; and c) a ribulose-5-P-4-epimerase, preferably a L-ribulose-5-P-4-epimerase. Preferably, in the cells of the invention, the ability to convert L-arabinose into D-xylulose 5-phosphate is the ability
15 to convert L-arabinose into D-xylulose 5-phosphate through the subsequent reactions of 1) isomerisation of arabinose into ribulose; 2) phosphorylation of ribulose to ribulose 5-phosphate; and, 3) epimerisation of ribulose 5-phosphate into D-xylulose 5-phosphate. Suitable nucleotide sequences encoding arabinose isomerases, a ribulokinases and ribulose-5-P-4-epimerases may be obtained from *Bacillus subtilis*, *Escherichia coli* (see
20 e.g. EP 1 499 708), *Lactobacilli*, e.g. *Lactobacillus plantarum* (see e.g. Wisselink et al. *supra*), or species of *Clavibacter*, *Arthrobacter* and *Gramella*, of which preferably *Clavibacter michiganensis*, *Arthrobacter aurescens* and *Gramella forsetii* (see WO2009/011591).

The transformed host cell of the invention further preferably comprises xylulose
25 kinase activity so that xylulose isomerised from xylose may be metabolised to pyruvate. Preferably, the cell contains endogenous xylulose kinase activity. More preferably, a cell of the invention comprises a genetic modification that increases the specific xylulose kinase activity. Preferably the genetic modification causes overexpression of a xylulose kinase, e.g. by overexpression of a nucleotide sequence
30 encoding a xylulose kinase. The gene encoding the xylulose kinase may be endogenous to the cell or may be a xylulose kinase that is heterologous to the cell. A nucleotide sequence that may be used for overexpression of xylulose kinase in the cells of the invention is e.g. the xylulose kinase gene from *S. cerevisiae* (*XKS1*) as described by

Deng and Ho (1990, Appl. Biochem. Biotechnol. 24-25: 193-199). Another preferred xylulose kinase is a xylose kinase that is related to the xylulose kinase from *Piromyces* (*xylB*; see WO 03/0624430). This *Piromyces* xylulose kinase is actually more related to prokaryotic kinase than to all of the known eukaryotic kinases such as the yeast kinase.

5 The eukaryotic xylulose kinases have been indicated as non-specific sugar kinases, which have a broad substrate range that includes xylulose. In contrast, the prokaryotic xylulose kinases, to which the *Piromyces* kinase is most closely related, have been indicated to be more specific kinases for xylulose, i.e. having a narrower substrate range. In the cells of the invention, a xylulose kinase to be overexpressed is
10 overexpressed by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to a strain which is genetically identical except for the genetic modification causing the overexpression. It is to be understood that these levels of overexpression may apply to the steady state level of the enzyme's activity, the steady state level of the enzyme's protein as well as to the steady state level of the transcript coding for the enzyme.

15 A cell of the invention further preferably comprises a genetic modification that increases the flux of the pentose phosphate pathway as described in WO 06/009434. In particular, the genetic modification causes an increased flux of the non-oxidative part pentose phosphate pathway. A genetic modification that causes an increased flux of the non-oxidative part of the pentose phosphate pathway is herein understood to mean a
20 modification that increases the flux by at least a factor 1.1, 1.2, 1.5, 2, 5, 10 or 20 as compared to the flux in a strain which is genetically identical except for the genetic modification causing the increased flux. The flux of the non-oxidative part of the pentose phosphate pathway may be measured as described in WO 06/009434.

Genetic modifications that increase the flux of the pentose phosphate pathway
25 may be introduced in the cells of the invention in various ways. These including e.g. achieving higher steady state activity levels of xylulose kinase and/or one or more of the enzymes of the non-oxidative part pentose phosphate pathway and/or a reduced steady state level of unspecific aldose reductase activity. These changes in steady state activity levels may be effected by selection of mutants (spontaneous or induced by
30 chemicals or radiation) and/or by recombinant DNA technology e.g. by overexpression or inactivation, respectively, of genes encoding the enzymes or factors regulating these genes.

In a preferred cell of the invention, the genetic modification comprises overexpression of at least one enzyme of the (non-oxidative part) pentose phosphate pathway. Preferably the enzyme is selected from the group consisting of the enzymes encoding for ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase. Various combinations of enzymes of the (non-oxidative part) pentose phosphate pathway may be overexpressed. E.g. the enzymes that are overexpressed may be at least the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate 3-epimerase; or at least the enzymes ribulose-5-phosphate isomerase and transketolase; or at least the enzymes ribulose-5-phosphate isomerase and transaldolase; or at least the enzymes ribulose-5-phosphate 3-epimerase and transketolase; or at least the enzymes ribulose-5-phosphate 3-epimerase and transaldolase; or at least the enzymes transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate 3-epimerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, transketolase and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, and transaldolase; or at least the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, and transketolase. In one embodiment of the invention each of the enzymes ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase are overexpressed in the cell of the invention. Preferred is a cell in which the genetic modification comprises at least overexpression of the enzyme transaldolase. More preferred is a cell in which the genetic modification comprises at least overexpression of both the enzymes transketolase and transaldolase as such a host cell is already capable of anaerobic growth on xylose. In fact, under some conditions we have found that cells overexpressing only the transketolase and the transaldolase already have the same anaerobic growth rate on xylose as do cells that overexpress all four of the enzymes, i.e. the ribulose-5-phosphate isomerase, ribulose-5-phosphate 3-epimerase, transketolase and transaldolase. Moreover, cells of the invention overexpressing both of the enzymes ribulose-5-phosphate isomerase and ribulose-5-phosphate 3-epimerase are preferred over cells overexpressing only the isomerase or only the 3-epimerase as overexpression of only one of these enzymes may produce metabolic imbalances.

There are various means available in the art for overexpression of enzymes in the host cells of the invention. In particular, an enzyme may be overexpressed by

increasing the copynumber of the gene coding for the enzyme in the cell, e.g. by integrating additional copies of the gene in the cell's genome, by expressing the gene from an episomal multicopy expression vector or by introducing a episomal expression vector that comprises multiple copies of the gene. The coding sequence used for
5 overexpression of the enzymes preferably is homologous to the host cell of the invention. However, coding sequences that are heterologous to the host cell of the invention may likewise be applied. Alternatively overexpression of enzymes in the cells of the invention may be achieved by using a promoter that is not native to the sequence coding for the enzyme to be overexpressed, i.e. a promoter that is
10 heterologous to the coding sequence to which it is operably linked. Although the promoter preferably is heterologous to the coding sequence to which it is operably linked, it is also preferred that the promoter is homologous, i.e. endogenous to the cell of the invention. Preferably the heterologous promoter is capable of producing a higher steady state level of the transcript comprising the coding sequence (or is capable of
15 producing more transcript molecules, i.e. mRNA molecules, per unit of time) than is the promoter that is native to the coding sequence, preferably under conditions where one or more of xylose, arabinose and glucose are available as carbon sources, more preferably as major carbon sources (i.e. more than 50% of the available carbon source consists of one or more of xylose, arabinose and glucose), most preferably as sole
20 carbon sources.

A further preferred cell of the invention comprises a genetic modification that reduces unspecific aldose reductase activity in the cell. Preferably, unspecific aldose reductase activity is reduced in the host cell by one or more genetic modifications that reduce the expression of or inactivates a gene encoding an unspecific aldose reductase.
25 Preferably, the genetic modifications reduce or inactivate the expression of each endogenous copy of a gene encoding an unspecific aldose reductase that is capable of reducing an aldopentose, including, xylose, xylulose and arabinose, in the cell's genome. A given cell may comprise multiple copies of genes encoding unspecific aldose reductases as a result of di-, poly- or aneu-ploidy, and/or a cell may contain
30 several different (iso)enzymes with aldose reductase activity that differ in amino acid sequence and that are each encoded by a different gene. Also in such instances preferably the expression of each gene that encodes an unspecific aldose reductase is reduced or inactivated. Preferably, the gene is inactivated by deletion of at least part of

the gene or by disruption of the gene, whereby in this context the term gene also includes any non-coding sequence up- or down-stream of the coding sequence, the (partial) deletion or inactivation of which results in a reduction of expression of unspecific aldose reductase activity in the host cell. A nucleotide sequence encoding an aldose reductase whose activity is to be reduced in the cell of the invention and amino acid sequences of such aldose reductases are described in WO 06/009434 and include e.g. the (unspecific) aldose reductase genes of *S. cerevisiae* *GRE3* gene (Träff et al., 2001, Appl. Environm. Microbiol. 67: 5668-5674) and orthologues thereof in other species.

10 A further preferred transformed host cell according to the invention may comprises further genetic modifications that result in one or more of the characteristics selected from the group consisting of (a) increased transport of xylose and/or arabinose into the cell; (b) decreased sensitivity to catabolite repression; (c) increased tolerance to ethanol, osmolarity or organic acids; and, (d) reduced production of by-products. By-
15 products are understood to mean carbon-containing molecules other than the desired fermentation product and include e.g. xylitol, arabinitol, glycerol and/or acetic acid. Any genetic modification described herein may be introduced by classical mutagenesis and screening and/or selection for the desired mutant, or simply by screening and/or selection for the spontaneous mutants with the desired characteristics. Alternatively, the
20 genetic modifications may consist of overexpression of endogenous genes and/or the inactivation of endogenous genes. Genes the overexpression of which is desired for increased transport of arabinose and/or xylose into the cell are preferably chosen from genes encoding a hexose or pentose transporter. In *S. cerevisiae* and other yeasts these genes include *HXT1*, *HXT2*, *HXT3*, *HXT4*, *HXT5*, *HXT7* and *GAL2*, of which *HXT7*,
25 *HXT5* and *GAL2* are most preferred (see Sedlack and Ho, Yeast 2004; 21: 671-684). Another preferred transporter for expression in yeast is the glucose transporter encoded by the *P. stipitis* *SUT1* gene (Katahira et al., 2008, Enzyme Microb. Technol. 43: 115-119). Similarly orthologues of these transporter genes in other species may be overexpressed. Other genes that may be overexpressed in the cells of the invention
30 include genes coding for glycolytic enzymes and/or ethanologenic enzymes such as alcohol dehydrogenases. Preferred endogenous genes for inactivation include hexose kinase genes e.g. the *S. cerevisiae* *HXK2* gene (see Diderich et al., 2001, Appl. Environ. Microbiol. 67: 1587-1593); the *S. cerevisiae* *MIG1* or *MIG2* genes; genes

coding for enzymes involved in glycerol metabolism such as the *S. cerevisiae* glycerol-phosphate dehydrogenase 1 and/or 2 genes; or (hybridising) orthologues of these genes in other species. Other preferred further modifications of host cells for xylose fermentation are described in van Maris et al. (2006, Antonie van Leeuwenhoek
5 90:391–418), WO2006/009434, WO2005/023998, WO2005/111214, and WO2005/091733. Any of the genetic modifications of the cells of the invention as described herein are, in as far as possible, preferably introduced or modified by self cloning genetic modification.

A preferred host cell according to the invention has the ability to grow on at least
10 one of xylose and arabinose as carbon/energy source, preferably as sole carbon/energy source, and preferably under anaerobic conditions, i.e. conditions as defined herein below for anaerobic fermentation process. Preferably, when grown on xylose as carbon/energy source the host cell produces essentially no xylitol, e.g. the xylitol produced is below the detection limit or e.g. less than 5, 2, 1, 0.5, or 0.3% of the carbon
15 consumed on a molar basis. Preferably, when grown on arabinose as carbon/energy source, the cell produces essentially no arabinitol, e.g. the arabinitol produced is below the detection limit or e.g. less than 5, 2, 1, 0.5, or 0.3 % of the carbon consumed on a molar basis.

A preferred host cell of the invention has the ability to grow on at least one of a
20 hexose, a pentose, glycerol, acetic acid and combinations thereof at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0,25 or 0,3 h⁻¹ under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions. Therefore, preferably the host cell has the ability to grow on at least one of xylose and arabinose as sole carbon/energy source at a rate of at least 0.01,
25 0.02, 0.05, 0.1, 0.2, 0,25 or 0,3 h⁻¹ under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions. More preferably, the host cell has the ability to grow on a mixture of a hexose (e.g. glucose) and at least one of xylose and arabinose (in a 1:1 weight ratio) as sole carbon/energy source at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0,25 or 0,3 h⁻¹
30 under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions. Most preferably, the host cell has the ability to grow on a mixture of a hexose (e.g. glucose), at least one of xylose and arabinose and glycerol (in a 1:1:1 weight ratio) as sole carbon/energy source

at a rate of at least 0.01, 0.02, 0.05, 0.1, 0.2, 0.25 or 0.3 h⁻¹ under aerobic conditions, or, more preferably, at a rate of at least 0.005, 0.01, 0.02, 0.05, 0.08, 0.1, 0.12, 0.15 or 0.2 h⁻¹ under anaerobic conditions.

In a one aspect, the invention relates to the use of a yeast cell according to the
5 invention for the preparation of a fermentation product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propane-diol, butanols and isoprenoid-derived products.

In another aspect the invention relates to a process for producing a fermentation
product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic
10 acid, acrylic acid, 1,3-propane-diol, butanols (1-butanol, 2-butanol, isobutanol) and isoprenoid-derived products. The process can be a process wherein formate is produced in addition to the fermentation product. The process preferably comprises the step of: a) fermenting a medium with a yeast cell, whereby the medium contains or is fed with: a) a source of at least one of a hexose, a pentose and glycerol and whereby the yeast cell
15 ferments the at least one of a hexose, pentose and glycerol to the fermentation product, and optionally to formate. The yeast cell preferably is a (host) cell as herein defined above. The process preferably comprise one or more further steps wherein the fermentation product and/or the formate are recovered. The process may be a batch process, a fed-batch process or a continuous process as are well known in the art.

20 In a preferred process the source of at least one of a hexose, a pentose and glycerol comprises or consist of: hexose and pentose; hexose and glycerol; pentose and glycerol; hexose, pentose and glycerol. In a further preferred process, the source of at least one of a hexose, a pentose and glycerol further comprises acetic acid.

In a preferred process the source of hexose comprises or consists of glucose.
25 Preferably the source of pentose comprises or consists of at least one of xylose and arabinose, of which xylose is preferred. Preferably, the medium fermented by the cells of the invention comprises or is fed with (fractions of) hydrolysed biomass comprising at least one at least one of a hexose and a pentose such as glucose, xylose and/or arabinose. The (fractions of) hydrolysed biomass comprising the hexoses and pentose
30 will usually also comprise acetic acid (or a salt thereof). An example of hydrolysed biomass to be fermented in the processes of the invention is e.g. hydrolysed lignocellulosic biomass. Lignocellulosic biomass is herein understood as plant biomass that is composed of cellulose, hemicellulose, and lignin. The carbohydrate polymers

(cellulose and hemicelluloses) are tightly bound to the lignin. Examples of lignocellulosic biomass to be hydrolysed for use in the present invention include agricultural residues (including e.g. empty fruit bunches (EFB) of oil palm, corn stover and sugarcane bagasse), wood residues (including sawmill and paper mill discards and
5 (municipal) paper waste. Methods for hydrolysis of biomass such as lignocelluloses are known in the art per se and include e.g. acids, such as sulphuric acid and enzymes such as cellulases and hemicellulases.

In the process of the invention, the sources of xylose, glucose and arabinose may be xylose, glucose and arabinose as such (i.e. as monomeric sugars) or they may be in
10 the form of any carbohydrate oligo- or polymer comprising xylose, glucose and/or arabinose units, such as e.g. lignocellulose, arabinans, xylans, cellulose, starch and the like. For release of xylose, glucose and/or arabinose units from such carbohydrates, appropriate carbohydrases (such as arabinases, xylanases, glucanases, amylases, cellulases, glucanases and the like) may be added to the fermentation medium or may
15 be produced by the modified host cell. In the latter case the modified host cell may be genetically engineered to produce and excrete such carbohydrases. An additional advantage of using oligo- or polymeric sources of glucose is that it enables to maintain a low(er) concentration of free glucose during the fermentation, e.g. by using rate-limiting amounts of the carbohydrases preferably during the fermentation. This, in turn,
20 will prevent repression of systems required for metabolism and transport of non-glucose sugars such as xylose and arabinose. In a preferred process the modified host cell ferments both the glucose and at least one of xylose and arabinose, preferably simultaneously in which case preferably a modified host cell is used which is insensitive to glucose repression to prevent diauxic growth. In addition to a source of at
25 least one of a hexose, a pentose and glycerol, as carbon source, the fermentation medium will further comprise the appropriate ingredients required for growth of the modified host cell. Compositions of fermentation media for growth of eukaryotic microorganisms such as yeasts are well known in the art.

In the process of the invention, the medium further preferably comprises and/or is
30 fed a source of glycerol. Glycerol for use in the process of the present invention may advantageously be glycerol that is generated as a by-product in biodiesel production from transesterification reactions using vegetable oils, such as palm oil, or animal fats and an alcohol.

The fermentation process may be an aerobic or an anaerobic fermentation process. An anaerobic fermentation process is herein defined as a fermentation process run in the absence of oxygen or in which substantially no oxygen is consumed, preferably less than 5, 2.5 or 1 mmol/L/h, more preferably 0 mmol/L/h is consumed (i.e. oxygen consumption is not detectable), and wherein organic molecules serve as both electron donor and electron acceptors. In the absence of oxygen, NADH produced in glycolysis and biomass formation, cannot be oxidised by oxidative phosphorylation. To solve this problem many microorganisms use pyruvate or one of its derivatives as an electron and hydrogen acceptor thereby regenerating NAD^+ . Thus, in a preferred anaerobic fermentation process pyruvate is used as an electron (and hydrogen acceptor) and is reduced to fermentation products such as ethanol, as well as non-ethanol fermentation products such as lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propane-diol, butanols (1-butanol, 2-butanol, isobutanol) and isoprenoid-derived products, preferably under concomitant production of formate. Anaerobic processes of the invention are preferred over aerobic processes because anaerobic processes do not require investments and energy for aeration and in addition, anaerobic processes produce higher product yields than aerobic processes. Alternatively, the fermentation process of the invention may be run under aerobic oxygen-limited conditions. Preferably, in an aerobic process under oxygen-limited conditions, the rate of oxygen consumption is at least 5.5, more preferably at least 6 and even more preferably at least 7 mmol/L/h.

The fermentation process is preferably run at a temperature that is optimal for the modified cells of the invention. Thus, for most yeasts cells, the fermentation process is performed at a temperature which is less than 42°C, preferably less than 38°C. For yeast cells, the fermentation process is preferably performed at a temperature which is lower than 35, 33, 30 or 28°C and at a temperature which is higher than 20, 22, or 25°C.

Because undissociated formic acid is more toxic to the host cells of the invention as compared to the dissociated acid ion, the fermentation process is preferably run at a pH that is higher than the pKa of formic acid, which is 3.75. The fermentation process is therefore preferably run at a pH of at least 3.8, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5 or 8.0. Preferably the fermentation process is run at at least the aforementioned pHs for the entire duration of the process. Alternatively the pH may be regulated so as to maintain the concentration of undissociated formic acid below a given concentration.

Preferably the pH of the medium is regulated during the fermentation process to maintain a concentration of undissociated formic acid that is not higher than 50.0, 30.0, 20.0, 18.1, 15.0, 10.0, 5.0, 2.0 or 1.0 mM.

A preferred fermentation process according to the invention is a process for the
5 production of ethanol and optionally formate, whereby the process comprises the step
of fermenting a medium with a yeast cell, whereby the medium contains or is fed with a
source of at least one of a hexose, a pentose and glycerol and whereby the yeast cell
ferments the at least one of a hexose, pentose and glycerol to ethanol and optionally
formate. Optionally the process comprises the step of recovery of at least one of ethanol
10 and optionally formate. The fermentation may further be performed as described above.
In the process the volumetric ethanol productivity is preferably at least 0.5, 1.0, 1.5,
2.0, 2.5, 3.0, 5.0 or 10.0 g ethanol per litre per hour. The ethanol yield on hexose and/or
pentose and/or glycerol (and/or acetate) in the process preferably is at least 50, 60, 70,
80, 90, 95 or 98%. The ethanol yield is herein defined as a percentage of the theoretical
15 maximum yield, which, for xylose, glucose and arabinose is 0.51 g. ethanol per g.
hexose or pentose. For glycerol the theoretical maximum yield is 0.50 g. ethanol per g.
glycerol and for acetic acid the theoretical maximum yield is 0.77 g. ethanol per g.
acetic acid.

In this document and in its claims, the verb "to comprise" and its conjugations is
20 used in its non-limiting sense to mean that items following the word are included, but
items not specifically mentioned are not excluded. In addition, reference to an element
by the indefinite article "a" or "an" does not exclude the possibility that more than one
of the element is present, unless the context clearly requires that there be one and only
one of the elements. The indefinite article "a" or "an" thus usually means "at least one".

25 All patent and literature references cited in the present specification are hereby
incorporated by reference in their entirety.

The following examples are offered for illustrative purposes only, and are not
intended to limit the scope of the present invention in any way.

Examples

1. Enzyme activity assays

Cell extracts for activity assays were prepared from exponentially growing aerobic or anaerobic batch cultures and analysed for protein content as described by
5 Abbot et al., (2009, Appl. Environ. Microbiol. 75: 2320-2325).

NAD⁺-dependent acetaldehyde dehydrogenase (EC 1.2.1.10) activity was measured at 30 °C by monitoring the oxidation of NADH at 340 nm. The reaction mixture (total volume 1 ml) contained 50 mM potassium phosphate buffer (pH 7.5), 0.15 mM NADH and cell extract. The reaction was started by addition of 0.5 mM
10 acetyl-Coenzyme A.

For glycerol 3-phosphate dehydrogenase (EC 1.1.1.8) activity determination, cell extracts were prepared as described above except that the phosphate buffer was replaced by triethanolamine buffer (10 mM, pH 5). Glycerol-3-phosphate dehydrogenase activities were assayed in cell extracts at 30 °C as described previously
15 (Blomberg and Adler, 1989, J. Bacteriol. 171: 1087-1092.9). Reaction rates were proportional to the amounts of cell extract added.

Acetyl-CoA synthase (EC 6.2.1.1) activity was measured as described by Frenkel and Kitchens (1977, J. Biol. Chem. 252: 504-507) which is a modification of the method of Webster (Webster, 1969, Methods Enzymol. 13: 375-381). NADH
20 formation measured is spectrophotometrically when the acetyl-CoA produced is coupled with citrate synthase and malate dehydrogenase reactions. The assay system contained 100 mM Tris-Cl(pH 7.6), 10 mM MgCl₂, 6 mM ATP, 5 mM malate, 1 mM NAD⁺, 0.1 mM NADH, 2.5 mM dithiothreitol or 2-mercaptoethanol, 0.2 mM coenzyme A, 25 µg citrate synthase (80 units/mg), 20 µg malate dehydrogenase (1000
25 units/mg), and 10 mM acetate and the reaction was measured rate was measured at 340 nm and calculated from the extinction coefficient of NADH (6.22×10^6 cm²/mol).

The activity of glycerol dehydrogenase and dihydroxyacetone kinase are measured at 30 °C in cell extracts, essentially as previously described (Gonzalez et al., 2008, Metab. Eng. 10, 234–245). Enzyme activities of glycerol dehydrogenase and
30 dihydroxyacetone kinase are reported as µmoles of substrate/min/mg of cell protein.

2. Strain construction

All modifications start with the xylose fermenting strain RN1041. RN1041 is a CEN.PK based strain with the following genotype:

Mat a, ura3-52, leu2-112, his3::loxP, gre3::loxP, loxP-Ptpi::TAL1, loxP-Ptpi::RK11, loxP-Ptpi-TKL1, loxP-Ptpi-RPE1, delta::Padh1XKS1Tcyc1-LEU2, delta::URA3-Ptpi-xylA-Tcyc1

Mat a = mating type a

ura3-52, leu2-112, his3::loxP mutations in the genes *ura3*, *leu2* and *his3*, the *ura3* is complemented by the *Piromyces xylA* overexpression construct, *leu2* is complemented by the *XKS1* overexpression construct. *his3* could be used for selection of additional plasmids, RN1041 needs histidine in the medium for growth.

gre3::loxP = deletion of the *gre3* gene encoding xylose reductase, *loxP* site is left after marker removal.

loxP-Ptpi= overexpression of the pentose phosphate pathway, *loxP* site upstream of constitutive promoter is left after marker removal

delta:: = integration of the construct after recombination on the long terminal repeats of the *Ty1* retrotransposon.

Constructs for expression of the *E.coli* pflA and pflB genes in yeast

The *E.coli* pflA and pflB genes were obtained as codon optimized synthetic genes. The pflA expression construct by ligating a TEF1 promoter-fragment (cut with the restriction enzymes *SalI* and *HindIII*; from plasmid collection Royal Nidalco), a pflA synthetic ORF (cut with *HindIII* and *BssHII*) and an ADH1 terminator-fragment (cut with *BssHII* and *BsiWI*; from plasmid collection Royal Nidalco) together into pCRII blunt (Invitrogen) to yield pRN613.

The pflB expression construct by ligating a PGK1 promoter-fragment (cut with the restriction enzymes *SpeI* and *PstI*; from plasmid collection Royal Nidalco), a pflB synthetic ORF (cut with *PstI* and *SalI*) and an PGI1 terminator-fragment (cut with *XhoI* and *BsiWI*; from plasmid collection Royal Nidalco) together into pCRII blunt (Invitrogen) to yield pRN614.

The LEU2 marker from plasmid pRS305 (Sikorski and Hieter, 1989, Genetics 122:19-27) is exchanged for the ZeoMX marker (*zeoMX* = resistance to phleomycin due to expression of the *Tn5 ble* gene with a TEF1 promoter and terminator sequence of

Ashbya gossypii). Yeast transformants are selected on media containing Zeomycin. pRS305 is cut with *Bst*BI and *Bsr*GI and a ZeoMX fragment (from plasmid collection Royal Nedalco) is cut with the same restriction enzymes. Ligation gives plasmid pRS30z.

- 5 A 2 μ ORI is introduced by cutting pRS30z with *Afl*III en *Mfe*I and cutting a 2 μ fragment (from plasmid collection Royal Nedalco) with *Afl*III and *Eco*RI. The fragments are ligated to yield pRN615 (6341bp)

For the final construction pRN615 is cut with *Xho*I and *Spe*I, this vector is combined with the inserts from pRN613 cut with *Sal*I and *Bsi*WI (1510bp) and pRN614cut with
10 *Spe*I and *Bsi*WI (3381bp) to produce pRN616 (SEQ ID NO: 31).

For expression of the *E.coli* pflA gene only, pRN613 is cut with *Sal*I and *Bsi*WI and this fragment is ligated into pRN615 cut with *Acc*65I and *Xho*I to produce pRN619 (SEQ ID NO: 32).

For expression of the *E.coli* pflB gene only, pRN614 is cut with *Xho*I and *Bsi*WI and
15 this fragment is ligated into pRN615 cut with *Acc*65I and *Xho*I to produce pRN620 (SEQ ID NO: 33).

Plasmid pRN616, pRN619 and pRN620 are used to transform the yeast strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

20

Constructs for overexpression of GCY1, gldA, DAK1 dhaK, and GUP1

GCY1

PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *Pst*I site 5' of the ATG and an *Afl*III site 3' of the TAA to produce the fragment of SEQ ID
25 NO: 25. A DNA fragment comprising the *S. cerevisiae* Actin1 promoter is ligated upstream of the GCY1 ORF and DNA fragment comprising the *S. cerevisiae* ADH1 terminator fragment is ligated downstream of the GCY1 ORF.

gldA

The construct for expression in yeast of the *E.coli* *gldA* was made by ligating a yeast
30 ACT1 promoter fragment (cut with the restriction enzymes *Spe*I and *Pst*I), a synthetic ORF (SEQ ID NO:50), encoding the *E.coli* *gldA*, (cut with *Pst*I en *Bss*HII) and a yeast CYC1 terminator fragment (cut with *Bss*HII and *Bsi*WI) together into pCRII blunt (Invitrogen) to yield pRNgldA (SEQ ID NO:51).

DAK1

PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *Xba*I site 5' of the ATG and a *Sal*I site 3' of the TAA to produce the fragment as contained in SEQ ID NO: 26. A DNA fragment comprising the *S. cerevisiae* TPI1 promoter is
5 ligated upstream of the DAK1 ORF and DNA fragment comprising the *S. cerevisiae* PGI1 terminator fragment is ligated downstream of the DAK1 ORF to produce the expression construct of SEQ ID NO: 26.

dhaK

The construct for expression in yeast of the *Citrobacter freundii* *dhaK* was made by
10 ligating the yeast TPI1 promoter fragment (cut with the restriction enzymes *Xho*I and *Xba*I), a synthetic ORF (SEQ ID NO:53), encoding the *C. freundii* *dhaK*, (cut with *Xba*I and *Sal*I) and a yeast PGI1 terminator fragment (cut with *Xho*I and *Bsi*WI) together into pCRII blunt (Invitrogen) to yield pRNdhaK (SEQ ID NO:54).

GUP1

15 PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *Hind*III site 5' of the ATG and a *Bam*HI site 3' of the TAA to produce the fragment as contained in SEQ ID NO: 27. A DNA fragment comprising the *S. cerevisiae* TDH3 promoter is ligated upstream of the GUP1 ORF and DNA fragment comprising the *S. cerevisiae* CYC1 terminator fragment is ligated downstream of the GUP1 ORF to produce the
20 expression construct of SEQ ID NO: 27.

Strains overexpressing GCY1, DAK1 and GUP1

The above-described expression constructs of GCY1, DAK1 and GUP1 are combined into a pRS315-based centromeric plasmid having the hygromycine resistance gene as selectable marker to produce pRN605 (SEQ ID NO: 29). Plasmid pRN605 is used to
25 transform the yeast strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

The above-described expression constructs of GCY1 and DAK1 are combined into a pRS315-based centromeric plasmid having the hygromycine resistance gene as selectable marker to produce pRN608 (SEQ ID NO: 30). is used to transform the yeast
30 strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strain RN1098, as indicated in Table 7.

Strains expressing *E.coli* gldA with or without expression of *C.freundii* dhaK

The *E.coli* gldA expression construct is cut from plasmid pRNgldA with the restriction enzymes *SpeI* and *BsiWI*. The *C.freundii* dhaK expression construct is cut from plasmid pRNdhaK with the restriction enzymes *BsiWI* and *XhoI*. These fragments are ligated into plasmid pRN595 cut with the restriction enzymes *SpeI* and *SalI* to yield
 5 pRN957 (SEQ ID NO: 55). Plasmid pRN958 (SEQ ID NO: 56) is subsequently made from plasmid pRN957 by deleting the *C.freundii* dhaK expression construct. To this end plasmid pRN957 is cut with the restriction enzymes *PspOMI* and *AvrII*, the single stranded parts are filled in with the DNA polymerase phusion (Finnzymes) and the plasmid is religated.

10 The expression constructs of DAK1 and GUP1 are combined into a pRS315-based centromeric plasmid having the hygromycine resistance gene as selectable marker to produce pRN607 (SEQ ID NO: 57).

Plasmids pRN957, pRN958 and pRN607 used to transform yeast together with the other plasmids indicated in Table 7 to produce strain RN1194, RN1195 and RN1196.

15

Construct for overexpression of FPS1

PCR is performed on genomic DNA of *S. cerevisiae* with primers introducing a *NsiI* site 5' of the ATG and a *BamHI* site 3' of the TAA to produce the fragment as contained in SEQ ID NO: 28. A DNA fragment comprising the *S. cerevisiae* ADH1(medium)
 20 promoter is ligated upstream of the FSP1 ORF and DNA fragment comprising the *S. cerevisiae* CYC1 terminator fragment is ligated downstream of the FSP1 ORF to produce the expression construct of SEQ ID NO: 28. Ligation of these fragments in pCRII blunt produces pRN617. The insert of pRN617 is cut with *BsiWI* and *XhoI* and cloned in pRN616 cut with *BsrGI* and *SalI* to produce pRN618 (SEQ ID NO: 34).
 25 Plasmid pRN618 is used to transform the yeast strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

Constructs for deletion of FDH1 and FDH2

Primers FDHuf, FDHur, FDHdf and FDHdr are used for amplification of genomic sequences fragments upstream and downstream of both FDH1 and FDH2 genes for
 30 their inactivation. The same primers are used for both FDH1 and FDH2 alleles.

For inactivation of the FDH1 allele, a 423bp upstream PCR fragment with a *BspEI* site at the 3'-end for ligation to the MX marker, and a *BbsI* site at the 5'-end for isolation of the deletion construct is amplified using FDHuf and FDHur and cloned in pCRII

(blunt) using topo-cloning (Invitrogen), which results in an additional *Bst*BI site at the 5'-end.

FDHuf: TCGAAGACTCCGAATGAAAAAGACATGCCAG (SEQ ID NO: 35)

FDHur: TCCGGATACCAAGTTCATTTTCAATACACCCCA (SEQ ID NO: 36)

5

A 378bp downstream PCR fragment with a *Nsi*I and a *Sph*I site at the 5'-end for ligation to the MX marker, and a *Bbs*I site at the 3'-end for isolation of the deletion construct is amplified using FDHdf and FDHdr.

10 FDHdf: ATGCATGCAGAATGGTTCTTATGCCAC (SEQ ID NO: 37)

FDHdr: GAAGACAGTTCTGTTATTAACGACGAGCCA (SEQ ID NO: 38)

For the final construction the downstream fragment cut with *Xho*I and *Nsi*I is ligated to the patMX marker (collection Royal Nedalco) cut with *Bsp*EI and *Nsi*I in the plasmid containing the upstream-fragment cut with *Bsp*EI and *Xho*I. The final plasmid, pRN621 (SEQ ID NO: 39), is cut with *Bbs*I prior to yeast transformation. Transformants are selected for phosphinotricin resistance. Correct integration of the disruption fragment is verified by PCR with specific primers.

FDH2 is found as a pseudogene in some strains (e.g. S288C, YPL275W and YPL276W). Two mutations restore the ORF homology with FDH1 (T436C and del A476). Sequence analysis shows that in the CEN.PK background (RN1041) FDH2 is one ORF homologous to FDH1.

For the final construction the downstream fragment cut with *Xho*I and *Sph*I is ligated to the kanMX marker (collection Royal Nedalco) cut with *Bsp*EI and *Sph*I into the plasmid containing the upstream fragment cut with *Bsp*EI and *Xho*I. The final plasmid, pRN622 (SEQ ID NO: 40), is cut with *Bbs*I prior to yeast transformation. Transformants are selected for G418 resistance. Correct integration of the disruption fragment is verified by PCR with specific primers.

Disruption fragments from pRN621 and pRN622 are used to transform the yeast strain RN1041 together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7.

Constructs for expression of the *E.coli* mhpF, adhE or the *Entamoeba histolytica* adh2 genes in yeast

For expression of the *E.coli* mhpF gene, a yeast PGK1 promoter fragment (*SpeI*-*PstI*) and an ADH1 terminator fragment (*AflIII*-*NotI*) (both from the Nedalco plasmid collection) were ligated onto the codon-optimized synthetic fragment encoding the *E.coli* mhpF (SEQ ID NO: 8).

- 5 pRS 303 with 2 μ ori (= pRN347, Royal Nedalco plasmid collection) was cut with *SpeI* and *NotI* and the mhpF expression construct was cloned into this vector to produce pRN558 (SEQ ID NO: 41). pRN558 is used to transform the yeast strain RN1041 (selection on medium lacking histidine) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Overexpression of mhpF is
10 verified by qPCR.

- For expression of the *E.coli* adhE gene, a codon optimized synthetic fragment encoding the *E.coli* adhE (SEQ ID NO: 10) is cut with *XbaI* and *AflIII* and ligated into pRN558 cut with *XbaI* and *AflIII* (replacing the *E.coli* mhpF gene in pRN558) to produce pRN595 (SEQ ID NO: 42). pRN595 is used to transform the yeast strain RN1041
15 (selection on medium lacking histidine) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Overexpression of adhE is verified by qPCR.

- For expression of the *Entamoeba histolytica* adh2, a codon optimized synthetic fragment encoding the *E. histolytica* adh2 (SEQ ID NO: 12) is cut with *XbaI* and *AflIII*
20 and ligated into pRN558 cut with *XbaI* and *AflIII* (replacing the *E.coli* mhpF gene in pRN558) to produce pRN596 (SEQ ID NO: 43). pRN596 is used to transform the yeast strain RN1041 (selection on medium lacking histidine) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Overexpression of adh2 is verified by qPCR.

25 Constructs for deletion of GPD2

- Primers GPD2uf, GPD2ur, GPD2df and GPD2dr are used for amplification of genomic sequences fragments upstream and downstream of the GPD2 gene for its inactivation. A 407bp upstream PCR fragment with an *AflIII* site at the 3'-end (derived from the GPD2 sequence) and a *BglII* site at the 5'-end (for isolation of the deletion construct) is
30 amplified using GPD2uf, GPD2ur and cloned in pCR2.1 (topo T/A, Invitrogen).

GPD2uf: GGTACCAGATCTTTTGCGGCGAGGTGCCG (SEQ ID NO: 44)

GPD2_{ur}: TCTAGACTTAAGGAATGTGTATCTTGTTAATCTTCTGACAGC (SEQ ID NO: 45)

A 417bp downstream PCR fragment with a *Xho*I site at the 5'-end and a *Bgl*II site at the 3'-end is amplified using GPD2_{df} and GPD2_{dr}.

GPD2_{df}: CTCGAGATAGTCTACAACAACGTCCGCA (SEQ ID NO: 46)

GPD2_{dr}: CCATGGAGATCTGCAGTGAAAAAGCTCGAAGAAACAGCT (SEQ ID NO: 47)

10

For the final construction the plasmid containing the upstream fragment is cut with *Afl*III and *Kpn*I, the downstream fragment is cut with *Xho*I and *Nco*I and the natMX marker (plasmid collection Royal Nedalco) is cut with *Afl*III and *Xho*I and the fragments are ligated to produce plasmid pRN594 (SEQ ID NO: 48). pRN594 is cut with *Bgl*II prior to yeast transformation. pRN594 is used to transform the yeast strain RN1041 (selection for nourseotricin resistance) together with the other plasmids as indicated in Table 7 to generate the strains as indicated in Table 7. Correct integration is verified by PCR using specific primers.

20 3. Fermentations with the constructed strains

3.1 Experimental set-up

Strains RN1041, RN1088, RN1089, RN1090, RN1091, RN1092, RN1093, RN1094, RN1095, RN1096, RN1097, RN1098, RN1099, RN1194, RN1195 and RN1196. are all tested for their ability to grow under both anoxic conditions (no oxygen present) at the expense of glycerol as carbon source. At the same time, formation of ethanol and formic acid is monitored. In several instances it is confirmed by experiments that the glycerol-consuming strains also retain their ability to ferment glucose and xylose.

Precultures of strains are prepared by inoculating a frozen glycerol stock culture of the yeast in an YP (Yeast extract Peptone) medium with addition of the sugar glucose (2.5% w/v) at 32 °C and pH 5.5. After 24 h incubation, this culture is used to inoculate the fermenter cultures. Cells are harvested by centrifugation and washed with cold dH₂O. Yeast inoculation used is 2 gram dry matter yeast per litre of fermentation medium.

Fermentations are performed as chemostat cultivation at 32 °C in 1-liter fermenters with a working volume of 800 ml. The dilution rate is set at 0.1 h⁻¹. These anoxic batch fermentations are performed in YP medium to which glycerol in combination with xylose and/or glucose is added. The culture pH is kept at pH = 7.0 by automatic
5 addition of 2M KOH. Cultures were stirred at 300 rpm.

Samples for analysis of glucose, xylose, ethanol, glycerol and formic acid production are taken at steady state situations which are obtained after 5 volume changes of the fermentation vessel. Ethanol, glycerol and formic acid concentrations are monitored by HPLC analysis. Glucose and xylose are determined by HPAEC (Dionex) analysis.

10 3.2 Results

Strain RN1041 (Tables 6 and 7) is a strain that is derived from strain RN1001 (RN1001 his3::loxP). Strain RN1001 has been described previously. If either glucose (25 mM) or xylose (25 mM) are added to the YP medium, then the anoxic fermentations under the conditions described above, leads to a full consumption of either glucose or xylose by
15 this strain. The concentrations of ethanol in the glucose and xylose media, respectively, are 40 mM and 33 mM at steady state. Small amounts of glycerol are formed both for glucose and xylose (1.8 and 1.5 mM, respectively). No formic acid formation is observed. In the absence of sugars in the YP medium, no ethanol nor glycerol or formic acid is formed. If glycerol was added to the medium at 50 mM concentration, then the
20 organism did not produce additional ethanol or formic acid as compared to the situations with glucose or xylose alone.

Strain RN1088 was derived from strain RN1041 by

1. upregulating DAK1, GCY1 and GUP1 allowing for enhanced uptake and consumption of glycerol under anoxic conditions
- 25 2. heterologous expression of pflA + pflB allowing for the conversion of pyruvate into formic acid and acetylCoA
3. heterologous expression of adhE2 allowing the conversion of acetylCoA into ethanol
4. deleting FDH1 and FDH2 preventing the oxidation of formic acid by this strain.

30 The organism is tested for growth and product formation in the same way as RN1041 as described above. With either glucose or xylose, similar results are obtained as for RN1041 in terms of ethanol production. However, the formation of glycerol was slightly lower and traces of formic acid are observed. If glycerol is added to either the

xylose- or glucose-containing medium at 50 mM concentration, then the organism does not consume more glycerol or produce additional ethanol and furthermore no additional formic acid is produced. Strains RN1089 and RN1090 have a similar genetic background as RN1088. The only difference is that adhE is not expressed in these organisms. Instead adh2 or mhpF are expressed. The characteristics of these strains during glycerol fermentations are similar to the characteristics of strain RN1088. Both strains showed similar characteristics as RN1088 for both glucose or xylose fermentation as strain RN1088.

Strains RN1091, RN1092 and RN1093 have the same genetic background as strain RN1088 with the exception of formic acid dehydrogenases. These strains have either 1 or 2 genes still present encoding for a formate dehydrogenase. Neither of the 3 organisms is able to perform good fermentations on glycerol. Only traces of either ethanol and formic acid were seen during fermentation experiments. The organisms do ferment both glucose or xylose at rates comparable to strain RN1088.

Strains RN1094, RN1095 and RN1096 have the same genetic background as strain RN1088 with the exception of the expression of either pflA or pflB. These strains lack either both these genes or either of the genes. Neither of the 3 organisms is able to perform good fermentations on glycerol. No additional ethanol or formic acid are seen as based on the addition of 50 mM glycerol to xylose-containing medium (25 mM) during fermentation experiments. The organisms do ferment both glucose or xylose at rates comparable to strain RN1088.

Strain RN1097 is derived from strain RN1088 by deleting GPD2. In doing so, a partial operation of a futile cycle might be prevented at the level of glycerol consumption/production. With glucose or xylose, the ethanol concentration at steady states are 38mM and 31 mM, respectively. Glycerol accumulates at 0.9mM and 0.5mM, respectively and formic acid had accumulates at 1.0 mM and 0.8 mM respectively. If glycerol is added to either the xylose- or glucose-containing medium at 50 mM concentration, then the organism does not consume more glycerol or produce additional ethanol and furthermore no additional formic acid is produced.

Strains RN1098 and RN1099 resemble strain RN1088 but differ genetically at the level of the glycerol transporter(s). Strain RN1098 has FPS1 overexpressed instead of GUP1, whereas strain RN1099 has both transporter genes expressed. The organisms are tested for growth and product formation in the same way as RN1041 as described above. With

either glucose or xylose, similar results are obtained as for RN1041 in terms of ethanol production. However, the formation of glycerol was slightly lower and traces of formic acid were observed. If glycerol was added to either the xylose- or glucose-containing medium at 50 mM concentration, then the RN1098 and RN1099 organisms did not
5 consume more glycerol or produce additional ethanol and furthermore no additional formic acid was produced.

RN1194 is tested for growth and product formation in the same way as RN1041 as described above. With either glucose or xylose, similar results are obtained as for RN1041 in terms of ethanol production. However, the formation of glycerol is slightly
10 lower and traces of formic acid were observed. If glycerol is added to either the xylose- or glucose-containing medium at 50 mM concentration, then the organism produces additional ethanol and furthermore formic acid is produced. The concentration of ethanol at steady state in the glycerol/glucose medium is 75 mM and the formic acid concentration is 32 mM. The remaining glycerol level is 7 mM. The concentration of
15 ethanol at steady state in the glycerol/xylose medium is 66 mM and the formic acid concentration is 30 mM. The remaining glycerol level is 8 mM. No attempts are made to test for any other fermentation products that might have been produced on the basis of glycerol. RN1196 produces similar results as RN1194 although glycerol consumption and production of ethanol and formic acid are somewhat reduced
20 compared to RN1194.

Strain RN1195 is derived from strain RN1194 by deleting GPD2. In doing so, a partial operation of a futile cycle might be prevented at the level of glycerol consumption/production. With glucose or xylose, the ethanol concentration at steady states are 38mM and 31 mM, respectively. Glycerol accumulates at 0.9mM and
25 0.5mM, respectively and formic acid accumulates at 1.0 mM and 0.8 mM, respectively. If glycerol is added to the xylose-medium (25 mM) at 50 mM concentration, then the organism additionally produces both ethanol and formic acid. The final concentration of ethanol at steady state was 72 mM, for glycerol 2 mM, and for formic acid 37 mM. No attempts were made to test for any other fermentation products that might have
30 been produced on the basis of glycerol. These results clearly indicate that deletion of the GPD2 gene has a beneficial effect on the fermentation of glycerol in strain RN1195 as compared to strain RN1194.

Table 6

Gene	GPD2	adhE	ade2	mphF	DAK1	GCY1	GUP1	FPS1	FDH1	FDH2	pflA+pflB	E.c gldA	C.f. dhak
Strain / marker		HIS3	HIS3	HIS3	hphMX	hphMX	hphMX	zeoMX	patMX	kanMX	zeoMX	HIS3	HIS3
RN1041	wt	absent	absent	absent	wt	wt	wt	wt	wt	wt	absent/absent	absent	absent
RN1088	wt	expr	absent	absent	up	up	up	wt	del	del	expr/expr	absent	absent
RN1089	wt	absent	expr	absent	up	up	up	wt	del	del	expr/expr	absent	absent
RN1090	wt	absent	absent	expr	up	up	up	wt	del	del	expr/expr	absent	absent
RN1091	wt	expr	absent	absent	up	up	up	wt	wt	wt	expr/expr	absent	absent
RN1092	wt	expr	absent	absent	up	up	up	wt	del	wt	expr/expr	absent	absent
RN1093	wt	expr	absent	absent	up	up	up	wt	wt	del	expr/expr	absent	absent
RN1094	wt	expr	absent	absent	up	up	up	wt	del	del	absent/absent	absent	absent
RN1095	wt	expr	absent	absent	up	up	up	wt	del	del	expr/absent	absent	absent
RN1096	wt	expr	absent	absent	up	up	up	wt	del	del	absent/expr	absent	absent
RN1097	del	expr	absent	absent	up	up	up	wt	del	del	expr/expr	absent	absent
RN1098	wt	expr	absent	absent	up	up	wt	up	del	del	expr/expr	absent	absent
RN1099	wt	expr	absent	absent	up	up	up	up	del	del	expr/expr	absent	absent
RN1194	wt	expr	absent	absent	wt	wt	wt	wt	del	del	expr/expr	expr	expr
RN1195	del	expr	absent	absent	wt	wt	wt	wt	del	del	expr/expr	expr	expr
RN1196	wt	expr	absent	absent	up	wt	up	wt	del	del	expr/expr	expr	absent

Table 7

	plasmids					
selection:	HIS3	hph	loxPNatMXlox P	nat	kan	zeo
Strain						
RN1041	pRN347					
RN1088	pRN595	pRN605		pRN621	pRN622	pRN616
RN1089	pRN596	pRN605		pRN621	pRN622	pRN616
RN1090	pRN558	pRN605		pRN621	pRN622	pRN616
RN1091	pRN595	pRN605				pRN616
RN1092	pRN595	pRN605		pRN621		pRN616
RN1093	pRN595	pRN605			pRN622	pRN616
RN1094	pRN595	pRN605		pRN621	pRN622	
RN1095	pRN595	pRN605		pRN621	pRN622	pRN619
RN1096	pRN595	pRN605		pRN621	pRN622	pRN620
RN1097	pRN595	pRN605	pRN594	pRN621	pRN622	pRN616
RN1098	pRN595	pRN608		pRN621	pRN622	pRN618
RN1099	pRN595	pRN605		pRN621	pRN622	pRN618
RN1194	pRN957			pRN621	pRN622	pRN616
RN1195	pRN957		pRN594	pRN621	pRN622	pRN616
RN1196	pRN958	pRN607		pRN621	pRN622	pRN616

pRN347 = pRS303+2 μ ORI

Claims

1. A yeast cell comprising:
 - a) an exogenous gene coding for an enzyme with the ability to convert pyruvate and
5 coenzyme-A into formate and acetyl-CoA;
 - b) a genetic modification that reduces specific NAD^+ -dependent formate dehydrogenase activity in the cell;
 - c) an exogenous gene coding for an enzyme with acetaldehyde dehydrogenase activity, which gene confers to the cell the ability to reduce acetylCoA into acetaldehyde; and,
 - 10 d) a genetic modification that increases the specific activity of NAD^+ -linked glycerol dehydrogenase.

2. A yeast cell according to claim 1, wherein the cell further comprises a genetic modification selected from the group consisting of:
 - 15 a) a genetic modification that increases the specific activity of dihydroxyacetone kinase;
 - b) a genetic modification that increases transport of glycerol into the cell;
 - an exogenous gene coding for an enzyme for activation of the pyruvate formate lyase;
 - c) a genetic modification that increases the specific acetyl-CoA synthetase activity in
20 the cell; and,
 - d) a genetic modification that reduces specific NAD^+ -dependent glycerol 3-phosphate dehydrogenase activity in the cell, while the yeast cell comprises a functional high-osmolarity glycerol response pathway.

- 25 3. A yeast cell according to claim 1 or 2, wherein the cell comprises at least one of:
 - i) an exogenous xylose isomerase gene, which gene confers to the cell the ability to isomerise xylose into xylulose; and,
 - ii) an exogenous genes coding for a L-arabinose isomerase, a L-ribulokinase and a L-ribulose-5-phosphate 4-epimerase, which genes together confer to the cell the
30 ability to convert L-arabinose into D-xylulose 5-phosphate.

4. A yeast cell according to any one of claims 1 - 3, wherein at least one of:

- a) the genetic modification that reduces specific NAD^+ -dependent formate dehydrogenase activity in the cell comprises the inactivation of at least one endogenous copy of a gene encoding a formate dehydrogenase in the cell's genome;
- b) the genetic modification that increases the specific activity of the NAD^+ -linked glycerol dehydrogenase is expression of a heterologous gene encoding an NAD^+ -linked glycerol dehydrogenase;
- c) the genetic modification that increases transport of glycerol into the cell is overexpression of a nucleotide sequence encoding at least one of a glycerol uptake protein and a glycerol channel; and,
- d) the genetic modification that reduces specific NAD^+ -dependent glycerol 3-phosphate dehydrogenase activity in the cell comprises the inactivation of at least one endogenous copy of a gene encoding a glycerol 3-phosphate dehydrogenase in the cell's genome and,
- e) the genetic modification that increases the specific acetyl-CoA synthetase activity is overexpression of a nucleotide sequence encoding an acetyl-CoA synthetase; and,
- f) the genetic modification that increases the specific activity of dihydroxyacetone kinase is overexpression of a nucleotide sequence encoding a dihydroxyacetone kinase.

5. A yeast cell according to any one of claims 1 - 4, wherein:
- a) the exogenous gene coding for an enzyme with pyruvate formate lyase activity comprises a nucleotide sequence coding for an amino acid sequence with at least 57% amino acid sequence identity with SEQ ID NO: 1;
- b) the exogenous gene coding for the pyruvate formate lyase activating enzyme comprises a nucleotide sequence coding for an amino acid sequence with at least 70% amino acid sequence identity with SEQ ID NO: 3;
- c) the gene encoding a formate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is a gene encoding a formate dehydrogenase having an amino acid sequence with at least 70% sequence identity to at least one of SEQ ID NO: 5 and 6;
- d) the exogenous gene coding for the enzyme with acetaldehyde dehydrogenase activity comprises a nucleotide sequence coding for an amino acid sequence with at least one of:
- i) at least 64% amino acid sequence identity with SEQ ID NO: 7,

- ii) at least 76% amino acid sequence identity with SEQ ID NO: 9 and
 - iii) at least 61% amino acid sequence identity with SEQ ID NO: 11;
 - e) the nucleotide sequence encoding the NAD⁺-linked glycerol dehydrogenase comprises a nucleotide sequence coding for an amino acid sequence with at least 50% amino acid sequence identity with SEQ ID NO: 49,
 - f) the nucleotide sequence encoding the dihydroxyacetone kinase comprises a nucleotide sequence coding for an amino acid sequence with at least 50% amino acid sequence identity with at least one of SEQ ID NO: 14, 15 and 52;
 - g) the nucleotide sequence encoding the glycerol uptake protein comprises a nucleotide sequence coding for an amino acid sequence with at least 50% amino acid sequence identity with at least one of SEQ ID NO: 16 and 17, and wherein the nucleotide sequence encoding the glycerol channel comprises a nucleotide sequence coding for an amino acid sequence with at least 30% amino acid sequence identity with the amino acid sequence between amino acids 250 and 530 of SEQ ID NO: 18;
 - h) the nucleotide sequence encoding the acetyl-CoA synthetase comprises an amino acid sequence with at least 70% amino acid sequence identity with at least one of SEQ ID NO's: 19 and 20; and,
 - i) the gene encoding a glycerolphosphate dehydrogenase whose activity is to be reduced or inactivated in the cell of the invention is a gene encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 21, while the cell has at least one functional copy of a endogenous gene encoding a glycerolphosphate dehydrogenase having an amino acid sequence with at least 70% sequence identity to SEQ ID NO: 22.
6. A yeast cell according to any one of claims 1 - 5, wherein the yeast cell comprises at least one further genetic modification that results in a characteristic selected from the group consisting of:
- a) increased xylulose kinase specific activity;
 - b) increased flux of the pentose phosphate pathway
 - c) reduced unspecific aldose reductase specific activity
 - d) increased transport of at least one of xylose and arabinose into the host cell;
 - e) decreased sensitivity to catabolite repression;
 - f) increased tolerance to ethanol, osmolarity or organic acids; and,

g) reduced production of by-products.

7. A yeast cell according to any one of claims 1 - 6, wherein the yeast cell is of a genus selected from the group consisting of *Saccharomyces*, *Kluyveromyces*, *Candida*,
5 *Pichia*, *Schizosaccharomyces*, *Hansenula*, *Kloeckera*, *Schwanniomyces*, and *Yarrowia*.

8. A yeast cell according to claim 7, wherein the yeast cell belongs to a species selected from the group consisting of *S. cerevisiae*, *S. exiguus*, *S. bayanus*, *K. lactis*, *K. marxianus* and *Schizosaccharomyces pombe*.
10

9. A yeast cell according any one of the preceding claims wherein at least one of the nucleotide sequences coding for the enzyme with pyruvate formate lyase activity, the pyruvate formate lyase activating enzyme, the enzyme with acetaldehyde dehydrogenase activity, the xylose isomerase, the L-arabinose isomerase, the L-
15 ribulokinase and the L-ribulose-5-phosphate 4-epimerase is a nucleotide sequence the codon usage of which has been optimised for expression in the yeast cell.

10. A process for producing a fermentation product selected from the group consisting of ethanol, lactic acid, 3-hydroxy-propionic acid, acrylic acid, 1,3-propanediol, a butanol and an isoprenoid-derived product, whereby the process comprises the steps of:
20

a) fermenting a medium with a yeast cell as defined in any one of claim 1 - 9, whereby the medium contains or is fed with a source of glycerol and with a source of at least one of a hexose and a pentose, whereby the yeast cell ferments the glycerol and the at least
25 one of the hexose and pentose to the fermentation product and formate; and optionally,
b) recovery of at least one of the fermentation product and formate.

11. A process according to claim 11, wherein the fermentation product is ethanol.

30 12. A process according to claim 10 or 11, wherein the medium contains or is fed with a lignocellulosic hydrolysate.

13. A process according to any one of the claims 10 - 12, wherein the yeast cell ferments under anaerobic conditions.

14. A process according to any one of the claims 10 - 12, wherein the pH of the
5 medium is regulated during the fermentation process to maintain a concentration of undissociated formic acid that is not higher than 20.0 mM.