Title: UREOHYDROLASES AS DOMINANT SELECTABLE MARKERS IN YEAST

Abstract: The invention relates to a nucleic acid molecule encoding a novel selection marker. Said marker is a guanidinobutyrase from Kluyveromyces lactis, which, when expressed in Saccharomyces, allows the growth of the yeast in the presence of guanidinobutyrate as the sole nitrogen source. Said marker can be used in a method for producing a microorganism having an altered genome. The invention further relates to a set of constructs, comprising a first construct comprising a recognition site for an endonuclease, a first region of homology with a target gene of a microorganism, and a first part of a nucleotide sequence encoding the selection marker, and a second construct comprising a second part of the nucleotide sequence encoding the selection marker, and a region of homology with the target gene of the microorganism, and a copy of the endonuclease recognition site. The invention further relates to methods for altering a target gene in a microorganism, to methods for producing a microorganism, and to microorganisms that are produced by the methods of the invention.
Title: Ureohydrolases as dominant selectable markers in yeast.

Field

The invention relates to the fields of molecular biology and genetic engineering of microorganisms, especially of yeast.

Introduction

Arginine metabolism has been subject of intensive biochemical studies. After discovery of the urea cycle for use of arginine as a nitrogen source (Krebs and Henseleit, (1932) Wochenschrift 11: 757-759; Krebs, (1973) Biochemical Education 1: 19-23), attention focused on its role as a precursor for the synthesis of polyamine and the signaling compounds γ-aminobutyrate (GABA) and nitric oxide (Knowles and Moncada (1994) Biochem J 298: 249-258; Pitkanen et al., (2001) Biochem Biophys Res Commun 287: 1051-1057). The most widely distributed pathway for arginine degradation that occurs across all three kingdoms (Abdelal, (1979) Annu Rev Microbiol 33: 139-168) is initiated by arginase (EC 3.5.3.1), an ureohydrolase that converts arginine to ornithine and urea. Its active site, which contains several Mn²⁺-binding sites, is also conserved in other ureohydrolases such as agmatinase (EC 3.5.3.11), formiminoglutamase (EC 3.5.3.8) and proclavaminate amidinohydrolase (EC 3.5.3.22), guanidinobutyrase (EC 3.5.3.7) and guanidinopropionase (EC 3.5.3.17) (Ouzounis and Kyrpides, (1994) J Mol Evol 39: 101-104). Genes encoding these enzymes are assumed to have emerged early in evolution (Hartman, (1975) Orig Life 6: 423-427) and have been used as markers in phylogenetic studies (Ouzounis and Kyrpides (1994) J Mol Evol 39: 101-104; Sekowska et al., (2000) Microbiology 146: 1815-1828).

In eukaryotes, only two types of ureohydrolase have hitherto been described. In addition to arginase, higher eukaryotes express agmatinase (Coleman et al., (2004) Biochem J 379: 849-855), which participates in an alternative pathway for arginine catabolism (Figure 1). In this pathway, arginine
is first decarboxylated to agmatine, which is converted to putrescine and urea by agmatinase. Putrescine can then either be converted to GABA or to the polyamines spermine and spermidine (Pegg, (2009) IUBMB Life 61: 880-894). The rapidly increasing number of whole genome sequences has enabled the putative identification of arginase and agmatinase genes in many eukaryotes. However, since such annotation is based on sequence homology only, it does not enable definitive conclusions on the catalytic function of the encoded proteins.

Much of the knowledge on fungal arginine metabolism is based on studies with the model organism Saccharomyces cerevisiae. In S. cerevisiae, arginine is transported into the cell and subsequently hydrolyzed by arginase (Car1) to yield ornithine and urea (Sumrada and Cooper, (1992) Yeast 8: 311-314; Cooper et al., (1992) J Bacteriol 700 174: 48-55; Shima 64 et al., (2003) Appl Environ Microbiol 69: 715-718). An ATP-dependent amidolyase (Dur1,2) then converts urea into ammonia and carbon dioxide. Ornithine is further converted by an ornithine specific transaminase (Car2) into glutamate-γ-semialdehyde (GSA), which spontaneously forms 1-pyrroline-5-carboxylate (P5C) (Martin et al., (2003) Appl Environ Microbiol 69: 1623-1628). Due to subcellular compartmentation, S. cerevisiae is unable to convert cytosolic P5C directly to glutamate (Davis, (1986) Microbiol Rev 50: 280-313). Instead, P5C is reduced to proline using pyrroline-5-carboxylate reductase (Pro3). Proline is then transported into the mitochondria (Brandriss and Falvey, (1992) Bacteriol 174: 5176), converted back to P5C by an oxidase (Put1) and, finally converted to glutamate by mitochondrial P5C dehydrogenase (Put2) (Davis, (1986) Microbiol Rev 50: 280-313). Since only very few physiological studies have been conducted on arginine metabolism in non-Saccharomyces yeasts, it is unknown whether the arginase pathway, which is essential for growth of S. cerevisiae on arginine as sole nitrogen source (Bossinger and Cooper, (1977) J Bacteriol 131: 163-173, is the only fungal pathway for arginine catabolism.

S. cerevisiae and Kluyveromyces lactis both belong to the Saccharomycetaceae family. These two related yeasts are considered to have
genetically separated before the whole genome duplication (WGD) event that reshaped the genome of *S. cerevisiae*, furthermore *K. lactis* is regarded as resembling a pre-WGD ancestor of *S. cerevisiae* (Dujon, (2010) Nat Rev Genet 11: 512-524). While many studies have been conducted on the differences in sugar metabolism between these two species, the differences in amino-acid metabolism have not been studied in detail. Nonetheless, the complete genome sequence of *K. lactis* revealed many putative orthologs of *S. cerevisiae* genes involved in arginine metabolism (Dujon et al., (2004) Nature 430: 35-44; Souciet et al., (2009) BMC Genomics 13: 517; Dias et al., (2012) BMC Genomics 13: 517).

The selection of a microorganism that is transformed with recombinant DNA is strongly facilitated by the use of a suitable selection marker. The molecular biologist working with *Saccharomyces cerevisiae* has access to a large number of selectable markers (Solís-Escalente et al., (2013) FEMS Yeast Research 13: 126–139). However, auxotrophic markers and antibiotic resistance markers are sometimes undesired, thereby vertiginously decreasing the number of suitable markers. Eventually, the molecular geneticist is left with a limited number when he wants to genetically access wild type, allopolyploid and/or aneuploid prototrophic yeast strains.

The study of arginine metabolism resulted in the identification of two new “gain of function” dominant markers, which can be used, for example, for introducing genomic alterations in microorganisms, preferably in laboratory, wild and industrial yeast strains, including *S. cerevisiae* strains.

Therefore, the invention provides a nucleic acid molecule comprising (a) a nucleotide sequence encoding a guanidinobutyrase selection marker; and/or (b) a nucleotide sequence encoding a guanidino-amide hydrolase selection marker, whereby the nucleotide sequence is operably linked to (heterologous) promoter and terminator sequences. This nucleic acid molecule, also termed dominant marker cassette, provides a convenient dominant selectable marker system suitable for use in microorganisms, preferably in yeast.
Guanidinobutyrase (guanidino-acid hydrolase) and guanidino-amide hydrolase belong to a larger protein family, the ureohydrolase, which comprises enzymes sharing a 3-layer alpha-beta-alpha structure and play important roles in arginine/agmatine metabolism, the urea cycle and histidine degradation.

Guanidinobutyrase catalyzes the hydrolysis of 4-guanidinobutanoate into 4-aminobutanoate and urea. Guanidino-amide hydrolase hydrolyses agmatine to urea and putrescine, the precursor for the biosynthesis of polyamines, spermidine and spermine. The ureohydrolase family comprises three further subgroups: i) the guanidino-amino acid hydrolase (arginase, EC3.5.3.1) that catalyses the conversion of arginine to urea and ornithine, ii) the proclavaminate amidohydrolase (EC 3.5.3.22), an activity that catalyses the conversion of amidino proclavaminate into urea and proclavaminate, an intermediate in clavulanic acid biosynthesis and iii) the formiminoglutamase (EC 3.5.3.8) that catalyzes the fourth step in histidine degradation, by hydrolysing N-formimidoyl-L-glutamate to L-glutamate and formamide.

Said guanidinobutyrase-encoding nucleotide sequence encodes guanidino-acid hydrolase (EC.3.5.3.7). The nucleotide sequence preferably encodes a guanidino-acid hydrolase having one or both conserved domains as indicated in Figure 2A and depicted in Figure 2B, and/or encodes the consensus sequence as depicted in Figure 2A. Said guanidino-acid hydrolase preferably comprises between 350 and 450 amino acids, preferably about 400 amino acids. Said guanidinobutyrase-encoding nucleotide sequence preferably encodes *Kluyveromyces lactis* NRRL Y-1140 hypothetical protein, having the sequence of GenBank XP_456325.1, as depicted in Figure 3.

Said guanidino-amide hydrolase-encoding nucleotide sequence encodes agmatine ureohydrolase (agmatinase) (EC.3.5.3.11). The nucleotide sequence preferably encodes an agmatine ureohydrolase having the conserved domain as indicated in Figure 4A and depicted in Figure 4B, and/or encodes the consensus sequence as depicted in Figure 4A. Said agmatine ureohydrolase preferably comprises between 290 and 330 amino acids, preferably about 306 amino acids.
Said agmatine ureohydrolase preferably encodes the sequence of GenBank AAC75974.1, as depicted in Figure 5.

Both guanidinobutyrase and agmatinase catalyze the formation of urea, a nitrogen source commonly assimilated by microorganisms such as *S. cerevisiae*. Therefore, these two ureohydrolase genes present the essential characteristics of a potential dominant “gain of function” selectable marker in microorganisms such as *S. cerevisiae*, when grown on guanidinobutyrate and/or agmatine as sole nitrogen source.

The invention thus provides a method of culturing a microorganism of the genus *Saccharomyces* in the presence of guanidinobutyrate as sole nitrogen source, comprising (a) introducing a nucleic acid molecule comprising a nucleotide sequence encoding a guanidinobutyrase into the microorganism, whereby the nucleotide sequence is operably linked to promoter and terminator sequences, (b) culturing the microorganism such that the nucleic acid molecule encoding the guanidinobutyrase is expressed in the microorganism, and (c) culturing the microorganism in the presence of guanidinobutyrate as sole nitrogen source.

Said guanidinobutyrase-encoding nucleotide sequence preferably encodes *Kluyveromyces* lactis NRRL Y-1140 hypothetical protein.

Said promoter directs expression of the selection marker in the microorganism. Said terminator mediates efficient mRNA 3’ end formation. Said promoter preferably is a yeast promoter, more preferably a yeast promoter selected from a glycolytic gene such as *PGI1* (phosphoglucone isomerase 1), *PFK1* (phosphofructokinase-1), *PFK2* (phosphofructokinase-2), *FBA1* (fructose-bisphosphate aldolase-1), *TPI1* (triosephosphate isomerase-1), *TDH1* (glyceraldehyde-3-phosphate dehydrogenase 1), *TDH3* (glyceraldehyde-3-phosphate dehydrogenase3), *PGK1* (phosphoglycerate kinase 1), *GPM1* (glycerate phosphomutase 1), *PYK1* (pyruvate kinase 1), *ENO1* (alpha-enolase), and/or *ENO2* (enolase 2) promoter, or selected from *ACT1* (actin 1), *TEF1* (translational elongation factor EF-1 alpha), *TEF2* (translational elongation factor EF-1 alpha
2), AgTEF2 (Ashbya gossypii TEF2 gene), PMA1 (plasma membrane P2-type H+-ATPase) promoter. Preferred promoter sequences are selected from promoter sequences of the PGI1, FBA1; TPI1; TDH3; PGK1; GPM1; ENO1; ENO2; and PYK1 genes. Terminators from a number of genes are known to the skilled person and have been employed, for example in expression vectors, including CYC1, TRP1, ADH1, MFI, FLP and D gene terminators (Romanos et al., 1992. Yeast 8: 423-488).

Preferred promoter sequences and terminator sequences do not comprise Saccharomyces sequences. Preferred promoter sequences and terminator sequences are from the Ashbya gossypii TEF gene encoding translation elongation factors and/or the regulatory sequences from K. lactis URA3 and LEU2 genes that encode the orotidine-5'-phosphate (OMP) decarboxylase and the beta-isopropylmalate dehydrogenase, respectively.

A nucleic acid molecule according to the invention, preferably encoding a guanidinobutyrase, is preferably provided as an amplified product or as a part of an amplified product. Said amplified product may further comprise sequences homologous to a first part of the genome of a microorganism, preferably an upstream part of a gene that is to be altered (termed target gene), and sequences homologous to second part of the genome of a microorganism, preferably a downstream part of the target gene, adjacent to the nucleic acid molecule. The term adjacent is used to indicate that the sequences homologous to the first part of the genome are located on one side of the nucleic acid molecule, while sequences homologous to the second part of the genome are located on the other side of the nucleic acid molecule. The sequences homologous to the first and second part of the genome preferably comprise between 20 and 1000 bp, more preferred between 30 and 500 bp, more preferred between 40 and 250 bp, more preferred between 50- to 80-bp of sequences. The sequences homologous to the upstream part of a gene preferably include the start codon. The sequences homologous to the downstream part of a gene preferably include the stop codon. Methods for producing an amplified product comprising a nucleic acid molecule according to the invention are known to a skilled person including, for example,
polymerase chain reaction and nucleic acid sequence based amplification (NASBA).

A nucleic acid molecule according to the invention is preferably comprised in a vector. A vector contains bacterial resistance genes that, for example, allow growth of bacteria in the presence of an antibiotic. A most preferred vector is a plasmid, a double-stranded DNA molecule that is capable of replicating in bacteria independent of the chromosomal DNA.

Said vector, preferably plasmid, preferably additionally comprises sequences homologous to a first part of the genome of a microorganism, preferably an upstream part of a gene that is to be altered (termed target gene), and sequences homologous to second part of the genome of a microorganism, preferably a downstream part of the target gene, adjacent to the nucleic acid molecule. The term adjacent is used to indicate that the sequences homologous to the first part of the genome are located on one side of the nucleic acid molecule, while sequences homologous to the second part of the genome are located on the other side of the nucleic acid molecule.

The sequences homologous to the first part of the genome preferably comprise between 20 and 1000 bp, more preferred between 30 and 500 bp, more preferred between 40 and 250 bp, more preferred between 50- to 80-bp of sequences that are homologous to a first part of the genome, preferably to an upstream part of the gene to be altered or deleted. The sequences homologous to the upstream part of a gene preferably include the start codon.

The sequences homologous to the second part of the genome preferably comprise between 20 and 1000 bp, more preferred between 30 and 500 bp, more preferred between 40 and 250 bp, more preferred between 50- to 80-bp of sequences that are homologous to a second part of the genome, preferably to the downstream part of the gene to be altered or deleted. The sequences homologous to the downstream part of a gene preferably include the stop codon.
Said sequences that are homologous to a first and/or second part of the genome, preferably a target gene, may comprise sequences that are altered, when compared to the sequences of the genome. The terms altering, alteration and altered refer to a replacement of one or more nucleotides, the insertion of one or more nucleotides, and/or the deletion of one or more nucleotides anywhere within the homologous sequences.

A replacement of one or more nucleotides can be accomplished by altering one or more nucleotides in first part and/or the second part, preferably in sequences that are homologous to an upstream and/or downstream part of a gene. When the first part of homology and the second part of homology cover adjacent regions on the genome, preferably on a target gene, the integration of the targeting vector will result in an alteration of the genome.

Said vector, preferably plasmid, preferably further comprises a recognition site for an endonuclease at one end of the nucleic acid molecule according to the invention, and a copy of this recognition site at the other end of the nucleic acid molecule according to the invention. Said recognition sites preferably are located directly adjacent to the nucleic acid molecule according to the invention. Said endonuclease preferably is a rare-cutting endonuclease such as, for example, PacI (target recognition sequence 5’-TTAATTAA); AscI (target recognition sequence 5’-GGCGCGCC), and AsiSI (target recognition sequence 5’-GCGATCGC). PacI, AscI and AsiSI are available from New England Biolabs. The endonuclease more preferably is a homing endonuclease. The term homing endonuclease refers to an endonucleases that is encoded either as freestanding genes within introns, as a fusion with a host protein, or as a self-splicing intein. A preferred list of homing endonucleases is provided in Table 1. Additional examples of homing nucleases are provided in Benjamin K (patent application US2012/052582), which is enclosed herein by reference.
Said vector, preferably plasmid, preferably further comprises a nucleic acid sequence of between 20 and 200 bp, preferably between 30 and 100 bp, more preferred about 40-50 bp, that is duplicated on either side of the nucleic acid molecule according to the invention, preferably outside of the recognition sites for an endonuclease. Said nucleic acid sequence of between 20 and 200 bp preferably is located in between the sequences homologous to the upstream part of a target gene and the recognition site for an endonuclease at one end of the nucleic acid molecule according to the invention, and in between the sequences homologous to the downstream part of the target gene and the recognition site for an endonuclease at the second end of the nucleic acid molecule according to the invention. The duplicated nucleic acid sequence preferably is identical to a region on the target genome, preferably on the target gene. The duplicated nucleic acid sequence of between 20 and 200 bp ensures seamless marker removal from the target genome by homologous recombination.

The invention further provides a method for producing a microorganism, preferably a yeast, comprising the nucleic acid molecule of the invention, the method comprising providing the microorganism with the nucleic acid molecule of the invention, selecting a microorganism having said nucleic acid molecule, thereby producing a microorganism comprising the nucleic acid molecule of the invention.

Methods for selecting a microorganism, preferably a yeast, having the nucleic acid according to the invention are known in the art and include Southern blotting and amplification of a nucleic acid product comprising at least a part of the nucleic acid molecule using at least one primer that is specific for the nucleic acid molecule. Alternatively, a vector that comprises the nucleic acid molecule according to the invention preferably further comprises a selectable marker that allows selection of a microorganism comprising the nucleic acid molecule of the invention.
The term specific, as used herein, refers to a primer or polynucleotide that will hybridize only to its target subsequence, typically in a complex mixture of nucleic acids, but to essentially no other sequences applying stringent conditions as is known to the skilled person. Stringent conditions are sequence-dependent and will be different in different circumstances. An extensive guide to the hybridization of nucleic acids is found in Tijssen (Tijssen, (1993) Hybridization with Nucleic Acid Probes, vol. 2, Laboratory techniques in biochemistry and molecular biology, Volume 24. Elsevier, Amsterdam). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the primers complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium).

Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., 10 to 50 nucleotides) and at least about 60°C for long probes (e.g., greater than 50 nucleotides). For selective or specific hybridization, a positive signal is at least two times background, preferably 10 times background hybridization.

A preferred method of the invention comprises selection of a microorganism, preferably a yeast, that functionally expresses the nucleic acid molecule encoding the selection marker. For example, a yeast may be selected that expresses guanidinobutyrase and/or agmatinase, as determined in an enzyme activity assay. Methods to determine guanidinobutyrase or agmatinase activity are known in the art. The amount of urea produced can be quantified using, for example, the Archibald method (Archibald, (1945) J Biol Chem 157: 507-518). A calibration line ranging from 0 - 1 mM of urea can be generated for quantification of the amount of urea.
The term “functionally express” in this specification is used to indicate that a gene of interest expresses the protein that is encoded by the gene, in an active form.

Said microorganism, preferably yeast, preferably of the genus *Saccharomycetaceae*, is preferably selected as it is able to grow on a substrate, preferably a synthetic medium, comprising, as sole nitrogen source, guanidinobutyrate and/or agmatine.

The invention further provides a microorganism, preferably yeast comprising the nucleic acid molecule of the invention. Said nucleic acid molecule of the invention preferably is integrated into the genome of said microorganism.

Said microorganism preferably is of the genus *Saccharomycetacea* and comprises a nucleotide sequence encoding a guanidinobutyrase, preferably a guanidinobutyrase-encoding nucleotide sequence encoding *Kluyveromyces lactis* NRRL Y-1140 hypothetical protein.

The method of the invention can be applied to all microorganisms. If a microorganism has endogenous guanidinobutyrase and/or agmatinase activity, mutants of this organisms can be provided in which this activity is inactivated. Methods to inactivate a gene encoding guanidinobutyrase or agmatinase in a microorganism are known in the art. Suitable methods have been described in, for example, Akada et al., (2002). Yeast 19: 393–402; McNabb et al., (1997). Biotechniques, 22: 1134–1139; Storici et al., (1999). Yeast 15: 271–283; Gueldener et al., (2002). Nucleic Acids Res 30: e23; and Iwaki and Takegawa, (2004). Biosci Biotechnol Biochem 68: 545–550. Any of these methods can be applied to generate a microorganism in which an endogenous gene encoding guanidinobutyrase and/or agmatinase can be functionally inactivated.

The term functionally inactivated is used herein to indicate a reduced functional presence of a protein product of a gene in a microorganism, which is due to either a reduced level of expression or a reduced level of activity of the
protein. Said reduced functional presence preferably results in a reduction of more than 90% of the protein amount and/or activity, more preferred a reduction of more than 95% of the protein amount and/or activity, most preferred a reduction of more than 99% of the protein amount and/or activity, compared to the corresponding protein activity in a related microorganism not comprising the functionally inactivated gene. In a most preferred embodiment, a functionally inactivated gene has no residual activity and is equivalent to a knock-out gene. The term knock-out gene refers to gene that has been made functionally inactive by partial or complete deletion of the coding region from the genomic DNA encoding said gene.

A microorganism, preferably yeast, of the invention preferably is a yeast of the genus Saccharomycetaceae. This genus includes Saccharomyces sensu stricto, Kazachstania, Naumovozyma, Nakaseomyces and Vanderwaltozyma. It has been proposed that these genera belonging to Saccharomycetaceae family have arisen after the whole genome duplication (post WGD) event that played a major role in the evolution of this subphylum.

A comparison of the nitrogen metabolism between Saccharomyces cerevisiae and Kluyveromyces lactis revealed substantial differences in arginine assimilation between these two yeast species. While deletion of the single S. cerevisiae arginase gene CAR1 was sufficient to abolish growth on arginine as sole nitrogen source, the corresponding Klcar1Δ mutant in K. lactis was still able to grow on arginine. This phenotypic difference was caused by the presence in K. lactis of a gene (KLLA0F27995g/KlGBU1) encoding for a guanidinobutyrase (EC.3.5.3.7), an enzyme not previously demonstrated in fungi. The presence of this enzyme provides the ability to grow on guanidinobutyrate as sole nitrogen source.

It was found that guanidinobutyrase orthologs are not present in the genus Saccharomycetaceae post WGD. Remarkably, one out of 34 annotated S. cerevisiae genome sequences, available from the SGD database
(www.yeastgenome.org/) belonging to strain EC1118, did harbour a gene whose predicted protein sequence shared 62% sequence identity with a guanidinobutyrase sequence of *Kluyveromyces lactis*. Further inspection of the genome structure of strain EC1118 revealed that this ortholog was found on a genomic region acquired by horizontal transfer, most probably from *Torulaspora delbrueckii*. This gene origin fits the phylogenetic distribution of guanidinobutyrase orthologs in pre WGD Saccharomycetaceae genera (*Zygosaccharomyces, Lachancea, Torulaspora, Kluyveromyces and Erethrothecium*), and in genera belonging to the CTG group and *Dipodascaceae* (Kurtzman, (2003) FEMS Yeast Res 4: 233-245; Dujon, (2010) Nat Rev Genet 11: 512-524).

A preferred yeast is a laboratory, wild and industrial Lager brewing yeast *Saccharomyces pastorianus*, a laboratory, wild and industrial *S. cerevisiae* strain, preferably a *Saccharomyces sensu stricto* (*Saccharomyces paradoxus, S. mikatae, S. bayanus, S. eubay anus, S. kudriavzevii, S. paradoxus, S. arboriculus*), or a strain of *Kazachstania, Naumovozyma, Nakaseomyces* or *Vanderwaltozyma*, most preferred a laboratory, wild and industrial Lager brewing yeast *Saccharomyces pastorianus* or a *S. cerevisiae* strain.

The invention additionally provides a method of altering the genome of a microorganism, preferably a yeast genome, preferably a yeast of the genus Saccharomycetaceae, comprising providing the nucleic acid molecule of the invention, preferably encoding guanidinobutyrase, more preferably encoding *Kluyveromyces lactis* NRRL Y-1140 hypothetical protein, to said microorganism, and selecting a microorganism in which the genome has been altered by insertion of the nucleic acid molecule of the invention into the genome. A preferred method comprises providing a set of constructs according to the invention to said microorganism, and selecting a microorganism in which the genome has been altered, preferably by selection of a microorganism that functionally expresses said guanidinobutyrase. A preferred method comprises selecting a microorganism by culturing in the presence of guanidinobutyrate as sole nitrogen source.
The present invention further provides a set of constructs, comprising a first construct comprising a first part of the nucleotide sequence encoding a selection marker as indicated in claim 1, and a second construct comprising a second part of the nucleotide sequence encoding a selection marker as indicated in claim 1, whereby a fragment of the first part of the selection marker overlaps with a fragment that is present in the second part of the nucleotide sequence, allowing recombination between the first and second part of the nucleotide sequence.

This set of constructs overcomes a low targeting efficiency by providing a set of targeting constructs, in which the correct expression of a selection marker depends on a recombination event between the targeting constructs. It was found that the occurrence of a recombination event between the targeting constructs is markedly enhanced after integration of the targeting constructs in the correct targeting locus. Therefore, the target system of the present invention, comprising a set of targeting constructs, greatly enhances the percentage of correctly integrated constructs in microorganisms that express the selection marker, compared to a one-vector targeting system. Splitting the marker on two separate constructs limits the occurrence of false positives due to single cross over events.

The split marker approach improves the ratio of true positives over false positives (Nielsen et al., 2006. Fungal Gen Biol 43: 54-64).

The term construct or targeting construct, as used herein, refers to an artificially constructed segment of nucleic acid. A preferred construct is a vector, preferably a vector that contains bacterial resistance genes for growth in bacteria. A most preferred construct is a plasmid, a linear or circular double-stranded DNA that is capable of replicating in bacteria independently of the chromosomal DNA.

The term overlap, as is used herein, refers to a duplicated region of the nucleotide sequence encoding a selection marker that is present on both set of constructs. The duplicated region is substantially identical and preferably is
between 40 and 400 bp, preferably about 200 bp. The term substantially, as is used herein, is used to indicate that the region is at least 90% identical to ensure efficient recombination between the targeting constructs, more preferred at least 95% identical, more preferred at least 99% identical, more preferred 100% identical.

The first construct in the set of targeting constructs preferably further comprises a recognition site for an endonuclease and a first region of homology with a target genome of a microorganism, and the second construct further comprises a second region of homology with the target genome of the microorganism, and a copy of the endonuclease recognition site, whereby a coding sequence that encodes the endonuclease and which is coupled to an inducible promoter is present on the first or second construct; and a part of the first region of homology with the target genome on the first construct is duplicated between the copy of the endonuclease recognition site and the second region of homology with the target genome on the second construct; or a part of the second region of homology with the target genome on the second construct is duplicated between first region of homology with the target genome and the endonuclease recognition site on the first construct.

Said duplicated region of homology with the target genome on the first and second targeting construct preferably is between 20 and 200 bp, preferably between 40 and 100 bp, preferably about 80 bp. Said duplicated region of homology with the target genome on the first and second targeting construct allows scarless removal of the marker from the target genome by homologous recombination.

The first construct preferably comprises, in this order, a first region of homology with a target genome of a microorganism, a recognition site for an endonuclease, and a first part of a selection marker. The second construct preferably comprises, in this order, a region of overlap with the first part of the nucleotide sequence encoding a selection marker followed by a second part of the
nucleotide sequence encoding the selection marker, a coding sequence that encodes the endonuclease and which is coupled to an inducible promoter, a copy of the endonuclease recognition site, a copy of a part of the first region of homology with the target genome that is present on the first construct, and a second region of homology with the target genome of the microorganism. This configuration is depicted in Figure 6.

The target genome can be any location, preferably a gene, on the genome of a microorganism, preferably of a yeast, preferably of a yeast of the genus Saccharomyces, of which the genomic sequence is to be altered. The term gene, as is used herein, refers to a part of the genome of the microorganism that comprises intronic and exonic parts of a gene, the promoter region of said gene, and genomic sequences that mediate the expression of said gene, such as, for example enhancer sequences.

The skilled person will understand that the targeting constructs are preferably used to alter a gene of a microorganism. Hence, the invention further provides a set of targeting constructs, comprising a first construct comprising a first region of homology with a target gene of a microorganism, a recognition site for an endonuclease, and a first part of a nucleotide sequence encoding a selection marker, and a second construct comprising a region of overlap with the first part of the nucleotide sequence encoding the selection marker followed by a nucleotide sequence encoding a second part of the selection marker, a copy of the endonuclease recognition site and a second region of homology with the target gene of the microorganism, whereby the overlapping fragments allow recombination between the first and second part of the nucleotide sequence encoding the selection marker; whereby a coding sequence that encodes the endonuclease and which is coupled to an inducible promoter is present on the first or second construct; and whereby a part of the first region of homology with the target gene on the first construct is duplicated between the copy of the endonuclease recognition site and the second region of homology with the target gene on the second construct; or a part of the second region of homology with the
target gene on the second construct is duplicated between first region of homology with the target gene and the endonuclease recognition site on the first construct.

5 Said duplicated region of homology with the target gene on the first and second targeting construct preferably is between 20 and 200 bp, preferably between 40 and 100 bp, preferably about 80 bp.

The term alteration of the genomic sequence includes a replacement of one or more nucleotides, the insertion of one or more nucleotides, and/or the deletion of one or more nucleotides anywhere within a genome, preferably within a gene.

For example, if the first and second region of homology with a target gene comprise adjacent genomic sequences of the gene, a replacement of one or more nucleotides in the first region of homology, and/or in the second region of homology, will result in an alteration of the gene following homologous targeting with the set of targeting constructs according to the invention. Said replacement of one or more nucleotides preferably is in the region of homology with the target gene that is present on the first and on the second construct.

Said alteration of the genomic sequence preferably is a deletion of one or more nucleotides, preferably anywhere within the gene. For example, if the first and second region of homology with a target gene comprise genomic sequences of the gene that are separated on the genome of the organism, an alteration of the gene following homologous targeting with the set of targeting constructs according to the invention will result in a deletion of the region that was located between the first and second region of homology on the parental chromosome.

30 Said first construct preferably comprises a first part, preferably the first two-third or first half, of a region that encodes the selection marker. For example, the guanidinobutyrase protein of \textit{K. lactis} has 410 amino acids, which is encoded
by a nucleic acid sequence of 1230 bp. Said first construct preferably comprises between 400 and 800 bp of the coding region of this protein, more preferred between 500 and 700 bp. The second construct preferably comprises between 400 and 800 bp of the coding region of this protein, more preferred between 500 and 700 bp.

The region of overlap between the first and second part of the selection marker preferably is between about 50 bp and about 600 bp, preferably about 200 bp.

The first or second targeting construct comprises a coding sequence that encodes an endonuclease and which is coupled to an inducible promoter. The endonuclease preferably is a rare-cutting endonuclease such as, for example, PacI (target recognition sequence 5'-TTAATTAA); AscI (target recognition sequence 5'-GGCGCGCC), and AsISI (target recognition sequence 5'-GCGATCGC). PacI, AscI and AsISI are available from New England Biolabs. The endonuclease more preferably is a homing endonuclease. The term homing endonuclease refers to an endonucleases that is encoded either as freestanding genes within introns, as a fusion with a host protein, or as a self-splicing intein. A preferred list of homing endonucleases is provided in Table 1. Additional examples of homing nucleases are I-Dirl, I-NjaI, I-NanI, I-NitI, F-TevI, F-TevII, F-CphI, PI-MgaI, I-CsmI, which are all known to the skilled person. Further examples of homing nucleases are provided in US patent application US 2012/052582, which is enclosed herein by reference.

A preferred homing nuclease is PI-PspI (New England Biolabs; recognition sequence 5’-TGGCAAAACAGCTATTATGGGTATTATGGGT) or PI-SceI (New England Biolabs; recognition sequence 5’-ATCTATGTCGGGTGGGAGAAAGGGTAAT). The coding sequences of most homing endonuclease are known. For example, the coding sequence of PI-SceI and of PI-PspI are available from public databases (GenBank accession number Z74233.1 and Genbank accession number U00707.1, respectively). The skilled
person will understand that a sequence that differs from the publicly available sequence for a nuclease, may still encode the nuclease. For example, the term PI-PspI coding region may include a sequence that deviates from the publicly available sequence, for example by codon optimization, but which still expresses an active endonuclease that recognizes and digests the indicated target recognition sequence.

Said endonuclease is preferably under control of an inducible promoter. The term inducible promoter, as is used herein, refers to a promoter of which the expression can be regulated. Inducible promoters are known to the skilled person. Examples of inducible promoters that have been employed in yeast are the GAL1 promoter and the GAL10 promoter, which both are inducible by galactose, the SUC2 promoter, which is inducible by sucrose, the MAL12 promoter, which is inducible by maltose; the CUP1 promoter, which is inducible by copper, and the tetO7 and tetO2 promoters, which are both inducible by tetracycline [Gari et al., (1997) Yeast 13: 837-48; Yen et al., (2003) Yeast 20 1255-62]. A preferred inducible promoter is the GAL1 promoter.

One recognition site comprising the target recognition sequence for the endonuclease, is located adjacent to (behind) the first region of homology with a target gene of a microorganism on the first construct. A copy of this recognition site is located adjacent to (in front of) the second region of homology with the target gene of the microorganism on the second construct. The skilled person will understand that when a part of the first region of homology with the target gene on the first construct is duplicated between the copy of the endonuclease recognition site and the second region of homology with the target gene on the second construct, said copy of the recognition site is located adjacent to (in front of) the duplication of the first region of homology with the target gene on the second construct. Alternatively, the recognition site is located adjacent to (behind) the duplicated part of the second region of homology with the target gene on the first construct when a part of the second region of homology with the target gene on the second construct is duplicated on the first construct. The
selection marker, including promoter and terminator sequences, and the coding region of the endonuclease, including the inducible promoter, are between the recognition site on the first construct and the copy of this recognition site on the second construct.

5

The invention further provides a method for altering a genome, preferably a target gene, in a microorganism, comprising providing the set of targeting constructs according to the invention to said microorganism, and selecting a microorganism in which the genome has been altered. Said selection of a microorganism in which the genome has been altered is preferably accomplished by selection of a microorganism that functionally expresses a recombined selection marker.

As is indicated herein above, the occurrence of a recombination event between the targeting constructs is markedly enhanced after integration of the targeting constructs in the correct targeting locus. Hence, the presence of a functionally recombined selection marker is highly indicative for the presence of correctly integrated targeting constructs in the target genome and, therefore, of an altered genome in the microorganism.

As is indicated herein above, the terms altering, alteration and altered refer to a replacement of one or more nucleotides, the insertion of one or more nucleotides, and/or the deletion of one or more nucleotides anywhere in the genome, preferably within a target gene.

A replacement of one or more nucleotides can be accomplished by altering one or more nucleotides in the first region of homology and/or in the second region of homology. When the first region of homology and the second region of homology with the target genome cover adjacent regions of the genome, preferably target gene, the integration of the targeting vectors will result in an alteration of the genome. When present, said replacement of one or more nucleotides is preferably accomplished by altering one or more nucleotides in the
overlapping region of homology with the genome that is present on the first and on the second construct.

Said alteration of a genomic sequence preferably is a deletion of one or more nucleotides anywhere within a genome, preferably within a gene. For example, if the first and second region of homology with a target genome comprise genomic sequences that are separated on the genome of the organism, an alteration of the genome following homologous targeting with the set of targeting constructs according to the invention will result in a deletion of the region that was located between the first and second region of homology on the parental chromosome.

The invention further provides a method for producing a microorganism comprising an altered genome, preferably an altered gene, the method comprising providing the set of targeting constructs according to the invention to said microorganism, and selecting a microorganism in which the genome has been altered and that functionally expresses a recombined selection marker.

The method for producing a microorganism comprising an altered genome preferably comprises inducing the inducible promoter for expression of the endonuclease, thereby removing the selection marker and the coding region of the endonuclease, including the inducible promoter, from the target genome.

The invention further provides a microorganism, comprising a genomic alteration that is produced by the methods of the invention. When present, the duplicated regions of homology with the target genome on the first and second targeting construct ensure seamless marker removal from the target genome by homologous recombination. The resulting microorganism comprises only the alteration or alterations that were present on the first and/or second targeting construct, or that were induced by recombination of the targeting constructs into the targeting genome, such as an insertion into the targeting genome or a deletion from the targeting genome.
The invention further provides a microorganism, comprising a genomic alteration, preferably an alteration of a target gene, the alteration comprising an insertion of a functionally recombined selection marker and a coding sequence for an endonuclease that is coupled to an inducible promoter, whereby the target genome comprises one copy of a recognition sequence for the endonuclease on both sites of the insertion.

The invention further provides a kit comprising the nucleic acid molecule of the invention, or the set of constructs of the invention. Said kit may further comprise methods and means for growth of a microorganism, preferably a yeast, preferably of the Saccharomycetaceae, in synthetic medium comprising guanidinobutyrate and/or agmatine, and/or the identification of guanidinobutyrase or agmatinase enzyme such as, for example, guanidinobutyric acid and/or agmatine.

The invention also provides a method of culturing a microorganism, preferably a yeast, preferably of the Saccharomycetaceae, in the presence of guanidinobutyrate or agmatine as sole nitrogen source, comprising: (a) introducing the nucleic acid molecule of the invention into the microorganism, and (b) culturing the microorganism such that the nucleotide molecule is expressed in the microorganism.

The invention also provides a method of culturing a microorganism, preferably a yeast, preferably of the Saccharomycetaceae, in the presence of guanidinobutyrate or agmatine as sole nitrogen source, comprising: (a) introducing the set of constructs of the invention into the microorganism, and (b) culturing the microorganism such that following recombination of the targeting constructs the nucleotide molecule encoding the selection marker is expressed in the microorganism.

The invention further provides a method for producing a microorganism comprising an altered genome, the method comprising providing a microorganism
comprising an alteration of the genome, preferably of a target gene, the alteration comprising an insertion of a functionally recombined nucleotide sequence encoding a selection marker and a coding sequence for an endonuclease that is coupled to an inducible promoter, whereby the target genome comprises one copy of a recognition sequence for the endonuclease on both sites of the insertion, and inducing the inducible promoter to remove the nucleic acid sequences in between the recognition sequences of the endonuclease. Again, when present, the duplicated regions of homology with the target gene on the first and second targeting constructs ensure seamless marker removal from the target genome by homologous recombination by providing the genomic DNA with a small homologous piece to re-connect the broken DNA strands efficiently. The resulting microorganism comprises only the alteration or alterations that were present on the first and/or second targeting construct, or that were induced by recombination of the targeting constructs into the targeting genome, such as an insertion into the targeting genome or a deletion from the genome, preferably an insertion into a targeted gene or a deletion of the targeted gene or a deletion from within the targeted gene.

For the purpose of clarity and a concise description, features are described herein as part of the same or separate embodiments, however, it will be appreciated that the scope of the invention may include embodiments having combinations of all or some of the features described.
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Recognition sequence</th>
<th>Cut</th>
<th>SF</th>
<th>Source</th>
<th>D</th>
<th>SCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-Anil</td>
<td>5’ TTGAGGGAGTTCTCTGTAAATAA 3’</td>
<td>5’ ---TTGAGGAGTTTC TCTGATAAATAA--- 3’</td>
<td>HI</td>
<td>Aspergillus nidulans</td>
<td>E</td>
<td>mito</td>
</tr>
<tr>
<td></td>
<td>3’ AACCTCTCCTAAGACATTATT 5’</td>
<td>3’ ---AACCTCTCCC AAGGACATTATT--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Ceul</td>
<td>5’ TAATCATCTACGGAACAGG 3’</td>
<td>5’ ---TAATCATCTACGGAACAGG--- 3’</td>
<td>HI</td>
<td>Chlamydomonas eugametos</td>
<td>E</td>
<td>chloro</td>
</tr>
<tr>
<td></td>
<td>3’ ATGGATATGAGGCCTTGAGTCTCAG 5’</td>
<td>3’ ---ATGGATATGAGGCCTTGAGTCTCAG--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Chul</td>
<td>5’ GAAGGTGTCAGCGACCTGGCTCTATC 3’</td>
<td>5’ ---GAAGGTGTCAGCGACCTGGCTCTATC--- 3’</td>
<td>HI</td>
<td>Chlamydomonas humicola</td>
<td>E</td>
<td>chloro</td>
</tr>
<tr>
<td></td>
<td>3’ CTCCCAAACGGGACCGTACGGTCGATG 5’</td>
<td>3’ ---CTCCCAAACGGGACCGTACGGTCGATG--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Cpal</td>
<td>5’ CGATCCGTAAGTGAGCAACAATTCA 3’</td>
<td>5’ ---CGATCCGTAAGTGAGCAACAATTCA--- 3’</td>
<td>HI</td>
<td>Chlamydomonas pallidostigmata</td>
<td>E</td>
<td>chloro</td>
</tr>
<tr>
<td></td>
<td>3’ GCCTCTAGCCATGCTGTTGACGTTCTTATT 5’</td>
<td>3’ ---GCCTCTAGCCATGCTGTTGACGTTCTTATT--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-CpalL</td>
<td>5’ CCCGCTAATCGTGCCG 3’</td>
<td>5’ ---CCCGCTAATCGTGCCG--- 3’</td>
<td>HI</td>
<td>Chlamydomonas pallidostigmata</td>
<td>E</td>
<td>chloro</td>
</tr>
<tr>
<td></td>
<td>3’ GGCCGAGTGGACGAGGTC 5’</td>
<td>5’ ---GGCCGAGTGGACGAGGTC--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Crel</td>
<td>5’ CTGSGTCTAACAAAACGTGGAACGGATGTGG 3’</td>
<td>5’ ---CTGSGTCTAACAAAACGTGGAACGGATGTGG--- 3’</td>
<td>HI</td>
<td>Chlamydomonas reinhardtii</td>
<td>E</td>
<td>chloro</td>
</tr>
<tr>
<td></td>
<td>3’ GACCCACAGGGGATGATGCTGCACATGCATGAC 5’</td>
<td>3’ ---GACCCACAGGGGATGATGCTGCACATGCATGAC--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dmol</td>
<td>5’ ATGCCTCGGCGGATGAGTGACGACGATG 3’</td>
<td>5’ ---ATGCCTCGGCGGATGAGTGACGACGATG--- 3’</td>
<td>Desulfurococcus mobilis</td>
<td>A</td>
<td>Chrml</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’ TACCACCTCGAATACGGCAGCGCGGCGTA 5’</td>
<td>3’ ---TACCACCTCGAATACGGCAGCGCGGCGTA--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H-Drel</td>
<td>5’ CAAACAGGCTGAAGAGCTACGGCGCGGCGG 3’</td>
<td>5’ ---CAAACAGGCTGAAGAGCTACGGCGCGGCGG--- 3’</td>
<td>HI</td>
<td>Escherichia coli</td>
<td>B</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’ GTTTTCGAGCGATTACAGGGCCCGC 5’</td>
<td>5’ ---GTTTTCGAGCGATTACAGGGCCCGC--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Hmul</td>
<td>5’ AGTATGGCTACCTGACG 3’</td>
<td>1 : 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’ TGTTATCTCGGATGGCTCTCC 5’</td>
<td>3’ ---TGTTATCTCGGATGGCTCTCC--- 5’</td>
<td>HIII</td>
<td>Bacillus subtilis SPO1</td>
<td>B</td>
<td>phage</td>
</tr>
<tr>
<td>L-Hmull</td>
<td>5’ AGTATGCTACGAGGGCTCAACAC 3’</td>
<td>3’ ---AGTATGCTACGAGGGCTCAACAC--- 5’</td>
<td>HIII</td>
<td>Bacillus subtilis phase SP82</td>
<td>B</td>
<td>phage</td>
</tr>
<tr>
<td></td>
<td>3’ CATTATTCTCGGATGGCTCTCC 5’</td>
<td>3’ ---CATTATTCTCGGATGGCTCTCC--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Lal</td>
<td>5’ CATGACCTTAATCGCTATCTATTTT 3’</td>
<td>5’ ---CATGACCTTAATCGCTATCTATTTT--- 3’</td>
<td>HIII</td>
<td>Lactococcus lactis</td>
<td>B</td>
<td>chrml</td>
</tr>
<tr>
<td></td>
<td>3’ GTGAGGATCTGTATAGCTAAATAAAA 5’</td>
<td>5’ ---GTGAGGATCTGTATAGCTAAATAAAA--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Msol</td>
<td>5’ CTGGGGCGTAACGGGACGGATGGTGG 3’</td>
<td>5’ ---CTGGGGCGTAACGGGACGGATGGTGG--- 3’</td>
<td>Monomastix sp.</td>
<td>E</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’ GACCCAGTGGGGGCTACAGACCTGCACGCAAA 5’</td>
<td>3’ ---GACCCAGTGGGGGCTACAGACCTGCACGCAAA--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PL-Pful</td>
<td>5’ GAAGATGGGGAGGAGGAGCCGAGCATTCT 3’</td>
<td>5’ ---GAAGATGGGGAGGAGGAGCCGAGCATTCT--- 3’</td>
<td>Pyrococcus furiosus</td>
<td>A</td>
<td>Ve1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3’ CTCTACCCTCTCCCTGCGCTGAGTTGGA 5’</td>
<td>3’ ---CTCTACCCTCTCCCTGCGCTGAGTTGGA--- 5’</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| PI-Pkol | 5' CAGTACTACGCGTTACG3'  
3' GTGATGATGCAACATG5' | 5' ---CAGTACTACG GTTAC--- 3'  
3' ---GTGATG ATGCAATG--- 5' | Pyrococcus  
badiaariensis  
KOD1 | A |
|---------|-------------------|-------------------|-----------------|------|
| I-Psol | 5' GCAGCCGGCTAAGCTGTTGTCACG5'  
3' CAGTTCGGCCATCCACACCCAGG5' | 5' ---GCAGCCGGCTAAGCTGTTGTCACG--- 3'  
3' ---CAGTTCGGCCATCCACACCCAGG--- 5' | Pyrobaculum  
organomethanolum | A | chrn |
| L-Ppol | 5' TAATCTAGCCCTCTTCTGTTACG5'  
3' CTGGTCTACCTGATCCATTCACGG5' | 5' ---TAATCTAGCCCTCTTCTGTTACG--- 3'  
3' ---CTGGTCTACCTGATCCATTCACGG--- 5' | Physarum  
polycephalum | E | nuclear |
| PI-Pspl | 5' TGCCCAACACGCTTACATTGCCGATG5'  
3' ACCTGGTTGCTAATCCACACCCAGG5' | 5' ---TGCCCAACACGCTTACATTGCCGATG--- 3'  
3' ---ACCTGGTTGCTAATCCACACCCAGG--- 5' | Pyrococcus sp. | A | chrn |
| L-Seal | 5' TGTGCACTTGATGGGATCTAGGCTAATAG5'  
3' ACAGTCTGATCCTTGGCGCCATCCAG5' | 5' ---TGTGCACTTGATGGGATCTAGGCTAATAG--- 3'  
3' ---ACAGTCTGATCCTTGGCGCCATCCAG--- 5' | Saccharomyces  
cerevisiae | E | mito |
| L-Seel | 5' AGTTAAGCTAGGATGAAACAGGATATCG5'  
3' CTAATCGGTCACCTTGGCCATCCAG5' | 5' ---AGTTAAGCTAGGATGAAACAGGATATCG--- 3'  
3' ---CTAATCGGTCACCTTGGCCATCCAG--- 5' | Saccharomyces  
cerevisiae | E | mito |
| PI-Seel | 5' ATCTATGTCGGGCGGAGAACAGGATATCG5'  
3' TAGATACAGCCGCTACTTCCACATCCAGG5' | 5' ---ATCTATGTCGGGCGGAGAACAGGATATCG--- 3'  
3' ---TAGATACAGCCGCTACTTCCACATCCAGG--- 5' | Saccharomyces  
cerevisiae | E | mito |
| L-SeeI | 5' TTGTGATTTGTCACCTTGGCCGCAAG5'  
3' AAACTAAAGAAACAGTGGGCTACTTGG5' | 5' ---TTGTGATTTGTCACCTTGGCCGCAAG--- 3'  
3' ---AAACTAAAGAAACAGTGGGCTACTTGG--- 5' | Saccharomyces  
cerevisiae | E | mito |
| L-SeeIII | 5' ATGGGAGGGTTTGGTAAACTATATATACG5'  
3' TAACCTCCGAAACACTTAGATATAAGG5' | 5' ---ATGGGAGGGTTTGGTAAACTATATATACG--- 3'  
3' ---TAACCTCCGAAACACTTAGATATAAGG--- 5' | Saccharomyces  
cerevisiae | E | mito |
| L-SeeIV | 5' TCTTTTCTTTGATGAGTACG5'  
3' AGGAAGAGAGAAAAGAAACAGTGGGGAGATCG5' | 5' ---TCTTTTCTTTGATGAGTACG--- 3'  
3' ---AGGAAGAGAGAAAAGAAACAGTGGGGAGATCG--- 5' | Saccharomyces  
cerevisiae | E | mito |
| L-SeeV | 5' AATAATTTCCTCTTCTTACG5'  
3' TTATTTAAAGAAGAACAGTGGGGAGATCG5' | 5' ---AATAATTTCCTCTTCTTACG--- 3'  
3' ---TTATTTAAAGAAGAACAGTGGGGAGATCG--- 5' | Saccharomyces  
cerevisiae | E | mito |
| L-SeeVI | 5' GTIAATTTAAATAGTCAGG5'  
3' CAAATAACAAAATACTCAGAAC5' | 5' ---GTIAATTTAAATAGTCAGG--- 3'  
3' ---CAAATAACAAAATACTCAGAAC--- 5' | Saccharomyces  
cerevisiae | E | mito |
| L-SeeVII | 5' GTCACTACGCGTTACG5'  
3' ACAGTCTGATCCTTGGCGCCATCCAG5' | Unknown | Saccharomyces  
cerevisiae | E | mito |
| **L-Ssp6808I** | 5' GTCGGGCTCATAACCCGAA  
3' CAGCCCQGATATTGCCCTT | 5' ---GTGGCCCT CATAACCCGAA--- 3'  
3' ---CAGCCCCGACT TTGGGCTT--- 5' | *Synechocystis* sp. PCC 6803 | B |
| **L-Tev1** | 5' AGTGOTATACACCGCTAGTATG  
3' TGCACATGAT GCCTAGCAATTCAC | 5' ---AGTGOTATACACCGCTAGTATG--- 3'  
3' ---TCGCACATGAT GCCTAGCAATTCAC--- 5' | *Escherichia coli* phage T4 | B |
| **L-Tev2** | 5' GCTTTAGTATGAAAGTGAACGGTTACGTTATTC  
3' CGAAATCTCATACCTTCAGTGAGAATAAAG | 5' ---GCTTTAGTATGAAAGTGAACGGTTACGTTATTC--- 3'  
3' ---CGAAATCTCATACCTTCAGTGAGAATAAAG--- 5' | *Escherichia coli* phage T4 | B |
| **L-Tev3** | 5' TATGTAATCTTTCGCCGTGCTACCTTTAACCTC  
3' ATACATAGAAACACCCATGAAATGGGAG | 5' ---T ATGTAATCTTTCGCCGTGCTACCTTTAACCTC--- 3'  
3' ---ATACATAGAAACACCCATGAAATGGGAG--- 5' | *Escherichia coli* phage RB3 | B |
| **PL-Thl** | 5' THGCGAYACNGACGGTTYYT  
3' ATRCGQCTRTGQCTGCTAARA | 5' ---THGCGAYACNGACGGTTYYT--- 3'  
3' ---ATRCQGQCTRTGQCTGCTAARA--- 5' | *Thermococcus litoralis* | A |
| **PL-Th2** | 5' AATCTGCTTGCAAAACAGCTATTACGGCTAT  
3' TTTACGAAACGTGTTCGCTGATA | Unknown **  
3' ---ATCTGCTTGCAAAACAGCTATTACGGCTAT--- 5'  
3' ---T TTACGAAACGTGTTCGCTGATA--- 5' | *Thermococcus litoralis* | A |
| **L-Tsp061I** | 5' CTTCAGTATGACCCGGAAC  
3' GAATGCTATGACTGCGTGGTTG | 5' ---CTTCAGTATGACCCGGAAC--- 3'  
3' ---GAATGCTATGACTGCGTGGTTG--- 5' | *Thermoproteus sp.* IC-061 | A |
| **L-Vdi141I** | 5' CCTGACTCTCCTTTAAGTCAGCCCAA  
3' GGCTGAGAGAATTCATCGCTTTT | 5' ---CCTGACTCTCCTTTAAGTCAGCCCAA--- 3'  
3' ---GGCTGAGAGAATTCATCGCTTTT--- 5' | Vulcaniseta distributa IC-141 | A |
Table 1: overview of homing endonucleases and their target sequences.

Abbreviations: SF Structural family; HI: LAGLIDADG family; HII: GIY-YIG family; HIII: H-N-H family; HIV: His-Cys box family.

D: Biological domain of the source: A: archaea; B: bacteria; E: eukarya.

Table 2: *Saccharomyces cerevisiae* strains used in this study.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEN.PK113-7D</td>
<td>Prototrophic reference strain MATa</td>
<td>(Nijkamp et al., 2012)</td>
</tr>
<tr>
<td>CEN.PK113-5D</td>
<td>MATa ura3-52</td>
<td>(Entian &amp; Kötter, 2007)</td>
</tr>
<tr>
<td>IMZ312</td>
<td>MATa ura3-52 pAG426GPD-ccdB</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>(TDH3pr-CYC1&lt;sub&gt;er&lt;/sub&gt;,URA3 2µ)</td>
<td></td>
</tr>
<tr>
<td>IME215</td>
<td>MATa ura3-52 pUDE264 (TDH3&lt;sub&gt;pr&lt;/sub&gt;,</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>KIGBU1-CYC1&lt;sub&gt;er&lt;/sub&gt;,URA3 2µ)</td>
<td></td>
</tr>
<tr>
<td>IMX598</td>
<td>MAT a ade2Δ::TDH3&lt;sub&gt;pr&lt;/sub&gt;,KIGBU1-CYC1&lt;sub&gt;er&lt;/sub&gt;</td>
<td>This study</td>
</tr>
</tbody>
</table>

References:
<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence 5’ to 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUG-RV</td>
<td>GGGAGATCTCTGGCCAAGCGAATTTGGAAGACCGTGCTGAGATGGAAGAAC</td>
</tr>
<tr>
<td></td>
<td>TTAAGGTTGGCTGAGCTGCC</td>
</tr>
<tr>
<td>pUg-FW</td>
<td>GGGCTCTAGATAGACAAAGGCGATCTGCTGTAATTTATGAGTCC</td>
</tr>
<tr>
<td></td>
<td>AGATTCACATAGTGGCTATG</td>
</tr>
<tr>
<td>pDS-RV</td>
<td>GGGTCTGACCGCCGCAAGCGAATTTGGAAGACCGTGCTGAGATGGAAGAAC</td>
</tr>
<tr>
<td></td>
<td>TGTATTTAAATGAGCTGAGCGTTTTAAAGGCTGCTGAGCGCTC</td>
</tr>
<tr>
<td>pDS-FW</td>
<td>GGGGATCTCTGACAAAGGCGAATTTGGAAGACCGTGCTGAGATGGAAGAAC</td>
</tr>
<tr>
<td></td>
<td>GATACATAGGCGAATTTGGAAGACCGTGCTGAGATGGAAGAAC</td>
</tr>
<tr>
<td>FK140 pUD1065 fw</td>
<td>GTATACACGAGGCCCTTTTC</td>
</tr>
<tr>
<td>SLT1_control_rv</td>
<td>CAATTCAACGCGTCCTGTGAC</td>
</tr>
<tr>
<td>KANMX4 fw</td>
<td>TCTTTCTGCGGTATCC</td>
</tr>
<tr>
<td>FK105-MP1</td>
<td>CTTCGTTGAGTTTTTCTCTTCA</td>
</tr>
<tr>
<td>GBU1 forward primer</td>
<td>CATTCCGAAACATAAACAACGGATGAGTTGCCCAGGATATTATTC</td>
</tr>
<tr>
<td>GBU1 reverse primer</td>
<td>CAAAGAATCTTTTATTTGTCAGTAATCTGACAGCTGTGTCCTT</td>
</tr>
<tr>
<td>Backbone forward</td>
<td>CAATATAAAATCTGGCAACCTTCATGAGTTGTTATATGTTGAGTG</td>
</tr>
<tr>
<td>Backbone reverse</td>
<td>GAACATTTTTTCGAGCGCTGATGACATGATCGTAATAAAAGATTCTTC</td>
</tr>
<tr>
<td>GBU1-ADE2ko-CENPK-fw</td>
<td>GTAAAAATGGGTGGTCTCTCTCTCTCTACTGACATCTCTACTTAAACATAAGA</td>
</tr>
<tr>
<td></td>
<td>AAAATCGGACAAAAACAAAACTACAGTATAGCTGGCTGAGCGTGCAAAAGCTTATTG</td>
</tr>
<tr>
<td>GBU1-ADE2ko-CENPK-rv</td>
<td>GATGTAATACATACAAAGGCTAATATATATATATATATATATATATATATATATATGAGATCACA</td>
</tr>
<tr>
<td></td>
<td>AGTATATACATACAAAGGCTAATATATATATATATATATATATATATATATATGAGATCACA</td>
</tr>
</tbody>
</table>
Table 4: Plasmids used in this study. a DR Direct Repeat.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Characteristic</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAG426GPDCcdB</td>
<td>2μ ori URA3 TDH3pr-ccdB-CYC1or</td>
<td>(Alberti et al., 2007)</td>
</tr>
<tr>
<td>pUDE264</td>
<td>2μ ori URA3 TDH3pr-KIGBU1-CYC1or</td>
<td>This study</td>
</tr>
<tr>
<td>pDS1</td>
<td>ori AmpR DRpr-I-Scel site-AgTEF2or-KanMX-AgTEF2or-I-Scel site-DR</td>
<td>This study</td>
</tr>
<tr>
<td>pDS8</td>
<td>ori AmpR DRpr-I-Scel site-AgTEF2or-KIGBU1-AgTEF2or-I-Scel site-DR</td>
<td>This study</td>
</tr>
</tbody>
</table>


Table 5: Guanidinobutyrase activities measured in cell extracts of S. cerevisiae strains IME215 and CEN.PK113-7D grown in batch cultures with arginine as sole nitrogen source. The S. cerevisiae strains and IME215 were pre-grown in synthetic medium with glucose and ammonium as sole nitrogen source. B.D. denotes below detection limit, which was estimated at 0.005 μmol min⁻¹ mg⁻¹ protein. Data represent the average ± mean deviation of independent biological duplicate cultures.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
<th>Activity (μmol min⁻¹ mg protein⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. cerevisiae IME215</td>
<td>MATa ura3-52 pUDE264 (TDH3pr-KIGBU1-CYC1or, URA3)</td>
<td>0.17 ± 0.006</td>
</tr>
<tr>
<td>S. cerevisiae CEN.PK113-7D</td>
<td>MATa Prototrophic reference</td>
<td>B.D.</td>
</tr>
</tbody>
</table>


Figure legends

Figure 1. Overview of the key reactions in eukaryotic arginine metabolism. Thick lines indicate ureohydrolase reactions. EC 3.5.3.1: arginase, EC 4.1.1.17, ornithine decarboxylase, EC 2.6.1.13: ornithine aminotransferase, EC 1.5.1.2: pyrrole-5-carboxylate reductase, EC 1.5.99.8: proline dehydrogenase, EC 1.5.1.12: 1-pyrrole-5-carboxylate dehydrogenase, EC 2.6.1.-aminotransferase, EC 4.1.1.75 2-oxo acid decarboxylase, EC 1.2.1.54 gamma-guanidinobutyraldehyde dehydrogenase, EC 3.5.3.7 guanidinobutyrate, EC 2.6.1.19 GABA transaminase, EC 1.2.1.16 succinate-semialdehyde dehydrogenase.

Figure 2. A- Alignment of guanidino-acid hydrolase (EC.3.5.3.7) amino acid sequences of Saccharomycotina yeasts. The amino acid sequences were aligned using Clustal W (V1.7). The amino acid residues conserved in all sequences are depicted with the * symbol. XP_456325: reference GBU1 from Kluyveromyces lactis NRRL Y-1140; XP_002498240: Zygosaccharomyces rouxii CBS 732; XP_716668: Candida albicans SC5314; XP_461566: Debaryomyces hansenii CBS767; EIF45280: Dekkera bruxellensis AWRI1499; BAO40383:

Kluyveromyces marxianus DMKU3-1042; AADMO1000201.1: Lachancea waltii NCYC 2644; XP_503530: Yarrowia lipolytica CLIB122; EFW95635.1: Ogataea parapolymorpha DL-1 (Hansenula polymorpha); XP_002552049: Lachancea thermotolerans CBS 6340 (Kluyveromyces thermotolerans CBS6340); XP_001523956: Lodderomyces elongisporus NRRL YB-4239; XP_001482640:

Meyerozyma guillermondii ATCC 6260 (Pichia guillermondii ATCC6260); XP_004196483: Millerozyma farinosa CBS 7064 (Pichia sorbitophila); XP_001385334: Scheffersomyces stipitis CBS 6054 (Pichia stipitis CBS 6054); XP_003679661: Torulaspora delbrueckii CBS 1146; AACE03000003.1: Lachancea kluyveri NRRL Y-12651. The consensus sequence is presented under the sequence alignment and amino acid conserved in at least 50% of the
sequences. Bold and underlined amino acid denote a conserved domains found in guanidino-acid hydrolase.

B- Weblogo (Schneider and Stephens (1990). Nucleic Acids Research 18, 6097–6100) representing the conserved guanidino acid hydrolase motives from the sequence alignment.

Figure 3. GenBank entry of the *K. lactis* protein XP_456325.1

Figure 4. Alignment of bacterial guanidino-amide hydrolase (EC.3.5.3.11) amino acid sequences. A. The amino acid sequences were aligned using Clustal W (V1.7). The amino acid residues conserved in all sequences are depicted with the * symbol. NP_289508: *Escherichia coli*; WP_000105576: *Shigella flexneri*; YP_005016506: *Klebsiella oxytocia* KCTC 1686; YP_001337000: *Klebsiella pneumoniae* subsp. *pneumoniae* MGH 78578; YP_003614749: *Enterobacter cloacae* subsp. *cloacae* ATCC 13047; YP_001455807: *Citrobacter koseri* ATCC BAA-895; XP_004532666: *Cerateis capitata*; WP_006734551: *Salmonella enterica*; AHE29794: *Burkholderia pseudomallei* NCTC 1317. B- Consensus sequence generated from the sequence alignment. The consensus sequence is presented under the sequence alignment and amino acid conserved in at least 50% of the sequences. Bold and underlined amino acid denote a conserved domains found in guanidino-amide hydrolase.

B- Weblogo representing the conserved guanidino amide hydrolase motives from the sequence alignment.

Figure 5. GenBank entry of the *E. coli* protein (agmatinase) AAC75974.1.

Figure 6. Vector 1 and 2 with all essential parts for the standard deletion cassette. The 400 base overlap in the selection marker *KlGBU1* (indicated by a cross) is designed to recombine due to the homology.
Figure 7. Map of the plasmid pDS8 comprising the KIGBU1YM marker module.

Figure 8. Typical growth profile of S. cerevisiae strains on guanidinobutyrate: the S. cerevisiae strains CEN.PK113-7D (closed circle) and IME215 (TDH3pr::KIGBU1::CYC1ter) (closed square) were grown in shake flasks on glucose synthetic medium containing guanidinobutyrate as sole nitrogen source.

Figure 9. Transformation of S. cerevisiae and S. pastorianus with pUDE264. The strains S. cerevisiae CEN.PK113-7D and S. pastorianus CBS1483 were transformed with the plasmid pUDE264 that carries the guanidinobutyrase gene from K. lactis. The transformed cells (50 µl) were plated on SM with different nitrogen sources ((NH₄)₂SO₄, guanidinobutyrate and agmatine). Untransformed strains (-DNA) were also plated on similar media as negative control.

Figure 10. Examples of deletion cassettes. The deletion cassettes contain several regions: A- (1) a 50- to 80-bp sequence homologous to the upstream part of the gene to be deleted, including the start codon, and a 50- to 80-bp sequence homologous to the downstream part of the gene to be deleted, including the stop codon and (4) the selectable cassette which includes the Ashbya gossypii TEF2 promoter, the K. lactis GBU1 gene and the A. gossypii TEF2 terminator. B- In addition to the region (1) and (4) the deletion cassette includes (2) a 40-bp sequence flanked by (3) an I-SceI restriction site located upstream and downstream of the marker module.
Examples

Example 1

Materials and Methods

Strains and maintenance

The *Saccharomyces cerevisiae* strains used in this study are listed in Table 2. The *S. cerevisiae* strains were constructed in the CEN.PK background (Nijkamp et al., (2012) Microb Cell Fact 11: 36; Entian & Kötter, (2007) Meth Microbiol 36: 629-666). Yeast strains that did not carry a plasmid were maintained on YPD medium (demineralized water; 10 g/L yeast extract; 20 g/L peptone; 20 g/L glucose). Yeast strains carrying plasmid were maintained on synthetic medium containing salts, trace elements and vitamins, prepared and sterilized as described previously (Verduyn et al., (1992) Yeast 8: 501-517) in which urea instead of ammonium sulfate was used when applicable. Culture stocks were prepared from shake flask cultures incubated at 30 °C and stirred at 200 rpm, by addition of 20 % (v/v) glycerol and were stored at -80 °C.

Media and culture conditions

Growth experiments were conducted in synthetic medium containing salts, trace elements and vitamins, prepared and sterilized as described previously (Verduyn et al., (1992) Yeast 8: 501-517). Glucose was added to a final concentration of 20.0 g/L. When ammonium sulfate was not the nitrogen source in the synthetic medium, it was replaced by guanidinobutyrate which was filter sterilized and added to sterile medium to concentrations of 2.9 g/L. Moreover, 3.3 g/L potassium sulfate was added to compensate for the removal of ammonium sulfate.

If required, 0.15 g/L uracil and/or 200 mg/L of G418 (Geneticin) were added to complete media. Selection agar plates were made by adding 20.0 g/L agar to these synthetic media.
Shake flask cultures were conducted in 500 ml or 250 ml shake flasks containing 100 ml or 20 ml of liquid medium respectively and incubated in an orbital shaker (New Brunswick Scientific, Edison, NJ) at 200 rpm at 30 °C.

Cloning and overexpression of K. lactis KLLA0F27995g in S. cerevisiae.
Genomic DNA of the prototrophic reference strain S. cerevisiae CEN.PK113-7D was prepared as described previously (Burke et al., 2000. Cold Spring Harbor Laboratory. Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual). ORF KLLA0F27995g (KIGBU1) was cloned from genomic DNA using Phusion Hot-Start polymerase (Finnzymes) and primers GBU1 forward primer/GBU1 reverse primer Table 3.
The PCR products was digested with SpeI and XhoI cloned into pAG426GPDccdB (Alberti et al., (2007) Yeast 24: 913-919; Table 4) preliminarily digested with the same enzymes, using T4 ligase (Life Technologies, Breda, The Netherlands) resulting in plasmid pUDE264.
The plasmid pUDE264 was transformed in S. cerevisiae CEN.PK113-5D (ura3-52) using the LiAc method as previously described in (Gietz & Woods, (2002) Methods Enzymol 350: 87-96) resulting in strain IME215 (MATa ura3-52 pUDE264 (TDH3pr-KIGBU1-CYC1or URA3 2μ). The backbone plasmid pAG426GPD-ccdB was transformed in S. cerevisiae CEN.PK113-5D (ura3-52) using the LiAc method as previously described in (Gietz & Woods, (2002) Methods Enzymol 350: 87-96) resulting in strain IMZ312 (MATa ura3-52 pAG426GPD-ccdB (TDH3pr-CYC1or URA3 2μ).

Construction of the plasmid pDS8
The KLLA0F27995g (KIGBU1) ORF was transferred from pUDE264 into pDS1 in place of the kanR gene. For the construction of the plasmid pDS1, the vector pUG6 (Guldener et al. (1996) Nucleic Acids Res 24:2519-24) was used as template for PCR using PhusionTM Hot Start II DNA Polymerase (Thermo Scientific, Waltham, MA) according to manufacturer’s instructions and the primer pairs pUGfw/pUGrv to amplify the backbone of the vector and
pDSfw/pDSrv to amplify the selection marker. The primer pUGfw contained the XbaI restriction site and a 40 bp synthetic sequence (repeat B) which was used to construct a 80 bp direct repeat. The primer pUGrv contained the BgIII restriction site and another 40 bp synthetic sequence (repeat A) to construct the second 80 bp direct repeat. The primer pDSfw contained the BgIII restriction site followed by to 40 bp B repeat. The primer pDSrv contained the XbaI restriction site and the 40 bp A repeat. All PCR’s were visualized with gel electrophoresis. The fragments were isolated from agarose gel using Zymoclean™ Gel DNA Recovery Kit. (Zymo Research, Irvine, CA) The PCR fragments were restricted with XbaI and BgIII and ligated with T4 DNA ligase (Thermo Scientific) according to manufacturer’s instructions. Chemical competent *Escherichia coli*, strain DH5α, was transformed with the ligation mix and a correct colony was selected by PCR using the following primers FK140 pUDI065 fw, SLT1_control_rv, KANMX4 fw and FK105-MP1.

The transfer of *KIGBU1* in place of the kanR sequence was achieved by Gibson assembly (Gibson et al., (2009) Nat Methods 6: 343-345). The fragment carrying the *KIGBU1* gene was PCR amplified using Phusion Hot-Start polymerase (Finnzymes) and primers GBU1-fw and GBU1-rv (Table 3). The plasmid pUDE264 was used as template. The backbone plasmid was PCR amplified using Phusion Hot-Start polymerase (Finnzymes) and primers Backbone-fw and Backbone-rv. The plasmid pDS1 was used as template. The two fragments shared at least 43 nucleotides identity at their flanks and were assembled in vitro using the Gibson assembly cloning kit from New England Biolabs (Ipswich, MA). The assembled mixture was transformed in *E. coli* DH5α and plated on LB plate containing ampicillin (100mg/L). The assembled plasmid which contains the marker module KIGBU1YM was verified by restriction analysis and sequencing and a correct clone was named pDS8.

*Deletion of ADE2 in S. cerevisiae.*

Gene deletions in *S. cerevisiae* were performed by integration of the KIGBU1YM (SceI site::AgTEF2_pr-KIGBU1-AgTEF2_tr::SceI site) cassettes via
the short-flanking-homology PCR method (Wach et al., (1994) Yeast 10: 1793-1808). Sequences of oligonucleotide primers are shown in Table 3. Deletion cassette for ADE2 was amplified using Phusion Hot-Start polymerase (Finnzymes, Landsmeer, The Netherlands) and the template plasmid pDS8 using primers GBU1-ADE2ko-CENPK-fw/ GBU1-ADE2ko-CENPK-rv. The transformation of S. cerevisiae CEN.PK113-7D with the ADE2 deletion cassettes was performed using the LiAc method as previously described in (Gietz & Woods, (2002) Methods Enzymol 350: 87-96) resulting in strain IMX598.

Correct integration of the KIGBU1SY cassette and replacement of the gene of interest was, verified by diagnostic PCR using a forward primer specific for the 5' UTR (untranslated region) of ADE2 and the reverse primer for the deletion cassette (Table 3).

Preparation of cell extracts
For preparation of cell extracts, culture samples were harvested by centrifugation, washed twice with 10 mM potassium phosphate buffer (pH 7.5) containing 2 mM EDTA and stored at -20 °C. Before cell disruption, samples were thawed at room temperature, washed, and resuspended in 100 mM potassium phosphate buffer (pH 7.5) containing 2 mM MgCl2 and 2 mM dithiothreitol. Extracts were prepared by sonication with 0.7 mm glass beads at 0 °C for 2 min at 0.5 min intervals with an MSE sonicator (Wolf Laboratories Limited, Pocklington, United Kingdom) (150 W output; 8 μm peak-to-peak amplitude) (Luttik et al., (2008) Metab Eng 10: 141-153). Unbroken cells and debris were removed by centrifugation at 4 °C (20 min; 36000 x g). The resulting cell extract was used for enzyme assays.

Enzyme activity assays
For the arginase enzymatic assay 50μL cell extract were activated in 950μL manganese maleate buffer (50 mM manganese sulfate, 50 mM maleic acid, pH
7) for 1 h at 37 °C (Messenguy et al., (1971) Eur J Biochem 22: 277-286). The reaction mixture for arginase assays, prepared in dark eppendorf tubes, contained 60 μL of activated cell extract, 400 μL 713 mM arginine solution (pH 9.5) and demineralized water up to 1 mL. The reaction mixture was incubated for 30 min at 37 °C. To stop the reaction, 0.7 mL sulfuric-phosphoric acid mixture (20% v/v concentrated sulfuric acid and 60% v/v syrupy phosphoric acid in demineralized water) was added to the reaction mixture. The amount of urea produced was quantified using the Archibald method (Archibald, (1945) J Biol Chem 157: 507-518) with a calibration line ranging from 0 until 0.6 mM of urea. 0.06 mL of a 4% v/v α-isonitroso-propiophenone in ethanol solution was added and samples were thoroughly mixed before boiling for 1 h in a 100 °C water bath to develop the color. The samples were cooled at room temperature for 15 min and the absorbance at 540 nm was measured in a Libra S11 spectrophotometer (Biochrom, Cambridge, United Kingdom).

The reaction mixture for guanidinobutyrase (GBU) enzyme assays was prepared in dark eppendorf tubes, containing in a 1 mL final volume: 50 mM glycine buffer (pH 9), 5 mM MnSO4 and 50 μL to 100 μL cell extract. The reaction was started by addition of 50 mM guanidinobutyric acid. After 30 min of incubation at 37 °C, the reaction was stopped by addition of 700 μL of sulfuric-phosphoric acid mixture (20% v/v concentrated sulfuric acid and 60% v/v syrupy phosphoric acid in demineralized water). The amount of urea produced was quantified using the Archibald method (Archibald, 1945) with a calibration line ranging from 0 until 0.6 mM of urea. 0.06 mL of a 4% v/v α-isonitrosopropiophenone in ethanol solution was added and samples were thoroughly mixed before boiling for 1 h in a 100 °C water bath to develop the color. The samples were cooled at room temperature for 15 min and the absorbance at 540 nm was measured with a Libra S11 spectrophotometer (Biochrom).
Results

Expression of KIGBU1 in S. cerevisiae confers the ability to grow on guanidinobutyrate as sole nitrogen source.

To characterize the function the \textit{K. lactis} putative ureohydrolase gene, KLLA0F27995g was cloned under the control of the strong constitutive \textit{TDH3} promoter (\textit{TDH3pr}) in an expression vector and transformed to \textit{S. cerevisiae} (Figure 8).

The expression plasmid pUDE264 was transformed in the laboratory strain CEN.PK113.7D (prototroph) and the wild brewing \textit{Saccharomyces pastorianus} strain CBS1483 (www.cbs.knaw.nl/Collections/). The transformed cells were plated on synthetic medium containing either guanidinobutyrate or agmatine or ammonium sulfate. Expectedly, for both strains all transformants could grow on non-selective plates containing ammonium sulfate. In line with the enzyme measurements no transformants was rescued on agmatine plates whereas more than thousand and hundred transformants were detected on guanidinobutyrate plates for CEN.PK113-7D and CBS1483, respectively (Figure 9). This demonstrated that \textit{KlGBU1} could be efficiently used as a dominant selectable marker in prototrophic strain of \textit{S. pastorianus}.

Plasmids and deletion cassettes construction.

The coding sequence of the \textit{K. lactis} GBU1 gene, flanked by the \textit{Ashbya. gossypii TEF2} promoter and terminator, was cloned into the vector pDS1 by replacing the KanMX gene, resulting in the plasmid pDS8 (Figure 7). The resulting KIGBU1YM module only contained heterologous sequences, thereby reducing the probability of mistargeted integration (Wach et al., (1994) Yeast 10: 1793-1808). The pDS8 plasmid can be easily used as template for deletion cassettes containing the new marker module KIGBU1YM and was used for the construction of all deletion cassettes used in this study.
The deletion cassettes contained three major regions (Fig. 10): (1) a 50- to 55-bp sequence homologous to the upstream part of the gene to be deleted, including the start codon, and a 50- to 55-bp sequence homologous to the downstream part of the gene to be deleted, including the stop codon. These regions were used for targeted homologous recombination (Baudin et al., 1993) Nucleic Acids Res 21: 3329-3330, (2) a 40-bp sequence flanked by (3) an I-SceI restriction site located upstream and downstream of the marker module and (4) the KIGBU1YM marker. Upon restriction by the endonuclease, homologous recombination of the direct repeat would be sufficient to pop-out and recycle the marker module.

**Gene deletion in S. cerevisiae using KIGBU1YM**

To evaluate whether the new marker KIGBU1YM was suitable for gene knockout in *S. cerevisiae*, it was attempted to delete a gene in the laboratory strain CEN.PK113-7D. *ADE2* was selected for this proof-of-principle experiment because the phenotype caused by *ADE2* deletion can be visually screened, giving a fast preliminary evaluation of targeted integration. *ADE2* codes for the enzyme phosphoribosylaminomimidazol carboxylase, which is involved in the biosynthesis of purine nucleotides. *ade2* mutants require an external source of adenine and accumulate precursors of purine nucleotides in the vacuole which give colonies a red color (Zonneveld & van der Zanden, 1995) Yeast 11: 823-827).

The potential of KIGBU1YM as dominant marker was tested by transforming a deletion cassette to disrupt *ADE2* in CEN.PK113-7D. After transformation, cells were grown on synthetic medium (SM) agar plates containing guanidinobutyrate as sole nitrogen source. Targeted gene deletion was confirmed by the inability of single colonies to grow on SM (in absence of adenine) and by PCR. The average transformation efficiency was 5 transformants per microg of DNA, with 100% of the colonies harboring the
correct integration, and being able to grow on guanidinobutyrate as sole nitrogen source.
Claims

1. A method of culturing a microorganism of the genus Saccharomycetaceae in the presence of guanidinobutyrate as sole nitrogen source, comprising:
   (a) introducing a nucleic acid molecule comprising a nucleotide sequence encoding a guanidinobutyrase into the microorganism, whereby the nucleotide sequence is operably linked to promoter and terminator sequences;
   (b) culturing the microorganism such that the nucleic acid molecule encoding the guanidinobutyrase is expressed in the microorganism; and
   (c) culturing the microorganism in the presence of guanidinobutyrate as sole nitrogen source.

2. The method according to claim 1, wherein said guanidinobutyrase-encoding nucleotide sequence encodes Kluyveromyces lactis NRRL Y-1140 hypothetical protein.

3. The method according to claim 1 or claim 2, wherein the promoter and/or terminator sequences are selected from a glycolytic gene.

4. The method according to claim 3, wherein the glycolytic gene is selected from PG11, PFK1, PFK2, FBA1, TPI1, TDH1, TDH3, PGK1, GPM1, PYK1, ENO1, and/or ENO2.

5. A microorganism of the genus Saccharomycetaceae that comprises a nucleotide sequence encoding a guanidinobutyrase.
6. The microorganism of the genus *Saccharomyces* according to claim 5, wherein said guanidinobutyrase-encoding nucleotide sequence encodes *Kluyveromyces lactis* NRRL Y-1140 hypothetical protein.

7. A set of constructs, comprising a first construct comprising a first part of a nucleotide sequence encoding a guanidinobutyrase, and a second construct comprising a second part of the nucleotide sequence encoding the guanidinobutyrase, whereby a fragment of the first part of the nucleotide sequence overlaps with a fragment that is present in the second part of the nucleotide sequence, allowing recombination between the first and second part of the nucleotide sequence.

8. The set of constructs according to claim 7, wherein the first construct further comprises a recognition site for an endonuclease and a first region of homology with a target genome of a microorganism, and the second construct further comprises a second region of homology with the target genome of the microorganism, and a copy of the endonuclease recognition site, whereby a coding sequence that encodes the endonuclease and which is coupled to an inducible promoter is present on the first or second construct; and a part of the first region of homology with the target genome on the first construct is duplicated between the copy of the endonuclease recognition site and the second region of homology with the target genome on the second construct; or a part of the second region of homology with the target genome on the second construct is duplicated between first region of homology with the target genome and the endonuclease recognition site on the first construct.

9. The set of constructs according to claim 7 or claim 8, wherein the overlapping fragment of the nucleotide sequence encoding the selection marker is about 200 base pairs.
10. The set of constructs according to any one of claims 7-9, wherein the duplicated region of homology with the target genome on the first and second construct preferably is between 20 bp and 200 bp.

11. A method for altering a genome, preferably a target gene, in a microorganism of the genus *Saccharomycetaceae*, comprising providing the set of constructs according to any one of claims 7-10 to said microorganism, and selecting a microorganism in which the genome has been altered, preferably by selection of a microorganism that functionally expresses said guanidinobutyrase.

12. The method according to claim 11, wherein a microorganism is selected by culturing a microorganism in the presence of guanidinobutyrate as sole nitrogen source.

13. A kit comprising the set of constructs of any one of claims 7-10.

14. A method of altering a genome of a microorganism of the genus *Saccharomycetaceae*, comprising providing the set of constructs according to any one of claims 7-10 to said microorganism, and selecting a microorganism in which the genome has been altered by insertion of a functional, recombined selection marker.

15. The method according to claim 14, wherein a microorganism is selected by culturing a microorganism in the presence of guanidinobutyrate as sole nitrogen source.
FIG. 2A

Consensus M

XP_456325
XP_002498240
XP_716668
XP_461566
EIF45280
BA040353
AAEM0100201.1
XP_503530
EWF95653
XP_002552049
XP_001232356
XP_001826240
XP_001385334
XP_003679661
AAEM03000003.1

Consensus M

L---WG-WP F-GI-TPAHL-H
FIG. 2A, Cont’d

XP_456325 HKCLIDMERE—KKFDIGVIPFPDTAVSVFGARFGPAIRKASQRQTSMRFNGPRAIDN
XP_002498240 HTCLVD—N—PEFDIALIGVIPFDTAVSVFRGARFPGPAIRKASQRQTNQLRGPNMARAGIN
XP_716668 FCQLESSE—KQFDIGVIPFPDTAVSVYRGPFAPRAIDASQRQTNQLRGPNKALFD
XP_461566 NKKLDDFD—QEDYDIGVIPFPDTATSYRGPFAPRAIRASQRQTSRGNQRADEN
EIF45620 TSKLLSPE—SKYDIGVIPFPDTASYRGPFAPQAIRLARASQRQTNQLRGPNRAIDN
BA043083 QKSSLMDG—KQFDIGVIPFPAVSYRGPAQGPIQKSERAQTSRQPNSRAGIN
AADM01000201.1 --------------------------MPFDTVSRYRGPAIRASQRQTNQLRGPNRAIDN
XP_503530 VSSCDPTNASQEQFADVLGAVPGTPATVRSGFAIGAKASQRQTNQLRGPNRAIDN
EY955635 TKCLVDFE—TSFDIGVIPFPDTATTYRGPFAPRAIRASQRQTNQLRGPNRAIDN
XP_002552049 EKCLMRS—LDFDIGVIPFPDTAVSTRGPFAPKASQRQTNQLRGPNRAIDN
XP_001523956 TQCLVNPK—EESPQAVGFPDCTAVSVRGPFAPRAIRASQRQTNQLRGPNRAIDN
XP_001482640 KVCQLQFD—E1YDFAIQGFPDCTAVSTRGPFAPRRAISAAQRQTNQLRGPNRAIDN
XP_004164683 FKBVQEGD—EKYDIGVIPFPDTAVSYRGPFAPRAIRASQRQTNQLRGPNRAIDN
XP_001385334 HCKCLQEG—NTYDVVLGFPDTAVSYRGPFAPRAIRASQRQTNQLRGPNRAIDN
XP_003679661 QKSSLDDK—KTFDIGVIPFPDASYRGPFAPRRAIRASQRQTNQLRGPNRAIDN
AACB030000003.1 --------------------------MRGPNRAIDN

Consensus —KCL--- --FDIG-IGVFPDTAVS-R-GARFGP-AIR-ASQRO-S-RGPN-RA-IN

XP_456325 PYQWSAVVVDGDPVPTPMNDCLALKPMTTAYENLSSHESQISDWN---PRFVILGDHG
XP_002498240 PYQWNMAIICDCDPVPTPMNDQALCMAAYDLV--NGTTTSGRNTAGPALRLVSLGDHG
XP_716668 PYQSWARIIIDCDPVPTPDMDNSAAAYQMSAFKDLNKRK--SNTNEIP--PRYIALGDHG
XP_461566 PYTSWAVIICDCPVDPMNDLSAFKQMKFQIEELIRARRSKRTV---PRYIALGDHG
EIF45620 PYQWASLVDCDPVPTPMKDQMLDQMTQFAEELLLRRMLGDNAP---PYVALGDHG
BAO43083 PYNNWASAVDCDPVPTPMNMLQMTAYNLILLSHEKAESENLP---PLRVLIGDH
AADM01000201.1 PYQWAKMDICDPVPTPMNDQLKMTSAYELNLHSTTSDKLPL---PRLVTLDHG
XP_503530 LYDDWAKFVDCGVDAMHPLENDRLYALNQLRKOMRAHNTTSLNH---PRAILMGDH
EY955653 PYQWAKVDCDPVPTPMNDLALQMTFKEELLIIKRNKVQDDGSP---PKLVILGDHG
XP_002552049 PYQWAKVLEDVDCDPVPTPMNDLQMTMGAYNHNLNNKLNSLQKALFL---PRFAILGDHG
XP_001523956 PYQSWAVIICDCPVDPMNHLFQQMIDAFDE---QRSSANDSRPV---PYVVLGDH
XP_001482640 PYTSWAVIICDCPVDPMNHLFQQMIDAFDIEELR---RSSKAP---PYVVLGDH
XP_004164683 PYQSWARIVDCDPVPTPDMLQMTAEEELLLRRMADSSMFL---PRHILGDHG
XP_001385334 PYASWAKQVICDGVPFPMNDQLEAFTOMRGFEBELLRRSLNSSELFL---PYVALGDHG
XP_003679661 PYQWAKVDCDPVPTPDMSLALEKMTAYENLDDRRDSEYKSMF---PRLVILGDHG
AACB030000003.1 PYQWAKVDCDPVPTPMNQLKMTAYENSLDRN87AEFLPL---PRFAILGDHG

* *** **** * * * * * *  

FIG. 2A, Cont’d

Consensus

**S-LP-LRAL-K-YG----VIHF**

**HLD**

**TW-P-XPSWHS----SEFTHG**

**SMLW-A----EGLL**
FIG. 2A, Cont’d

XP_456325  SKIPSVDVVYISVDVLDPSAAPGTGTMVGGWMTRELIRIIRELDELNVLGADIVEEVS
XP_002498240  QIVPKNPVPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_716668  ATIPDFTATYISVDVLDPSAAPGTPGSGFRLPLLYLRSIDGTLTVGVADVEEVS
XP_461566  BTIPDPFTTPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
EIF45280  KIRPAPHVPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
BAO40383  SKIPDNAPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
AADM01000201.1  ERVPKIDPVYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_505350  DRYGNTNVYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
EWW55653  ERIKPDPPVPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_002552049  DTVSDKPDPVYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_001523956  ETIPDPSFTPVYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_001482640  ATIPDAPFTPVYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_000419683  KVVPANPFTPVPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_001385534  DTVPDSFTPVPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
XP_000367961  QLLPKQPPVPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS
AACB030000003.1  EHVPKDKPVPYISVDVLDPSAAPGTGTVBPGWLTRELSISLRQLQLDPLVVGAVVEEVS

Consensus

---P-D-P YISVDIDVLDP--APGTTG-E-GGW--RELI----R-LE-L--VGAD-VEEVS

XP_456325  PFPDP-PTETILSLAGAQIAYELITNMVKKPGIDPELIKHN-----LELSDLKT-QQQL--
XP_002498240  PFPDP-HADVTAMAQVAYEIIITNMVKPLEELTRKF-------------------
XP_716668  PFPDP-IAEITATNGAQIAYELITSVKKKNIDSDKLVSV---VHVFDP---
XP_461566  PFPDP-TAEITATNGAQVFEIITSVKKGSVG-HLYKNNPFKELLEIKSKNSDKQ---
EIF45280  PFPDP-QAETANNAQVFEIITSVKKGK-------------------
BAO40383  PFPDP-PSEITTLAGAQVAYELITNMVGGPLDPMVKNR-----LESKNNH-FQM---
AADM01000201.1  PFPDP-SGQDTLSLAQIAYELITNMVKPGAVAEIEKKN-----KQIAENLA-ENHVV--
XP_505350  PFPDP-SGQGPDVILLAAQVDISLALMNQGML-----S---TPQSTK--
EWW95653  PFPDP-PAEITANGAQVAYELITSVKKGK-------------------
XP_002552049  PFPDP-QSDITSTASSQIVYELITNMVGGLDPMQAN------K---NSEM-QQDKP--
XP_001523956  PFPDP-VSEVTATNGAQFELITSMVGGKNDNKVRTD---IBI---
XP_001482640  PFPDP-NAETATNGAQVAYDELITSMVGGKVD---LLVQR-----EEATPIK-SSEEV--
XP_000419683  PFPDP-IAEITATNGAQVFEIITSVKKGK-------------------
XP_001385534  PFPDP-TAEITSTNGAQIAFEIITSVKKGKIDFAVKN------KRELVKIV-HVNELEQ
XP_000367961  PFPDP-QSEITAIASSIAYELITSMVGGPIDPQMPIQEN-----GLEWNLAL-QDNNV--
AACB030000003.1  PFPDP-HADITSTASSQIVYELITNMVGGPDPFAVKNAN------*---**---*---*

Consensus

P---D---E-T-GAQ-ALEYI--MKKGP--
FIG. 2A, Cont’d

XP_456325
XM_002498240
XP_716668
XP_461566
EIF45280
BA040383
AADM1000201.1
XP_503530
EFW95653
XP_002552049
XP_001523956
XP_001482640
XP_004196483
XP_001385334
XP_003679661
AACS003000003.1

-------LGFSST-------DELNDKIQ-------KE-------QFV-------LQA-------
-------FSNM-------
-------
-------LOKQEIN-------RLIENKLQ-------BF-------ENIKFNLSEIEQLRST-------
-------KPL-------PVRED-------

------------Q--------NE--------QNY---------LHI-------

-------RVTEOCT-------TKILERLI-------QDAQQAV-------LHATNPPLL

-------QNLQNE-------YQSLGECQ-------RQ-------NFMESL-------LQGRV-------
-------FBS-------

-------LTLERL-------TSEIRQLK-------NM-------QAL-------FN-------

-------SSAALQD-------DKIKQXQDVLTRKN-------PFV-------ISN-------

-------EKKDFIDLEKAKKIEQKLKELKSERS-------SQ--------LLE-------LREIPF------

-------DFSAAXP-------DTYDKALL-------

Consensus

------------------------------------------
FIG. 3

1 mkvagfilga liqfsletgh vegenanatit enwgedwpfs qigtahalp hkclidmekk
61 fdigvgipv dfavsfrrga rfgpgairka srgqtsmgrf nfradinpyq dwasvvdcdg
121 vpvtmdncil alkmttaaye nllshesgqs dnnlprpvf llgdshisilp alralrkyg
181 rlavihfdsh ltdwpasky sfwhdsstef thqsmlwiah negllitenh ihaglrrlrs
241 gssfedydd dkgvfrhria deimddgiks ivekikskip sdpvpsiysd ivdipvsaap
301 gtgtnavgw mttrllirr eledlnlvga divspppf dteitslqag qiayelitnm
361 vkkpgipdel ikhnlksdk lttqgqllglf saptdelndk iqkegfvlga

FIG. 4A

Consensus

NF_289508
WF_000105576
YP_005016506
YP_001337000
YP_003614749
YP_001455807
-P_004532666
WF_006734551
AHE29794

<table>
<thead>
<tr>
<th></th>
<th>MSILGHQYDNSLVSNAGFLRPMNFQYPDSDADNWTVGPFDMAIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF_289508</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>WF_000105576</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>YP_005016506</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>YP_001337000</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>YP_003614749</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>YP_001455807</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>-P_004532666</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>WF_006734551</td>
<td>GRAGGRHGPAAIRQSTNLWAEHRNRFPMNFMERLAIVDVCDLVYAFQCGAREMEKQLQ</td>
</tr>
<tr>
<td>AHE29794</td>
<td>FRSQARGLPAVRAASVQLA-ELRYPYPGFDPPDDLAVICDQWDAHNPLGIAKPAIVE</td>
</tr>
</tbody>
</table>

Consensus

NF_289508
WF_000105576
YP_005016506
YP_001337000
YP_003614749
YP_001455807
-P_004532666
WF_006734551
AHE29794

<table>
<thead>
<tr>
<th></th>
<th>MSILGHQYDNSLVSNAGFLRPMNFQYPDSDADNWTVGPFDMAIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF_289508</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>WF_000105576</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_005016506</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_001337000</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_003614749</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_001455807</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>-P_004532666</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>WF_006734551</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>AHE29794</td>
<td>HARTLQGQAMILSTLQGHTPPYPLLAHARAGRKGPLSLILHFDHAKDWDADAPSSLNHG</td>
</tr>
</tbody>
</table>

Consensus

NF_289508
WF_000105576
YP_005016506
YP_001337000
YP_003614749
YP_001455807
-P_004532666
WF_006734551
AHE29794

<table>
<thead>
<tr>
<th></th>
<th>MSILGHQYDNSLVSNAGFLRPMNFQYPDSDADNWTVGPFDMAIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>NF_289508</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>WF_000105576</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_005016506</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_001337000</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_003614749</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>YP_001455807</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>-P_004532666</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>WF_006734551</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
<tr>
<td>AHE29794</td>
<td>HAEEKLLAAKGRLMSFGGDFHVTLPLLRAHAKKFGK-MALVHFDAHDTTYANGC-EDDHG</td>
</tr>
</tbody>
</table>
FIG. 4A, Cont’d

NP_289508  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
WP_000105576  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
YP_005016506  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
YP_001337000  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
YP_003614749  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
YP_001455807  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
-P_004532666  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
WP_006734551  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP
AHE29794  TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP

Consensus

TMFYTAPEGLIDPNHSQIGIRTEFDKNGTFVLDAQVNDRSVDVIAQVKIQVGDMP

NP_289508  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
WP_000105576  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
YP_005016506  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
YP_001337000  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
YP_003614749  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
YP_001455807  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
-P_004532666  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
WP_006734551  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI
AHE29794  VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI

Consensus

VYLTFIDCLDPAPAGTPGIPVIGGLTSRIAKLVRLGLKLNIVGMIVDVAPAYDQSEI

NP_289508  TALAAATLALMELYIQAAKGE---
WP_000105576  TALAAATLALMELYIQAAKGE---
YP_005016506  TALAAATLALMELYIQAAKGE---
YP_001337000  TALAAATLALMELYIQAAKGE---
YP_003614749  TALAAATLALMELYIQAAKGE---
YP_001455807  TALAAATLALMELYIQAAKGE---
-P_004532666  TALAAATLALMELYIQAAKGE---
WP_006734551  TALAAATLALMELYIQAAKGE---
AHE29794  TALAAATLALMELYIQAAKGE---

Consensus

TALAAATLALMELYIQAAKGE---
FIG. 7
FIG. 10
INTERNATIONAL SEARCH REPORT

International application No
PCT/NL2015/050238

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/55 C12N9/78 C12N15/81
ADD. C12N15/65

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category*</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>
| X         | DATABASE UniProt [Online]  
16 August 2004 (2004-08-16), "SubName: Full=KLLA0F27995p;", XP002731320, retrieved from EBI accession no. UNIPROT:Q6CB4  
Database accession no. Q6CB4 cited in the application  
the whole document ----- /-- | 5,6 |

X Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:
*A* document defining the general state of the art which is not considered to be of particular relevance
*E* earlier application or patent but published on or after the international filing date
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
*O* document referring to an oral disclosure, use, exhibition or other means
*P* document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search  
3 August 2015

Date of mailing of the international search report  
21/08/2015

Name and mailing address of the ISA/  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel: (+31-70) 340-2040, Fax: (+31-70) 340-3016

Authorized officer  
Madruga, Jaime
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>DUJON BERNARD ET AL: &quot;Genome evolution in yeasts&quot;&lt;br&gt;NATURE, NATURE PUBLISHING GROUP, UNITED KINGDOM,&lt;br&gt;vol. 430, no. 6995,&lt;br&gt;1 July 2004 (2004-07-01), pages 35-44,&lt;br&gt;XP002730312,&lt;br&gt;ISSN: 0028-0836</td>
<td>5,6</td>
</tr>
<tr>
<td>Y</td>
<td>the whole document</td>
<td>1-15</td>
</tr>
</tbody>
</table>
| Y        | TAN GUIHONG ET AL: "SMC, a simple method to rapidly assemble multiple fragments into one construct."
FRONTIERS IN BIOSCIENCE (ELITE EDITION) 2010,<br>vol. 2, 1 January 2010 (2010-01-01), pages 1105-1114, XP009180831,<br>ISSN: 1945-0508 the whole document | 1-15                 |
| Y        | Victoria Lundblad ET AL: "Manipulation of Cloned Yeast DNA"
| Y        | ROTHSTEIN RODNEY: "Targeting, disruption, replacement and allele rescue: integrative DNA transformation in yeast"
METHODS IN ENZYMOLOGY, ACADEMIC PRESS, US,<br>vol. 1974, 1 January 1991 (1991-01-01), pages 281-301, XP002958952,<br>ISSN: 0076-6879, DOI: 10.1016/0076-6879(91)94022-5 the whole document | 1-15                 |
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>

---
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>