Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3015440 C 2024/06/18

(11)(21) 3 015 440

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2017/02/21
(87) Date publication PCT/PCT Publication Date: 2017/08/31
(45) Date de délivrance/Issue Date: 2024/06/18
(85) Entrée phase nationale/National Entry: 2018/08/22
(86) N° demande PCT/PCT Application No.: FI 2017/050114
(87) N° publication PCT/PCT Publication No.: 2017/144777
(30) Priorité/Priority: 2016/02/22 (FI20165137)

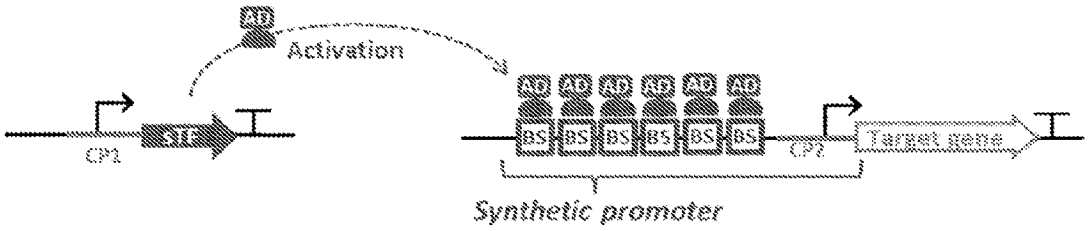
(51) Cl.Int./Int.Cl. *C12N 15/79* (2006.01)

(72) Inventeurs/Inventors:
MOJZITA, DOMINIK, FI;
RANTASALO, ANSSI, FI;
JANTTI, JUSSI, FI;
LANDOWSKI, CHRISTOPHER, FI;
KUIVANEN, JOOSU, FI

(73) Propriétaire/Owner:
TEKNOLOGIAN TUTKIMUSKESKUS VTT OY, FI

(74) Agent: GOWLING WLG (CANADA) LLP

(54) Titre : SYSTEME D'EXPRESSION POUR DES ORGANISMES EUCARYOTES
(54) Title: EXPRESSION SYSTEM FOR EUKARYOTIC ORGANISMS



(57) Abrégé/Abstract:
The present invention provides an expression system for a eukaryotic host, which comprises 1) an expression cassette comprising a core promoter, said core promoter controlling the expression of a DNA sequence encoding a synthetic transcription factor (sTF), and 2) one or more expression cassettes each comprising a DNA sequence encoding a desired product operably linked to a synthetic promoter, said synthetic promoter comprising a core promoter, and sTF-specific binding sites upstream of the core promoter. The present invention also provides a method for identifying universal core promoters for eukaryotic hosts, expression systems using universal core promoters, hosts comprising said systems, and methods for producing protein products in eukaryotic hosts .

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 August 2017 (31.08.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/144777 A1

(51) **International Patent Classification**:
C12N 15/79 (2006.01)

(21) **International Application Number**:
PCT/FI2017/050114

(22) **International Filing Date**:
21 February 2017 (21.02.2017)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
20165137 22 February 2016 (22.02.2016) FI

(71) **Applicant**: **TEKNOLOGIAN TUTKIMUSKESKUS VTT OY** [FI/FI]; Vuorimiehentie 3, 02150 Espoo (FI).

(72) **Inventors**: **MOJZITA, Dominik**; c/o VTT, PL 1000, 02044 VTT (FI). **RANTASALO, Anssi**; c/o VTT, PL 1000, 02044 VTT (FI). **JÄNTTI, Jussi**; c/o VTT, PL 1000, 02044 VTT (FI). **LANDOWSKI, Christopher**; c/o VTT, PL 1000, 02044 VTT (FI). **KUIVANEN, Joosu**; c/o VTT, PL 1000, 02044 VTT (FI).

(74) **Agent**: **BERGGREN OY**; P.O. Box 16 (Eteläinen Rautatiekatu 10 A), 00101 Helsinki (FI).

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

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

Published:

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

— *with sequence listing part of description (Rule 5.2(a))*

WO 2017/144777 A1

(54) **Title**: EXPRESSION SYSTEM FOR EUKARYOTIC ORGANISMS

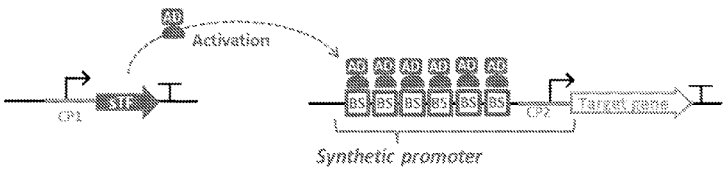


Fig. 1

(57) **Abstract**: The present invention provides an expression system for a eukaryotic host, which comprises 1) an expression cassette comprising a core promoter, said core promoter controlling the expression of a DNA sequence encoding a synthetic transcription factor (sTF), and 2) one or more expression cassettes each comprising a DNA sequence encoding a desired product operably linked to a synthetic promoter, said synthetic promoter comprising a core promoter, and sTF-specific binding sites upstream of the core promoter. The present invention also provides a method for identifying universal core promoters for eukaryotic hosts, expression systems using universal core promoters, hosts comprising said systems, and methods for producing protein products in eukaryotic hosts.

Expression system for eukaryotic organisms

Technical field

The present invention relates to an expression system for a eukaryotic host (such as a microorganism host), a host comprising said expression system, and a method for producing a desired protein product by using said host. Furthermore the present invention relates to a method for identifying a universal core promoter, a universal core promoter obtainable by said method, and an expression system, a eukaryotic organism host (such as a microorganism host) and method for producing a protein product by using a universal core promoter.

Background

Controlled and predictable gene expression is very difficult to achieve even in well-established hosts, especially in terms of stable expression in diverse cultivation conditions or stages of growth. In addition, for many potentially interesting industrial hosts, there is a very limited (or even absent) spectrum of tools and/or methods to accomplish expression of heterologous genes. In many instances, this prohibits the use of these (often very promising hosts) in industrial applications. In some hosts, specific inducing conditions need to be in place to achieve desirable expression of target genes. This results in specific requirements for culture media or downstream processing that ultimately increase production costs. Another problem in industrial hosts is the establishment of complex expression programs where it is desired to have specific expression levels of multiple genes simultaneously. This is, for instance, important for metabolic pathway engineering, where the individual genes encoding enzymes in production pathways need to be expressed (and the corresponding enzymes produced) in balanced ratio to ensure optimal metabolic flux towards the desired products.

In order to achieve predictable and/or stable expression patterns of the target genes in a host organism (in variable conditions) it is important that the expression of these genes is minimally affected by the intrinsic regulatory mechanisms of the host. This can be accomplished by use of non-native (heterologous) components (promoters, transcription factors, and inducing agents) in the engineered target gene expression systems. These expression systems are called orthogonal, if they

are not influenced by the host and also if they are not influencing the host in other ways than intended. The orthogonal expression systems still, however, rely on the host endogenous cellular functions, such as transcription and translation, so they have to fulfil certain criteria permitting their functionality in the host. These criteria are to some extent species (host)-specific, which makes it difficult to design an orthogonal system functional across a broad variety of very dissimilar species.

Typically, the current strategies for expression of heterologous genes employ use of endogenous (host specific) promoters in specific hosts (Hubmann et al. 2014 and Blumhoff et al. 2012). These promoters can be either inducible, or so-called constitutive, but in neither case are they orthogonal, because their function is dependent on specific factors existing in the host organism. Also, the use of host specific promoters prevents the inter-species transfer of these expression systems, which results in the necessity to develop customized expression systems for each host. The existing examples of inter-species transferable expression systems, based on the native host promoters, are limited to a narrow spectrum of closely related organisms, in which the promoters works. These include some yeast promoters, such as *Kluyveromyces lactis URA3* and *LEU2*, or *Schizosaccharomyces pombe HIS5* promoters functional in *Saccharomyces cerevisiae*. In filamentous fungi, for instance *gpdA* promoter of *Aspergillus nidulans* has been successfully used in *Aspergillus niger*, *Aspergillus fumigatus*, and *Trichoderma reesei*. These promoters are, however, mainly used for expression of selection marker genes in these organisms. They are not suitable for target gene expression (encoding a desired protein) and especially not for simultaneous expression of multiple genes (encoding a metabolic pathway), because their activity is strongly influenced by growth conditions or they confer an insufficient spectrum of transcriptional activities.

Several studies have reported the characterization and engineering of gene expression systems that employ synthetic (orthogonal) transcription factors (sTFs) and engineered sTF-dependent promoters to control the expression of target genes. The sTF-dependent promoters are composed of a variable number of sTF-binding sites linked to a core promoter. The number of binding sites in combination with a specific core promoter defines the level of expression of the target gene and it represents a significant improvement in expression level control compared to the systems which utilize host-specific promoters for the target gene expression. The sTFs used in these expression systems are, however, expressed from native (host-specific) promoters or modified native promoters, which makes these sys-

tems only partially orthogonal, and which prohibits their use in diverse species. Examples of the partially orthogonal expression systems include:

1) Expression system developed for *S. cerevisiae* , where the sTF is expressed from the *S. cerevisiae TDH3* promoter or from promoter combining the *TDH3* UAS and the *S. cerevisiae CYC1* core promoter, and the target genes are expressed from synthetic promoters containing a diverse number of sTF binding sites and *TDH3* or *CYC1* core promoters (Ito et al., 2015).

2) Expression system developed for *A. nidulans* and *A. niger*, where the sTF is expressed from the *A. nidulans gpdA* promoter, and the target gene is expressed from a synthetic promoter containing three binding sites for the sTF, *S. cerevisiae URA3* core promoter, and a 94 bp random sequence derived from *E. coli* (Pachlinger et al., 2005).

3) Expression system developed for *Arabidopsis thaliana*, where the sTF is expressed from the *A. thaliana* 35S promoter, or other *A. thaliana* promoter, or from a synthetic promoter containing four binding sites for the sTF and *A. thaliana* 35S minimal promoter. The minimal promoter probably refers to a core promoter in the referred publication. The target gene is expressed from the synthetic promoter containing four binding sites for the sTF and the *A. thaliana* 35S minimal promoter (US2002081667).

Although several gene expression systems have been disclosed in the prior art, there is still a need for gene expression systems for eukaryotic organism hosts (e.g. eukaryotic microorganism hosts) that can provide robust and stable expression, a broad spectrum of expression levels, and can be used in several different eukaryotic organism species and genera such as in several different eukaryotic microorganism species and genera. This would e.g. enable efficient transfer to and testing of engineered metabolic pathways simultaneously in several potential production hosts for functionality evaluation. Furthermore, a true orthogonal expression system would provide benefits to the scientific community who study eukaryotic organisms.

Summary

One objective of the present invention is to provide orthogonal expression systems which are functional (transferable) in a large spectrum of eukaryotic organisms such as eukaryotic microorganisms. Such expression systems would overcome

the need to use host-native DNA sequences in constructing the expression systems and, therefore, establishing expression systems not dependent on the intrinsic transcriptional regulation of the expression host.

A further objective of the invention is to provide expression systems, which allow robust, stable, and predictable expression levels of target genes, and which are not influenced by the cultivation conditions or developmental or growth stages of the host organism.

The motivation for the present invention is based on the finding that 1) the use of the host-specific promoters, or their parts, for expressing the sTFs, and 2) the use of species-specific core promoters in the sTF-dependent promoters controlling the expression of the target genes are the main reasons why the current expression systems based on sTFs cannot be transferred between diverse species without loss of their function.

The present invention shows that it is advantageous to use a core promoter alone for the expression of a sTF. This allows low, constitutive expression of sTF in the host (e.g. microorganism host).

Furthermore, the present invention shows that it is possible to develop a method to identify core promoters that are functional in distant species.

In addition, the present invention shows that it is possible to construct expression systems based on these core promoters functional in diverse species, which allow tunable expression levels of target genes across a large spectrum of eukaryotic organisms (e.g. eukaryotic microorganisms).

Hence, the present invention provides an expression system for a eukaryotic host (e.g. microorganism host), which comprises:

(a) an expression cassette comprising a core promoter,
said core promoter being the only “promoter” controlling the expression of a DNA sequence encoding synthetic transcription factor (sTF), and

(b) one or more expression cassettes each comprising a DNA sequence encoding a desired protein product operably linked to a synthetic promoter,
said synthetic promoter comprising a core promoter identical to (a) or another core promoter, and sTF-specific binding sites upstream of the core promoter.

The present invention provides also a eukaryotic host, such as a eukaryotic microorganism host, comprising the expression system.

Furthermore, the present invention provides a method for producing a desired protein product (or multiple desired protein products simultaneously) in a eukaryotic host comprising cultivating the eukaryotic host under suitable cultivation conditions.

Furthermore, the present invention provides a method for producing a desired protein product (or multiple desired protein products simultaneously) in a eukaryotic microorganism host comprising cultivating the eukaryotic microorganism host under suitable cultivation conditions.

The present invention provides also a method for identifying universal core promoters for eukaryotic hosts.

The identification method comprises the following steps:

- constitutively expressing a synthetic transcription factor, sTF, in *Saccharomyces cerevisiae*,
- in the same host co-expressing a reporter gene operably linked to a sTF-dependent test promoter, said sTF-dependent test promoter comprising a core promoter to be tested, and sTF binding sites upstream to that,
- allowing said reporter gene to be expressed under the test promoter in the presence of activation by the sTF,
- assessing the level of expression of the reporter gene, and
- selecting from the tested core promoters, core promoters showing at least 40% as high expression of the reporter gene as obtained with *S. cerevisiae PGK1* core promoter tested in the same reporter system;
- In specific cases, also selecting core promoters showing lower than 40% expression of the reporter gene as compared to the reporter gene expression obtained with *S. cerevisiae PGK1* core promoter tested in the same reporter system.

Furthermore, the present invention provides a universal core promoter (UCP). The universal core promoter is obtainable by the disclosed identification method.

A universal core promoter (UCP) typically comprises a DNA sequence containing the 5´-upstream region of a eukaryotic gene, starting 10 – 50 bp upstream of a TATA-box and ending 9 bp upstream of the ATG start codon. The distance between the TATA-box and the start codon is preferably no greater than 180 bp and no smaller than 80 bp. The UCP typically comprises also a DNA sequence comprising random 1-20 bp at its 3'-end. In one embodiment a UCP typically comprises a DNA sequence having at least 90% sequence identity to said 5´-upstream region of a eukaryotic gene, and a DNA sequence comprising random 1-20 bp at its 3'-end.

Furthermore, the present invention provides an expression system for a eukaryotic host, which comprises

(a) an expression cassette comprising a UCP,
said UCP controlling the expression of a DNA sequence encoding synthetic transcription factor (sTF), and

(b) one or more expression cassettes each comprising a DNA sequence encoding a desired protein product operably linked to a synthetic promoter,

said synthetic promoter comprising a UCP identical to (a) or another UCP, and sTF-specific binding sites upstream of the UCP.

In addition, the present invention provides a eukaryotic host (e.g. a eukaryotic microorganism host) comprising an expression system using universal core promoters.

The present invention provides also a method for producing a desired protein product (or multiple desired protein products simultaneously) in a eukaryotic host (e.g. a eukaryotic microorganism host) using an expression system with universal core promoters.

The present invention thus provides an orthogonal expression system which is functional (transferable) in a large spectrum of eukaryotic organisms or eukaryotic microorganisms, which allows robust, stable, and predictable expression levels of target genes, and is not influenced by cultivation conditions or developmental or growth stages of the host organism.

The expression system provided by the present invention simplifies and focuses the genetic tools needed for constructing new expression hosts. Currently there is a wide array of expression systems that are highly organism and species specific. With the present invention, industry and wider scientific community working on eukaryotic organisms can adopt a smaller, common set of orthogonal expression tools. This would benefit the community and drive forward new innovations in the field.

Brief description of Figures

Figure 1 depicts a scheme of an expression system for expression of a single gene in a eukaryotic organism (e.g. a eukaryotic microorganism).

Figures 2A, 2B and 2C depict a scheme of the screening method for selecting UCPs from the candidate core promoters.

Figure 3 depicts a scheme of an expression system utilizing the UCPs for simultaneously regulating the expression of multiple genes in a eukaryotic organism such as a eukaryotic microorganism.

Figure 4 depicts examples of the expression systems functional/transferable in diverse organisms or microorganisms.

Figures 5A and 5B depict testing of different versions of the sTFs and assessment of modulation of the expression system's performance in *Saccharomyces cerevisiae* by fluorometry.

Figures 6A and 6B depict the analysis of the expression systems in diverse fungal hosts. Quantitative analysis of the reporter gene expression determined by fluorescence flow cytometry (6A) and by fluorometry (6B).

Figures 7A, 7B, 7C and 7D depict the analysis of the tunable expression levels in different hosts (*Pichia kudriavzevii*, *Aspergillus niger*, and *Trichoderma reesei*) by fluorescence flow cytometry and western blotting.

Figures 8A and 8B depict the scheme of the expression system (8A) and the analysis of a reporter gene expression in *Kazachstania exigua* (8B) by quantitative real-time PCR (qPCR).

Figures 9A, 9B, 9C and 9D depict the analysis of the protein production in diverse expression hosts (*Trichoderma reesei* and *Pichia pastoris*) containing the expression system.

Detailed description

Definitions

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

DNA refers to deoxyribonucleic acid.

Codon is a tri-nucleotide unit which is coding for a single amino acid in the genes that code for proteins. The codons encoding one amino acid may differ in any of their three nucleotides. Different organisms have different frequency of the codons in their genomes, which has implications for the efficiency of the mRNA translation and protein production.

Coding sequence refers to a DNA sequence that encodes a specific RNA or polypeptide (i.e. a specific amino acid sequence). The coding sequence could, in some instances, contain introns (i.e. additional sequences interrupting the reading frame, which are removed during RNA molecule maturation in a process called RNA splicing). If the coding sequence encodes a polypeptide, this sequence contains a **reading frame**.

Reading frame is defined by a start codon (AUG in RNA; corresponding to ATG in the DNA sequence), and it is a sequence of consecutive codons encoding a polypeptide (protein). The reading frame is ending by a stop codon (one of the three: UAG, UGA, and UAA in RNA; corresponding to TAG, TGA, and TAA in the DNA sequence). A person skilled in the art can predict the location of open reading frames by using generally available computer programs and databases.

Eukaryotic **Promoter** is a region of DNA necessary for initiation of transcription of a gene. It is upstream of a DNA sequence encoding a specific RNA or polypeptide (**coding sequence**). It contains an upstream activation sequence (UAS) and a core promoter. A person skilled in the art can predict the location of a promoter by using generally available computer programs and databases.

Core promoter (CP) is a part of a eukaryotic promoter and it is a region of DNA immediately upstream (5'-upstream region) of a coding sequence which encodes a polypeptide, as defined by the start codon. The core promoter comprises all the general transcription regulatory motifs necessary for initiation of transcription, such as a TATA-box, but does not comprise any specific regulatory motifs, such as UAS sequences (binding sites for native activators and repressors).

Core promoter is defined for the purpose of the present invention as a DNA sequence containing: 1) a 5´-upstream region of a **highly expressed gene** starting 10-50 bp upstream of the **TATA box** and ending 9 bp upstream of the start codon, where the distance between the TATA box and the start codon is no greater than 180 bp and no smaller than 80 bp, 2) random 1-20 bp, typically 5 to 15 or 6 to 10, which are located in place of the 9bp of the DNA region (1) immediately upstream of the start codon; or as a DNA sequence containing : 1) a DNA sequence having at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% sequence identity to said 5´-upstream region and 2) random 1-20 bp, typically 5 to 15 or 6 to 10, which are located in place of the 9bp of the DNA region (1) immediately upstream of the start codon.

A highly expressed gene in an organism in the context of this invention is a gene which has been shown in that organism to be expressed among the top 3% or 5% of all genes in any studied condition as determined by transcriptomics analysis, or a gene, in an organism where the transcriptomics analysis has not been performed, which is the closest sequence homologue to the highly expressed gene.

TATA-box is **defined for the purpose of the present invention** as a DNA sequence (TATA) upstream of the start codon, where the distance of the TATA sequence and the start codon is no greater than 180 bp and no smaller than 80 bp. In case of multiple sequences fulfilling the description, the TATA-box is defined as the TATA sequence with smallest distance from the start codon.

Transcription factor refers to a protein that binds to specific DNA sequences present in the UAS, thereby controlling the rate of transcription, which is performed by RNA II polymerase. Transcription factors perform this function alone or with other proteins in a complex, by promoting (as an activator), or blocking (as a repressor) the recruitment of RNA polymerase to core promoters of genes.

Synthetic transcription factor (sTF) refers to a protein which functions as a transcription factor, but is not a native protein of a host organism. In the context of this

invention, the sTF is an artificial protein which typically comprises a DNA-binding protein of prokaryotic origin, a nuclear localization signal, and a transcription activation domain of viral origin.

Synthetic promoter refers to a region of DNA which functions as a eukaryotic promoter, but it is not a naturally occurring promoter of a host organism. It contains an upstream activation sequence (UAS) and a core promoter, wherein the UAS, or the core promoter, or both elements, are not native to the host organism. In the context of this invention, the synthetic promoter comprises (usually 1-10, typically 1, 2, 4 or 8) sTF-specific binding sites (synthetic UAS – sUAS) linked to a core promoter.

DNA binding domain or DBD refers to the region of a protein, typically specific protein domain, which is responsible for interaction (binding) of the protein with a specific DNA sequence.

Universal core promoter (UCP) is a core promoter which confers sufficient (usually but not necessarily at least 40% of) reporter expression or activity level, such as fluorescence level, obtained with the *Saccharomyces cerevisiae PKG1* core promoter tested in **a CP-screening system** as disclosed in the present invention. A core promoter selected by using this system typically provides sufficient expression of a transcription factor in various species and genera of eukaryotic organisms.

An orthogonal expression system means here an expression system consisting of heterologous (non-native) core promoters, transcription factor(s), and transcription-factor-specific binding sites. Typically, **the orthogonal expression system** is functional (transferable) in diverse eukaryotic organisms such as eukaryotic microorganisms.

CP-screening system is constructed in *Saccharomyces cerevisiae* and it comprises a *Saccharomyces cerevisiae* strain constitutively expressing a sTF and preferably a centromeric type reporter plasmid assembled with the core promoter to be tested. The reporter plasmid typically contains binding sites specific for the sTF, a reporter gene, such as mCherry gene, and a terminator, such as the *ADH1* terminator for the mCherry gene. The tested core promoter is inserted between the sTF binding sites and the reporter gene. The tested core promoter typically comprises at its 3'-end a sequence comprising 1-20 random nucleotides, such as sequence TTAATTAAA, and typically including restriction sites. The function of the

core promoter is assessed by a reporter measurement, such as fluorescence measurement of the resulting strain and compared to a control strain where the core promoter is the *Saccharomyces cerevisiae PKG1* core promoter.

A centromeric plasmid refers here to a single or low copy number plasmid used in *S. cerevisiae*. This plasmid is containing DNA regions functional as a centromere (CEN sequence) and as an autonomously replicating sequence (ARS) in *S. cerevisiae*. The ARS sequence provides replication origin and the CEN sequence regulates replication and distribution of the plasmids during cell division which makes the centromeric plasmid analogous to a chromosome.

Sufficient expression of a transcription factor is defined as an expression level of a transcription factor which leads to transcription activation of a gene or genes which are under the control of the transcription factor-dependent promoter(s).

Eukaryotic organism is defined in the context of this invention as an organism belonging to: 1) Fungal kingdom, including yeast, such as classes *Saccharomycetales*, including but not limited to species *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Candida krusei (Pichia kudriavzevii)*, *Pichia pastoris* (*Komagataella pastoris*), *Eremothecium gossypii*, *Kazachstania exigua*, *Yarrowia lipolytica*, and others; or *Schizosaccharomycetes*, such as *Schizosaccharomyces pombe*; filamentous fungi, such as classes *Eurotiomycetes*, including but not limited to species *Aspergillus niger*, *Aspergillus nidulans*, *Penicillium chrysogenum*, and others; *Sordariomycetes,* including but not limited to species *Trichoderma reesei*, *Myceliophthora thermophile*, and others; or *Mucorales*, such as *Mucor indicus* and others. 2) Plant kingdom, including flowering plants, such as orders *Solanales*, including but not limited to genus *Nicotiana* (*N. benthamiana*), *Solanum* (*S. tuberosum*), *Lycopersicon* (*L. esculentum*), *Capsicum* (*C. anuum*) and others; *Brassicales* including but not limited to genus *Arabidopsis* (*A. thaliana*), *Brassica* (*B. napus*), and others; *Poales* including but not limited to species *Avena sativa*, *Secale cereale*, *Zea mays*, *Triticum* spp., *Oryza sativa*, *Hordeum vulgare*, *Sorghum bicolor*, *Saccharum officinarum*, and others; *Fabales* including but not limited to species *Phaseolus* spp., *Vigna* spp., *Glycine max*, *Pisum sativum*, *Lens culinaris, Cicer arietinum* and others; *Malpighiales*, including but not limited to genus *Populus*, and others; *Pinales*, including but not limited to genus *Pinus*, and others; or *Arecales* including but not limited to species *Elaeis guineensis*, *Cocos nucifera*, and others; and green algae, such as classes *Chlorophyceae*, including but not limited to genus *Chlamydomonas* (*C. reinhardtii*); or *Trebouxiophyceae*, including but not lim-

ited to species *Chlorella* spp., and others. 3) Animal kingdom, including mammals (Mammalia), including but not limited to species *Mus musculus* (mouse), *Cricetulus griseus* (hamster), *Homo sapiens* (human), and others; insects, including but not limited to species *Mamestra brassicae*, *Spodoptera frugiperda*, *Trichoplusia ni*, *Drosophila melanogaster,* and others.

Eukaryotic microorganism is defined in the context of the invention as a microorganism including yeast, such as classes *Saccharomycetales*, including but not limited to species *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Candida krusei (Pichia kudriavzevii)*, *Pichia pastoris* (*Komagataella pastoris*), *Eremothecium gossypii*, *Kazachstania exigua*, *Yarrowia lipolytica*, and others; *Schizosaccharomycetes*, such as *Schizosaccharomyces pombe*; and filamentous fungi, such as classes *Eurotiomycetes*, including but not limited to species *Aspergillus niger*, *Aspergillus nidulans*, *Penicillium chrysogenum*, and others; *Sordariomycetes,* including but not limited to species *Trichoderma reesei*, *Myceliophthora thermophile*, and others; *Mucorales*, such as *Mucor indicus* and others.

The present invention provides an expression system for a eukaryotic host, which comprises

(a) an expression cassette comprising a core promoter;
the core promoter being the only promoter for controlling the expression of a DNA sequence encoding synthetic transcription factor (sTF), and

(b) one or more expression cassettes each comprising a DNA sequence encoding a desired protein product operably linked to a synthetic promoter;
the synthetic promoter comprises a core promoter, which is identical to the core promoter in (a) or another core promoter, and one or more sTF-specific binding sites upstream of the core promoter.

The core promoter typically comprises a DNA sequence containing the 5´-upstream region of a eukaryotic gene, starting 10 – 50 bp upstream of a TATA-box and ending 9 bp upstream of the ATG start codon. The distance between the TATA-box and the start codon is no greater than 180 bp and no smaller than 80 bp. The core promoter typically comprises also a DNA sequence comprising random 1-20 bp at its 3'-end. In one embodiment the core promoter typically comprises a DNA sequence having at least 90% sequence identity to said 5´-upstream re-

gion of a eukaryotic gene, and a DNA sequence comprising random 1-20 bp at its 3'-end.

The DNA sequence encoding the synthetic transcription factor (sTF) typically comprises a prokaryotic transcription regulator, a nuclear localization signal, and a transcription activation domain.

The CPs used in the expression system can be different, or the first one, CP1, can be identical to the second one CP2, (or the third one CP3, or the fourth one CP4). This is illustrated in Figures 1 and 3.

The two expression cassettes ((a) and (b)) can be introduced to a eukaryotic host (typically integrated into a genome) as two individual DNA molecules, or as one DNA molecule in which the two (or more) expression cassettes are connected (fused) to form a single DNA.

In specific applications, where the target gene is a native (homologous) gene of a host organism, the synthetic promoter can also be inserted immediately upstream of the target gene coding region in the genome of the host organism, possibly replacing the original (native) promoter of the target gene.

More specifically, the expression system thus comprises two DNA-parts, which are assembled into the expression system comprising at least two individual expression cassettes:

(a) a sTF – synthetic transcription factor – cassette, which comprises a CP controlling expression of a gene encoding a fusion protein (sTF), the sTF itself, and a terminator. The sTF comprises a DNA-binding protein derived from prokaryotic origin, typically bacterial transcription regulators, such as from the TetR family; nuclear localization signal, such as the SV40 NLS; and a transcription activation domain, such as the VP16 or VP64 activation domain; and

(b) a target gene expression cassette, which comprises a synthetic promoter, which comprises a variable number of sTF-binding sites, usually 1 to 10, typically 1, 2, 4 or 8, separated by 0-20, typically 5 -15, random nucleotides, a CP, a target gene, and a terminator.

The composition of the example expression system is illustrated in Figure 1.

The present invention is based on the idea to use a core promoter (CP), instead of a full promoter, for expression of a synthetic transcription factor (sTF). Some CPs

can sustain low level of transcription when placed in front of a gene. Due to the absence of specific regulatory sequences required for conditional transcription control which are present in full promoters (typically in the upstream activating sequence – UAS), this transcription is constitutive – that is constant in all growth or metabolic conditions. Because the general transcription machinery is evolutionarily conserved, some of the CPs can function in very diverse species. These features are used in the invention for construction of species-transferable expression systems.

The constitutive low expression of the sTF gene facilitated by a CP provides a sufficient amount of a synthetic transcription factor, which binds to its specific binding sites on the synthetic promoter of the target gene and activates its expression. The number of the binding sites is proportional to the expression level of the target gene(s), where more binding sites results in higher expression. The synthetic promoter comprises, in addition to the sTF-binding sites, also a CP. The choice of the CP in the synthetic promoter controlling the expression of the target gene(s) is also important for the expression level of the target gene(s). The combination of the sTF-binding sites and the CP can result in a range of expression levels which can be modulated from very low to very high. At the high end, the expression achieved by this system exceeds the expression levels of the most highly expressed native genes in a host organism.

Figure 1 illustrates an example of a scheme of an expression system for expression of a single gene in a eukaryotic organism or microorganism. The synthetic transcription factor (sTF) expression cassette contains a CP (CP1), a sTF coding sequence, and a terminator. The CP1 provides constitutive low expression of the sTF. Therefore the sTF is present in a host cell in a constant level all the time, in all growth conditions, and all developmental and growth stages. The target gene expression cassette contains a synthetic promoter, a target gene coding sequence, and a terminator. The synthetic promoter comprises multiple sTF-specific binding sites (usually 1-10, typically 1, 2, 4 or 8; forming a synthetic upstream activating sequence – sUAS), and a CP (CP2). The target gene encodes a protein product of interest.

The transcription activity of the CP1, the “signal”, is “amplified” by the sTF bound to the sUAS. This leads to activation of transcription on the CP2, resulting in expression of the target gene. As discussed above, the two expression cassettes can be introduced into a eukaryotic host (typically integrated into a genome) as

two individual DNA molecules, or as one DNA molecule in which the two cassettes are connected (fused) into a single DNA. In specific applications, where the target gene is a native (homologous) gene of a host organism, the synthetic promoter can also be inserted immediately upstream of the target gene coding region in the genome of the host organism. The CPs used in the expression system can be different, or the CP1 can be identical to the CP2.

The present invention also provides a eukaryotic host (e.g. a eukaryotic microorganism host) which comprises the expression system as disclosed herein.

A eukaryotic organism refers here in particular to 1) fungal species including yeast, such as species from classes *Saccharomycetales*, including but not limited to *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Candida krusei (Pichia kudriavzevii)*, *Pichia pastoris* (*Komagataella pastoris*), *Eremothecium gossypii*, *Kazachstania exigua*, *Yarrowia lipolytica*, and others; *Schizosaccharomycetes*, such as *Schizosaccharomyces pombe*; and filamentous fungi species, such as those from classes *Eurotiomycetes*, including but not limited to *Aspergillus niger*, *Aspergillus nidulans*, *Penicillium chrysogenum*, and others; *Sordariomycetes,* including but not limited to *Trichoderma reesei*, *Myceliophthora thermophile*, and others; *Mucorales*, such as *Mucor indicus* and others; 2) plant species including flowering plants, such as species from orders *Solanales*, including but not limited to *Nicotiana benthamiana*, *Solanum tuberosum*, *Lycopersicon esculentum*, *Capsicum anuum* and others; *Brassicales*, including but not limited to *Arabidopsis thaliana*, *Brassica napus*, and others; *Poales*, including but not limited to *Avena sativa*, *Secale cereale*, *Zea mays*, *Triticum* spp. *Oryza sativa*, *Hordeum vulgare*, *Sorghum bicolor*, *Saccharum officinarum*, and others; *Fabales* including but not limited to *Phaseolus* spp., *Vigna* spp., *Glycine max*, *Pisum sativum*, *Lens culinaris, Cicer arietinum* and others; *Malpighiales*, including but not limited to *Populus* sp., and others; *Pinales*, including but not limited to *Pinus* sp., and others; or *Arecales* including but not limited to *Elaeis guineensis*, *Cocos nucifera*, and others; and green algae species, including but not limited to *Chlamydomonas reinhardtii*, *Chlorella* spp. and others; 3) Animal species including but not limited to mammals (Mammalia), including but not limited to species *Mus musculus* (mouse), *Cricetulus griseus* (hamster), *Homo sapiens* (human), and others; insect species, including but not limited to species *Mamestra brassicae*, *Spodoptera frugiperda*, *Trichoplusia ni*, *Drosophila melanogaster,* and others.

The present invention also provides a method for producing a desired protein product in a eukaryotic host (e.g. microorganism host) comprising cultivating the host under suitable cultivation conditions.

By suitable cultivation conditions are meant any conditions allowing survival or growth of the host organism, and/or production of the desired product in the host organism. Desired product can be a product of the target gene or genes (protein or proteins), or compound produced by a protein (enzyme) or by a metabolic pathway. In the present context the desired product is typically a protein (enzyme) product.

The present invention also provides a gene expression system which is functional in several different eukaryotic species and genera. The key element in the system is a core promoter which facilitates expression in several species. Such a core promoter is here called **universal core promoter** – UCP.

This property, so called basal transcription activity, is based on efficient recruitment of the RNA polymerase II complex to the core promoter; and it results in low but stable expression level in all cultivation and growth (developmental) conditions. This low constitutive signal is amplified by a synthetic transcription factor (sTF), whose expression is controlled by the UCP, to adjustable expression level of target genes (native or heterologous). Each target gene is under the control of an engineered promoter and comprises a selected number of sTF-specific binding sites and a UCP. The combination of the sTF-specific binding sites and the UCP defines the expression level of the target gene.

This provides means to control expression in diverse hosts, including those with undeveloped know-how. Applications of the use of UCPs are protein production, metabolic engineering and artificial genetic regulatory networks.

Furthermore, the system can be used as a platform to identify new UCPs with novel properties.

The present invention provides a method for identifying a universal core promoter for eukaryotic hosts. The method comprises

- constitutively expressing a synthetic transcription factor, sTF, in *Saccharomyces cerevisiae*,

- co-expressing in the same host a reporter gene operably linked to a sTF-dependent test promoter, said sTF-dependent test promoter comprising a core promoter to be tested, and sTF binding sites upstream to that,
- allowing said reporter gene to be expressed under the test promoter in the presence of activation by the sTF,
- assessing the level of expression of the reporter gene, and
- selecting from the tested core promoters, core promoters showing at least 40% as high expression of the reporter gene as obtained with *S. cerevisiae PGK1* core promoter tested in the same reporter system.
- in specific cases, also selecting core promoters showing lower than 40% level of reporter expression

More specifically, the method optimally comprises the use of a circular centromeric plasmid comprising sTF specific binding sites operably linked to the tested core promoter, and a reporter gene.

The DNA sequence encoding synthetic transcription factor (sTF) typically comprises a DNA sequence encoding a DNA-binding protein of prokaryotic origin, a nuclear localization signal, and a transcription activation domain. The sTF comprises a DNA-binding protein derived from prokaryotic, typically bacterial origin, transcription regulators, such as a protein from the TetR family; a nuclear localization signal, such as the SV40 NLS; and a transcription activation domain, such as the VP16 or VP64 activation domain.

The promoter to be tested is selected from the promoters of eukaryotic genes expressed to the level of the highest 3% or 5% of all genes in any condition in the given eukaryotic organism.

The present invention provides a universal core promoter (UCP), in which the core promoter is obtainable by the identification method as disclosed herein.

Typically a universal core promoter comprises a DNA sequence containing 1) the 5´-upstream region of a eukaryotic gene, starting 10 – 50 bp upstream of a TATA-box, and ending 9 bp upstream of the ATG start codon; and 2) a random 1-20 bp DNA sequence which is located in place of the 9bp of the DNA region (1) immediately upstream of the start codon. The distance between the TATA-box and the start codon of the original eukaryotic gene is no greater than 180 bp and no

smaller than 80 bp. In one embodiment the core promoter comprises a DNA sequence having at least 90% sequence identity to said 5′-upstream region, and a random 1-20 bp DNA sequence which is located in place of the 9bp of the DNA region (1) immediately upstream of the start codon.

The selection of the CPs functional in distant organisms is carried out in *Saccharomyces cerevisiae*, and the sources of the candidate CPs are preferably (but not necessarily) industrially relevant organisms, preferably (but not necessarily) distant in terms of evolutionary divergence or in other features, such as genome architecture or GC-content.

The selection of the candidate CPs is based on the level of expression of the genes in the selected source organisms, containing the candidate CP in their promoters. Another selection criterion is the presence of a TATA-box in the candidate CP (Figure 2A).

In one embodiment the screen for functional CPs is advantageously performed by *in vivo* assembling the candidate CP with the sTF-dependent reporter cassette expressed in a *S. cerevisiae* strain constitutively expressing the sTF (Figure 2B). The resulting strains are tested for a level of a reporter, preferably fluorescence, and these levels are compared to a control strain where the *S. cerevisiae PGK1* core promoter is used in the reporter construct. The candidate CPs, which facilitate sufficient reporter, preferably fluorescence, levels (usually but not necessarily higher than 40% of) the control strain (Figure 2C), and therefore fulfil the criteria of the screening are called universal core promoters, UCPs. The selected CPs and UCPs are used for constructing of expression systems.

The resulting expression systems are functional in eukaryotic hosts. These hosts include all eukaryotic organisms, in particular: 1) Fungal microorganisms including filamentous fungi and yeasts, in particular organisms from the following taxa: **A)** *Saccharomycetales*, including but not limited to species *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Candida krusei (Pichia kudriavzevii)*, *Pichia pastoris* (*Komagataella pastoris*), *Eremothecium gossypii*, *Kazachstania exigua*, *Yarrowia lipolytica*, and others; *Schizosaccharomycetes*, such as *Schizosaccharomyces pombe*; **B)** *Eurotiomycetes*, including but not limited to species *Aspergillus niger*, *Aspergillus nidulans*, *Penicillium chrysogenum*, and others; **C)** *Sordariomycetes*, including but not limited to species *Trichoderma reesei*, *Myceliophthora thermophile*, and others; **D)** *Mucorales*, such as *Mucor indicus* and others. 2) Plant organisms, including flowering plants and green algae, in particular organisms from

the following taxa: **E)** *Solanales*, including but not limited to species *Nicotiana benthamiana*, *Solanum tuberosum*, *Lycopersicon esculentum*, *Capsicum anuum*, and others; **F)** *Brassicales*, including but not limited to species *Arabidopsis thaliana*, *Brassica napus*, and others; **G)** *Poales*, including but not limited to species *Avena sativa*, *Secale cereale*, *Zea mays*, *Triticum* spp., *Oryza sativa*, *Hordeum vulgare*, *Sorghum bicolor*, *Saccharum officinarum*, and others; **H)** *Fabales* including but not limited to species *Phaseolus* spp., *Vigna* spp., *Glycine max*, *Pisum sativum*, *Lens culinaris, Cicer arietinum* and others; **I)** *Malpighiales*, including but not limited to species *Populus* sp., and others; **J)** *Pinales*, including but not limited to species *Pinus* sp., and others; **K)** *Arecales* including but not limited to species *Elaeis guineensis*, *Cocos nucifera*, and others; **L)** *Chlorophyceae*, including but not limited to species *Chlamydomonas reinhardtii*, and others; **M)** *Trebouxiophyceae*, including but not limited to species *Chlorella* spp., and others. 3) Animal organisms, in particular organisms from the following taxa: **N)** mammals (*Mammalia*), including but not limited to species *Mus musculus* (mouse), *Cricetulus griseus* (hamster), *Homo sapiens* (human), and others; **O)** insects (*Insecta*), including but not limited to species *Mamestra brassicae*, *Spodoptera frugiperda*, *Trichoplusia ni*, *Drosophila melanogaster,* and others.

Figures 2A, B and C illustrate an example of a scheme of the screening method used for selecting UCPs from the candidate core promoters.

Figure 2A illustrates a scheme of selection of a candidate core promoter in a eukaryotic organism. The DNA region immediately upstream of a gene of any eukaryotic organism, which belongs to a group of top 3% or 5% most highly expressed genes in any condition, is analyzed for presence of TATA sequence (TATA-box) within -180bp and -80bp upstream of a start codon (ATG). If more than one TATA sequence appears in this region, then the one closest to the ATG (start codon) is chosen as a TATA-box. The sequence starting 10-50bp upstream of the TATA-box and ending 9bp upstream of the ATG (start codon) is selected for the core promoter screen.

Figure 2B illustrates a *Saccharomyces cerevisiae* strain constitutively expressing a sTF. It is co-transformed typically with a linearized centromeric (single or low copy number) plasmid, and a library, or individual versions, of the core promoters to be tested. The centromeric plasmid contains typically for example 4 sTF binding sites, the reporter gene, such as mCherry gene, and it is linearized between these two features as shown in the figure. Each core promoter DNA fragment comprises

the selected DNA sequence (Figure 2A) followed by a sequence comprising random nucleotides, typically containing restriction sites, here one useful sequence is for example TTAATTAAA, and flanked by 20-50 bp long DNA sequences on each end. These flanking sequences are homologous to each end of the linearized plasmid, the 5′-flanking sequence is homologous to a region partly covering the sTF binding sites in the linearized plasmid, and the 3′-flanking sequence is homologous to the 5′-end of the reporter gene, such as mCherry gene, open reading frame. After the transformation, the plasmid is assembled *in vivo* by an intrinsic homologous recombination machinery of the *Saccharomyces cerevisiae* yeast, resulting in the circular centromeric plasmid comprising the sTF binding sites, followed by a core promoter including a sequence, such as TTAATTAAA, and the reporter gene, such as mCherry gene. The resulting strains are analyzed for the reporter, such as a red fluorescence caused by the produced mCherry protein. The level of intensity of the reporter, such as fluorescence, is corresponding to the level of expression of the reporter gene, such as mCherry gene, which is corresponding to the function of the tested core promoter.

Figure 2C) The transformed strains, which confer sufficient level of reporter, such as fluorescence, which is typically 40% or higher than the reporter, such as fluorescence of the strain containing the *PGK1* core promoter assembled in the same centromeric plasmid (highlighted by arrow in the figure), are selected. The centromeric plasmids are isolated from the selected strains, the plasmid DNA is purified and sequenced, and the selected core promoters are used for subsequent constructions of expression cassettes for testing in other eukaryotic organisms. In specific cases, also core promoters which do not confer 40% or higher level of reporter expression are used for constructions of expression cassettes for eukaryotic organisms. In case the core promoter is functional in the core-promoter-donor host and also in at least one other host which is different species than the core-promoter-donor host, then the core promoter is assigned as universal core promoter, UCP.

The present invention provides a universal core promoter (UCP), which is obtainable by the disclosed method. Typically the UCP comprises a DNA sequence containing: 1) the 5′-upstream region of a eukaryotic gene, starting 10 – 50 bp upstream of a TATA-box and ending 9 bp upstream of the ATG start codon, and wherein the distance between the TATA-box and the start codon is no greater than 180 bp and no smaller than 80 bp. 2) and a DNA sequence comprising random 1-20 bp which is located at the 3’-end of the DNA sequence (1). In one em-

bodiment the universal core promoter comprises 1) a DNA sequence having at least 90% sequence identity to said 5′-upstream region and 2) a DNA sequence comprising random 1-20 bp which is located at the 3'-end of the DNA sequence.

The present invention provides also an expression system for a eukaryotic host, which comprises

(a) an expression cassette comprising an UCP,

said UCP controlling the expression of a DNA sequence encoding synthetic transcription factor (sTF), and

(b) one or more expression cassettes each comprising a DNA sequence encoding a desired protein product operably linked to a synthetic promoter,

said synthetic promoter comprising UCP of (a) or another UCP, and sTF-specific binding sites upstream of the UCP.

It is possible to construct multiple synthetic promoters with different numbers of binding sites (usually 1-10, typically 1, 2, 4 or 8, separated by 0-20, typically 5 -15 random nucleotides) controlling different target genes simultaneously by one sTF. This would for instance result in a set of differently expressed genes forming a metabolic pathway.

Figure 3 illustrates an example of a scheme of an expression system utilizing the UCPs for a simultaneous regulation of expression of multiple genes in a eukaryotic organism (e.g. microorganism). The scheme depicts a hypothetical metabolic pathway, but the approach could also be used for other multi-gene expression systems (signaling, transport, or glycosylation pathways, simultaneous protein production, etc.) or their combinations.

The synthetic transcription factor (sTF) expression cassette (**A**) in Figure 3 is analogous to the one shown in Figure 1, fulfilling the same purpose. The target gene expression cassettes can be present in variable number ranging from 1 to 20. Each target gene expression cassette contains a synthetic promoter, which can have either classical (mono-directional) architecture (**B**), or a bidirectional design (**C**). The synthetic promoter (mono- or bidirectional) consists of multiple sTF-specific binding sites (usually 1-10, typically 1, 2, 4 or 8), and a UCP. The target genes (Gene A, B, C, D) encode proteins of interest which can form a metabolic

pathway or its part, or encode any combination of proteins depending on the application.

The function of the expression system illustrated in Figure 3 is analogous to the one presented in Figure 1, transcription activity of the UCP1 is "modulated" by sTF bound to the sUASs of the target genes. The occupancy of each sUAS in combination with UCPs leads to specific expression levels of each individual gene, resulting in specific levels of the target proteins. The expression cassettes can be introduced to a eukaryotic host (typically integrated into a genome) as individual DNA molecules or as larger DNA molecules where the individual expression cassettes are fused together. In specific applications, where the target genes are native (homologous) genes of a host organism – the synthetic promoters can also be inserted immediately upstream of each target gene coding region in the genome of the host organism. The UCPs used in the expression system can be different, or some or all of the UCPs can also be identical.

The present invention provides a eukaryotic host comprising the disclosed expression system. These hosts include all eukaryotic organisms, in particular fungal microorganisms, including filamentous fungi and yeasts, plant hosts, including flowering plans and algae, and animal hosts, including mammals and insects.

The present invention provides also a method for producing a desired protein product in a eukaryotic host (e.g. microorganism host) comprising cultivating the host under suitable cultivation conditions.

The tuning of the expression system for different expression levels can be carried out in *S. cerevisiae* where a multitude of options, including choices of UCPs, sTFs, different numbers of BSs, and target genes, can be tested rapidly. The established optimal set of differently expressed genes can be directly transferred into destination host, where it retains its function. The high level of expression achieved by this system can also be utilized in the protein (enzyme) production hosts. The advantage of using *S. cerevisiae* is the availability of well-established and fast methods for genetic modifications, DNA transformation, screening, analyses, cultivations, and *in silico* modelling. This will speed up the process of industrial host development and enable the use of novel hosts which have high potential for specific purposes, but very limited spectrum of tools for genetic engineering.

Figure 4 illustrates examples of the expression systems functional/transferable in diverse eukaryotic organisms. The expression systems assembled in a single DNA

molecule comprises two expression cassettes: 1) sTF expression cassette, which comprises different UCPs, exemplified here either with the 533cp or 008cp (see Table 1 and 2), the sTF version with the DNA-binding protein, exemplified here by BM3R1, and a terminator, exemplified here by the *Trichoderma reesei TEF1* terminator. And 2) the target gene expression cassette comprises a number of sTF specific binding sites, exemplified here by eight BM3R1-specific binding sites, different UCPs, exemplified here by either 114cp or 201cp (see Table 1 and 2), the reporter gene coding region, exemplified here by the mCherry (red fluorescent protein) coding region, and a terminator, exemplified here by the *S. cerevisiae ADH1* terminator. The coding region of the DNA binding protein, here BM3R1, was codon-optimized to fit the codon usage of *Aspergillus niger*. In the Example 3, the expression system version containing 201cp and 008cp is referred to as "version A", and the expression system version containing 114cp and 533cp is referred to as "version B".

Figures 5A and 5B illustrate an example of a test of different versions of the sTFs and assessment of modulation of the expression systems performance in *Saccharomyces cerevisiae*.

Figure 5A) The expression systems analogous to the one presented in Figure 4 were constructed with following modifications to the above described system: 1) different DNA-binding proteins were used as parts of the sTFs (LexA, SrpR, PhlF, TetR, BM3R1, and TarA, see Example 1); 2) different numbers of the sTF-specific binding sites were used in the synthetic promoters of the target gene expression cassettes (the version with 8 binding sites shown in the figure); 3) the individual cassettes (the sTF expression cassette and the reporter cassette) were integrated each in a single copy into the *Saccharomyces cerevisiae* genome in two separate genomic loci (exemplified here by *URA3* and *LEU2*); 4) the sTFs were expressed from *S. cerevisiae* core promoter, here exemplified by *TDH3* core promoter; 5) the core promoter used in the reporter expression cassette was here exemplified by *S. cerevisiae ENO1*cp.

Figure 5B) The strains with both expression cassettes integrated were tested for level of fluorescence. Control expression systems were tested which have eight sTF-specific binding sites and which lack the sTF expression cassettes (shown as "wo sTF" in the figure). The DNA-binding proteins, SrpR, PhlF, TetR, BM3R1, and TarA, were codon-optimized to the codon usage of *Saccharomyces cerevisiae*. In most of the cases, a clear modulation of the expression level of the target gene is

demonstrated, which reflects the number of sTF-specific binding sites (0-8 as specified in the figure) in the synthetic promoters of the target gene expression cassettes.

Figures 6A and 6B depict examples of the analysis of the expression systems in diverse fungal hosts. Quantitative analysis of the reporter gene expression determined by fluorescence flow cytometry (6A) and by fluorometry (6B). The constructed expression systems (Table 2) were integrated in a single copy into the genomes of: *Saccharomyces cerevisiae*, *Aspergillus niger*, *Trichoderma reesei*, *Yarrowia lipolytica*, *Candida krusei (Pichia kudriavzevii)*, and *Pichia pastoris* (*Komagataella pastoris*). The functionality of the system in all these organisms was confirmed by fluorescent analysis of the transformed strains. The expression systems used for each organism are identical to those presented in Figure 4. The strain identifiers in the figures (6A and 6B) mean the following: “WT” represents a background strain of each expression host to which the expression systems were not transformed; “A” represents strains with a version of the expression system (integrated in the genome in single copy) shown in Figure 4 containing 201cp and 008cp; “A*” represents strains with the expression system version A where the DNA-binding part of the sTF was codon-optimized to match codons frequent in *Saccharomyces cerevisiae*; “B” represents strains with a version of the expression system (integrated in the genome in single copy) shown in Figure 4 containing 114cp and 533cp; “B*” represents strains with the expression system version B where the DNA-binding part of the sTF was codon-optimized to match codons frequent in *Saccharomyces cerevisiae*; “A_NC” and “B_NC” represent strains with the negative-control-versions of the expression systems (A or B) (integrated in the genome in single copy) where the sTF expression cassette was absent (deleted), leaving only the target gene expression cassette (exemplified here with 8 BS + 201/114cp + mCherry + Sc-ADH1 terminator).

Figure 6A depicts the flow-cytometry analysis (DB FACSAria III instrument) of the mCherry expression in the hosts. It was performed on cells (for the unicellular fungi - *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *Pichia kudriavzevii*, *Pichia pastoris*) or spores (for the filamentous fungi - *Aspergillus niger*, *Trichoderma reesei*). The graphs show the fluorescence intensity (mCherry) normalized by the particle (cell/spore) size (FSC – forward scatter) for 10000 cells/spores from each strain. The horizontal line (inside the grey box) represents the median value, the grey box represents the interquartile range (IQ range), the bottom of grey box represents the 25% percentile value, the top of grey box represents the 75% percentile value,

and the whiskers in box plot represent values that extend from 25% / 75% percentile values to the highest and lowest values which are no greater than 1.5 times the IQ range, which, together with the IQ range, represent about 99% of all measured instances (cells/spores) in these experiments.

Figure 6B depicts an example of the analysis of the mCherry expression in the hosts. It was performed by fluorometry measurement using the Varioskan instrument (Thermo Electron Corporation), on cell/mycelium suspensions after growing 18 hours in SCD medium. The graphs show fluorescence intensity (mCherry) normalized by the optical density of cell/mycelium suspensions used for the fluorometric analysis. The columns represent average values and the error bars standard deviations from at least 3 experimental replicates.

Figures 7A, 7B, 7C and 7D depict examples of the analysis of the tunable expression levels in different hosts (*Pichia kudriavzevii*, *Aspergillus niger*, and *Trichoderma reesei*).

Figure 7A depicts an example (shown as a scheme) of the expression system with variable number of sTF-binding sites used for modulation of reporter gene expression in *Pichia kudriavzevii* and *Aspergillus niger*. The expression system assembled in a single DNA molecule comprises two expression cassettes (analogous to those in Figure 4): 1) sTF expression cassette, which comprises a UCP, exemplified here with the 008cp (see Table 1), the sTF version with the DNA-binding protein, exemplified here by BM3R1, and the activation domain, exemplified here by VP16), and a terminator, exemplified here by the *Trichoderma reesei* TEF1 terminator. And 2) the target gene expression cassettes, which comprise different number of sTF specific binding sites, exemplified here by 0, 1, 2, 4, and 8 BM3R1-specific binding sites, different UCP, exemplified here by 201cp (see Table 1), the reporter gene coding region, exemplified here by the mCherry (red fluorescent protein) coding region, and a terminator, exemplified here by the *S. cerevisiae* ADH1 terminator. The coding region of the DNA binding protein, here BM3R1, was codon-optimized to fit the codon usage of *Aspergillus niger*.

Figure 7B depicts the flow-cytometry analysis (DB FACSAria III instrument) of the mCherry expression in *Pichia kudriavzevii* and *Aspergillus niger* containing the expression systems with variable number of sTF-binding sites (0, 1, 2, 4, and 8). It was performed on cells obtained from 18 hours cultivation in SCD medium (for

Pichia kudriavzevii), or on spores obtained after 4 days of cultivation on PDA agar plates (*Aspergillus niger*). The graphs show fluorescence intensity (mCherry) normalized by particle (cell/spore) size (FSC – forward scatter) for 10000 cells from each strain. The horizontal line (inside the grey box) represents the median value, the grey box represents the interquartile range (IQ range), the bottom of grey box represents the 25% percentile value, the top of grey box represents the 75% percentile value, and the whiskers in box plot represent values that extend from 25% / 75% percentile values to the highest and lowest values which are no greater than 1.5 times the IQ range, which, together with the IQ range, represent about 99% of all measured instances (cells/spores) in these experiments.

Figure 7C depicts an example (shown as a scheme) of the expression system with variable number of sTF-binding sites used for modulation of the CBH1 protein production in *Trichoderma reesei*. The expression system assembled in a single DNA molecule comprises two expression cassettes (analogous to those in Figure 4 and 7A): 1) sTF expression cassette, which comprises a UCP, exemplified here with the 533cp (see Table 1), the sTF version with the DNA-binding protein, exemplified here by BM3R1, and the activation domain, exemplified here by VP16, and a terminator, exemplified here by the *Trichoderma reesei* TEF1 terminator. And 2) the target gene expression cassettes, which comprise different number of sTF specific binding sites, exemplified here by 0, 1, 2, 4, and 8 BM3R1-specific binding sites, different UCP, exemplified here by 201cp (see Table 1), the target gene coding region, exemplified here by the *Trichoderma reesei* CBH1 coding region (including introns occurring in the native *Trichoderma reesei* CBH1 gene), and a terminator, exemplified here by the *S. cerevisiae* ADH1 terminator. The coding region of the DNA binding protein, here BM3R1, was codon-optimized to fit the codon usage of *Aspergillus niger*.

Figure 7D depicts western blot analyses of the CBH1 protein produced by *Trichoderma reesei* to different levels with use of the expression systems with variable number of sTF-binding sites (0, 1, 2, 4, and 8). Two different culture conditions were used, 1) a medium with spent-grain extract and lactose (“SGE-lactose” in the Figure), which leads to strong upregulation of the native CBH1 gene expression (in the background strain – WT), and 2) SCD (“SCD” in the Figure), which has a strong inhibitory effect on the native CBH1 gene expression (in the background strain – WT). Equivalent of 15 µl of the 3-days-culture supernatant from each culture was loaded on a gel (4-20% gradient). The gel was transferred onto a nitrocel-

lulose membrane, and the CBH1 protein was detected with specific (mouse) anti-CBH1 primary antibody (and anti-mouse-IR680-conjugated secondary antibody), and the visualization of the signal was performed on the Odyssey CLx Imaging System instrument (LI-COR Biosciences).

Figures 8A and 8B depict the scheme of the expression system (8A) and the analysis of a reporter gene expression in *Kazachstania exigua* (8B).

Figure 8A depicts an example (shown as a scheme) of the expression system used for *Kazachstania exigua*. The expression system assembled in a single DNA molecule comprises two expression cassettes (Table 3): 1) sTF expression cassette, which comprises a UCP, exemplified here with the Sc-TDH3cp (see Table 1), the sTF version with the DNA-binding protein, exemplified here by TetR, and the activation domain, exemplified here by VP16, and a terminator, exemplified here by the *Kazachstania exigua* g706 terminator. And 2) the target gene expression cassette comprises a number of sTF specific binding sites, exemplified here by eight TetR-specific binding sites, different UCP, exemplified here by Sc-ENO1cp (see Table 1), the reporter gene coding region, exemplified here by the Venus (yellow fluorescent protein) coding region, and a terminator exemplified here by the *S. cerevisiae PDC1* terminator. The coding region of the DNA binding protein, here TetR, and the coding region of the target gene, here Venus reporter, were codon-optimized to fit the codon usage of *Saccharomyces cerevisiae*.

Figure 8B depicts an example of an analysis of the Venus expression in *Kazachstania exigua* containing the expression system (described in Figure 8A, Table 3). The expression cassette was integrated into the genome of the background strain of *Kazachstania exigua* ("WT" in the Figure), replacing the native g706 coding region, to obtain the tested strain ("SES" in the Figure). The two strains (WT and SES) were cultivated in the SCD medium for 10 hours, and the SES strain for 22 to reach the stationary phase ("SES_stat" in the Figure). The transcription of Venus and *ADH1* genes were analysed by qPCR, with the *ALG9* gene being the normalization control for expression quantification. The columns represent the average values and the error bars the standard deviations from 3 experimental replicates.

Figures 9A, 9B, 9C and 9D depict the analysis of the protein production in diverse expression hosts (*Trichoderma reesei* and *Pichia pastoris*) containing the expression system.

Figure 9A depicts an example (shown as a scheme) of the expression system for the CBH1 protein production in *Trichoderma reesei*. The expression system assembled in a single DNA molecule comprises two expression cassettes (analogous to those in Figure 4 and 7C): 1) sTF expression cassette, which comprises a UCP, exemplified here with the 533cp (see Table 1), the sTF version with the DNA-binding protein, exemplified here by BM3R1, and the activation domain, exemplified here by VP16, and a terminator, exemplified here by the *Trichoderma reesei* TEF1 terminator. And 2) the target gene expression cassettes, which comprise a number of sTF specific binding sites, exemplified here by eight BM3R1-specific binding sites, different UCP, exemplified here by 114cp (see Table 1), the target gene coding region, exemplified here by *the Trichoderma reesei* CBH1 coding region (including introns occurring in the native *Trichoderma reesei* CBH1 coding region), and a terminator, exemplified here by the *Trichoderma reesei* PDC1 terminator. The coding region of the DNA binding protein, here BM3R1, was codon-optimized to fit the codon usage of *Aspergillus niger*.

Figure 9B depicts an example of an analysis of the production of CBH1 in *Trichoderma reesei* containing the expression system (described in Figure 9A). The background strain of *Trichoderma reesei* ("WT" in the Figure) was a mutant strain harboring multiple deletions of the genes encoding 8 diverse proteases. The expression cassette was integrated into the genome of the background strain to obtain the tested strain ("SES" in the Figure). The two strains (WT and SES) were cultivated in the bioreactor (fermentor) in two different conditions: 1) in a medium containing spent grain, spent-grain extract, and lactose ("SGM" in the Figure) which leads to a strong upregulation of the native CBH1 gene expression (in the background strain – WT) and other cellulolytic genes expression (in both strains), and 2) in a medium containing yeast extract and glucose ("glucose" in the Figure) which has a strong inhibitory effect on the native CBH1 gene expression (in the background strain – WT) and other cellulolytic genes expression (in both strains). The supernatants from these cultures were analyzed by the SDS-PAGE (SDS-polyacrylamide gel electrophoresis) analysis, for total protein content (Coomassie) and for the specific CBH1 content (western blot). For the total protein content analysis, equivalent of 1,5 µl of different time-points culture supernatants, and the range of purified CBH1 protein (loading control), were loaded on a gel (4-20% gra-

dient), and the gel was stained with colloidal coomassie (PageBlue Protein Staining Solution; Thermo Fisher Scientific). For the CBH1 specific analysis, equivalent of 0,075 µl of different time-points culture supernatants, and the range of purified CBH1 protein (loading control), were loaded on a gel (4-20% gradient), and the CBH1 protein (after transfer onto a nitrocellulose membrane) was detected with specific (mouse) anti-CBH1 primary antibody (and anti-mouse-IR680-conjugated secondary antibody). For both analyses, the visualization of the signal was performed on the Odyssey CLx Imaging System instrument (LI-COR Biosciences). The protein concentration (in the culture supernatants) was estimated from the range of purified CBH1 loaded on each gel, and the values are shown in the Figure.

Figure 9C depicts an example (shown as a scheme) of the expression system used for *Pichia pastoris* for production of a protein product. The expression system assembled in a single DNA molecule comprises two expression cassettes (analogous to those in Figure 4): 1) sTF expression cassette, which comprises a UCP, exemplified here with the 008cp (see Table 1), the sTF version with the DNA-binding protein, exemplified here by BM3R1, and the activation domain, exemplified here by VP16, and a terminator, exemplified here by the *Trichoderma reesei* TEF1 terminator. And 2) the target gene expression cassette comprises a number of sTF specific binding sites, exemplified here by eight BM3R1-specific binding sites, different UCP, exemplified here by 201cp (see Table 1), the target gene coding region, exemplified here by the coding region of the fusion protein comprising *S. cerevisiae* secretion signal (α-factor), KEX/spe13 protease cleavage site, carbohydrate-binding module (CBM), elastin-like protein (ELP5), and another CBM, and a terminator exemplified here by the *S. cerevisiae ADH1* terminator. The coding region of the DNA binding protein, here BM3R1, was codon-optimized to fit the codon usage of *Saccharomyces cerevisiae*.

Figure 9D depicts an analysis of the production of CBM-ELP5-CBM in *Pichia pastoris* containing the expression system (described in Figure 9C). The strain was cultivated in diverse conditions, and the supernatants from these cultures were analyzed by the western blot. Equivalent of 22.5 µl of the culture supernatants were loaded on a gel (4-20% gradient), and the CBM-ELP5-CBM protein (after transfer onto a nitrocellulose membrane) was detected with specific (mouse) anti-CBM primary antibody (and anti-mouse-IR680-conjugated secondary antibody). The visualization of the signal was performed on the Odyssey CLx Imaging System instrument (LI-COR Biosciences).

Table 1:

Selection of core promoters tested in *Saccharomyces cerevisiae* and other organisms. The shaded sequences are the 3'-flanking regions added to the core promoter sequences for screening or cloning purposes. The ATG (start codon) is underlined. Sc - *Saccharomyces cerevisiae* origin; An – *Aspergillus niger* origin; Tr – *Trichoderma reesei* origin; At - *Arabidopsis thaliana* origin; Cr - *Chlamydomonas reinhardtii* origin; Mm - *Mus musculus* origin

DNA sequences of the selected UCPs and other CPs used for constructing the expression systems	
Sc-THI4cp	ATCATGAAATTGATTTTTTGATTTTCAATTTATGAACTACCCAGATATATAAATATTGGAATAAATTGTGTATTAA GTAGTCGGGAAATATCTTTTATGTTCTCTTTCTTATCATCTAGAAATAATAAATCACAACCAAAAAAATCAACTAA CTTAATTAAAATG (SEQ ID NO: 1)
Sc-TEF1cp	CTCTTTCGATGACCTCCCATTGATATTTAAGTTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCATTTTTCT TGTTCTATTACAACTTTTTTTACTTCTTGCTCATTAGAAAGAAAGCATAGCAATCTAATCTAAGTTTTAATTAAAA TG (SEQ ID NO: 2)
Sc-TDH3cp	AGCTGAAAAAAAAGGTTGAAACCAGTTCCCTGAAATTATTCCCCTACTTGACTAATAAGTATATAAAGACGGTA GGTATTGATTGTAATTCTGTAAATCTATTTCTTAAACTTCTTAAATTCTACTTTTATAGTTAGTCTTTTTTTTAGTT TTAAAACACCAAGAACTTAGTTTCGAATAAACACACATAATTAATTAAAATG (SEQ ID NO: 3)
Sc-ENO1cp	TCTCCCCGGAAACTGTGGCCTTTTCTGGCACACATGATCTCCACGATTTCAACATATAAATAGCTTTTGATAAT GGCAATATTAATCAAATTTATTTTACTTCTTTCTTGTAACATCTCTCTTGTAATCCCTTATTCCTTCTAGCTATTTT TCATAAAAAACCAAGCAACTGCTTATCAACACACAAACACTTAATTAAAATG (SEQ ID NO: 4)
Sc-PGK1cp	AAGGGGGTGGTTTAGTTTAGTAGAACCTCGTGAAACTTACATTTACATATATATAAACTTGCATAAATTGGTCAA TGCAAGAAATACATATTTGGTCTTTTCTAATTCGTAGTTTTTCAAGTTCTTAGATGCTTTCTTTTTCTCTTTTTTAC AGATCATCAAGGAAGTAATTATCTACTTTTTACAACAAATTAATTAAAATG (SEQ ID NO: 5)
An-201205cp (201cp)	TTCTCTTTTCTTAAGAATATGTTCAAAGACTAGGATGGATAAATGGGGTATATAAAGCACCCTGACTCCCTTCCT CCAAGTTCTATCTAACCAGCCATCCTACACTCTACATATCCACACCAATCTACTACAATTATTAATTAAAATG (SEQ ID NO: 6)
An-53301cp (533cp)	CGCCCCAAGAGAGCTGAAGATGCTGAGTAGGGTTGTCCAGGCAGCACATATATAAGATGCTTCGTCCCCTCC CATCGAGTCCTTCTTTTCTCTCTCTCATCAATCACTCTACTTCCTACTCTACCTTAAACTCTTCACTACTTCATAC ATTAATTAAAATG (SEQ ID NO: 7)

An-205017cp	TATAGTACTATTGATTTAGTATTGTTGTTGGATGTGCTGGTAGGTGTGTAGTATATATAGGAGATAGTAGAGGC AGATGATGATGATGGTACTATTTTGAATCACCTCAAACGATACTATTCGCATCTTTGATAAAGATATCAAGAAAC CAGAACAATCATTACTACTCTCCATAAGGATATATATATACTTTACATCTTAATTAAAATG (SEQ ID NO: 8)
An-00850cp (008cp)	AACCCAAAGTAATAAGTCTGTAGTAATTGGTCTCGCCCTGAATTCCAAACTATAAATCAACCACTTTCCCTCCT CCCCCCCGCCCCCACTTGGTCGATTCTTCGTTTTCTCTCTACCTTCTTTCTATTCGGTTTTCTTCTTCTTTTATTT TCCCTCTCCCATCAATCAAATTCATATTTGAAAAAAATTAACATTAATTAAAATG (SEQ ID NO: 9)
An-1114556cp	GGGGCGGAAACTTGAAACTGGACGCCTTGTGAACGGCGTATGTGGTATATAAGGAACCAAGTCCCGCTGTAG TCTTCGGTTCATCAGACCCAGCACAGCACAGCAACACAACATTACAGCATAGCAAGCACTTCTCTATATTTCTA CACATCACAGCACATTTCTATACAGTTTACGTCTAATTATCTCCTGTTAATTAAAATG (SEQ ID NO: 10)
An-1147651cp (114cp)	GCCCTGCAGTGCCTGATCACCTTATCAAGTGGCCAAATATCCCACTATAAAAGGCTTGGGAACCCCTCGTTCT GTCTTACCTTCTATCATCTTACCAAATCCACTCCTCTTCCTTCATACATCAATCTTACCAATCAACTACCTCTACA ACTCCAATACACTTAATTAAAATG (SEQ ID NO: 11)
An-1178623cp	GGCTACTCGGGTTTTAAGCCGTCTTAAAAGCCGACACGAATTAGTTATAAAAGACTCTGTACTTGAGCAGGATA TTCCTTCATTCTTTTCATTTAGATTGATATCGAATTCATTCTACAAGGATCGGATACTCTTCCATCCTTTATTTTG TCTCTGTGAATCAAACTTAATTAAAATG (SEQ ID NO: 12)
An-57241cp	AGGTAATGAATATTGGTTGCTGGCGGGCTGATCTTCTCCCGACACGTCTATATAAACTGGTCACCTTCTGGCC CTTCCTTTCTATCTCTTCCTTCTCATCATCAGTCTCAAACAAGCCTCTTTCTCTCCTACCTTCACTCTCCACTTTC TCCTTTCGAAAGGGATAAAACTCTCCTCCTCATTCTCACCTATATATACCTTGTGCTTTAATTAAAATG (SEQ ID NO: 13)
An-06590cp	GATTTCTAGAAATTTCTGCCCTTTACTTGCCTTCCCTCTTTGTCAACAAATATAAAGAGACTCCAATTCCCCTTC TCTGATTTCCAACATTTTTCATTCTCCACTTCAGAACCATCTGAAGGAGCTTGGCTGTCTCGCTTCTTCTTCTTT CCTTCTTTACTAACATCCCTACCCCTCCTTAGAAAACCAAGTCTCTCCTCCTTTAATTAAAATG (SEQ ID NO: 14)
An-1141688cp	ACTTGGATGATGGAGGAGTTGATCGAGGTCAATGAGGAGAGGCTTGCAAGTATAAGAAGAGACTGCTCGACC AGCAGAATGGATCTTCTTGTTCATCAACCAAGAGTCCAAGGCTTCTTTGTCTGGTTCTATCTCTTCTCCGAACT CTCTTGCTTGACATTCTCTTAATTAAAATG (SEQ ID NO: 15)
Tr-123979cp	TAGCCAGCAGTGAAGAAGAGGGGAAGAAGATAAACCTGTAGGTTGGACAGAGTGTATAAAAGGGAGGGCTGT GCCCAACGAGGAGCGAGATTAACTTTGGATTTGGAGCAGAACAATATTGGAATCACAAGAAGAAGGATCTCTG TCTTTAATTAAAATG (SEQ ID NO: 16)
Tr-	TAGCCACATCCTTGGAGATCAGTTGCAGTCTATTCATTCAGGCTCAACATATAAAGATGGGATACTTCCAACAG ATGATAGTTGTCAAACAACCTCTTTGATCCTACACAATTTGGCCCAAGACACACAAGACGCTCACATCTCCTAC

112258cp	CTAACCAAACAAAGAAAAAAACATCCACCAACTTAATTAAAATG (SEQ ID NO: 17)
Tr-123236cp	ATCTTACAAAGTTGCTTGGCAGTAAACCGTGCAATGGACACCAGGTATAAAGTCAGTGATATCCTCCCCGAATTCAAAGTTTCATCACCAAGCTCCTCAATCAACTCTACTTGAACAATACTACAAACAACCAAACCTCATTCAACAACTTAATTAAAATG (SEQ ID NO: 18)
Tr-123989cp	TAAACGGAATGAGCTAGTAGGCAAAGTCAGCGAATGTGTATATATAAAGGTTCGAGGTCCGTGCCTCCCTCATGCTCTCCCCATCTACTCATCAACTCAGATCCTCCAGGAGACTTGTACACCATCTTTTGAGGCACAGAAACCCAATAGTCAACCGCGGACTTAATTAAAATG (SEQ ID NO: 19)
Tr-119989cp	AACAGCCTGCGAGAGCTGGAAGATGAAGAGGGCCAGAAAAAAAAGTATAAAGAAGACCTCGATTCCCGCCATCCAACAATCTTTTCCATCCTCATCAGCACACTCATCTACAACCATCACCACATTCACTCAACTCCTCTTTCTCAACTCTCCAAACACAAACATTCTTTGTTGAATACCAACCATCACCACTTAATTAAAATG (SEQ ID NO: 20)
Tr-123232cp	GGTCTGGATGAAACGTCTTGGCCAAATCGTGATCGATTGATACTCGCATCTATAAGATGGCACAGATCGACTCTTGATTCACAGACATCCGTCAGCCCTCAAGCCGTTTGCAAGTCCACAAACACAAGCACAAGCATATTAATTAAAATG (SEQ ID NO: 21)
Tr-73638cp	CCGGCACAAATCAGGAGCAACAGGCACTGCAAAATGACCTGGCAGTATATATAGACCTGACCGTATGAGTCTATTGTAGACATTCTAGCTAAGAGATCCGAGCCTAGTTCATAATACAGTAGTTGAGTTCATAGCAACTTCACTCTCTAGCTGAACAAATTATCTTTAATTAAAATG (SEQ ID NO: 22)
Tr-123818cp	GAGACGAGGCAAGCTTGATGAGGCCAAATTATCCGTCAACTGTCTTATAAAGGAGCCCATGCCAAACCCCCCCTAAAGACTCAAGAAGCCAAACCTGAACAACCCCAGCACCTGAACAGTCATACAACCCCTCCAAGCCCAAAAGACACAACAACTCCTACTAGCTGAAGCAAGAAGTTAATTAAAATG (SEQ ID NO: 23)
Tr-123979cp	CAGCAGTGAAGAAGAGGGGAAGAAGATAAACCTGTAGGTTGGACAGAGTGTATAAAAGGGAGGGCTGTGCCCAACGAGGAGCGAGATTAACTTTGGATTTGGAGCAGAACAATATTGGAATCACAAGAAGAAGGATCTCTGTCTTTAATTAAAATG (SEQ ID NO: 24)
Tr-69465cp	GAAAAATGGTGAGGAGATCTGCCTTCGAGTGCGTGTAGAAAAATGTATATAAGGATGTGTTTCACTCAACTTGTCTTAAGAATCGGTTCTCTAGCCGCGCTTTCAATTACTTCGAGACTTTCGCTTAAAATCGCCCTGCCATTTAATTAAAATG (SEQ ID NO: 25)
Tr-49976cp	TGCCCCTGGCGTTGCAAGCCGCGTACAACTGCCCTTTTACCTAGGTATAAAAGACCTGTAGTAACCAACTACTATTGCAATTCTTCTTCACGTGGGCATCTATTCGTATCTTACACAAGGGCGCTGCAACTAATTGACTTGATCTTCCATCTCGTGTCTTGCTTGTAACCATTAATTAAAATG (SEQ ID NO: 26)
Tr-123946cp	CTGTTAGGCTGTGAGTTATAAAGGTTGATGGATTGGGTCGAGGTTGTCAATGTCAGAGCATCTTACCTCTCACGCTTCAATCTTACCTACACGCTTCCTCTCAATCCTTGAACACCAATTGTTGCTCTAGCGCCTATCCTTCACTCAT

	CACTCGCCTCGTACACTAAACTCTTCATCCCGAACAGACACGGCTTAATTAAAATG (SEQ ID NO: 27)
At-CRA1cp	AAGTCATAAATAGCAATTTAAGTGAAGTGTAAATTGTACATAGTCGACTCTATATACCTGGTTCTTATCTCATTC AATTTATCCTCAACAACTTTAATAGAAAAATATCAAATAAATTCCCTATAAATAGCTTCACATAATGCAAGTGAGA AACCACAAAAAGTAAGAAATATAAGATTAATTAAAATG (SEQ ID NO: 28)
At-RPL41Dcp	ATCCCCTCTGGCAAATTCTTATCCATTTGGGTTTTATTGGGCTTTTGAAATAATAAAGCCCATTAAGTTAGTTAC TAGGGTTTTGTTGTTGTTTAAAGGAGGAATAAGAGCGTAAGCTACAAAATCTTTCTATTCATCTCCGCCGCTCC TCATCCTGTAAAGCTAAACAAATAATCAGAGGAACGAAGGAGACAGCTTCTGCTTAATTAAAATG (SEQ ID NO: 29)
At-ATTI7cp	GAATTTGTGGTTCTCGTGAAGTCGTGATAATAGTTTGTCCAAGCGATAAATATAAAATAGTATTGCACCTCAAC AAGTGTTAAGCATGCAAATCCATTTACGCATACATATTAACTCCGAGTGAAATATAAATATTAGAGAGTAGAGA CAGAGAAAAAGACAGAGACAAAGTTAATTAAAATG (SEQ ID NO: 30)
At-THI1cp	ATCGTTACTTTCCATTGATGGCTAAAAATTAAAATAATCACGATAAATATTAATAATACAAAAAACAATTAAAATA ACAAAAAAAGATCAAAAATTCTCTAACCCTTCATTCCTTATCTCTGACGTGGCCATCAATCTTCAGATTTTCTTC TTCTTCTAATTTAAATACTCAACAACCACTCTTCACTTCACCATCAGCATCACTAAACTCGAACCCTAAAGTTAA TTAAAATG (SEQ ID NO: 31)
At-MT2Bcp	GTGGACAAAGATCGTTGACACGTGGACGGTCTACAAATTCTAATTTTGCCTATAAATATCAAAGCTCCTGAATA TGTAAGTTTCATTCACTGATTATCGTTTAAGGCAAATTAAGATCATCTTCATAAATCTTCTCAGATCTCTTCCAAT TTTCTTTAATTAAAATG (SEQ ID NO: 32)
At-TCTP1cp	CCAAAATTGTAATTTACCGAGAATTGTAAATTTACCTGAAAACCCTACGCTATAGTTTCGACTATAAATACCAAA CTTAGGACCTCACTTCAGAATCCCCTCGTCGCTGCGTCTCTCTCCCGCAACCTTCGATTTTCGTTTATTCGCAT CCATCGGAGAGAGAAAACAATCAATTAATTAAAATG (SEQ ID NO: 33)
At-RPL26Acp	GAATTAACTTTTACTAGGCCAGAAGTGTAGCTAACATAGAAGAGGCCCATTATAAAACTCTTTAAAATCAAAATC TAAAACAGGCCCAGCCCATTCATAACAAAGCCCTAATATATCGAGTAAACCTAGCTCCACTCAAAACCTAACTA TATAACCTTCACACACACTCATAACCTCTTCCTCATCCCCTTAAAAAACCCTAAGAGTAGAGACTCTCTCAATCC CGTTAATTAAAATG (SEQ ID NO: 34)
At-MED37Ecp	ACCAATTTTTGACCGTCCGATGGAAACTCTAGCCTCAACCCAAAACTCTATATAAAGAAATCTTTTCCTTCGTTA TTGCTTACCAAATACAAACCCTAGCCGCCTTATTCGTCTTCTTCGTTCTCTAGTTTTTTCCTCAGTCTCTGTTCT TAGATCCCTTGTAGTTTCCAAATCTTTTAATTAAAATG (SEQ ID NO: 35)
At-AT1G1527	ATACACTTTCAGAGCCCATTTAATAGGTTGCGTTGTTACTACGAACTCATTATAAATATGAACCGTAGCCCCAAT CAGAGAGATTCGATACCGTCTGCAACTCTCAGCTACTTTTTCCCCAATTTTGAGCTCAACATCGAACCCTAGCT

0cp	CAACTTAATTAAAATG (SEQ ID NO: 36)
At-FKBP12cp	TCTGCTTCTTAATTCGGTCTGGTACAGTATTATATTATCCACCTTTGAGAAAGAATAAATAATGGGCCTAAATTT CATCGAATTGGGTTTTGGATTATTGTTAGGCCCAGATAGGGTTTAGATCAAACAGCATGATAATTGATAAATAA CAAAATATATAGGCAAAAGCTACTCCGAGATTCGAAGCTGCAAAGAACGCGAAACAGTGAGAGAGACAGAGA GAATTAATTAAAATG (SEQ ID NO: 37)
At-AT4G25140 cp	CAATTGCATGATGTCTCCATTGACACGTGACTTCTCGTCTCCTTTCTTAATATATCTAACAAACACTCCTACCTC TTCCAAAATATATACACATCTTTTTGATCAATCTCTCATTCAAAATCTCATTCTCTCTAGTAAACAAGTTAATTAA AATG (SEQ ID NO: 38)
At-DRT112cp	ATGCCATGTCACGACACAGTATCTAAAATCAACCAATCACAACGCGTCTTTATAGATAACTTGTTTTTTTATGGA GTTTGCTTTTAGAGCCATCCATTGTCCTATCTCACTTTCTCTCTTTCACCACATAAAAACTCATAAACTCGATCG AACCAAAGCTAAACGAAAAACTTAAAACCCAAATCTTATCACTACTCTAAAAGATTAATTAAAATG (SEQ ID NO: 39)
At-AT1G13930	TTTTCACATTTACGTCTACAATCACAATGTATGTTATTTAGAACAATAATTATAGTGGCTTAAAAATCATTAATGA AAGTAGATAATAGTATACTTTTTCTTTTTCTTTGTGTGGCCAACATATCCATTTTCTAGTCTATATATACACATAT CCATCTCTTAACTCTTCCATCCAAAAAAAACAAAACAAAAAATTATATTCAAGAGAAATTAATTAAAATG (SEQ ID NO: 40)
At-RPL14Bcp	TATTTAGTAAAGATAGGCCCAAACCACAAAACCCTAGAATGAAGATTATATATAGTGCAAAACCTAATCGATTTT TTCCTCTGCTGTCGCTCGTCTACATTTACACTCGGAGCTTAGACCTTCCAATCTACCGTTAATTAAAATG (SEQ ID NO: 41)
At-PDF2cp	TTAAAAATGCAATTCTCTAATAGACTATCAAATATCCCGATACCTCTTTATATAGTGCCATCTTCATCCTTAGTAA TGTACACACACACACATAACACTTATTTCCAACTCTGTCTCTCTCAATTTTCTTTCTCTTTTAATTAAAATG (SEQ ID NO: 42)
Cr-eIF-5Acp	CGACGAAGGGATGTCTCCGCAAGGCAAGTATATAACGGCTAGCAACGTATGCCTTAGCATAGTAGAGCAATTA GTTGTCTATGTGCCTCGGTGCAAGCGCACACGCCGGGAATAATGCGGCATGGGGGCTTCTGTTGGCCCCATG CGAGCCCCCAGGAAGAAAAGTCGCGCGGCGCCCGTATTCTGCCCTCTTGCTGTGCCAACCTCCTAGTCGCTT CTTCGCACTTTTTAATTAAAATG (SEQ ID NO: 43)
Cr-RPS27E1cp	GGATGGTGCCAAGGGTCCGGGTCACCGAGTAGCATTGGCCCACTCTAAGATATAAGTTGAGCCGTGTTTAAC TTGTTGCAACATCAGGCCTCGCGCGCAACGTCAGGAATGGCTGCATGGGGCCACCGTACCATGGCGCGGAG GGAGGGTATTGTCGCTGAGCGCGACTCAGAGCTCCCTCTCCTTTTGCTGACCGCGGAGCCTGCCCCTCTTAA TTAAAATG (SEQ ID NO: 44)
Cr-RPS8	CTGTTCTGGACGGGTCCAATGAGCCCGCTCTAAATAAACGTCTAGGAGAAGCAGTTAACCTAAGGGAAGGTG

cp	GCAGCAGGGGAGAGAGGGAGAGAGGGAGCGGGTCCATTGCTCCAGGGCGAGCCGGAATGGGGCCGGCGC TGCCTGGGCTTTGTCCGGCTGCAGACACACGGCTCTTTCTCTTTCCATTCCTTAGGGGCTGGGAACGGTTAAT TAAAATG (SEQ ID NO: 45)
Cr-RPS3A cp	CCGAACTTGCTCTCGGTGTCATATTGCACCATCCCATCTTGTATAACCGATATAACATAGCTTCGAGTGTGCCG ATAAATTATTGTGAGGGCGTCGGGGGGGCGAGCTGAGGGAAATGGAGGGGGCACTCATCTCGGCCGCCCCTC CCATCGCGACCTCGGCGCTCAAGCGGGGGTCCCGCACTCGCTTCGGTCTCTTTTGGTCAGCAGCCGTTTGTT GACTACCGTTAATTAAAATG (SEQ ID NO: 46)
Cr-RPL17 cp	AAGCGGACAGAAATTTAGTTCAGGAAGAATTGTCAGATTTGCTACTGGCATATAATTTTTTCTGCAGGGTCTGG CGTGGAAGAATGCCAAATGGCGCGGAGCTGGCTGCATGGGGCGCCACCTCCCAGCAAGGGCCACCACTGCA ACCTGCTCTTTCTCTTTCGTCGCGCCTTGCACGTAGCGTTAATTAATTAAAATG (SEQ ID NO: 47)
Cr-RPL19 cp	AGAACGCGCCTAGTACTCATGCCACGAGAGTTTCATCATTCCAGCATGCATAATAAATTTGTCACTCAGGCAG AGCATTTGCGGGGCGCGCAATGTTTAGCGGGGCCCAAAGTCGCCATCGCGGTCGCGCCCCCATGCAGCGTT CCACCCTGGCTTTCAGGCGCGGGCGCACCTGGACTATCCCTTTCTTTGCGTCGTCCGCTTGCAAACAGATTAA TTAAAATG (SEQ ID NO: 48)
Cr-RPS24 cp	AGCGAGCGAACTAGTCTCGCGCCCCGGCCCCCCCGTCGCCAACAGACCGCTATAACCAAGTAATTTTGTGTG GCTTTATTTGTATTGCTAAAAACCCCCGAGCGGGGTGAGCCCAAGACAAGAAACGAAGGCGGCCCCCTCCTG GAGCAATGGGCGTCTGAGAGACGGGGCAAGACCAGGGAGAGTCCCAGCCTCCTCCCTCTTTCTCTTGGCAC ACTTAATTAAAATG (SEQ ID NO: 49)
Cr-RPS15 cp	TCTAGGAGGGGTTTTCCCTTGTTTAGCCCTATATAACGTAAAGCTCACACTTTGAATGAGCAACATAAATTATAT TTAGTGCGAAAGCCGGCTATGAAAATGGACATGGGGATCGCGATGGGCGCCCCCGCGCCCTGGCGGCGGTG TGCACAGGAGCGAGGCCCTCGCCTGCTCCTTCCTCTTTCTCTCGCCGTCAGGCTCGTAGTTTCAAAAGTTAAT TAAAATG (SEQ ID NO: 50)
Cr-ATPC cp	CGTGCAAGCTACTCCCAGGCTCCTGCATTCTATAAGCGTAATTTTATGCCGGGTATGCTTGTGATTTGACGAA GATCTACTCGACGGCGTTCTGGTGGGCAAAATCGGAGGCAAACCCAATTGGCCCCCCTGGAGTGATAAGTCC TGGGTGCCAAGTGCGCAAGTGAAGCCTTGAACTGCGCCTTTCCTTGCACCTTGTTCGCCGCTCTTTCTATTAA TTAAAATG (SEQ ID NO: 51)
Cr-RPS9 cp	GCAGAAGGGTGCCAAAGGGAGGCTCTAAGCAGTCCAGGCGGGCACAAACATATAAAGCTGAAGCTAGTGGTA TACCTAAATTAATTTTGGGGCGCTCCACTGAAAAATGGACTGCCTGCATGGGCCCTGTACGGCTTCGCCGAGC GAGCCCGGTGCAAGGGCCGCGACCGTGCATAAGTCTCTCTCTTTCAAGTTGGCAGAGGGAGCGCCAGTTAAT TAAAATG (SEQ ID NO: 52)
Cr-RPL10a	CTGGGCCTTACTTTCATCATAGGGAAAGCATAAATCATAACAGTGTAGTTTATATTATGCATGATGTCTTCCGC AGAAGAGGCACCGTGATGCCCACCGCCCCCATGCATCAATTGTGAGGGTCAAGAGCGCCCGCGGACCCCTG

cp	GACATTCCTTTTCCTTGGGTGATTAATTAAAATG (SEQ ID NO: 53)
Cr-HSP70A cp	CGCGCGGCGTCCAGAAGGCGCCATACGGCCCGCTGGCGGCACCCATCCGGTATAAAAGCCCGCGACCCCG AACGGTGACCTCCACTTTCAGCGACAAACGAGCACTTATACATACGCGACTATTCTGCCGCTATACATAACCA CTCAACTCGCTTAAGAGTTAATTAAAATG (SEQ ID NO: 54)
Cr-VIP1 cp	CAGCTGCGGGCAGGCCGGCTGCATGGCTTGCTTCTGGAGAGGGCCAATTGTAATTACCGCTTTCCTGCCTTT CCAAGAGCCCCCTACAACCTACGCACTTTAAAATCACATACAGCCTGTGGCCCAACTTCCTTGTTAGTCCTTAA TTAAAATG (SEQ ID NO: 55)
Cr-NPC1 cp	CCTGGTAAATATTCTGCGCCGCTTTCGTAACAGGTGCAGGCGCAGGTAGCTATACAAATATGGTCGCGGCTG CAAATGCGGGGGGAGGAGGAGTACTTGCATGGGTCGCCCGCGATCGGCACTCCCGCTCGGTCCCCGACTGA ACACCCGCGCGAGCCCCGTGGTTCCCCCCTTTTCAACATTAGCCAACTCGACCCCAGTCGACTTTTCTCGTCG TTTTAATTAAAATG (SEQ ID NO: 56)
Cr-AAA1 cp	TCGCCATTGGGGGCCGCATGGGGCCCTGGAGCACCGAAAGTGCAGAGCTCTATAGAGCGCCACTCGTTCTT CTTGCCTCTTCACTAGCCCGCCCACAATAATTGGGTTGCAGTCAAGTGAGTGCGTAGCTTCACAGCAGGGTCT ATAGGGCCCCGACACTTGCACCAAACCTGCCGATCACAAGCATTAATTAAAATG (SEQ ID NO: 57)
Mm-Eef2cp	CCGGACGAGCACCCGGCGCCGTCACGTGACGCACCCAACCGGCGTCGACCTATAAAAGGCCGGGCGTTGAC GTCAGCGGTCTCTTCCGCCGCAGCCGCCGCCATCGTCGGCGCGCTTCCCTGTTCACCTCTGACTCTGAGAAT CCGTCGCCATTAATTAAAATG (SEQ ID NO: 58)
Mm-Col1a1cp	TCCCCCTCTCCGAGAGGCAGGGTTCCTCCCAGCTCTCCATCAAGATGGTATAAAAGGGGCCCAGGCCAGTCG TCGGAGCAGACGGGAGTTTCTCCTCGGGACGGAGCAGGAGGCACGCGGAGTGAGGCCACGCATGAGCCGA AGCTAACCCCCCACCCCAGCCGCAAAGAGTCTACATGTCTAGTTAATTAAAATG (SEQ ID NO: 59)
Mm-Rpl4cp	CTCTAATTTGATTTTGATAAGGGGCAGGATGCGGAAGACGAGTGGAAGGATATATAGAGTACAAGTGACAAGT CTTTCCTTTTCCTGTGGGAGCAGCCGGGTAGAGAGGAGCGTGGCCTTCTCCTTTAATTAAAATG (SEQ ID NO: 60)
Mm-Fabp9cp	TAGAGAGGCTCCCTGAATTAAACTTTGCAGGTTAGTTTCTGTTGGTGGTATATAAAATGAGTCAACCGCCGGT GCCTGGAATCTCAGATTCCTGGCAGCCCATTGGTTGCTTCAGGAATGCCACGTGACAAAGATGTTCTTTAAAA GAAGGGGCTGGCGGCACAGCGATGCCCACACTTCATGGTTTTTTAATTAAAATG (SEQ ID NO: 61)
Mm-Vimcp	GCTCGGCGGCTAGGATGGCAGTGGGAGGGGACCCTCTTTCCTAACAGTGTTATAAAAGCAGCGCCCTTGGC GTTGTCCAGTCCTCTGCCACTCTTGCTCCGGGACCCCAGAGACCCCAGCGCTCCTACGATTCACAGCCACCG CGCCCTCATTCCCTTGTTGCAGTTTTTCCAGCCGCAGCAAGCCAGCCCACCTTAATTAAAATG (SEQ ID NO: 62)

Mm-Rplp1cp	GTCGGGCATCGATTGGTCGGCGCGGTCCCATAAGAAGCTGCGCGCAGGCGTATATGATCTTTTCCTCAGCTG CCGCCAAGGTGCTCGGTCCTTCCGAGGAAGCTAAGGCCGCGTTGGGGTGAGGCCCTCACTTCATCCGGCGA CTAGCACCGTGCCGGCATTAATTAAAATG (SEQ ID NO: 63)
Mm-Atp5bcp	ATTGGCACCAGTTTAGACCAATAGCTGATAAGCTCCGAGTTTTTTTACCCTATAGAAGCGTTAGTGGTGATGAC GAACAGCAAAATCACCCAATTACTGTGCCTACGGCGGAGGTTGCCCCGCCCCAGCTGCAGGACCGGCGGAG AGGACCGCTTCGGCGCTCAGTCTCCACCCGTTAATTAAAATG (SEQ ID NO: 64)
Mm-Ppt1cp	TGTAAAATGAGAGCAGTGCATAAGATCAATTAAAAGATGGAAAGCCCTTATATAGTAGAGTCTGGTGGGTGTTA AGCAATGAATAAACGTCTCTGTCAGGATGCAGAGCCCGGCAGGGGGCGTGGCCACCGGAATTACTTTGGTCC ACAGTCCCCGCGGTCATGTGTTAATTAAAATG (SEQ ID NO: 65)
Mm-Lgals1cp	GACTGGTCACCTCTGCTCCAAAATTGACTTTATAATCCATAGACTCCACTTCTGGTGGCCCCCATCCTTGTCCT GACATGCAATTGGCTGAACTCCCGGGGAGGGGCGGGACTCACCGGGTCTGATCCAGTTAAAAAGGTCGGAG CGGGCCTGGGGCCCGTCTCTCGGGTGGAGTCTTCTGACTGCTGGTGGAGCAGGTCTCAGGAATCTCTTCGC TTCATTAATTAAAATG (SEQ ID NO: 66)
Mm-Fth1cp	GGCCAGCGCTCGCCTGACGCAGGATCCCGCTATAAGTGCGGCCCGCTGTCCCCTCCTGCGCCAGACGTTCT CGCCCAGAGTCGCCGCGGTTTCCTGCTTCAACAGTGCTTGAACGGAACCCGGTGCTCGACCCCTCCGACCC CCGCCGGCCGCTTCGAGCCTGAGCCCTTTGCAACTTCGTCGTTCCGCCGCTCCAGCGTCGCCACCGCGCCT CGCCCCTTAATTAAAATG (SEQ ID NO: 67)

Table 2:

DNA sequences of some of the tested inter-species transferable expression systems. The functional DNA parts are indicated: 8×BM3R1 binding site (white text, black highlight); core promoters (underlined); mCherry coding region (white text, grey highlight); terminators (*italics, grey highlight*); BM3R1-sTF (grey highlight).

DNA sequences of the tested inter-species transferable expression systems	
8BS(BM3R1)-201cp-mCherry + 008cp-BM3R1_sTF	GCATTTGCTCGGCTAGTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGG ACTCTAGATAAGCACGGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTA GTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGGACTCTAGATAAGCAC GGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTAGTTCTCCCCGGAAAC TGTGGCCATATGTTCAAAGACTAGGATGGATAAATGGGGTATATAAAGCACCCTGACTCCCTTCCTCCAAGTT CTATCTAACCAGCCATCCTACACTCTACATATCCACACCAATCTACTACAATTATTAATTAAA[redacted]TAATGAGGATCC*GAATTT CTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGG TTTTAAAACGAAAATTCTTATTCTTGAGTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGG TCGCTCTTATTGACCACACCTCTACCGG*CCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGTCGAGGCT AACAACCCAAAGTAATAAGTCTGTAGTAATTGGTCTCGCCCTGAATTCCAAACTATAAATCAACCACTTTCCCT CCTCCCCCCCGCCCCCACTTGGTCGATTCTTCGTTTTCTCTCTACCTTCTTTCTATTCGGTTTTCTTCTTCTTTT ATTTTCCCTCTCCCATCAATCAAATTCATATTTGAAAAAAATTAACCATTAATTAACAATGGAGTCCACACCCAC GAAACAAAAAGCTATTTTTTCTGCCTCGCTCCTTCTGTTCGCCGAACGCGGGTTTGACGCCACTACGATGCCG ATGATCGCTGAAAATGCTAAGGTCGGCGCAGGAACGATTTACCGATACTTTAAGAATAAGGAGAGTCTGGTCA ACGAGCTGTTCCAGCAGCACGTTAATGAATTTTTGCAATGTATCGAGAGTGGCTTGGCGAACGAAAGGGACG GTTATCGCGATGGGTTCCATCATATCTTCGAGGGAATGGTCACATTCACAAAGAACCATCCGCGCGCCTTGGG ATTTATCAAGACACATTCCCAAGGTACATTCCTAACCGAAGAGTCACGCCTTGCATACCAAAAACTTGTTGAGT TCGTCTGCACCTTCTTTCGAGAGGGACAGAAACAGGGCGTAATTCGAAACTTGCCCGAGAATGCCCTGATCG CCATCCTATTCGGATCGTTTATGGAGGTCTATGAGATGATCGAAAACGATTATCTCTCTCTAACGGATGAGTTG CTTACGGGGGTAGAGGAATCGCTCTGGGCTGCTCTCTCCCGACAATCGGCTAGCCCTCCCAAGAAGAAGCGC AAGGTCAGCACGGCCCCCCCCACGGACGTCTCCCTCGGCGACGAGCTCCACCTGGACGGCGAGGACGTCG CCATGGCCCACGCCGACGCCCTCGACGACTTCGACCTCGACATGCTGGGCGACGGCGACAGCCCCGGCCC CGGCTTTACCCCCCACGACTCCGCCCCCTACGGCGCCCTGGACATGGCCGACTTCGAGTTTGAGCAGATGTT CACCGACGCCCTGGGCATTGACGAGTACGGCGGCTGAGGCCGGCC*GCGATACCCATCATCAACACCTGATG TTCTGGGGTCCCTCGTGAGGTTTCTCCAGGTGGGCACCACCATGCGCTCACTTCTACGACGAAACGATCAAT GTTGCTATGCATGAGCACTCGACTATGAATCGAGGCACGTTAATTGAGAGGCTGGGAATAAGGGTTCCATCAG AACTTCTCTGGGAATGCAAAACAAAAGGGAACAAAAAAACTAGATAGAAGTGAATTCATGACTTCGACAACCAA ATCATCTTGTCTCCGTCTGCATACGTGAAGCTTGTGACGATTATTCTCGCGATGCCACGACAAAGGTTGTGCG ACCGTATCTTGTCCACTGTCGTCCAGTCTGCCTATTCCCCCTCCAGTGCTGCCATGTGTCGTACCTTGAGGTA GGTAGTCTACCTAGGCCAGGGAGCTGTTAGTGCCCGGCTACTGGGTAATTTGTAGCGCTGGAGCGATTCGGT CACAGGCGTCAAGAGTGCTGTAGCAATGTCCGACGCCATTGATCCTGATATCAAATACCACCTGGGCAGGTCT GGGTATGTGAGGTCTTGTCGGATGTGTCGAGTTCTTCTCCAACGTAGTGTTCATTCGCGCTCAT*GCCC (SEQ ID NO: 68)
8BS(BM3R1)-114cp-mCherry + 533cp-BM3R1_sTF	GCATTTGCTCGGCTAGTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGG ACTCTAGATAAGCACGGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTA GTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGGACTCTAGATAAGCAC GGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTAGTTCTCCCCGGAAAC TGTGGCCATATGCCCTGCAGTGCCTGATCACCTTATCAAGTGGCCAAATATCCCACTATAAAAGGCTTGGGAA CCCCTCGTTCTGTCTTACCTTCTATCATCTTACCAAATCCACTCCTCTTCCTTCATACATCAATCTTACCAATCA ACTACCTCTACAACTCCAATACACTTAATTAAA[redacted]

F	TAATGAGGATCC*GAATTTCTTATGATTTATGATTTTTATTATTAAA TAAGTTATAAAAAAAATAAGTGTATACAAATTTTAAAGTGACTCTTAGGTTTTAAAACGAAAATTCTTATTCTTGA GTAACTCTTTCCTGTAGGTCAGGTTGCTTTCTCAGGTATAGCATGAGGTCGCTCTTATTGACCACACCTCTACC GG*CCAGCTTTTGTTCCCTTTAGTGAGGGTTAATTGCGCGTCGAGGCTAGCCGCCCCAAGAGAGCTGAAGATG CTGAGTAGGGTTGTCCAGGCAGCACATATATAAGATGCTTCGTCCCCTCCCATCGAGTCCTTCTTTTCTCTCTC TCATCAATCACTCTACTTCCTACTCTACCTTAAACTCTTCACTACTTCATACGATTAACAATGGAGTCCACACCC ACGAAACAAAAAGCTATTTTTTCTGCCTCGCTCCTTCTGTTCGCCGAACGCGGGTTTGACGCCACTACGATGC CGATGATCGCTGAAAATGCTAAGGTCGGCGCAGGAACGATTTACCGATACTTTAAGAATAAGGAGAGTCTGGT CAACGAGCTGTTCCAGCAGCACGTTAATGAATTTTTGCAATGTATCGAGAGTGGCTTGGCGAACGAAAGGGAC GGTTATCGCGATGGGTTCCATCATATCTTCGAGGGAATGGTCACATTCACAAAGAACCATCCGCGCGCCTTGG GATTTATCAAGACACATTCCCAAGGTACATTCCTAACCGAAGAGTCACGCCTTGCATACCAAAAACTTGTTGAG TTCGTCTGCACCTTCTTTCGAGAGGGACAGAAACAGGGCGTAATTCGAAACTTGCCCGAGAATGCCCTGATCG CCATCCTATTCGGATCGTTTATGGAGGTCTATGAGATGATCGAAAACGATTATCTCTCTCTAACGGATGAGTTG CTTACGGGGGTAGAGGAATCGCTCTGGGCTGCTCTCTCCCGACAATCGGCTAGCCCTCCCAAGAAGAAGCGC AAGGTCAGCACGGCCCCCCCCACGGACGTCTCCCTCGGCGACGAGCTCCACCTGGACGGCGAGGACGTCG CCATGGCCCACGCCGACGCCCTCGACGACTTCGACCTCGACATGCTGGGCGACGGCGACAGCCCCGGCCC CGGCTTTACCCCCCACGACTCCGCCCCCTACGGCGCCCTGGACATGGCCGACTTCGAGTTTGAGCAGATGTT CACCGACGCCCTGGGCATTGACGAGTACGGCGGCTGAGGCCGGCC*GCGATACCCATCATCAACACCTGATG TTCTGGGGTCCCTCGTGAGGTTTCTCCAGGTGGGCACCACCATGCGCTCACTTCTACGACGAAACGATCAAT GTTGCTATGCATGAGCACTCGACTATGAATCGAGGCACGTTAATTGAGAGGCTGGGAATAAGGGTTCCATCAG AACTTCTCTGGGAATGCAAAACAAAAGGGAACAAAAAAACTAGATAGAAGTGAATTCATGACTTCGACAACCAA ATCATCTTGTCTCCGTCTGCATACGTGAAGCTTGTGACGATTATTCTCGCGATGCCACGACAAAGGTTGTGCG ACCGTATCTTGTCCACTGTCGTCCAGTCTGCCTATTCCCCCTCCAGTGCTGCCATGTGTCGTACCTTGAGGTA GGTAGTCTACCTAGGCCAGGGAGCTGTTAGTGCCCGGCTACTGGGTAATTTGTAGCGCTGGAGCGATTCGGT CACAGGCGTCAAGAGTGCTGTAGCAATGTCCGACGCCATTGATCCTGATATCAAATACCACCTGGGCAGGTCT GGGTATGTGAGGTCTTGTCGGATGTGTCGAGTTCTTCTCCAACGTAGTGTTCATTCGCGCTCAT*GCCC (SEQ ID NO: 69)

Table 3:

DNA sequence of the expression system tested in *Kazachstania exigua* and *Saccharomyces cerevisiae*. The functional DNA parts are indicated: 8×TetR binding site (white text, black highlight); core promoters (underlined); Venus coding region (white text, grey highlight); terminators (*italics, grey highlight*); TetR-sTF (grey highlight).

DNA sequence of the expression system tested in *Kazachstania exigua* and *Saccharomyces cerevisiae*	
8BS(TetR)-ENO1cp-Venus + TDH3cp-TetR_sTF	GCTAGCTCTCTATCACTGATAGGGAGTATTGACAAGCTTTCTCTATCACTGATAGGAGTGGCTTATCTAGATCT CTATCACTGATAGGGAGTTCACATCCTAGGTCTCTATCACTGATAGGGAGTACTAGCTCTCTATCACTGATAGG GAGTATTGACAAGCTTTCTCTATCACTGATAGGAGTGGCTTATCTAGATCTCTATCACTGATAGGGAGTTCACA TCCTAGGTCTCTATCACTGATAGGGAGTACTAGTTCTCCCCGGAAACTGTGGCCTTTTCTGGCACACATGATC TCCACGATTTCAACATATAAATAGCTTTTGATAATGGCAATATTAATCAAATTTATTTTACTTCTTTCTTGTAACAT CTCTCTTGTAATCCCTTATTCCTTCTAGCTATTTTTCATAAAAAACCAAGCAACTGCTTATCAACACACAAACAC TTAATTAAA

	TAAGTCGA*CGCTAATTAACATAAAACTCATGATTCAACGTTTGTGTATTTTTTTACTTTTGAAGGTT ATAGATGTTTAGGTAAATAATTGGCATAGATATAGTTTTAGTATAATAAATTTCTGATTTGGTTTAAAATATCAAC TATTTTTTTTCACATATGTTCTTGTAATTACTTTTCTGTCCTGTCTTCCAGGTTAAAGATTAGCTTCTAATATTTTA GGTGGTTTATTATTTAATTTTATGCTGATTAATTTATTTACTTTCGTATTCGGTTTTGTACCTTTAGCTATGATCTT AGCTAATTGAAG*GGGCCTCGAGGCTAGCAGCTGAAAAAAAAAGGTTGAAACCAGTTCCCTGAAATTATTCCCCT ACTTGACTAATAAGTATATAAAGACGGTAGGTATTGATTGTAATTCTGTAAATCTATTTCTTAAACTTCTTAAATT CTACTTTTATAGTTAGTCTTTTTTTTAGTTTTAAAACACCAAGAACTTAGTTTCGAATAAACACACATAATTAATTA AATCTAGACAATGAGTAGATTAGACAAATCAAAAGTGATAAATTCTGCATTAGAATTGTTGAATGAAGTAGGCAT TGAAGGTTTGACTACCCGTAAGTTAGCTCAGAAACTAGGTGTTGAACAACCTACATTATACTGGCACGTTAAAA ATAAAAGGGCATTGTTGGATGCGCTTGCCATTGAGATGTTGGATAGGCATCATACCCACTTTTGCCCATTAGAA GGAGAGTCTTGGCAGGACTTTTTGAGGAATAATGCCAAGTCATTTAGATGTGCATTGTTGTCTCATAGAGATG GGGCCAAGGTTCATCTAGGTACCCGTCCTACGGAAAAACAATATGAGACGTTGGAAAATCAGTTAGCGTTCTT ATGCCAACAAGGCTTTAGCTTGGAAAATGCTTTATATGCTCTATCAGCTGTCGGTCATTTTACATTGGGATGCG TTTTAGAAGACCAGGAGCACCAGGTGGCAAAGGAAGAAAGAGAAACACCAACAACTGATTCAATGCCACCCCT ACTGAGACAAGCTATCGAATTATTTGATCATCAAGGTGCGGAACCTGCCTTCTTGTTTGGCCTAGAATTGATCA TTTGTGGTTTAGAAAAGCAGTTAAAATGTGAGAGTGGCTCAGAATTCCCTCCCAAGAAGAAGCGCAAGGTCAG CACGGCCCCCCCCACGGACGTCTCCCTCGGCGACGAGCTCCACCTGGACGGCGAGGACGTCGCCATGGCC CACGCCGACGCCCTCGACGACTTCGACCTCGACATGCTGGGCGACGGCGACAGCCCCGGCCCCGGCTTTAC CCCCCACGACTCCGCCCCCTACGGCGCCCTGGACATGGCCGACTTCGAGTTTGAGCAGATGTTCACCGACG CCCTGGGCATTGACGAGTACGGCGGCTGA (SEQ ID NO: 70)

Table 4:

DNA sequences of the expression systems tested in *Nicotiana benthamiana*. The functional DNA parts are indicated: 8×sTF binding site (white text, black highlight); core promoters (underlined); mCherry coding region (white text, grey highlight); terminators (*italics, grey highlight*); sTFs (grey highlight).

DNA sequence of the expression system tested in *Nicotiana benthamiana*	
8BS(BM3R1)-At-ATTI7cp-mCherry+At-RPL41Dcp–BM3R1_sTF	CATTTGCTCGGCTAGTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGGA CTCTAGATAAGCACGGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTAG TCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGGACTCTAGATAAGCACG GAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTAGTTCTCCCCGGAAACT GGAATTTGTGGTTCTCGTGAAGTCGTGATAATAGTTTGTCCAAGCGATAAATATAAAATAGTATTGCACCTCAA CAAGTGTTAAGCATGCAAATCCATTTACGCATACATATTAACTCCGAGTGAAATATAAATATTAGAGAGTAGAG ACAGAGAAAAAGACAGAGACAAAGTTAATTAAA[redacted]TAA*TAGCTATATATCTTTCTTACATCATTATTGTAATCTGTTCTCCTTCTGTG TATTCGTTTCAATGTTGCAGCAATGAACTTTTGGATAAAAGTCAAATTTGTTGTTTCCTTAATTCGAAAGACGAT TGAGACTTGAAATCATAACACTAAGCTTCATTGAATCAAGATTCAATAGTATTCATCAATTCATAATATAATAGTG TACTAAACTCGAGCTTGCATATTCTGAGTTAATTGAAATACCTCACTGTAATACCTAGAACGAACTTACCTTACG AGCAAATCAAGCATGTATTTACTCTCGGATGTATAATTCACCTTATCAACCTTCACAACAGTCATCTTCACTCTT TGTTCATCCCCATACGATTCCTCTTTGATCTTCAGCTTC*ATTTAAATGCGATCCCCTCTGGCAAATTCTTATCCA TTTGGGTTTTATTGGGCTTTTGAAATAATAAAGCCCATTAAGTTAGTTACTAGGGTTTTGTTGTTGTTTAAAGGA GGAATAAGAGCGTAAGCTACAAAATCTTTCTATTCATCTCCGCCGCTCCTCATCCTGTAAAGCTAAACAAATAA TCAGAGGAACGAAGGAGACAGCTTCTGCTTAATTAAAATGGAGAGTACACCAACCAAACAGAAAGCTATTTTTA GCGCAAGCCTGTTATTATTTGCTGAGCGTGGCTTTGACGCTACGACGATGCCCATGATAGCCGAAAATGCTAA AGTTGGGGCTGGAACCATATACCGATACTTCAAAAACAAAGAAAGTCTGGTGAATGAACTGTTTCAACAACAC GTAAACGAGTTCTTGCAGTGCATCGAGTCTGGACTCGCTAACGAGCGTGACGGCTATAGAGATGGATTTCATC ATATATTTGAGGGCATGGTCACCTTCACAAAAAACCACCCAAGGGCTCTGGGTTTCATAAAAACGCACAGTCA AGGCACATTCCTTACGGAGGAGAGCAGATTAGCATATCAGAAATTAGTGGAGTTCGTATGTACTTTCTTTAGAG AAGGACAGAAGCAAGGAGTTATTCGAAACCTGCCGGAAAACGCCTTAATTGCCATCCTGTTTGGGTCTTTCAT GGAAGTTTACGAAATGATAGAGAATGATTACCTCTCCCTTACCGACGAATTGTTGACTGGCGTCGAAGAATCAT

	TGTGGGCAGCATTGTCTAGGCAATCAGAATTCCCACCTAAGAAGAAAAGGAAAGTATCCGCCTCCGGTTCTGG GCGTGCCGACGCTCTGGACGATTTCGACCTCGATATGTTGGGGTCAGACGCATTGGACGACTTTGATTTGGA CATGTTAGGTAGTGATGCTTTAGACGATTTCGACTTGGACATGTTAGGCTCCGATGCATTGGACGATTTTGACT TAGATATGTTGATTAATAGTAGGTAA*TGAGTCGACTCTTTAATCAAAATGTAATATGAATAAAAGTTGATGTGGG CTCATCTATTGAGCTCATGTCTCTCTTATTACTACTCTCTAGTATGGTGTGATGTAATGGGTTATGACCCTTCTT TCCCTTCCCTATAAAACTAAAGAAACTTGCAAGATAATTGAAAAGATGGTTTCTTTTTATTATCAATCGCATCAA AATGGGATTTTGTATCAAATGCATACATTATCTCTTGCTTTTATACCCTAAACCCATACCGGGTGATGAACAATC TTCTGTTGCTCATTCCTTTTGATGATCCATCAAATTACGTATTAGAAAAAGAAAAAAAAAGTATCAGACGTTGAAA CCTTCTGCTCGGAGACAAATTTTATGAGCCTCGATCC*ATTTAAATCAAGCCCGGG (SEQ ID NO: 71)
8BS(TetR)-At-ATTI7cp-mCherry + At-RPL41Dcp–TetR_sTF	CATTTGCTCGGCTAGCTCTCTATCACTGATAGGGAGTATTGACAAGCTTTCTCTATCACTGATAGGAGTGGCTT ATCTAGATCTCTATCACTGATAGGGAGTTCACATCCTAGGTCTCTATCACTGATAGGGAGTACTAGCTCTCTAT CACTGATAGGGAGTATTGACAAGCTTTCTCTATCACTGATAGGAGTGGCTTATCTAGATCTCTATCACTGATAG GGAGTTCACATCCTAGGTCTCTATCACTGATAGGGAGTACTAGTTCTCCCCGGAAACTGGAATTTGTGGTTCT CGTGAAGTCGTGATAATAGTTTGTCCAAGCGATAAATATAAAATAGTATTGCACCTCAACAAGTGTTAAGCATG CAAATCCATTTACGCATACATATTAACTCCGAGTGAAATATAAATATTAGAGAGTAGAGACAGAGAAAAAGACA GAGACAAAGTTAATTAAA[redacted] TAA*TAGCTATATATCTTTCTTACATCATTATTGTAATCTGTTCTCCTTCTGTGTATTCGTTTCAATGTT GCAGCAATGAACTTTTGGATAAAAGTCAAATTTGTTGTTTCCTTAATTCGAAAGACGATTGAGACTTGAAATCAT AACACTAAGCTTCATTGAATCAAGATTCAATAGTATTCATCAATTCATAATATAATAGTGTACTAAACTCGAGCTT GCATATTCTGAGTTAATTGAAATACCTCACTGTAATACCTAGAACGAACTTACCTTACGAGCAAATCAAGCATG TATTTACTCTCGGATGTATAATTCACCTTATCAACCTTCACAACAGTCATCTTCACTCTTTGTTCATCCCCATAC GATTCCTCTTTGATCTTCAGCTTC*ATTTAAATGCGATCCCCTCTGGCAAATTCTTATCCATTTGGGTTTTATTGG GCTTTTGAAATAATAAAGCCCATTAAGTTAGTTACTAGGGTTTTGTTGTTGTTTAAAGGAGGAATAAGAGCGTAA GCTACAAAATCTTTCTATTCATCTCCGCCGCTCCTCATCCTGTAAAGCTAAACAAATAATCAGAGGAACGAAGG AGACAGCTTCTGCTTAATTAAAATGTCAAGATTAGACAAAAGCAAAGTAATCAATAGTGCATTAGAACTTTTAAA CGAGGTCGGAATAGAGGGATTAACTACACGTAAACTCGCCCAGAAGCTCGGAGTTGAACAACCTACGCTGTA TTGGCATGTTAAAAATAAACGAGCATTATTGGATGCTCTGGCTATTGAGATGCTCGATAGGCACCATACCCACT TTTGCCCTCTGGAGGGGGAATCTTGGCAAGACTTCTTGCGTAACAACGCCAAGTCATTCAGATGTGCCTTGCT GAGTCACCGTGACGGCGCTAAAGTCCATCTCGGAACCCGACCGACCGAGAAGCAATACGAGACCTTAGAAAA CCAATTAGCCTTTCTTTGCCAGCAAGGGTTTTCATTAGAGAATGCTCTCTACGCCCTTTCCGCTGTTGGGCATT TCACCCTGGGTTGCGTCTTGGAGGATCAGGAACATCAAGTAGCAAAGGAGGAACGAGAGACACCTACTACGG ATTCTATGCCGCCCCTCCTCAGGCAGGCAATTGAACTGTTCGATCATCAGGGAGCTGAACCTGCTTTTCTGTT TGGCCTGGAATTGATAATATGCGGACTGGAGAAACAGTTAAAGTGCGAGAGCGGTAGCGAATTCCCACCTAAA AAAAAGAGAAAAGTTTCCACTGCCCCCCCAACGGACGTCTCTCTGGGCGACGAGCTGCACCTCGATGGTGAG GACGTGGCTATGGCTCATGCTGATGCCTTAGACGACTTCGACTTAGATATGCTGGGAGATGGCGACTCACCG GGACCAGGGTTTACACCTCATGATTCTGCTCCTTACGGAGCTTTAGATATGGCCGATTTTGAATTTGAGCAAAT GTTCACTGACGCCCTTGGAATAGACGAATACGGGGGCTAA*TGAGTCGACTCTTTAATCAAAATGTAATATGAAT AAAAGTTGATGTGGGCTCATCTATTGAGCTCATGTCTCTCTTATTACTACTCTCTAGTATGGTGTGATGTAATG GGTTATGACCCTTCTTTCCCTTCCCTATAAAACTAAAGAAACTTGCAAGATAATTGAAAAGATGGTTTCTTTTTA TTATCAATCGCATCAAAATGGGATTTTGTATCAAATGCATACATTATCTCTTGCTTTTATACCCTAAACCCATAC CGGGTGATGAACAATCTTCTGTTGCTCATTCCTTTTGATGATCCATCAAATTACGTATTAGAAAAAGAAAAAAAA GTATCAGACGTTGAAACCTTCTGCTCGGAGACAAATTTTATGAGCCTCGATCC*ATTTAAATCAAGCCCGGG (SEQ ID NO: 72)

Table 5:

DNA sequences of the expression systems tested in *Chlamydomonas reinhardtii*. The functional DNA parts are indicated: 8×sTF binding site (white text, black highlight); core promoters (underlined); mCherry coding region (white text, grey highlight); terminators (*italics, grey highlight*); sTFs (grey highlight).

DNA sequence of the expression system tested in *Chlamydomonas reinhardtii*	
8BS(BM3R1)- Cr-eIF-5Acp - mCherry + Cr-RPS3Acp – BM3R1_sTF	GCATTTGCTCGGCTAGTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGG ACTCTAGATAAGCACGGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTA GTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGGACTCTAGATAAGCAC GGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTAGTTCTCCCCGGAAAC TGCGACGAAGGGATGTCTCCGCAAGGCAAGTATATAACGGCTAGCAACGTATGCCTTAGCATAGTAGAGCAAT TAGTTGTCTATGTGCCTCGGTGCAAGCGCACACGCCGGGAATAATGCGGCATGGGGGCTTCTGTTGGCCCCA TGCGAGCCCCCAGGAAGAAAAGTCGCGCGGCGCCCGTATTCTGCCCTCTTGCTGTGCCAACCTCCTAGTCGC TTCTTCGCACTTTTTAATTAAA[redacted]TAA*ATGGAGGCGCTCGTTGATCTGAGCCTTGCCCCCTGACGAACGGCGGTG* *GATGGAAGATACTGCTCTCAAGTGCTGAAGCGGTAGCTTAGCTCCCCGTTTCGTGCTGATCAGTCTTTTTCAA* *CACGTAAAAAGCGGAGGAGTTTTGCAATTTTGTTGGTTGTAACGATCCTCCGTTGATTTTGGCCTCTTTCTCCA* *TGGGCGGGCTGGGCGTATTTGAAGCGCTTTTGGAAAAGTTGCTGCGGGGTTCATCAGCTGAAGGGGACTCG* *GTTCGCAGATCAGTTACACACTAAAGAACGGCGGGTAGCAACACCAGCAAACGTGACGAAACGGAACCGTGC* *AGC*ATTTAAATGGCCCGAACTTGCTCTCGGTGTCATATTGCACCATCCCATCTTGTATAACCGATATAACATAG CTTCGAGTGTGCCGATAAATTATTGTGAGGGCGTCGGGGGGGCGAGCTGAGGGAAATGGAGGGGGCACTCAT CTCGGCCGCCCCTCCCATCGCGACCTCGGCGCTCAAGCGGGGGTCCCGCACTCGCTTCGGTCTCTTTTGGT CAGCAGCCGTTTGTTGACTACCGTTAATTAAAATGGAGAGCACCCCTACCAAGCAGAAGGCGATCTTTTCGGC TTCGCTGCTGCTGTTTGCCGAGCGCGGGTTTGATGCTACCACCATGCCCATGATCGCTGAGAACGCTAAGGT CGGGGCGGGCACGATTTACCGGTACTTCAAGAACAAGGAGTCGCTCGTCAACGAGCTGTTTCAGCAGCATGT GAACGAGTTCCTGCAGTGCATTGAGTCCGGTCTCGCCAACGAGCGGGATGGCTACCGCGATGGTTTCCATCA CATCTTCGAGGGCATGGTCACGTTTACGAAGAACCATCCTCGCGCTCTCGGTTTTATCAAGACCCATTCCCAG GGGACCTTTCTCACGGAGGAGTCGCGGCTGGCTTACCAGAAGCTGGTCGAGTTTGTCTGCACCTTTTTCCGG GAGGGTCAGAAGCAGGGTGTGATTCGGAACCTGCCGGAGAACGCTCTCATTGCTATCCTCTTTGGCTCGTTT ATGGAGGTCTACGAGATGATTGAGAACGATTACCTGTCCCTGACGGACGAGCTGCTCACGGGCGTCGAGGAG AGCCTCTGGGCTGCTCTGTCGCGGCAGTCGGAGCTCCCCCCTAAGAAGAAGCGCAAGGTGTCGGCCTCCGG GAGCGGCCGGGCTGATGCTCTGGATGACTTCGACCTGGACATGCTGGGTAGCGACGCTCTCGACGATTTTGA CCTGGACATGCTCGGCTCGGATGCCCTGGACGATTTCGATCTGGATATGCTGGGTAGCGACGCGCTCGACGA CTTTGATCTGGACATGCTCATTAACAGCCGCTAAACGCGT*GGCCCCACCGTTGCGTGTGCGCCCGCGGTGCG* *CTGCGCGGTCGGCAGCTTGGGTGTGGCATCCGGTGCGGCTTGTCCCGCCGGCATGTAGCTCTTATGTAACG* *GGCTGTCTGTACTCACTTGTGTCCAACCGCCTCTCTGGATGTCTGGTTCATGACCAACAGCTAAGCAAAGAAG* *CAGCTGGGACACCAGGGGACGCTGACAATGGAGTGGGCAGCCGACGCAGCAGAGGGGGGACTGCGAGTTA* *TACGGTATTAGGCTGGGCTGGCAGGTCCGGTAGACGGTAATGCGACACACAAGCCGTGGGAGAAGGTTGCG* *TCAGGAAGTCCAAGCAGGTTCTGT*ATTTAAATGCGGCCGC (SEQ ID NO: 73)
8BS(TetR)-Cr-eIF-5A-cp-mCherry + Cr-RPS3Acp –TetR_sTF	TTGCTCGGCTAGCTCTCTATCACTGATAGGGAGTATTGACAAGCTTTCTCTATCACTGATAGGAGTGGCTTATC TAGATCTCTATCACTGATAGGGAGTTCACATCCTAGGTCTCTATCACTGATAGGGAGTACTAGCTCTCTATCAC TGATAGGGAGTATTGACAAGCTTTCTCTATCACTGATAGGAGTGGCTTATCTAGATCTCTATCACTGATAGGGA GTTCACATCCTAGGTCTCTATCACTGATAGGGAGTACTAGTTCTCCCCGGAAACTGCGACGAAGGGATGTCTC CGCAAGGCAAGTATATAACGGCTAGCAACGTATGCCTTAGCATAGTAGAGCAATTAGTTGTCTATGTGCCTCG GTGCAAGCGCACACGCCGGGAATAATGCGGCATGGGGGCTTCTGTTGGCCCCATGCGAGCCCCCAGGAAGA AAAGTCGCGCGGCGCCCGTATTCTGCCCTCTTGCTGTGCCAACCTCCTAGTCGCTTCTTCGCACTTTTTAATT AAA[redacted]TAA*ATGGAGGCGCTCGTTGATCTGAGCCTTGCCCCCTGACGAACGGCGGTGGATGGAAGATACTGCTCTC* *AAGTGCTGAAGCGGTAGCTTAGCTCCCCGTTTCGTGCTGATCAGTCTTTTTCAACACGTAAAAAGCGGAGGAG* *TTTTGCAATTTTGTTGGTTGTAACGATCCTCCGTTGATTTTGGCCTCTTTCTCCATGGGCGGGCTGGGCGTATT* *TGAAGCGCTTTTGGAAAAGTTGCTGCGGGGTTCATCAGCTGAAGGGGACTCGGTTCGCAGATCAGTTACACA* *CTAAAGAACGGCGGGTAGCAACACCAGCAAACGTGACGAAACGGAACCGTGCAGC*ATTTAAATGGCCCGAAC TTGCTCTCGGTGTCATATTGCACCATCCCATCTTGTATAACCGATATAACATAGCTTCGAGTGTGCCGATAAAT TATTGTGAGGGCGTCGGGGGGCGAGCTGAGGGAAATGGAGGGGGCACTCATCTCGGCCGCCCCTCCCATC GCGACCTCGGCGCTCAAGCGGGGGTCCCGCACTCGCTTCGGTCTCTTTTGGTCAGCAGCCGTTTGTTGACTA

	CCGTTAATTAAAATGAGCCGGCTGGATAAGTCCAAGGTCATCAACTCCGCGCTCGAGCTGCTCAACGAGGTC GGGATCGAGGGCCTGACGACCCGGAAGCTGGCGCAGAAGCTGGGGGTGGAGCAGCCGACCCTGTACTGGC ACGTCAAGAACAAGCGGGCCCTGCTCGATGCCCTCGCTATCGAGATGCTCGATCGGCATCATACGCATTTTT GCCCTCTCGAGGGGGAGTCCTGGCAGGACTTTCTGCGGAACAACGCCAAGTCGTTCCGGTGCGCCCTGCTG TCCCATCGGGATGGTGCTAAGGTCCATCTCGGGACGCGGCCTACCGAGAAGCAGTACGAGACCCTGGAGAA CCAGCTCGCTTTTCTGTGCCAGCAGGGGTTCTCCCTGGAGAACGCTCTCTACGCCCTCTCCGCTGTGGGTCA TTTTACCCTCGGTTGCGTCCTGGAGGATCAGGAGCATCAGGTCGCCAAGGAGGAGCGGGAGACCCCTACCA CGGACTCGATGCCCCCTCTCCTCCGGCAGGCTATTGAGCTGTTTGACCATCAGGGCGCGGAGCCTGCCTTTC TCTTTGGGCTCGAGCTGATTATTTGCGGCCTCGAGAAGCAGCTCAAGTGCGAGTCCGGTTCGGAGCTCCCTC CTAAGAAGAAGCGCAAGGTGAGCACCGCCCCCCCCACGGATGTGTCCCTGGGTGATGAGCTGCATCTCGAC GGCGAGGATGTCGCCATGGCTCATGCCGATGCCCTGGATGATTTCGATCTCGATATGCTGGGTGATGGTGAC TCCCCCGGTCCGGGTTTTACGCCTCACGATAGCGCCCCTTACGGCGCTCTGGACATGGCCGATTTTGAGTTT GAGCAGATGTTCACGGACGCGCTCGGCATCGACGAGTACGGCGGTTAAACGCGT*GGCCCCACCGTTGCGTG TGCGCCCGCGGTGCGCTGCGCGGTCGGCAGCTTGGGTGTGGCATCCGGTGCGGCTTGTCCCGCCGGCATG TAGCTCTTATGTAACGGGCTGTCTGTACTCACTTGTGTCCAACCGCCTCTCTGGATGTCTGGTTCATGACCAA CAGCTAAGCAAAGAAGCAGCTGGGACACCAGGGGACGCTGACAATGGAGTGGGCAGCCGACGCAGCAGAG GGGGGACTGCGAGTTATACGGTATTAGGCTGGGCTGGCAGGTCCGGTAGACGGTAATGCGACACACAAGCC GTGGGAGAAGGTTGCGTCAGGAAGTCCAAGCAGGTTCTGT*ATTTAAATGCGGCCGCG (SEQ ID NO: 74)

Table 6:

DNA sequence of the expression systems tested in Chinese hamster ovary cells (CHO cells - *Cricetulus griseus*). The functional DNA parts are indicated: 8×sTF binding site (white text, black highlight); core promoters (underlined); mCherry coding region (white text, grey highlight); terminators (*italics, grey highlight*); sTF (grey highlight).

DNA sequence of the expression system tested in the CHO cells (*Cricetulus griseus*)	
8BS(BM3R1)- Mm-Eef2-cp - mCherry + Mm-Atp5b-cp – BM3R1_VP64	TTTGCTCGGCTAGTCGGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGGACT CTAGATAAGCACGGAATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTAGTC GGAATGAACATTCATTCCGAGACCTAGGATGTGACGGAATGAAGGTTCATTCCGGACTCTAGATAAGCACGGA ATGAACTTTCATTCCGCTGAAGCTTGTCAATCGGAATGAAGGTTCATTCCGGCTAGTTCTCCCCGGAAACTGC CGGACGAGCACCCGGCGCCGTCACGTGACGCACCCAACCGGCGTTGACCTATAAAAGGCCGGGCGTTGACG TCAGCGGTCTCTTCCGCCGCAGCCGCCGCCATCGTCGGCGCGCTTCCCTGTTCACCTCTGACTCTGAGAATC CGTCGCCATCCGCCACC[redacted]TAAGATATC*TTCCCCAAAGCCACGTGACTTTACTGGTCACTGAGGCAGTGCATGCATGTCAGGCTGCC TTCATCTTTTCTATAAGTTGCACCAAAACATCTGCTTAAGTTCTTTAATTTGTACCATTTCTTCAAATAAAGAATT TTGGTACCCAGCTTCTTTTCTTTGTGATTGAGGATAAGCATTCCAGCTTCCAGTTGCTTCACCGCCAGTTATAC TAATCACACTGAAACACCTAAAAGAATATTCACGTTTATTAAACTCCTTAGTTTGGGAAAGATCGTAAAATACAG GTGTTTTCAGGCAGGACTATTAAGTACTCTTGGTTCTGAGTTACATGCTAGACTGTCGTGGGAACACACTCCT GGGTGTCGCTGCTTGTGTGCCTTTGACTGGGTCAGTG*ATTTAAATATTGGCACCAGTTTAGACCAATAGCTGA TAAGCTCCGAGTTTTTTTACCCTATAGAAGCGTTAGTGGTGATGACGAACAGCAAAATCACCCAATTACTGTGC CTACGGCGGAGGTTGCCCCGCCCCAGCTGCAGGACCGGCGGAGAGGACCGCTTCGGCGCTCAGTCTCCAC CCGGATTCCGCCATGGAAAGCACACCAACAAAGCAAAAAGCAATATTTTCAGCCTCACTTCTTTTGTTTGCCGA GAGGGGTTTCGACGCTACAACAATGCCCATGATAGCCGAAAATGCCAAAGTAGGAGCCGGGACAATATACAG GTATTTTAAAAACAAGGAAAGTCTGGTCAATGAACTTTTCCAGCAGCACGTAAATGAGTTTCTTCAATGTATTGA ATCTGGCCTGGCTAACGAACGCGACGGTTATCGTGATGGCTTTCATCACATATTTGAGGGAATGGTCACTTTC ACCAAAAATCACCCTAGGGCCTTGGGCTTTATCAAAACACATTCTCAGGGTACATTCCTCACCGAGGAATCTC GACTCGCCTATCAAAAGCTCGTTGAGTTTGTCTGTACTTTCTTTAGGGAGGGACAAAAGCAAGGCGTAATCCG AAACCTCCCAGAGAACGCCTTGATCGCTATTCTCTTCGGATCTTTTATGGAGGTCTATGAGATGATCGAAAATG ACTATCTTAGTCTGACAGATGAGCTTCTGACAGGTGTTGAAGAATCATTGTGGGCTGCTTTGTCTAGACAGAGT GAATTCCCTCCCAAGAAGAAACGAAAGGTAAGCGCCTCTGGTTCAGGTCGTGCTGACGCTCTGGATGATTTTG

	ATCTCGACATGCTTGGTTCTGATGCTCTTGACGACTTCGACCTTGATATGCTGGGCAGTGACGCATTGGACGA CTTTGATTTGGACATGTTGGGAAGCGATGCCTTGGACGACTTTGACCTTGATATGCTGATAAATAGTCGCTAAT AG*CTCGAGCGCCTCCTCCTCCCATAGCCGATGGCCACAGTCAATTCACCACCCCAGGGTCCTCAGCTAGGAG GAGGACAGAGTGTGGAAAGTAGACAGTTTCCACTTCCTTTTCCCTACATCTTTCAGTATGAGGGTACCATATCC TGCTCCACCCAGGTCCTGTGGATAACAATAAAAAAGGAAGTGTGTGTGCCTTTGTATGTGTTCCCCTCACGTC TTTGACAATGGGGTTGGGGAGGTCTGGGGTCAGAGAGAATTGCGTTGTGGGATTTTGAGTTAACTGCTTTTGG CTTTAGAGATCGACAGTCTAAGAGGTAAAATTAGATGTGAATTAGTTGGGAAGCTGCCAAGTGTCCCAGAGCT TTGGACACCCACTCTAGGGACACATTGTCCCCTT*ATTTAAATAGGGCCCGTTTAAACC (SEQ ID NO: 75)

Examples

Example 1

The bacterial DNA-binding proteins and their binding sites used in the expression systems:

- LexA (transcription repressor from *Escherichia coli*; GenBank: EDV67321.1)

  LexA binding sites (regardless of the DNA strand):

  CTGTATATAAACACAG (SEQ ID NO: 76);

  CTGTATATATACCCAG (SEQ ID NO: 77);

  CTGTATATAAAACCAG (SEQ ID NO: 78);

  GTGGTTATATATACAG (SEQ ID NO: 79)

- SrpR (transcriptional regulator from *Pseudomonas putida*; NCBI Reference Sequence: WP_019437727.1)

  SrpR binding sites (regardless of the DNA strand):

  ATATACATACATGCTTGTTTGTTTGTAAAC (SEQ ID NO: 80);

  ATTTACATACATTCTTGTTTGTTTGTAAAC (SEQ ID NO: 81)

- PhlF (transcriptional regulator from *Pseudomonas protegens*; GenBank: AAF20928.1)

  PhlF binding sites (regardless of the DNA strand):

  ATGATACGAAACGTACCGTATCGTTAAGGT (SEQ ID NO: 82);

  ATGATACGGAACGTTACGTATCGTTAAGCT (SEQ ID NO: 83);

  ATGATACGGAAGCTACCGTATCGTAAAGGT (SEQ ID NO: 84);

  ATGATACGTAACGTACCGTATCGTAAAGGT (SEQ ID NO: 85)

- TetR (transcriptional regulator from *Escherichia coli*, GenBank: EFK45326.1)

  TetR binding site (regardless of the DNA strand):

  ACTCCCTATCAGTGATAGAGA (SEQ ID NO: 86)

- BM3R1 (transcriptional regulator from *Bacillus megaterium*; NCBI Reference Sequence: WP_013083972.1)

  BM3R1 binding sites (regardless of the DNA strand):

  CGGAATGAAGGTTCATTCCG (SEQ ID NO: 87);

  CGGAATGAACTTTCATTCCG (SEQ ID NO: 88);

CGGAATGAACATTCATTCCG (SEQ ID NO: 89);
CGGAATGAACGTTCATTCCG (SEQ ID NO: 90)

- TarA (transcriptional regulator *Streptomyces lavenduligriseus*; NCBI Reference Sequence: WP_030788560.1)

TarA binding sites (regardless of the DNA strand):
AACATACCGTGTGGTATGTT (SEQ ID NO: 91);
AACATACCGAGTGGTATGTT (SEQ ID NO: 92);
AACATACCGTGAGGTATGTT (SEQ ID NO: 93);
AAACATACCGTGTGGTATGTTC (SEQ ID NO: 94)

- LacI (lac repressor from *Escherichia coli*; NCBI Reference Sequence: WP_048339836.1)

LacI binding site (regardless of the DNA strand):
AATTGTGAGCGGCTCACAATT (SEQ ID NO: 95)

Example 2

Test of different versions of the sTFs and assessment of modulation of the expression system performance in *Saccharomyces cerevisiae*. (Figure 5)

The expression systems (individual expression cassettes for the sTFs and for the reporters) were constructed as two separate DNA molecules (plasmids) (Figure 5A). The plasmids with the expression cassettes for each sTF contained: 1) the *Saccharomyces cerevisiae* codon-optimized coding region of the DNA binding protein (LexA, PhlF, SrpR, TetR, BM3R1, and TarA; Example 1) in each sTF coding region, 2) NLS and the VP16 activation domain in each sTF coding region, 3) the Sc-TDH3cp (Table 1) controlling the expression of each sTF, 4) the *URA3* selection marker gene (of *Kluyveromyces lactis* origin), 5) the flanks for integration into the genome by homologous recombination into the *ura3-52* locus (for replacing the mutated coding region of the locus), 6) regions needed for propagation of the plasmid in *E. coli*. The plasmids with the reporter cassettes contained 1) the *Saccharomyces cerevisiae* codon-optimized coding region of the Venus (yellow fluorescent) protein, 2) the Sc-ENO1cp (Table 1) controlling the expression of Venus together with 3) upstream positioned sTF-specific binding sites (0, 1, 2, 4, or 8) (Example 1), 4) the *LEU2* selection marker gene (of *Kluyveromyces lactis* origin), 5) the flanks for integration into the genome by homologous recombination into the *leu2-3_112* locus (replacing the mutated coding region of the locus), and 6) regions needed for propagation of the plasmid in *E. coli*.

Saccharomyces cerevisiae CEN.PK (MATα, *ura3-52 leu2-3_112 his3Δ1 MAL2-8C SUC2*) was used as the parental strain. The expression cassettes (Figure 5A) were introduced into cells through transformation of the linearized integrative plasmids, the sTF and the corresponding reporter expression cassettes were transformed into a single strain. Each integration cassette was released by *Not*I restriction endonuclease from the plasmid prior to the transformation. Transformations were performed using the standard lithium acetate protocol. The single copy integrations were confirmed by qPCR, where the qPCR signal of the Venus gene was compared to a qPCR signal of a unique native sequence in each strain.

For all cultivations, 6.7 g/L of yeast nitrogen base (YNB, Becton, Dickinson and Company), synthetic complete amino acid mixture lacking uracil and leucine supplemented with 20 g/L D-glucose (SCD-LU) was used. In case of agar plate cultivations, 20 g/L agar was used in addition to the above mentioned components.

Pre-cultures of the tested strains were grown for 24-48 hours on the SCD-LU agar plates prior to inoculation of 4 ml of SCD-LU in 24-well cultivation plates to OD600=0.2. The cultures were grown for 18 hours at 800 rpm (Infors HT Microtron) and 28°C in triplicates, centrifuged, washed, and resuspended in 0.2 ml of sterile water. Two hundred µl of the cell suspension was analysed in black 96-well (Black Cliniplate; Thermo Scientific) using the Varioskan (Thermo Electron Corporation) fluorimeter. The settings for Venus were 510 nm (excitation) and 530 nm (emission), respectively. For normalization of the fluorescence results, the analyzed cell-suspensions were diluted 100× and OD600 was measured in transparent 96-well microtiter plates (NUNC) using Varioskan (Thermo Electron Corporation).

The results from the fluorescent analysis are shown in Figure 5B.

Example 3

Quantitative analysis of the expression system performance in diverse fungal hosts (Figure 6)

The expression systems (Table 2, Figure 4) and their negative control versions (the expression systems with deleted regions spanning the core promoter controlling the sTF and the sTF itself) were cloned into plasmids introducing selection markers and genome-integration flanks for 6 different species. **1) *Saccharomyces cerevisiae*** CEN.PK strain was used as the parental strain. The expression sys-

tems (including the versions with the *Saccharomyces cerevisiae* codon-optimized coding region of the DNA binding protein, BM3R1), including the *LEU2* selection marker gene (of *Kluyveromyces lactis* origin), were integrated into the *leu2-3_112* locus (replacing the mutated coding region of the locus) using the corresponding flanking regions for homologous recombination. The transformations were done by the standard lithium acetate protocol. **2) *Aspergillus niger*** ATCC1015 strain was used as the parental strain. The expression systems, including a hygromycin-resistance selection marker gene with a suitable promoter and terminator, were integrated into *gaaC* locus (replacing the native coding region) using the corresponding flanking regions for homologous recombination. The transformations were carried out by using the CRISPR transformation protocol (see below), including: protoplasts of the *A. niger* strain, linear donor DNA (expression cassette with the selection marker and the integration flanks), protein Cas9 (IDT) and mix of synthetic crRNA and tracrRNA (IDT). Cas9, crRNA and tracrRNA form a ribonucleoprotein (RNP) complex that generates a double-stranded brake at the target genomic locus which is then repaired with the linear donor DNA by homologous recombination. **3) *Trichoderma reesei*** strain M124 (VTT culture collection) was used as the parental strain. The expression systems, including a hygromycin-resistance selection marker gene with a suitable promoter and terminator, were integrated into *pep4* locus (replacing the native coding region) using the corresponding flanking regions for homologous recombination. The transformations were done by using the CRISPR transformation protocol (see below), including: protoplasts of the *T. reesei* strain, linear donor DNA (expression cassette with the selection marker and flanking regions) and RNP complex that generates a double-stranded brake at the target genomic locus which is then repaired with the linear donor DNA. **4) *Pichia kudriavzevii*** ATCC 32196 strain was used as the parental strain. The expression systems, including a hygromycin-resistance selection marker gene with a suitable promoter and terminator, were integrated into *PDC1* locus (replacing the native coding region) using corresponding flanking regions for homologous recombination. The transformations were done by using the standard lithium acetate protocol. **5) *Pichia pastoris*** X-33 strain (Invitrogen) was used as the parental strain. The expression systems (with the coding region of the DNA binding protein, BM3R1, that was codon-optimized to fit the codon usage of *Saccharomyces cerevisiae*), including a zeocin-resistance selection marker gene with suitable promoter and terminator, were integrated into the *AOX1* locus (integration into the *AOX1* promoter region) using the corresponding flanking regions for homologous recombination. The transformations were done by using the standard

lithium acetate protocol. **6)** ***Yarrowia lipolytica*** C-00365 (VTT culture collection) was used as the parental strain. The expression systems, including a nourseothricin-resistance selection marker gene with a suitable promoter and terminator, were integrated into the *ANT1* locus (replacing the native coding region) using the corresponding flanking regions for homologous recombination. The transformations were done by using the standard lithium acetate protocol.

The CRISPR transformation protocol: Isolated protoplasts were suspended into 200 µl of STC solution (1.33 M sorbitol,10 mM Tris-HCl, 50 mM $CaCl_2$, pH 8.0). One hundred µl of protoplast suspension was mixed with 3.5 µg of donor DNA and 20 µl of RNP-solution (1µM Cas9 protein (IDT), 1µM synthetic crRNA (IDT), and 1µM tracrRNA (IDT)) and 100 µl of the transformation solution (25% PEG 6000, 50 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5). The mixture was incubated on ice for 20 min. Two ml of transformation solution was added and the mixture was incubated 5 min at room temperature. Four ml of STC was added followed by addition of 7 ml of the molten (50°C) top agar (200g/L D-sorbitol, 6.7 g/L of yeast nitrogen base (YNB, Becton, Dickinson and Company), synthetic complete amino acid, 20 g/L D-glucose, 400 mg/L (for *A. niger*) or 100 mg/L (for *T. reesei*) of hygromycin B, and 20g/L agar). The mixture was poured onto a hygromycin selection plate (200 g/L D-sorbitol, 6.7 g/L of yeast nitrogen base (YNB, Becton, Dickinson and Company), synthetic complete amino acid, 20 g/L D-glucose, 400 mg/L (for *A. niger*) or 100 mg/L (for *T. reesei*) of hygromycin B, 20 g/L agar). Cultivation was done at +28 °C for five or seven days, colonies were picked and re-cultivated on the YPD plates containing 400 mg/L (for *A. niger*) or 100 mg/L (for *T. reesei*) of hygromycin B.

The correct integrations were confirmed by PCR of the genomic DNA of each transformed strain, where the amplicon (amplified DNA region) spanned the integrated construct and the genomic DNA outside of the integration flanks. The single copy integrations were confirmed by qPCR, where the qPCR signal of the mCherry gene was compared to a qPCR signal of a unique native sequence in each host.

For the liquid cultivations, 6.7 g/L of yeast nitrogen base (YNB, Becton, Dickinson and Company), synthetic complete amino acid supplemented with 20 g/L D-glucose (SCD) was used. In case of agar plate cultivations, solidified medium containing 20 g/L agar, 20 g/L bacto peptone (Becton Dickinson), 10 g/L yeast extract, and 20 g/L D-glucose (YPD plates). To obtain spores of the filamentous fungi,

PDA agar plates were used for sporulation (39 g/L BD-Difco Potato dextrose agar).

For the flow-cytometry analysis of the mCherry production in the tested strains (Figure 6A), pre-cultures of the tested yeast strains were grown for 24-48 hours on YPD agar plates and the filamentous fungi (*A. niger* and *T. reesei* strains) were sporulated on PDA plates (7 days), spores collected, and diluted in 1x PBS prior to analysis. In case of yeasts strains, 4 ml of SCD medium in 24-well cultivation plates was inoculated from pre-cultures to OD600=0.2 by every tested yeast strain. The cultures were grown for 18 hours at 800 rpm (Infors HT Microtron) and 28°C. One hundred µL of the culture was combined with 1,5 mL of 1x PBS prior to analysis. Measurements were done with FACSAria III (BD), where 10000 events were recorded and results were normalized by dividing mCherry fluorescence values by cell size (forward scatter, FSC-A). The results from the flow-cytometry fluorescent analysis are shown in Figure 6A.

For the quantitative fluorometry analysis (Figure 6B), pre-cultures of the tested yeast strains were grown for 24-48 hours on YPD agar plates and pre-cultures (inoculated by spores) of *Trichoderma reesei* strains were grown for 24 hours in YPG medium (20 g/L bacto peptone, 10 g/L yeast extract, and 30 g/L gelatin). Four ml of the SCD medium in 24-well cultivation plates was inoculated to OD600=0.2 by every tested yeast strain (OD600=0.5 in case of *T. reesei*). The cultures were grown for 18 hours at 800 rpm (Infors HT Microtron) and 28°C in triplicates, centrifuged, washed, and resuspended in 0.2 ml of sterile water. Two hundred µl of each cell suspension was analyzed in black 96-well plates (Black Cliniplate; Thermo Scientific) using the Varioskan (Thermo Electron Corporation) fluorimeter. The settings for mCherry were 587 nm (excitation) and 610 nm (emission), respectively. For normalization of the fluorescence results, the analyzed cell-suspensions were diluted 100× and OD600 was measured in transparent 96-well microtiter plates (NUNC) using Varioskan (Thermo Electron Corporation). The results from the analysis are shown in Figure 6B.

Example 4

Analysis of the adjustable expression levels in different hosts (*Pichia kudriavzevii*, *Aspergillus niger*, and *Trichoderma reesei*) (Figure 7)

The expression systems for *Pichia kudriavzevii* and *Aspergillus niger* with diverse numbers of the sTF-specific binding sites (0, 1, 2, 4, or 8) (Figure 7A) were constructed analogously to the Example 3 (the version of the expression system A – shown in Figure 4 – was used). In case of *Pichia kudriavzevii*, the ATCC 32196 strain was used as the parental strain. The expression systems, including hygromycin-resistance selection marker gene with a suitable promoter and terminator, were integrated into *PDC1* locus (replacing the native coding region) using the corresponding flanking regions for homologous recombination. The transformations were done using the standard lithium acetate protocol. In case of *Aspergillus niger*, the ATCC1015 strain was used as the parental strain. The expression systems, including a hygromycin-resistance selection marker gene with a suitable promoter and terminator, were integrated into the *gaaC* locus (replacing the native coding region) using the corresponding flanking regions for homologous recombination. The transformations were done using the CRISPR transformation protocol, including: protoplasts of the *A. niger* strain, linear donor DNA (expression cassette with the selection marker and the integration flanks) and RNP complex that generates a double-stranded brake at the target genomic locus which is then repaired with the linear donor DNA.

The DNA molecule, containing the expression systems for *Trichoderma reesei* for adjustable expression of the CBH1 gene (Figure 7C), contained the 201cp (Table 1) together with upstream positioned BM3R1-specific binding sites (0, 1, 2, 4, or 8) controlling the expression of the CBH1 coding region, the *T. reesei* PDC1 terminator, 533cp (Table 1) controlling the expression of the sTF coding region, the sTF coding region, the *Trichoderma reesei* TEF1 terminator, the hygromycin-resistance selection marker gene with suitable promoter and terminator, and the flanks for integration into the genome by homologous recombination into the *CBH1* locus (replacing the native coding region). The *T. reesei* strain M124 (VTT culture collection) was used as the parental strain. The transformations were done by the protoplast transformation protocol (see below), including: protoplasts of the *T. reesei* strain and linear donor DNA (expression cassette with the selection marker and the integration flanks).

The protoplast transformation protocol: Isolated protoplasts were suspended into 200 µl of STC solution (1.33 M sorbitol,10 mM Tris-HCl, 50 mM $CaCl_2$, pH 8.0). One hundred µl of protoplast suspension was mixed with 10 µg of the donor DNA and 100 µl of the transformation solution (25% PEG 6000, 50 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5). The mixture was incubated on ice for 20 min. Two ml of transformation solution was added and the mixture was incubated 5 min at room temperature. Four ml of STC was added followed by addition of 7 ml of the molten top agar (200 g/L D-sorbitol, 6.7 g/L of yeast nitrogen base (YNB, Becton, Dickinson and Company), synthetic complete amino acid, 20 g/L D-glucose, 100 mg/L hygromycin B, 20 g/L agar). The mixture was poured onto a selection plate (200 g/L D-sorbitol, 6.7 g/L of yeast nitrogen base (YNB, Becton, Dickinson and Company), synthetic complete amino acid, 20 g/L D-glucose, 100 mg/L hygromycin B, and 20 g/L agar). Cultivation was done at +28 °C for five days; colonies were picked and re-cultivated on the YPD plates containing 100 mg/L hygromycin B.

The correct integrations were confirmed by PCR of the genomic DNA of each transformed strain, where the amplicon (amplified DNA region) spanned the integrated construct and the genomic DNA outside of the integration flanks. The single copy integrations were confirmed by qPCR, where the qPCR signal of the mCherry gene (for *Pichia kudriavzevii* and *Aspergillus niger* strains) or the BM3R1 coding region (for *Trichoderma reesei* strains) was compared to a qPCR signal of a unique native sequence in each host.

For liquid cultivations, 6.7 g/L of yeast nitrogen base (YNB, Becton, Dickinson and Company), synthetic complete amino acid supplemented with 20 g/L D-glucose (SCD) was used. In case of agar plate cultivations, solidified medium containing 20 g/L agar, 20 g/L bacto peptone (Becton Dickinson), 10 g/L yeast extract, and 20 g/L D-glucose (YPD plates) was used. To obtain spores of the filamentous fungi, PDA agar plates were used for sporulation (39 g/L BD-Difco Potato dextrose agar).

For the flow-cytometry analysis of mCherry production in the tested strains (Figure 7B), pre-cultures of the *Pichia kudriavzevii* strains were grown for 24-48 hours on the YPD agar plates and the *Aspergillus niger* strains were sporulated on PDA plates (for 7 days), spores collected, and diluted in 1x PBS prior to analysis. In case of the *Pichia kudriavzevii* strains, 4 ml of the SCD medium in 24-well cultivation plates was inoculated from pre-culture to OD600=0.2. The cultures were grown for 18 hours at 800 rpm (Infors HT Microtron) and 28 °C. One hundred µL of

the culture was combined with 1,5 mL of 1x PBS prior to analysis. Measurements were done with FACSAria III (BD), where 10000 events were recorded and results were normalized by dividing mCherry fluorescence values by cell size (forward scatter, FSC-A). The results from flow-cytometry fluorescent analysis are shown in Figure 7B.

For the western blot analysis of the CBH1 production in the *Trichoderma reesei* strains, pre-cultures (inoculated by spores) were grown for 24 hours in YPG medium (20 g/L bacto peptone, 10 g/L yeast extract, and 30 g/L gelatine). Four ml of either SGE-lactose (15 g/L KH_2PO_4, 5.4 g/L Na_2SO_4, 1 mL/L trace elements (3.7 mg/L $CoCl_2$, 5 mg/L $FeSO_4.7H_2O$, 1.4 mg/L $ZnSO_4.7H_2O$, 1.6 mg/L $MnSO_4.7H_2O$), 40 g/L lactose, 333.25 g/L spent grain extract, 8.6 g/L $(NH_4)_2$-citrate, 100 mM PIPPS, 2.4 mM $MgSO_4$, and 4.1 mM $CaCl_2$, pH adjusted to 4.8 with KOH) or the SCD medium in 24-well cultivation plates was inoculated to OD600=0,5 for each tested strain. The cultures were grown for 3 days at 800 rpm (Infors HT Microtron) and 28°C, centrifuged, and the supernatant transferred into a clean tube. Fifteen µL of each supernatant was mixed with 4 µL of 4x SDS loading buffer (400 ml/L glycerol, 100 ml/L β-mercaptoethanol, 2 g/L OrangeG dye (Sigma), 40 g/L SDS, and 125 mM Tris-HCl pH 6.8), boiled and loaded on the 4–20% SDS-PAGE gradient gel. The gel was transferred onto a nitrocellulose membrane, and the CBH1 protein was detected with specific (mouse) anti-CBH1 primary antibody (and anti-mouse-IR680-conjugated secondary antibody), and visualization of the signal was performed on the Odyssey CLx Imaging System instrument (LI-COR Biosciences). The results from the analysis are shown in Figure 7D.

Example 5

Test of the expression system in *Kazachstania exigua* (Figure 8)

The expression system used for *Kazachstania exigua* (Table 3, Figure 8A) was cloned into a plasmid containing flanking regions for the *K. exigua* gene g706 encoding a homolog of *S. cerevisiae* ALD2. In the resulting construct, *K. exigua* g706 3'-UTR flanking region formed a terminator sequence for the sTF in the expression system. The expression system, including flanking regions for homologous recombination, was integrated into the g706 locus (replacing the native coding region).

Kazachstania exigua C-02458 (VTT culture collection) strain was modified by the replacement of both KU70 loci with the Cas9 expression cassette (containing suit-

able promoter and terminator). The resulting strain (*MAT a/α ura3Δ/ura3Δ ku70Δ::Cas9/ku70::Cas9*) was used as the parental strain (WT in Figure 8B). Transformation of the expression system (donor DNA) into the background strain was carried out together with a centromeric plasmid containing *URA3* selection marker and an expression cassette for a sgRNA that targets Cas9 into the g706 locus (sgRNA was expressed under the control of *S. cerevisiae* RNA polymerase III promoter *SNR52* and terminator *SUP4*). The resulting strain is shown as "SES" in Figure 8B.

Transformation was done by the electroporation protocol: Cells were inoculated in YPD medium and cultivated overnight at 250 rpm and 30 °C. The overnight culture was diluted to an OD600=0.2 and grown to an OD600=1.3. The harvested and washed cells were resuspended in 10 mL Tris-EDTA (pH 7.5) containing 10 mM dithiothreitol and incubated at 30 °C for 30 minutes. Forty mL of ice cold water was added to cells followed by centrifugation. This was followed with two washing steps, first with 50 mL of ice cold sterile water, then with 10 mL of ice cold 1 M sorbitol. Finally, cells were resuspended in 125 µL of ice cold 1 M sorbitol. Fifty µL of cell suspension was combined with 15 µL of a DNA mix (containing 5 µg of the donor DNA and 5 µg of the gRNA plasmid). Electroporation was performed in 2 mm cuvettes at 1.25 kV, 200 Ω and 25 µF. Nine hundred fifty µL of recovery solution (10 g/L yeast extract, 10 g/L Bacto peptone, 20 g/L glucose, 1 M sorbitol) was added immediately after electroporation. The cells were recovered for 30 minutes at 250 rpm and 30 °C before plating on SCD medium lacking uracil.

For expression analysis, the two strains (WT and SES) were cultivated in triplicates in SCD medium for 10 hours, and the SES strain also for 22 hours to reach stationary phase when all glucose had been consumed ("SES_stat" in the Figure 8B). Total RNA was isolated from the strains (RNeasy Kit - QIAGEN), cDNA was produced (Transcriptor First Strand cDNA Synthesis Kit – Roche), and transcription of the Venus and the *ADH1* (a glycolytic gene highly expressed in exponential growth phase and down-regulated in the absence of glucose) genes were analyzed by qPCR with primers specific for each gene. The *ALG9* gene was used as the normalization control for expression quantification. The results from the analysis are shown in Figure 8B.

Example 6

The expression system used for production of a secreted protein in fungi (*Trichoderma reesei* and *Pichia pastoris*) (Figure 9)

The DNA molecule, containing the expression system for *Trichoderma reesei* (Figure 9A), contained the 114cp (Table 1) together with upstream positioned eight BM3R1-specific binding sites controlling the expression of the *CBH1* coding region, the coding region for the CBH1 gene, the *Trichoderma reesei PDC1* terminator, 533cp (Table 1) controlling the expression of the sTF coding region, the sTF coding region, the *Trichoderma reesei TEF1* terminator, the hygromycin-resistance selection marker gene with a suitable promoter and terminator, and flanking regions for genomic integration into the *CBH1* locus (replacing the native coding region) by homologous recombination. The *T. reesei* strain M1763 (VTT culture collection) was used as the parental strain ("WT" in Figure 9B). Transformations were done by the protoplast transformation protocol (Example 4), using protoplasts of the *T. reesei* strain and linear donor DNA (expression cassette with the selection marker and integration flanks).

The correct integrations were confirmed using PCR from genomic DNA, where the amplicon (amplified DNA region) spanned the integrated construct and the genomic DNA outside of the integration flanks. Single copy integration was tested using qPCR, where the qPCR signal from the BM3R1 coding region was compared to the signal from a unique native sequence in the host. The strain containing the expression cassette ("SES" in the Figure 9B) was analyzed for the CBH1 production and compared to the background strain in cellulase-inducing and –repressing conditions.

The CBH1 production in *Trichoderma reesei* strains was carried out in 1 L bioreactors. Pre-cultures (inoculated with spores) for the cellulase-inducing conditions cultivations were grown for 24 hours in SGE-lactose medium (Example 4) to produce sufficient amount of mycelium for bioreactor inoculations. Pre-cultures (inoculated with spores) for the cellulase-repressing conditions cultivations were grown for 24 hours in YE-glucose-A medium (20 g/L glucose, 10 g/L yeast extract, 15 g/L KH_2PO_4, 5 g/L $(NH_4)_2SO_4$, 1 mL/L trace elements (3.7 mg/L $CoCl_2$, 5 mg/L $FeSO_4.7H_2O$, 1.4 mg/L $ZnSO_4.7H_2O$, 1.6 mg/L $MnSO_4.7H_2O$), 2.4 mM $MgSO_4$, and 4.1 mM $CaCl_2$, pH adjusted to 4.8). The cellulase-inducing bioreactor cultivations were inoculated in SGM medium ("SGM" in Figure 9B) (20 g/L spent grain

extract, 20 g/L sold spent grain, 60 g/L lactose, 5 g/l KH_2PO_4, 5 g/l NH_4SO_4, 1 mL/L trace elements, 2.4 mM $MgSO_4$, and 4.1 mM $CaCl_2$, 1mL/L Antifoam J647, pH 4.8), air flow at 0.5 slpm (0.4-0.6 vvm), and stirring at 600 rpm. The cellulase-repressing bioreactor cultivations were inoculated in the YE-glucose-B medium ("glucose" in Figure 9B) (10 g/L glucose, 20 g/L yeast extract, 5 g/L KH_2PO_4, 5 g/L NH_4SO_4, 1 mL/L trace elements, 2.4 mM $MgSO_4$, and 4.1 mM $CaCl_2$, 1mL/L Antifoam J647, pH 4.8), and these cultures were continuously fed with glucose (300 g/L glucose with flow rate at 4.4 g/h), air flow at 0.5 slpm (0.4-0.6 vvm), and stirring at 900 rpm. The cultivation was carried out for 150 hours, samples taken at various times, subset shown in Figure 9B.

For the coomassie stain analysis (Figure 9B – upper left panel), 1.5 µL of each culture (time-point) supernatant was mixed with 15 µL of 1x SDS loading buffer (100 ml/L glycerol, 25 ml/L β-mercaptoethanol, 0.5 g/L OrangeG dye (Sigma), 10 g/L SDS, and 31.2 mM Tris-HCl pH 6.8), boiled and loaded on the 4–20% SDS-PAGE gradient gel together with dilutions of purified CBH1 protein as a standard. The gel was stained with colloidal coomassie stain (PageBlue Protein Staining Solution; Thermo Fisher Scientific) according to the manufacture's protocol. The visualization of the stained gel was performed on the Odyssey CLx Imaging System instrument (LI-COR Biosciences). Protein concentration in the culture supernatant was estimated from the CBH1 standard in the same gel (Figure 9B – upper right table). For the western analysis (Figure 9B – lower left panel), 0.075 µL of each culture (time-point) supernatant was mixed with 15 µL of 1x SDS loading buffer, boiled and loaded on the 4–20% SDS-PAGE gradient gel together with dilutions of purified CBH1 protein. The gel was transferred onto a nitrocellulose membrane, and the CBH1 protein was detected with specific (mouse) anti-CBH1 primary antibody (and anti-mouse-IR680-conjugated secondary antibody), and the visualization of the signal was performed on the Odyssey CLx Imaging System instrument (LI-COR Biosciences). The CBH1 concentration in the culture supernatant was estimated (Figure 9B – lower right table).

The DNA molecule, containing the expression system for *Pichia pastoris* (Figure 9C), consisted of the 201cp (Table 1) together with upstream positioned eight BM3R1-specific binding sites controlling expression of the fusion protein coding region, the coding region for the fusion protein (consisting of N-terminal *Saccharomyces cerevisiae* secretion signal (α-factor), KEX/spe13 protease cleavage site, carbohydrate-binding module (CBM), elastin-like protein (ELP5), and another CBM), followed by the *S. cerevisiae ADH1* terminator, 008cp (Table 1) controlling

the expression of the sTF coding region, the sTF coding region (the coding region of the DNA binding protein, BM3R1, of the sTF was codon-optimized to fit the codon usage of *Saccharomyces cerevisiae*), the *Trichoderma reesei TEF1* terminator, the zeocin-resistance selection marker gene with suitable promoter and terminator, and the flanks for integration into the genome by homologous recombination into the *AOX1* locus (integration into the *AOX1* promoter region). Transformations were done by using the standard lithium acetate protocol. The strain containing a single copy of the expression cassette was tested for production of the protein in diverse conditions (Figure 9D).

The CBM-ELP5-CBM production in *Pichia pastoris* was carried out in Erlenmeyer flasks. The pre-culture was done for 24 hours in the YPP medium (10 g/L yeast extract, 20 g/L peptone, 13.4 g/L yeast nitrogen base, 0.4 mg/L biotin, 20 g/L glycerol, 13.2 mM K_2HPO_4, and 86.8 mM KH_2PO_4, pH=6.0) with 20 g/L glycerol (YPP-Gly). To test the effect of different carbon sources, glycerol was replaced either by 20 g/L glucose ("YPP-Glc" in Figure 9D) or by 20 g/L ethanol ("YPP-EtOH" in Figure 9D). The cells from the pre-culture were inoculated (to OD600=1.0) in YPP-Gly, YPP-Glc, or YPP-EtOH and cultured for 2 days, also addition of protease inhibitors (chymostatin and pepstatin) was tested. The pre-culture was also cultivated for additional two days (three days in total).

For the western analysis (Figure 9D), 22.5 µL of each culture supernatant was mixed with 7.5 µL of 4x SDS loading buffer, boiled and loaded on the 4–20% SDS-PAGE gradient gel. The gel was transferred onto a nitrocellulose membrane, and the CBM-ELP5-CBM protein was detected with specific (mouse) anti-CBM primary antibody (and anti-mouse-IR680-conjugated secondary antibody), and the visualization of the signal was performed on the Odyssey CLx Imaging System instrument (LI-COR Biosciences) (Figure 9D).

Example 7

Test of the expression system performance in plant organism (*Nicotiana benthamiana*)

Two expression systems are tested in *Nicotiana benthamiana* (Table 4): The expression systems assembled in single DNA molecules comprise two expression cassettes: 1) sTF expression cassette, which comprises a core promoter used for the sTF expression control, exemplified here with the At-RPL41D_cp (Table 1); the

sTF version with the DNA-binding protein, exemplified here by either BM3R1 or TetR, and with the activation domain, exemplified here by either VP16AD or VP64AD; and a terminator, exemplified here by the *Arabidopsis thaliana* MT3 terminator. And 2) the target gene expression cassette, which comprises a number of sTF specific binding sites, exemplified here by either eight BM3R1-specific binding sites or by eight TetR-specific binding sites; another core promoter, exemplified here by At-ATTI7_cp (see Table 1), the target gene coding region, exemplified here by the mCherry (red fluorescent protein reporter) coding region; and a terminator, exemplified here by the *Arabidopsis thaliana* PSBX terminator. The coding regions of the sTFs and the mCherry are codon-optimized to fit the codon usage of *Nicotiana benthamiana*. Also, negative control versions (the expression systems with deleted regions spanning the At-RPL41D_cp, the sTF, and the MT3 terminator) are constructed.

The expression systems (and the negative control versions) are cloned into a plasmid containing the plant (NptII) selectable marker coding region with suitable promoter and terminator, and the sequences for propagation in *Agrobacterium tumefaciens*, including kanamycin selection marker. The plasmids are transformed into *Agrobacterium tumefaciens* (strain EHA105) by electroporation (2 mm cuvettes; with settings: 1.25 kV, 200 Ω and 25 µF), and the transformants are grown in presence of kanamycin and rifampicin prior to infection of *Nicotiana benthamiana* leaves. The leaves of 6-weeks-old plants are infiltrated with the 1:1 mixture of the *Agrobacterium tumefaciens* cultures, one with the strain carrying the expression system and the other with a strain carrying an expression vector for post transcriptional gene silencing inhibitor p19 (Silhavy et al., 2002) (both cultures diluted to OD600=0.7 with 10mM MgCl2 + 10mM MES – pH=5.8). The infiltrated leaf discs (corresponding to the infiltrated area) are harvested after 6 days incubation in a greenhouse, grinded in 1xPBS, and the crude extracts are analysed for mCherry fluorescence using the Varioskan instrument (Thermo Electron Corporation).

Example 8

Test of the expression system performance in green algae (*Chlamydomonas reinhardtii*)

Two expression systems are tested in *Chlamydomonas reinhardtii* (Table 5): The expression systems assembled in single DNA molecules comprise two expression

cassettes: 1) sTF expression cassette, which comprises a core promoter used for the sTF expression control, exemplified here with the Cr-eIF-5A_cp (Table 1); the sTF version with the DNA-binding protein, exemplified here by either BM3R1 or TetR, and with the activation domain, exemplified here by either VP16AD or VP64AD; and a terminator, exemplified here by the *Chlamydomonas reinhardtii* RPS27A terminator. And 2) the target gene expression cassette, which comprises a number of sTF specific binding sites, exemplified here by either eight BM3R1-specific binding sites or by eight TetR-specific binding sites; another core promoter, exemplified here by Cr-RPS3A_cp (Table 1), the target gene coding region, exemplified here by the mCherry (red fluorescent protein reporter) coding region; and a terminator, exemplified here by the *Chlamydomonas reinhardtii* RBCS2 terminator. The coding regions of the sTFs and the mCherry are codon-optimized to fit the codon usage of *Chlamydomonas reinhardtii*. Also, negative control versions (the expression systems with deleted regions spanning the Cr-eIF-5A_cp, the sTF, and the RPS27A terminator) are constructed.

The expression systems (and the negative control versions) are cloned into the NcoI site of the plasmid pChlamy_4 (Invitrogen). The resulting plasmids (including unmodified pChlamy_4 plasmid), after linearization, are transformed into *Chlamydomonas reinhardtii* (strain 137c; Invitrogen). The transformations are performed according to protocol in the GeneArt Chlamydomonas Protein Expression Kit manual (Invitrogen). The transformants are grown in the Gibco Tap Growth medium in presence of Zeocin and analyzed for mCherry fluorescence using the Varioskan instrument (Thermo Electron Corporation).

Example 9

Test of the expression system performance in CHO cells (*Cricetulus griseus*)

Expression system for *Cricetulus griseus* (Table 6) assembled in single DNA molecules comprises two expression cassettes: 1) sTF expression cassette, which comprises a core promoter used for the sTF expression control, exemplified here with the Mm-Atp5b_cp (Table 1); the sTF version with the DNA-binding protein, exemplified here by BM3R1, and with the activation domain, exemplified here by VP64AD; and a terminator, exemplified here by the *Mus musculus* INHA terminator. And 2) the target gene expression cassette, which comprises a number of sTF specific binding sites, exemplified here by either eight BM3R1-specific binding sites; another core promoter, exemplified here by Mm-Eef2_cp (Table 1), the tar-

get gene coding region, exemplified here by the mCherry (red fluorescent protein reporter) coding region; and a terminator, exemplified here by the *Mus musculus* FTH1 terminator. The coding regions of sTF and mCherry are codon-optimized to fit the codon usage of *Cricetulus griseus*. Also, negative control versions (the expression systems with deleted regions spanning the Mm-Atp5b_cp, the sTF, and the INHA terminator) are constructed.

The expression system (and the negative control version) is cloned between MluI and XbaI sites of the plasmid pcDNA3.1 (Invitrogen). The resulting plasmids are transfected into Chinese hamster ovary cells (CHO-K1; American Type Culture Collection (Rockville, MD)). Prior to the transformation, the CHO-K1 cells are cultured in Ham's F-12K (Kaighn's) Medium (Gibco) containing 2 mM L-glutamine and 1500 mg/L sodium bicarbonate, supplemented with 10% fetal bovine serum (FBS), 100 U/ml of penicillin, and 100 mg/mL of streptomycin. The cells are maintained in an atmosphere of 5% CO_2 and 90% relative humidity at 37 °C. A flask of cells are cultured, split, and $3x10^5$ cells are seeded into 6-well culture plates and grown in 2 ml of medium until 70% confluent. The transfection is done with FuGene 6 (Roche) according to the manufacturer's instructions with approximately 1 µg of plasmid DNA added per well. The transfected cells are allowed to continue growing for up to 5 days. The mCherry expression is monitored daily post transfection by fluorescence microscopy.

Literature references:

1. Hubmann G, Thevelein J, Nevoigt E (2014) Natural and Modified Promoters for Tailored Metabolic Engineering of the Yeast *Saccharomyces cerevisiae*. In: Mapelli V, editor. Yeast Metabolic Engineering: Springer New York. pp. 17-42.

2. Blumhoff M, Steiger MG, Marx H, Mattanovich D, Sauer M (2013) Six novel constitutive promoters for metabolic engineering of Aspergillus niger. Appl Microbiol Biotechnol 97(1):259-67.

3. Ito Y, Yamanishi M, Ikeuchi A, Matsuyama T (2015) A highly tunable system for the simultaneous expression of multiple enzymes in *Saccharomyces cerevisiae*. ACS Synth Biol 4: 12-16.

4. Pachlinger R, Mitterbauer R, Adam G, Strauss J (2005) Metabolically independent and accurately adjustable Aspergillus sp. expression system. Appl Environ Microbiol 71: 672-678.

5. Silhavy D, Molnar A, Lucioli A, Szittya G, Hornyik C, Tavazza M, Burgyan J. (2002) A viral protein suppresses RNA silencing and binds silencing-generated, 21- to 25-nucleotide double-stranded RNAs. EMBO J. 21(12):3070-80.

Patent references

US2002081667

Claims

1. An expression system for a eukaryotic host, which comprises

(a) an expression cassette comprising

a core promoter and a DNA sequence encoding synthetic transcription factor (sTF), said core promoter being the only regulatory sequence for controlling expression of the DNA sequence encoding the sTF, and

(b) one or more expression cassettes, each comprising a DNA sequence encoding a desired product and being operably linked to a synthetic promoter,

said synthetic promoter comprising the core promoter identified in (a), and sTF-specific binding sites upstream of the core promoter identified in (a) or upstream of the other core promoter different from the core promoter identified in (a).

2. The expression system according to claim 1, wherein the core promoter comprises a DNA sequence containing the 5'-upstream region of a eukaryotic gene, starting 10 – 50 bp upstream of a TATA-box and ending 9 bp upstream of the ATG start codon, and wherein the distance between the TATA-box and the start codon is no greater than 180 bp and no smaller than 80 bp, and a DNA sequence at its 3'-end comprising random 1-20 bp; or the core promoter comprises a DNA sequence having at least 90% sequence identity to said 5'-upstream region, and a DNA sequence comprising random 1-20 bp which is located at the 3'-end of the DNA sequence.

3. The expression system according to claim 1 or 2, wherein the core promoter is a universal core promoter (UCP) functional in diverse eukaryotic organisms.

4. The expression system according to any one of claims 1-3, wherein said synthetic transcription factor (sTF) comprises a prokaryotic transcription regulator, a nuclear localization signal, and a transcription activation domain.

5. The expression system according to any one of claims 1-4, wherein the eukaryotic host is selected from the group consisting of fungal species, plant species, and animal species.

Date Reçue/Date Received 2023-04-17

6. The expression system according to claim 5, wherein the fungal species is yeast or filamentous fungi.

7. The expression system according to claim 5 or 6, wherein the plant species is flowering plants or green algae species.

8. A eukaryotic cell comprising the expression system of any one of claims 1 to 7.

9. The eukaryotic cell according to claim 8, wherein the core promoter is a universal core promoter (UCP) functional in diverse eukaryotic organisms and originates from another eukaryotic host species than *Saccharomyces cerevisiae*.

10. The eukaryotic cell according to claim 8 or 9, wherein the eukaryotic host is selected from the group consisting of fungal species, plant species, and animal species.

11. The eukaryotic cell according to claim 10, wherein the fungal species is yeast or filamentous fungi.

12. The eukaryotic cell according to claim 10 or 11, wherein the plant species is a flowering plant or a green algae species.

13. An *in vitro* method for producing a desired protein product in a eukaryotic cell comprising cultivating the cell according to claim 8 or 9 under suitable cultivation conditions.

14. Use of the eukaryotic cell of any one of claims 8 to 12 to produce a desired protein product.

Date Reçue/Date Received 2023-04-17

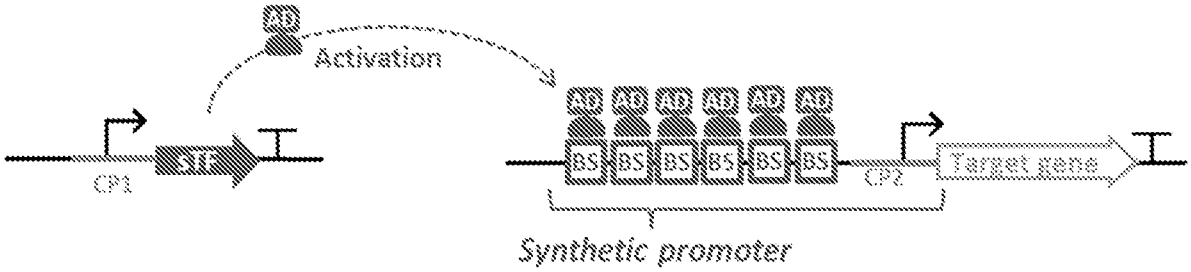


Fig. 1

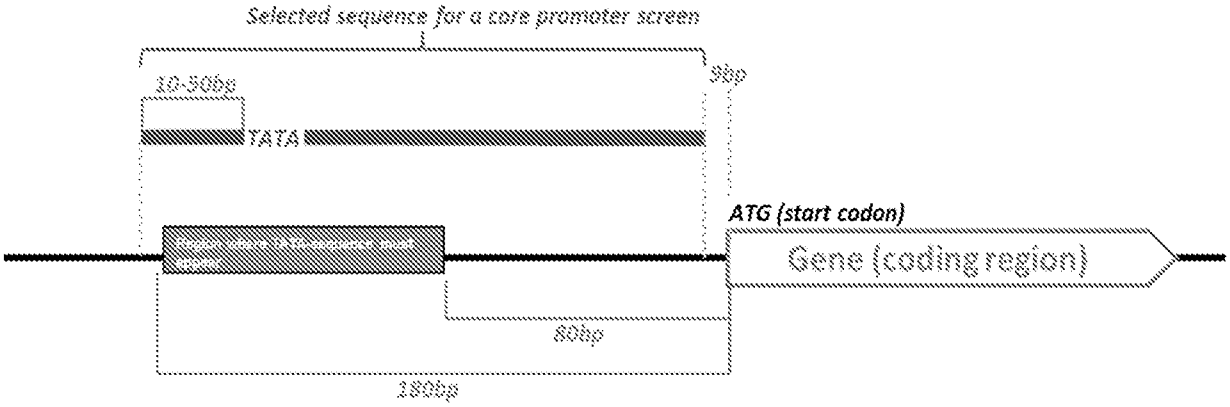


Fig. 2A

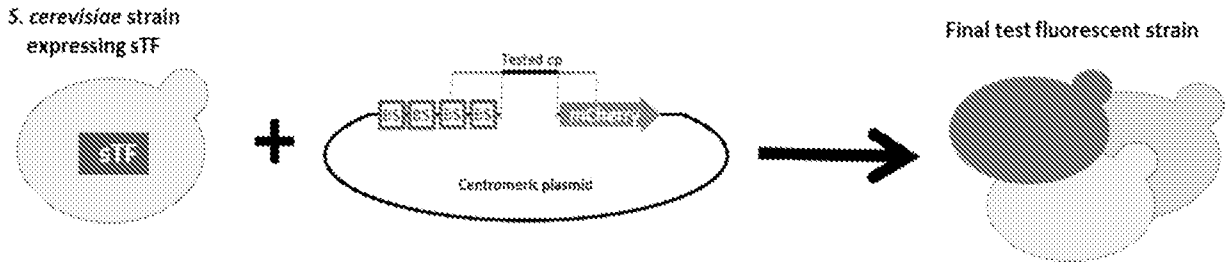


Fig. 2B

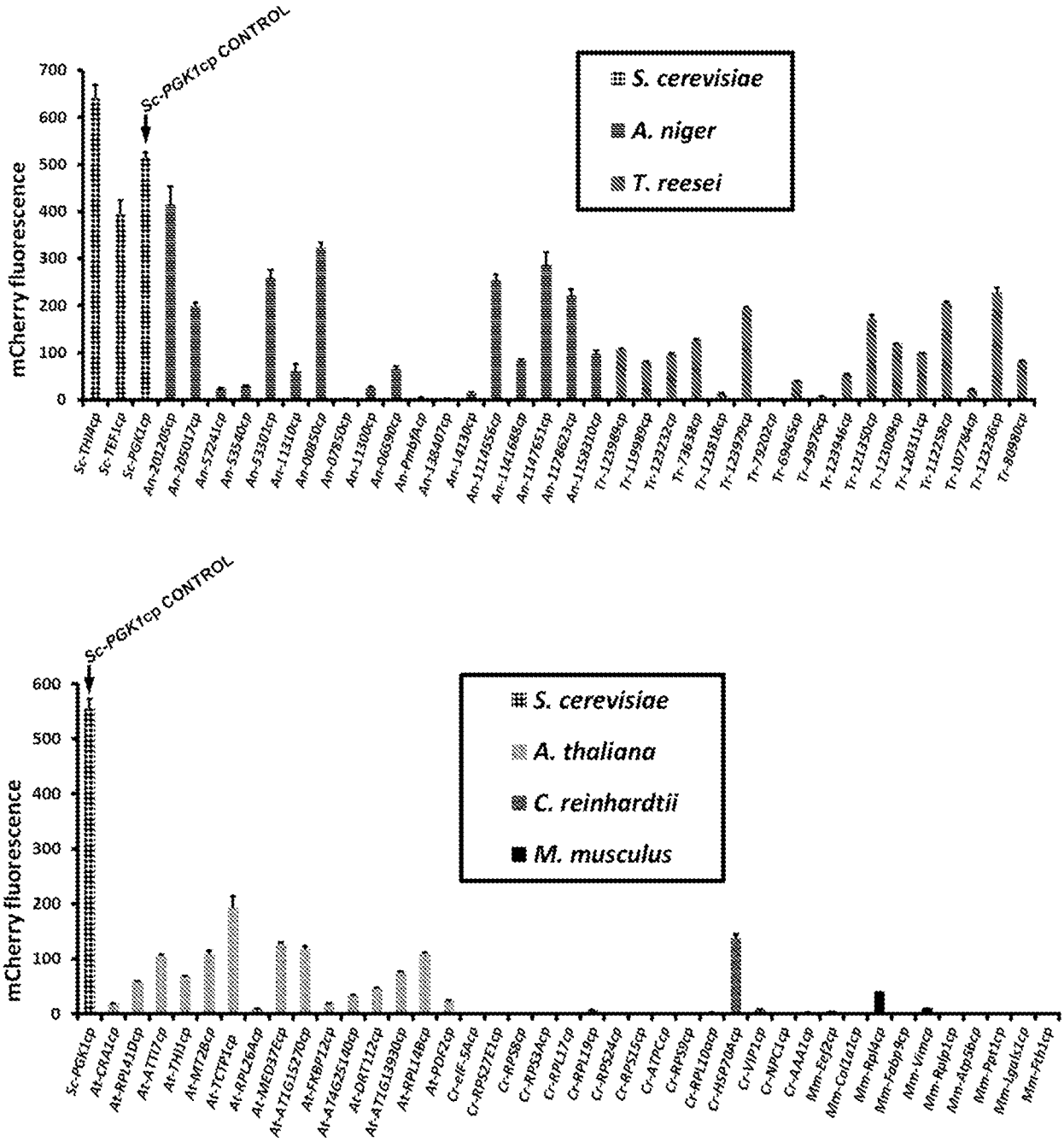

mCherry fluorescence
Sc-PGK1cp CONTROL
S. cerevisiae
A. niger
T. reesei
A. thaliana
C. reinhardtii
M. musculus

Fig. 2C

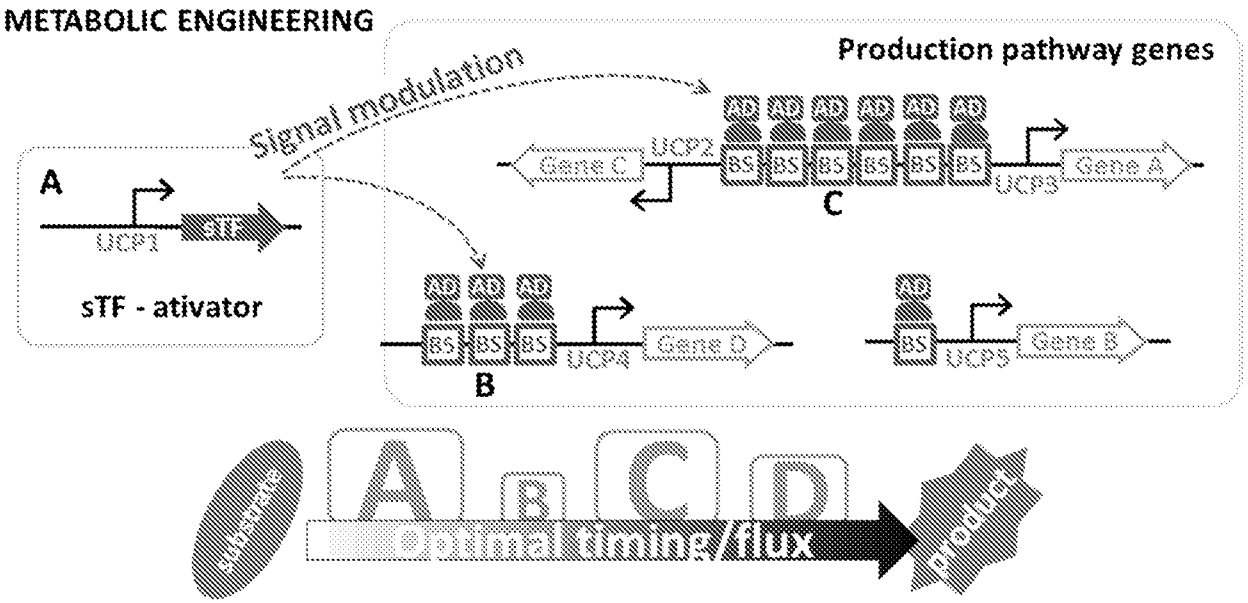

METABOLIC ENGINEERING
Production pathway genes
Signal modulation
A
UCP1
sTF
sTF - ativator
AD
BS
Gene C
UCP2
C
UCP3
Gene A
B
UCP4
Gene D
UCP5
Gene B
A
B
C
D
Optimal timing/flux
Product

Fig. 3

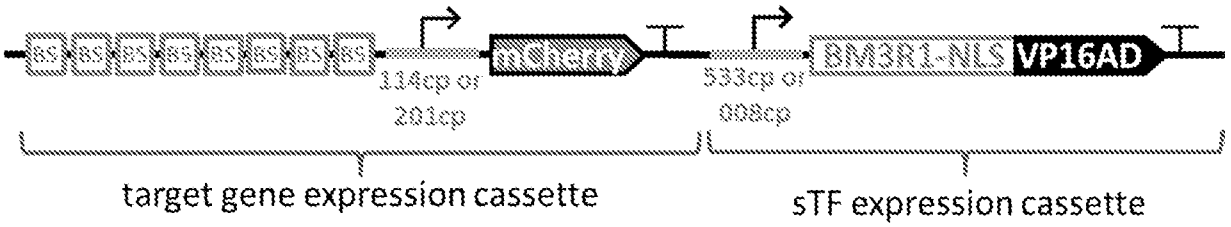
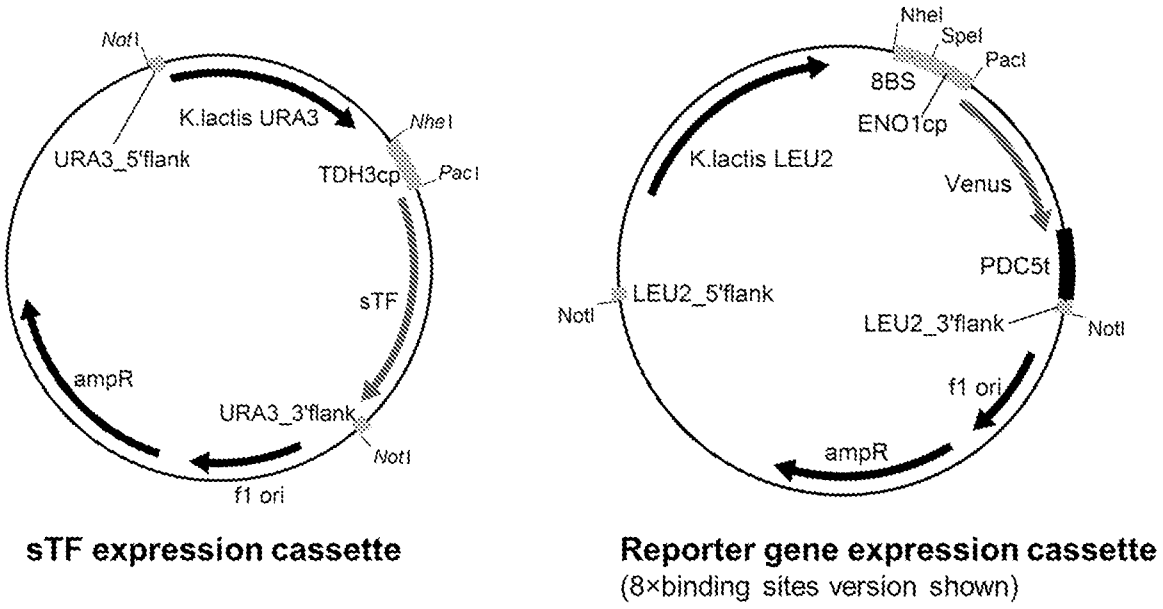

BS
114cp or
201cp
mCherry
533cp or
808cp
BM3R1-NLS
VP16AD
target gene expression cassette
sTF expression cassette

Fig. 4



sTF expression cassette

Reporter gene expression cassette
(8×binding sites version shown)

Fig. 5A

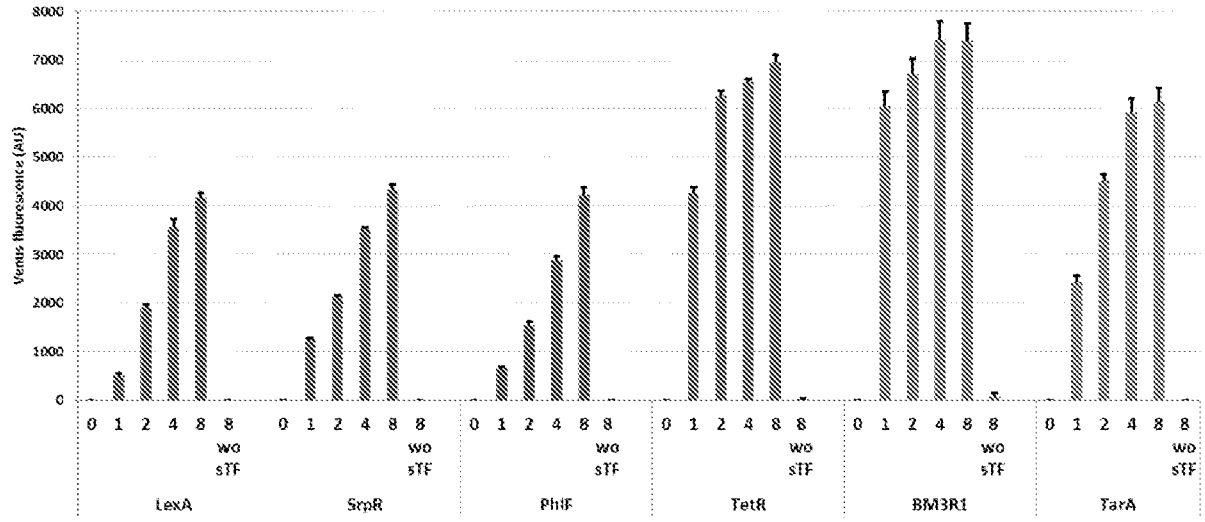


Fig. 5B

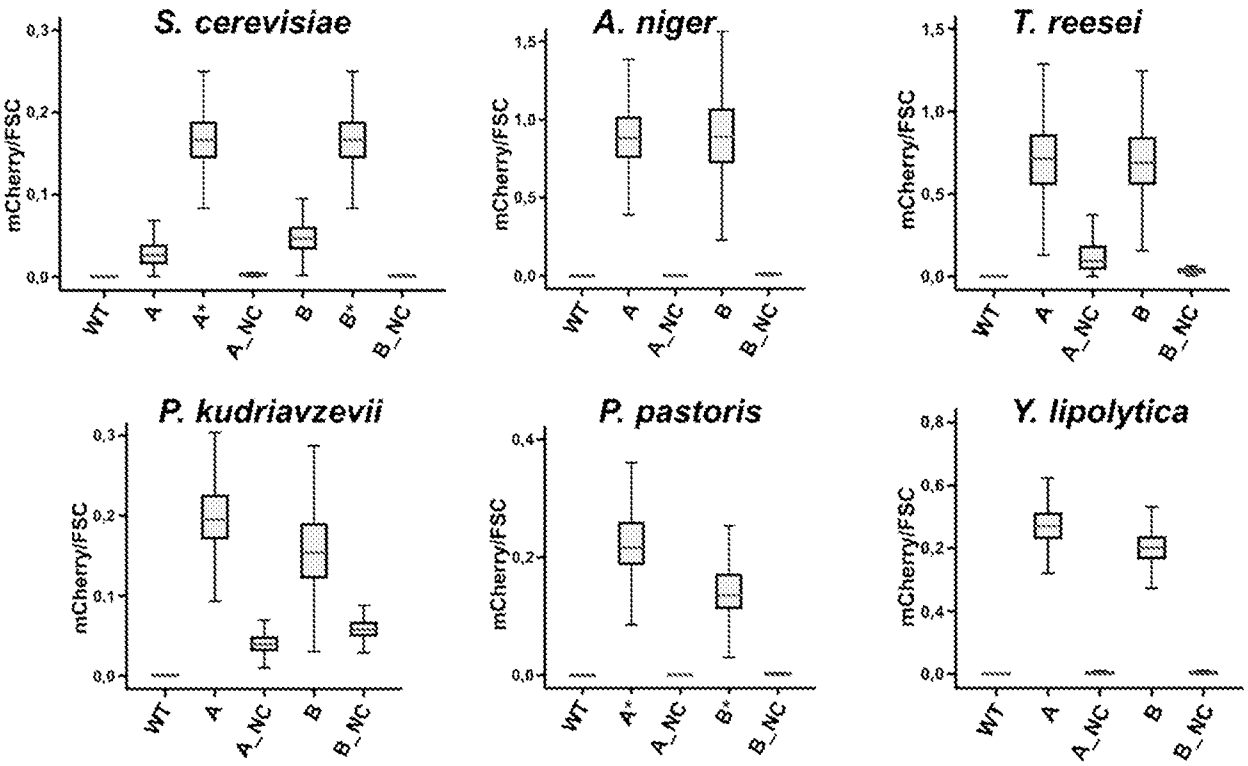

S. cerevisiae
A. niger
T. reesei
P. kudriavzevii
P. pastoris
Y. lipolytica
mCherry/FSC
WT
A
A*
A_NC
B
B*
B_NC

Fig. 6A

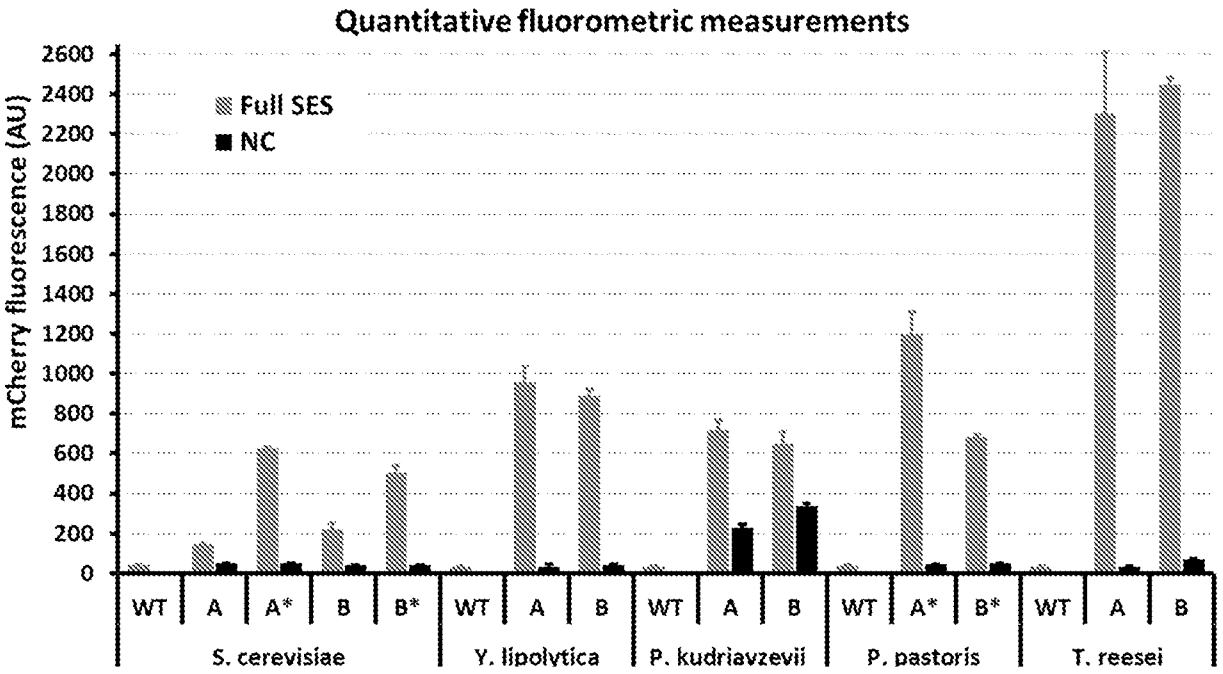

Quantitative fluorometric measurements
mCherry fluorescence (AU)
Full SES
NC
WT
A
A*
B
B*
S. cerevisiae
Y. lipolytica
P. kudriavzevii
P. pastoris
T. reesei

Fig. 6B

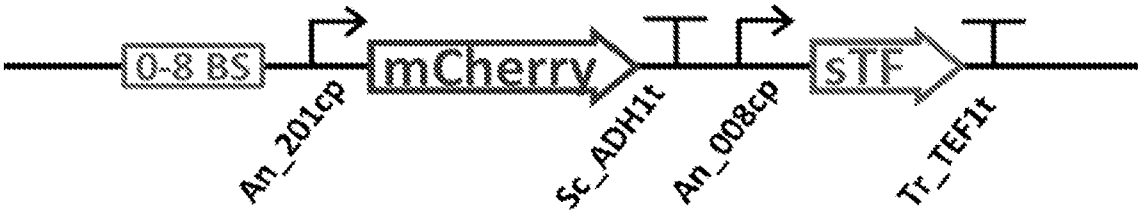

0-8 BS
An_201cp
mCherry
Sc_ADH1t
An_008cp
sTF
Tr_TEF1t

Fig. 7A

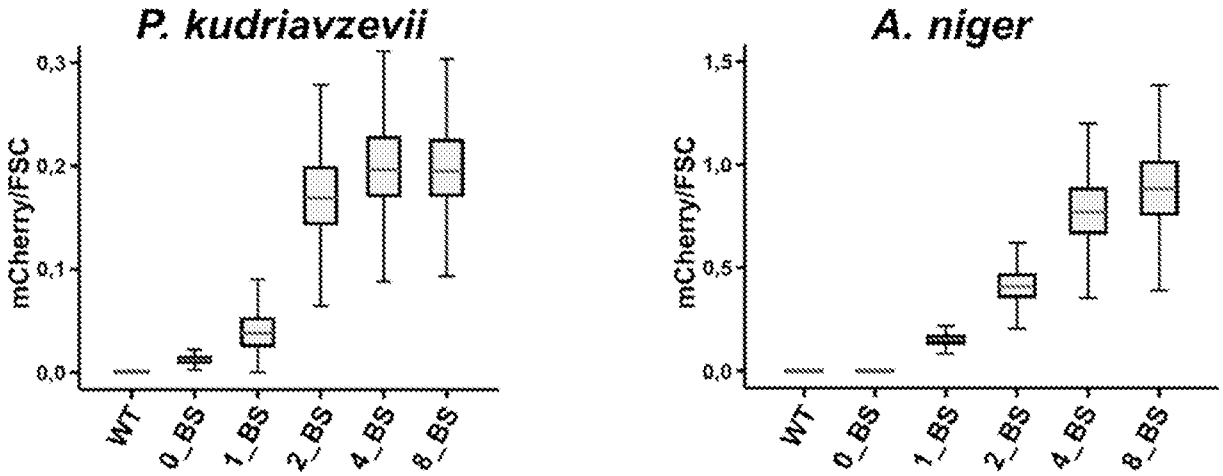

P. kudriavzevii
A. niger
mCherry/FSC
0,0
0,1
0,2
0,3
0,5
1,0
1,5
WT
0_BS
1_BS
2_BS
4_BS
8_BS

Fig. 7B

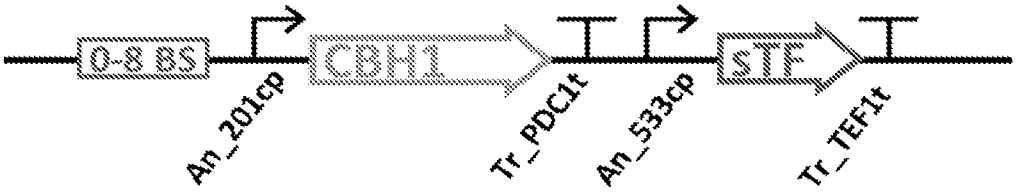


Fig. 7C

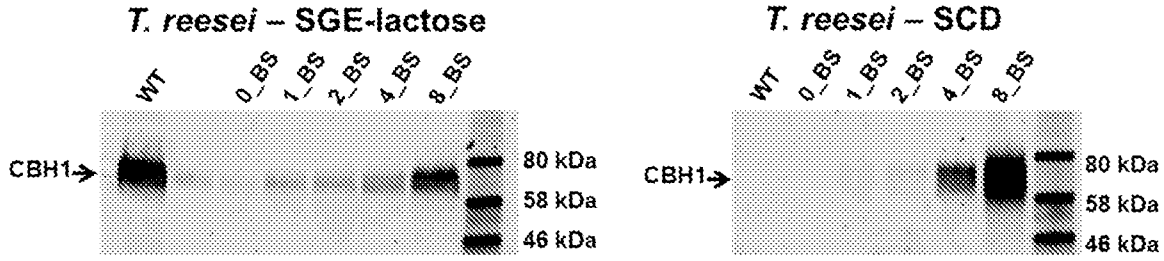


Fig. 7D

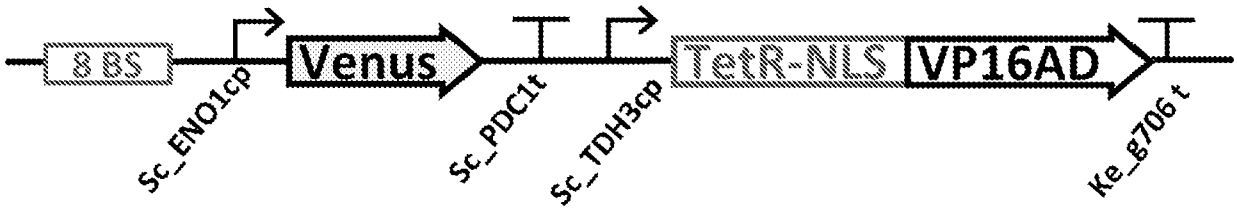


Fig. 8A

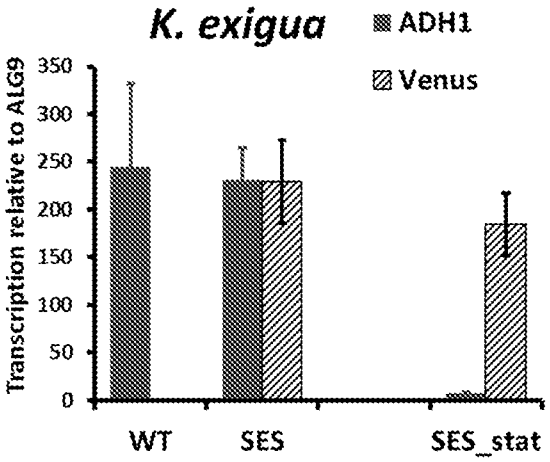


Fig. 8B

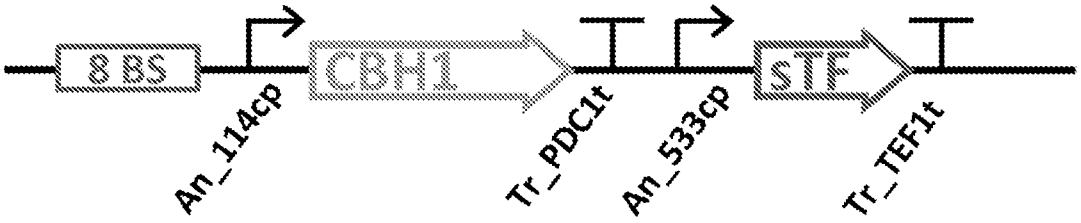
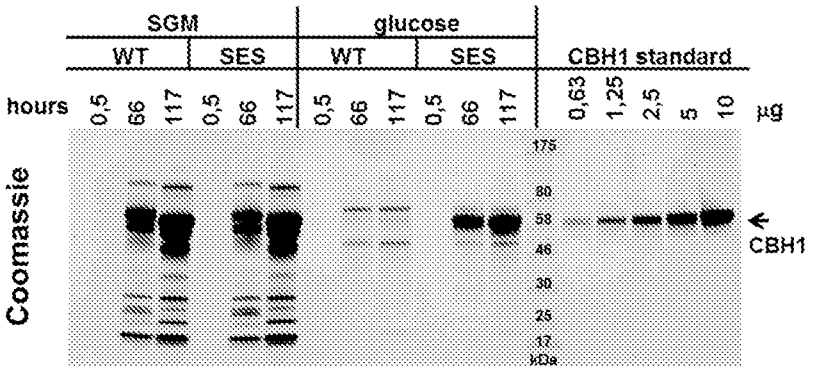
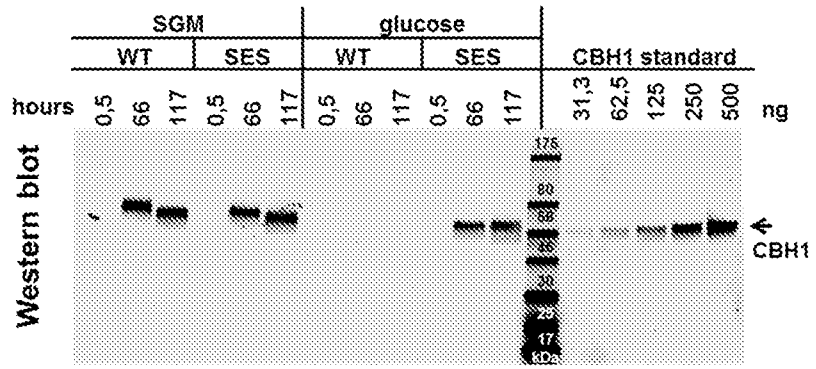


Fig. 9A



SGM	WT	0,5 h	0,6 g/L
		66 h	9,4 g/L
		117 h	17,8 g/L
	SES	0,5 h	0,6 g/L
		66 h	8,4 g/L
		117 h	17,6 g/L
Glucose	WT	0,5 h	<0,5 g/L
		66 h	0,9 g/L
		117 h	1,2 g/L
	SES	0,5 h	<0,5 g/L
		66 h	3,7 g/L
		117 h	5,9 g/L



SGM	WT	0,5 h	<0,1 g/L
		66 h	4,9 g/L
		117 h	4,8 g/L
	SES	0,5 h	<0,1 g/L
		66 h	4,3 g/L
		117 h	5,4 g/L
Glucose	WT	0,5 h	<0,1 g/L
		66 h	<0,1 g/L
		117 h	<0,1 g/L
	SES	0,5 h	<0,1 g/L
		66 h	3,2 g/L
		117 h	4,1 g/L

Fig. 9B

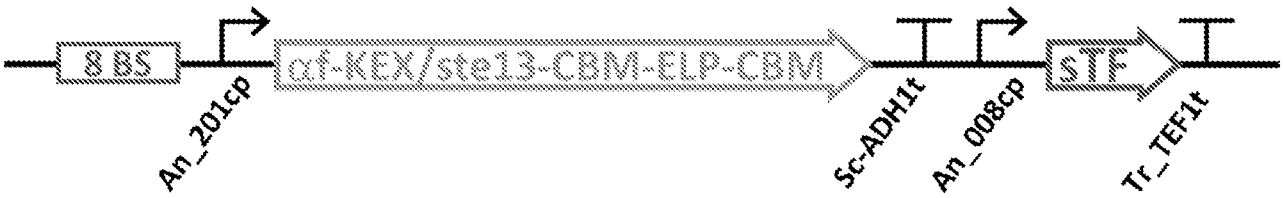

8 BS
An_201cp
αf-KEX/Ste13-CBM-ELP-CBM
Sc-ADH1t
An_008cp
STF
Tr_TEF1t

Fig. 9C

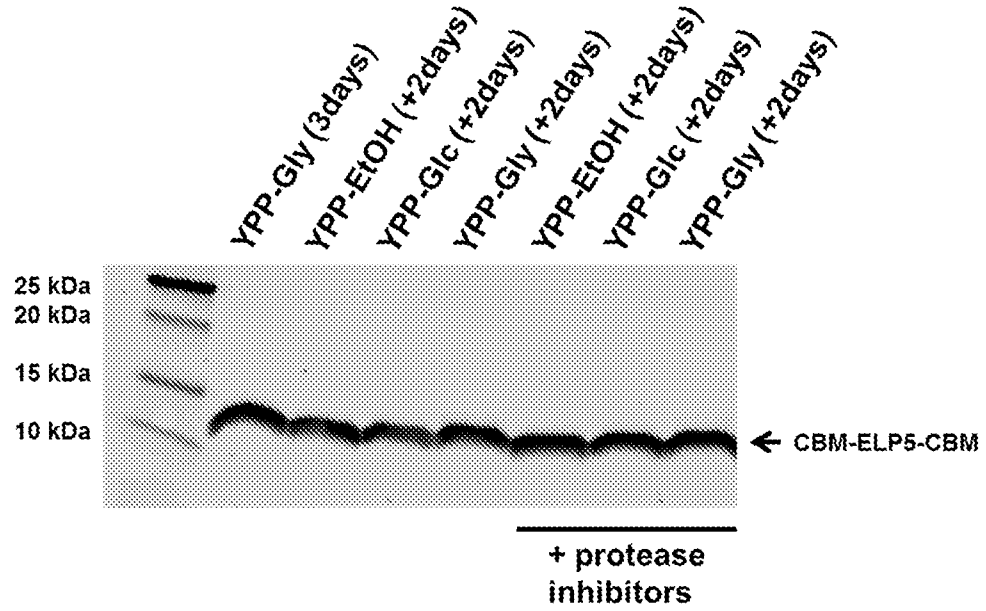
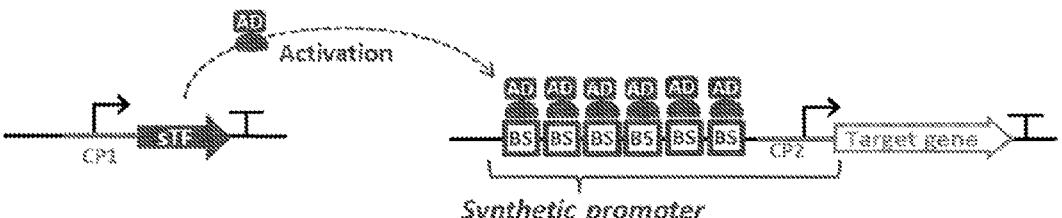

YPP-Gly (3days)
YPP-EtOH (+2days)
YPP-Glc (+2days)
YPP-Gly (+2days)
YPP-EtOH (+2days)
YPP-Glc (+2days)
YPP-Gly (+2days)
25 kDa
20 kDa
15 kDa
10 kDa
CBM-ELP5-CBM
+ protease inhibitors

Fig. 9D


AD
Activation
CP1
ST
AD AD AD AD AD AD
BS BS BS BS BS BS
CP2
Target gene
Synthetic promoter