(51) International Patent Classification:
C12N 15/70 (2006.01) C12P 19/02 (2006.01)

(21) International Application Number:
PCT/US2017/039997

(22) International Filing Date:
29 June 2017 (29.06.2017)

(25) Filling Language:
English

(26) Publication Language:
English

(30) Priority Data:
62/356,924 30 June 2016 (30.06.2016) US


(72) Inventors: MANCHESTER, Shawn; 278 38th Street, Oakland, California 94611 (US). MELLIN, Jeffrey; 3878 Whittle Avenue, Oakland, California 94602 (US).

(54) Title: METHODS FOR GENERATING A GLUCOSE PERMEASE LIBRARY AND USES THEREOF

(57) Abstract: The present disclosure describes methods for generating microbial strains expressing a heterologous bacterial glucose permease gene that produce biomolecules of interest. In aspects, the disclosure provides novel bacterial strains, which express a heterologous bacterial glucose permease gene whose expression is controlled by a native Corynebacterium glutamicum promoter or a mutant promoter derived therefrom. Also provided herein are methods for producing a library of bacterial glucose permease genes using a promoter ladder comprising a plurality of promoters derived from Corynebacterium glutamicum.

FIG. 1

![Image of a diagram illustrating the performance of glucose permeases in evaluation methods]
Published:

— with international search report (Art. 21(3))
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
— with sequence listing part of description (Rule 5.2(a))
METHODS FOR GENERATING A GLUCOSE PERMEASE LIBRARY AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of priority from U.S. Provisional Application Serial No. 62/356,924, filed June 30, 2016, which is hereby incorporated by reference in its entirety for all purposes.

STATEMENT REGARDING SEQUENCE LISTING

[0002] The Sequence Listing associated with this application is provided in text format in lieu of a paper copy, and is hereby incorporated by reference into the specification. The name of the text file containing the Sequence Listing is ZYMR_005_01WO_SeqList_ST25.txt. The text file is 49 KB, was created on June 28, 2017, and is being submitted electronically via EFS-Web.

FIELD

[0003] The present disclosure is directed to microbial genomic engineering. The disclosed genomic engineering method entails the generation of a library of glucose permease genes and/or glucokinase genes and introducing said library into microbial hosts in order to produce strains with a desired phenotype (e.g. microbial production of commercial products).

BACKGROUND

[0004] Glucose transport in some microorganisms such as, for example, Corynebacterium glutamicum is natively accomplished using the phosphotransferase transport system (PTS). In this system, phosphorylation of glucose is carried out simultaneously to transport. The phospho donor is phosphoenolpyruvate (PEP), therefore linking transport directly to glycolytic flux. In addition, the PTS system is natively regulated by a number of transcriptional processes in ways that are not always ideal for the production of commercial products.

[0005] Microbial processes for the production of various commercial products from glucose strive to maximize the efficiency with which the carbon skeleton of glucose is converted into the desired product. Control of glucose flux is critical for the production of products in ways that are dependent
on the fermentation process, strain of microbial host being used (e.g., C. glutamicum), and small molecule being produced. If there is too much flux through glycolysis under high concentrations of glucose, glycolytic by-products (usually organic acids) are produced which decrease yield of product. If there is too little transport of glucose into the cell, then it is difficult to produce product at high rates. The genotypes of strains which are engineered in various ways to produce specific products interact with process conditions to lead to situations in which more or less glucose transport occurs than would be ideal to maximize yield or productivity.

[0006] Microbial strain improvement has been attempted by the expression of different glucose permeases and glucokinases which may alter glucose transport in such a way as to increase yield or productivity of commercial products. This has been demonstrated in a number of cases. For example, deletion of the native PTS system for glucose transport and overexpression of a native C. glutamicum permease along with a native C. glutamicum kinase led to the increased yield of lysine production from glucose (see Linder et al. Appl. Environ. Microbiol. June 2011 vol. 77, no. 11 pp 3571-3581, the contents of which are hereby incorporated by reference in their entirety). In another example, overexpression of the glucose permease and glucokinase from Z. mobilis in C. glutamicum was used for the production of small molecules (see US5602030, the contents of which are hereby incorporated by reference in its entirety).

[0007] However, the selection of a particular glucose permease to create the ideal level of glucose transport for a given metabolic process to produce a specific commercial product relies on a good understanding of a number of interacting factors, including the interaction of the genotype of a strain with the process environment in which fermentation takes place. Further, the correct expression, affinity, and transport rate, in combination with glucose and other carbon source concentrations may be required to deliver a balanced flux of glucose into the cell to match the flux through the pathway of interest. Understanding these parameters a priori and then choosing a single permease which embodies them can be difficult or impossible.

[0008] Thus, there is a great need in the art for new methods of engineering industrial microbes for producing specific commercial products, which do not suffer from the aforementioned drawbacks inherent with traditional strain improvement programs.

SUMMARY OF THE DISCLOSURE
[0009] In one aspect, provided herein is a host cell comprising a heterologous glucose permease gene functionally linked to a first promoter polynucleotide, wherein the first promoter polynucleotide comprises a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. In some cases, the glucose permease gene is a bacterial glucose permease gene. In some cases, the bacterial glucose permease gene is a gene that encodes a polypeptide sequence selected from SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9 and SEQ ID NO: 14. In some cases, the bacterial glucose permease gene is a gene with a nucleotide sequence selected from SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19 and SEQ ID NO: 24. In some cases, the host cell further comprises a hexokinase gene functionally linked to a second promoter polynucleotide, wherein the second promoter polynucleotide comprises a nucleotide sequence selected from SEQ ID NO: 2, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. In some cases, the hexokinase gene is a glucokinase gene. In some cases, the glucokinase gene is a bacterial glucokinase gene. In some cases, the bacterial glucokinase gene is a gene that encodes a polypeptide with an amino acid sequence selected from SEQ ID NO: 15 and SEQ ID NO: 16. In some cases, the bacterial glucokinase gene is a gene with a nucleotide sequence selected from SEQ ID NO: 25 and SEQ ID NO: 26. In some cases, the first promoter polynucleotide and the second promoter polynucleotide are different. In some cases, the first promoter polynucleotide and the second promoter polynucleotide are identical. In some cases, the host cell belongs to the genus Corynebacterium. In some cases, the host cell is Corynebacterium glutamicum. In some cases, the host cell is used in a method of producing a biomolecule from glucose comprising culturing the host cell under conditions suitable for producing the biomolecule. In some cases, the biomolecule is a small molecule, a nucleotide, an amino acid, an organic acid, or an alcohol. In some cases, the amino acid is tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine. In some cases, the organic acid is succinate, lactate or pyruvate. In some cases, the alcohol is ethanol or isobutanol.

[0010] In another aspect, provided herein is a method for generating a microorganism capable of increased production of a biomolecule from glucose, the method comprising: a) genetically modifying a host microorganism, wherein the modifying comprises introducing a glucose permease gene from a library of glucose permease genes into the genome of the host.
microorganism, wherein each glucose permease gene from the library of glucose permease genes is functionally linked to a promoter comprising a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8 and wherein the modification generates a strain of the host microorganism expressing the glucose permease gene; b) repeating step a) for a plurality of rounds until a plurality of strains of the host microorganism are generated, wherein each strain of the plurality of strains of the host microorganism expresses a separate glucose permease gene from the library of glucose permease genes; c) contacting each strain of the plurality of strains of the host microorganism with a carbon source comprising glucose under fermentative conditions; and d) selecting each strain of the host microorganism that produces an increased amount of a biomolecule from glucose as compared to the amount of the biomolecule produce from glucose from a control microorganism, wherein the control microorganism does not express a glucose permease gene from the library of glucose permease genes. In some cases, each of the glucose permease genes in the library of glucose permease genes is a bacterial glucose permease gene. In some cases, the library of bacterial glucose permease genes comprises genes that encode polypeptide sequences of SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9, SEQ ID NO: 14 or a combination thereof. In some cases, the library of bacterial glucose permease genes comprises genes with a nucleotide sequence of SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19, SEQ ID NO: 24 or a combination thereof. In some cases, the method further comprises introducing a hexokinase gene from a library of hexokinase genes, wherein each hexokinase gene from the library of hexokinase genes is functionally linked to a promoter polynucleotide, wherein the promoter polynucleotide comprises a sequence selected from SEQ ID NO: 2, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. In some cases, the introduction of each hexokinase gene from the library of hexokinase genes is concurrent with the introduction of each glucose permease gene from the library of glucose permease genes. In some cases, each hexokinase gene from the library of hexokinase genes is present in a chimeric construct comprising a glucose permease gene from the library of glucose permease genes. In some cases, the hexokinase gene is a glucokinase gene. In some cases, the glucokinase gene is a bacterial glucokinase gene. In some cases, the library of bacterial glucokinase genes comprises genes that encode polypeptide sequences of SEQ ID NO: 15 and/or SEQ ID NO: 16. In some cases, the library of bacterial glucokinase genes comprises
genes with nucleotide sequences of SEQ ID NO: 25 and/or SEQ ID NO: 26. In some cases, the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are different. In some cases, the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are identical. In some cases, the host microorganism belongs to the genus Corynebacterium. In some cases, the host microorganism is Corynebacterium glutamicum. In some cases, the introducing is performed by transformation, transduction or electroporation. In some cases, the biomolecule is a small molecule, an amino acid, a nucleotide, an organic acid, or an alcohol. In some cases, the amino acid is tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine. In some cases, the organic acid is succinate, lactate or pyruvate. In some cases, the alcohol is ethanol or isobutanol.

[0011] In yet another aspect, provided herein is a library of glucose permease genes, wherein each glucose permease gene in the library of glucose permease genes is functionally linked to a promoter comprising a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. In some cases, each glucose permease gene is a bacterial glucose permease gene. In some cases, the library of bacterial glucose permease genes comprises genes that encode polypeptide sequences of SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9, SEQ ID NO: 14 or a combination thereof. In some cases, the library of bacterial glucose permease genes comprises genes with nucleotide sequences of SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19, SEQ ID NO: 24 or a combination thereof. In some cases, each glucose permease gene in the library of glucose permease genes is a first portion of a chimeric construct, wherein the chimeric construct comprises a second portion, wherein the second portion is a hexokinase gene. In some cases, the hexokinase gene is functionally linked to a promoter polynucleotide, wherein the promoter polynucleotide comprises a sequence selected from SEQ ID NO: 2, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. In some cases, the hexokinase gene is a glucokinase gene. In some cases, the glucokinase gene is a bacterial glucokinase gene. In some cases, the library of bacterial glucokinase genes comprises genes that encode polypeptide sequences of SEQ ID NO: 15 and/or SEQ ID NO: 16. In some cases, the library of bacterial glucokinase genes comprises genes with
nucleotides sequences of SEQ ID NO: 25 and/or SEQ ID NO: 26. In some cases, the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are different. In some cases, the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are identical. In some cases, the library is used in a method of producing a biomolecule comprising introducing a glucose permease gene from the library into a host cell and culturing the host cell under conditions suitable for producing the biomolecule. In some cases, the biomolecule is an amino acid, a nucleotide, an organic acid, or an alcohol. In some cases, the amino acid is tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine. In some cases, the organic acid is succinate, lactate or pyruvate. In some cases, the alcohol is ethanol or isobutanol. In some cases, the host cell belongs to the genus Corynebacterium. In some cases, the host cell is Corynebacterium glutamicum. In some cases, the introducing is performed by transformation, transduction or electroporation.

[0012] In another aspect, provided herein is an isolated, synthetic or recombinant polynucleotide comprising a codon optimized polynucleotide selected from SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19 and SEQ ID NO: 24, wherein the polynucleotide is codon optimized for expression in a host cell. In some cases, the host cell is E. coli and/or C. glutamicum.

[0013] In a further aspect, provided herein is an isolated, synthetic or recombinant polynucleotide comprising a codon optimized polynucleotide selected from SEQ ID NO: 25 and SEQ ID NO: 26, wherein the polynucleotide is codon optimized for expression in a host cell. In some cases, the host cell is E. coli and/or C. glutamicum.

[0014] In yet another aspect, provided herein is an isolated, synthetic or recombinant polynucleotide comprising a first codon optimized polynucleotide and a second codon optimized polynucleotide, wherein the first polynucleotide and the second polynucleotide are each codon optimized for expression in a host cell, and wherein the first codon optimized polynucleotide encodes a polypeptide with glucose permease activity and the second codon optimized polynucleotide encodes a polypeptide with glucokinase activity. In some cases, the first codon optimized polynucleotide is selected from SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19 and SEQ ID NO: 24. In some cases, the second codon optimized
polynucleotide is selected from SEQ ID NO: 25 and SEQ ID NO: 26. In some cases, the polypeptide with glucose permease activity comprises a sequence selected from SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9 and SEQ ID NO: 14. In some cases, the polypeptide with glucokinase activity comprises a sequence selected from SEQ ID NO: 15 and SEQ ID NO: 16. In some cases, the host cell is *E. coli* and/or *C. glutamicum*.

**BRIEF DESCRIPTION OF THE FIGURES**

**[0015]** FIG. 1 illustrates performance of glucose permeases in evaluation method as described in Example 1.

**[0016]** FIG. 2 illustrates performance of glucose permeases in desired fermentation conditions as described in Example 1.

**[0017]** FIG. 3 illustrates assembly of transformation plasmids of the present disclosure, and their integration into host organisms. The insert sequence insert DNA is generated by combining one or more synthesized oligonucleotides in an assembly reaction. DNA inserts contain desired promoter sequence flanked by direct repeat region (i.e., homology arms) designed for looping out DNA in subsequent steps. Assembled plasmids contain the insert DNA (permease gene and/or glucokinase gene functionally linked to promoters provided herein), and optionally, one or more selection markers.

**[0018]** FIG. 4 illustrates a procedure for looping-out selected regions of DNA from host strains. Direct repeat (DR) regions of the inserted DNA form a loop with corresponding sequences in the host strain’s genome. Cells counter selected for selection marker exhibit DNA deletion of loop DNA.

**DETAILED DESCRIPTION**

**Definitions**

**[0019]** While the following terms are believed to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the presently disclosed subject matter.

**[0020]** The term “a” or “an” refers to one or more of that entity, *i.e.* can refer to a plural referents. As such, the terms “a” or “an”, “one or more” and “at least one” are used interchangeably herein. In addition, reference to “an element” by the indefinite article “a” or “an” does not exclude the
possibility that more than one of the elements is present, unless the context clearly requires that there is one and only one of the elements.

[0021] Unless the context requires otherwise, throughout the present specification and claims, the word “comprise” and variations thereof, such as, “comprises” and “comprising” are to be construed in an open, inclusive sense that is as “including, but not limited to”.

[0022] Reference throughout this specification to “one embodiment” or “an embodiment” means that a particular feature, structure or characteristic described in connection with the embodiment may be included in at least one embodiment of the present disclosure. Thus, the appearances of the phrases “in one embodiment” or “in an embodiment” in various places throughout this specification may not necessarily all refer to the same embodiment. It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination.

[0023] As used herein the terms “cellular organism” “microorganism” or “microbe” should be taken broadly. These terms can be used interchangeably and include, but may not be limited to, the two prokaryotic domains, Bacteria and Archaea, as well as certain eukaryotic fungi and protists. In some embodiments, the disclosure refers to the “microorganisms” or “cellular organisms” or “microbes” of lists/tables and figures present in the disclosure. This characterization can refer to not only the identified taxonomic genera of the tables and figures, but also the identified taxonomic species, as well as the various novel and newly identified or designed strains of any organism in said tables or figures. The same characterization holds true for the recitation of these terms in other parts of the Specification, such as in the Examples.

[0024] The term “prokaryotes” is art recognized and refers to cells which contain no nucleus or other cell organelles. The prokaryotes are generally classified in one of two domains, the Bacteria and the Archaea. The definitive difference between organisms of the Archaea and Bacteria domains is based on fundamental differences in the nucleotide base sequence in the 16S ribosomal RNA.

[0025] The term “Archaea” refers to a categorization of organisms of the division Mendosicutes, typically found in unusual environments and distinguished from the rest of the prokaryotes by several criteria, including the number of ribosomal proteins and the lack of muramic acid in cell
walls. On the basis of ssrRNA analysis, the Archaea consist of two phylogenetically-distinct
groups: Crenarchaeota and Euryarchaeota. On the basis of their physiology, the Archaea can be
organized into three types: methanogens (prokaryotes that produce methane); extreme halophiles
(prokaryotes that live at very high concentrations of salt (NaCl), and extreme (hyper) thermophilus
(prokaryotes that live at very high temperatures). Besides the unifying archaean features that
distinguish them from Bacteria (i.e., no murein in cell wall, ester-linked membrane lipids, etc.),
these prokaryotes exhibit unique structural or biochemical attributes which adapt them to their
particular habitats. The Crenarchaeota consists mainly of hyperthermophilic sulfur-dependent
prokaryotes and the Euryarchaeota contains the methanogens and extreme halophiles.

[0062] “Bacteria” or “eubacteria” refers to a domain of prokaryotic organisms. Bacteria include at
least 11 distinct groups as follows: (1) Gram-positive (gram+) bacteria, of which there are two
major subdivisions: (1) high G+C group (Actinomycetes, Mycobacteria, Micrococcus, others) (2)
low G+C group (Bacillus, Clostridia, Lactobacillus, Staphylococci, Streptococci, Mycoplasmas);
(2) Proteobacteria, e.g., Purple photosynthetic+non-photosynthetic Gram-negative bacteria
(includes most “common” Gram-negative bacteria); (3) Cyanobacteria, e.g., oxygenic
phototrophs; (4) Spirochetes and related species; (5) Planctomyces; (6) Bacteroides,
Flavobacteria; (7) Chlamydia; (8) Green sulfur bacteria; (9) Green non-sulfur bacteria (also
anaerobic phototrophs); (10) Radioresistant micrococcii and relatives; (11) Thermotoga and Thermosipho thermophiles.

[0072] A “eukaryote” is any organism whose cells contain a nucleus and other organelles enclosed
within membranes. Eukaryotes belong to the taxon Eukarya or Eukaryota. The defining feature
that sets eukaryotic cells apart from prokaryotic cells (the aforementioned Bacteria and Archaea)
is that they have membrane-bound organelles, especially the nucleus, which contains the genetic
material, and is enclosed by the nuclear envelope.

[0082] The terms “genetically modified microorganism,” “recombinant microorganism,”
“recombinant host cell,” and “recombinant strain” can be used interchangeably herein and can
refer to microorganisms that have been genetically modified. Thus, the terms include a
microorganism (e.g., bacteria, yeast cell, fungal cell, etc.) that has been genetically altered,
modified, or engineered, such that it exhibits an altered, modified, or different genotype and/or
phenotype (e.g., when the genetic modification affects coding nucleic acid sequences of the
microorganism), as compared to the naturally-occurring microorganism from which it was derived.
It is understood that the terms refer not only to the particular recombinant microorganism in question, but also to the progeny or potential progeny of such a microorganism.

[0029] The term “wild-type microorganism” can describe a cell that occurs in nature, *i.e.* a cell that has not been genetically modified.

[0030] The term “genetically engineered” may refer to any manipulation of a microorganism’s genome (*e.g.* by insertion or deletion of nucleic acids).

[0031] As used herein, the term “allele(s)” can mean any of one or more alternative forms of a gene, all of which alleles relate to at least one trait or characteristic. In a diploid cell, the two alleles of a given gene can occupy corresponding loci on a pair of homologous chromosomes. Since the present disclosure, in embodiments, relates to QTLs, *i.e.* genomic regions that may comprise one or more genes or regulatory sequences, it is in some instances more accurate to refer to “haplotype” (*i.e.* an allele of a chromosomal segment) instead of “allele”, however, in those instances, the term “allele” should be understood to comprise the term “haplotype”.

[0032] As used herein, the term “locus” (loci plural) can mean a specific place or places or a site on a chromosome where for example a gene or genetic marker is found.

[0033] As used herein, the term “genetically linked” can refer to two or more traits that are co-inherited at a high rate during breeding such that they are difficult to separate through crossing.

[0034] A “recombination” or “recombination event” as used herein can refer to a chromosomal crossing over or independent assortment. The term “recombinant” can refer to an organism having a new genetic makeup arising as a result of a recombination event.

[0035] As used herein, the term “phenotype” can refer to the observable characteristics of an individual cell, cell culture, organism, or group of organisms which results from the interaction between that individual’s genetic makeup (*i.e.*, genotype) and the environment.

[0036] As used herein, the term “chimeric” or “recombinant” when describing a nucleic acid sequence or a protein sequence can refer to a nucleic acid, or a protein sequence, that links at least two heterologous polynucleotides, or two heterologous polypeptides, into a single macromolecule, or that can re-arrange one or more elements of at least one natural nucleic acid or protein sequence. For example, the term “recombinant” can refer to an artificial combination of two otherwise separated segments of sequence, *e.g.*, by chemical synthesis or by the manipulation of isolated segments of nucleic acids by genetic engineering techniques.
[0037] As used herein, a “synthetic nucleotide sequence” or “synthetic polynucleotide sequence” can be a nucleotide sequence that is not known to occur in nature or that is not naturally occurring. Generally, such a synthetic nucleotide sequence will comprise at least one nucleotide difference when compared to any other naturally occurring nucleotide sequence.

[0038] As used herein, the term “nucleic acid” can refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides, or analogs thereof. This term can refer to the primary structure of the molecule, and thus includes double- and single-stranded DNA, as well as double- and single-stranded RNA. It can also include modified nucleic acids such as methylated and/or capped nucleic acids, nucleic acids containing modified bases, backbone modifications, and the like. The terms “nucleic acid” and “nucleotide sequence” can be used interchangeably.

[0039] As used herein, the term “gene” can refer to any segment of DNA associated with a biological function. Thus, genes can include, but are not limited to, coding sequences and/or the regulatory sequences required for their expression. Genes can also include non-expressed DNA segments that, for example, form recognition sequences for other proteins. Genes can be obtained from a variety of sources, including cloning from a source of interest or synthesizing from known or predicted sequence information, and may include sequences designed to have desired parameters.

[0040] As used herein, the term “homologous” or “homologue” or “ortholog” is known in the art and can refer to related sequences that share a common ancestor or family member and are determined based on the degree of sequence identity. The terms “homology,” “homologous,” “substantially similar” and “corresponding substantially” can be used interchangeably herein. They can refer to nucleic acid fragments wherein changes in one or more nucleotide bases do not affect the ability of the nucleic acid fragment to mediate gene expression or produce a certain phenotype. These terms can also refer to modifications of the nucleic acid fragments of the instant disclosure such as deletion or insertion of one or more nucleotides that do not substantially alter the functional properties of the resulting nucleic acid fragment relative to the initial, unmodified fragment. It is therefore understood, as those skilled in the art will appreciate, that the disclosure can encompass more than the specific exemplary sequences. These terms can describe the relationship between a gene found in one species, subspecies, variety, cultivar or strain and the corresponding or equivalent gene in another species, subspecies, variety, cultivar or strain. For purposes of this disclosure homologous sequences can be compared. “Homologous sequences” or
“homologues” or “orthologs” can be thought, believed, or known to be functionally related. A functional relationship may be indicated in any one of a number of ways, including, but not limited to: (a) degree of sequence identity and/or (b) the same or similar biological function. Preferably, both (a) and (b) are indicated. Homology can be determined using software programs readily available in the art, such as those discussed in Current Protocols in Molecular Biology (F.M. Ausubel et al., eds., 1987) Supplement 30, section 7.718, Table 7.71. Some alignment programs are MacVector (Oxford Molecular Ltd, Oxford, U.K.), ALIGN Plus (Scientific and Educational Software, Pennsylvania) and AlignX (Vector NTI, Invitrogen, Carlsbad, CA). Another alignment program is Sequencher (Gene Codes, Ann Arbor, Michigan), using default parameters.

[0041] As used herein, the term “nucleotide change” can refer to, e.g., nucleotide substitution, deletion, and/or insertion, as is well understood in the art. For example, mutations contain alterations that produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded protein or how the proteins are made.

[0042] As used herein, the term “protein modification” can refer to, e.g., amino acid substitution, amino acid modification, deletion, and/or insertion, as is well understood in the art.

[0043] As used herein, the term “at least a portion” or “fragment” of a nucleic acid or polypeptide can mean a portion having the minimal size characteristics of such sequences, or any larger fragment of the full length molecule, up to and including the full length molecule. A fragment of a polynucleotide of the disclosure may encode a biologically active portion of a genetic regulatory element. A biologically active portion of a genetic regulatory element can be prepared by isolating a portion of one of the polynucleotides of the disclosure that comprises the genetic regulatory element and assessing activity as described herein. Similarly, a portion of a polypeptide may be 4 amino acids, 5 amino acids, 6 amino acids, 7 amino acids, and so on, going up to the full length polypeptide. The length of the portion to be used can depend on the particular application. A portion of a nucleic acid useful as a hybridization probe may be as short as 12 nucleotides; in some embodiments, it is 20 nucleotides. A portion of a polypeptide useful as an epitope may be as short as 4 amino acids. A portion of a polypeptide that performs the function of the full-length polypeptide can generally be longer than 4 amino acids.

[0044] Variant polynucleotides also encompass sequences that can be derived from a mutagenic and recombinogenic procedure such as DNA shuffling. Strategies for such DNA shuffling are known in the art. See, for example, Stemmer (1994) PNAS 91:10747-10751; Stemmer (1994)


[0046] The term “primer” as used herein can refer to an oligonucleotide which is capable of annealing to the amplification target allowing a DNA polymerase to attach, thereby serving as a point of initiation of DNA synthesis when placed under conditions in which synthesis of primer extension product is induced, i.e., in the presence of nucleotides and an agent for polymerization such as DNA polymerase and at a suitable temperature and pH. The (amplification) primer is preferably single stranded for maximum efficiency in amplification. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the agent for polymerization. The exact lengths of the primers will depend on many factors, including temperature and composition (A/T vs. G/C content) of primer. A pair of bi-directional primers consists of one forward and one reverse primer as commonly used in the art of DNA amplification such as in PCR amplification.

[0047] The terms “stringency” or “stringent hybridization conditions” can refer to hybridization conditions that affect the stability of hybrids, e.g., temperature, salt concentration, pH, formamide concentration and the like. These conditions can be empirically optimized to maximize specific binding and minimize non-specific binding of primer or probe to its target nucleic acid sequence. The terms as used can include reference to conditions under which a probe or primer will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g. at least 2-fold over
background). Stringent conditions can be sequence dependent and will be different in different circumstances. Longer sequences can hybridize specifically at higher temperatures. Generally, stringent conditions can be selected to be about 5° C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm can be the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe or primer. Typically, stringent conditions may be those in which the salt concentration is less than about 1.0 M Na+ ion, typically about 0.01 to 1.0 M Na+ ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C for short probes or primers (e.g. 10 to 50 nucleotides) and at least about 60° C for long probes or primers (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Exemplary low stringent conditions or “conditions of reduced stringency” can include hybridization with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37° C and a wash in 2×SSC at 40° C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37° C, and a wash in 0.1×SSC at 60° C. Hybridization procedures are well known in the art and are described by e.g. Ausubel et al., 1998 and Sambrook et al., 2001. In some embodiments, stringent conditions are hybridization in 0.25 M Na2HPO4 buffer (pH 7.2) containing 1 mM Na2EDTA, 0.5-20% sodium dodecyl sulfate at 45°C, such as 0.5%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19% or 20%, followed by a wash in 5×SSC, containing 0.1% (w/v) sodium dodecyl sulfate, at 55°C to 65°C.

[0048] As used herein, “promoter” or “promoter polynucleotide” can refer to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. The promoter sequence consists of proximal and more distal upstream elements, the latter elements can often be referred to as enhancers. Accordingly, an “enhancer” can be a DNA sequence that can stimulate promoter activity, and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. It is further
recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of some variation may have identical promoter activity.

[0049] As used herein, the phrases “recombinant construct”, “expression construct”, “chimeric construct”, “construct”, and “recombinant DNA construct” can be used interchangeably herein. A recombinant construct comprises an artificial combination of nucleic acid fragments, e.g., regulatory and coding sequences that are not found together in nature. For example, a chimeric construct may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. In some cases, a chimeric construct can be a recombinant construct comprising a plurality of regulatory (e.g., promoter) and coding sequences (e.g., glucose permease gene and hexokinase gene (glucokinase gene)). Each coding sequence in a chimeric construct comprising a plurality of coding sequences can be controlled by or functionally linked to a separate regulatory sequence. Such constructs described herein may be used by itself or may be used in conjunction with a vector. If a vector is used then the choice of vector can be dependent upon the method that will be used to transform host cells as is well known to those skilled in the art. For example, a plasmid vector can be used. The skilled artisan is well aware of the genetic elements that must be present on the vector in order to successfully transform, select and propagate host cells comprising any of the isolated nucleic acid fragments of the disclosure. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) EMBO J. 4:2411-2418; De Almeida et al., (1989) Mol. Gen. Genetics 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, immunoblotting analysis of protein expression, or phenotypic analysis, among others. Vectors can be plasmids, viruses, bacteriophages, pro-viruses, phagemids, transposons, artificial chromosomes, and the like, that replicate autonomously or can integrate into a chromosome of a host cell. A vector can also be a naked RNA polynucleotide, a naked DNA polynucleotide, a polynucleotide composed of both DNA and RNA within the same strand, a poly-lysine-conjugated DNA or RNA, a peptide-conjugated DNA or RNA, a liposome-conjugated DNA, or the like, that is not autonomously replicating. As used herein, the term “expression” refers to the production of a functional end-product e.g., an mRNA or a protein (precursor or mature).
“Operably linked” or “functionally linked” can mean in this context the sequential arrangement of the promoter polynucleotide according to the disclosure with a further oligo- or polynucleotide (e.g., glucose permease gene and/or glucokinase gene), resulting in transcription of said further polynucleotide (e.g., glucose permease gene and/or glucokinase gene). In other words, “operably linked” or “functionally linked” can mean the promoter controls the transcription of the gene (e.g. glucose permease gene and/or glucokinase gene) adjacent or downstream or 3’ to said promoter.

The term “carbon source” generally can refer to a substance suitable to be used as a source of carbon for cell growth. Carbon sources can include, but are not limited to, biomass hydrolysates, starch, sucrose, cellulose, hemicellulose, xylose, and lignin, as well as monomeric components of these substrates. Carbon sources can comprise various organic compounds in various forms, including, but not limited to polymers, carbohydrates, acids, alcohols, aldehydes, ketones, amino acids, peptides, etc. These can include, for example, various monosaccharides such as glucose, dextrose (D-glucose), maltose, oligosaccharides, polysaccharides, saturated or unsaturated fatty acids, succinate, lactate, acetate, ethanol, etc., or mixtures thereof. Photosynthetic organisms can additionally produce a carbon source as a product of photosynthesis. In some embodiments, carbon sources may be selected from biomass hydrolysates and glucose.

The term “feedstock” can be defined as a raw material or mixture of raw materials supplied to a microorganism or fermentation process from which other products can be made. For example, a carbon source, such as biomass or the carbon compounds derived from biomass can be a feedstock for a microorganism that produces a product of interest (e.g. small molecule, peptide, synthetic compound, fuel, alcohol, etc.) in a fermentation process. However, a feedstock may contain nutrients other than a carbon source.

The term “volumetric productivity” or “production rate” can be defined as the amount of product formed per volume of medium per unit of time. Volumetric productivity can be reported in gram per liter per hour (g/L/h).

The term “specific productivity” can defined as the rate of formation of the product. To describe productivity as an inherent parameter of the microorganism and not of the fermentation process, productivity can herein further be defined as the specific productivity in gram product per gram of cell dry weight (CDW) per hour (g/g CDW/h). Using the relation of CDW to OD600 for the
given microorganism specific productivity can also be expressed as gram product per liter culture medium per optical density of the culture broth at 600 nm (OD) per hour (g/L/h/OD).

[0055] The term “yield” can be defined as the amount of product obtained per unit weight of raw material and may be expressed as g product per g substrate (g/g). Yield may be expressed as a percentage of the theoretical yield. “Theoretical yield” is defined as the maximum amount of product that can be generated per a given amount of substrate as dictated by the stoichiometry of the metabolic pathway used to make the product.

[0056] The term “titre” or “titer” can be defined as the strength of a solution or the concentration of a substance in solution. For example, the titre of a product of interest (e.g. small molecule, peptide, synthetic compound, fuel, alcohol, etc.) in a fermentation broth can be described as g of product of interest in solution per liter of fermentation broth (g/L).

[0057] The term “total titer” can be defined as the sum of all product of interest produced in a process, including but not limited to the product of interest in solution, the product of interest in gas phase if applicable, and any product of interest removed from the process and recovered relative to the initial volume in the process or the operating volume in the process.

[0058] As used herein, the term “glucose permease” can refer to any transporter (e.g., myo-inositol transporter and/or glucose permease) that exhibits an affinity for glucose and subsequently facilitates its transport across the cell membrane of a host cell. The glucose permease can be a transmembrane protein. The transport can be passive transport whereby glucose diffuses in or out of the host cell as facilitated by the glucose permease. The glucose permease can be derived from a prokaryotic cell (i.e., Bacteria or Archaea) or a eukaryotic cell (e.g., a fungal cell). The prokaryotic glucose permease protein can be from any genus and species of bacteria or Archaea known in the art. The eukaryotic glucose permease protein can be, for example, from any genus and species of fungus known in the art. The term “bacterial glucose permease” as used herein can refer to a glucose permease as described herein derived from a bacteria.

[0059] As used herein, the term “glucose permease gene” can refer to any nucleic acid (e.g., genomic DNA, cDNA and/or mRNA) that when transcribed and/or translated encodes a glucose permease protein as described herein. The term “bacterial glucose permease gene” as used herein can refer to a bacterial glucose permease protein as described herein derived from a bacteria.

[0060] As used herein, the term “hexokinase” can refer to any protein derived from a prokaryotic cell (i.e., Bacteria or Archaea) or a eukaryotic cell (e.g., a fungal cell) that is an enzyme that can
facilitate the phosphorylation of a hexose (six-carbon sugar). As provided herein, a hexokinase can be a glucokinase. As used herein, the term “glucokinase” can refer to any protein derived from a prokaryotic cell (i.e., Bacteria or Archaea) or a eukaryotic cell (e.g., a fungal cell) that is an enzyme that can facilitate the phosphorylation of glucose to glucose-6-phosphate. The prokaryotic hexokinase or glucokinase protein can be from any genus and species of bacteria or Archaea known in the art. The eukaryotic hexokinase or glucokinase protein can be, for example, from any genus and species of fungus known in the art. The term “bacterial hexokinase” as used herein can refer to a hexokinase as described herein derived from a bacteria. The term “bacterial glucokinase” as used herein can refer to a glucokinase as described herein derived from a bacteria.

[0061] As used herein, the “hexokinase gene” can refer to any nucleic acid (e.g., genomic DNA, cDNA and/or mRNA) that when transcribed and/or translated encodes a hexokinase protein as described herein. As used herein, the “glucokinase gene” can refer to any nucleic acid (e.g., genomic DNA, cDNA and/or mRNA) that when transcribed and/or translated encodes a glucokinase protein as described herein. The term “bacterial hexokinase gene” as used herein can refer to a bacterial hexokinase protein as described herein derived from a bacteria. The term “bacterial glucokinase gene” as used herein can refer to a bacterial glucokinase protein as described herein derived from a bacteria.

[0062] The term "product of interest" or "biomolecule" as used herein refers to any product produced by microbes from feedstock. In some cases, the product of interest may be a small molecule, enzyme, peptide, amino acid, organic acid, synthetic compound, fuel, alcohol, etc. For example, the product of interest or biomolecule may be any primary or secondary extracellular metabolite. The primary metabolite may be, inter alia, ethanol, citric acid, lactic acid, glutamic acid, glutamate, lysine, threonine, tryptophan and other amino acids, vitamins, polysaccharides, etc. The secondary metabolite may be, inter alia, an antibiotic compound like penicillin, or an immunosuppressant like cyclosporin A, a plant hormone like gibberellin, a statin drug like lovastatin, a fungicide like griseofulvin, etc. The product of interest or biomolecule may also be any intracellular component produced by a microbe, such as: a microbial enzyme, including: catalase, amylase, protease, pectinase, glucose isomerase, cellulase, hemicellulase, lipase, lactase, streptokinase, and many others. The intracellular component may also include recombinant proteins, such as: insulin, hepatitis B vaccine, interferon, granulocyte colony-stimulating factor, streptokinase and others.
Overview

[0063] The present disclosure provides a microbial genomic engineering method that does not suffer from the myriad of problems associated with traditional microbial strain improvement programs.

[0064] One aspect provided herein is a method for generating a microorganism (e.g., bacteria) that is capable of increased production of a biomolecule or product of interest. In general, the methods for generating a microorganism for use in producing any biomolecule as provided herein can entail genetically modifying a host microorganism by introducing a member of a library of target genes into said host microorganism to generate a genomically engineered strain of said microorganism, culturing said engineered strain under conditions suitable to produce the biomolecule or product of interest, and selecting said engineered strain if said engineered strain produces an increased amount of the biomolecule or product of interest. The increased amount can be as compared to a wild-type strain of the host microorganism. The increased amount can be as compared to a strain of the host microorganism that does not contain a member of the library of target genes. The library of target genes can comprise a plurality of vectors, wherein each vector in the library comprises a chimeric construct comprising at least one promoter polynucleotide functionally linked or coupled to a target gene.

[0065] An exemplary workflow of one of the embodiments of the invention entails selecting a target gene, acquiring or synthesizing nucleic acid (e.g., DNA) for the target gene, and cloning said acquired or synthesized target gene into a suitable vector. Any method known in the art and/or provided herein can be used to assemble or clone the target gene or target genes into a suitable vector. The vector can be any vector known in the art and/or provided herein that is compatible with the host microorganism to be utilized. Once the vector comprising the target gene(s) is assembled, it can be introduced into the host microorganism. The introduction of the vector can be using any method known in the art and/or provided herein. The host microorganism can be any host microorganism provided herein. Once introduced into the host microorganism, genetically modified hosts can be selected and the insertion of the target gene(s) can be evaluated. The target gene(s) can be engineered to be inserted into specific locations of the host microorganism’s genome. In some cases, the target gene(s) is inserted into a neutral site of the genome that facilitates expression of the target gene(s) without perturbing unintended pathways/processes within the host microorganism. In some cases, the target gene(s) replace specific gene(s) within the host
microorganism. The specific gene can be the homologous target gene normally present in the host microorganism. The integration site, such as, for example, the neutral integration site can be determined empirically such that various sites can tested and a site that permits expression of the integrated target gene(s) without being detrimental to the host cell can be chosen. Integration into a desired site (e.g., neutral site) can be facilitated by cloning the target gene(s) into a vector comprising portions of sequence homologous to the desired integration site (i.e., homologous arms) and subsequently performing a recombination event in the host cell. The target gene(s) can be inserted between the portions of homologous sequence. In one embodiment, the vector comprises about 2 kb of sequence homologous to the desired integration site. The sequence homologous to the desired site can flank a glucose permease gene insert and/or glucose permease-gluco kinase gene insert such that a first portion of the sequence is upstream (i.e., 5’) of the gene insert and a second portion of the sequence is downstream (i.e., 3’) of the gene insert. In another embodiment, the vector comprises about 4 kb of sequence homologous to the desired integration site. In this embodiment, the vector comprises about 2 kb of sequence homologous to the desired integration site upstream (i.e., 5’) to a glucose permease gene insert and/or glucose permease-gluco kinase gene insert and about 2 kb of sequence homologous to the desired integration site downstream (i.e., 3’) to a glucose permease gene insert and/or glucose permease-gluco kinase gene insert. In one embodiment, integration is performed by a single-cross-over integration and subsequent loop out of the plasmid backbone facilitated by counter-selection on a marker present in the vector backbone. In one embodiment, the target gene is any bacterial glucose permease gene known in the art and/or provided herein. In one embodiment, the target gene is any bacterial gluco kinase gene known in the art and/or provided herein. In one embodiment, target genes are any bacterial glucose permease gene known in the art and/or provided herein and any bacterial gluco kinase gene known in the art and/or provided herein.

[0066] Evaluation of the insertion can be performed using any method know in the art such as, for example, amplifying and/or sequencing of the genetically modified microorganism’s genome or portions thereof. In some cases, the methods provided herein also entail the removal or looping out of selection markers through counter selection as described herein. The looping out can be performed using any of the methods provided herein.

[0067] Following the evaluation of the insertion of the target gene(s) and, optional, removal of selection markers, the genetically modified strain can be evaluated for its ability to produce a
biomolecule or product of interest. Prior to evaluation an optional step can be expanding the strain. Expansion can entail culturing the genetically modified strain on plates or in wells in a multi-well plate in growth media suitable for expansion. The evaluation step can entail culturing the genetically modified strain on plates or in wells in a multi-well plate comprising growth media/conditions designed to mimic actual conditions for producing a biomolecule or product of interest. In some cases, the growth media in this step is suitable for the production of biomolecules or products of interest derived from the metabolic processing of glucose. If the genetically modified strain possesses or is predicted to produce a desired or threshold rate of production or yield of the biomolecule or product of interest as determined from the evaluation step, the strain can be selected and placed in cold storage. The prediction can be based on measuring the amount of product of interest and biomass formed at various time points during culturing of the strain and using said measurements to predict how said strain will perform under expanded or larger scale conditions (e.g., fermentation conditions). In one embodiment, the prediction is based on a linear regression analysis of the performance of the strain during the evaluation method.

[0068] In some cases, a genetically modified strain possessing or predicted to produce a desired or threshold rate of production or yield of the biomolecule or product of interest is transferred to or grown in a larger culture under conditions for producing the biomolecule or product of interest (e.g., fermentation conditions). This step can be used in order to determine if the selected strain can perform as predicted under actual conditions for the production of the biomolecule or product of interest. In some cases, the steps provided herein for the introduction and evaluation of each target gene from a library of target genes such as those provided herein are repeated for each target gene from the library in order to select one or more strains of genetically modified microorganisms that produce a desired or threshold yield and/or productivity rate of a biomolecule or product of interest.

[0069] In one embodiment, the biomolecule or product of interest is derived from glucose and the metabolic processing thereof by the microorganism such that the methods provided herein entail the generation of a strain or strains of microorganisms that produce an increased amount of a biomolecule or product of interest derived from the metabolic processing of glucose by the strain or strains. In one embodiment, the methods provided herein entail the introduction of one or more target genes involved in glucose transport and/or metabolism. In one embodiment, the one or more target genes are utilized in a phosphotransferase system (PTS). In one embodiment, the target gene
is a glucose permease gene such that a glucose permease gene is introduced into the host microorganism in the methods provided herein. The glucose permease gene can be a heterologous gene in the host microorganism. In one embodiment, the target gene is a hexokinase gene such that a hexokinase gene is introduced into the host microorganism in the methods provided herein. In one embodiment, both a glucose permease gene and a hexokinase gene are introduced into the host microorganism in the methods provided herein. In one embodiment, the introduction of a glucose permease gene and/or hexokinase gene into the host microorganism produces a non-PTS recombinant glucose uptake system in the host microorganism. The recombinant glucose uptake system can serve to uncouple glucose transport from phosphoenolpyruvate (PEP) utilization, thereby producing more PEP for the synthesis of biomolecules or products of interest. The biomolecules or products of interest produced by the methods provided herein can be any commercial product produced from glucose. In some cases, the biomolecule or product of interest is a small molecule, an amino acid, an organic acid, or an alcohol. The amino acid can be tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine. The organic acid can be succinate, lactate or pyruvate. The alcohol can be ethanol or isobutanol.

[0070] In one embodiment, the disclosed microbial genomic engineering method utilizes a library of glucose permease genes and/or hexokinase genes. A glucose permease gene can be selected based on the glucose permeases affinity for glucose and/or glucose transport rate. In some cases, the microbes are engineered utilizing a glucose permease library, a hexokinase (e.g., glucokinase) library or a combination of glucose permease and hexokinase (e.g., glucokinase) libraries. In one embodiment, the library contains a plurality of chimeric construct inserts such that each insert in the library comprises a glucose permease gene and a hexokinase (e.g., glucokinase) gene. Following engineering, the microbes can be efficiently screened or evaluated for resultant outcome, e.g. production of a product from glucose as provided herein. This process of utilizing the libraries provided herein to define particular genomic alterations and then testing/screening host microbial genomes harboring the alterations can be implemented in an efficient and iterative manner and can be used to identify specific combinations of glucose permease/hexokinase genes (e.g., glucokinase genes) whose expression in a host cell produces a desired or threshold level of a biomolecule or product of interest form glucose.

[0071] In one embodiment, each glucose permease gene or hexokinase gene (glucokinase gene) as provided herein for use in the methods provided herein is under the control of or functionally
linked to a native promoter or any of the promoter polynucleotides provided herein. A “promoter polynucleotide” or a “promoter” or a “polynucleotide having promoter activity” can mean a polynucleotide, preferably deoxyribopolynucleotide, or a nucleic acid, preferably deoxyribonucleic acid (DNA), which when functionally linked to a polynucleotide to be transcribed determines the point and frequency of initiation of transcription of the coding polynucleotide (e.g., glucose permease gene or glucokinase gene), thereby enabling the strength of expression of the controlled polynucleotide to be influenced. In one embodiment, each glucose permease gene and/or hexokinase gene (e.g., glucokinase gene) in a library comprising glucose permease genes and/or hexokinase genes (e.g., glucokinase genes) is under the control of the same or an identical promoter. In one embodiment, each glucose permease gene and/or hexokinase gene (e.g., glucokinase gene) in a library comprising glucose permease genes and/or hexokinase genes (e.g., glucokinase genes) is under the control of separate or different promoter. In yet another embodiment, each target gene in a chimeric construct in a library of chimeric constructs comprising the target genes are under the control of the same or an identical promoter. In a further embodiment, each target gene in a chimeric construct in a library of chimeric constructs comprising the target genes are under the control of a separate or different promoter.

[0072] In one embodiment, provided herein is a promoter ladder for use in generating a library of glucose permease genes or hexokinase genes or glucokinase genes. The term “promoter ladder” as used herein refers to a plurality of promoters with incrementally increasing levels of promoter activity. The term “promoter activity” as used herein refers to the ability of the promoter to initiate transcription of a polynucleotide sequence into mRNA. Methods of assessing promoter activity are well known to those of skill in the art and can include, for example the methods described in Example 2 of US62/264,232, filed on December 7, 2015 and PCT/US16/65464 (i.e., PCT Publication No. WO2017/100376), each of which is herein incorporated by references in its entirety. The term “constitutive promoter” as used herein can refer to a promoter that directs the transcription of its associated genes at a constant rate regardless of the internal or external cellular conditions.

Promoters
[0073] In some embodiments, the present disclosure teaches methods of selecting promoters with optimal expression properties to modulate RNA degradation function and produce beneficial effects on overall-host strain productivity.

[0074] Promoters regulate the rate at which genes are transcribed and can influence transcription in a variety of ways. Constitutive promoters, for example, direct the transcription of their associated genes at a constant rate regardless of the internal or external cellular conditions, while regulatable promoters increase or decrease the rate at which a gene is transcribed depending on the internal and/or the external cellular conditions, e.g., growth rate, temperature, responses to specific environmental chemicals, and the like. Promoters can be isolated from their normal cellular contexts and engineered to regulate the expression of virtually any gene, enabling the effective modification of cellular growth, product yield and/or other phenotypes of interest.

[0075] In some embodiments, the present disclosure teaches methods of identifying one or more promoters and/or generating variants of one or more promoters within a host cell, which exhibit a range of expression strengths (e.g. promoter ladders discussed infra), or superior regulatory properties (i.e., tighter regulatory control for selected genes). A particular combination of these identified and/or generated promoters can be grouped together as a promoter ladder for use in the RNA degradation perturbation experiments explained in more detail below.

[0076] In some embodiments, promoter ladders are created by identifying natural, native, or wild-type promoters associated with a target gene of interest that have a range of expression strengths. These identified promoters can be grouped together as a promoter ladder.

[0077] In some embodiments, promoter ladders are created by: identifying natural, native, or wild-type promoters associated with a target gene of interest and then mutating said promoter to derive multiple mutated promoter sequences. Each of these mutated promoters is tested for effect on target gene expression. In some embodiments, the edited promoters are tested for expression activity across a variety of conditions, such that each promoter variant’s activity is documented/characterized/annotated and stored in a database. The resulting edited promoter variants are subsequently organized into promoter ladders arranged based on the strength of their expression (e.g., with highly expressing variants near the top, and attenuated expression near the bottom, therefore leading to the term “ladder”).

[0078] In some embodiments, the present disclosure teaches promoter ladders that are a combination of identified naturally occurring promoters and mutated variant promoters.
In some embodiments, the present disclosure teaches methods of identifying natural, native, or wild-type promoters that satisfied both of the following criteria: 1) represented a ladder of constitutive promoters; and 2) could be encoded by short DNA sequences, ideally less than 100 base pairs. In some embodiments, constitutive promoters of the present disclosure exhibit constant gene expression across two selected growth conditions (typically compared among conditions experienced during industrial cultivation). In some embodiments, the promoters of the present disclosure will consist of a ~60 base pair core promoter, and a 5’ UTR between 26- and 40 base pairs in length.

In some embodiments, one or more of the aforementioned identified naturally occurring promoter sequences are chosen for gene editing. In some embodiments, the natural promoters are edited via any known genetic mutation methods. In other embodiments, the promoters of the present disclosure are edited by synthesizing new promoter variants with the desired sequence.


A non-exhaustive list of the promoters of the present disclosure is provided in Table 1 below. Each of the promoter sequences in Table 1 can be referred to as a heterologous promoter or heterologous promoter polynucleotide.

Table 1. Selected promoter sequences of the present disclosure.

<table>
<thead>
<tr>
<th>SEQ ID No.</th>
<th>Promoter Short Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>P1</td>
</tr>
<tr>
<td>2</td>
<td>P2</td>
</tr>
<tr>
<td>3</td>
<td>P3</td>
</tr>
<tr>
<td>4</td>
<td>P4</td>
</tr>
<tr>
<td>5</td>
<td>P5</td>
</tr>
<tr>
<td>6</td>
<td>P6</td>
</tr>
<tr>
<td>7</td>
<td>P7</td>
</tr>
<tr>
<td>8</td>
<td>P8</td>
</tr>
</tbody>
</table>

In some embodiments, the promoters of the present invention exhibit at least 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, or 75% sequence identity with a promoter sequences from Table 1.
Glucose Permeases

[0084] Provided herein is a library of glucose permease genes for use in the methods provided herein. The library of glucose permease genes can comprise one or more glucose permease genes. Each glucose permease gene in the library can be a native form of the glucose permease gene or a mutated form. The mutated form can comprise one or more mutations selected from an insertion, deletion, single nucleotide polymorphism (SNP), or translocation. Each glucose permease gene in the library can be a bacterial glucose permease gene. The glucose permease gene can be any glucose permease gene from a prokaryotic cell (i.e., Bacteria and/or Archaea) known in the art. The glucose permease gene can be any glucose permease gene from a eukaryotic cell (e.g., fungal) known in the art. A glucose permease can be considered any protein comprising glucose permease activity. For example, a glucose permease for use herein can be any transporter (e.g., myo-inositol transporter) that exhibits an affinity for glucose and subsequently facilitates its transport across the cell membrane of a host cell. The host cell can be any host cell provided herein. In one embodiment, the library of glucose permease genes comprises glucose permease genes from any strain/species/sub-species of Mycobacterium (e.g., Mycobacterium smegmatis), Streptomyces (e.g., Streptomyces coelicolor), Zymomonas (e.g., Zymomonas mobilis), Synechocystis (e.g., Synechocystis sp. PCC6803), Bifidobacterium (e.g., Bifidobacterium longum), Escherichia (e.g., Escherichia coli), Bacillus (e.g., Bacillus subtilis), Corynebacterium (e.g., Corynebacterium glutamicum), Saccharomyces (e.g., S. cerevisiae) or a combination thereof. In one embodiment, the library of glucose permease genes comprises glucose permease genes that encode polypeptide sequences selected from SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14 or a combination thereof.

[0085] In some embodiments, the permeases of the present invention exhibit at least 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, or 75% sequence identity with a permease provided herein.

[0086] In one embodiment, the library of glucose permease genes comprises glucose permease genes selected from SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24 or a combination thereof.

[0087] In some embodiments, the permease genes of the present invention exhibit at least 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%,
82%, 81%, 80%, 79%, 78%, 77%, 76%, or 75% sequence identity with a permease gene provided herein.

[0088] Each glucose permease in the library can be functionally linked or under the control of its native promoter or a mutated form of its native promoter. Each glucose permease gene in the library can be functionally linked to or controlled by any promoter provided herein. Each glucose permease gene in the library can be controlled by a promoter polynucleotide sequence that comprises a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 8. Each glucose permease gene in the library can be controlled by a promoter polynucleotide sequence that contains a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. In one embodiment, each glucose permease gene in the library is present as a set of glucose permease genes, wherein each set has one glucose permease gene functionally linked to SEQ ID NO. 1, one glucose permease gene functionally linked to SEQ ID NO. 2, one glucose permease gene functionally linked to SEQ ID NO. 3, one glucose permease gene functionally linked to SEQ ID NO. 4, one glucose permease gene functionally linked to SEQ ID NO. 5, one glucose permease gene functionally linked to SEQ ID NO. 6, one glucose permease gene functionally linked to SEQ ID NO. 7 and one glucose permease gene functionally linked to SEQ ID NO. 8 or a combination thereof. Each glucose permease gene in a library of glucose permease genes can be present in a chimeric construct such that the gene can be flanked by one or more regulatory sequences and/or sequence homologous to sequence present in the genome of a host cell. The sequence homologous to sequence present in the host cell can facilitate integration of the glucose permease gene into a site or locus of the host cell genome that comprises complementary sequence. Integration can be via a recombination event. The regulatory sequence can be any regulatory sequence known in the art or provided herein such as, for example, a promoter, start, stop, signal, secretion and/or termination sequence used by the genetic machinery of the host cell.

**Hexokinases**

[0089] Provided herein is a library of hexokinase genes for use in the methods provided herein. The library of hexokinase genes can comprise one or more hexokinase genes. Each hexokinase gene in the library can be a native form or a mutated form of the gene. The mutated form can
comprise one or more mutations selected from an insertion, deletion, single nucleotide polymorphism (SNP), or translocation. Each hexokinase gene can be a glucokinase gene. Each glucokinase gene in the library can be a bacterial glucokinase gene. The glucokinase gene can be any glucokinase gene from a prokaryotic cell (i.e., Bacteria and/or Archaea) known in the art. The glucokinase gene can be any glucokinase gene from a eukaryotic cell (e.g., fungal) known in the art. A glucokinase can be considered any kinase known in the art that can utilize glucose as a substrate and phosphorylate glucose to produce glucose-6-phosphate. In one embodiment, the library of glucokinase genes comprises glucokinase genes from any strain/species/sub-species of Corynebacterium (e.g., C. glutamicum), Zymomonas (e.g., Zymomonas mobilis), Staphylococcus (e.g., S. aureus glkA), Enterococcus (e.g., E. faecalis), Escherichia (e.g., E. coli), Clostridium (e.g., C. difficile), Streptococcus (e.g., S. pneumoniai), Bacillus (e.g., B. anthracis), Renibacterium (e.g., R. salmoninarium), Saccharomyces (e.g., S. cerevisiae) or a combination thereof. In one embodiment, the library of glucokinase genes comprises glucokinase genes that encode polypeptide sequences selected from SEQ ID NO: 15 and/or SEQ ID NO: 16.

[0090] In some embodiments, the hexokinases (e.g., glucokinases) of the present invention exhibit at least 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, or 75% sequence identity with a hexokinase (e.g., glucokinase) provided herein.

[0091] In one embodiment, the library of glucokinase genes comprises glucokinase genes selected from SEQ ID NO: 25 and/or SEQ ID NO: 26.

[0092] In some embodiments, the hexokinase genes (e.g., glucokinase genes) of the present invention exhibit at least 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 79%, 78%, 77%, 76%, or 75% sequence identity with a hexokinase gene (e.g., glucokinase gene) provided herein.

[0093] Each hexokinase gene (e.g., glucokinase gene) in the library can be functionally linked or under the control of its native promoter or a mutated form of its native promoter. Each hexokinase gene (e.g., glucokinase gene) in the library can be functionally linked to or controlled by any promoter provided herein. Each hexokinase gene (e.g., glucokinase gene) in the library can be controlled by a promoter polynucleotide sequence that comprises a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 8. Each hexokinase gene (e.g., glucokinase gene) in the library can be controlled by
a promoter polynucleotide sequence that contains a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. In one embodiment, each hexokinase gene (e.g., glucokinase gene) in the library is present as a set of hexokinase genes (e.g., glucokinase genes), wherein each set has one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 1, one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 2, one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 3, one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 4, one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 5, one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 6, one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 7 and one hexokinase gene (e.g., glucokinase gene) functionally linked to SEQ ID NO. 8 or a combination thereof. Each hexokinase gene in a library of hexokinase genes can be present in a chimeric construct such that the gene can be flanked by one or more regulatory sequences and/or sequence homologous to sequence present in the genome of a host cell. The sequence homologous to sequence present in the host cell can facilitate integration of the hexokinase gene into a site or locus of the host cell genome that comprises complementary sequence. Integration can be via a recombination event. The regulatory sequence can be any regulatory sequence known in the art or provided herein such as, for example, a promoter, start, stop, signal, secretion and/or termination sequence used by the genetic machinery of the host cell.

[0094] Provided herein is a library comprising glucose permease genes and hexokinase genes for use in the methods provided herein. In one embodiment, the glucose permease genes and the hexokinase genes are present in a single chimeric insert. Each glucose permease gene or hexokinase gene (e.g., glucokinase gene) in a chimeric construct in the library can be a native form or a mutated form of either gene. A mutated form of either gene can comprise one or more mutations selected from an insertion, deletion, single nucleotide polymorphism (SNP), or translocation. The glucose permease gene can be a bacterial glucose permease gene. The glucose permease gene can be any bacterial glucose permease gene known in the art. In one embodiment, a glucose permease gene in a chimeric construct comprises a glucose permease gene from any strain/species/sub-species of Mycobacterium (e.g., Mycobacterium smegmatis), Streptomyces (e.g., Streptomyces coelicolor), Zymomonas (e.g., Zymomonas mobilis), Synechocystis (e.g., Synechocystis sp. PCC6803), or Bifidobacterium (e.g., Bifidobacterium longum) Escherichia (e.g.,
Escherichia coli), Bacillus (e.g., Bacillus subtilis) or Corynebacterium (e.g., Corynebacterium glutamicum). In one embodiment, the glucose permease gene in a chimeric construct is a gene that encodes a polypeptide sequence selected from SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13 or SEQ ID NO: 14. In one embodiment, the glucose permease gene in a chimeric construct is selected from SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24 or a combination thereof. The hexokinase gene in each chimeric construct can be a glucokinase gene. Each glucokinase gene in each chimeric construct can be a bacterial glucokinase gene. The glucokinase gene can be any bacterial glucokinase gene known in the art. In one embodiment, a glucokinase gene in a chimeric construct comprises a glucokinase gene from any strain/species/sub-species of Corynebactium (e.g., C. glutamicum), Zymomonas (e.g., Zymomonas mobilis), Staphylococcus (e.g., S. aureus gK), Enterococcus (e.g., E. faecalis), Escherichia (e.g., E. coli), Clostridium (e.g., C. difficile), Streptococcus (e.g., S. pneumonia), Bacillus (e.g., B. anthracis) or Renibacterium (e.g., R. salmoninarium). In one embodiment, the glucokinase gene in the chimeric construct comprises a glucokinase gene that encodes a polypeptide sequence selected from SEQ ID NO: 15 and/or SEQ ID NO: 16. In one embodiment, the glucokinase gene in the chimeric construct is selected from SEQ ID NO: 25 and/or SEQ ID NO: 26. In the chimeric construct as provided herein the glucose permease gene can be any glucose permease gene provided herein, while the glucokinase gene can be any glucokinase gene provided herein. In one embodiment, a library comprising chimeric glucose permease gene and glucokinase gene constructs comprises a plurality of constructs, whereby the plurality comprises each possible combination of glucose permease genes and glucokinase genes provided herein.

[0095] Each glucose permease gene and/or hexokinase gene (e.g., glucokinase gene) in a chimeric construct as provided herein can be functionally linked or under the control of its native promoter or a mutated form of its native promoter. Each glucose permease gene and/or hexokinase gene (e.g., glucokinase gene) in a chimeric construct as provided herein can be functionally linked to or controlled by any promoter provided herein. Each glucose permease genes in a chimeric construct as provided herein can be controlled by a promoter polynucleotide sequence that comprises or contains a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 8. Each hexokinase gene (e.g., glucokinase gene) in a chimeric construct as provided herein can be controlled by a promoter polynucleotide sequence
that comprises or contains a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8. The glucose permease gene and the hexokinase gene (e.g., glucokinase gene) in a chimeric construct as provided herein can each be functionally linked to a promoter that comprises or contains the same sequence. The glucose permease gene and the hexokinase gene (e.g., glucokinase gene) in a chimeric construct as provided herein can each be functionally linked to a promoter that comprises or contains different sequence.

**Generating Mutated forms of Glucose Permease and/or Hexokinase Genes**

[0096] As provided herein, a glucose permease gene and/or a hexokinase gene (e.g., glucokinase gene) for use in the methods provided herein can be a mutated form of the gene from which it is derived. The mutated gene can be mutated in any way known in the art or provided herein.

[0097] In some embodiments, the present disclosure teaches mutating cell populations by introducing, deleting, or replacing selected portions of genomic DNA. Thus, in some embodiments, the present disclosure teaches methods for targeting mutations to a specific locus (e.g., glucose permease or glucokinase). In other embodiments, the present disclosure teaches the use of gene editing technologies such as ZFNs, TALENS, or CRISPR, to selectively edit target DNA regions. Following mutation of the cell populations, the targeted mutations can be isolated from the cells and subsequently used for generating a library of glucose permease and/or hexokinase genes as described herein.

[0098] In some embodiments, the present disclosure teaches mutating selected DNA regions (e.g., glucose permease genes or glucokinase genes) outside of the host organism. For example, in some embodiments, the present disclosure teaches mutating native glucose permease genes or hexokinase genes (e.g., glucokinase gene).

[0099] In some embodiments, the selected regions of DNA are produced in vitro via gene shuffling of natural variants, or shuffling with synthetic oligos, plasmid-plasmid recombination, virus plasmid recombination, or virus-virus recombination. In other embodiments, the genomic regions are produced via error-prone PCR or site-directed mutagenesis.

[0100] In some embodiments, generating mutations in selected genetic regions containing a glucose permease or hexokinase gene is accomplished by “reassembly PCR.” Briefly, oligonucleotide primers (oligos) are synthesized for PCR amplification of segments of a nucleic
acid sequence of interest (e.g., glucose permease gene or glucokinase gene), such that the sequences of the oligonucleotides overlap the junctions of two segments. The overlap region is typically about 10 to 100 nucleotides in length. Each of the segments is amplified with a set of such primers. The PCR products are then “reassembled” according to assembly protocols. In brief, in an assembly protocol, the PCR products are first purified away from the primers, by, for example, gel electrophoresis or size exclusion chromatography. Purified products are mixed together and subjected to about 1-10 cycles of denaturing, reannealing, and extension in the presence of polymerase and deoxynucleoside triphosphates (dNTP's) and appropriate buffer salts in the absence of additional primers (“self-priming”). Subsequent PCR with primers flanking the gene are used to amplify the yield of the fully reassembled and shuffled genes.

[00101] In some embodiments of the disclosure, mutated permease or hexokinase DNA regions, such as those discussed above, are enriched for mutant sequences so that the multiple mutant spectrum, i.e. possible combinations of mutations, is more efficiently sampled. In some embodiments, mutated sequences are identified via a mutS protein affinity matrix (Wagner et al., Nucleic Acids Res. 23(19):3944-3948 (1995), Su et al., Proc. Natl. Acad. Sci. (U.S.A.), 83:5057-5061(1986)) with a preferred step of amplifying the affinity-purified material in vitro prior to an assembly reaction. This amplified material is then put into an assembly or reassembly PCR reaction.

Generation of Libraries comprising glucose permease and/or hexokinase genes

[00102] In some embodiments, the present disclosure teaches inserting and/or replacing and/or deleting a DNA segment comprising a glucose permease and/or glucokinase gene of the host organism (e.g.,). In some aspects, the methods taught herein involve building an oligonucleotide of interest (i.e. a glucose permease or glucose permease-hexokinase segment), which can be incorporated into the genome of a host organism. In some embodiments, the glucose permease or glucose permease-hexokinase DNA segments of the present disclosure can be obtained via any method known in the art, including, copying or cutting from a known template, mutation, or DNA synthesis. In some embodiments, the present disclosure is compatible with commercially available gene synthesis products for producing DNA sequences (e.g., GeneArt™, GeneMaker™, GenScript™, Anagen™, Blue Heron™, Entelechon™, GeNOsys, Inc., or Qiagen™).
[00103] In some embodiments, the glucose permease or glucose permease-hexokinase DNA segment is designed to incorporate the glucose permease or glucose permease-hexokinase DNA segment into a selected DNA region of the host organism (e.g., adding a beneficial non-PTS glucose transport system). The selected DNA region can be a neutral integration site. In other embodiments, the glucose permease or glucose permease-hexokinase DNA segment is designed to remove the native permease and/or hexokinase gene from the DNA of the host organisms (e.g., removing a native PTS glucose transport system).

[00104] In some embodiments, the glucose permease gene, hexokinase gene or glucose permease-hexokinase genes used in the inventive methods can be synthesized in stages as oligonucleotides using any of the methods of enzymatic or chemical synthesis known in the art. The oligonucleotides may be synthesized on solid supports such as controlled pore glass (CPG), polystyrene beads, or membranes composed of thermoplastic polymers that may contain CPG. Oligonucleotides can also be synthesized on arrays, on a parallel microscale using microfluidics (Tian et al., Mol. BioSyst., 5, 714-722 (2009)), or known technologies that offer combinations of both (see Jacobsen et al., U.S. Pat. App. No. 2011/0172127).

[00105] Synthesis on arrays or through microfluidics offers an advantage over conventional solid support synthesis by reducing costs through lower reagent use. The scale required for gene synthesis is low, so the scale of oligonucleotide product synthesized from arrays or through microfluidics is acceptable. However, the synthesized oligonucleotides are of lesser quality than when using solid support synthesis (See Tian infra.; see also Staehler et al., U.S. Pat. App. No. 2010/0216648).

[00106] A great number of advances have been achieved in the traditional four-step phosphoramidite chemistry since it was first described in the 1980’s (see for example, Sierzchala, et al. J. Am. Chem. Soc., 125, 13427-13441 (2003) using peroxy anion deprotection; Hayakawa et al., U.S. Pat. No. 6,040,439 for alternative protecting groups; Azhayev et al., Tetrahedron 57, 4977-4986 (2001) for universal supports; Kozlov et al., Nucleosides, Nucleotides, and Nucleic Acids, 24 (5-7), 1037-1041 (2005) for improved synthesis of longer oligonucleotides through the use of large-pore CPG; and Damha et al., NAR, 18, 3813-3821 (1990) for improved derivatization).

[00107] Regardless of the type of synthesis, the resulting oligonucleotides may then form the smaller building blocks for longer polynucleotides (i.e., glucose permease gene, hexokinase gene or glucose permease-hexokinase genes). In some embodiments, smaller oligonucleotides can be
joined together using protocols known in the art, such as polymerase chain assembly (PCA), ligase chain reaction (LCR), and thermodynamically balanced inside-out synthesis (TBIO) (see Czar et al. Trends in Biotechnology, 27, 63-71 (2009)). In PCA, oligonucleotides spanning the entire length of the desired longer product are annealed and extended in multiple cycles (typically about 55 cycles) to eventually achieve full-length product. LCR uses ligase enzyme to join two oligonucleotides that are both annealed to a third oligonucleotide. TBIO synthesis starts at the center of the desired product and is progressively extended in both directions by using overlapping oligonucleotides that are homologous to the forward strand at the 5’ end of the gene and against the reverse strand at the 3’ end of the gene.

[00108] Another method of synthesizing a larger double stranded DNA fragment is to combine smaller oligonucleotides through top-strand PCR (TSP). In this method, a plurality of oligonucleotides spans the entire length of a desired product and contain overlapping regions to the adjacent oligonucleotide(s). Amplification can be performed with universal forward and reverse primers, and through multiple cycles of amplification a full-length double stranded DNA product is formed. This product can then undergo optional error correction and further amplification that results in the desired double stranded DNA fragment end product.

[00109] In one method of TSP, the set of smaller oligonucleotides that will be combined to form the full-length desired product are between 40-200 bases long and overlap each other by at least about 15-20 bases. For practical purposes, the overlap region should be at a minimum long enough to ensure specific annealing of oligonucleotides and have a high enough melting temperature (Tm) to anneal at the reaction temperature employed. The overlap can extend to the point where a given oligonucleotide is completely overlapped by adjacent oligonucleotides. The amount of overlap does not seem to have any effect on the quality of the final product. The first and last oligonucleotide building block in the assembly should contain binding sites for forward and reverse amplification primers. In one embodiment, the terminal end sequence of the first and last oligonucleotide contain the same sequence of complementarity to allow for the use of universal primers.

Assembling/Cloning Plasmids

[00110] In some embodiments, the present disclosure teaches methods for constructing vectors capable of inserting desired glucose permease genes and/or glucokinase genes DNA sections into
the genome of host organisms. In some embodiments, the present disclosure teaches methods of cloning vectors comprising the insert DNA (e.g., glucose permease gene and/or glucokinase gene), homology arms, and at least one selection marker. (see FIG. 3).

[0011] In some embodiments, the present disclosure is compatible with any vector suited for transformation into the host organism. In some embodiments, the present disclosure teaches use of shuttle vectors compatible with a host cell. In one embodiment, a shuttle vector for use in the methods provided herein is a shuttle vector compatible with an E. coli and/or Corynebacterium host cell. Shuttle vectors for use in the methods provided herein can comprise markers for selection and/or counter-selection as described herein. The markers can be any markers known in the art and/or provided herein. The shuttle vectors can further comprise any regulatory sequence(s) and/or sequences useful in the assembly of said shuttle vectors as known in the art. The shuttle vectors can further comprise any origins of replication that may be needed for propagation in a host cell as provided herein such as, for example, E. coli or C. glutamicum. The regulatory sequence can be any regulatory sequence known in the art or provided herein such as, for example, a promoter, start, stop, signal, secretion and/or termination sequence used by the genetic machinery of the host cell. The termination sequence can be SEQ ID NO: 17 or 18. In certain instances, the target DNA can be inserted into vectors, constructs or plasmids obtainable from any repository or catalogue product, such as a commercial vector (see e.g., DNA2.0 custom or GATEWAY® vectors).

[00113] In some embodiments, the present disclosure teaches cloning vectors with at least one selection marker. Various selection marker genes are known in the art often encoding antibiotic resistance function for selection in prokaryotic (e.g., against ampicillin, kanamycin, tetracycline, chloramphenicol, zeocin, spectinomycin/streptomycin) or eukaryotic cells (e.g. geneticin, neomycin, hygromycin, puromycin, blasticidin, zeocin) under selective pressure. Other marker systems allow for screening and identification of wanted or unwanted cells such as the well-known blue/white screening system used in bacteria to select positive clones in the presence of X-gal or fluorescent reporters such as green or red fluorescent proteins expressed in successfully transduced host cells. Another class of selection markers most of which are only functional in prokaryotic systems relates to counter selectable marker genes often also referred to as “death genes” which express toxic gene products that kill producer cells. Examples of such genes include sacB, rpsL(strA), tetAR, pheS, thyA, gata-1, or ccdB, the function of which is described in (Reyrat et al. 1998 “Counterselectable Markers: Untapped Tools for Bacterial Genetics and Pathogenesis.” Infect Immun. 66(9): 4011-4017).

[00114] In one embodiment, the vector into which the target DNA segment is cloned into comprises a promoter polynucleotide from a promoter ladder or library as provided herein. In one embodiment, provided herein is promoter ladder comprising or containing a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, or SEQ ID NO: 8.

[00115] In one embodiment, the vector comprises a first promoter polynucleotide and a second promoter polynucleotide. The first and/or second promoter polynucleotide can comprise or contain a sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, or SEQ ID NO: 8. The promoter polynucleotide can be used in each case for over-expressing or under-expressing a glucose permease and/or hexokinase in a host microorganism.

[00116] In some embodiments, each generated strain comprising a heterologous glucose permease gene or glucose permease gene-glucokinase gene is cultured and analyzed under one or more criteria of the present disclosure (e.g., productivity of a biomolecule or product of interest). Data from each of the analyzed host strains is associated / correlated with a particular glucose permease gene or glucose permease gene/glucokinase gene combination, and is recorded for future use. Thus, the present disclosure enables the creation of large and highly annotated genetic diversity
libraries/depositories that identify the effect of a glucose permease gene or combination of glucose permease gene/glucokinase gene on any number of microbial genetic or phenotypic traits of interest.

[00117] In some embodiments, the present disclosure teaches the use of vectors for cloning the glucose permease gene and/or hexokinase gene with start and/or stop codon variants such that the cloned gene utilizes the start and/or stop codon variant. For example, typical stop codons for S. cerevisiae and mammals are UAA and UGA, respectively. The typical stop codon for monocotyledonous plants is UGA, whereas insects and E. coli commonly use UAA as the stop codon (Dalphin et al. (1996) Nucl. Acids Res. 24: 216-218).

[00118] In one embodiment, the methods of the provided disclosure comprise codon optimizing one or more genes expressed by the host organism. Methods for optimizing codons to improve expression in various hosts are known in the art and are described in the literature (see U.S. Pat. App. Pub. No. 2007/0292918, incorporated herein by reference in its entirety). Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host (see also, Murray et al. (1989) Nucl. Acids Res. 17:477-508) can be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence.

[00119] In some embodiments, a glucose permease gene or polynucleotide provided herein comprises a molecule codon optimized for translation in a host cell provided herein, such as, for example, E. coli and/or C. glutamicum. The gene or polynucleotide can be an isolated, synthetic or recombinant nucleic acid. The codon optimized glucose permease gene or polynucleotide can be selected from SEQ ID NO: 19, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24. In some cases, provided herein is a permease gene or polynucleotide that is codon optimized to encode a polypeptide sequence selected from SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13 or SEQ ID NO: 14. The codon optimized glucose permease gene or polynucleotide provided herein can be generated using a method known in the art for generating codon optimized polynucleotides such as, for example, GenScript’s OptimumGene™ gene design system or DNA2.0 GeneGPS® Expression Optimization technology.

[00120] In some embodiments, a hexokinase (e.g., glucokinase) gene or polynucleotide provided herein comprises a molecule codon optimized for translation in a host cell provided herein, such
as, for example, *E. coli* and/or *C. glutamicum*. The gene or polynucleotide can be an isolated, synthetic or recombinant nucleic acid. The codon optimized hexokinase gene (e.g., glucokinase gene) can be selected from SEQ ID NO: 25 or SEQ ID NO: 26. In some cases, provided herein is a hexokinase (e.g., glucokinase) gene or polynucleotide that is codon optimized to encode a polypeptide sequence selected from SEQ ID NO: 15 or SEQ ID NO: 16. The codon optimized hexokinase (e.g., glucokinase) gene or polynucleotide provided herein can be generated using a method known in the art for generating codon optimized polynucleotides such as, for example, GenScript’s OptimumGene™ gene design system or DNA2.0 GeneGPS® Expression Optimization technology.

[00121] Protein expression is governed by a host of factors including those that affect transcription, mRNA processing, and stability and initiation of translation. Optimization can thus address any of a number of sequence features of any particular gene. As a specific example, a rare codon induced translational pause can result in reduced protein expression. A rare codon induced translational pause includes the presence of codons in the polynucleotide of interest that are rarely used in the host organism may have a negative effect on protein translation due to their scarcity in the available tRNA pool.

[00122] Alternate translational initiation also can result in reduced heterologous protein expression. Alternate translational initiation can include a synthetic polynucleotide sequence inadvertently containing motifs capable of functioning as a ribosome binding site (RBS). These sites can result in initiating translation of a truncated protein from a gene-internal site. One method of reducing the possibility of producing a truncated protein, which can be difficult to remove during purification, includes eliminating putative internal RBS sequences from an optimized polynucleotide sequence.

[00123] Repeat-induced polymerase slippage can result in reduced heterologous protein expression. Repeat-induced polymerase slippage involves nucleotide sequence repeats that have been shown to cause slippage or stuttering of DNA polymerase which can result in frameshift mutations. Such repeats can also cause slippage of RNA polymerase. In an organism with a high G+C content bias, there can be a higher degree of repeats composed of G or C nucleotide repeats. Therefore, one method of reducing the possibility of inducing RNA polymerase slippage, includes altering extended repeats of G or C nucleotides.
[00124] Interfering secondary structures also can result in reduced heterologous protein expression. Secondary structures can sequester the RBS sequence or initiation codon and have been correlated to a reduction in protein expression. Stemloop structures can also be involved in transcriptional pausing and attenuation. An optimized polynucleotide sequence can contain minimal secondary structures in the RBS and gene coding regions of the nucleotide sequence to allow for improved transcription and translation.

[00125] For example, the optimization process can begin by identifying the desired amino acid sequence to be expressed by the host. From the amino acid sequence a candidate polynucleotide or DNA sequence can be designed. During the design of the synthetic DNA sequence, the frequency of codon usage can be compared to the codon usage of the host expression organism and rare host codons can be removed from the synthetic sequence. Additionally, the synthetic candidate DNA sequence can be modified in order to remove undesirable enzyme restriction sites and add or remove any desired signal sequences, linkers or untranslated regions. The synthetic DNA sequence can be analyzed for the presence of secondary structure that may interfere with the translation process, such as G/C repeats and stem-loop structures.

Transformation of Host Cells

[00126] In some embodiments, the vectors of the present disclosure may be introduced into the host cells using any of a variety of techniques, including transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Particular methods include calcium phosphate transfection, DEAE-Dextran mediated transfection, lipofection, or electroporation (Davis, L., Dibner, M., Battey, I., 1986 “Basic Methods in Molecular Biology”). Other methods of transformation include for example, lithium acetate transformation and electroporation See, e.g., Gietz et al., Nucleic Acids Res. 27:69-74 (1992); Ito et al., J. Bacterol. 153:163-168 (1983); and Becker and Guarente, Methods in Enzymology 194:182-187 (1991). In some embodiments, transformed host cells are referred to as recombinant host strains.

[00127] In some embodiments, the present disclosure teaches high throughput transformation of cells using 96-well plate robotics platform and liquid handling machines known in the art.

[00128] In some embodiments, the present disclosure teaches screening transformed cells with one or more selection markers. In one such embodiment, cells transformed with a vector comprising a kanamycin resistance marker (KanR) are plated on media containing effective amounts of the
kanamycin antibiotic. Colony forming units visible on kanamycin-laced media are presumed to have incorporated the vector cassette into their genome. Insertion of the desired sequences can be confirmed via PCR, restriction enzyme analysis, and/or sequencing of the relevant insertion site.

**Looping Out of Selected Sequences**

[00129] In some embodiments, the present disclosure teaches methods of looping out selected regions of DNA from the host organisms. The looping out method can be as described in Nakashima et al. 2014 “Bacterial Cellular Engineering by Genome Editing and Gene Silencing.” Int. J. Mol. Sci. 15(2), 2773-2793. In some embodiments, the present disclosure teaches looping out selection markers from positive transformants. Looping out deletion techniques are known in the art, and are described in (Tear et al. 2014 “Excision of Unstable Artificial Gene-Specific inverted Repeats Mediates Scar-Free Gene Deletions in Escherichia coli.” Appl. Biochem. Biotech. 175:1858-1867). The looping out methods used in the methods provided herein can be performed using single-crossover homologous recombination or double-crossover homologous recombination. In one embodiment, looping out of selected regions as described herein can entail using single-crossover homologous recombination as described herein.

[00130] First, loop out vectors are inserted into selected target regions within the genome of the host organism (e.g., via homologous recombination, CRISPR, or other gene editing technique). In one embodiment, single-crossover homologous recombination is used between a circular plasmid or vector and the host cell genome in order to loop-in the circular plasmid or vector such as depicted in **FIG. 3**. The inserted vector can be designed with a sequence which is a direct repeat of an existing or introduced nearby host sequence, such that the direct repeats flank the region of DNA slated for looping and deletion. Once inserted, cells containing the loop out plasmid or vector can be counter selected for deletion of the selection region (e.g., see **FIG. 4**; lack of resistance to the selection gene).

**Host Microorganisms**

[00131] The genomic engineering methods provided herein are exemplified with industrial microbial cell cultures, but can be applicable to any organism where desired traits can be identified in a population of genetic mutants.
Thus, as used herein, the term “microorganism” should be taken broadly. It includes, but is not limited to, the two prokaryotic domains, Bacteria and Archaea, as well as certain eukaryotic fungi and protists. However, in certain aspects, “higher” eukaryotic organisms such as insects, plants, and animals can be utilized in the methods taught herein.

Suitable host cells include, but are not limited to: bacterial cells, algal cells, plant cells, fungal cells, insect cells, and mammalian cells. In one illustrative embodiment, suitable host cells include *E. coli* (e.g., SHuffle™ competent *E. coli* available from New England BioLabs in Ipswich, Mass.).

Other suitable host organisms of the present disclosure include microorganisms of the genus *Corynebacterium*. In some embodiments, preferred *Corynebacterium* strains/species include: *C. efficiens*, with the deposited type strain being DSM44549, *C. glutamicum*, with the deposited type strain being ATCC13032, and *C. ammoniagenes*, with the deposited type strain being ATCC6871. In some embodiments, the preferred host of the present disclosure is *C. glutamicum*.

Suitable host strains of the genus *Corynebacterium*, in particular the known wild-type strains: *Corynebacterium glutamicum* ATCC13032, *Corynebacterium acetoglutamicum* ATCC15806, *Corynebacterium acetoacidophilum* ATCC13870, *Corynebacterium melassecola* ATCC17965, *Corynebacterium thermoaminogenes* FERM BP-1539, *Brevibacterium flavum* ATCC14067, *Brevibacterium lactofermentum* ATCC13869, and *Brevibacterium divaricatum* ATCC14020; and L-amino acid-producing mutants, or strains, prepared therefrom, such as, for example, the L-lysine-producing strains: *Corynebacterium glutamicum* FERM-P 1709, *Brevibacterium flavum* FERM-P 1708, *Brevibacterium lactofermentum* FERM-P 1712, *Corynebacterium glutamicum* FERM-P 6463, *Corynebacterium glutamicum* FERM-P 6464, *Corynebacterium glutamicum* DM58-1, *Corynebacterium glutamicum* DG52-5, *Corynebacterium glutamicum* DSM5714, and *Corynebacterium glutamicum* DSM12866.

The term “*Micrococcus glutamicus*” has also been in use for *C. glutamicum*. Some representatives of the species *C. efficiens* have also been referred to as *C. thermoaminogenes* in the prior art, such as the strain FERM BP-1539, for example.

In some embodiments, the host cell of the present disclosure is a eukaryotic cell. Suitable eukaryotic host cells include, but are not limited to: fungal cells, algal cells, insect cells, animal
cells, and plant cells. Suitable fungal host cells include, but are not limited to: *Ascomycota, Basidiomycota, Deuteromycota, Zygomyctota, Fungi imperfecti*. Certain preferred fungal host cells include yeast cells and filamentous fungal cells. Suitable filamentous fungi host cells include, for example, any filamentous forms of the subdivision *Eumycotina* and *Oomycota*. (see, e.g., Hawksworth et al., In Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 1995, CAB International, University Press, Cambridge, UK, which is incorporated herein by reference). Filamentous fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, cellulose and other complex polysaccharides. The filamentous fungi host cells are morphologically distinct from yeast.

[00138] In certain illustrative, but non-limiting embodiments, the filamentous fungal host cell may be a cell of a species of: *Achlya, Acremonium, Aspergillus, Aureobasidium, Bjerkandera, Ceriporiopsis, Cephalosporium, Chrysosporium, Cochliobolus, Corynascus, Cryphonectria, Cryptococcus, Coprinus, Coriolus, Diplodia, Endothia, Fusarium, Gibberella, Gliocladium, Humincola, Hypocrea, Myceliophthora* (e.g., *Myceliophthora thermophila*), *Mucor, Neurospora, Penicillium, Podospora, Phlebia, Piromyces, Pyricularia, Rhizomucor, Rhizopus, Schizophyllum, Scytalidium, Sporotrichum, Talaromyces, Thermoascus, Thielavia, Trametes, Tolypocladium, Trichoderma, Verticillum, Volvariella*, or teleomorphs, or anamorphs, and synonyms or taxonomic equivalents thereof.

[00139] Suitable yeast host cells include, but are not limited to: *Candida, Hansenula, Saccharomyces, Schizosaccharomyces, Pichia, Kluveromyces, and Yarrowia*. In some embodiments, the yeast cell is *Hansenula polymorpha, Saccharomyces cerevisiae, Saccharomyces carlsbergensis, Saccharomyces diastaticus, Saccharomyces norbensis, Saccharomyces kluyveri, Schizosaccharomyces pombe, Pichia pastoris, Pichia finlandica, Pichia trehalophila, Pichia kodamae, Pichia membranaefaciens, Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia quercuum, Pichia pijperi, Pichia stipitis, Pichia methanolica, Pichia angusta, Kluveromyces lactis, Candida albicans, or Yarrowia lipolytica*.

[00140] In certain embodiments, the host cell is an algal such as, *Chlamydomonas* (e.g., *C. Reinhardtii*) and *Phormidium* (P. sp. ATCC29409).

[00141] In other embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells include gram positive, gram negative, and gram-variable bacterial cells. The host cell may be a species of, but not limited to: *Agrobacterium, Alicyclobacillus, Anabaena, Anacystis, Acinetobacter,*
Acidothermus, Arthrobacter, Azobacter, Bacillus, Bifidobacterium, Brevibacterium, Butyrivibrio, Buchnera, Campestris, Campylobacter, Clostridium, Corynebacterium, Chromatium, Coprococcus, Escherichia, Enterococcus, Enterobacter, Erwinia, Fusobacterium, Faecalibacterium, Francisella, Flavobacterium, Geobacillus, Haemophilus, Helicobacter, Klebsiella, Lactobacillus, Lactococcus, Ilyobacter, Micrococcus, Microbacterium, Mesorhizobium, Methylobacterium, Methylobacterium, Mycobacterium, Neisseria, Pantoea, Pseudomonas, Prochlorococcus, Rhodobacter, Rhodopseudomonas, Rhodopseudomonas, Roseburia, Rhodospirillum, Rhodococcus, Scenedesmus, Streptomyces, Streptococcus, Syneccocus, Saccharomonospora, Staphylococcus, Serratia, Salmonella, Shigella, Thermoanaerobacterium, Tropheryma, Tularensis, Temecula, Thermosynechococcus, Thermococcus, Ureaplasma, Xanthomonas, Xylella, Yersinia, and Zymomonas. In some embodiments, the host cell is Corynebacterium glutamicum.

[00142] In some embodiments, the bacterial host strain is an industrial strain. Numerous bacterial industrial strains are known and suitable in the methods and compositions described herein.

[00143] In some embodiments, the bacterial host cell is of the Agrobacterium species (e.g., A. radiobacter, A. rhizogenes, A. rubi), the Arthrobacter species (e.g., A. aurescens, A. citreus, A. globiformis, A. hydrocarbobglutamicus, A. mysorens, A. nicotianae, A. paraffineus, A. protephonniae, A. roseoparaffinus, A. sulfureus, A. ureafaciens), the Bacillus species (e.g., B. thuringiensis, B. anthracis, B. megaterium, B. subtilis, B. lentus, B. circulars, B. pumilus, B. lautus, B. coagulans, B. brevis, B. firmus, B. alkaapios, B. licheniformis, B. clausii, B. stearothermophilus, B. halodurans and B. amyloliqufaciens). In particular embodiments, the host cell will be an industrial Bacillus strain including but not limited to B. subtilis, B. pumilus, B. licheniformis, B. megaterium, B. clausii, B. stearothermophilus and B. amyloliqufaciens. In some embodiments, the host cell will be an industrial Clostridium species (e.g., C. acetobutylicum, C. tetani E88, C. liltuseburense, C. saccharobutylicum, C. perfringens, C. heijerinkii). In some embodiments, the host cell will be an industrial Corynebacterium species (e.g., C. glutamicum, C. acetoacidophillum). In some embodiments, the host cell will be an industrial Escherichia species (e.g., E. coli). In some embodiments, the host cell will be an industrial Erwina species (e.g., E. uredovora, E. carotovora, E. ananas, E. herbicola, E. punctata, E. terreus). In some embodiments, the host cell will be an industrial Pantoea species (e.g., P. citrea, P. agglomerans). In some embodiments, the host cell will be an industrial Pseudomonas species, (e.g., P. putida, P.
aeruginosa, P. mevalonii). In some embodiments, the host cell will be an industrial Streptococcus species (e.g., S. equisimiles, S. pyogenes, S. uberis). In some embodiments, the host cell will be an industrial Streptomyces species (e.g., S. ambofaciens, S. achronogenes, S. avermitilis, S. coelicolor, S. aureofaciens, S. aureus, S. fungicidicus, S. griseus, S. lavidans). In some embodiments, the host cell will be an industrial Zymomonas species (e.g., Z. mobilis, Z. lipolytica), and the like.

[00144] In various embodiments, strains that may be used in the practice of the disclosure including both prokaryotic and eukaryotic strains, are readily accessible to the public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

[00145] In some embodiments, the methods of the present disclosure are also applicable to multicellular organisms. For example, the platform could be used for improving the performance of crops. The organisms can comprise a plurality of plants such as Gramineae, Fagaceae, Poaceae, Agrostis, Phleum, Dactylis, Sorgum, Setaria, Zea, Oryza, Triticum, Secale, Avena, Hordeum, Saccharum, Poa, Festuca, Stenotaphrum, Cynodon, Coix, Olyreae, Phareae, Compositae or Leguminosae. For example, the plants can be corn, rice, soybean, cotton, wheat, rye, oats, barley, pea, beans, lentil, peanut, yam bean, cowpeas, velvet beans, clover, alfalfa, lupine, vetch, lotus, sweet clover, wisteria, sweet pea, sorghum, millet, sunflower, canola or the like. Similarly, the organisms can include a plurality of animals such as non-human mammals, fish, insects, or the like.

Cell Fermentation and Culture

[00146] Microorganisms of the present disclosure including those genetically engineered as described herein can be cultured in conventional nutrient media modified as appropriate for any desired biosynthetic reactions or selections. In some embodiments, the present disclosure teaches culture in inducing media for activating promoters. In some embodiments, the present disclosure teaches media with selection agents, including selection agents of transformants (e.g., antibiotics), or selection of organisms suited to grow under inhibiting conditions (e.g., high ethanol conditions). In some embodiments, the present disclosure teaches growing cell cultures in media optimized for
cell growth. In other embodiments, the present disclosure teaches growing cell cultures in media optimized for product yield such as, for example, products or biomolecules of interest derived from metabolic processing of glucose. In some embodiments, the present disclosure teaches growing cultures in media capable of inducing cell growth and also contains the necessary precursors for final product production (e.g., high levels of sugars for ethanol production). The biomolecules or products of interest produced by the methods provided herein can be any commercial product produced from glucose. In some cases, the biomolecule or product of interest is a small molecule, an amino acid, an organic acid, or an alcohol. The amino acid can be tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine. The organic acid can be succinate, lactate or pyruvate. The alcohol can be ethanol or isobutanol.

Plant Culture Catalogue and supplement also from Sigma-Aldrich, Inc (St Louis, Mo.) (“Sigma-PCCS”), all of which are incorporated herein by reference.

[00148] The culture medium or fermentation medium to be used must in a suitable manner satisfy the demands of the respective strains. Descriptions of culture media for various microorganisms are present in the “Manual of Methods for General Bacteriology” of the American Society for Bacteriology (Washington D.C., USA, 1981). The terms culture medium and fermentation medium are interchangeable.

[00149] In some embodiments, the present disclosure teaches that the microorganisms produced may be cultured continuously—as described, for example, in WO 05/021772—or discontinuously in a batch process (batch cultivation) or in a fed-batch or repeated fed-batch process for the purpose of producing the desired organic-chemical compound. A summary of a general nature about known cultivation methods is available in the textbook by Chmiel (Bioprozeßtechnik. 1: Einführung in die Bioverfahrenstechnik (Gustav Fischer Verlag, Stuttgart, 1991)) or in the textbook by Storhas (Bioreaktoren and periphere Einrichtungen (Vieweg Verlag, Braunschweig/Wiesbaden, 1994)).

[00150] In some embodiments, the cells of the present disclosure are grown under batch or continuous fermentations conditions. Classical batch fermentation is a closed system, wherein the compositions of the medium is set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a fed-batch fermentation which also finds use in the present disclosure. In this variation, the substrate is added in increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. Continuous fermentation is a system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned medium is removed simultaneously for processing and harvesting of desired proteins. In some embodiments, continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. In some embodiments, continuous fermentation generally maintains the cultures at a stationary or late log/stationary, phase growth. Continuous fermentation systems strive to maintain steady state growth conditions.
[00151] Methods for modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

[00152] For example, a non-limiting list of carbon sources for the cultures of the present disclosure include, sugars and carbohydrates such as, for example, glucose, sucrose, lactose, fructose, maltose, molasses, sucrose-containing solutions from sugar beet or sugar cane processing, starch, starch hydrolysate, and cellulose; oils and fats such as, for example, soybean oil, sunflower oil, groundnut oil and coconut fat; fatty acids such as, for example, palmitic acid, stearic acid, and linoleic acid; alcohols such as, for example, glycerol, methanol, and ethanol; and organic acids such as, for example, acetic acid or lactic acid.

[00153] A non-limiting list of the nitrogen sources for the cultures of the present disclosure include, organic nitrogen-containing compounds such as peptones, yeast extract, meat extract, malt extract, corn steep liquor, soybean flour, and urea; or inorganic compounds such as ammonium sulfate, ammonium chloride, ammonium phosphate, ammonium carbonate, and ammonium nitrate. The nitrogen sources can be used individually or as a mixture.

[00154] A non-limiting list of the possible phosphorus sources for the cultures of the present disclosure include, phosphoric acid, potassium dihydrogen phosphate or dipotassium hydrogen phosphate or the corresponding sodium-containing salts. The culture medium may additionally comprise salts, for example in the form of chlorides or sulfates of metals such as, for example, sodium, potassium, magnesium, calcium and iron, such as, for example, magnesium sulfate or iron sulfate, which are necessary for growth. Finally, essential growth factors such as amino acids, for example homoserine and vitamins, for example thiamine, biotin or pantothenic acid, may be employed in addition to the abovementioned substances.

[00155] In some embodiments, the pH of the culture can be controlled by any acid or base, or buffer salt, including, but not limited to sodium hydroxide, potassium hydroxide, ammonia, or aqueous ammonia; or acidic compounds such as phosphoric acid or sulfuric acid in a suitable manner. In some embodiments, the pH is generally adjusted to a value of from 6.0 to 8.5, preferably 6.5 to 8.

[00156] In some embodiments, the cultures of the present disclosure may include an anti-foaming agent such as, for example, fatty acid polyglycol esters. In some embodiments the cultures of the
present disclosure are modified to stabilize the plasmids of the cultures by adding suitable selective substances such as, for example, antibiotics.

[00157] In some embodiments, the culture is carried out under aerobic conditions. In order to maintain these conditions, oxygen or oxygen-containing gas mixtures such as, for example, air are introduced into the culture. It is likewise possible to use liquids enriched with hydrogen peroxide. The fermentation is carried out, where appropriate, at elevated pressure, for example at an elevated pressure of from 0.03 to 0.2 MPa. The temperature of the culture is normally from 20°C to 45°C and preferably from 25°C to 40°C, particularly preferably from 30°C to 37°C. In batch or fed-batch processes, the cultivation is preferably continued until an amount of the desired organic-chemical compound sufficient for being recovered has formed. In some embodiments, the culture is carried out under anaerobic conditions.

Product Recovery and Quantification

[00158] Methods for screening for the production of products of interest are known to those of skill in the art and are discussed throughout the present specification. Such methods may be employed when screening the strains of the disclosure. The biomolecules or products of interest produced by the methods provided herein can be any commercial product produced from glucose. In some cases, the biomolecule or product of interest is an amino acid, an organic acid, or an alcohol. The amino acid can be tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine. The organic acid can be succinate, lactate or pyruvate. The alcohol can be ethanol or isobutanol.

[00159] In some embodiments, the present disclosure teaches methods of improving strains designed to produce non-secreted intracellular products. For example, the present disclosure teaches methods of improving the robustness, yield, efficiency, or overall desirability of cell cultures producing intracellular enzymes, oils, pharmaceuticals, or other valuable small molecules or peptides. The recovery or isolation of non-secreted intracellular products can be achieved by lysis and recovery techniques that are well known in the art, including those described herein.

[00160] For example, in some embodiments, cells of the present disclosure can be harvested by centrifugation, filtration, settling, or other method. Harvested cells are then disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well known to those skilled in the art.
[00161] The resulting product of interest, e.g. a polypeptide, may be recovered/isolated and optionally purified by any of a number of methods known in the art. For example, a product polypeptide may be isolated from the nutrient medium by conventional procedures including, but not limited to: centrifugation, filtration, extraction, spray-drying, evaporation, chromatography (e.g., ion exchange, affinity, hydrophobic interaction, chromatofocusing, and size exclusion), or precipitation. Finally, high performance liquid chromatography (HPLC) can be employed in the final purification steps. (See for example Purification of intracellular protein as described in Parry et al., 2001, *Biochem. J.*.353:117, and Hong et al., 2007, *Appl. Microbiol. Biotechnol.*. 73:1331, both incorporated herein by reference).


[00163] In some embodiments, the present disclosure teaches the methods of improving strains designed to produce secreted products. For example, the present disclosure teaches methods of improving the robustness, yield, efficiency, or overall desirability of cell cultures producing valuable small molecules or peptides.

[00164] In some embodiments, immunological methods may be used to detect and/or purify secreted or non-secreted products produced by the cells of the present disclosure. In one example approach, antibody raised against a product molecule (e.g., against an insulin polypeptide or an immunogenic fragment thereof) using conventional methods is immobilized on beads, mixed with cell culture media under conditions in which the endoglucanase is bound, and precipitated. In some embodiments, the present disclosure teaches the use of enzyme-linked immunosorbent assays (ELISA).
[00165] In other related embodiments, immunochromatography is used, as disclosed in U.S. Pat. No. 5,591,645, U.S. Pat. No. 4,855,240, U.S. Pat. No. 4,435,504, U.S. Pat. No. 4,980,298, and Se-Hwan Paek, et al., “Development of rapid One-Step Immunochromatographic assay, Methods”, 22, 53-60, 2000), each of which are incorporated by reference herein. A general immunochromatography detects a specimen by using two antibodies. A first antibody exists in a test solution or at a portion at an end of a test piece in an approximately rectangular shape made from a porous membrane, where the test solution is dropped. This antibody is labeled with latex particles or gold colloidal particles (this antibody will be called as a labeled antibody hereinafter). When the dropped test solution includes a specimen to be detected, the labeled antibody recognizes the specimen so as to be bonded with the specimen. A complex of the specimen and labeled antibody flows by capillarity toward an absorber, which is made from a filter paper and attached to an end opposite to the end having included the labeled antibody. During the flow, the complex of the specimen and labeled antibody is recognized and caught by a second antibody (it will be called as a tapping antibody hereinafter) existing at the middle of the porous membrane and, as a result of this, the complex appears at a detection part on the porous membrane as a visible signal and is detected.

[00166] In some embodiments, the screening methods of the present disclosure are based on photometric detection techniques (absorption, fluorescence). For example, in some embodiments, detection may be based on the presence of a fluorophore detector such as GFP bound to an antibody. In other embodiments, the photometric detection may be based on the accumulation on the desired product from the cell culture. In some embodiments, the product may be detectable via UV of the culture or extracts from said culture.

[00167] In some embodiments, the product recovery methods allow for the quantitative determination of the effect on performance of each candidate glucose permease gene and/or glucokinase gene. In some embodiments, the product recovery methods allow for the quantitative determination of the effect on performance of each candidate glucose permease gene/glucokinase gene combination, allowing for comparison of each and selection for the optimal combination.

Selection Criteria and Goals

[00168] The selection of a particular strain of host cell expressing a heterologous glucose permease or glucose permease and glucokinase can be based on specific goals. For example, in some
embodiments, the program goal may be to maximize single batch yields of reactions with no immediate time limits. In other embodiments, the program goal may be to rebalance biosynthetic yields to produce a specific product, or to produce a particular ratio of products. In some embodiments, the program goal may be to improve performance characteristics such as yield, titer, productivity, by-product elimination, tolerance to process excursions, optimal growth temperature and growth rate. In some embodiments, the program goal is improved host performance as measured by volumetric productivity, specific productivity, yield or titre, of a product of interest produced by a microbe.

[00169] In other embodiments, the program goal may be to optimize synthesis efficiency of a commercial strain in terms of final product yield per quantity of inputs (e.g., total amount of ethanol produced per pound of sucrose). In other embodiments, the program goal may be to optimize synthesis speed, as measured for example in terms of batch completion rates, or yield rates in continuous culturing systems. In one embodiment, the program goal is to optimize final product yield and/or production rate of a biomolecule or product of interest. The biomolecules or products of interest produced by the methods provided herein can be any commercial product produced from glucose. In some cases, the biomolecule or product of interest is a small molecule, an amino acid, an organic acid, or an alcohol. The amino acid can be tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine. The organic acid can be succinate, lactate or pyruvate. The alcohol can be ethanol or isobutanol.

[00170] Persons having ordinary skill in the art will recognize how to tailor strain selection criteria to meet the particular project goal. For example, selections of a strain’s single batch max yield at reaction saturation may be appropriate for identifying strains with high single batch yields. Selection based on consistency in yield across a range of temperatures and conditions may be appropriate for identifying strains with increased robustness and reliability.

[00171] In some embodiments, the selection criteria for the initial phase and the tank-based validation will be identical. In other embodiments, tank-based selection may operate under additional and/or different selection criteria.
EXAMPLES

[00172] The present invention is further illustrated by reference to the following Examples. However, it should be noted that these Examples, like the embodiments described above, is illustrative and is not to be construed as restricting the scope of the invention in any way.

Example 1: Transformation of Corynebacterium with Glucose Permease and Glucokinase Library

Generation of Glucose Permease Libraries

[00173] A number of glucose permeases from various bacteria were selected for generation of a glucose permease library based on their affinity and transport rates for glucose as reported in the literature. The glucose permease selected for inclusion in the library were the glucose permease genes from Mycobacterium smegmatis that encodes SEQ ID NO: 9; from Bifidobacterium longum (BL_1631 in FIG. 1) that encodes SEQ ID NO: 11; from Zymomonas mobilis (glf in FIG. 1) that encodes SEQ ID NO: 13; from Synechocystis sp. PCC6803 (glcP in FIG. 1) that encodes SEQ ID NO: 12; from Streptomyces coelicolor (SC05578 in FIG. 1) that encodes SEQ ID NO: 10; and the myo-inositol transporter gene from Corynebacterium glutamicum that encodes SEQ ID NO: 14. Additionally, two glucokinases were selected for use in the generation of the glucose permease library based on their compatibility with the host cell. The glucokinases chosen were the ppgK glucokinase gene from Corynebacterium glutamicum that encodes SEQ ID NO: 16 and the glk kinase gene from Z. mobilis that encodes SEQ ID NO: 15.

[00174] For generation of the glucose permease library, each glucose permease described above was paired with a glucokinase described above such that each gene was cloned into a single C. glutamicum / Escherichia coli compatible expression vector using type IIIs restriction and ligation cloning techniques. More specifically, genes that encode the M. smegmatis glucose permease (SEQ ID NO: 9), the B. longum glucose permease (BL_1631; SEQ ID NO: 11), the Synechocystis sp. PCC6803 glucose permease (glcP; SEQ ID NO: 12), the C. glutamicum myo-inositol transporter (iolT1; SEQ ID NO: 14) and the S. coelicolor glucose permease (SCO5578; SEQ ID NO: 10) were all individually paired with the gene that encodes C. glutamicum ppgK glucokinase (SEQ ID NO: 16), while the gene that encodes the Z. mobilis glucose permease (glf; SEQ ID NO: 10) was paired with the gene that encodes the Z. mobilis glk kinase (SEQ ID NO: 15) and separately paired with
the gene that encodes *C. glutamicum* ppgK glucokinase (SEQ ID NO: 16). In addition, within each glucose permease-glucokinase construct, a P1 promoter (SEQ ID NO: 1) was cloned in front of the respective glucose permease gene, while a P2 promoter (SEQ ID NO: 2) was cloned in front of the respective glucokinase gene such that each respective permease or glucokinase gene was functionally linked to the preceding or upstream promoter. Finally, each permease gene in a construct ended with a T1 termination sequence (SEQ ID NO: 17), while each glucokinase gene in a construct ended with a T2 termination sequence (SEQ ID NO: 18).

**Transformation of Assembled Clones into E.coli**

[00175] Vectors containing the glucose permease-glucokinase genes were each individually transformed into *E.coli* in order to identify correctly assembled clones, and to amplify vector DNA for *Corynebacterium* transformation. Amplified DNA was validated via PCR. Positive clones were saved at -20°C fridge for future use.

**Transformation of Assembled Clones into Corynebacterium**

[00176] Validated clones were then individually transformed into *Corynebacterium glutamicum* host cells via electroporation. In order to test the effect of strain background on construct performance, two different strain backgrounds (i.e., Parent 1/background 2 and parent 2/background 1 in FIGs. 1 and 2) of *C. glutamicum* were used with each construct being transformed into each background. Each vector was designed to integrate into a neutral integration site within the *C. glutamicum* genome that was empirically determined to permit expression of the heterologous glucose permease and glucokinase genes but not be detrimental to the host cell. To facilitate integration, the expression vector further comprised about 2 kb of sequence homologous (i.e., homology arms) to the desired integration site whereby each glucose permease-glucokinase gene cassette described above was inserted between. Integration into the genome occurred by single-crossover integration and then loop-out of the plasmid backbone facilitated by counter-selection on a second marker included in the plasmid backbone.

[00177] Transformed bacteria were then tested for assembly success (correct integration into the genome). Colonies from each *Corynebacterium* transformation plate were cultured and tested for correct integration via PCR. This process was repeated for each of the transformations conducted.
for each glucose permease-glucokinase construct. Genomic integration of each transformation was also analyzed with respect to the targeted genome location for each plasmid.

**Evaluation of Individual Glucose Permease-Glukokinase Constructs in Corynebacterium**

[00178] The phenotype of each transformant was then tested in an evaluation method designed to mimic or simulate a specific fermentation process for producing a desired fermentation end product in order to determine the effects the expression of each construct in each host cell background had on the desired phenotype (i.e., improved ability to produce a desired fermentation end product). Briefly, the evaluation method was an experiment where the transformants were cultured in a 96 well plate format under conditions that were meant to mimic fermentation conditions. The amount of product and biomass formed at various time points was measured and used to predict how each strain would perform under fermentation conditions. This prediction was a linear regression generated from testing strains with various fermentation performance in the evaluation method and determining the correlation of measurements to performance.

[00179] The rate of production and yield of the desired fermentation end product was determined for each permease-glucokinase transformant, some examples of which are shown in FIG. 1. As shown in FIG. 1, for the specific permease-glucokinase inserts shown, the productivity (top) in a fermentation process was predicted to increase in each host background for each permease-glucokinase insert shown vs. the respective control host cell, while the yield (bottom) was predicted to be similar (glcP, B1.1631), increased (SCO5578), or decreased (ggl) vs. the respective control host cell. Please note that the AU units in FIG. 1 are the output of a linear regression that takes as inputs various measurements made on cultures at small scale and predicts the performance of strains under fermentation conditions.

**Assessment of Individual Glucose-Permease-Glucokinase Constructs Under Fermentation Conditions**

[00180] Following evaluation as described above, transformants with heterologous glucose permease-glucokinase genes with predicted increased performance (i.e., increased predicted productivity and/or predicted yield) were selected and subsequently grown in medium containing glucose under conditions designed to facilitate fermentation and the production of desired fermentation end products. Following growth of each transformant for a predetermined length of
time under fermentation conditions designed to produce a desired end-product, the yield and volumetric productivity of the end-product for each transformant was then determined. Briefly, high-performance liquid chromatography (HPLC) was used to determine the amount of product (i.e., avg yield) produced for a certain amount of substrate fed. Productivity (i.e., avg productivity) was similarly determined with the addition of time and volume data.

[00181] As shown in FIG. 2, the BL_1631 glucose permease- C. glutamicum ppgK kinase construct increased productivity of the host cell with background 1 by 15% as well as increased the yield by about 1%. In addition, in host cell with background 2, both the glk permease-glk kinase construct and the glcP permease- C. glutamicum ppgK kinase construct increased productivity by more than 30%, but did not affect yield. Accordingly, this example shows that the methods provided herein can be used to increase the performance of microbial strains in terms of producing fermentation end products.

INCORPORATION BY REFERENCE

[00182] The following applications are hereby incorporated by reference in their entirety, including all descriptions, references, figures, and claims for all purposes: U.S. Application No. 15/396,230, filed on December 30, 2016; International Application No. PCT/US2016/065465, filed on December 07, 2016; U.S. Application No. 15/140,296, filed on April 27, 2016; U.S. Provisional Application No. 62/368,786, filed on July 29, 2016; and U.S. Provisional Application No. 62/264,232, filed on December 07, 2015.

[00183] All references, articles, publications, patents, patent publications, and patent applications cited herein are incorporated by reference in their entireties for all purposes.

[00184] However, mention of any reference, article, publication, patent, patent publication, and patent application cited herein is not, and should not be taken as an acknowledgment or any form of suggestion that they constitute valid prior art or form part of the common general knowledge in any country in the world.

*****
CLAIMS

What is claimed is:

1. A host cell comprising a heterologous glucose permease gene functionally linked to a first promoter polynucleotide, wherein the first promoter polynucleotide comprises a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

2. The host cell of claim 1, wherein the glucose permease gene is a bacterial glucose permease gene.

3. The host cell of claim 2, wherein the bacterial glucose permease gene is a gene that encodes a polypeptide with an amino acid sequence selected from SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9 and SEQ ID NO: 14.

4. The host cell of claim 2, wherein the bacterial glucose permease gene is a gene with a nucleotide sequence selected from SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19 and SEQ ID NO: 24.

5. The host cell of any one of claims 1-4, further comprising a hexokinase gene functionally linked to a second promoter polynucleotide, wherein the second promoter polynucleotide comprises a nucleotide sequence selected from SEQ ID NO: 2, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

6. The host cell of claim 5, wherein the hexokinase gene is a glucokinase gene.

7. The host cell of claim 6, wherein the glucokinase gene is a bacterial glucokinase gene.

8. The host cell of claim 7, wherein the bacterial glucokinase gene is a gene that encodes a polypeptide sequence selected from SEQ ID NO: 15 and SEQ ID NO: 16.

9. The host cell of claim 7, wherein the bacterial glucokinase gene is a gene with a nucleotide sequence selected from SEQ ID NO: 25 and SEQ ID NO: 26.

10. The host cell of any one of claims 5-9, wherein the first promoter polynucleotide and the second promoter polynucleotide are different.

11. The host cell of any one of claims 5-9, wherein the first promoter polynucleotide and the second promoter polynucleotide are identical.

12. The host cell of any one of claims 1-11, wherein the host cell belongs to the genus _Corynebacterium._
13. The host cell of any one of claims 1-11, wherein the host cell is *Corynebacterium glutamicum*.

14. A method of producing a biomolecule from glucose comprising culturing a host cell of any one of claims 1-13 under conditions suitable for producing the biomolecule.

15. The method of claim 14, wherein the biomolecule is a small molecule, an amino acid, a nucleotide, an organic acid, or an alcohol.

16. The method of claim 15, wherein the amino acid is tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine.

17. The method of claim 15, wherein the organic acid is succinate, lactate or pyruvate.

18. The method of claim 15, wherein the alcohol is ethanol or isobutanol.

19. A method for generating a microorganism capable of increased production of a biomolecule from glucose, the method comprising:
   a) genetically modifying a host microorganism, wherein the modifying comprises introducing a glucose permease gene from a library of glucose permease genes into the genome of the host microorganism, wherein each glucose permease gene from the library of glucose permease genes is functionally linked to a promoter comprising a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8 and wherein the modification generates a strain of the host microorganism expressing the glucose permease gene;
   b) repeating step a) for a plurality of rounds until a plurality of strains of the host microorganism are generated, wherein each strain of the plurality of strains of the host microorganism expresses a separate glucose permease gene from the library of glucose permease genes;
   c) contacting each strain of the plurality of strains of the host microorganism with a carbon source comprising glucose under fermentative conditions; and
   d) selecting each strain of the host microorganism that produces an increased amount of a biomolecule from glucose as compared to the amount of the biomolecule produced from glucose from a control microorganism, wherein the control microorganism does not express a glucose permease gene from the library of glucose permease genes.
20. The method of claim 19, wherein each of the glucose permease genes in the library of glucose permease genes is a bacterial glucose permease gene.

21. The method of claim 20, wherein the library of bacterial glucose permease genes comprises genes that encode polypeptide sequences of SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9, SEQ ID NO: 14 or a combination thereof.

22. The method of claim 20, wherein the library of bacterial glucose permease genes comprises genes with a nucleotide sequence of SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19, SEQ ID NO: 24 or a combination thereof.

23. The method of any one of claims 19-22, further comprising introducing a hexokinase gene from a library of hexokinase genes, wherein each hexokinase gene from the library of hexokinase genes is functionally linked to a promoter polynucleotide, wherein the promoter polynucleotide comprises a sequence selected from SEQ ID NO: 2, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

24. The method of claim 23, wherein the introduction of each hexokinase gene from the library of hexokinase genes is concurrent with the introduction of each glucose permease gene from the library of glucose permease genes.

25. The method of claim 24, wherein each hexokinase gene from the library of hexokinase genes is present in a chimeric construct comprising a glucose permease gene from the library of glucose permease genes.

26. The method of any of claims 23-25, wherein the hexokinase gene is a glucokinase gene.

27. The method of claim 26, wherein the glucokinase gene is a bacterial glucokinase gene.

28. The method of claim 27, wherein the library of bacterial glucokinase genes comprises genes that encode polypeptide sequences of SEQ ID NO: 15 and/or SEQ ID NO: 16.

29. The method of claim 27, wherein the library of bacterial glucokinase genes comprises genes with nucleotide sequences of SEQ ID NO: 25 and/or SEQ ID NO: 26.

30. The method of any one of claims 23-29, wherein the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are different.
31. The method of any one of claims 23-29, wherein the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are identical.

32. The method of any one of claims 19-31, wherein the host microorganism belongs to the genus *Corynebacterium*.

33. The method of any one of claims 19-31, where the host microorganism is *Corynebacterium glutamicum*.

34. The method of any one of claims 19-32, wherein the introducing is performed by transformation, transduction or electroporation.

35. The method of any one of claims 19-34, wherein the biomolecule is a small molecule, an amino acid, a nucleotide, an organic acid, or an alcohol.

36. The method of claim 35, wherein the amino acid is tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine.

37. The method of claim 35, wherein the organic acid is succinate, lactate or pyruvate.

38. The method of claim 35, wherein the alcohol is ethanol or isobutanol.

39. A library of glucose permease genes, wherein each glucose permease gene in the library of glucose permease genes is functionally linked to a promoter comprising a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

40. The library of claim 39, wherein each glucose permease gene is a bacterial glucose permease gene.

41. The library of claim 40, wherein the library of bacterial glucose permease genes comprises genes that encode polypeptide sequences of SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9, SEQ ID NO: 14 or a combination thereof.

42. The library of claim 40, wherein the library of bacterial glucose permease genes comprises genes with nucleotide sequences of SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19, SEQ ID NO: 24 or a combination thereof.

43. The library of any one of claims 39-42, wherein each glucose permease gene in the library of glucose permease genes is a first portion of a chimeric construct, wherein the chimeric construct comprises a second portion, wherein the second portion is a hexokinase gene.
44. The library of claim 43, wherein the hexokinase gene is functionally linked to a promoter polynucleotide, wherein the promoter polynucleotide comprises a sequence selected from SEQ ID NO: 2, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, and SEQ ID NO: 8.

45. The library of any of claims 43-44, wherein the hexokinase gene is a glucokinase gene.

46. The library of claim 45, wherein the glucokinase gene is a bacterial glucokinase gene.

47. The library of claim 46, wherein the library of bacterial glucokinase genes comprises genes that encode polypeptide sequences of SEQ ID NO: 15 and/or SEQ ID NO: 16.

48. The library of claim 46, wherein the library of bacterial glucokinase genes comprises genes with nucleotide sequences of SEQ ID NO: 25 and/or SEQ ID NO: 26.

49. The library of any one of claims 44-48, wherein the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are different.

50. The library of any one of claims 44-48, wherein the promoter polynucleotide functionally linked to the glucose permease gene and the promoter polynucleotide functionally linked to the hexokinase gene are identical.

51. A method of producing a biomolecule comprising introducing a glucose permease gene from the library of any one of claims 39-50 into a host cell and culturing the host cell under conditions suitable for producing the biomolecule.

52. The method of claim 51, wherein the biomolecule is an amino acid, a nucleotide, an organic acid, or an alcohol.

53. The method of claim 52, wherein the amino acid is tyrosine, phenylalanine, tryptophan, aspartic acid, asparagine, threonine, isoleucine, methionine, or lysine.

54. The method of claim 52, wherein the organic acid is succinate, lactate or pyruvate.

55. The method of claim 52, wherein the alcohol is ethanol or isobutanol.

56. The method of any one of claims 51-55, wherein the host cell belongs to the genus Corynebacterium.

57. The method of any one of claims 51-55, wherein the host cell is Corynebacterium glutamicum.

58. The method of any one of claims 51-57, wherein the introducing is performed by transformation, transduction or electroporation.
59. An isolated, synthetic or recombinant polynucleotide comprising a codon optimized polynucleotide selected from SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19 and SEQ ID NO: 24, wherein the polynucleotide is codon optimized for expression in a host cell.

60. An isolated, synthetic or recombinant polynucleotide comprising a codon optimized polynucleotide selected from SEQ ID NO: 25 and SEQ ID NO: 26, wherein the polynucleotide is codon optimized for expression in a host cell.

61. An isolated, synthetic or recombinant polynucleotide comprising a first codon optimized polynucleotide and a second codon optimized polynucleotide, wherein the first polynucleotide and the second polynucleotide are each codon optimized for expression in a host cell, and wherein the first codon optimized polynucleotide encodes a polypeptide with glucose permease activity and the second codon optimized polynucleotide encodes a polypeptide with glucokinase activity.

62. The isolated, synthetic or recombinant polynucleotide of claim 61, wherein the first codon optimized polynucleotide is selected from SEQ ID NO: 23, SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 19 and SEQ ID NO: 24.

63. The isolated, synthetic or recombinant polynucleotide of claim 61 or 62, wherein the second codon optimized polynucleotide is selected from SEQ ID NO: 25 and SEQ ID NO: 26.

64. The isolated, synthetic or recombinant polynucleotide of claim 61, wherein the polypeptide with glucose permease activity comprises a sequence selected from SEQ ID NO: 13, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 9 and SEQ ID NO: 14.

65. The isolated, synthetic or recombinant polynucleotide of claim 61 or 64, wherein the polypeptide with glucokinase activity comprises a sequence selected from SEQ ID NO: 15 and SEQ ID NO: 16.

66. The isolated, synthetic or recombinant polynucleotide of any of claims 59-65, wherein the host cell is E. coli and/or C. glutamicum.
FIG. 1

Non-PTS glucose transporters and glucose kinases from various species increase performance of different strain backgrounds in evaluation method.

BL1631: glucose permease from B. longum
glf: glucose permease from Z. mobilis
glpP: glucose permease from Synechocystis sp. PCC6803
SC05578: glucose permease from S. coelicolor
Insert DNA e.g., a heterologous promoter functionally linked to permease and/or glucokinase gene

FIG. 3

Cloning into Vector

Generate Insert DNA

Assemble Plasmid

Transformation and Selection

Integrate into Target Genome

Genomic DNA

Inserted DNA
Insert DNA

e.g., a heterologous promoter functionally linked to permease and/or glucokinase gene

Loop-out marker

Counter Selection

Final Genomic DNA
**INTERNATIONAL SEARCH REPORT**

**A. CLASSIFICATION OF SUBJECT MATTER**
**IPC(8) -** C12N 15/70, C12P 19/02 (2017.01)
**CPC -** C12N 15/70, C12P 19/02

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

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

See Search History Document

Electronic database consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category*</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y A</td>
<td>US 2011/0277179 A1 (Puzio et al.) 10 November 2011 (10.11.2011) para [0007], [2829], SEQ ID NO: 12063</td>
<td>61, 64, 65</td>
</tr>
<tr>
<td>Y A</td>
<td>US 2007/0118916 A1 (Puzio et al.) 24 May 2007 (24.05.2007) para [0278], [0824], SEQ ID NO: 50139</td>
<td>61, 64, 65</td>
</tr>
<tr>
<td>A P</td>
<td>US 2017/0159045 A1 (Zymergen, Inc.) 08 June 2017 (08.06.2017) para [0357], Table 1, SEQ ID NO: 1</td>
<td>1-9, 39-44, 59, 60, 62, 63</td>
</tr>
</tbody>
</table>

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

**Date of the actual completion of the international search**
31 October 2017

**Date of mailing of the international search report**
09 NOV 2017

**Name and mailing address of the ISA/US**
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450
Facsimile No. 571-273-8300

**Authorized officer:** Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)
**INTERNATIONAL SEARCH REPORT**

**Box No. II  Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [ ] Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. [ ] Claims Nos.:
   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. [ ] Claims Nos.: 10-18, 30-38, 45-58, 66
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box No. III  Observations where unity of invention is lacking (Continuation of item 3 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:
- Please see extra sheet for Box No. III Observations where unity of invention is lacking -

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. [x] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-9, 39-44, 59-65 limited to SEQ ID NOs: 1, 2, 13, 15, 23, 25

**Remark on Protest**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
</table>

- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2015)
Continuation of:
Box NO III. Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I+: Claims 1-9, 39-44, 59-65, directed to a host cell, a library or an isolated nucleotide comprising a heterologous glucose permease gene functionally linked to a first promoter polynucleotide, and a hexokinase gene functionally linked to a second promoter polynucleotide. The host cell composition will be searched to the extent that the first promoter encompasses SEQ ID NO: 1; glucose permease protein encompasses SEQ ID NO: 13; glucose permease gene encompasses SEQ ID NO: 23; the second promoter encompasses SEQ ID NO: 2; glucokinase protein encompasses SEQ ID NO: 15; glucokinase gene encompasses SEQ ID NO: 25. It is believed that claims 1-9, 39-44, 59-65 encompass this first named invention, and thus these claims will be searched without fee to the extent that they encompass SEQ ID NOs: 1, 2, 13, 15, 23, 25. Additional first and second promoter(s), glucose permease(s) and glucokinase(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected promoter(s), glucose permease(s) and glucokinase(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "*" group(s) will result in only the first claimed invention to be searched. An exemplary election would host a cell comprising the first promoter encompasses SEQ ID NO: 2; glucose permease protein encompasses SEQ ID NO: 10; glucose permease gene encompasses SEQ ID NO: 20; the second promoter encompasses SEQ ID NO: 1; glucokinase protein encompasses SEQ ID NO: 16; glucokinase gene encompasses SEQ ID NO: 26 (Claims 1-9, 39-44, 59-65).

Group II+: Claims 19-29, directed to a method for generating a microorganism capable of increased production of a biomolecule from glucose. Group II+ will be searched upon payment of additional fees. The method may be searched, for example, to the extent that the first promoter encompasses SEQ ID NO: 1; glucose permease protein encompasses SEQ ID NO: 13; glucose permease gene encompasses SEQ ID NO: 23; the second promoter encompasses SEQ ID NO: 2; glucokinase protein encompasses SEQ ID NO: 15; glucokinase gene encompasses SEQ ID NO: 25. It is believed that claims 19-29 read on this exemplary invention. Additional first and second promoter(s), glucose permease(s) and glucokinase(s) will be searched upon the payment of additional fees. Applicants must specify the claims that encompass any additionally elected promoter(s), glucose permease(s) and glucokinase(s). Failure to clearly identify how any paid additional invention fees are to be applied to the "*" group(s) will result in only the first claimed invention to be searched. Another exemplary election would be a microorganism comprising the first promoter encompasses SEQ ID NO: 2; glucose permease protein encompasses SEQ ID NO: 10; glucose permease gene encompasses SEQ ID NO: 20; the second promoter encompasses SEQ ID NO: 1; glucokinase protein encompasses SEQ ID NO: 16; glucokinase gene encompasses SEQ ID NO: 26 (Claims 19-29).

The inventions listed as Groups I+ and II+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Special Technical Features

Group I+ includes the special technical feature of a host cell, a library or isolated nucleotides comprising promoter, glucose permease, and glucokinase, not required by Group II+.

Group II+ includes the special technical feature of a method for generating a microorganism capable of increased production of a biomolecule from glucose, not required by Group I+.

The technical feature of each of the inventions listed as Groups I+ and II+ is the specific promoter, glucose permease, and glucokinase, recited therein. Each invention requires a specific set of promoter, glucose permease, and glucokinase having distinctive nucleotide (polypeptide) sequence, respectively, not required by any of the other inventions.

Common Technical Features

The inventions of Groups I+ and II+ share the technical feature of a microorganism comprising a heterologous glucose permease gene functionally linked to a first promoter polynucleotide, and a glucokinase gene functionally linked to a second promoter polynucleotide. However, these shared technical features do not represent a contribution over prior art in view of US 2009/0325248 A1 to Marx et al. (hereinafter 'Marx'). Marx teaches (see claim 1) a host cell comprising a heterologous glucose permease gene functionally linked to a first promoter polynucleotide, and a glucokinase gene functionally linked to a second promoter polynucleotide (abstract, a cell which is genetically modified in relation to its wild type.; para [0028], if the increase in enzymatic activity is brought about by increasing the expression of an enzyme, then for example...the promoter region and regulatory region or the ribosome binding site located upstream from the structural gene is mutated.; [0138]-[0141]). Genetically modified cells which...the activity of the glucose permease or enzyme combinations is increased...E23 to be a glucose transporter, preferably a glucose transporter encoded by a gene selected from...a glucose permease (2.7.1.69), E24 to be a glucokinase (2.7.1.2)).

Marx teaches (see claim 59) an isolated polynucleotide encoding glucose permease and glucokinase (para [0140], [0141]). Marx does not specifically teach codon optimized polynucleotide, however, it is well known to one of ordinary skill in the art that codon optimization of a polynucleotide for expressing in a specific host is needed to ensure that the host cell expresses said heterologous nucleic acid. Thus, one of ordinary skill in the art would have codon optimized the glucose permease and glucokinase of Marx.

The inventions of Groups I+ and II+ share the technical feature of a library of glucose permease genes. US 2009/0221442 A1 to Dower et al. (hereinafter 'Dower') teaches a library of glucose permease genes (para [0021], screen for new transport proteins using cells that have been transformed with a DNA library.; [0049], Some examples of transporter proteins...facilitated glucose transporter;...sodium/glucose transporter; see instant Specification, para [0058], the term 'glucose permease' can refer to any transporter (e.g., myo-inositol transporter and/or glucose permease) that exhibits an affinity for glucose and subsequently facilitates its transport across the cell membrane of a host cell.).

****continued on next sheet****
The inventions of Groups I+ share the technical feature of a method for generating a microorganism capable of increased production of a biomolecule from glucose. Mark teaches (abstract, methods for producing a genetically modified cell, to the genetically modified cells obtainable by this method, to methods for producing 3-hydroxypropionic acid, to a method for producing acrylic acid, to a method for producing polyacrylates; [0126]. Cells of this type can be cultured in a nutrient medium which contains carbohydrates such as, for example, glucose as carbon source.).

As said technical features were known in the art at the time of the invention, these cannot be considered special technical feature that would otherwise unify the groups.

Groups I+ and II+ therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.