

ENHANCING MICROBIAL METABOLISM OF C5 ORGANIC CARBON

5 CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of priority to U.S. Provisional Application No. 62/191,983, filed July 13, 2015, and U.S. Provisional Application No. 62/354,444, filed June 24, 2016, which are incorporated by reference herein in their entireties.

10 BACKGROUND OF THE INVENTION

Heterotrophic fermentation of microorganisms is an efficient way of generating high value oil and biomass products. Under certain cultivation conditions, microorganisms synthesize intracellular oil, which can be extracted and used to produce fuel (e.g., biodiesel, bio-jetfuel, and the like) and nutritional lipids (e.g., polyunsaturated fatty acids such as DHA, EPA, and DPA). The biomass of some microorganisms is of great nutritional value due to high polyunsaturated fatty acid (PUFA) and protein content, and can be used as a nutritional supplement for animal feed. Thraustochytrids are eukaryotic, single-cell, microorganisms which can be used in such fermentation processes to produce lipids. Heterotrophic fermentations with Thraustochytrids convert organic carbon provided in the growth medium to lipids, which are harvested from the biomass at the end of the fermentation process. However, existing microorganism fermentations use mainly expensive carbohydrates, such as glucose, as the carbon source.

BRIEF SUMMARY OF THE INVENTION

Provided herein are recombinant microorganisms having two or more copies of a nucleic acid sequence encoding xylose isomerase, wherein the nucleic acid encoding the xylose isomerase is an exogenous nucleic acid. Optionally, the recombinant microorganisms include at least one nucleic acid sequence encoding a xylulose kinase and/or at least one nucleic acid sequence encoding a xylose transporter. The provided recombinant microorganisms are capable of growing on xylose as a carbon source.

30

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic of the xylose metabolism pathway.

Figure 2 is a graph showing expression of xylose isomerase in WT ONC-T18 during cycles of glucose starvation.

5 Figure 3 is a graph showing expression of the putative xylulose kinase in WT ONC-T18 during cycles of glucose starvation.

Figure 4 is a schematic showing an alpha-tubulin *ble*-isomerase plasmid construct.

Figure 5 is a schematic showing an alpha-tubulin *hygro-xyIB* plasmid construct.

10 Figure 6 is a schematic showing a nucleic acid construct having an alpha-tubulin promoter a *ble* sequence a 2A sequence an xylose isomerase sequence and an alpha-tubulin terminator.

Figure 7 is an image of a Southern blot to probe the xylose isomerase His-tagged gene within recombinant ONC-T18 strains "6" and "16".

15 Figure 8 is a graph showing the qPCR determination of the number of xylose isomerase His-tagged gene insertions in recombinant ONC-T18 strains.

Figure 9 is an image of a Southern blot to probe the *xyIB* gene within recombinant ONC-T18 strains containing both xylose isomerase and xylulose kinase referred to in the graph as "7-3" and "7-7".

20 Figure 10 is a graph of qPCR determination of the number of *xyIB* gene insertions in recombinant 7-3 and 7-7 ONC-T18 strains.

Figure 11 is a graph showing the expression of the xylose isomerase gene transcript in recombinant ONC-T18 strains "6" and "16."

Figure 12 is a graph showing the *in vitro* xylose isomerase activity in Wt ONC-T18 and recombinant ONC-T18 strains "6" and "16."

25 Figure 13 is a graph showing the combined xylose isomerase and xylulose kinase activity *in vitro* of recombinant ONC-T18 strain "16" encoding only xylose isomerase and recombinant ONC-T18 strains "7-3" and "7-7" encoding xylose isomerase and xylulose kinase.

30 Figures 14A and 14B are graphs showing xylose uptake improvement and decreased xylitol production in recombinant ONC-T18 strain "16" (squares). The Wild Type (WT) strain is represented by diamonds.

Figures 15A and 15B are graphs showing xylose usage improvement and decreased xylitol production in recombinant ONC-T18 strain "16" (squares) and recombinant ONC-T18

strains “7-3” (triangles) and “7-7” (asterisks). The Wild Type (WT) strain is represented by diamonds.

Figure 16 is a graph showing accumulation of xylitol during a glucose:xylose fermentation with recombinant ONC-T18 strain “16” and recombinant ONC-T18 strain “7-7.”

Figure 17 is a schematic of different versions of the constructs used for transformation of ONC-T18.

Figure 18 is a graph showing the alignment of the *xylB* sequence from *E. coli* (SEQ ID NO:20) with the codon optimized version of *E. coli xylB* (SEQ ID NO:5).

Figures 19A, 19B, and 19C are graphs showing xylose usage (Fig. 19A), glucose usage (Fig. 19B) and percent xylitol made (Fig. 19C) in strains comprising xylose isomerase, xylulose kinase and the sugar transporter Gxs1. WT is wild-type; IsoHis XylB “7-7” contains the xylose isomerase and *xylB* sequences, 36-2, 36-9 and 36-16 are transformants containing Gxs1, xylose isomerase and the *xylB* sequences (xylulose kinase).

Figures 20A and 20B are graphs showing the impact of temperature incubation on the activity of isomerase from T18 (Figure 20A) and *E. coli* (Figures 20B) with xylose (diamond) and xylulose (square).

Figures 21A and 21B are graphs showing dose dependency of isomerase from T18 (Figure 21A) and *E. coli* (Figure 21B) with xylose (diamond) and xylulose (square).

Figures 22A and 22B are graphs showing xylose use (Figure 22A) and decreased xylitol production (Figure 22B) in a T18B strain engineered with xylose isomerases (“16” (squares), “B” (x), and “6” (crosses)). Figures 22C (xylose) and 22D (xylitol production) show the same data expressed relative to wild type (diamonds) at 4 (gray) and 7 (black) days.

Figures 23A and 23B are graphs showing xylose use and decreased xylitol production in a T18B strain engineered with a xylose isomerase “16” (squares) and strains engineered to express a xylose isomerase and xylulose kinase “7-7” (x) and “7-3” (triangles). Figures 23C (xylose) and 23D (xylitol production) show the same data relative to wild type (diamonds) at 9 (gray) and 11 (black) days.

Figure 24 is a graph showing improved xylose usage and decreased xylitol production in a T18B strain engineered to express a xylose isomerase and xylulose kinase “7-7” in fermentation. The wild type strain is represented by diamonds and the dotted line and the strain “7-7” is represented by circles.

Figure 25 is a schematic showing α -tubulin aspTx-neo and α -tubulin gxs1-neo constructs.

Figure 26A is an image of a Southern blot to probe the Gxs1 gene within “7-7” T18B strains engineered with the xylose transporter Gxs1. Figure 26B is an image of a Southern blot to probe the AspTx gene within “7-7” T18B strains engineered with the xylose transporter AspTx.

5 Figures 27A is a graph showing the use of xylose in T18 strains engineered with a xylose isomerase, a xylulose kinase and either the Gxs1 transporter (triangles) or AspTx transporter (circles). Strain “7-7” is represented by diamonds. Figure 27B is a bar graph of the ratio of xylitol production versus xylose use for each of the 3 modified strains. Figure 27C is a bar graph showing xylose use relative to strain “7-7.” Figures 27D is a bar graph
10 showing xylitol production made relative to strain “7-7.”

 Figures 28 is a graph showing growth of wild type (WT) (diamonds), isohis strain “16” (squares), strain “7-7” (x), and transporter strains Gxs1 (asterisks) and AspTx (triangles) in media containing xylose as sole carbon source.

 Figure 29A is a graph showing remaining glucose in alternative feedstock containing
15 glucose and xylose during growth of WT (squares), strain “7-7” (triangles), and transporter strains Gxs1 (asterisks) and AspTx (crosses). Figure 29B is a graph showing xylose remaining and xylitol produced over time when WT (squares) strain “7-7” (triangles) and transporter strains Gxs1 (asterisks) and AspTx (crosses) are grown on alternative feedstock containing glucose and xylose.

20 **DETAILED DESCRIPTION OF THE INVENTION**

 Microorganisms such as Thraustochytrids encode genes required for the metabolism of xylose. However, the microorganism’s innate metabolic pathways produce a large amount of the sugar alcohol, xylitol, which is secreted and potentially hinders growth of the microorganisms (see Figure 14, WT). Furthermore, carbon atoms sequestered into xylitol are
25 atoms that are diverted away from the target product in this process, namely, lipid production. In nature, two xylose metabolism pathways exist, the xylose reductase/xylitol dehydrogenase pathway and the xylose isomerase/xylulose kinase pathway (Figure 1). Thraustochytrids have genes that encode proteins active in both pathways; however, the former pathway appears to be dominant as evidenced by a build-up of xylitol when grown in a xylose
30 medium. In other organisms, the build-up of xylitol has been shown to be due to a redox co-factor imbalance required for xylose reductase/xylitol dehydrogenase pathway. Since the isomerase/kinase pathway does not depend on redox co-factors, over-expression of the isomerase gene removes co-factor dependence in the conversion of xylose to xylulose. As shown herein, transcriptomic studies with ONC-T18 showed that its xylose isomerase and

putative xylulose kinase genes are mostly expressed during glucose starvation (Figure 2 and Figure 3); whereas, the putatively identified genes encoding for the xylose reductase and xylitol dehydrogenase are constitutively expressed. To increase the expression of the isomerase and kinase throughout all growth stages, microorganisms were engineered to include ONC-T18 isomerase gene and an *E. coli* xylulose kinase gene (*xylB*) such that they are under the control of the constitutively expressed promoter and terminator, e.g., an α -tubulin promoter and terminator. Optionally, the genes can be under the control of a inducible promoter and/or terminator.

The provided recombinant microorganisms demonstrate a level of control of the amount of expression of a gene of interest via the number of integrated transgene copies. As shown in the examples below, a recombinant ONC-T18 strain (Iso-His #16) harbouring eight (8) transgene copies demonstrates higher levels of xylose isomerase transcript expression, enzyme activity and xylose metabolism than a strain harbouring a single copy of the transgene (Iso-His #6). When Iso-His #16 was further modified to incorporate the *xylB* gene, a similar phenomenon is observed. Multiple copies of the *xylB* gene conferred greater enzyme activity and xylose metabolism productivity compared to single insertions. Thus, unexpectedly, it was not only necessary to recreate a xylose metabolism pathway, but to do so with multiple copies of the necessary transgenes. It was not anticipated that the Thraustochytrid genome could accommodate multiple transgene copies and remain viable; therefore, it was not expected to observe such variability in expression levels amongst transformant strains. However, as provided herein, recombinant microorganisms can be produced that allow for controlled expression levels of transgenes indirectly by selecting among transformant strains that possess a transgene copy number "tailored" to a particular expression level optimized for the metabolic engineering of a particular pathway, e.g., the xylose pathway.

Provided herein are nucleic acids encoding one or more genes involved in xylose metabolism. The present application provides recombinant microorganisms, methods for making the microorganisms, and methods for producing oil using the microorganisms that are capable of metabolizing xylose. Specifically, provided herein are nucleic acids and polypeptides encoding xylose isomerase, xylulose kinase and xylose transporters for modifying microorganisms to be capable of metabolizing xylose and/or growing on xylose as the sole carbon source. Thus, provided are nucleic acids encoding a xylose isomerase. The nucleic acid sequences can be endogenous or heterologous to the microorganism. Exemplary nucleic acids sequences of xylose isomerases include, but are not limited to, those from

Piromyces sp., *Streptococcus sp.*, and Thraustochytrids. For example, exemplary nucleic acid sequences encoding xylose isomerases include, but are not limited to, SEQ ID NO:2 and SEQ ID NO:15; and exemplary polypeptide sequences of xylose isomerase include, but are not limited to, SEQ ID NO:16. Exemplary nucleic acids sequences of xylulose kinases include, but are not limited to, those from *E. coli*, *Piromyces sp.*, *Saccharomyces sp.*, and *Pichia sp.* For example, exemplary nucleic acid sequences encoding xylulose kinases include, but are not limited to, SEQ ID NO:5, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 and SEQ ID NO:20. Exemplary nucleic acid sequences encoding sugar transporters, e.g., xylose transporters, include, but are not limited to, those from *Aspergillus sp.*, Gfx1, Gxs1 and Sut1. For example, exemplary nucleic acid sequences encoding xylose transporters include, but are not limited to, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, and SEQ ID NO:24.

Nucleic acid, as used herein, refers to deoxyribonucleotides or ribonucleotides and polymers and complements thereof. The term includes deoxyribonucleotides or ribonucleotides in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs). Unless otherwise indicated, conservatively modified variants of nucleic acid sequences (e.g., degenerate codon substitutions) and complementary sequences can be used in place of a particular nucleic acid sequence recited herein. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al., Nucleic Acid Res. 19:5081 (1991); Ohtsuka et al., J. Biol. Chem. 260:2605-2608 (1985); Rossolini et al., Mol. Cell. Probes 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

A nucleic acid is operably linked when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA that encodes a presequence or secretory leader is operably linked to DNA that encodes a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a

ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, operably linked means that the DNA sequences being linked are near each other, and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. For example, a nucleic acid sequence that
5 is operably linked to a second nucleic acid sequence is covalently linked, either directly or indirectly, to such second sequence, although any effective three-dimensional association is acceptable. A single nucleic acid sequence can be operably linked to multiple other sequences. For example, a single promoter can direct transcription of multiple RNA species. Linking can be accomplished by ligation at convenient restriction sites. If such sites do not
10 exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

The terms identical or percent identity, in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (i.e.,
15 about 60% identity, preferably 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher identity over a specified region, when compared and aligned for maximum correspondence over a comparison window or designated region) as measured using a BLAST or BLAST 2.0 sequence comparison algorithms with default parameters described below, or by manual alignment and visual inspection (see, e.g., NCBI
20 web site or the like). Such sequences are then said to be substantially identical. This definition also refers to, or may be applied to, the complement of a test sequence. The definition also includes sequences that have deletions and/or additions, as well as those that have substitutions. As described below, the preferred algorithms can account for gaps and the like. Preferably, identity exists over a region that is at least about 25 amino acids or
25 nucleotides in length, or more preferably over a region that is 50-100 amino acids or nucleotides in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if
30 necessary, and sequence algorithm program parameters are designated. Preferably, default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters.

A comparison window, as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981); by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970); by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988); by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI); or by manual alignment and visual inspection (see, e.g., *Current Protocols in Molecular Biology* (Ausubel et al., eds. 1995 supplement)).

A preferred example of an algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., *Nuc. Acids Res.* 25:3389-3402 (1977), and Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), respectively. BLAST and BLAST 2.0 are used, with the parameters described herein, to determine percent sequence identity for nucleic acids or proteins. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information, as known in the art. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of a selected length (W) in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the

sensitivity and speed of the alignment. The Expectation value (E) represents the number of different alignments with scores equivalent to or better than what is expected to occur in a database search by chance. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1989)), alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The term polypeptide, as used herein, generally has its art-recognized meaning of a polymer of at least three amino acids and is intended to include peptides and proteins. However, the term is also used to refer to specific functional classes of polypeptides, such as, for example, desaturases, elongases, etc. For each such class, the present disclosure provides several examples of known sequences of such polypeptides. Those of ordinary skill in the art will appreciate, however, that the term polypeptide is intended to be sufficiently general as to encompass not only polypeptides having the complete sequence recited herein (or in a reference or database specifically mentioned herein), but also to encompass polypeptides that represent functional fragments (i.e., fragments retaining at least one activity) of such complete polypeptides. Moreover, those in the art understand that protein sequences generally tolerate some substitution without destroying activity. Thus, any polypeptide that retains activity and shares at least about 30-40% overall sequence identity, often greater than about 50%, 60%, 70%, or 80%, and further usually including at least one region of much higher identity, often greater than 90% or even 95%, 96%, 97%, 98%, or 99% in one or more highly conserved regions, usually encompassing at least 3-4 and often up to 20 or more amino acids, with another polypeptide of the same class, is encompassed within the relevant term polypeptide as used herein. Those in the art can determine other regions of similarity and/or identity by analysis of the sequences of various polypeptides described herein. As is known by those in the art, a variety of strategies are known, and tools are available, for performing comparisons of amino acid or nucleotide sequences in order to assess degrees of identity and/or similarity. These strategies include, for example, manual alignment, computer assisted sequence alignment and combinations thereof. A number of algorithms (which are generally computer implemented) for performing sequence alignment are widely available, or can be produced by one of skill in the art. Representative algorithms include, e.g., the local homology algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2: 482); the homology alignment algorithm of Needleman and Wunsch (J. Mol. Biol., 1970, 48: 443); the search for

similarity method of Pearson and Lipman (Proc. Natl. Acad. Sci. (USA), 1988, 85: 2444); and/or by computerized implementations of these algorithms (e.g., GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.). Readily available computer programs incorporating
5 such algorithms include, for example, BLASTN, BLASTP, Gapped BLAST, PILEUP, CLUSTALW, etc. When utilizing BLAST and Gapped BLAST programs, default parameters of the respective programs may be used. Alternatively, the practitioner may use non-default parameters depending on his or her experimental and/or other requirements (see for example, the Web site having URL www.ncbi.nlm.nih.gov).

10 As discussed above, the nucleic acids encoding the xylose transporter, xylulose kinase and xylose isomerase, can be linked to a promoter and/or terminator. Examples of promoters and terminators include, but are not limited to, tubulin promoters and terminators. By way of example, the promoter is a tubulin promoter, e.g., an alpha-tubulin promoter. Optionally, the promoter is at least 80% identical to SEQ ID NO:25 or SEQ ID NO:26. Optionally, the
15 terminator is a tubulin terminator. Optionally, the terminator is at least 80% identical to SEQ ID NO:27, SEQ ID NO:28, or SEQ ID NO:30.

As used herein, the terms promoter, promoter element, and regulatory sequence refer to a polynucleotide that regulates expression of a selected polynucleotide sequence operably linked to the promoter, and that effects expression of the selected polynucleotide sequence in
20 cells. The term *Thraustochytrium* promoter, as used herein, refers to a promoter that functions in a *Thraustochytrium* cell. In some embodiments, a promoter element is or comprises untranslated regions (UTR) in a position 5' of coding sequences. 5' UTRs form part of the mRNA transcript and so are an integral part of protein expression in eukaryotic organisms. Following transcription 5'UTRs can regulate protein expression at both the
25 transcription and translation levels.

As used herein, the term terminator refers to a polynucleotide that abrogates expression of, targets for maturation (e.g., adding a polyA tail), or imparts mRNA stability to a selected polynucleotide sequence operably linked to the terminator in cells. A terminator sequence may be downstream of a stop codon in a gene. The term *Thraustochytrium*
30 terminator, as used herein, refers to a terminator that functions in a *Thraustochytrium* cell. Provided herein are also nucleic acid constructs that include nucleic acid sequences encoding xylose isomerase, xylulose kinase and xylose transporter as well as promoters, terminators, selectable markers, 2A peptides or any combination thereof. By way of example, provided is a first nucleic acid construct including a promoter, a selectable marker, a nucleic acid

sequence encoding a 2A peptide, a nucleic acid sequence encoding a xylose isomerase, and a terminator. Also provided is a second nucleic acid construct including a promoter, selectable marker, a nucleic acid sequence encoding a 2A peptide, a nucleic acid sequence encoding a xylulose kinase, and a terminator. Further provided is a third nucleic acid construct including
5 a promoter, a nucleic acid sequence encoding a xylose transporter, a nucleic acid sequence encoding a 2A peptide, a selectable marker, and a terminator. These constructs are exemplary and the nucleic acid sequences encoding the xylose isomerase, xylulose kinase and xylose transporter can be included on the same construct under control of the same or different promoters. Optionally, each of the nucleic acid sequences encoding the xylose
10 isomerase, xylulose kinase and xylose transporter are on the same construct and are separated by 2A polypeptide sequences, e.g., as shown in SEQ ID NO:6. Thus, by way of example, a nucleic acid construct can include a tubulin promoter, a nucleic acid sequences encoding a xylose isomerase, xylulose kinase, and xylose transporter separated by a nucleic acid sequence encoding SEQ ID NO:6, a tubulin terminator and a selectable marker. Optionally,
15 the selectable marker is the ble gene. Optionally, the selectable marker comprises SEQ ID NO:29.

The phrase selectable marker, as used herein, refers either to a nucleotide sequence, e.g., a gene, that encodes a product (polypeptide) that allows for selection, or to the gene product (e.g., polypeptide) itself. The term selectable marker is used herein as it is generally
20 understood in the art and refers to a marker whose presence within a cell or organism confers a significant growth or survival advantage or disadvantage on the cell or organism under certain defined culture conditions (selective conditions). For example, the conditions may be the presence or absence of a particular compound or a particular environmental condition such as increased temperature, increased radiation, presence of a compound that is toxic in
25 the absence of the marker, etc. The presence or absence of such compound(s) or environmental condition(s) is referred to as a selective condition or selective conditions. By growth advantage is meant either enhanced viability (e.g., cells or organisms with the growth advantage have an increased life span, on average, relative to otherwise identical cells), increased rate of proliferation (also referred to herein as growth rate) relative to otherwise
30 identical cells or organisms, or both. In general, a population of cells having a growth advantage will exhibit fewer dead or nonviable cells and/or a greater rate of cell proliferation than a population of otherwise identical cells lacking the growth advantage. Although typically a selectable marker will confer a growth advantage on a cell, certain selectable markers confer a growth disadvantage on a cell, e.g., they make the cell more susceptible to

the deleterious effects of certain compounds or environmental conditions than otherwise identical cells not expressing the marker. Antibiotic resistance markers are a non-limiting example of a class of selectable marker that can be used to select cells that express the marker. In the presence of an appropriate concentration of antibiotic (selective conditions),
5 such a marker confers a growth advantage on a cell that expresses the marker. Thus, cells that express the antibiotic resistance marker are able to survive and/or proliferate in the presence of the antibiotic while cells that do not express the antibiotic resistance marker are not able to survive and/or are unable to proliferate in the presence of the antibiotic.

Examples of selectable markers include common bacterial antibiotics, such as but not
10 limited to ampicillin, kanamycin and chloramphenicol, as well as selective compounds known to function in microalgae; examples include *rrnS* and *AadA* (Aminoglycoside 3'-adenylyltransferase), which may be isolated from *E. coli* plasmid R538-1, conferring resistance to spectinomycin and streptomycin, respectively in *E. coli* and some microalgae (Hollingshead and Vapnek, Plasmid 13:17-30, 1985; Meslet-Cladière and Vallon, Eukaryot
15 Cell. 10(12):1670-8 2011). Another example is the 23S RNA protein, *rrnL*, which confers resistance to erythromycin (Newman, Boynton et al., Genetics, 126:875-888 1990; Roffey, Golbeck et al., Proc. Natl Acad. Sci. USA, 88:9122-9126 1991). Another example is *Ble*, a GC rich gene isolated from *Streptoalloteichus hindustanus* that confers resistance to zeocin (Stevens, Purton et al., Mol. Gen. Genet., 251:23-30 1996). *Aph7* is yet another example,
20 which is a *Streptomyces hygroscopicus*-derived aminoglycoside phosphotransferase gene that confers resistance to hygromycin B (Berthold, Schmitt et al., Protist 153(4):401-412 2002). Additional examples include: *AphVIII*, a *Streptomyces rimosus* derived aminoglycoside 3'-phosphotransferase type VIII that confers resistance to Paromycin in *E. coli* and some microalgae (Sizova, Lapina et al., Gene 181(1-2):13-18 1996; Sizova, Fuhrmann et al., Gene
25 277(1-2):221-229 2001); *Nat* & *Sat-1*, which encode nourseothricin acetyl transferase from *Streptomyces noursei* and streptothricin acetyl transferase from *E. coli*, which confer resistance to nourseothricin (Zaslavskaja, Lippmeier et al., Journal of Phycology 36(2):379-386, 2000); *Neo*, an aminoglycoside 3'-phosphotransferase, conferring resistance to the aminoglycosides; kanamycin, neomycin, and the analog G418 (Hasnain, Manavathu et al.,
30 Molecular and Cellular Biology 5(12):3647-3650, 1985); and *Cry1*, a ribosomal protein S14 that confers resistance to emetine (Nelson, Savereide et al., Molecular and Cellular Biology 14(6):4011-4019, 1994).

Other selectable markers include nutritional markers, also referred to as auto- or auxotrophic markers. These include photoautotrophy markers that impose selection based on the

restoration of photosynthetic activity within a photosynthetic organism. Photoautotrophic markers include, but are not limited to, AtpB, TscA, PetB, NifH, psaA and psaB (Boynton, Gillham et al., Science 240(4858):1534-1538 1988; Goldschmidt-Clermont, Nucleic Acids Research 19(15):4083-4089, 1991; Kindle, Richards et al., PNAS, 88(5):1721-1725, 1991; 5 Redding, MacMillan et al., EMBO J 17(1):50-60, 1998; Cheng, Day et al., Biochemical and Biophysical Research Communications 329(3):966-975, 2005). Alternative or additional nutritional markers include ARG7, which encodes argininosuccinate lyase, a critical step in arginine biosynthesis (Debuchy, Purton et al., EMBO J 8(10):2803-2809, 1989); NIT1, which encodes a nitrate reductase essential to nitrogen metabolism (Fernández, Schnell et al., 10 PNAS, 86(17):6449-6453, 1989); THI10, which is essential to thiamine biosynthesis (Ferris, Genetics 141(2):543-549, 1995); and NIC1, which catalyzes an essential step in nicotinamide biosynthesis (Ferris, Genetics 141(2):543-549, 1995). Such markers are generally enzymes that function in a biosynthetic pathway to produce a compound that is needed for cell growth or survival. In general, under nonselective conditions, the required compound is present in 15 the environment or is produced by an alternative pathway in the cell. Under selective conditions, functioning of the biosynthetic pathway, in which the marker is involved, is needed to produce the compound.

The phrase selection agent, as used herein refers to an agent that introduces a selective pressure on a cell or populations of cells either in favor of or against the cell or population of 20 cells that bear a selectable marker. For example, , the selection agent is an antibiotic and the selectable marker is an antibiotic resistance gene. Optionally, zeocin is used as the selection agent.

Suitable microorganisms that can be transformed with the provided nucleic acids encoding the genes involved in xylose metabolism and nucleic acid constructs containing the 25 same include, but are not limited to, algae (e.g., microalgae), fungi (including yeast), bacteria, or protists. Optionally, the microorganism includes Thraustochytrids of the order Thraustochytriales, more specifically Thraustochytriales of the genus *Thraustochytrium*. Optionally, the population of microorganisms includes Thraustochytriales as described in U.S. Patent Nos. 5,340,594 and 5,340,742, which are incorporated herein by reference in 30 their entireties. The microorganism can be a Thraustochytrium species, such as the Thraustochytrium species deposited as ATCC Accession No. PTA-6245 (i.e., ONC-T18) as described in U.S. Patent No. 8,163,515, which is incorporated by reference herein in its entirety. Thus, the microorganism can have an 18s rRNA sequence that is at least 95%, 96%,

97%, 98%, 99%, 99.1%, 99.2%, 99.3%, 99.4%, 99.5%, 99.6%, 99.7%, 99.8%, 99.9% or more (e.g., including 100%) identical to SEQ ID NO:1.

Microalgae are acknowledged in the field to represent a diverse group of organisms. For the purpose of this document, the term microalgae will be used to describe unicellular
 5 microorganisms derived from aquatic and/or terrestrial environments (some cyanobacteria are terrestrial/soil dwelling). Aquatic environments extend from oceanic environments to freshwater lakes and rivers, and also include brackish environments such as estuaries and river mouths. Microalgae can be photosynthetic; optionally, microalgae are heterotrophic. Microalgae can be of eukaryotic nature or of prokaryotic nature. Microalgae can be non-
 10 motile or motile.

The term thraustochytrid, as used herein, refers to any member of the order Thraustochytriales, which includes the family Thraustochytriaceae. Strains described as thraustochytrids include the following organisms: Order: Thraustochytriales; Family: Thraustochytriaceae; Genera: *Thraustochytrium* (Species: sp., *arudimentale*, *aureum*,
 15 *benthicola*, *globosum*, *kinnei*, *motivum*, *multirudimentale*, *pachydermum*, *proliferum*, *roseum*, *striatum*), *Ulkenia* (Species: sp., *amoeboidea*, *keruelensis*, *minuta*, *profunda*, *radiata*, *sailens*, *sarkariana*, *schizochytrids*, *visurgensis*, *yorkensis*), *Schizochytrium* (Species: sp., *aggregatum*, *limnaceum*, *mangrovei*, *minutum*, *octosporuni*), *Japonochytrium* (Species: sp., *marinum*), *Aplanochytrium* (Species: sp., *haliotidis*, *keruelensis*, *profunda*, *stocchinoi*),
 20 *Althornia* (Species: sp., *crouchii*), or *Elina* (Species: sp., *marisalba*, *sinorifica*). Species described within *Ulkenia* will be considered to be members of the genus *Thraustochytrium*. Strains described as being within the genus *Thraustochytrium* may share traits in common with and also be described as falling within the genus *Schizochytrium*. For example, in some taxonomic classifications ONC-T18 may be considered within the genus *Thraustochytrium*,
 25 while in other classifications it may be described as within the genus *Schizochytrium* because it comprises traits indicative of both genera.

The term transformation, as used herein refers to a process by which an exogenous or heterologous nucleic acid molecule (e.g., a vector or recombinant nucleic acid molecule) is introduced into a recipient cell or microorganism. The exogenous or heterologous nucleic
 30 acid molecule may or may not be integrated into (i.e., covalently linked to) chromosomal DNA making up the genome of the host cell or microorganism. For example, the exogenous or heterologous polynucleotide may be maintained on an episomal element, such as a plasmid. Alternatively or additionally, the exogenous or heterologous polynucleotide may become integrated into a chromosome so that it is inherited by daughter cells through

chromosomal replication. Methods for transformation include, but are not limited to, calcium phosphate precipitation; Ca^{2+} treatment; fusion of recipient cells with bacterial protoplasts containing the recombinant nucleic acid; treatment of the recipient cells with liposomes containing the recombinant nucleic acid; DEAE dextran; fusion using polyethylene glycol (PEG); electroporation; magnetoporation; biolistic delivery; retroviral infection; lipofection; and micro-injection of DNA directly into cells.

The term transformed, as used in reference to cells, refers to cells that have undergone transformation as described herein such that the cells carry exogenous or heterologous genetic material (e.g., a recombinant nucleic acid). The term transformed can also or alternatively be used to refer to microorganisms, strains of microorganisms, tissues, organisms, etc. that contain exogenous or heterologous genetic material.

The term introduce, as used herein with reference to introduction of a nucleic acid into a cell or organism, is intended to have its broadest meaning and to encompass introduction, for example by transformation methods (e.g., calcium-chloride-mediated transformation, electroporation, particle bombardment), and also introduction by other methods including transduction, conjugation, and mating. Optionally, a construct is utilized to introduce a nucleic acid into a cell or organism.

The microorganisms for use in the methods described herein can produce a variety of lipid compounds. As used herein, the term lipid includes phospholipids, free fatty acids, esters of fatty acids, triacylglycerols, sterols and sterol esters, carotenoids, xanthophylls (e.g., oxycarotenoids), hydrocarbons, and other lipids known to one of ordinary skill in the art. Optionally, the lipid compounds include unsaturated lipids. The unsaturated lipids can include polyunsaturated lipids (i.e., lipids containing at least 2 unsaturated carbon-carbon bonds, e.g., double bonds) or highly unsaturated lipids (i.e., lipids containing 4 or more unsaturated carbon-carbon bonds). Examples of unsaturated lipids include omega-3 and/or omega-6 polyunsaturated fatty acids, such as docosahexaenoic acid (i.e., DHA), eicosapentaenoic acid (i.e., EPA), and other naturally occurring unsaturated, polyunsaturated and highly unsaturated compounds.

Provided herein are recombinant microorganisms engineered to express polypeptides for metabolizing C5 carbon sugars such as xylose. Specifically, provided is a recombinant microorganism having one or more copies of a nucleic acid sequence encoding xylose isomerase, wherein the nucleic acid encoding xylose isomerase is a exogenous nucleic acid. Optionally, the recombinant microorganism comprises two or more copies of the nucleic acid sequence encoding xylose isomerase. Optionally, the recombinant

microorganisms also contains one or two copies of an endogenous nucleic acid sequence encoding xylose isomerase. By way of example, the recombinant microorganisms can contain one or two copies of an endogenous nucleic acid sequence encoding xylose isomerase and one copy of an exogenous nucleic acid sequence encoding xylose isomerase. Optionally, 5 the recombinant microorganism includes three copies of a nucleic acid sequence encoding xylose isomerase, one being exogenously introduced and the other two being endogenous. The term recombinant when used with reference to a cell, nucleic acid, polypeptide, vector, or the like indicates that the cell, nucleic acid, polypeptide, vector or the like has been modified by or is the result of laboratory methods and is non-naturally occurring. Thus, for 10 example, recombinant microorganisms include microorganisms produced by or modified by laboratory methods, e.g., transformation methods for introducing nucleic acids into the microorganism. Recombinant microorganisms can include nucleic acid sequences not found within the native (non-recombinant) form of the microorganisms or can include nucleic acid sequences that have been modified, e.g., linked to a non-native promoter.

15 As used herein, the term exogenous refers to a substance, such as a nucleic acid (e.g., nucleic acids including regulatory sequences and/or genes) or polypeptide, that is artificially introduced into a cell or organism and/or does not naturally occur in the cell in which it is present. In other words, the substance, such as nucleic acid or polypeptide, originates from outside a cell or organism into which it is introduced. An exogenous nucleic acid can have a 20 nucleotide sequence that is identical to that of a nucleic acid naturally present in the cell. For example, a *Thraustochytrid* cell can be engineered to include a nucleic acid having a *Thraustochytrid* or *Thraustochytrium* regulatory sequence. In a particular example, an endogenous *Thraustochytrid* or *Thraustochytrium* regulatory sequence is operably linked to a gene with which the regulatory sequence is not involved under natural conditions. Although 25 the *Thraustochytrid* or *Thraustochytrium* regulatory sequence may naturally occur in the host cell, the introduced nucleic acid is exogenous according to the present disclosure. An exogenous nucleic acid can have a nucleotide sequence that is different from that of any nucleic acid that is naturally present in the cell. For example, the exogenous nucleic acid can be a heterologous nucleic acid, i.e., a nucleic acid from a different species or organism. Thus, 30 an exogenous nucleic acid can have a nucleic acid sequence that is identical to that of a nucleic acid that is naturally found in a source organism but that is different from the cell into which the exogenous nucleic acid is introduced. As used herein, the term endogenous, refers to a nucleic acid sequence that is native to a cell. As used herein, the term heterologous refers to a nucleic acid sequence that is not native to a cell, i.e., is from a different organism

than the cell. The terms exogenous and endogenous or heterologous are not mutually exclusive. Thus, a nucleic acid sequence can be exogenous and endogenous, meaning the nucleic acid sequence can be introduced into a cell but have a sequence that is the same as or similar to the sequence of a nucleic acid naturally present in the cell. Similarly, a nucleic acid sequence can be exogenous and heterologous meaning the nucleic acid sequence can be introduced into a cell but have a sequence that is not native to the cell, e.g., a sequence from a different organism.

As discussed above, the provided recombinant microorganisms contain at least two copies of a nucleic acid sequence encoding a xylose isomerase. The provided microorganisms optionally also contain at least one nucleic acid sequence encoding a xylulose kinase. Optionally, the recombinant microorganisms comprise at least one nucleic acid sequence encoding a xylose transporter. The nucleic acid sequences encoding the xylose isomerase, xylulose kinase, and/or xylose transporter are, optionally, exogenous nucleic acid sequences. Optionally, the nucleic acid sequence encoding the xylose isomerase is an endogenous nucleic acid sequence. Optionally, the nucleic acid sequence encoding the xylulose kinase and/or xylose transporter is a heterologous nucleic acid. Optionally, the microorganism contains at least two copies of a nucleic acid sequence encoding a xylose isomerase, at least two copies of a nucleic acid sequence encoding a xylulose kinase, and at least one nucleic acid sequence encoding a xylose transporter. Optionally, the heterologous nucleic acid sequence encoding the xylose isomerase is at least 90% identical to SEQ ID NO:2. Optionally, the heterologous nucleic acid sequence encoding the xylulose kinase is at least 90% identical to SEQ ID NO:5. As noted above, optionally, the nucleic acid encoding the xylose transporter is a heterologous nucleic acid. Optionally, the xylose transporter encoded by the heterologous nucleic acid is GXS1 from *Candida intermedia*. Optionally, the heterologous nucleic acid sequence encoding the xylose transporter is at least 90% identical to SEQ ID NO:23.

The provided recombinant microorganisms not only contain nucleic acid sequences encoding genes involved in xylose metabolism, they can include multiple copies of such sequences. Thus, the microorganism comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 copies of the nucleic acid sequence encoding xylose isomerase. Optionally, the microorganism comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 copies of the nucleic acid sequence encoding the xylulose kinase. Optionally, the microorganism

comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 copies of the nucleic acid sequence encoding the xylose transporter.

In the provided microorganisms, the nucleic acids, e.g., xylose isomerase, xylulose
5 kinase or xylose transporter can be operably linked to a promoter and/or terminator.
Optionally, the exogenous nucleic acid sequence encoding the xylose isomerase is operably
linked to a promoter. Optionally, the nucleic acid sequence encoding the xylulose kinase
and/or the nucleic acid sequence encoding the xylose transporter are also operably linked to a
promoter. Optionally, the promoter is a tubulin promoter. Optionally, the promoter is at least
10 80% identical to SEQ ID NO:25 or SEQ ID NO:26. Optionally, the exogenous nucleic acid
sequence encoding the xylose isomerase comprises a terminator. Optionally, the nucleic acid
sequence encoding the xylulose kinase comprises a terminator. Optionally, the nucleic acid
sequence encoding the xylose transporter comprises a terminator. Optionally, the terminator
is a tubulin terminator. Optionally, the terminator is at least 80% identical to SEQ ID NO:27,
15 SEQ ID NO:28, or SEQ ID NO:30.

The provided microorganisms can include a selectable marker to confirm
transformation of genes of interest. Thus, the microorganism can further include a selectable
marker. Optionally, the selectable marker is an antibiotic resistance gene. Optionally, the
antibiotic is zeocin, hygromycin B, kanamycin or neomycin. Optionally, the microorganism
20 is either a *Thraustochytrium* or a *Schizochytrium* microorganism. Optionally, the
microorganism is ONC-T18.

The provided microorganisms have distinguishing features over wild type
microorganisms. For example, the recombinant microorganisms can have increased xylose
transport activity as compared to a non-recombinant control (or wild type) microorganism,
25 increased xylose isomerase activity as compared to a non-recombinant control (or wild type)
microorganism, increased xylulose kinase activity as compared to a non-recombinant control
(or wild type) microorganism, or any combination of these activities. Optionally, the
recombinant microorganism grows with xylose as the sole carbon source.

Also provided are methods of making the recombinant microorganisms. Thus,
30 provided is a method of making a recombinant xylose-metabolizing microorganism including
providing one or more nucleic acid constructs comprising a nucleic acid sequence encoding a
xylose isomerase, a nucleic acid sequence encoding a xylulose kinase and a nucleic acid
sequence encoding a xylose transporter; transforming the microorganism with the one or
more nucleic acid constructs; and isolating microorganisms comprising at least two copies of

the nucleic acid sequences encoding the xylose isomerase. Optionally, the methods further include isolating microorganisms comprising at least two copies of the nucleic acid sequence encoding the xylulose kinase. Optionally, the method includes isolating microorganisms comprising at least one copy of the xylose transporter. Optionally, the one or more nucleic acid constructs further comprise a selectable marker.

In the provided methods, the nucleic acid sequences encoding the xylose isomerase, xylulose kinase and xylose transporter can be located on the same or different constructs. Optionally, the method includes providing a first nucleic acid construct comprising a nucleic acid sequence encoding a xylose isomerase, a second nucleic acid construct comprising a nucleic acid sequence encoding a xylulose kinase and a third nucleic acid construct comprising a nucleic acid sequence encoding a xylose transporter. Optionally, the first, second and third nucleic acid constructs comprise the same selectable marker. Optionally, the first nucleic acid construct comprises a promoter, a selectable marker, a nucleic acid sequence encoding a 2A peptide, the nucleic acid sequence encoding the xylose isomerase, and a terminator. Optionally, the second nucleic acid construct comprises a promoter, selectable marker, a nucleic acid sequence encoding a 2A peptide, the nucleic acid sequence encoding the xylulose kinase, and a terminator. Optionally, the third nucleic acid construct comprises a promoter, the nucleic acid sequence encoding the xylose transporter, a nucleic acid sequence encoding a 2A peptide, a selectable marker, and a terminator. As noted above, selectable markers include, but are not limited to, antibiotic resistance genes. Optionally, the antibiotic is zeocin, hygromycin B, kanamycin or neomycin. Promoters used for the constructs include, but are not limited to, a tubulin promoter. Optionally, the promoter is at least 80% identical to SEQ ID NO:25 or SEQ ID NO:26. Terminators used for the constructs include, but are not limited to, a tubulin terminator. Optionally, the terminator is at least 80% identical to SEQ ID NO:27, SEQ ID NO:28, or SEQ ID NO:30.

In the provided methods, the isolated recombinant microorganisms can include one or more copies of the xylose isomerase, xylulose kinase and xylose transporter. Optionally, the isolated recombinant microorganism comprise at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 copies of the nucleic acid sequence encoding xylose isomerase. Optionally, the xylose isomerase is an endogenous xylose isomerase or a heterologous xylose isomerase. Optionally, the nucleic acid sequence encoding the xylose isomerase is at least 90% identical to SEQ ID NO:2. Optionally, the isolated recombinant microorganism comprises at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,

30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 copies of the nucleic acid sequence encoding the xylulose kinase. Optionally, the xylulose kinase is a heterologous xylulose kinase.

Optionally, the nucleic acid sequence encoding the xylulose kinase is at least 90% identical to SEQ ID NO:5. Optionally, the isolated recombinant microorganism comprises at least 2, 3,
5 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30,
31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 copies of the nucleic acid sequence encoding the xylose transporter. Optionally, the xylose transporter is a heterologous xylose transporter. Optionally, the xylose transporter is GXS1 from *Candida intermedia*. Optionally, the nucleic acid sequence encoding the xylose transporter is at least 90% identical to SEQ ID NO:23.

10 Optionally, the microorganism is either a *Thraustochytrium* or a *Schizochytrium* microorganism. Optionally, the microorganism is ONC-T18.

As noted above, the isolated recombinant microorganisms can have increased xylose transport activity as compared to a control non-recombinant microorganism, increased xylose isomerase activity as compared to a control non-recombinant microorganism, increased
15 xylulose kinase activity as compared to a control non-recombinant microorganism, or a combination thereof. Optionally, the isolated recombinant microorganism grows with xylose as the sole carbon source.

As described herein, a control or standard control refers to a sample, measurement, or value that serves as a reference, usually a known reference, for comparison to a test
20 sample, measurement, or value. For example, a test microorganism, e.g., a microorganism transformed with nucleic acid sequences encoding genes for metabolizing xylose can be compared to a known normal (wild-type) microorganism (e.g., a standard control microorganism). A standard control can also represent an average measurement or value gathered from a population of microorganisms (e.g., standard control microorganisms) that
25 do not grow or grow poorly on xylose as the sole carbon source or that do not have or have minimal levels of xylose isomerase activity, xylulose kinase activity and/or xylose transport activity. One of skill will recognize that standard controls can be designed for assessment of any number of parameters (e.g., RNA levels, polypeptide levels, specific cell types, and the like).

30 Provided herein are also methods of producing oil using the recombinant microorganisms. The method includes providing the recombinant microorganism, wherein the microorganism grows on xylose as the sole carbon source, and culturing the microorganism in a culture medium under suitable conditions to produce the oil. Optionally, the oil comprises triglycerides. Optionally, the oil comprises alpha linolenic acid,

arachidonic acid, docosahexanoic acid, docosapentaenoic acid, eicosapentaenoic acid, gamma-linolenic acid, linoleic acid, linolenic acid, or a combination thereof. Optionally, the method further includes isolating the oil.

5 The provided methods include or can be used in conjunction with additional steps for culturing microorganisms according to methods known in the art. For example, a Thraustochytrid, e.g., a *Thraustochytrium* sp., can be cultivated according to methods described in U.S. Patent Publications 2009/0117194 or 2012/0244584, which are herein incorporated by reference in their entireties for each step of the methods or composition used therein.

10 Microorganisms are grown in a growth medium (also known as culture medium). Any of a variety of medium can be suitable for use in culturing the microorganisms described herein. Optionally, the medium supplies various nutritional components, including a carbon source and a nitrogen source, for the microorganism. Medium for Thraustochytrid culture can include any of a variety of carbon sources. Examples of carbon sources include fatty
15 acids, lipids, glycerols, triglycerols, carbohydrates, polyols, amino sugars, and any kind of biomass or waste stream. Fatty acids include, for example, oleic acid. Carbohydrates include, but are not limited to, glucose, cellulose, hemicellulose, fructose, dextrose, xylose, lactulose, galactose, maltotriose, maltose, lactose, glycogen, gelatin, starch (corn or wheat), acetate, m-inositol (e.g., derived from corn steep liquor), galacturonic acid (e.g., derived from
20 pectin), L-fucose (e.g., derived from galactose), gentiobiose, glucosamine, alpha-D-glucose-1-phosphate (e.g., derived from glucose), cellobiose, dextrin, alpha-cyclodextrin (e.g., derived from starch), and sucrose (e.g., from molasses). Polyols include, but are not limited to, maltitol, erythritol, and adonitol. Amino sugars include, but are not limited to, N-acetyl-D-galactosamine, N-acetyl-D-glucosamine, and N-acetyl-beta-D-mannosamine.

25 Optionally, the microorganisms provided herein are cultivated under conditions that increase biomass and/or production of a compound of interest (e.g., oil or total fatty acid (TFA) content). Thraustochytrids, for example, are typically cultured in saline medium. Optionally, Thraustochytrids can be cultured in medium having a salt concentration from about 0.5 g/L to about 50.0 g/L. Optionally, Thraustochytrids are cultured in medium having
30 a salt concentration from about 0.5 g/L to about 35 g/L (e.g., from about 18 g/L to about 35 g/L). Optionally, the Thraustochytrids described herein can be grown in low salt conditions. For example, the Thraustochytrids can be cultured in a medium having a salt concentration from about 0.5 g/L to about 20 g/L (e.g., from about 0.5 g/L to about 15 g/L). The culture

medium optionally includes NaCl. Optionally, the medium includes natural or artificial sea salt and/or artificial seawater.

The culture medium can include non-chloride-containing sodium salts as a source of sodium. Examples of non-chloride sodium salts suitable for use in accordance with the present methods include, but are not limited to, soda ash (a mixture of sodium carbonate and sodium oxide), sodium carbonate, sodium bicarbonate, sodium sulfate, and mixtures thereof. See, e.g., U.S. Pat. Nos. 5,340,742 and 6,607,900, the entire contents of each of which are incorporated by reference herein. A significant portion of the total sodium, for example, can be supplied by non-chloride salts such that less than about 100%, 75%, 50%, or 25% of the total sodium in culture medium is supplied by sodium chloride.

Medium for Thraustochytrids culture can include any of a variety of nitrogen sources. Exemplary nitrogen sources include ammonium solutions (e.g., NH_4 in H_2O), ammonium or amine salts (e.g., $(\text{NH}_4)_2\text{SO}_4$, $(\text{NH}_4)_3\text{PO}_4$, NH_4NO_3 , $\text{NH}_4\text{OOCH}_2\text{CH}_3$ (NH_4Ac)), peptone, tryptone, yeast extract, malt extract, fish meal, sodium glutamate, soy extract, casamino acids and distiller grains. Concentrations of nitrogen sources in suitable medium typically range between and including about 1 g/L and about 25 g/L.

The medium optionally includes a phosphate, such as potassium phosphate or sodium-phosphate. Inorganic salts and trace nutrients in medium can include ammonium sulfate, sodium bicarbonate, sodium orthovanadate, potassium chromate, sodium molybdate, selenous acid, nickel sulfate, copper sulfate, zinc sulfate, cobalt chloride, iron chloride, manganese chloride calcium chloride, and EDTA. Vitamins such as pyridoxine hydrochloride, thiamine hydrochloride, calcium pantothenate, p-aminobenzoic acid, riboflavin, nicotinic acid, biotin, folic acid and vitamin B12 can be included.

The pH of the medium can be adjusted to between and including 3.0 and 10.0 using acid or base, where appropriate, and/or using the nitrogen source. Optionally, the medium can be sterilized.

Generally a medium used for culture of a microorganism is a liquid medium. However, the medium used for culture of a microorganism can be a solid medium. In addition to carbon and nitrogen sources as discussed herein, a solid medium can contain one or more components (e.g., agar or agarose) that provide structural support and/or allow the medium to be in solid form.

Optionally, the resulting biomass is pasteurized to inactivate undesirable substances present in the biomass. For example, the biomass can be pasteurized to inactivate compound degrading substances. The biomass can be present in the fermentation medium or isolated

from the fermentation medium for the pasteurization step. The pasteurization step can be performed by heating the biomass and/or fermentation medium to an elevated temperature. For example, the biomass and/or fermentation medium can be heated to a temperature from about 50°C to about 95°C (e.g., from about 55°C to about 90°C or from about 65°C to about 80°C). Optionally, the biomass and/or fermentation medium can be heated from about 30 minutes to about 120 minutes (e.g., from about 45 minutes to about 90 minutes, or from about 55 minutes to about 75 minutes). The pasteurization can be performed using a suitable heating means, such as, for example, by direct steam injection.

Optionally, no pasteurization step is performed. Stated differently, the method taught herein optionally lacks a pasteurization step.

Optionally, the biomass can be harvested according to a variety of methods, including those currently known to one skilled in the art. For example, the biomass can be collected from the fermentation medium using, for example, centrifugation (e.g., with a solid-ejecting centrifuge) or filtration (e.g., cross-flow filtration). Optionally, the harvesting step includes use of a precipitation agent for the accelerated collection of cellular biomass (e.g., sodium phosphate or calcium chloride).

Optionally, the biomass is washed with water. Optionally, the biomass can be concentrated up to about 20% solids. For example, the biomass can be concentrated to about 5% to about 20% solids, from about 7.5% to about 15% solids, or from about solids to about 20% solids, or any percentage within the recited ranges. Optionally, the biomass can be concentrated to about 20% solids or less, about 19% solids or less, about 18% solids or less, about 17% solids or less, about 16% solids or less, about 15% solids or less, about 14% solids or less, about 13% solids or less, about 12% solids or less, about 11% solids or less, about 10% solids or less, about 9% solids or less, about 8% solids or less, about 7% solids or less, about 6% solids or less, about 5% solids or less, about 4% solids or less, about 3% solids or less, about 2% solids or less, or about 1% solids or less.

The provided methods, optionally, include isolating the polyunsaturated fatty acids from the biomass or microorganisms. Isolation of the polyunsaturated fatty acids can be performed using one or more of a variety of methods, including those currently known to one of skill in the art. For example, methods of isolating polyunsaturated fatty acids are described in U.S. Patent No. 8,163,515, which is incorporated by reference herein in its entirety. Optionally, the medium is not sterilized prior to isolation of the polyunsaturated fatty acids. Optionally, sterilization comprises an increase in temperature. Optionally, the polyunsaturated fatty acids produced by the microorganisms and isolated from the provided

methods are medium chain fatty acids. Optionally, the one or more polyunsaturated fatty acids are selected from the group consisting of alpha linolenic acid, arachidonic acid, docosahexaenoic acid, docosapentaenoic acid, eicosapentaenoic acid, gamma-linolenic acid, linoleic acid, linolenic acid, and combinations thereof.

5 Oil including polyunsaturated fatty acids (PUFAs) and other lipids produced according to the method described herein can be utilized in any of a variety of applications exploiting their biological, nutritional, or chemical properties. Thus, the provided methods optionally include isolating oil from the harvested portion of the threshold volume. Optionally, the oil is used to produce fuel, e.g., biofuel. Optionally, the oil can be used in
10 pharmaceuticals, food supplements, animal feed additives, cosmetics, and the like. Lipids produced according to the methods described herein can also be used as intermediates in the production of other compounds.

By way of example, the oil produced by the microorganisms cultured using the provided methods can comprise fatty acids. Optionally, the fatty acids are selected from the
15 group consisting of alpha linolenic acid, arachidonic acid, docosahexaenoic acid, docosapentaenoic acid, eicosapentaenoic acid, gamma-linolenic acid, linoleic acid, linolenic acid, and combinations thereof. Optionally, the oil comprises triglycerides. Optionally, the oil comprises fatty acids selected from the group consisting of palmitic acid (C16:0), myristic acid (C14:0), palmitoleic acid (C16:1(n-7)), cis-vaccenic acid (C18:1(n-7)),
20 docosapentaenoic acid (C22:5(n-6)), docosahexaenoic acid (C22:6(n-3)), and combinations thereof.

Optionally, the lipids produced according to the methods described herein can be incorporated into a final product (e.g., a food or feed supplement, an infant formula, a pharmaceutical, a fuel, etc.). Suitable food or feed supplements into which the lipids can be
25 incorporated include beverages such as milk, water, sports drinks, energy drinks, teas, and juices; confections such as candies, jellies, and biscuits; fat-containing foods and beverages such as dairy products; processed food products such as soft rice (or porridge); infant formulae; breakfast cereals; or the like. Optionally, one or more produced lipids can be incorporated into a dietary supplement, such as, for example, a vitamin or multivitamin.
30 Optionally, a lipid produced according to the method described herein can be included in a dietary supplement and optionally can be directly incorporated into a component of food or feed (e.g., a food supplement).

Examples of feedstuffs into which lipids produced by the methods described herein can be incorporated include pet foods such as cat foods; dog foods; feeds for aquarium fish,

cultured fish or crustaceans, etc.; feed for farm-raised animals (including livestock and fish or crustaceans raised in aquaculture). Food or feed material into which the lipids produced according to the methods described herein can be incorporated is preferably palatable to the organism which is the intended recipient. This food or feed material can have any physical
5 properties currently known for a food material (e.g., solid, liquid, soft).

Optionally, one or more of the produced compounds (e.g., PUFAs) can be incorporated into a nutraceutical or pharmaceutical product. Examples of such a nutraceuticals or pharmaceuticals include various types of tablets, capsules, drinkable agents, etc. Optionally, the nutraceutical or pharmaceutical is suitable for topical application.
10 Dosage forms can include, for example, capsules, oils, granula, granula subtilae, pulveres, tabellae, pilulae, trochisci, or the like.

The oil or lipids produced according to the methods described herein can be incorporated into products as described herein in combination with any of a variety of other agents. For instance, such compounds can be combined with one or more binders or fillers,
15 chelating agents, pigments, salts, surfactants, moisturizers, viscosity modifiers, thickeners, emollients, fragrances, preservatives, etc., or any combination thereof.

Disclosed are materials, compositions, and components that can be used for, can be used in conjunction with, can be used in preparation for, or are products of the disclosed methods and compositions. These and other materials are disclosed herein, and it is
20 understood that when combinations, subsets, interactions, groups, etc. of these materials are disclosed that while specific reference of each various individual and collective combinations and permutations of these compounds may not be explicitly disclosed, each is specifically contemplated and described herein. For example, if a method is disclosed and discussed and a number of modifications that can be made to a number of molecules including the method
25 are discussed, each and every combination and permutation of the method, and the modifications that are possible are specifically contemplated unless specifically indicated to the contrary. Likewise, any subset or combination of these is also specifically contemplated and disclosed. This concept applies to all aspects of this disclosure including, but not limited to, steps in methods using the disclosed compositions. Thus, if there are a variety of
30 additional steps that can be performed, it is understood that each of these additional steps can be performed with any specific method steps or combination of method steps of the disclosed methods, and that each such combination or subset of combinations is specifically contemplated and should be considered disclosed.

Publications cited herein and the material for which they are cited are hereby specifically incorporated by reference in their entireties.

The examples below are intended to further illustrate certain aspects of the methods and compositions described herein, and are not intended to limit the scope of the claims.

5

Examples

Example 1. C5 Carbon Metabolism by Recombinant *Thraustochytrids*

In nature, two xylose metabolism pathways exist, the xylose reductase/xylitol dehydrogenase pathway and the xylose isomerase/xylulose kinase pathway (Figure 1). ONC-T18 encodes genes from both pathways, and, as described above, the xylose reductase/xylitol dehydrogenase pathway is dominant, as evidenced by a build-up of xylitol when grown in a xylose medium. Since the isomerase/kinase pathway does not depend on redox co-factors, over-expression of ONC-T18's isomerase gene removes co-factor dependence in the conversion of xylose to xylulose. As shown herein in Figures 2 and 3, transcriptomic studies with ONC-T18 showed that its xylose isomerase and putative xylulose kinase genes were mostly expressed during glucose starvation; whereas, the putatively identified genes encoding for the xylose reductase and xylitol dehydrogenase were constitutively expressed.

T18 isomerase was purified by metal-affinity chromatography following his-tagging and over-expression in yeast INVSc1. As a positive control, his-tagged XylA from *E. coli* strain W3110 was over-expressed and purified from *E. coli* strain BL21(DE3)plysS. The protein concentration of purified proteins was determined by a standard Bradford assay. The impact of temperature on the activity of T18 isomerase and *E. coli* isomerase was determined using 5 µg of protein and 0.75 g/L of either xylose or xylulose in 5mM MgATP, 50mM Hepes (pH 7.4), 10 mM MgCl₂. Reactions were incubated overnight at 10 °C, 25 °C, 30 °C, 37 °C, 50 °C, 60 °C, and 80 °C. Reactions were stopped by heat inactivation at 95 °C for 5 mins. Reactions were analyzed by HPLC and the concentration of the sugars present was determined from the area under the peak relative to a standard curve. T18 isomerase had higher activity on both xylose and xylulose at temperatures at and above 37 °C (Figure 20A). This is in contrast to *E. coli* isomerase, which had higher activity at temperatures between 25 °C and 30°C (Figure 20B).

Dose-dependency was determined by incubating increasing protein concentrations of the isomerase with 0.75 g/L xylose or xylulose in 5mM MgATP, 50mM Hepes (pH 7.4), 10 mM MgCl₂. Reactions were incubated overnight at 30°C (*E. coli*) or 50°C (T18) then stopped by heat inactivation at 95 °C for 5 mins. Reactions were analyzed by HPLC and the concentration of the sugars present was determined from the area under the peak relative to a

standard curve. Observed was a dose dependency of T18 isomerase on both xylose and xylulose (Figures 21A and 21B).

This example describes the use of a *Thraustochytrium* *ONC-T18*-derived (ONC-T18) alpha-tubulin promoter to express endogenous and/or heterologous xylose metabolism transgenes in Thraustochytrid species, including ONC-T18. However, as discussed throughout, other regulatory elements can be used. Figures 4 and 5 show constructs of the plasmids containing the xylose isomerase and xylulose kinase genes, respectively. As described herein, the xylose metabolism transgenes were present in multiple (≥ 8) copies within the genome of the host. In the case of ONC-T18, the modified organisms demonstrated an increased metabolism of xylose compared to wild-type (WT) cells. For example, a strain modified to express an endogenous xylose isomerase gene (SEQ ID NO:2) (strain Iso-His #16) and a strain modified to express an endogenous xylose isomerase gene (SEQ ID NO:2) and a xylulose kinase gene (SEQ ID NO:5) (Iso-His+xylB, strain 7-7) both used 40% more xylose than the WT strain. Both Iso-His #16 and 7-7 converted less xylose to xylitol than the WT strain, 40% less and 420% less, respectively. The constructs used for transformation of ONC-T18 are shown in Figures 4 and 5. ONC-T18 transformants were created using standard biolistics protocols as described by BioRad's Biolistic PDS-1000/He Particle Delivery System (Hercules, CA). Briefly, 0.6 μm gold particles were coated with 2.5 μg of linearized plasmid DNA (EcoRI, 37°C, overnight). The coated gold particles were used to bombard plates previously spread with 1 ml of ONC-T18 cells at an OD600 of 1.0. The bombardment parameters included using a helium pressure of 1350 or 1100 psi with a target distance of 3 or 6 cm. After an overnight recovery, the cells were washed off the plate and plated on media containing selection antibiotics (Zeo 250 $\mu\text{g}/\text{mL}$ and hygro 400 $\mu\text{g}/\text{mL}$). Plates were incubated for 1 week at 25°C to identify resistant colonies. The resulting transformants were screened by PCR and Southern blot.

Southern blots were performed using standard protocols. Briefly, approximately 20 μg of genomic DNA were digested with 40 units of BamHI restriction enzyme in a total volume of 50 μL overnight at 37°C. 7.2 μg of each digested sample was run on a 1.0% agarose gel at 50V for approximately 1.5. hours, with a digoxigenin (DIG) DNA molecular-weight marker II (Roche, Basel, Switzerland). DNA was depurinated in the gel by submerging the gel in 250 mM HCl for 15 minutes. The gel was further denatured by incubation in a solution containing 0.5 M NaOH and 1.5 M NaCl (pH 7.5) for two 15 minute washes. The reaction was then neutralized by incubation in 0.5 M Tris-HCl (pH 7.5) for two 15 minute washes. Finally, the gel was equilibrated in 20X saline-sodium citrate (SSC) buffer

for 15 minutes. DNA was transferred to a positively charged nylon membrane using a standard transfer apparatus. DNA was fixed to the membrane using a UV Stratalinker at an exposure of 120,000 μ J. Southern blot probe was generated using a PCR DIG Probe Synthesis Kit (Roche, Basel, Switzerland) to generate a DIG-labelled probe according to the manufacturer's instructions. The DNA affixed to the nylon membrane was prehybridized with 20 mL of DIG EasyHyb solution (DIG EasyHyb Granules, Roche, Basel, Switzerland). The DIG-labelled probe was denatured by adding 40 μ L of the ble-probe reaction mixture to 300 μ L of ddH₂O and incubated at 99°C for 5 minutes. This solution was then added to 20 mL of DIG hybridization solution to create the probe solution. The probe solution was then added to the DNA-affixed nylon membrane and incubated at 53 °C overnight. The following day, the membrane was washed twice in 2X SSC, 0.1% SDS at room temperature. The membrane was further washed twice in 0.1X SSC, 0.1% SDS at 68°C for 15 minutes. For detection, the membrane was washed and blocked using DIG Wash and Block Buffer set (Roche, Basel, Switzerland) according to the manufacturer's instructions. An anti-DIG-AP conjugated antibody from a DIG Nucleic Acid Detection Kit (Roche, Basel, Switzerland) was used for detection. 2 μ L of the antibody solution was added to 20mL detection solution and incubated with the membrane at room temperature for 30 minutes. The blot was then immersed in a washing buffer provided with the kit. CDP-Star (Roche, Basel, Switzerland) was used for visualization. 10 μ L of the CDP-star solution was incubated on the membrane in 1 mL of detection solution, which was covered in a layer of 'sheet-protector' plastic to hold the solution to the membrane. Signal was immediately detected using a ChemiDoc imaging system (BioRad Laboratories, Hercules, CA).

The codon optimized ble gene was cloned under the control of T18B α -tubulin promoter and terminator elements (Figure 6). The isomerase gene was cloned from T18B in such a way as to add a six-histidine tag on the N-terminus of the expressed protein (Iso-His). Xylose isomerase enzymatic activity was confirmed by over-expression and purification of the histidine-tagged protein in yeast. The isomerase gene (along with the introduced six-histidine tag) was cloned under the control of the α -tubulin promoter and terminators by cloning the gene downstream of the ble gene and a 2A sequence (Figure 4 and Figure 6). Biolistic transformation of T18B with this plasmid (pALPHTB-B2G-hisIso) resulted in Zeocin (zeo) resistant transformants. Many transformant strains were obtained from this procedure. Two of these strains are shown as example #6 containing one copy of the transgene and example #16 containing eight copies of the transgene (Figure 7).

The insertion of the Iso-His transgene within the T18B genome was confirmed by PCR and Southern blot analysis (Figure 7). Qualitatively, these data showed the presence of a single copy of the transgene in strain #6 and multiple, concatameric, transgene copies, at a single site, in strain #16. The precise number of Iso-His transgene insertions was determined by qPCR on genomic DNA (Figure 8). These data showed the presence of one copy of the transgene in strain #6 and eight copies of the transgene in strain #16 (Figure 8). To test whether an increase in copy number correlated with an increase in expression level, mRNA was isolated from WT, Iso-His #6 and Iso-His #16 T18B cells and qRT-PCR was performed. Figure 11 shows significantly increased expression of the Iso-His transcript in strain #16 cells, containing eight copies of the transgene, compared to strain #6, containing a single copy of the transgene. No Iso-His transcript is detectable in WT cells (Figure 11). To assess whether increased mRNA expression correlated with increased isomerase enzymatic activity, cell extracts were harvested from WT, Iso-His #6 and Iso-His #16 cells. Enhanced isomerase enzyme activity is observed in strain #16 cells compared with strain #6 and WT cells (Figure 12). Finally, the ability of strain #16 to metabolize xylose was examined in xylose depletion assays (Figure 14) and compared with WT cells. These flask fermentations demonstrated the ability to metabolise xylose and quantify the amount of xylose converted to xylitol. Thus, Figure 14 shows an increase in xylose metabolism in Iso-His strain #16 compared with WT cells and significantly less production of xylitol.

For flask assays, cells were grown in media for 2 to 3 days. Pellets were washed twice in Media 2 (9 g/L NaCl, 4 g/L MgSO₄, 100 mg/L CaCl₂, 5 mg/L FeCl₃, 20 g/L (NH₄)₂SO₄, 0.86 g/L KH₂PO₄, 150 µg/L vitamin B12, 30 µg/L biotin, 6 mg/L thiamine hydrochloride, 1.5 mg/L cobalt (II) chloride, 3 mg/L manganese chloride) containing no sugar. Then, minimal media containing 20 g/L glucose & 50 g/L xylose was inoculated to an OD₆₀₀ of 0.05 with the washed cells. Samples were taken at various time points and the amount of sugar remaining in the supernatant was analyzed by HPLC. As shown in Figures 22A, 22B, 22C and 22D, with increased xylose isomerase gene copy number, up to 40% more xylose usage and 20% decrease in xylitol production when compared to WT.

Iso-His strain #16 was then used as the parent strain for a second round of transformation to introduce the *E. coli xylB* gene. This gene was introduced under hygromycin (hygro) selection. The hygro gene from pChlamy_3, the 2A sequence, and the T18B codon optimized W3110 *E. coli xylB* gene were cloned under the control of the T18B α -tubulin promoter and terminator elements for expression in T18B iso-his #16 (Figure 5). The *in vitro* ability of the *E. coli* xylulose kinase to work in concert with the T18B isomerase

was confirmed by over-expression and purification of the histidine-tagged proteins in yeast followed by enzymatic reactions with xylose and xylulose. Biolistic transformation of T18B iso-his strain #16 with the xylB plasmid (pJB47) resulted in hygro and zeo resistant transformants. The insertion of the hygro-2A-xylB genes within the T18B genome was confirmed by PCR and Southern blot analysis (Figure 9). Qualitatively, these data show the presence of a single copy of the transgene in strain #7-3 and multiple, concatameric, transgene copies, at a single site, in strain #7-7. The number of xylB gene insertions was determined by qPCR on genomic DNA isolations (Figure 10). Figure 10 shows sixteen insertions of the transgene in strain 7-7 and one copy in strain 7-3. To determine whether multiple copies of the transgene confer enhanced xylose metabolism *in vitro*, cell extract assays were performed and the ability of the cells extracts to metabolise xylose was analysed (Figure 13). The ability of the transformant cells to metabolize xylose was examined through flask-based xylose depletion assays (Figure 15). In this experiment, WT cells consumed the least amount of xylose and made the most xylitol. Strain Iso-His #16, 7-3 and 7-7 all consumed similar amounts of xylose; however, only 7-7, containing multiple copies of the xylB transgene, did not make significant amounts of xylitol. Finally, strains Iso-His #16 and 7-7 were tested at in 5L fermentation vessels in media containing glucose and xylose. During a seventy-seven (77) hour fermentation, strain Iso-His #16 converted approximately 8% of xylose to xylitol, whereas strain 7-7 converted approximately 2% of xylose to xylitol. Xylitol accumulation in this fermentation is shown in Figure 16.

For flasks assays, cells were grown in media for 2 to 3 days. Pellets were washed twice in media containing no sugar. Media containing 20 g/L: 50 g/L glucose: xylose was inoculated to an OD600 of 0.05 with the washed cells. Samples were taken at various time points and the amount of sugar remaining in the supernatant was analyzed by HPLC. As shown in Figures 23A, 23B, 23C and 23D, up to 50% more xylose was used and an 80% reduction in xylitol was observed in strains over-expressing both a xylose isomerase and a xylulose kinase when compared to WT.

To further analyze these strains, the strains were grown in parallel 5L Sartorius fermenters. Initial media contained 20g/L Glucose and 50g/L xylose along with other basal media components. Both cultures were maintained at 28°C and 5.5 pH, with constant mixing at 720 RPM and constant aeration at 1 Lpm of environmental air. The cultures were fed glucose for 16 hrs followed by 8hr starvation period. This cycle was completed 3 times. During starvation periods, 10mL samples were taken every 0.5 hr. Glucose, xylose and xylitol concentrations were quantified in these samples by HPLC. Larger 50mL samples were

taken periodically for further biomass and oil content quantification. Glucose feed rates matched glucose consumption rates, which was quantified by CO₂ detected in the culture exhaust gas. As shown in Figure 24, the 7-7 strain used up to 52% more xylose than WT under these conditions.

5 By Southern blot analysis, it was observed that strain Iso-His #16 contains eight (8) insertions of the isomerase transgene (Figure 8). This unexpected multiple insertion resulted in an increase in isomerase gene expression relative to strains harbouring a single copy (Figure 11) as well as increased isomerase *in vitro* activity (Figure 12). Strain Iso-His #16 demonstrated increased xylose productivity than strains harbouring a single copy of the
10 isomerase transgene (Figure 14).

Similarly, within the *Iso-His + xylB* transformants, one of the clones (*Iso-His + xylB* 7-7) also had multiple insertions of the *xylB* gene (Figure 10), which resulted in increased *in vivo* activity of both the xylose isomerase and xylulose kinase within the cell (Figure 13). This clone was capable of using either as much or more xylose than the parental strain, Iso-
15 His #16, while producing significantly less xylitol (Figure 15). Furthermore the *Iso-His + xylB* 7-7 produced more biomass than WT in the presence of xylose. These two strains showed that, not only is the presence of both the isomerase and the kinase genes important, but the number of insertions is as well.

To further optimize the iso-his & xylB containing “7-7” strain, this strain was
20 transformed with a xylose transporter. Figure 17 shows exemplary constructs for transformation. Examples of xylose transporters to be used include, but are not limited to, At5g17010 and At5g59250 (*Arabidopsis thaliana*), Gfx1 and GXS1 (*Candida*), AspTx (*Aspergillus*), and Sut1 (*Pichia*). Gxs1 (SEQ ID NO:23) was selected for transformation. The results are shown in Figures 19A, 19B, and 19C. The transformants 36-2, 36-9, and 36-
25 16, containing GXS1 use more xylose than 7-7 and WT strains. They also use glucose slower than WT and 7-7 strains. The data demonstrate both xylose and glucose being used in the earlier stages by the GXS1 containing strains. Further, the percent of xylitol made by the GXS1 containing strains is lower than both WT and 7-7 strains.

To further analyze the effect of sugar transporters on the metabolism of xylose,
30 codon optimized xylose transporters AspTX from *Aspergillus* (An11g01100) and Gxs1 from *Candida* were introduced in the 7-7 strain (*isohis + xylB*). Figure 25 shows the alpha-tubulin aspTx-neo and alpha-tubulin gxs1-neo constructs. T18 transformants were created using standard biolistics protocols as described by BioRad’s Biolistic PDS-1000/He Particle Delivery System. Briefly, 0.6 µm gold particles were coated with 2.5 µg of linearized

plasmid DNA (EcoRI, 37°C, o/n), The coated gold particles were used to bombard WD plates previously spread with 1 ml of T18 cells at an OD600 of 1.0. The bombardment parameters included using a Helium pressure of 1350 or 1100 psi with a target distance of 3 or 6 cm.

After an overnight recovery, the cells were washed off the plate and plated on media

- 5 containing selection antibiotics (G418 at 2 mg/mL). Plates were incubated for 1 week at 25°C to identify resistant colonies. The resulting transformants were screened by PCR and Southern blot (Figure 26).

- Southern blots were performed using standard protocols. Briefly, approximately 20 µg of genomic DNA were digested with 40 units of BamHI restriction enzyme in a total
10 volume of 50 µL o/n/ at 37°C. 7.2 µg of each digested sample was run on a 1.0% agarose gel at 50V for approximately 1.5. hours, with a digoxigenin (DIG) DNA molecular-weight marker II (Roche). DNA was depurinated in the gel by submerging the gel in 250 mM HCl for 15 minutes. The gel was further denatured by incubation in a solution containing 0.5 M NaOH and 1.5 M NaCL (pH 7.5) for two 15 minute washes. The reaction was then
15 neutralized by incubation in 0.5 M Tris-HCl (pH 7.5) for two 15 minute washes. Finally, the gel was equilibrated in 20X saline-sodium citrate (SSC) buffer for 15 minutes. DNA was transferred to a positively charged nylon membrane (Roche) using a standard transfer apparatus. DNA was fixed to the membrane using a UV Stratalinker at an exposure of 120,000 µJ. Southern blot probe was generated using a PCR DIG Probe Synthesis Kit
20 (Roche) to generate a DIG-labelled probe according to the manufacturer's instructions. The DNA affixed to the nylon membrane was prehybridised with 20 mL of DIG EasyHyb solution (DIG EasyHyb Granules, Roche). The DIG-labelled probe was denatured by adding 40 µL of the ble-probe reaction mixture to 300 µL of ddH₂O and incubated at 99°C for 5 minutes. This solution was then added to 20 mL of DIG hybridization solution to create the
25 probe solution. The probe solution was then added to the DNA-affixed nylon membrane and incubated at 53 °C overnight. The following day, the membrane was washed, twice, in 2X SSC, 0.1% SDS at RT. The membrane was further washed, twice, in 0.1X SSC, 0.1% SDS at 68°C for 15 minutes. For detection, the membrane was washed and blocked using DIG Wash and Block Buffer set (Roche) according to the manufacturer's instructions. An anti-DIG-AP
30 conjugated antibody from a DIG Nucleic Acid Detection Kit (Roche) was used for detection. 2 µL of the antibody solution was added to 20mL detection solution and incubated with the membrane at RT for 30 minutes. The blot was then immersed in a washing buffer provided with the kit. CDP-Star (Roche) was used for visualization. 10 µL of the CDP-star solution was incubated on the membrane in 1 mL of detection solution, which was covered in a layer

of 'sheet-protector' plastic to hold the solution to the membrane. Signal was immediately detected using a ChemiDoc imaging system.

For flasks assays, cells were grown in media for 2 to 3 days. Pellets were washed twice in media 2 (9 g/L NaCl, 4 g/L MgSO₄, 100 mg/L CaCl₂, 5 mg/L FeCl₃, 20 g/L (NH₄)₂SO₄, 0.86 g/L KH₂PO₄, 150 µg/L vitamin B12, 30 µg/L biotin, 6 mg/L thiamine hydrochloride, 1.5 mg/L cobalt (II) chloride, 3 mg/L manganese chloride) containing no sugar. Then, Media 2 containing 20 g/L Glucose and 20 g/L Xylose was inoculated to an OD₆₀₀ of 0.05 with the washed cells. As shown in Figures 27A, 27B, 27C and 27D, the expression of the xylose isomerase, xylulose kinase, and either xylose transporters resulted in up to 71% more xylose used and 40% less xylitol produced than the parental strain 7-7.

For flasks assays, cells were grown in media for 2 to 3 days. Pellets were washed twice in saline. Then, media containing 60 g/L xylose instead of glucose was inoculated to an OD₆₀₀ of 0.05 with the washed cells. Samples were taken at various time points and the amount of sugar remaining in the supernatant was analyzed by HPLC. Figure 28 shows T18 growth in media containing xylose as the main carbon source requires over-expression of both an isomerase and a kinase. The expression of the transporters in this background did not significantly increase xylose usage in this media.

Enhanced xylose usage by T18 7-7 and transporter strains was observed in media containing carbon from alternative feed stocks. For flasks assays, cells were grown in media for 2 to 3 days. Pellets were washed twice in 0.9% saline solution. Media 2 containing 20 g/L glucose : 50 g/L xylose as a combination of lab grade glucose and glucose and xylose from an alternative feedstock from forestry, was inoculated to an OD_{600nm} of 0.05 with the washed cells. Samples were taken at various time points and the amount of sugar remaining in the supernatant was analyzed by HPLC. As shown in Figures 29A and 29B, in media containing sugars from an alternative feedstock, the T18 7-7 strains encoding for transporters used more xylose than wild-type, or T18 7-7.