



Open Access Articles

Genetic resources for advanced biofuel production described with the Gene Ontology

The Faculty of Oregon State University has made this article openly available.
Please share how this access benefits you. Your story matters.

Citation	Torto-Alalibo, T., Purwantini, E., Lomax, J., Setubal, J. C., Mukhopadhyay, B., & Tyler, B. M. (2014). Genetic resources for advanced biofuel production described with the Gene Ontology. <i>Frontiers in Microbiology</i> , 5, 528. doi:10.3389/fmicb.2014.00528
DOI	10.3389/fmicb.2014.00528
Publisher	Frontiers Research Foundation
Version	Version of Record
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse



Genetic resources for advanced biofuel production described with the Gene Ontology

Trudy Torto-Alalibo^{1,2}, Endang Purwantini^{1,2}, Jane Lomax³, João C. Setubal^{2,4}, Biswarup Mukhopadhyay^{1,2,5} and Brett M. Tyler^{2,6} *

¹ Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

² Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

³ European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, UK

⁴ Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

⁵ Department of Biological Sciences, Oregon State University, Corvallis, OR, USA

⁶ Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA

Edited by:

Katherine M. Pappas, University of Athens, Greece

Reviewed by:

Kesen Ma, University of Waterloo, Canada

Christopher J. Brigham, University of Massachusetts at Dartmouth, USA

David E. Graham, Oak Ridge National Laboratory, USA

*Correspondence:

Brett M. Tyler, Center for Genome Research and Biocomputing, Oregon State University, 4750 Campus Way ALS3021, Corvallis, OR 97331, USA
e-mail: brett.tyler@oregonstate.edu

Dramatic increases in research in the area of microbial biofuel production coupled with high-throughput data generation on bioenergy-related microbes has led to a deluge of information in the scientific literature and in databases. Consolidating this information and making it easily accessible requires a unified vocabulary. The Gene Ontology (GO) fulfills that requirement, as it is a well-developed structured vocabulary that describes the activities and locations of gene products in a consistent manner across all kingdoms of life. The Microbial ENergy processes Gene Ontology (<http://www.mengo.biochem.vt.edu>) project is extending the GO to include new terms to describe microbial processes of interest to bioenergy production. Our effort has added over 600 bioenergy related terms to the Gene Ontology. These terms will aid in the comprehensive annotation of gene products from diverse energy-related microbial genomes. An area of microbial energy research that has received a lot of attention is microbial production of advanced biofuels. These include alcohols such as butanol, isopropanol, isobutanol, and fuels derived from fatty acids, isoprenoids, and polyhydroxyalkanoates. These fuels are superior to first generation biofuels (ethanol and biodiesel esterified from vegetable oil or animal fat), can be generated from non-food feedstock sources, can be used as supplements or substitutes for gasoline, diesel and jet fuels, and can be stored and distributed using existing infrastructure. Here we review the roles of genes associated with synthesis of advanced biofuels, and at the same time introduce the use of the GO to describe the functions of these genes in a standardized way.

Keywords: Gene Ontology, advanced biofuels, synthetic biology, cellulosome, advanced alcohols, fatty acid-derived fuel, isoprenoid-derived fuel

INTRODUCTION

Depletion of the world's fossil fuel resources and environmental concerns associated with the emission of greenhouse gases has fueled interest in renewable and environmentally friendly alternatives (Köne and Büke, 2010; Hansen et al., 2013; Suranovic, 2013). In this context, advanced biofuels have been of growing interest as these compounds can be generated from non-food cellulosic biomass, can be added directly to gasoline or diesel or sometimes used as stand-alone fuel, and can be stored and distributed using existing infrastructure (Dürre, 2007; Mehta et al., 2010; Weber et al., 2010). Advanced biofuels include alcohols such as butanol, isopropanol, and isobutanol, and fuels derived from fatty acids, isoprenoids, and polyhydroxyalkanoates (PHAs). Advanced biofuels from lignocellulose feedstock begins with biomass deconstruction. The cellulosic component of the biomass is degraded into pentoses and hexoses. The multi-enzyme complexes involved in the degradation of cellulosic biomass are discussed in this review. The native and engineered pathways leading to the production of advanced biofuels have been

studied extensively. As such the primary published literature is rich in information derived from research on these fuels but this information has yet to be aggregated in a manner that is easily accessible and amenable to computational analysis. The well-established Gene Ontology (GO; Ashburner et al., 2000) provides a basis for harnessing this information. The GO provides sets of standardized terms to describe molecular functions, biological processes and cellular components across all kingdoms of life (Ashburner et al., 2000; Harris et al., 2004; Torto-Alalibo et al., 2009; Dimmer et al., 2012). The Microbial ENergy processes Gene Ontology (MENGO) consortium¹, an associate of the GO, initiated an effort in 2011 to extend the GO by developing missing terms associated with bioenergy processes. This effort has generated over 600 bioenergy-related GO terms, most of which are described in this review and the MENGO website¹. These new terms, together with existing ones in the GO, were used to annotate gene products involved in bioenergy processes. Our

¹<http://www.mengo.biochem.vt.edu>

emphasis is on selected native pathways leading to the generation of advanced biofuels. The scope of the review is summarized in Figure 1.

THE GENE ONTOLOGY

The GO is a structured, species-neutral ontology that describes the attributes of gene products. The GO consists of three distinct ontologies: “GO:0003674 molecular function,” “GO:0008150 biological process” and “GO:0005575 cellular component,” which are made up of terms (otherwise known as classes) arranged in a graph structure. Terms in the graph are related to one another via relationships, which are of a given type. Examples of relations used in GO include “is_a,” “part_of,” “regulates,” “has_part” and “occurs_in.” Some relationships, namely “is_a” (is a type of) and “part_of,” are taxonomic such that “child terms” are more specific than the more general “parent” terms (Ashburner et al., 2000; Harris et al., 2004; Figure 2). Terms can have one or more “parent terms” and gene products annotated with specific child terms are automatically associated with the corresponding “parent terms” in the graph. Currently, GO uses the CHEBI database² as the source of all primary chemical names (Hill et al., 2013). Alternative names are added as synonyms. We made several additions to the CHEBI as it lacked most of the chemical entities used by the MENGO consortium. The GO also actively interacts

with and maintains mappings to EC, MetaCyc, Rhea, Reactome, and several other systems³. GO is working with Rhea and Reactome on a system of automatic import for enzymatic reactions to avoid duplication of curation effort and to further improve interoperability.

The GO is widely recognized as a powerful tool for the annotation of gene products of all organisms including important biofuel producers like *Saccharomyces cerevisiae* and cyanobacteria (Christie et al., 2009; Beck et al., 2012). The ongoing maintenance and expansion of the GO results from international collaborative efforts that are managed by the Gene Ontology consortium (GOC). The GOC encourages and works with associated groups, such as MENGO, on term development and annotation in focused subject areas.

The MENGO team has produced over 600 bioenergy-related terms^{4,5}. Included in this set are some key terms relevant to the description of advanced biofuel production. For example, terms for the microbial production of isobutanol “GO:1901961 isobutanol biosynthetic process,” isopropanol “GO:1902640 propan-2-ol biosynthetic process” and isopentenol “GO:1902934 isopentenol biosynthetic process” were additions made by the MENGO group. When specific gene products are annotated using GO terms, the

³http://wiki.geneontology.org/index.php/Enzymes_and_EC_mappings

⁴<http://www.mengo.biochem.vt.edu/>

⁵<http://tinyurl.com/o2l7nsy>

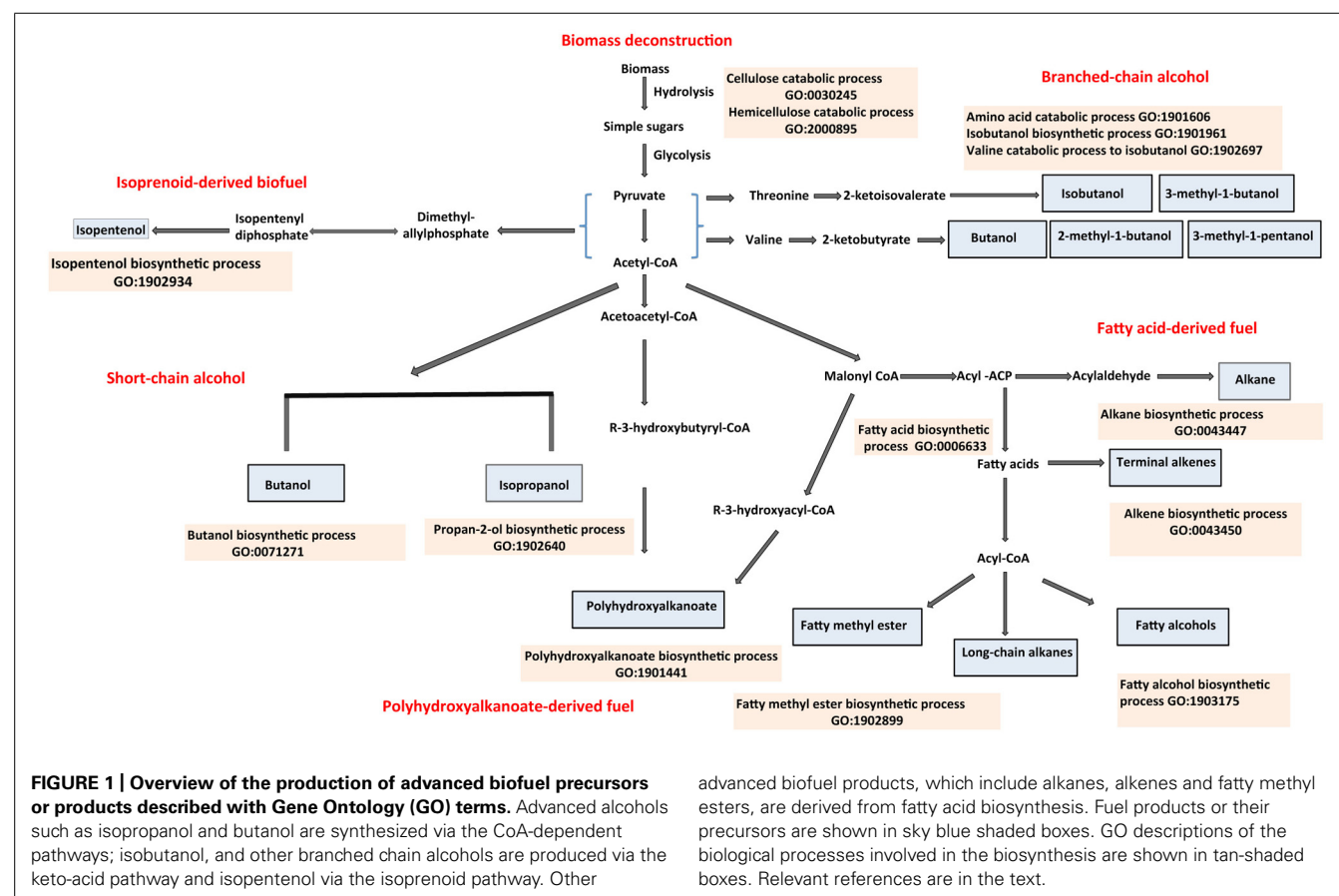


FIGURE 1 | Overview of the production of advanced biofuel precursors or products described with Gene Ontology (GO) terms. Advanced alcohols such as isopropanol and butanol are synthesized via the CoA-dependent pathways; isobutanol, and other branched chain alcohols are produced via the keto-acid pathway and isopentenol via the isoprenoid pathway. Other

advanced biofuel products, which include alkanes, alkenes and fatty methyl esters, are derived from fatty acid biosynthesis. Fuel products or their precursors are shown in sky blue shaded boxes. GO descriptions of the biological processes involved in the biosynthesis are shown in tan-shaded boxes. Relevant references are in the text.

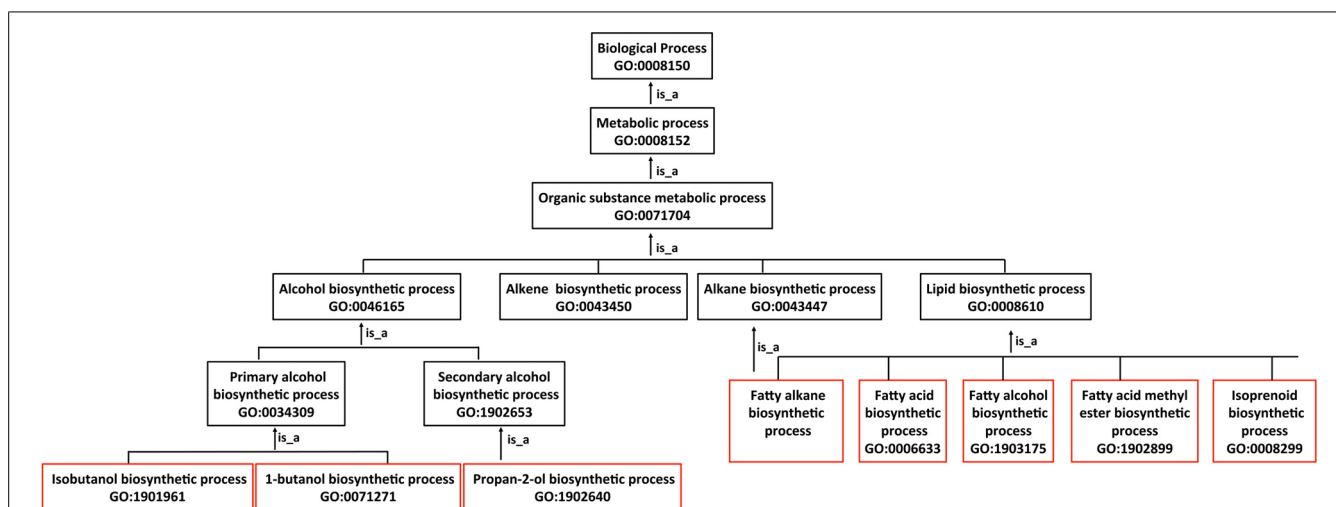


FIGURE 2 | Description of advanced biofuels using biological process terms in the Gene Ontology graph. The GO has three ontologies, Molecular Function, Biological Process and Cellular Component, which are made up of terms (otherwise known as classes) arranged in the form of a graph. Each

graph consists of a set of more specific subclasses (child) terms that form relationships (in this case “is_a”) with more general (parent) terms. As an example, biosynthesis terms for fuel and fuel-like products boxed in red are subclasses of the more general terms boxed in black.

information is stored in tabular form as “gene association” statements⁶, which are consolidated into a central GO database to enable public access to the annotations (Christie et al., 2009; Beck et al., 2012). Information stored in a gene association statement includes the gene product ID, the GO terms specifying the function of the gene product, its biological context and/or the location of activity. Experimental or bioinformatics evidence is represented by terms from the Evidence Code Ontology⁷ (ECO). Also included is the taxonomic source of the gene product and the reference(s) (often a PubMed identity number) indicating the publication(s) from which information was obtained. Selected gene products, annotated with GO bioenergy related terms are provided in **Table 1** and Supplementary Table 1. All other bioenergy related annotations made by MENGO can be found at the MENGO website⁸ or the GO website⁹ (AMIGO).

The GO has also been used extensively in the analysis of high-throughput data including genomic, transcriptomic, proteomic, and metagenomic data (Marinković et al., 2012; Molnar et al., 2012; Benso et al., 2013; Plewniak et al., 2013; Schmidt et al., 2014). Currently, the GO is limited to terms for describing natural processes (Blake and Harris, 2008). It does not provide terms for non-natural processes such as disease states, nor does it provide terms for novel biological functions that have been produced by genetic manipulations or synthetic biology. This is a current limitation of the GO in the context of bioenergy research. The MENGO team is currently discussing with the community the creation of a complementary set of GO-like terms suitable for synthetic biological functions (which we provisionally call SYNGO). To

underscore the anticipated broad scope of SYNGO, at the end of this review we provide brief descriptions of some synthetic processes and cell parts that are driving the development of this resource.

MULTIENZYME COMPLEXES FOR DECONSTRUCTION OF BIOMASS: CELLULOSOMES AND XYLANOSOMES

Complete hydrolysis of cellulose to glucose requires the synergistic action of three general types of glycoside hydrolases (Doi, 2008; Gomez del Pulgar and Saadeddin, 2014): (i) Cellulases or endo-1,4- β -glucanases randomly hydrolyzes internal bonds in a cellulose chain releasing products of varying chain lengths. Non-processive cellulases are truly random and processive cellulases, upon binding a cellulose chain continue cutting through the bound substrate. (ii) Exo-1,4- β -glucanases “GO:0031217 glucan 1,4-beta-glucosidase activity” works from either the reducing or non-reducing end of the cellulose polymer to release cellobiose. They are sometimes called exocellulases and their processive forms are known as cellobiohydrolases “GO:0016162 cellulose 1,4-beta-cellobiosidase activity.” A variation of this group, glucohydrolase “GO:0080079 cellobiose glucosidase activity,” acts on the non-reducing end and releases glucose. (iii) β -glucosidases “GO:0008422 β -glucosidase activity” hydrolyze cellobiose released by other enzymes to glucose and are also known as cellobiases (Wilson, 2011, 2012). Activities of these enzymes on oligosaccharides have also been reported and their actions from the non-reducing end release glucose, endoglucanases, exoglucanases, and cellobiases (Chang et al., 2013).

The major component of hemicellulose is xylan. Two key enzymes collectively called xylanases “GO:0097599 xylanase activity” are responsible for hydrolysis of xylan. Endo-xylanase (endo-1,4- β -xylanase) “GO:0031176 endo-1,4- β -xylanase activity” acts on the homopolymeric backbone of 1,4-linked β -D-xylopyranose

⁶<http://www.geneontology.org/GO.format.annotation.shtml>

⁷<http://www.evidenceontology.org/>

⁸<http://www.mengo.biochem.vt.edu/>

⁹<http://tinyurl.com/o2l7nys>

Table 1 | Abbreviated Gene Ontology table showing annotations of selected experimentally characterized gene products associated with advanced biofuel production.

Gene product	Ontology*	GO term	Evidence†	Taxon	Reference
Butanol biosynthetic process					
GO:0071271					
Thl	F	GO:0003985: acetyl-CoA acetyltransferase activity	IDA	1488	PMID:16347774
Hbd	F	GO:0008691: 3-hydroxybutyryl-CoA dehydrogenase activity	IDA	1520	PMID:1444364
Crt	F	GO:0003859: 3-hydroxybutyryl-CoA dehydratase activity	IDA	1488	PMID:16346566
Bcd	F	GO:0004085: butyryl-CoA dehydrogenase	IDA	1488	PMID:8655474
BYDH	F	GO:0004029: aldehyde dehydrogenase (NAD) activity	IDA	1488	PMID:8300540
BDH	F	GO:0004022: alcohol dehydrogenase activity	IDA	272562	PMID:11790753
Isobutanol biosynthetic process					
GO:1901961					
Bat1	F	GO:0004084: branched-chain aminotransferase activity	IDA	4932	PMID:8798704
Bat1	C	GO:0005739: mitochondrion	IDA	4932	PMID:8798704
Bat2	F	GO:0004084: branched-chain aminotransferase	IDA	4932	PMID:8798704
Bat2	C	GO:0005829: cytosol	IDA	4932	PMID:8798704
PDC	F	GO:0004737: pyruvate decarboxylase	IMP	4932	PMID:9546164
ADH2	F	GO:0004022: alcohol dehydrogenase activity	IMP	4932	PMID:3546317
Isopropanol biosynthetic process					
GO:1902639					
THL	F	GO:0003985: acetyl-CoA acetyltransferase activity	IDA	1488	PMID:16347774
CTF	F	GO:0008410: CoA-transferase activity	IDA	1488	PMID:2719476
ADC	F	GO:0047602: Acetoacetate decarboxylase activity	IDA	272562	PMID:2268159
SADH	F	GO:0004022: alcohol dehydrogenase activity	IDA	1520	PMID:8349550
Tolerance					
ARI1	P	GO:1990370: process resulting in tolerance to aldehyde			
Spo0A	P	GO:1990336: process resulting in tolerance to butanol	IDA	1488	PMID:15028679
Cfa	P	GO:1990336: process resulting in tolerance response to butanol	IDA	1350	PMID:24014527
Cfa	P	GO:1990337: process resulting in tolerance to isobutanol	IDA	1350	PMID:24014527
SII0690	P	GO:1990336: process resulting in tolerance to butanol	IMP	1148	PMID:23883549
Slr0947	P	GO:1990336: process resulting in tolerance to butanol	IMP	1148	PMID:23883549
slr1295	P	GO:1990336: process resulting in tolerance to butanol	IMP	1148	PMID:23883549

* Molecular function (F), biological process (P), cellular component (C).

† Inferred from direct assay (IDA), inferred from mutant phenotype (IMP).

producing xylooligomers, and β -xylosidase (xylan-1,4- β -xylosidase) “GO:0009044 xylan-1,4- β -xylosidase activity” act on the xylooligomers releasing xylose (Biely et al., 1985; Ahmed et al., 2009). Additionally, accessory enzymes such as acetyl xylan esterases “GO:0046555 acetylxylan esterase activity” act to remove the side chain substitution along the xylan backbone (Juturu and Wu, 2013). The GO uses the CHEBI database¹⁰ as the source of all primary chemical names and all alternate names are listed as synonyms. These hydrolytic enzymes exist as free independent enzymes or as multienzyme complexes called cellulosomes “GO:0043263 cellulosome” and xylanosomes “GO:1990358 xylanosome” (Jiang et al., 2004, 2006). This section will focus on using the GO to describe these multienzyme

complexes that mediate degradation of polysaccharides. Other aspects of deconstruction of biomass for biofuel can be found in several other reviews (Dodd and Cann, 2009; Blanch et al., 2011; Chundawat et al., 2011; Blumer-Schuette et al., 2014).

The cellulosome of *Clostridium thermocellum* serve as a paradigm for this enzymatic nanomachine, and thus a model for studies of the structure and assembly process. The central component of the cellulosome is a non-catalytic “scaffoldin” subunit, which mediates a highly specific interaction between the enzyme-bearing type I dockerin modules and the resident type I cohesin modules (Bras et al., 2012; Smith and Bayer, 2013; Hong et al., 2014). The cellulose-binding domain (CBD) of the scaffoldin subunit aids in binding the complex to the cellulosic substrate. The type II dockerin module, which forms

¹⁰<http://www.ebi.ac.uk/chebi/>

part of the primary scaffoldin in turn, binds to the type II cohesin of anchoring scaffoldins, which connect to the bacteria cell surface (**Figure 3**; Bayer et al., 1985; Bayer and Lamed, 1986). Recently, a third type of dockerin–cohesin interaction (type III) has been characterized in *Ruminococcus flavefaciens*. The type III dockerins and cohesins show a high degree of sequence divergence compared to their type I and II counterparts (Rincon et al., 2005; Karpol et al., 2013). Prior to our work, GO terms related to cellulosomes were “GO:0043263 cellulosome,” its regulation terms, and the different hydrolytic enzyme terms. However, the latter terms were not necessarily linked to the cellulosome term in a way that clearly indicated their associations. We used the “domain binding” aspect of the GO to make the multiple functions of the scaffoldin more obvious. To start with, the ability of the non-enzymatic scaffoldin to bind to several cellulose degradative enzymes via the cohesin domain makes it a complex. Therefore we introduced the term “GO:1990296 scaffoldin complex” which has a “part of” relationship to “GO:0043263 cellulosome.” In addition, domain-specific terms were also developed, namely: “GO:1990311 type I cohesin domain binding,” “GO:1990308 type I dockerin domain binding,” “GO:1990312 type II cohesin domain binding,” “GO:1990309 type II dockerin domain binding,” “GO:1990313 type III cohesin domain binding” and “GO:1990310 type III dockerin domain binding.” Gene products binding to each of

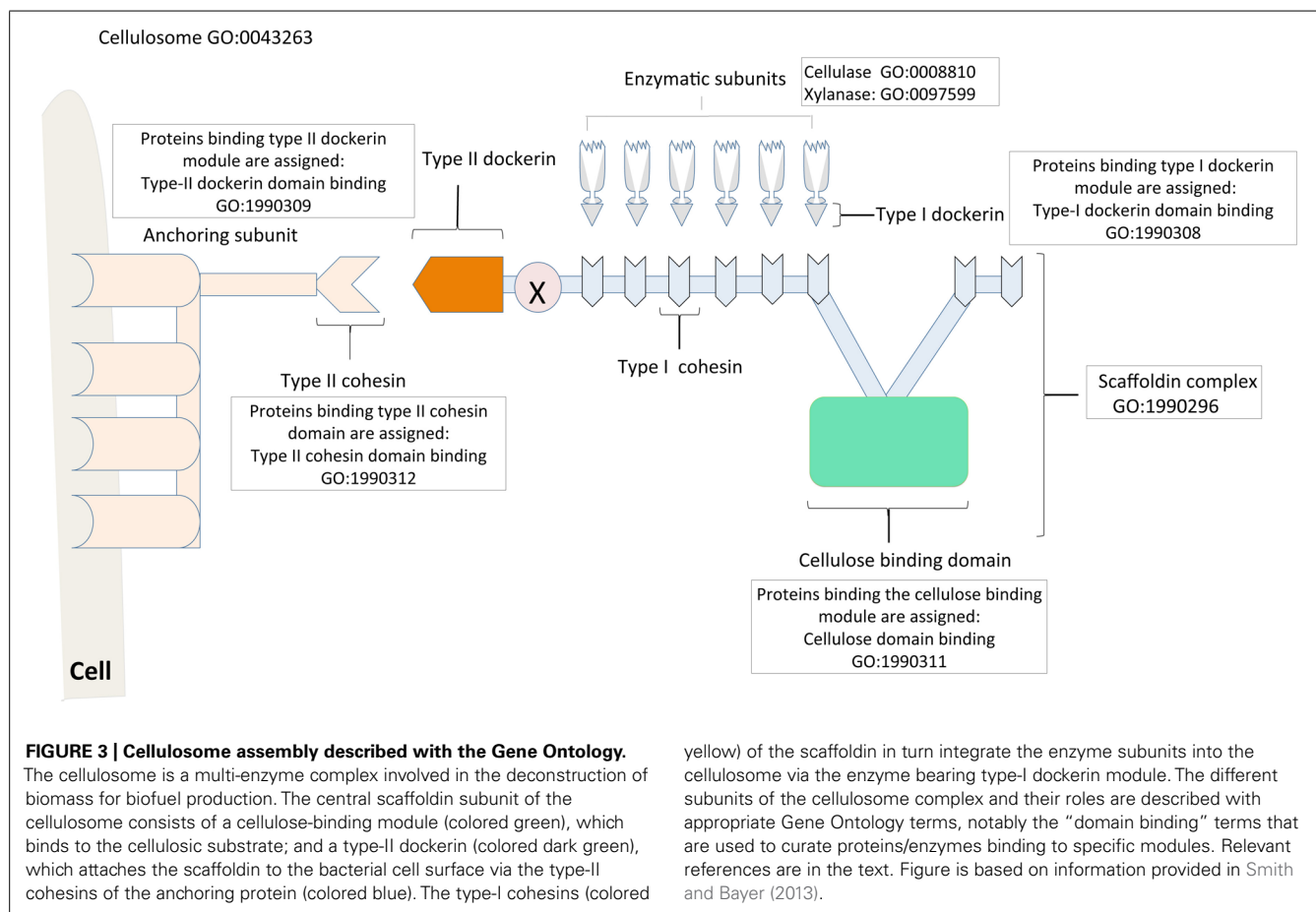
these modules were assigned the appropriate domain binding term.

The xylanosome occurs naturally in anaerobic fungi and bacteria. However, besides containing several hemicellulases, not much is known about its assembly. We introduced the term “GO:1990358 xylanosome” into the GO. A modular nature, if any, has not been described for the xylanosome. Therefore for now, the only way to identify components of the xylanosome will be at the annotation level where the hemicellulases are annotated with their unique enzymatic activity terms using the Molecular Function ontology and also with the cellular component term “GO:1990358 xylanosome” where there is experimental evidence of an association with the xylanosome.

The synergy and combined action of hemicellulases and cellulases within these multi-enzyme systems makes them efficient in the degradation of biomass (Morais et al., 2010). Systematic annotation of the ever-increasing repertoire of dockerin–cohesin pairs of the cellulosome, using the GO terms described above, will serve to inform their use in engineering more efficient cellulosomes, xylanosomes, and cellulosome-like complexes.

PRODUCTION OF ISOPROPANOL AND BUTANOL

Clostridia can produce isopropanol and butanol from sugars via pyruvate and acetyl-CoA [for a review of early work in this area see Stephenson (1949) and references therein]. Short chain



In *Clostridium*, the physiological transition from acidogenesis to solventogenesis is associated with the stationary phase of the organism's growth cycle. The *Clostridium* Spo0A gene product has been shown to regulate solventogenesis (Ravagnani et al., 2000; Harris et al., 2002; Dürre and Hollergschwandner, 2004; Hu et al., 2011) and activate the downstream genes of the acidogenic process,

in addition to controlling sporulation “GO:0043934 sporulation.” We annotated Spo0A with the term “GO:1902930 regulation of alcohol biosynthetic process” which has “regulation of solventogenesis” as a synonym, as well as with the term “GO:0043937 regulation of sporulation.” Another gene product, SpoIIE was shown to control sporulation in *C. acetobutylicum* but was not associated with solvent synthesis (Scotcher and Bennett, 2005; Bi et al., 2011). The multiple roles of Spo0A are thus documented as separate entries in the GO association table with its role in sporulation overlapping that of SpoIIE. Biosynthesis of butanol and isopropanol share the same metabolic pathway from pyruvate to acetyl-CoA and thereafter follow respective branches (Figure 4).

Microbial production of butanol occurs naturally in certain *Clostridium* species as part of the acetone butanol ethanol (ABE) pathway (Lee et al., 2008a; Qureshi et al., 2010). Some of these bacteria can use both pentose and hexose sugars as substrates. Typically, the breakdown of pentose and hexose sugars to pyruvate follows different paths but subsequent steps from pyruvate to butanol production may comprise the same reactions. The GO terms “GO:1990284 hexose catabolic process to 1-butanol” and “GO:1990290 pentose catabolic process to 1-butanol” are child terms of “GO:0071271 1-butanol biosynthetic process.” In the so-called CoA-dependent pathway for butanol production, the generation of butanol is initiated with two molecules of acetyl-CoA. Six enzymes encoded by seven genes mediate reactions leading to the production of butanol. Two molecules of acetyl-CoA are condensed, reduced, and dehydrated to form crotonyl-CoA, which is then reduced to butanol. The enzymes responsible for these reactions are shown in Figure 4 and Table 1, with appropriate GO annotations. The key enzymes leading to butanol production are each described with a distinct molecular function term. The transition from acidogenesis to solventogenesis in some *Clostridium* species occurs at low pH as a result of acid accumulation. Butyrate, produced during acidogenesis, becomes a substrate for butanol production in the solventogenesis phase. This process is mediated by the enzymes butyrate acetoacetate CoA-transferase “GO:0047371 butyrate acetoacetate CoA-transferase activity,” butyraldehyde dehydrogenase and butanol dehydrogenase “GO:1990362 butanol dehydrogenase activity.”

Isopropanol is naturally produced by *Clostridium beijerinckii* and *Clostridium aurantibutyricum* (George et al., 1983), although the natural yield is very low (Chen and Hiu, 1986). It can be used directly as an additive to gasoline or as a feedstock for biodiesel production (Lee et al., 1995); in the latter case esterification with isopropanol is used to reduce the chances of crystallization at low temperatures. The native pathway involves the condensation of two molecules of acetyl-CoA into a molecule of acetoacetyl-CoA; the coenzyme A is then transferred to acetate or butyrate, catalyzed by CoA transferase “GO:0008410 CoA-transferase activity.” The two child terms of CoA transferase are “GO:0008775 acetate CoA transferase” and “GO:0047371 butyrate-acetoacetate CoA-transferase.” A reaction catalyzed by acetoacetate decarboxylase “GO:0047602 acetoacetate decarboxylase activity” converts acetoacetate to acetone. Subsequently acetone is converted to isopropanol in a

NADPH-dependent reaction catalyzed by a secondary alcohol dehydrogenase (ADH) “GO:0050009 isopropanol dehydrogenase activity.”

ISOBUTANOL PRODUCTION FROM VALINE AND GLYCINE

The branched-chain alcohol isobutanol, (2-methylpropan-1-ol) exhibits superior physiochemical properties similar to butanol. Isobutanol is naturally produced during fermentation by *S. cerevisiae* albeit in low amounts (Dickinson et al., 1998; Branduardi et al., 2013). Dickinson and coworkers examined the metabolic pathways used in the degradation of valine to isobutanol. Catalytic breakdown of valine to isobutanol is mediated by the Ehrlich pathway, which starts with transamination of valine to α -ketovalerate via the branched chain amino acid aminotransferase, Bat2, which is also known as Twt2p and Eca40p. The alternate names are curated in the GO association table under “database object symbol synonyms.” The location of activity of gene products is described with the GO Cellular Component ontology. Bat2 has been shown to be localized in the cytosol “GO:0005829 cytosol” and a homolog Bat1 is localized to the mitochondrial matrix “GO:0005759 mitochondrial matrix,” both having the same molecular function “GO:0004084 branched chain amino acid aminotransferase” (Table 1). The subsequent decarboxylation step resulting in the formation of isobutyraldehyde is mediated by ketoacid decarboxylase. All of the three isozymes of pyruvate decarboxylase (Pdc1p, Pdc5p and Pdc6p) have been shown to be capable of decarboxylating α -ketovalerate (Dickinson et al., 1998). Finally a reduction step mediated by ADH converts the aldehyde to isobutanol (Dickinson et al., 1998). The PDCs are each assigned the same GO Molecular Function terms “GO:0004737 pyruvate decarboxylase activity” in separate entries in the GO association table and the aldehyde dehydrogenase gene is assigned “GO:0004022 ADH activity.” All these gene products are also associated with the biological process term “GO:1901961 isobutanol biosynthetic process.”

Another pathway was recently discovered in *S. cerevisiae* for *de novo* isobutanol biosynthesis. A study by Villas-Bôas et al. (2005) suggested that glycine deamination led to the generation of glyoxylate and subsequently the formation of α -ketovalerate and α -isoketovalerate. Based on the above work, Branduardi et al. (2013) conducted a study to decipher the components of the pathways leading to the formation of butanol and isobutanol with glycine as substrate. Briefly, glycine is converted into serine through serine hydroxymethyltransferase (Shm2; McNeil et al., 1994), which is deaminated by serine deaminase CHA1 to form pyruvate. Pyruvate is then converted to isoketovalerate and then to isobutanol following the reactions as described above. These findings emphasize that it is important to define formation of a product based on the substrate utilized, as sometimes the same end is reached from multiple starting points. The terms “GO:1902697 valine catabolic process to isobutanol” and “glycine catabolic process to isobutanol” can appropriately capture this difference and these two terms are linked to the parent term “GO:1901961 isobutanol biosynthetic process.” The pathway also produces butanol, but since a high level of isobutanol production was observed, McNeil et al. (1994) hypothesized that carbon flux was being shunted away

from butanol production toward isobutanol production by the isomerization of α -ketovalerate into α -isoketovalerate. Analogous to these natural processes in *S. cerevisiae*, Atsumi et al. (2008) employed a metabolic engineering approach, which involved diverting 2-keto acid intermediates from the amino acid biosynthetic pathway of *Escherichia coli* to produce other branched-chain alcohols such as 2-methyl-1-butanol, 3-methyl-1-butanol and 2-phenylethanol.

FATTY ACID BIOSYNTHESIS AND FATTY ACID-DERIVED BIOFUELS

Fatty acid metabolism has attracted the most attention as a biological route to convert sugars to liquid transportation fuels (Zhou and Zhao, 2011; Lennen and Pfleger, 2013; Janssen and Steinbuchel, 2014). It is particularly suited to provide precursors for advanced biofuel because of its high efficiency and the high-energy content of the end product (Atsumi and Liao, 2008; Rude and Schirmer, 2009; Peralta-Yahya and Keasling, 2010; Wen et al., 2013). Additionally, natural metabolic pathways have been identified that convert these precursors into the biofuel product. Generally, the first committed step in fatty acid biosynthesis in *E. coli* is the conversion of acetyl-CoA to malonyl-CoA catalyzed by acetyl-CoA carboxylase (AccABCD) “GO:0003989 acetyl-CoA carboxylase activity,” which is a protein complex (GO:0009317 acetyl CoA carboxylase complex) comprising four subunits (Li and Cronan, 1992). Malonyl-CoA is then transferred to acyl carrier protein (ACP) via a malonyl-CoA:ACP transacylase (FabD, GO:0004314 malonyl-CoA:ACP transacylase activity; Campbell and Cronan, 2001; Oefner et al., 2006). Subsequently cycles of fatty acid elongation are initiated by malonyl-ACP and acetyl CoA catalyzed by β -ketoacyl-ACP synthase III (FabH, GO:0033818 beta-ketoacyl-ACP synthase III activity; Lai and Cronan, 2003). FabB and FabF use acyl-ACP as substrates to initiate successive chain elongation reactions (Feng and Cronan, 2009). A β -hydroxyacyl-ACP is the product of the second step in elongation, generated by β -ketoacyl ACP reductase (FabG, GO:0004316 beta-ketoacyl ACP reductase activity) while expending one molecule of NADPH. FabA and FabZ catalyze formation of an enoyl-ACP (Mohan et al., 1994). Following enoyl-ACP formation, the last intermediate in the fatty acid elongation cycle, acyl-ACP is formed by FabI with the consumption of NADPH (Heath and Rock, 1995). Two thioesterases (TEs) in *E. coli*, TesA and TesB release the fatty acid chains from the ACP to produce free fatty acids (Lee et al., 2006). In the biofuel field there are preferences for fatty acids of various chain lengths (Lennen and Pfleger, 2012; Torella et al., 2013). For example, longer chain products (C_{12} – C_{20}) fall within the diesel range, which have high energy densities (Wen et al., 2013). The GO has terms describing the synthesis of fatty acids of different chain lengths: for short-chain fatty acids, “GO:0051790 short-chain fatty acid biosynthetic process”; for medium-chain fatty acids, “GO:0051792 medium-chain fatty acid biosynthetic process”; for long-chain fatty acids “GO:0042759 long-chain fatty acid biosynthetic process”; and for very long-chain fatty acids, “GO:0042761 very long-chain fatty acid biosynthetic process.” While free fatty acids are valuable, they cannot be used directly as fuels and must first be converted either to fatty acid alkyl esters (for biodiesel), or to fatty acid-derived alkanes, alkenes

or fatty alcohols (Zhang et al., 2011; Peralta-Yahya et al., 2012; Wen et al., 2013).

FATTY ACID-DERIVED ALKANES AND ALKENES

Alkanes, an integral part of fossil fuels (gasoline, diesel and jet fuel), are naturally found in diverse organisms including plants, insects and microbial species, but the genetic and biochemical bases behind the production of alkanes have been elusive. Those with C4–C23 carbon chain length possess higher energy densities, hydrophobic properties and compatibilities with existing liquid fuel infrastructure. Most evidence supporting the decarbonylation of aldehydes (a fatty acid metabolite) as the primary mechanism for alkane production have been obtained in eukaryotic systems (Cheesbrough and Kolattukudy, 1984; Dennis and Kolattukudy, 1991, 1992). This in part informed the identification of the *Arabidopsis cer* gene as encoding a protein with decarbonylase activity (GO:0071771 decarbonylase activity) involved in alkane biosynthesis (Aarts et al., 1995). It is only recently that the pathway of alkane biosynthesis was elucidated in cyanobacteria (Figure 5; Schirmer et al., 2010; Li et al., 2011, 2012) making this class of hydrocarbons eligible to be categorized as a next generation biofuel (advanced). In this pathway, fatty acids are reduced by a two-step reaction to alkanes. Fatty acyl-ACP (fatty acid metabolite) is reduced to a fatty aldehyde via a fatty acyl-ACP reductase. This is then followed by a deformylation step catalyzed by an aldehyde deformylating oxidase “GO:1990465 aldehyde oxygenase (deformylating) activity,” resulting in alkane production (Schirmer et al., 2010; Li et al., 2011, 2012; Warui et al., 2011; Coates et al., 2014). Commonly found hydrocarbons in cyanobacteria are heptadecane (GO:1900636 heptadecane biosynthetic process) and methyl heptadecane (Schirmer et al., 2010; Zhang et al., 2011), which have cetane numbers of 105 and 66 respectively (requirement for US diesel or ASTM standard, 47 minimum (Rashid et al., 2008) making the hydrocarbon products from cyanobacteria ideal candidates for diesel fuel applications.

Another group of hydrocarbons produced from fatty acid derivatives are olefins (alkenes, GO:0043450 alkene biosynthetic process). In *Jeotgalicoccus* sp., the terminal alkenes, 8-methyl-1-nonadecene and 17-methyl-1-nonadecene have been identified. A terminal olefin-forming fatty acid decarboxylase belonging to the cytochrome P450 family of enzymes (OleT) has been identified as the key enzyme involved in the production of these olefins in *Jeotgalicoccus* sp. (Rude et al., 2011). Other studies proposed a head-to-head condensation of fatty acids, which involves the formation of a carbon-to-carbon bond between the carboxyl carbon of one fatty acid and the α -carbon of another fatty acid as another mechanism to generate long chain (C_{23} – C_{33}) olefins. OleA, a homolog of the condensing enzyme in fatty acid biosynthesis (FabH-3-oxoacyl-ACP ketosynthase (KS) III), was identified as a key enzyme in long chain olefin production (Beller et al., 2010; Sukovich et al., 2010). In cyanobacteria, the olefin-producing pathway (OLS) involves a polyketide synthase (GO:0034081 polyketide synthase complex) that first elongates fatty acyl-CoA with two carbons from malonyl-CoA via KS and acyl transferase (AT) domains. This is followed by reduction to the hydroxyacid by a ketoreductase (KR, GO:0045703 ketoreductase activity; Mendez-Perez et al.,

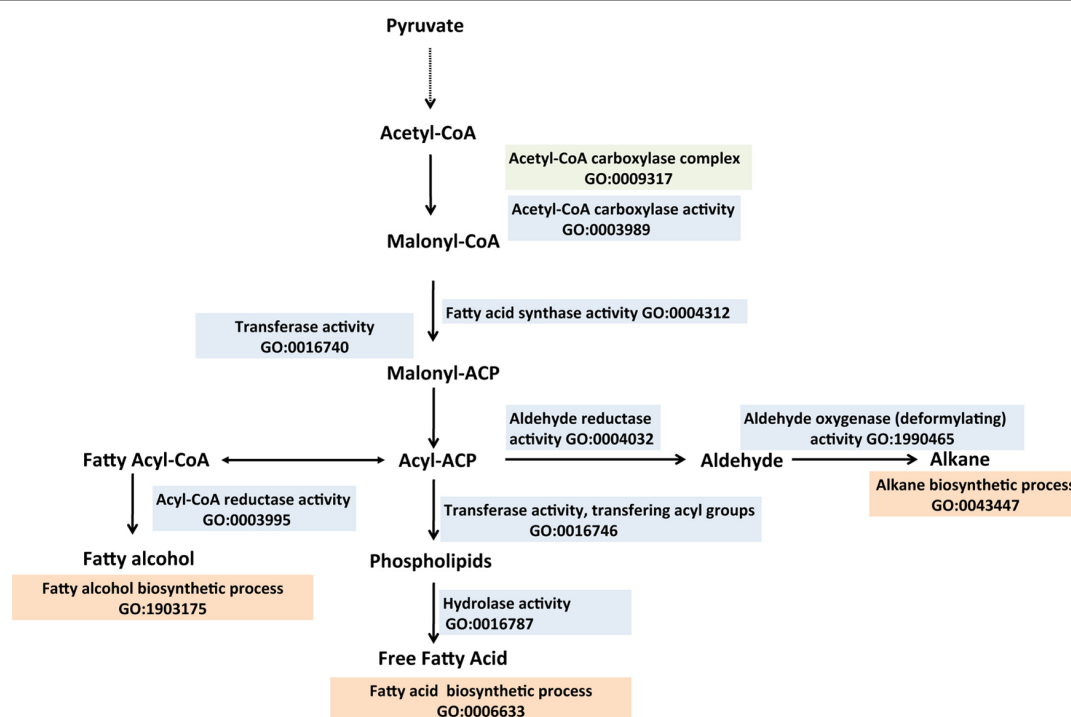


FIGURE 5 | Biosynthesis of fatty acids, fatty alkanes and fatty alcohols described with the Gene Ontology. Fatty acids are generated from acetyl-CoA and malonyl-CoA precursors through fatty acid synthases. Acyl-ACPs, which are intermediates of the fatty acid biosynthetic process, are reduced to fatty alcohols or deformylated to form alkanes. Enzymes

involved in the biosynthesis of the fatty acid derivatives are described with appropriate GO Molecular Function terms (colored sky blue), Biological Process terms (tan) and the Cellular Component terms (green). Relevant references are in the text. Figure is based on information provided in Lennen and Pfleger (2013).

2011). In the final step, sulfotransferases (ST, GO:0008146 sulfotransferase activity) activate the β -hydroxy group via sulfonation, and then a TE acts on this substrate to catalyze decarboxylation and loss of sulfate to form the terminal alkene (McCarthy et al., 2012).

FATTY ACID ALKYL ESTERS (BIODIESEL)

Biodiesel is a substitute for petroleum-based diesel fuel. Like its counterparts described above, biodiesel has properties similar to those of diesel, and therefore, can be used in the diesel engines and stored and distributed using the existing infrastructure. Other advantages of biodiesel include reduced fuel toxicity and increased lubricity. Additionally the use of biodiesel leads to lower carbon monoxide and soot emissions than conventional diesel fuels. Biodiesel is traditionally produced by the trans-esterification of mostly plant-derived triacylglycerols yielding glycerol and fatty acid alkyl esters (FAAE), particularly fatty acid methyl esters (FAMEs; “GO:1902899 FAME biosynthetic process”). However, biodiesel production is limited by the availability of inexpensive vegetable oil feedstocks. This has prompted a search for sustainable alternatives. Direct microbial production of FAAEs is an area of intense research as it bypasses the transesterification step, reducing cost and energy and also avoids the use of methanol, (Kalscheuer et al., 2006; Nawabi et al., 2011) which is an expensive and toxic feedstock.

ISOPRENOID-DERIVED BIOFUEL

Isoprenoids, also called terpenes, have been evaluated for pharmaceutical, nutritional and fuel products (Bohlmann and Keeling, 2008; Peralta-Yahya and Keasling, 2010). Two independent biosynthetic pathways for isoprenoid production “GO:0008299 isoprenoid biosynthetic process” are found in nature (Figure 6). The methylerythritol phosphate (MEP) pathway “GO:1902768 isoprenoid biosynthetic process via 1-deoxy-D-xylulose 5-phosphate” and the mevalonate pathway “GO:1902767 isoprenoid biosynthetic process via mevalonate” have evolved for the production of the key five-carbon isoprenoid intermediates, isopentenyl diphosphate (IPP), and dimethylallyl diphosphate (DMAPP; Beytia and Porter, 1976; Edwards et al., 1992; Figure 6). The MEP pathway consists of seven steps resulting in the conversion of glyceraldehyde-3-phosphate and pyruvate to IPP and DMAPP and it is found in most bacteria, chloroplast, unicellular eukaryotes, and certain parasites (Zhao et al., 2013; Jarchow-Choy et al., 2014; Figure 6). The mevalonate pathway is responsible for all the isoprenoid production in archaea, some bacteria and most eukaryotes (Miziorko, 2011; Figure 6). It converts acetyl CoA in six steps to IPP via the key intermediate mevalonate. Specifically, acetyl-CoA and acetoacetyl-CoA are condensed into 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), which is reduced to mevalonate via HMG-CoA reductase (HMGR, GO:0042282 HMG-CoA reductase activity). Following this are two phosphorylation steps, which convert mevalonate

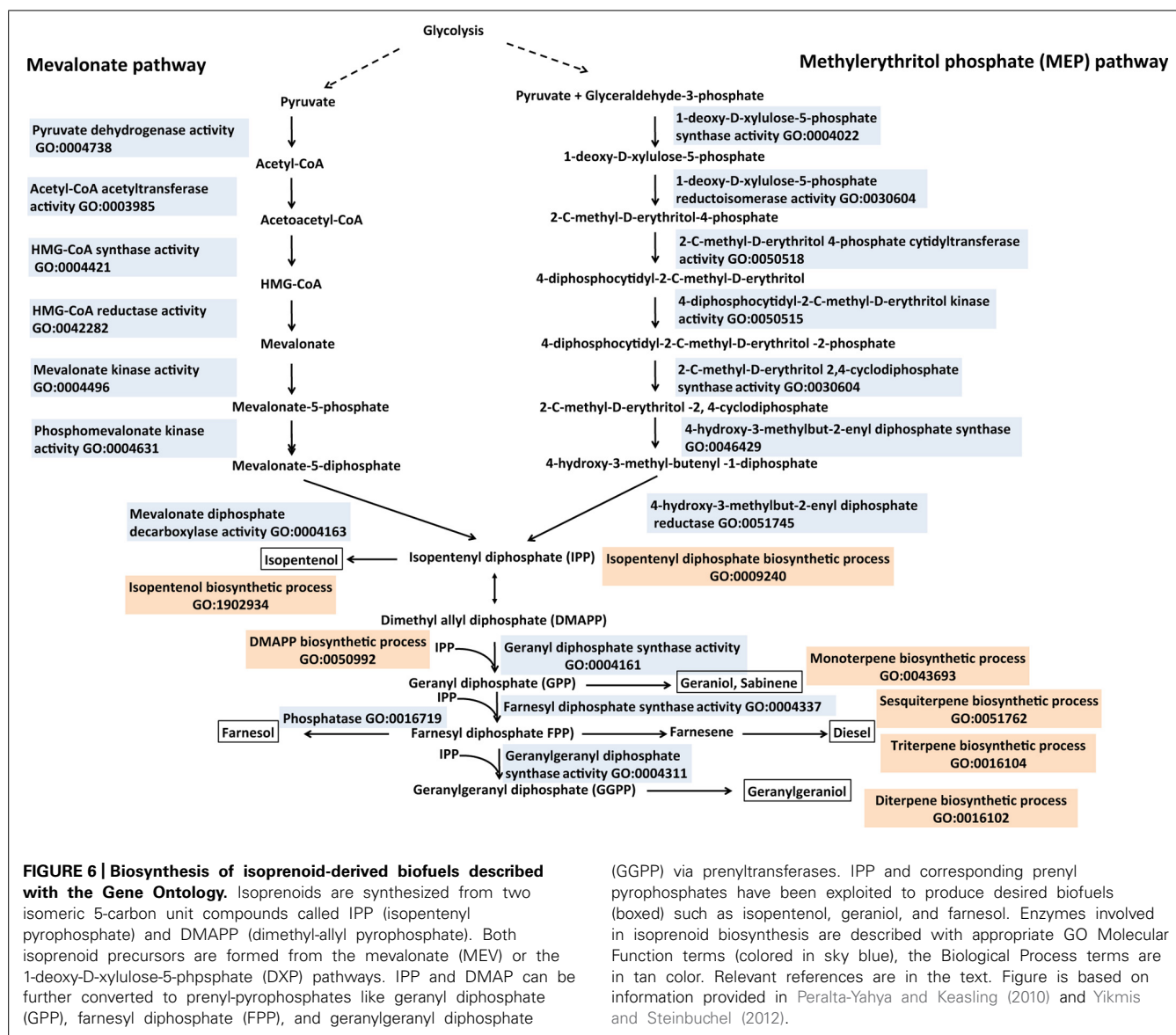


FIGURE 6 | Biosynthesis of isoprenoid-derived biofuels described with the Gene Ontology. Isoprenoids are synthesized from two isomeric 5-carbon unit compounds called IPP (isopentenyl pyrophosphate) and DMAPP (dimethyl-allyl pyrophosphate). Both isoprenoid precursors are formed from the mevalonate (MEV) or the 1-deoxy-D-xylulose-5-phosphate (DXP) pathways. IPP and DMAPP can be further converted to prenyl-pyrophosphates like geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) via prenyltransferases. IPP and corresponding prenyl pyrophosphates have been exploited to produce desired biofuels (boxed) such as isopentenol, geraniol, and farnesol. Enzymes involved in isoprenoid biosynthesis are described with appropriate GO Molecular Function terms (colored in sky blue), the Biological Process terms are in tan color. Relevant references are in the text. Figure is based on information provided in Peralta-Yahya and Keasling (2010) and Yikmis and Steinbuchel (2012).

pyrophosphatase “GO:0016462 pyrophosphatase activity,” nudF from *B. subtilis*, was shown to produce isopentenol in *E. coli*. Curating the biosynthesis of the intermediates and the vast diversity of terpene synthases “GO:0010333 terpene synthase activity” therefore will provide a useful resource, which will inform decisions in the synthesis of terpene-based fuels. A new class of terpenes (C_{35}), the sesquiterpenes has recently been classified (Sato, 2013) and it will be interesting to know if these compounds could provide precursors for terpene-based biofuels.

POLYHYDROXYALKANOATE-DERIVED BIOFUEL

Microbially produced PHAs have attracted attention as biodegradable polyesters and quite recently as biofuels (Zhang et al., 2009). PHAs are synthesized inside cells during oxygen, phosphorous or nitrogen starvation in the presence of excess carbon and found as insoluble cytoplasmic inclusions called polyhydroxyalkanoate

granules (GO:0070088 PHA granule; Pfeiffer and Jendrossek, 2012; Jendrossek and Pfeiffer, 2013). Synthesis of a PHA could occur through several metabolic pathways and from a variety of carbon precursors including glucose, glycerol, and fatty acids (Park et al., 2012; Shahid et al., 2013). PHAs comprise over 150 both homo- and hetero-polymers of which poly-3-hydroxybutyrate (poly (3HB), P (3HB) or PHB) being the most abundant naturally produced and most studied polyhydroxyalkanoate “GO:0042618 polyhydroxybutyrate biosynthetic process” (Laycock et al., 2014). On the basis of the carbon chain lengths of the monomers, PHAs are divided into two main groups: short chain length PHAs (scl-PHA) containing monomer units with 3–5 carbon atoms, and medium chain length PHAs (mcl-PHA) composed of monomer units with 6–18 carbon atoms. Based on the bacterial strains, PHA synthases and carbon sources involved, products of PHA biosynthesis can be a homopolymer, copolymer, block polymer or even blends (Lu et al., 2009). Functional groups such as unsaturated bonds, benzene, halogens and cyclic chemicals and epoxides can also modify PHA structures, opening up the potential for the production of a vast number of diverse PHAs. The biosynthesis of PHA “GO:190144 polyhydroxyalkanoate biosynthetic process” is quite complex and involves several enzymes that are directly or indirectly involved in PHA synthesis (Reddy et al., 2003; Park et al., 2012). PHA synthesis is well studied in the model organism *Ralstonia eutropha* (Riedel et al., 2014). In this organism, the carbon source is converted into coenzyme A thioesters of (R)-hydroxyalkanoic acids. Following this step, β ketothiolase “GO:0003988 acetyl-CoA C-acyltransferase activity” catalyzes the condensation of two coenzyme-A thioester monomers such as acetyl-CoA and propionyl-CoA. An (R)-specific reduction step involving acetoacetyl-CoA reductase “GO:0018454 acetoacetyl-CoA reductase activity” produces (R)-3 hydroxybutyryl-CoA or (R)-3 hydroxyvaleryl-CoA which is then converted into PHA by the action of PHA synthase. In regards to biofuels, 3-hydroxyalkanoates (3HAs; mcl PHAs) are linked by ester bonds formed with the hydroxyl group (–OH) of one monomer and the carboxyl (–COOH) group of the other monomer through the catalysis by various PHA synthases (Reddy et al., 2003). These hydroxyalkanoate esters (3HA esters) have been found to be similar to methyl esters of long chain fatty acids (in biodiesel). As such they can be used as fuel additives. Hydroxybutyrate and hydroxyalkanoate methyl esters (3HBME and 3HAME), generated from esterification of scl PHB and mcl PHA with methanol, respectively, are considered equivalent to ethanol (Gao et al., 2011). These esters can also be used as gasoline and biodiesel additive (Gao et al., 2011).

TOLERANCE TO BIOFUELS AND BY-PRODUCTS

Accumulation of biofuels as well as by-products often affect the integrity of the fuel producing organism's cell membrane and also impairs metabolic pathways associated with cell growth, thus compromising product titers (Heipieper et al., 2007; Segura et al., 2012). Most native producers of biofuels are sensitive to the solvents they produce. Advances are being made toward removing this block through investigations on the mechanisms of tolerance

to product accumulation (McEvoy et al., 2004; Fujita et al., 2006; Liu and Qureshi, 2009a; Ghiaci et al., 2013; Zingaro and Terry Papoutsakis, 2013) and several organisms are being investigated as better choices for biofuel production from this standpoint (Dunlop, 2011). These efforts will be substantially aided by curating all gene products associated with native or improved tolerance to biofuels and their by-products. Tolerance is considered a phenotype, not a process and is therefore out of the scope for GO Biological Process ontology. To circumvent this problem, we created the terms “process resulting in tolerance to x” instead of “tolerance to x” (x is any compound). Here we present some examples.

The common inhibitors to microbial growth that are found in lignocellulose hydrolysates include aldehydes, ketones, phenols and organic acids (Klinke et al., 2004; Jonsson et al., 2013). A *S. cerevisiae* gene, *ARI1*, encoding a novel NADPH-dependent aldehyde reductase has been shown to be involved in providing tolerance to inhibitors such as furfural, vanillin and cinnamaldehyde that arise from lignocellulose hydrolysis (Liu and Moon, 2009b). We annotated the gene product ARI1 with GO terms for its molecular role “GO:0018455 aldehyde reductase (NADPH/NADH) activity” and a biological process term “GO:1990370 process resulting in tolerance to aldehyde” to describe its ability to confer tolerance to several aldehydes. Other notable *S. cerevisiae* enzymes with efficient aldehyde reduction activities include ADHs (ADH1, ADH6 and ADH7; GO:0004022 ADH activity), aldehyde dehydrogenase (ALD4; GO:0004030 aldehyde dehydrogenase activity) and methylglyoxal reductases (GRE2 and GRE3; GO:0043892 methylglyoxal reductase activity; Petersson et al., 2006; Laadan et al., 2008).

Tolerance to alcohols such as butanol in *C. acetobutylicum* is influenced by the multifunctional Spo0A protein, which, as discussed in an earlier section, controls sporulation and solventogenesis (Alsaker et al., 2004). Kanno et al. (2013) obtained several aerobic and anaerobic bacterial isolates from soil, spanning diverse genera that tolerate butanol and isobutanol levels greater than 2% (vol/vol). The *cfa* gene, which encodes cyclopropane fatty acid (CFA) synthase in one of these bacteria (belonging to the *Firmicutes* phylum), confers solvent tolerance in recombinant *E. coli* (Kanno et al., 2013). CFA is annotated with the terms “GO:1990336 process resulting in tolerance to butanol” and “GO:1990337 process resulting in tolerance to isobutanol.” A quantitative transcriptomic analysis revealed that in the cyanobacterium *Synechocystis* sp. PCC 6803, which uses solar energy and carbon dioxide as sole energy and carbon sources, over 250 genes are induced upon exposure to butanol (Zhu et al., 2013). Of these, three, *slr0690*, *slr0947*, and *slr1295* were further characterized using knock-out mutants. The results indicated butanol sensitivity in strains that lacked any of these genes, indicating their involvement in resistance to butanol. Using a genome-scale analysis in *S. cerevisiae*, González-Ramos et al. (2013) demonstrated the role of protein degradation in tolerance to C3 and C4 alcohols (butanol, 2-butanol, isobutanol, and isopropanol). Specifically, the YLR224W gene was found to be associated with increased butanol tolerance and it encodes a subunit of the Skp-Cullin_F-box (SCF) ubiquitin ligase that recognizes damaged proteins. The

efflux pumps involved in the extrusion of toxins from cells have been considered appropriate candidates for studies on solvent tolerance (Segura et al., 2012). Such a pump consists of multiple proteins forming a multicomponent complex “GO:1990281 efflux pump complex,” which span the inner to the outer membrane of bacteria cells (Andersen, 2003; Du et al., 2014). An exposure to solvents induces the expression of *srpABC* genes in *P. putida* (Kieboom et al., 1998b). The SrpABC share considerable sequence similarities to multi-drug efflux pump proteins (Kieboom et al., 1998a). Specifically, SrpA functions as a periplasmic linker protein, SrpB as an inner membrane transporter, and SrpC, as an outer membrane channel. For these reasons, besides annotations for their roles in efflux pump and cellular locations (“GO:0015562 efflux transmembrane transporter activity” from Molecular Function ontology; “GO:1990281 efflux pump complex,” “GO:0042597 periplasmic space,” “GO:0009276 Gram-negative-bacterium type cell wall,” and “GO:0019867 outer membrane from the Cellular Component ontology”), we have qualified these proteins with a biological process term “GO:1990367 process resulting in tolerance to organic substance.” Rojas et al. (2001) have identified three different solvent efflux pumps in *P. putida*, TtgABC, TtgDEF, and TtgGHI, which extrude toluene. TtgABC also could extrude antibiotics (Ramos et al., 1998; Mosqueda and Ramos, 2000; Rojas et al., 2001). All these multi-protein efflux pumps are assigned the GO term “GO:1990281 efflux pump complex.” Hydrocarbons including nonane, decane, and undecane were shown to be toxic to *S. cerevisiae* when these solvents accumulated inside the cell. A transcriptomic analysis identified modified cell membrane and efflux pumps as contributing to alkane export and tolerance. Specifically efflux pumps Snp2p and Pdr5 were shown to reduce intracellular levels of decane and undecane thereby enhancing tolerance to the alkanes “GO:1990373 process resulting in tolerance to alkane” (Ling et al., 2013).

SYNTHETIC BIOLOGY

The yields of most fuels or fuel derivatives in the native biological systems discussed above are insufficient to be cost-competitive with petroleum-derived fuels, and therefore the technologies of metabolic engineering and synthetic biology are being employed to increase production and/or generate entirely new fuels with superior qualities. As mentioned in the introduction, currently the GO is limited to terms for describing natural processes. The MENGO team is currently discussing with the community the creation of a complementary set of GO-like terms suitable for the annotation of synthetic biological systems, which we provisionally call SYNGO. To underscore the anticipated broad scope of SYNGO, we provide here brief descriptions of selected synthetic processes and parts that are driving the development of commercially viable routes for the production of advanced biofuels.

Based on the modular nature of cellulosomes and the availability of a variety of dockerin-cohesin pairs, efforts are underway to construct designer cellulosomes, xylanosomes, and cellulosome-like structures. In such a precision-engineered multienzyme complex, the molecular architecture and enzyme content are well controlled, and the result is enhanced synergistic deconstruction of biomass (Mitsuzawa et al., 2009; Nordon et al., 2009; Smith and Bayer, 2013; Srikrishnan et al., 2013; Vazana

et al., 2013). For example, Mitsuzawa et al. (2009) modified a thermostable group II chaperonin (18-subunit self-assembling protein complex called rosettasome), from the archaeon *Sulfolobus shibatae*, for use as a scaffold to assemble selected hydrolytic enzymes. A cohesin module was fused to each of the eighteen subunits which in turn was combined with dockerin-containing cellulases from *C. thermocellum* to build an 18 enzyme cellulosome-like structure they termed rosettazyme. Truncated cellulosomes also called minicellulosomes have been constructed in an effort to fully understand the relationship between cellulosome structure and enzymatic activity (Arai et al., 2007; Cha et al., 2007; Mingardon et al., 2007). Examples are the mini-CipA and mini-CipC1 cellulosomes from *Clostridium cellulovorans* and *Clostridium cellulolyticum*, respectively (Murashima et al., 2002; Perret et al., 2004). The engineered minicellulosome platforms can be extended to build complex designer cellulosomes. Designer cellulosomes are not described in the GO and a complementary term “synthetic multi-cellular complex” would be useful to capture designer multienzyme complexes such as the rosettazyme, mini-CipA and mini-CipC1 and their components.

Generally, yield is one of the focal points for improving alcohol production by means of synthetic biology technology (Peralta-Yahya et al., 2012; Gronenberg et al., 2013). For example, low butanol yield in the native *Clostridium* species is not economically competitive, and thus the *Clostridium* CoA-dependent pathway has been heterologously expressed in genetically tractable industrial microbes hosts such as *E. coli* and *S. cerevisiae* (Inui et al., 2008; Steen et al., 2008) for facile manipulation toward better yield and productivity. For example, *E. coli* strain bearing the CoA-dependent pathway was further modified by substituting the reversible, flavin-dependent butyryl-CoA dehydrogenase (Bcd) with an irreversible trans-enoyl-CoA reductase (Ter) for the reduction of crotonyl-CoA (Bond-Watts et al., 2011). The increased flux to butanol improved product yield (Bond-Watts et al., 2011). While Bcd is associated with “butanol biosynthetic process” in the GO, Ter is not, as it is part of a non-natural (synthetic) pathway. This limitation in the GO could potentially be addressed with complementary SYNGO terms, for example, Ter could be assigned “enoyl-CoA reductase involved in increased butanol biosynthesis.”

The natural ability of *E. coli* to produce fatty acids and creation of new biochemical reactions through synthetic biology have provided the means to divert fatty acid metabolism toward the production of fuels (Clomburg and Gonzalez, 2010; Wen et al., 2013). This approach is a more sustainable alternative than, for example, the production of biodiesel from plant oils. Direct microbial production of fatty acid esters (biodiesel) eliminates the need for a subsequent chemical transesterification step. A pathway which includes an acyl-CoA ligase and the broad specificity acyltransferase (WS/DGAT; AtfA), together with the enzymes that provide ethanol for esterification [pyruvate decarboxylase (pdc) and alcohol dehydrogenase B (adhB)] were incorporated into a fatty acid overproducing *E. coli* strain for the production of fatty acid ethyl esters (Stoveken et al., 2005; Steen et al., 2010). Curating the components of the engineered pathway with complementary SYNGO terms,

associating them with fatty acid ethyl ester biosynthesis, should be a powerful tool for researchers working to improve these processes.

CONCLUSION

The MENGO group has created and continues to create GO terms relevant to microbial bioenergy research. In addition, the team has used the terms to produce high quality “gold standard” annotations of gene products based on the experimental scientific literature, which are useful for engineering of microbial strains to optimize bioenergy production. These resources provide easy access to otherwise dispersed information in the scientific literature and also aids in the computational analysis of large datasets. The GO annotations can also be used to uncover metabolic pathways, and to understand bioenergy-relevant microbes at the system biology level. Community involvement in MENGO term development and annotations is welcome and will create a more comprehensive resource for use by all. Since the GO currently is restricted to natural processes, a need has arisen for a complementary set of GO-like terms to describe engineered processes and the gene products that comprise them. Community involvement in this other development is also essential.

ACKNOWLEDGMENTS

We thank the editors at the GOC for reviewing the MENGO terms. This work was supported by grant DE-SC0005011 from the US Department of Energy. Jane Lomax is funded by European Molecular Biology Laboratories (EMBL) core funds. João C. Setubal is funded by CNPq and FAPESP. The authors wish to thank all participants in the MENGO workshops held at various venues in 2011 and 2012.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00528/abstract>

REFERENCES

- Aarts, M. G., Keijzer, C. J., Stiekema, W. J., and Pereira, A. (1995). Molecular characterization of the CER1 gene of arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* 7, 2115–2127. doi: 10.1105/tpc.7.12.2115
- Ahmed, S., Riaz, S., and Jamil, A. (2009). Molecular cloning of fungal xylanases: an overview. *Appl. Microbiol. Biotechnol.* 84, 19–35. doi: 10.1007/s00253-009-2079-4
- Alsaker, K. V., Spitzer, T. R., and Papoutsakis, E. T. (2004). Transcriptional analysis of spo0A overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *J. Bacteriol.* 186, 1959–1971. doi: 10.1128/JB.186.7.1959-1971.2004
- Andersen, C. (2003). Channel-tunnels: outer membrane components of type I secretion systems and multidrug efflux pumps of Gram-negative bacteria. *Rev. Physiol. Biochem. Pharmacol.* 147, 122–165. doi: 10.1007/s10254-003-0008-y
- Arai, T., Matsuoka, S., Cho, H. Y., Yukawa, H., Inui, M., Wong, S. L., et al. (2007). Synthesis of *Clostridium cellulovorans* minicellulosomes by intercellular complementation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1456–1460. doi: 10.1073/pnas.0610740104
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. The Gene Ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Atsumi, S., Hanai, T., and Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451, 86–89. doi: 10.1038/nature06450
- Atsumi, S., and Liao, J. C. (2008). Metabolic engineering for advanced biofuels production from *Escherichia coli*. *Curr. Opin. Biotechnol.* 19, 414–419. doi: 10.1016/j.copbio.2008.08.008
- Bayer, E. A., and Lamed, R. (1986). Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J. Bacteriol.* 167, 828–836.
- Bayer, E. A., Setter, E., and Lamed, R. (1985). Organization and distribution of the cellulosome in *Clostridium thermocellum*. *J. Bacteriol.* 163, 552–559.
- Beck, C., Knoop, H., Axmann, I. M., and Steuer, R. (2012). The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. *BMC Genomics* 13:56. doi: 10.1186/1471-2164-13-56
- Beller, H. R., Goh, E. B., and Keasling, J. D. (2010). Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Appl. Environ. Microbiol.* 76, 1212–1223. doi: 10.1128/AEM.02312-09
- Benso, A., Di Carlo, S., Ur Rehman, H., Politano, G., Savino, A., and Suravajhala, P. (2013). A combined approach for genome wide protein function annotation/prediction. *Proteome Sci.* 11, S1. doi: 10.1186/1477-5956-11-S1-S1
- Beytia, E. D., and Porter, J. W. (1976). Biochemistry of polyisoprenoid biosynthesis. *Annu. Rev. Biochem.* 45, 113–142. doi: 10.1146/annurev.bi.45.070176.000553
- Bi, C., Jones, S. W., Hess, D. R., Tracy, B. P., and Papoutsakis, E. T. (2011). SpoIIE is necessary for asymmetric division, sporulation, and expression of sigmaF, sigmaE, and sigmaG but does not control solvent production in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 193, 5130–5137. doi: 10.1128/JB.05474-11
- Biely, P., Mislovicova, D., and Toman, R. (1985). Soluble chromogenic substrates for the assay of endo-1,4-beta-xylanases and endo-1,4-beta-glucanases. *Anal. Biochem.* 144, 142–146. doi: 10.1016/0003-2697(85)90095-8
- Blake, J. A., and Harris, M. A. (2008). The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. *Curr. Protoc. Bioinformatics* Chap. 7, Unit 7.2. doi: 10.1002/0471250953.bi0702s23
- Blanch, H. W., Simmons, B. A., and Klein-Marcuschamer, D. (2011). Biomass deconstruction to sugars. *Biotechnol. J.* 6, 1086–1102. doi: 10.1002/biot.201000180
- Blumer-Schuette, S. E., Brown, S. D., Sander, K. B., Bayer, E. A., Kataeva, I., Zurawski, J. V., et al. (2014). Thermophilic lignocellulose deconstruction. *FEMS Microbiol. Rev.* 38, 393–448. doi: 10.1111/1574-6976.12044
- Bohlmann, J., and Keeling, C. I. (2008). Terpenoid biomaterials. *Plant J.* 54, 656–669. doi: 10.1111/j.1365-313X.2008.03449.x
- Bond-Watts, B. B., Bellerose, R. J., and Chang, M. C. (2011). Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat. Chem. Biol.* 7, 222–227. doi: 10.1038/nchembio.537
- Branduardi, P., Longo, V., Berterame, N. M., Rossi, G., and Porro, D. (2013). A novel pathway to produce butanol and isobutanol in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 6, 68. doi: 10.1186/1754-6834-6-68
- Bras, J. L., Alves, V. D., Carvalho, A. L., Najmudin, S., Prates, J. A., Ferreira, L. M., et al. (2012). Novel *Clostridium thermocellum* type I cohesin-dockerin complexes reveal a single binding mode. *J. Biol. Chem.* 287, 44394–44405. doi: 10.1074/jbc.M112.407700
- Campbell, J. W., and Cronan, J. E. Jr. (2001). Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu. Rev. Microbiol.* 55, 305–332. doi: 10.1146/annurev.micro.55.1.305
- Cha, J., Matsuoka, S., Chan, H., Yukawa, H., Inui, M., and Doi, R. H. (2007). Effect of multiple copies of cohesins on cellulase and hemicellulase activities of *Clostridium cellulovorans* mini-cellulosomes. *J. Microbiol. Biotechnol.* 17, 1782–1788.
- Chang, J. J., Ho, F. J., Ho, C. Y., Wu, Y. C., Hou, Y. H., Huang, C. C., et al. (2013). Assembling a cellulase cocktail and a cellobextrin transporter into a yeast host for CBP ethanol production. *Biotechnol. Biofuels* 6:19. doi: 10.1186/1754-6834-6-19
- Cheesbrough, T. M., and Kolattukudy, P. E. (1984). Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from *Pisum sativum*. *Proc. Natl. Acad. Sci. U.S.A.* 81, 6613–6617. doi: 10.1073/pnas.81.2.6613
- Chen, J.-S., and Hiu, S. F. (1986). Acetone-butanol-isopropanol production by *Clostridium beijerinckii* (Synonym. *Clostridium butylicum*). *Biotechnol. Lett.* 8, 371–376. doi: 10.1007/BF01040869
- Christie, K. R., Hong, E. L., and Cherry, J. M. (2009). Functional annotations for the *Saccharomyces cerevisiae* genome: the knowns and the

- known unknowns. *Trends Microbiol.* 17, 286–294. doi: 10.1016/j.tim.2009.04.005
- Chundawat, S. P., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* 2, 121–145. doi: 10.1146/annurev-chembioeng-061010-114205
- Clomburg, J. M., and Gonzalez, R. (2010). Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol.* 86, 419–434. doi: 10.1007/s00253-010-2446-1
- Coates, R. C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D. H., et al. (2014). Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. *PLoS ONE* 9:e85140. doi: 10.1371/journal.pone.0085140
- Dennis, M., and Kolattukudy, P. E. (1992). A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5306–5310. doi: 10.1073/pnas.89.12.5306
- Dennis, M. W., and Kolattukudy, P. E. (1991). Alkane biosynthesis by decarbonylation of aldehyde catalyzed by a microsomal preparation from *Botryococcus braunii*. *Arch. Biochem. Biophys.* 287, 268–275. doi: 10.1016/0003-9861(91)90478-2
- Dickinson, J. R., Harrison, S. J., and Hewlins, M. J. (1998). An investigation of the metabolism of valine to isobutyl alcohol in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273, 25751–25756. doi: 10.1074/jbc.273.40.25751
- Dimmer, E. C., Huntley, R. P., Alam-Faruque, Y., Sawford, T., O'donovan, C., Martin, M. J., et al. (2012). The UniProt-GO annotation database in 2011. *Nucleic Acids Res.* 40, D565–D570. doi: 10.1093/nar/gkr1048
- Dodd, D., and Cann, I. K. (2009). Enzymatic deconstruction of xylan for biofuel production. *Glob. Change Biol. Bioenergy* 1, 2–17. doi: 10.1111/j.1757-1707.2009.01004.x
- Doi, R. H. (2008). Cellulases of mesophilic microorganisms: cellulosome and noncellulosome producers. *Ann. N. Y. Acad. Sci.* 1125, 267–279. doi: 10.1196/annals.1419.002
- Du, D., Wang, Z., James, N. R., Voss, J. E., Klimont, E., Ohene-Agyei, T., et al. (2014). Structure of the AcrAB-TolC multidrug efflux pump. *Nature* 509, 512–515. doi: 10.1038/nature13205
- Dunlop, M. J. (2011). Engineering microbes for tolerance to next-generation biofuels. *Biotechnol. Biofuels* 4:32. doi: 10.1186/1754-6834-4-32
- Dürre, P. (2007). Biobutanol: an attractive biofuel. *Biotechnol. J.* 2, 1525–1534. doi: 10.1002/biot.200700168
- Dürre, P., and Hollerschwandner, C. (2004). Initiation of endospore formation in *Clostridium acetobutylicum*. *Anaerobe* 10, 69–74. doi: 10.1016/j.anaerobe.2003.11.001
- Edwards, P. A., Ashby, M. N., Spear, D. H., Marrero, P. F., Joly, A., and Popjak, G. (1992). Polyisoprenoid synthesis and metabolism. *Biochem. Soc. Trans.* 20, 475–479.
- Feng, Y., and Cronan, J. E. (2009). *Escherichia coli* unsaturated fatty acid synthesis: complex transcription of the fabA gene and in vivo identification of the essential reaction catalyzed by FabB. *J. Biol. Chem.* 284, 29526–29535. doi: 10.1074/jbc.M109.023440
- Fujita, K., Matsuyama, A., Kobayashi, Y., and Iwashashi, H. (2006). The genome-wide screening of yeast deletion mutants to identify the genes required for tolerance to ethanol and other alcohols. *FEMS Yeast Res.* 6, 744–750. doi: 10.1111/j.1567-1364.2006.00040.x
- Gao, X., Chen, J. C., Wu, Q., and Chen, G. Q. (2011). Polyhydroxyalkanoates as a source of chemicals, polymers, and biofuels. *Curr. Opin. Biotechnol.* 22, 768–774. doi: 10.1016/j.copbio.2011.06.005
- George, H. A., Johnson, J. L., Moore, W. E., Holdeman, L. V., and Chen, J. S. (1983). Acetone, Isopropanol, and Butanol production by *Clostridium beijerinckii* (syn. *Clostridium butylicum*) and *Clostridium aurantibutyricum*. *Appl. Environ. Microbiol.* 45, 1160–1163.
- Ghiaci, P., Norbeck, J., and Larsson, C. (2013). Physiological adaptations of *Saccharomyces cerevisiae* evolved for improved butanol tolerance. *Biotechnol. Biofuels* 6:101. doi: 10.1186/1754-6834-6-101
- Gomez del Pulgar, E. M., and Saadeddin, A. (2014). The cellulolytic system of *Thermobifida fusca*. *Crit. Rev. Microbiol.* 40, 236–247. doi: 10.3109/1040841X.2013.776512
- González-Ramos, D., Van Den Broek, M., Van Maris, A. J. A., Pronk, J. T., and Daran, J.-M. G. (2013). Genome-scale analyses of butanol tolerance in *Saccharomyces cerevisiae* reveal an essential role of protein degradation. *Biotechnol. Biofuels* 6:48. doi: 10.1186/1754-6834-6-48
- Gronenberg, L. S., Marcheschi, R. J., and Liao, J. C. (2013). Next generation biofuel engineering in prokaryotes. *Curr. Opin. Chem. Biol.* 17, 462–471. doi: 10.1016/j.cbpa.2013.03.037
- Hansen, J., Kharecha, P., Sato, M., Masson-Delmotte, V., Ackerman, F., Beerling, D. J., et al. (2013). Assessing “dangerous climate change”: required reduction of carbon emissions to protect young people, future generations and nature. *PLoS ONE* 8:e81648. doi: 10.1371/journal.pone.0081648
- Harris, L. M., Welker, N. E., and Papoutsakis, E. T. (2002). Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 184, 3586–3597. doi: 10.1128/JB.184.13.3586-3597.2002
- Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261. doi: 10.1093/nar/gkh036
- Heath, R. J., and Rock, C. O. (1995). Enoyl-acyl carrier protein reductase (fabI) plays a determinant role in completing cycles of fatty acid elongation in *Escherichia coli*. *J. Biol. Chem.* 270, 26538–26542. doi: 10.1074/jbc.270.44.26538
- Heipieper, H. J., Neumann, G., Cornelissen, S., and Meinhardt, F. (2007). Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems. *Appl. Microbiol. Biotechnol.* 74, 961–973. doi: 10.1007/s00253-006-0833-4
- Hill, D. P., Adams, N., Bada, M., Batchelor, C., Berardini, T. Z., Dietze, H., et al. (2013). Dovetailing biology and chemistry: integrating the Gene Ontology with the ChEBI chemical ontology. *BMC Genomics* 14:513. doi: 10.1186/1471-2164-14-513
- Hong, W., Zhang, J., Feng, Y., Mohr, G., Lambowitz, A. M., Cui, G. Z., et al. (2014). The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. *Biotechnol. Biofuels* 7, 80. doi: 10.1186/1754-6834-7-80
- Hu, S., Zheng, H., Gu, Y., Zhao, J., Zhang, W., Yang, Y., et al. (2011). Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 2018. *BMC Genomics* 12:93. doi: 10.1186/1471-2164-12-93
- Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., et al. (2008). Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 77, 1305–1316. doi: 10.1007/s00253-007-1257-5
- Janssen, H. J., and Steinbuchel, A. (2014). Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnol. Biofuels* 7, 7. doi: 10.1186/1754-6834-7-7
- Jarchow-Choy, S. K., Koppisch, A. T., and Fox, D. T. (2014). Synthetic routes to methylerythritol phosphate pathway intermediates and downstream isoprenoids. *Curr. Org. Chem.* 18, 1050–1072. doi: 10.2174/1385272819666140501001101
- Jendrossek, D., and Pfeiffer, D. (2013). New insights in the formation of polyhydroxyalkanoate (PHA) granules (carbonosomes) and novel functions of poly(3-hydroxybutyrate; PHB). *Environ. Microbiol.* 16, 2357–2373. doi: 10.1111/1462-2920.12356
- Jiang, Z., Dang, W., Yan, Q., Zhai, Q., Li, L., and Kusakabe, I. (2006). Subunit composition of a large xylanolytic complex (xylanosome) from *Streptomyces olivaceoviridis* E-86. *J. Biotechnol.* 126, 304–312. doi: 10.1016/j.jbiotec.2006.05.006
- Jiang, Z. Q., Deng, W., Li, L. T., Ding, C. H., Kusakabe, I., and Tan, S. S. (2004). A novel, ultra-large xylanolytic complex (xylanosome) secreted by *Streptomyces olivaceoviridis*. *Biotechnol. Lett.* 26, 431–436. doi: 10.1023/B:BILE.0000018266.75248.03
- Jonsson, L. J., Alriksson, B., and Nilvebrant, N. O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol. Biofuels* 6:16. doi: 10.1186/1754-6834-6-16
- Juturu, V., and Wu, J. C. (2013). Insight into microbial hemicellulases other than xylanases: a review. *J. Chem. Technol. Biotechnol.* 88, 353–363. doi: 10.1002/jctb.3969
- Kalscheuer, R., Stolling, T., and Steinbuchel, A. (2006). Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology* 152, 2529–2536. doi: 10.1099/mic.0.29028-0
- Kanno, M., Katayama, T., Tamaki, H., Mitani, Y., Meng, X. Y., Hori, T., et al. (2013). Isolation of butanol- and isobutanol-tolerant bacteria and physiological characterization of their butanol tolerance. *Appl. Environ. Microbiol.* 79, 6998–7005. doi: 10.1128/AEM.02900-13

- Karpol, A., Jobby, M. K., Slutzki, M., Noach, I., Chitayat, S., Smith, S. P., et al. (2013). Structural and functional characterization of a novel type-III dockerin from *Ruminococcus flavefaciens*. *FEBS Lett.* 587, 30–36. doi: 10.1016/j.febslet.2012.11.012
- Kieboom, J., Dennis, J. J., De Bont, J. A., and Zylstra, G. J. (1998a). Identification and molecular characterization of an efflux pump involved in *Pseudomonas putida* S12 solvent tolerance. *J. Biol. Chem.* 273, 85–91. doi: 10.1074/jbc.273.1.85
- Kieboom, J., Dennis, J. J., Zylstra, G. J., and De Bont, J. A. (1998b). Active efflux of organic solvents by *Pseudomonas putida* S12 is induced by solvents. *J. Bacteriol.* 180, 6769–6772.
- Klinke, H. B., Thomsen, A. B., and Ahring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pretreatment of biomass. *Appl. Microbiol. Biotechnol.* 66, 10–26. doi: 10.1007/s00253-004-1642-2
- Köne, A. C., and Büke, T. (2010). Forecasting of CO₂ emissions from fuel combustion using trend analysis. *Renew. Sust. Energ. Rev.* 14, 2906–2915. doi: 10.1016/j.rser.2010.06.006
- Laadan, B., Almeida, J. R., Radstrom, P., Hahn-Hagerdal, B., and Gorwa-Grauslund, M. (2008). Identification of an NADH-dependent 5-hydroxymethylfurfural-reducing alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Yeast* 25, 191–198. doi: 10.1002/yea.1578
- Lai, C. Y., and Cronan, J. E. (2003). Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. *J. Biol. Chem.* 278, 51494–51503. doi: 10.1074/jbc.M308638200
- Laycock, B., Halley, P., Pratt, S., Werker, A., and Lant, P. (2014). The chemomechanical properties of microbial polyhydroxyalkanoates. *Prog. Polym. Sci.* 39, 397–442. doi: 10.1016/j.progpolymsci.2013.06.008
- Lee, I., Johnson, L. A., and Hammond, E. G. (1995). Use of branched-chain esters to reduce the crystallization temperature of biodiesel. *J. Am. Oil Chem. Soc.* 72, 1155–1160. doi: 10.1007/BF02540982
- Lee, L. C., Lee, Y. L., Leu, R. J., and Shaw, J. F. (2006). Functional role of catalytic triad and oxyanion hole-forming residues on enzyme activity of *Escherichia coli* thioesterase I/protease I/phospholipase L1. *Biochem. J.* 397, 69–76. doi: 10.1042/BJ20051645
- Lee, S. M., Cho, M. K., Park, C. H., Chung, Y. C., Kim, J. H., Sang, B. I., et al. (2008a). Continuous butanol production using suspended and immobilized *Clostridium beijerinckii* NCIMB 8052 with supplementary butyrate. *Energ. Fuel.* 22, 3459–3464. doi: 10.1021/ef800076j
- Lee, S. Y., Park, J. H., Jang, S. H., Nielsen, L. K., Kim, J., and Jung, K. S. (2008b). Fermentative butanol production by *Clostridia*. *Biotechnol. Bioeng.* 101, 209–228. doi: 10.1002/bit.22003
- Lennen, R. M., and Pfleger, B. F. (2012). Engineering *Escherichia coli* to synthesize free fatty acids. *Trends Biotechnol.* 30, 659–667. doi: 10.1016/j.tibtech.2012.09.006
- Lennen, R. M., and Pfleger, B. F. (2013). Microbial production of fatty acid-derived fuels and chemicals. *Curr. Opin. Biotechnol.* 24, 1044–1053. doi: 10.1016/j.copbio.2013.02.028
- Li, N., Chang, W. C., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger, J. M. Jr. (2012). Evidence for only oxygenative cleavage of aldehydes to alk(a)enes and formate by cyanobacterial aldehyde decarbonylases. *Biochemistry* 51, 7908–7916. doi: 10.1021/bi300912n
- Li, N., Norgaard, H., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger, J. M. Jr. (2011). Conversion of fatty aldehydes to alk(a)enes and formate by a cyanobacterial aldehyde decarbonylase: cryptic redox by an unusual dimetal oxygenase. *J. Am. Chem. Soc.* 133, 6158–6161. doi: 10.1021/ja2013517
- Li, S. J., and Cronan, J. E. Jr. (1992). The gene encoding the biotin carboxylase subunit of *Escherichia coli* acetyl-CoA carboxylase. *J. Biol. Chem.* 267, 855–863.
- Ling, H., Chen, B., Kang, A., Lee, J. M., and Chang, M. W. (2013). Transcriptome response to alkane biofuels in *Saccharomyces cerevisiae*: identification of efflux pumps involved in alkane tolerance. *Biotechnol. Biofuels* 6:95. doi: 10.1186/1754-6834-6-95
- Liu, S., and Qureshi, N. (2009a). How microbes tolerate ethanol and butanol. *N. Biotechnol.* 26, 117–121. doi: 10.1016/j.nbt.2009.06.984
- Liu, Z. L., and Moon, J. (2009b). A novel NADPH-dependent aldehyde reductase gene from *Saccharomyces cerevisiae* NRRL Y-12632 involved in the detoxification of aldehyde inhibitors derived from lignocellulosic biomass conversion. *Gene* 446, 1–10. doi: 10.1016/j.gene.2009.06.018
- Lu, J. N., Tappel, R. C., and Nomura, C. T. (2009). Mini-review: biosynthesis of poly(hydroxyalkanoates). *J. Macromol. Sci.* 49, 226–248.
- Marinković, M., De Leeuw, W. C., De Jong, M., Kraak, M. H., Admiraal, W., Breit, T. M., et al. (2012). Combining next-generation sequencing and microarray technology into a transcriptomics approach for the non-model organism *Chironomus riparius*. *PLoS ONE* 7:e48096. doi: 10.1371/journal.pone.0048096
- McCarthy, J. G., Eisman, E. B., Kulkarni, S., Gerwick, L., Gerwick, W. H., Wipf, P., et al. (2012). Structural basis of functional group activation by sulfotransferases in complex metabolic pathways. *ACS Chem. Biol.* 7, 1994–2003. doi: 10.1021/cb300385m
- McEvoy, E., Wright, P. C., and Bustard, M. T. (2004). The effect of high concentration isopropanol on the growth of a solvent-tolerant strain of *Chlorella vulgaris*. *Enzyme Microb. Tech.* 35, 140–146. doi: 10.1016/j.enzmictec.2004.01.012
- McNeil, J. B., McIntosh, E. M., Taylor, B. V., Zhang, F. R., Tang, S., and Bognar, A. L. (1994). Cloning and molecular characterization of three genes, including two genes encoding serine hydroxymethyltransferases, whose inactivation is required to render yeast auxotrophic for glycine. *J. Biol. Chem.* 269, 9155–9165.
- Mehta, R. N., Chakraborty, M., Mahanta, P., and Parikh, P. A. (2010). Evaluation of fuel properties of butanol-biodiesel-diesel blends and their impact on engine performance and emissions. *Ind. Eng. Chem. Res.* 49, 7660–7665. doi: 10.1021/ie1006257
- Mendez-Perez, D., Begemann, M. B., and Pfleger, B. F. (2011). Modular synthase-encoding gene involved in alpha-olefin biosynthesis in *Synechococcus* sp. strain PCC 7002. *Appl. Environ. Microbiol.* 77, 4264–4267. doi: 10.1128/AEM.00467-11
- Mingardon, F., Chanal, A., Lopez-Contreras, A. M., Dray, C., Bayer, E. A., and Fierobe, H. P. (2007). Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. *Appl. Environ. Microbiol.* 73, 3822–3832. doi: 10.1128/AEM.00398-07
- Mitsuzawa, S., Kagawa, H., Li, Y., Chan, S. L., Paavola, C. D., and Trent, J. D. (2009). The rosettazyme: a synthetic cellulosome. *J. Biotechnol.* 143, 139–144. doi: 10.1016/j.jbiotec.2009.06.019
- Miziorko, H. M. (2011). Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* 505, 131–143. doi: 10.1016/j.abb.2010.09.028
- Mohan, S., Kelly, T. M., Eveland, S. S., Raetz, C. R., and Anderson, M. S. (1994). An *Escherichia coli* gene (FabZ) encoding (3R)-hydroxymyristoyl acyl carrier protein dehydrase. Relation to fabA and suppression of mutations in lipid biosynthesis. *J. Biol. Chem.* 269, 32896–32903.
- Molnar, I., Lopez, D., Wisecaver, J. H., Devarenne, T. P., Weiss, T. L., Pellegrini, M., et al. (2012). Bio-crude transcriptomics: gene discovery and metabolic network reconstruction for the biosynthesis of the terpenome of the hydrocarbon oil-producing green alga, *Botryococcus braunii* race B (Showa). *BMC Genomics* 13:576. doi: 10.1186/1471-2164-13-576
- Morais, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., et al. (2010). Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. *MBio* 1, e00285-10. doi: 10.1128/mBio.00285-10
- Mosqueda, G., and Ramos, J. L. (2000). A set of genes encoding a second toluene efflux system in *Pseudomonas putida* DOT-T1E is linked to the tod genes for toluene metabolism. *J. Bacteriol.* 182, 937–943. doi: 10.1128/JB.182.4.937-943.2000
- Murashima, K., Chen, C. L., Kosugi, A., Tamaru, Y., Doi, R. H., and Wong, S. L. (2002). Heterologous production of *Clostridium cellulovorans* engB, using protease-deficient *Bacillus subtilis*, and preparation of active recombinant cellulosomes. *J. Bacteriol.* 184, 76–81. doi: 10.1128/JB.184.1.76-81.2002
- Nawabi, P., Bauer, S., Kyrpides, N., and Lykidis, A. (2011). Engineering *Escherichia coli* for biodiesel production utilizing a bacterial fatty acid methyltransferase. *Appl. Environ. Microbiol.* 77, 8052–8061. doi: 10.1128/AEM.05046-11
- Nordon, R. E., Craig, S. J., and Foong, F. C. (2009). Molecular engineering of the cellulosome complex for affinity and bioenergy applications. *Biotechnol. Lett.* 31, 465–476. doi: 10.1007/s10529-008-9899-7
- Oefner, C., Schulz, H., D'arcy, A., and Dale, G. E. (2006). Mapping the active site of *Escherichia coli* malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 62, 613–618. doi: 10.1107/S0907444906009474
- Park, S. J., Kim, T. W., Kim, M. K., Lee, S. Y., and Lim, S. C. (2012). Advanced bacterial polyhydroxyalkanoates: towards a versatile and sustainable platform for unnatural tailor-made polyesters. *Biotechnol. Adv.* 30, 1196–1206. doi: 10.1016/j.biotechadv.2011.11.007

- Peralta-Yahya, P. P., and Keasling, J. D. (2010). Advanced biofuel production in microbes. *Biotechnol. J.* 5, 147–162. doi: 10.1002/biot.200900220
- Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., and Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature* 488, 320–328. doi: 10.1038/nature11478
- Perret, S., Belaich, A., Fierobe, H. P., Belaich, J. P., and Tardif, C. (2004). Towards designer cellulosomes in *Clostridia*: mannanase enrichment of the cellulosomes produced by *Clostridium cellulolyticum*. *J. Bacteriol.* 186, 6544–6552. doi: 10.1128/JB.186.19.6544-6552.2004
- Petersson, A., Almeida, J. R., Modig, T., Karhumaa, K., Hahn-Hagerdal, B., Gorwa-Grauslund, M. F., et al. (2006). A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* ADH6 gene conveys HMF tolerance. *Yeast* 23, 455–464. doi: 10.1002/yea.1370
- Pfeiffer, D., and Jendrossek, D. (2012). Localization of poly(3-hydroxybutyrate) (PHB) granule-associated proteins during PHB granule formation and identification of two new phasins, PhaP6 and PhaP7, in *Ralstonia eutropha* H16. *J. Bacteriol.* 194, 5909–5921. doi: 10.1128/JB.00779-12
- Plewniak, F., Koehler, S., Navet, B., Dugat-Bony, E., Bouchez, O., Peyret, P., et al. (2013). Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments. *Mol. Ecol.* 22, 4870–4883. doi: 10.1111/mec.12432
- Qureshi, N., Sahal, B. C., Hector, R. E., and Cotta, M. A. (2010). Production of butanol (a biofuel) from agricultural residues: Part 1-Use of barley straw hydrolysate. *Biomass Bioenerg.* 34, 559–565. doi: 10.1016/j.biombioe.2009.12.024
- Ramos, J. L., Duque, E., Godoy, P., and Segura, A. (1998). Efflux pumps involved in toluene tolerance in *Pseudomonas putida* DOT-T1E. *J. Bacteriol.* 180, 3323–3329.
- Rashid, U., Anwar, F., Moser, B. R., and Knothe, G. (2008). Moringa oleifera oil: a possible source of biodiesel. *Bioresour. Technol.* 99, 8175–8179. doi: 10.1016/j.biortech.2008.03.066
- Ravagnani, A., Jennert, K. C., Steiner, E., Grunberg, R., Jefferies, J. R., Wilkinson, S. R., et al. (2000). Spo0A directly controls the switch from acid to solvent production in solvent-forming *Clostridia*. *Mol. Microbiol.* 37, 1172–1185. doi: 10.1046/j.1365-2958.2000.02071.x
- Reddy, C. S., Ghai, R., Rashmi, and Kalia, V. C. (2003). Polyhydroxyalkanoates: an overview. *Bioresour. Technol.* 87, 137–146. doi: 10.1016/S0960-8524(02)00212-2
- Riedel, S. L., Lu, J., Stahl, U., and Brigham, C. J. (2014). Lipid and fatty acid metabolism in *Ralstonia eutropha*: relevance for the biotechnological production of value-added products. *Appl. Microbiol. Biotechnol.* 98, 1469–1483. doi: 10.1007/s00253-013-5430-8
- Rincon, M. T., Cepeljnik, T., Martin, J. C., Lamed, R., Barak, Y., Bayer, E. A., et al. (2005). Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface. *J. Bacteriol.* 187, 7569–7578. doi: 10.1128/JB.187.22.7569-7578.2005
- Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J. L., et al. (2001). Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *J. Bacteriol.* 183, 3967–3973. doi: 10.1128/JB.183.13.3967-3973.2001
- Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B., and Schirmer, A. (2011). Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl. Environ. Microbiol.* 77, 1718–1727. doi: 10.1128/AEM.02580-10
- Rude, M. A., and Schirmer, A. (2009). New microbial fuels: a biotech perspective. *Curr. Opin. Microbiol.* 12, 274–281. doi: 10.1016/j.mib.2009.04.004
- Sato, T. (2013). Unique biosynthesis of sesquiterpenes (C₁₅ terpenes). *Biosci. Biotechnol. Biochem.* 77, 1155–1159. doi: 10.1271/bbb.130180
- Schirmer, A., Rude, M. A., Li, X., Popova, E., and Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science* 329, 559–562. doi: 10.1126/science.1187936
- Schmidt, A., Forne, I., and Imhof, A. (2014). Bioinformatic analysis of proteomics data. *BMC Syst. Biol.* 8 Suppl. 2:S3. doi: 10.1186/1752-0509-8-S2-S3
- Scotcher, M. C., and Bennett, G. N. (2005). SpoIIE regulates sporulation but does not directly affect solventogenesis in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 187, 1930–1936. doi: 10.1128/JB.187.6.1930-1936.2005
- Segura, A., Molina, L., Fillet, S., Krell, T., Bernal, P., Munoz-Rojas, J., et al. (2012). Solvent tolerance in Gram-negative bacteria. *Curr. Opin. Biotechnol.* 23, 415–421. doi: 10.1016/j.copbio.2011.11.015
- Shahid, S., Mosrati, R., Ledauphin, J., Amiel, C., Fontaine, P., Gaillard, J. L., et al. (2013). Impact of carbon source and variable nitrogen conditions on bacterial biosynthesis of polyhydroxyalkanoates: evidence of an atypical metabolism in *Bacillus megaterium* DSM 509. *J. Biosci. Bioeng.* 116, 302–308. doi: 10.1016/j.jbiosc.2013.02.017
- Smith, S. P., and Bayer, E. A. (2013). Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* 23, 686–694. doi: 10.1016/j.sbi.2013.09.002
- Srikrishnan, S., Chen, W., and Da Silva, N. A. (2013). Functional assembly and characterization of a modular xylanosome for hemicellulose hydrolysis in yeast. *Biotechnol. Bioeng.* 110, 275–285. doi: 10.1002/bit.24609
- Steen, E. J., Chan, R., Prasad, N., Myers, S., Petzold, C. J., Redding, A., et al. (2008). Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb. Cell Fact.* 7:36. doi: 10.1186/1475-2859-7-36
- Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., et al. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463, 559–562. doi: 10.1038/nature08721
- Stephenson, M. (1949). *Microbial Metabolism*. Cambridge: MIT Press.
- Stoveken, T., Kalscheuer, R., Malkus, U., Reichelt, R., and Steinbuchel, A. (2005). The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from *Acinetobacter* sp. strain ADP1: characterization of a novel type of acyltransferase. *J. Bacteriol.* 187, 1369–1376. doi: 10.1128/JB.187.4.1369-1376.2005
- Sukovich, D. J., Seffernick, J. L., Richman, J. E., Gralnick, J. A., and Wackett, L. P. (2010). Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Appl. Environ. Microbiol.* 76, 3850–3862. doi: 10.1128/AEM.00436-10
- Suranovic, S. (2013). Fossil fuel addiction and the implication for climate change policy. *Global Environ. Chang.* 23, 598–608. doi: 10.1016/j.gloenvcha.2013.02.006
- Torella, J. P., Ford, T. J., Kim, S. N., Chen, A. M., Way, J. C., and Silver, P. A. (2013). Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11290–11295. doi: 10.1073/pnas.1307129110
- Torto-Alalibo, T., Collmer, C. W., and Gwinn-Giglio, M. (2009). The Plant-Associated Microbe Gene Ontology (PAMGO) consortium: community development of new Gene Ontology terms describing biological processes involved in microbe-host interactions. *BMC Microbiol.* 9(Suppl. 1):S1. doi: 10.1186/1471-2180-9-S1-S1
- Vazana, Y., Barak, Y., Unger, T., Peleg, Y., Shamshoum, M., Ben-Yehzekel, T., et al. (2013). A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. *Biotechnol. Biofuels* 6:182. doi: 10.1186/1754-6834-6-182
- Villas-Bôas, S. G., Kesson, M., and Nielsen, J. (2005). Biosynthesis of glyoxylate from glycine in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 5, 703–709. doi: 10.1016/j.femsyr.2005.03.001
- Warui, D. M., Li, N., Norgaard, H., Krebs, C., Bollinger, J. M. Jr., and Booker, S. J. (2011). Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarboxylase. *J. Am. Chem. Soc.* 133, 3316–3319. doi: 10.1021/ja111607x
- Weber, C., Farwick, A., Benisch, F., Brat, D., Dietz, H., Subtil, T., et al. (2010). Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. *Appl. Microbiol. Biotechnol.* 87, 1303–1315. doi: 10.1007/s00253-010-2707-z
- Wen, M., Bond-Watts, B. B., and Chang, M. C. (2013). Production of advanced biofuels in engineered *E. coli*. *Curr. Opin. Chem. Biol.* 17, 472–479. doi: 10.1016/j.cbpa.2013.03.034
- Wilson, D. B. (2011). Microbial diversity of cellulose hydrolysis. *Curr. Opin. Microbiol.* 14, 259–263. doi: 10.1016/j.mib.2011.04.004
- Wilson, D. B. (2012). Processive and nonprocessive cellulases for biofuel production—lessons from bacterial genomes and structural analysis. *Appl. Microbiol. Biotechnol.* 93, 497–502. doi: 10.1007/s00253-011-3701-9
- Xue, C., Zhao, X. Q., Liu, C. G., Chen, L. J., and Bai, F. W. (2013). Prospective and development of butanol as an advanced biofuel. *Biotechnol. Adv.* 31, 1575–1584. doi: 10.1016/j.biotechadv.2013.08.004
- Yikmis, M., and Steinbuchel, A. (2012). Historical and recent achievements in the field of microbial degradation of natural and synthetic rubber. *Appl. Environ. Microbiol.* 78, 4543–4551. doi: 10.1128/AEM.00001-12
- Zhang, F., Rodriguez, S., and Keasling, J. D. (2011). Metabolic engineering of microbial pathways for advanced biofuels production. *Curr. Opin. Biotechnol.* 22, 775–783. doi: 10.1016/j.copbio.2011.04.024

- Zhang, X., Luo, R., Wang, Z., Deng, Y., and Chen, G. Q. (2009). Application of (R)-3-hydroxyalkanoate methyl esters derived from microbial polyhydroxyalkanoates as novel biofuels. *Biomacromolecules* 10, 707–711. doi: 10.1021/bm801424e
- Zhao, L., Chang, W. C., Xiao, Y., Liu, H. W., and Liu, P. (2013). Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annu. Rev. Biochem.* 82, 497–530. doi: 10.1146/annurev-biochem-052010-100934
- Zhou, Y. J., and Zhao, Z. K. (2011). Advanced biofuel-oriented engineering of fatty acid pathway: a review. *Sheng Wu Gong Cheng Xue Bao* 27, 1261–1267.
- Zhu, H., Ren, X., Wang, J., Song, Z., Shi, M., Qiao, J., et al. (2013). Integrated OMICS guided engineering of biofuel butanol-tolerance in photosynthetic *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels* 6:106. doi: 10.1186/1754-6834-6-106
- Zingaro, K. A., and Terry Papoutsakis, E. (2013). GroESL overexpression imparts *Escherichia coli* tolerance to i-, n-, and 2-butanol, 1,2,4-butanetriol and ethanol with complex and unpredictable patterns. *Metab. Eng.* 15, 196–205. doi: 10.1016/j.ymben.2012.07.009
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 28 July 2014; accepted: 22 September 2014; published online: 10 October 2014.
- Citation: Torto-Alalibo T, Purwantini E, Lomax J, Setubal JC, Mukhopadhyay B and Tyler BM (2014) Genetic resources for advanced biofuel production described with the Gene Ontology. *Front. Microbiol.* 5:528. doi: 10.3389/fmicb.2014.00528
- This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.
- Copyright © 2014 Torto-Alalibo, Purwantini, Lomax, Setubal, Mukhopadhyay and Tyler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.