

LIPASES FOR USE IN PRODUCTION OF A FATTY ACID ESTER

The present application claims priority to Singapore patent application number 10202107587W titled "Lipases" filed on 9 July 2021 which is incorporated by reference
5 herein in its entirety.

Technical Field

The present invention relates to lipases and their uses.

10 Background

Lipases are a group of enzymes capable of catalyzing the hydrolysis of triacylglycerides into free fatty acids during lipid metabolism. In the presence of short-chain alcohols, some of these enzymes can catalyze the transesterification of plant oils into biodiesel. Biodiesel has recently gained traction to be a green alternative to fossil
15 fuel. Efficient production of biodiesel at an industrial scale is limited by the reusability of the enzyme and its tolerance towards high concentrations of organic solvents.

Biodiesel is a form of biofuel derived from organic materials that consists of long-chain fatty acid alkyl esters (FAAEs), for example fatty acid methyl esters (FAMEs). In recent years, biodiesel has gained traction to be used as a green alternative to fossil fuels
20 due to its renewability and the minimal impact it has to the environment. It has been estimated that the production of clean, renewable energy through green manufacturing will soon become essential in most countries by 2050, as the worldwide need for energy continues to increase. This will involve the redesigning and restructuring of our current manufacturing processes, by incorporating automation and other technologies that are
25 environmentally friendly and more energy efficient.

The majority of the current world's supply of biodiesel is derived from a chemical process that utilizes an alkali to catalyze the transesterification of vegetable oil and alcohol. However, this process produces a large amount of alkali-rich wastewater, which creates additional burden to the environment. Enzyme-mediated FAAE production with
30 lipases provides a good alternative to chemical methods due to the ability of these enzymes to tolerate a wide variety of feedstocks that are usually not compatible with alkali

hydrolysis (e.g. non-edible and waste oil). However, the main drawback associated with the use of lipases for biodiesel production is the low tolerance of these enzymes towards organic solvents. Although high-throughput screening and protein engineering have been used extensively to identify lipases with improved thermostability and methanol tolerance, the cost of enzyme-mediated production of biodiesel remains high.

To date, most of the high-throughput screening of lipases are carried out by functional screens of metagenomic libraries to identify enzymes with properties suitable for industrial applications. However, such libraries are often tedious to construct and have a strong bias towards the codon usage of the native host. This can lead to the isolation of disparate clones when different hosts are used for the screen.

The listing or discussion of an apparently prior-published document in this specification should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

Any document referred to herein is hereby incorporated by reference in its entirety.

Summary of Invention

In a first aspect of the invention, there is provided a method of producing a fatty acid ester, the method comprising: (a) providing a mixture comprising a triacylglyceride and/or a fatty acid, a short-chained alcohol, and an enzyme comprising a sequence selected from any one of SEQ ID Nos: (Sequence Identification Numbers) 1 to 6 and 8 to 114; and (b) incubating the mixture for a time and under conditions sufficient to allow for formation of the fatty acid ester. Preferably, the enzyme consists essentially or consist of the sequence.

By "alcohol", it includes a hydrocarbon compound containing one or more hydroxy groups, and includes alcohols such as methanol, ethanol, propanol, butanol, isopropanol, isobutanol, t-butanol, pentanol, cyclopentanol, isopentanol, neopentanol, hexanol, isohexanol, cyclohexanol, 3-methylpentanol, 2,2-dimethylbutanol, 2,3-dimethylbutanol and isomers thereof. In various embodiments, the alcohol is selected from the group consisting of methanol, ethanol, n-propanol, and i-propanol. Preferably, the alcohol is methanol or ethanol, more preferably methanol.

In various embodiments, the alcohol is present from 10% v/v to 70% v/v. In various embodiments, the alcohol may be present in at least 20% v/v, preferably at least 30% v/v, more preferably at least 40% v/v, even more preferably at least 50% v/v. In an embodiment, the alcohol may be present at about 40% v/v.

5 In various embodiments, the enzyme is selected from the group consisting of: SEQ ID NO: 90, SEQ ID NO: 5, SEQ ID NO: 22, SEQ ID NO: 27, SEQ ID NO: 35, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ
10 ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 100, SEQ ID NO: 107, SEQ ID NO: 108, and SEQ ID NO: 109. Preferably, the lipase comprises SEQ ID NO: 90.

In various embodiments, the triacylglyceride is present in the mixture and the triacylglyceride is a long chain triacylglyceride. Preferably, the triacylglyceride is triolein. Preferably, the enzyme is selected from the group consisting of: SEQ ID NO: 90, SEQ ID
15 NO: 53, SEQ ID NO: 27, SEQ ID NO: 55, SEQ ID NO: 60, SEQ ID NO: 73, and SEQ ID NO: 107.

In various embodiments, the fatty acid is present in the mixture and the fatty acid is a long chain fatty acid. The fatty acid may be oleic acid. Preferably, the enzyme is selected from the group consisting of: SEQ ID NO: 90, SEQ ID NO: 57, SEQ ID NO: 74, SEQ ID
20 NO: 92, SEQ ID NO: 93, and SEQ ID NO: 108.

In various embodiments, wherein the triacylglyceride and/or fatty acid is palm oil fatty acid distillate.

Preferably, a yield of the fatty acid ester is at least 80%. Preferably the yield of the fatty acid ester is at least 85%, more preferably at least 90%, and even more preferably
25 at least 95%.

In a second aspect of the invention, there is provided a method of hydrolyzing an ester, the method comprising providing a mixture comprising a triacylglyceride, water and an enzyme comprising a sequence selected from the group consisting of: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID
30 NO: 23, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 39,

SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 88, SEQ ID NO: 91, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, and SEQ ID NO: 113.

10 In a third aspect of the invention, there is provided a product obtained by the method in the first and second aspects.

In a fourth aspect of the invention, there is provided an isolated lipase comprising a sequence selected from any one of SEQ ID Nos. 1 to 114. Preferably, the isolated lipase comprising essentially, consists essentially or consists of the sequence. Preferably, the lipase is purified. Preferably, the sequence is selected from the group consisting of SEQ ID NO: 90, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110 and SEQ ID NO: 113.

Preferably, the sequence is selected from the group consisting of SEQ ID NO: 30, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 46, SEQ ID NO: 85, and SEQ ID NO: 113. Advantageously, these sequences are active in acidic medium.

Preferably, the sequence is selected from the group consisting of SEQ ID NO: 54, SEQ ID NO: 98, SEQ ID NO: 107, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 59, SEQ ID NO: 68, SEQ ID NO: 73, SEQ ID NO: 83, SEQ ID NO: 86, SEQ ID NO: 97, and SEQ ID NO: 101. Advantageously, these sequences are active in alkaline
5 medium.

Preferably, the sequence is selected from the group consisting of SEQ ID NO: 96, SEQ ID NO: 70, SEQ ID NO: 5, SEQ ID NO: 35, SEQ ID NO: 22, SEQ ID NO: 30, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 53, SEQ ID NO: 57, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, SEQ ID NO: 105, and SEQ ID NO: 108. Advantageously, these sequences have a melting
10 temperature of at least 60°C.

Preferably, the lipase comprises SEQ ID NO: 90.

In a fifth aspect, the invention provides for a polynucleotide encoding the lipase according to the fourth aspect of the invention. It follows too that the invention provides
15 for a vector comprising the polynucleotide, and a host cell comprising said polynucleotide or vector.

Other technical advantages may become readily apparent to one of ordinary skill in the art after review of the following figures and description.

20 Description of Figures (FIGs)

FIG. 1a shows a schematic design of identifying enzymes for biodiesel production.

FIG. 1b shows a schematic scalable workflow for the identification, screening and characterization of enzymes using an automated pipeline (HT: high-throughput, MT: medium-throughput).
25

FIG. 2 shows the sequence similarity network of selected T1.2-like lipases with an e-value of 10^{-130} . The cluster numbers are indicated by reference numbers and arrows in FIG. 2. Clusters 1, 2 and 3 show the *Staphylococcus* spp., *Bacillus* spp., and *Geobacillus* spp. respectively. Clusters 4 and 5 show the *Clostridium* spp. Clusters 6 to 13 show the
30 *Deinococcus* spp., *Massilia* spp., *Paenibacillus* spp., *Clostridium* spp., *Bacillus* spp., *Pelosinus* spp., *Firmicutes* spp., and *Clostridium* spp. respectively. Panel A shows the

selected and active targets from the phenol red assay. Panel B shows the expression yield of T1.2-like lipases. Panel C shows the specificity of T1.2-like lipases against short-chain and medium-chain substrates. Panel D shows the specificity of T1.2-like lipases against triesters and monoesters. Panel E shows the optimal pH of T1.2-like lipases. Panel F shows the melting temperature (T_m) of T1.2-like lipases. Panel G shows the methanol-tolerance of T1.2-like lipases. Panel H shows the biodiesel production from T1.2-like lipases.

FIG. 3 shows the results of a high-throughput assay to detect lipase activity from cell lysate using phenol red indicator. In FIG. 3 “c” refers to a negative control, “+” refers to the T1.2 lipase, and “*” refers to detectable lipase activity. The following shows the lipase contained in the wells of each plate in FIG. 3 and is in the following format – (plate number).(well number range) (lipase ortholog). As each lipase is tested in triplicate, each well number represents a range of 3 well numbers, and each row of 24 wells is used for 1 or 2 enzymes with four different substrates (tributyrin, tricaprylin, triolein, vinyl laurate) i.e. B1 refers to B1-B3, B7-B9, B13-B15, and B19-B21; B4 refers to B4-B6, B10-B12, B16-B18, and B22-B24 for tributyrin, tricaprylin, triolein, vinyl laurate respectively and so forth for the other rows and plates. Hence, 1.B1 (8) means plate 1, well numbers B1-B3, B7-B9, B13-B15, and B19-B21 contain lipase ortholog 8 for tributyrin, tricaprylin, triolein, vinyl laurate respectively, 1.B4 (24) means plate 1, well numbers B4-B6, B10-B12, B16-B18, and B22-B24 contain lipase ortholog 24 for tributyrin, tricaprylin, triolein, vinyl laurate respectively for tributyrin, tricaprylin, triolein, vinyl laurate respectively for tributyrin, tricaprylin, triolein, vinyl laurate respectively, and so forth. Plate 1.B1 (8) 1.B4 (24) 1.C1 (1) 1.C4 (16) 1.D1 (9) 1.D4 (25) 1.E1 (2) 1.E4 (17) 1.F1 (10) 1.F4 (26) 1.G1 (3) 1.G4 (18) 1.H1 (11) 1.H4 (27) 1.I1 (4) 1.I4 (19) 1.J1 (23) 1.J4 (28) 1.K1 (5) 1.K4 (20) 1.L1 (13) 1.L4 (29) 1.M1 (6) 1.M4 (21) 1.N1 (14) 1.N4 (30) 1.O1 (12) 1.O4 (22) 1.P1 (15) 1.P4 (31) 2.B1 (40) 2.B4 (56) 2.C1 (32) 2.C4 (48) 2.D1 (41) 2.D4 (57) 2.E1 (33) 2.E4 (49) 2.F1 (42) 2.F4 (58) 2.G1 (34) 2.G4 (50) 2.H1 (43) 2.H4 (59) 2.I1 (35) 2.I4 (51) 2.J1 (44) 2.J4 (60) 2.K1 (36) 2.K4 (52) 2.L1 (45) 2.L4 (61) 2.M1 (37) 2.M4 (53) 2.N1 (46) 2.N4 (62) 2.O1 (38) 2.O4 (54) 2.P1 (47) 2.P4 (63) 3.B1 (72) 3.B4 (88) 3.C1 (64) 3.C4 (80) 3.D1 (73) 3.D4 (89) 3.E1 (65) 3.E4 (81) 3.F1 (74) 3.F4 (90) 3.G1 (66) 3.G4 (82) 3.H1 (75) 3.H4 (91) 3.I1 (67) 3.I4 (83) 3.J1 (76) 3.J4 (92) 3.K1 (68) 3.K4 (84) 3.L1 (77) 3.L4 (93) 3.M1 (69) 3.M4 (85) 3.N1

(78) 3.N4 (94) 3.O1 (70) 3.O4 (86) 3.P1 (79) 3.P4 (95) 4.B1 (104) 4.B4 (112) 4.C1 (96) 4.C4 (39) 4.D1 (105) 4.D4 (113) 4.E1 (97) 4.E4 (55) 4.F1 (106) 4.F4 (114) 4.G1 (98) 4.G4 (71) 4.H1 (107) 4.H4 (103) 4.I1 (99) 4.I4 (87) 4.J1 (108) 4.K1 (100) 4.L1 (109) 4.M1 (101) 4.N1 (110) 4.O1 (102) 4.P1 (111).

5 FIG. 4 shows the SDS-PAGE results of the purified lipases with the expected size of each protein marked with a white asterisk. In FIGs. 4 to 9, the number at the top of each panel refers to the SEQ ID number.

FIG. 5 shows the results of the TLC assay to detect the transesterification of triolein and methanol to produce FAME. A white asterisk indicates sequences with higher
10 conversion to FAME. In FIGs. 5, 6 and 7, "S" refers to standards, "B" refers to blank, "E" refers to Eversa®, "1" refers to methyl oleate on the TLC plate, "2" refers to triolein on the TLC plate, "3" refers to oleic acid on the TLC plate, "4" refers to 1,2-diolein and 1,3-diolein on the TLC plate, and "5" refers to 1-monoolein on the TLC plate. The reaction mixture contained 8 μ L triolein, 1 μ L methanol and 1 μ L enzyme, and was carried out at 37 °C for
15 24 hours.

FIG. 6 shows the results of the TLC assay to detect the esterification of oleic acid and methanol to produce FAME. A white asterisk indicates sequences with higher conversion to FAME. The reaction mixture contained 8 μ L oleic acid, 1 μ L methanol and 1 μ L enzyme, and was carried out at 37 °C for 24 hours.

20 FIG. 7 shows the results of the TLC assay to detect the transesterification of triolein (Panels A to D) and esterification of oleic acid (Panel E) with methanol to produce FAME. The reaction contained 8 μ L triolein/oleic acid, 3 μ L methanol and 3 μ L enzyme, and was carried out at 37 °C for 72 h.

FIG. 8 shows the results of the TLC assay to detect the transesterification of triolein
25 with methanol to produce FAME at different temperatures. In FIG. 8, "S" refers to Standards, "B" refers to Triolein + buffer, "Ev" refers to Triolein + Eversa, "a" refers to methyl oleate on the TLC plate, "b" refers to triolein on the TLC plate, "c" refers to oleic acid on the TLC plate, "d" refers to 1,2-diolein and 1,3-diolein on the TLC plate, and "e" refers to 1-monoolein on the TLC plate.

30 FIG. 9 shows the results of the TLC assay to detect the transesterification of triolein and methanol to produce FAME at 37 °C under different methanol concentrations, after

24 h, 48 h, and 72h. SEQ ID NO: 90 is still able to produce FAME in the presence of 70% methanol. In FIG. 9, "S" refers to standards, "1" refers to methyl oleate on the TLC plate, "2" refers to triolein on the TLC plate, "3" refers to oleic acid on the TLC plate, "4" refers to 1,2-diolein and 1,3-diolein on the TLC plate, and "5" refers to 1-monoolein on the TLC plate.

FIG. 10 shows a heat map of the expression level, substrate specificity, pH range, melting temperature, methanol tolerance and biodiesel production from the T1.2-like lipases.

Some of the figures contain one or more panels, and the reference of the figure number and alphabet refers to the panel in the figure.

Detailed Description of Embodiments of the Invention

With the rapid decline in the cost of DNA synthesis, it is now feasible to tap into the richness of the uncharacterized/unknown enzymes available in protein databases, to identify enzymes with potential industrial applications. The Enzyme Function Initiative (EFI) was recently established to facilitate the assignment of function(s) to unknown members of the diverse enzyme superfamilies, such as the enolases and isoprenoid synthases. This platform provides users with various bioinformatics and computational tools to analyze the sequence-function relationships of protein superfamilies.

The present invention identifies useful lipases that may be used to hydrolyze fatty acids. Such methods disclosed in the present invention are useful for the production of biodiesel. By "biodiesel", it is meant to include fatty acid alkyl esters suitable for use as a transportation and power generation fuel or additives, in particular fatty acid methyl esters.

Described herein is a scalable workflow that integrates web-based tools and automation to identify a group of 114 orthologous bacterial lipases for expression, purification and characterization using a high-throughput platform. The activity profile of these enzymes revealed many targets with different substrate specificities and optimal pH range. Most of these enzymes are thermostable and can tolerate up to 40% methanol (v/v). Among the 114 lipases that were targeted, 22 were found to be able to produce

methyl oleate from triolein in the presence of 50% methanol. This demonstrates the utility of the workflow to identify lipase candidates for the industrial production of biodiesel.

The lipases of the present invention may be synthesized in whole or part using techniques that are well-known in the art.

5 Methods which are well known to those skilled in the art can be used to construct expression vectors containing the polypeptide coding sequence and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. See, for example, the techniques described in Green, M. R. and Sambrook, P. (2012).
10 *Molecular cloning: A Laboratory Manual (Fourth Edition)*. Cold Spring Harbor Laboratory, New York. Cold Spring Harbor Laboratory Press; and Ausubel, F. M. et al. (1994) *Current Protocols in Molecular Biology*. John Wiley & Sons, Secaucus, N.J.

A variety of host-expression vector systems may be utilized to express the relevant lipase coding sequence. These include but are not limited to microorganisms such as
15 bacteria transformed with recombinant bacteriophage DNA, plasmid DNA or cosmid DNA expression vectors containing the coding sequence; yeast transformed with recombinant yeast expression vectors containing the domain coding sequence; insect cell systems infected with recombinant virus expression vectors (e.g., baculovirus) containing the coding sequence; plant cell systems infected with recombinant virus expression vectors
20 (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with recombinant plasmid expression vectors (e.g., Ti plasmid) containing the coding sequence; or animal cell systems. The expression elements of these systems vary in their strength and specificities.

Depending on the host/vector system utilized, any of a number of suitable
25 transcription and translation elements, including constitutive and inducible promoters, may be used in the expression vector. For example, when cloning in bacterial systems, inducible promoters such as pL of bacteriophage λ , plac, ptrp, ptac (ptrp-lac hybrid promoter) and the like may be used; when cloning in insect cell systems, promoters such as the baculovirus polyhedrin promoter may be used; when cloning in plant cell systems,
30 promoters derived from the genome of plant cells (e.g., heat shock promoters; the promoter for the small subunit of RUBISCO; the promoter for the chlorophyll a/b binding

protein) or from plant viruses (e.g., the ³⁵S RNA promoter of CaMV; the coat protein promoter of TMV) may be used; when cloning in mammalian cell systems, promoters derived from the genome of mammalian cells (e.g., metallothionein promoter) or from mammalian viruses (e.g., the adenovirus late promoter; the vaccinia virus 7.5K promoter) may be used; when generating cell lines that contain multiple copies of the kinase domain DNA, SV4O-, BPV- and EBV-based vectors may be used with an appropriate selectable marker.

Exemplary methods describing methods of DNA manipulation, vectors, various types of cells used, methods of incorporating the vectors into the cells, expression techniques, protein purification and isolation methods, and protein concentration methods are disclosed in detail in PCT publication WO 96/18738. This publication is incorporated herein by reference in its entirety, including any drawings. Those skilled in the art will appreciate that such descriptions are applicable to the present invention and can be easily adapted to it.

The use of a scalable workflow that integrates web-based tools and automation includes: (1) Web-based bioinformatics tools from the EFI were used to identify candidate enzymes from a diverse group of bacterial lipases for gene synthesis and characterization; (2) A high-throughput automated pipeline was used to clone, express and functionally characterize these enzymes to determine their substrate specificity, thermostability and tolerance towards methanol (FIG. 1b). In recent years, the availability of low cost and reliable liquid handlers (OT-2s from Opentrons) allows tasks that are repetitive and tedious to be replaced by automation. This improves the reproducibility and scalability of the research without significant increase in cost. From the screen, we were able to identify many targets that possess high transesterification and esterification activity against triolein and oleic acid, respectively. These enzymes have the potential to be developed as catalysts for the industrial production of biodiesel.

In order that the present invention may be fully understood and readily put into practical effect, there shall now be described by way of non-limitative examples only preferred embodiments of the present invention, the description being with reference to the accompanying illustrative figures.

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs.

The terms “about”, “approximately”, “substantially” must be read with reference to the context of the application as a whole, and have regard to the meaning a particular technical term qualified by such a word usually has in the field concerned. For example, it may be understood that a certain parameter, function, effect, or result can be performed or obtained within a certain tolerance, and the skilled person in the relevant technical field knows how to obtain the tolerance of such term.

The phrase “at least one of A and B” means it requires only A alone, B alone, or A and B, i.e. only one of A or B is required.

Where a combination is disclosed, each subcombination of the elements of that combination is also specifically disclosed and is within the scope of the invention. Conversely, where different elements or groups of elements are disclosed, combinations thereof are also disclosed. Where any element of an invention is disclosed as having a plurality of alternatives, examples of that invention in which each alternative is excluded singly or in any combination with the other alternatives are also hereby disclosed; more than one element of an invention can have such exclusions, and all combinations of elements having such exclusions are hereby disclosed.

Where a range of values is recited, it is to be understood that each intervening integer value, and each fraction thereof, between the recited upper and lower limits of that range is also specifically disclosed, along with each subrange between such values. The upper and lower limits of any range can independently be included in or excluded from the range, and each range where either, neither or both limits are included is also encompassed within the invention. Where a value being discussed has inherent limits, for example where a component can be present at a concentration of from 0 to 100%, or where the pH of an aqueous solution can range from 1 to 14, those inherent limits are specifically disclosed. Where a value is explicitly recited, it is to be understood that values which are about the same quantity or amount as the recited value are also within the scope of the invention, as are ranges based thereon.

Materials and Methods

Generating sequence similarity network (SSN) and target selection

The SSN of T1.2-like lipases was generated using the Sequence BLAST function of the EFI-EST server with the amino acid sequence of T1.2 lipase (SEQ ID NO: 118) as the input sequence. The SSN was visualized with Cytoscape v3.7.2 at an e-value of 10^{-130} in order to separate the bacterial lipases based on their genus classification. All the targets were selected based on the length of the protein (between 360 and 540 residues) and share a sequence identity of less than 75% (Table 1 and sequence listing). The genes were synthesized by Twist Bioscience and cloned within the BsaI site of the HcKanOR vector (SEQ ID NO: 115). All genes were synthesized without the signal peptide.

Workflow used for the discovery and characterization of lipases

Putative lipase genes were assembled by Golden Gate assembly and transformed into *Escherichia coli* by heat-shock transformation using a PCR machine. The transformed cells were plated onto agar plates with Rotor HDA (Singer Instruments) to obtain single colonies. Picking and inoculation of the colonies were carried out using PiXL (Singer Instruments). For high-throughput screening of lipase activity, induced cells were lysed by sonication using a microplate horn (Qsonica) and the colorimetric assays were set up by OT-2 (Opentrons). To purify the enzymes for further characterization, an autolysis *E. coli* strain XJb (DE3) (Zymo Research) was used to increase the throughput of cell lysis and facilitate purification. Up to six different enzymes can be purified from 50 mL culture each time using Proteomist DTII (CellFree Sciences). The melting temperatures of the purified proteins (T_m) were measured by fluorescence spectroscopy using UNcle (Unchained Labs). The details for each method are listed below.

Cloning, transformation and expression of T1.2-like lipases

Golden Gate assembly was used to subclone the lipase genes from HcKanOR vector (SEQ ID NO: 115) into p15mGG2 vector (SEQ ID NO: 116). In short, each reaction contained 25 ng of BsaI-linearized p15mGG2, 75 ng of lipase-containing HcKanOR vector, 1x T4 DNA Ligase buffer (NEB), 1x bovine serum albumin (NEB), 4U of BsaI-HF (NEB) and 6U of T4 DNA Ligase (NEB) in a 5 mL reaction. The reactions were set up in

a PCR plate and incubated at 37 °C for 5 min and 18 °C for 5 min, for a total of 70 cycles. Heat inactivation of the enzymes was carried out at 50 °C for 15 min followed by 80 °C for 5 min.

High-throughput transformation of the subcloned vectors was carried out by mixing 5
5 μ L of the Golden Gate reaction with 30 μ L of Acella *E. coli* competent cells (MoBiTec) on a PCR plate. Acella cells can be used for the propagation of plasmids for long term storage and expression of target genes due to the *recA* and *endA* deletions in the BL21(DE3) host strain. The mixture was incubated on ice for 20 min before heat shock at 42 °C for 35 s and back on ice for an additional 2 min. The cells were transferred to a
10 deep well block containing 70 μ L of SOC medium for recovery at 37 °C. After 1 h, 100 μ L of LB containing 200 μ g/mL of ampicillin was added to the cells to enrich the positive clones (37 °C for 6 h). Each sample was then diluted serially at 5x, 25x and 125x before plating on a PlusPlate with long 96-well Repads and Rotor HDA. Three biological replicates were picked using PiXL and inoculated into LB containing 100 μ g/mL ampicillin
15 for storage and high-throughput screening.

High-throughput screening with phenol red assay

Small-scale expression of the lipases was carried out by inoculation of cells from glycerol stock into 500 μ L of autoinduction media (AIM) containing Terrific Broth (Bio
20 Basic), 50 mM NaCl, 5 mM Na₂SO₄, 0.015% glucose, 1.2% glycerol, 0.6% lactose and 2 mM MgSO₄. The cells were grown at 30 °C for 24 h before harvesting and resuspension in 500 μ L of storage buffer containing 2.5 mM MOPS, pH 7.2 and 100 mM NaCl. 100 μ L of cells were transferred to a PCR plate and the cells were lysed using a microplate horn with the following program: 90% amplitude, 36 cycles of 5 s on, 15 s off. The clarified cell
25 lysate was diluted 10x with storage buffer and screened for lipase activity using the phenol red assay.

The phenol red assay was set up with OT-2 in a 384-well microplate. Each reaction contained 10 μ L of diluted cell lysate, 5 μ L of substrate (100 mM tributyrin, tricaprylin, triolein or vinyl laurate dissolved in 10% tert-butanol) and 45 μ L of buffer (2.5 mM MOPS,
30 pH 7.2, 5% gum arabic and 0.75 mM phenol red). The samples were incubated at 25 °C for 2 h and the absorbance readings were measured at 560 nm. Cells containing

p15mGG2 (SEQ ID NO: 116) and T1.2 (SEQ ID NO: 118) were used as negative and positive controls for the screen, respectively.

Expression and purification of active lipases

5 To express and purify the active enzymes, Golden Gate assembly was used to subclone the genes from the HcKanOR vector (SEQ ID NO: 115) into the p15mGG6 expression vector (SEQ ID NO: 117), following the steps mentioned above. The constructed vectors were transformed into XJb (DE3) *E. coli* competent cells by heat shock transformation and inoculated into 50 mL LB medium (3 mM arabinose, 1 mM
10 MgCl₂) for expression. After the cells reached stationary phase, 0.1 mM of IPTG was added and expression was carried out at 25 °C for 24 h. The cells were then harvested, resuspended in 5 mL binding buffer (20 mM Tris, pH 7.9, 500 mM NaCl, 5 mM Imidazole) and stored at -80 °C. For cell lysis, the frozen pellets were thawed rapidly in a 37 °C water bath for 10 min. 125U of Benzonase (Sigma) was added to reduce the viscosity of the
15 cell lysate and the clarified supernatant was loaded onto Protomist DTII for HIS-tagged protein purification. Bound proteins were eluted in 1 mL of elute buffer (20 mM Tris, pH 7.9, 500 mM NaCl, 100 mM L-Histidine) and concentrated for storage at -80 °C. The concentration of the purified enzymes was determined from the absorbance reading at 280 nm. Proteins with low purity were quantitated using the His Tag ELISA detection kit
20 (GenScript). The protein yield for selected lipases is shown in Table 2a. The unit of the protein yield is mg/L (yield of protein from 1L of *E. coli* culture. Some of the proteins have a yield lower than 0.1 mg/L which were rounded off to "0.0".

Substrate specificity assays

25 Determination of substrate specificity for each enzyme was carried out using a pH indicator assay adapted from Camacho-Ruiz *et al.* Each reaction contained 2.5 mM MOPS, pH 7.2, 100 mM NaCl, 0.42 mM *p*-nitrophenol and 4.2 mM substrate. Enzyme within linear range was added last to a final volume of 120 µL. The reaction was incubated at 30 °C and the absorbance readings at 405 nm were recorded once every 30 s for a
30 total of 10 min. The amount of fatty acids produced was calculated from the standard curves and specific activity (U/mg) for each enzyme and substrate was calculated from

the initial rate based on the equation reported. The specificity of each enzyme against various substrates can be calculated using the following equations:

$$\text{Chain – length specificity} = \log\left(\frac{SA_{TB}}{SA_{TC}} \times \frac{SA_{VB}}{SA_{VD}}\right)$$

$$\text{Triester/Monoester specificity} = \log\left(\frac{SA_{TB}}{SA_{VB}} \times \frac{SA_{TC}}{SA_{VD}}\right)$$

where SA_{TB} is specific activity against tributyrin, SA_{TC} is specific activity against tricaprylin, SA_{VB} is specific activity against vinyl butyrate and SA_{VD} is specific activity against vinyl decanoate. For SSN mapping, cutoff values for enzymes with specificity towards short chain and medium chain substrates are greater than two (> 2) and less than negative two (< -2), respectively. The cutoff values for enzymes with specificity against triesters and monoesters are also greater than two and less than negative two, respectively. The results of the specific activities for different substrates and the substrate specificities for selected lipases are shown in Tables 2a and 2b.

15 Optimal pH of T1.2-like lipases

Different pH indicator and buffer pairs were used to determine the specific activity of purified lipases at different pH values. For pH 5.0, 1.5 mM of malate buffer and 0.11 mM bromocresol green ($\lambda_{\max} = 620$ nm) were used. For pH 6.0, 2.5 mM of MES buffer and 0.21 mM of chlorophenol red ($\lambda_{\max} = 575$ nm) were used. 2.5 mM of Tris buffer and 0.21 mM of cresol red ($\lambda_{\max} = 575$ nm) were used at pH 8.2 while 2.5 mM of CHES buffer and 0.42 mM of thymol blue ($\lambda_{\max} = 600$ nm) were used at pH 9.2. The reaction condition is the same as the substrate specificity assay described above. The specific activities for selected lipases at different pH values are shown in Table 2b.

25 Determination of melting temperature (T_m)

The melting temperature of proteins (T_m) were measured using UNcle (Unchained Labs). 9 μ L of purified proteins (2 mg/mL) were loaded into quartz wells in duplicates and sealed within an aluminum thermal transfer holder before loading into the instrument. The instrument was programmed to heat the samples from 20 °C to 95 °C at a rate of 0.3 °C/min. Fluorescence spectra from 250 – 720 nm were acquired 4 times every 5 s. UNcle

measures the T_m of proteins using differential scanning fluorimetry, which is the fluorescence shift of intrinsic tyrosine or tryptophan residues as the protein unfolds when temperature increases. Data were analyzed using UNcle analysis software, and T_m were calculated using the barycentric mean method. The melting temperature for selected lipases are shown in Table 2c.

Methanol tolerance assay

The methanol tolerance of the purified enzymes was measured with resorufin lipase substrate (Sigma). 1 – 5 μ g of each enzyme was incubated in 100 μ L of 20 mM Tris, pH 8 containing 0 – 40% methanol for 1 h at 25 °C. The enzyme was then diluted 1 – 1000X using the same buffer and 100 μ L of the diluted enzyme was added to the substrate solution (100 μ L) containing 20 μ g/mL resorufin, 10 μ M CaCl_2 , 1 mM sodium citrate, 20 mM Tris (pH 8) and 0 – 40% methanol. The reaction was incubated at 30 °C for 10 min in a microplate reader and fluorescence readings (λ_{ex} : 560 nm; λ_{em} : 610 nm) were measured at a 30 s interval. The initial rates of the enzymes in various methanol concentration were calculated and normalized against the activity of the enzyme in the absence of methanol. The methanol tolerance results for selected lipases are shown in Table 2c.

TLC assay

For biodiesel production, the transesterification of triolein with methanol was carried out in a 10 μ L reaction containing 8 μ L triolein, 1 μ L methanol and 1 μ L enzyme (1 – 5 μ g in 20 mM MOPS, pH 7.2 and 100 mM NaCl). The reaction was incubated at 37 °C for 24 h and stopped with 90 μ L of isopropanol. For esterification reaction, oleic acid was used instead of triolein. The products were analyzed by spotting 1 μ L on a Silica gel 60 F₂₅₄ TLC plate (Merck) and developed with a solvent containing hexane, diethyl ether and formic acid (80:20:2). After 10 min, the plate was removed to air dry and stained with iodine. To drive the completion of triolein transesterification and oleic acid esterification, the reaction was carried out for 3 days, with fresh enzyme (1 μ L) and methanol (1 μ L) added once a day.

Results and Discussion

Sequence similarity network of bacterial T1.2-like lipases

Lipases isolated from thermophilic microorganisms often exhibit a higher thermostability and increased tolerance towards organic solvents, making them suitable targets for biodiesel production. To date, more than twenty bacterial lipases have been characterized, and many were discovered to possess a wide range of physiochemical properties suitable for various industrial applications. Although a number of thermostable lipases from *Geobacillus* spp. have been reported, most of the assays used for the detection of lipase activity rely heavily on the measurement of substrate hydrolysis. To better determine if certain lipases are able to produce biodiesel, the selected targets were characterized by not only substrate hydrolysis, but also transesterification activity in the presence of high methanol concentration.

Computational tools are widely used to facilitate the discovery of novel enzymes with unknown functions. In particular, the EFI-EST web-based tool allows users to create sequence similarity networks (SSN) based on protein sequences available in public databases, which can then be used to assist the discovery of closely related enzymes from a known target. Based on the selection criteria determined by the user, targeted genes can be synthesized and subjected to an automated workflow to identify the enzyme with the most desired properties. More recently, machine-learning-based tools have been developed to allow the automated mining of soluble enzymes with diverse catalytic properties. As it remains challenging to determine *a priori* if an enzyme can be produced in its active form in non-native organisms, experimental screening of unknown enzymes expressed in various heterologous hosts still constitutes an essential part of the enzyme discovery workflow.

To identify lipases suitable for biodiesel production, an SSN was created based on a thermostable lipase isolated from a *Geobacillus* sp (T1.2, SEQ ID NO: 118). This enzyme shares a high (97%) sequence identity with a putative lipase from *Geobacillus* sp. B4113_201601 (UniProt: A0A150MY45). The SSN of the T1.2-like lipases contains a total of 705 unique sequences and is grouped into 13 major clusters (identified with reference numbers 1 to 13 in FIG. 2) with more than five nodes in each cluster at an e-value of 10^{-130} (40% sequence identity). Most of the *Geobacillus* lipases (54 nodes) fall under cluster

3 with a sequence identity of at least 87%. The other major clusters include lipases from *Staphylococcus* spp., *Bacillus* spp., *Clostridium* spp., *Deinococcus* spp., *Massilia* spp., *Paenibacillus* spp., *Pelosinus* spp. and *Firmicutes* spp.. From the SSN, a total of 114 targets were selected for gene synthesis using the following criteria: the target must (1) have a sequence length between 360 and 540 residues and (2) shares a sequence identity of less than 75%. The only exceptions are lipases from *Geobacillus* spp. with known PDB structures (FIG. 2). They were included despite having high sequence identity (more than 91%).

High-throughput cloning, expression, screening, and purification of T1.2-like lipases

A workflow was developed to clone, express, transform and screen all the 113 synthesized bacterial lipases for their ability to hydrolyze various natural and unnatural substrates (FIG. 1b). Lipase SEQ ID NO: 7 was not synthesized successfully and was not included in the assay in FIG. 3. Golden Gate cloning was used to subclone the synthesized gene from the donor vector (HcKanOR, SEQ ID NO: 115) to the expression vector (p15mGG2, SEQ ID NO: 116) in a one-pot reaction. Golden Gate assembly is highly accurate and efficient, eliminating the need to sequence each of the individual constructs after assembly. The method is also amenable to automation and hence allows the scalability of the workflow. Since *E. coli* cannot secrete lipases efficiently into the extracellular space, the expression vector was designed with a T7 promoter for expression and contained a RFP reporter gene to facilitate the counter-selection of negative clones. All the reactions were set up in a 96-well plate format and transformed directly into Acella (DE3) *E. coli* cells. The transformed cells were plated using Rotor HDA and white colonies were picked with PiXL. Use of Rotor HDA for cell plating and PiXL for colony picking significantly reduced the amount of time and resources required when compared to traditional methods. Such an approach can be used to clone and transform hundreds of genes into *E. coli* for screening within one week.

To detect the hydrolytic activity of lipases, cells were induced for expression using autoinduction media and lysed by sonication in a 96-well plate. Phenol red was used as a pH indicator to detect the release of free fatty acids from hydrolysis due to its high sensitivity and low background from cell lysate. Four different substrates (tributyryl, tricaprylin, triolein and vinyl laurate) were used for the screen and lipase activity was

detected in 72 of the targets (FIG. 3). The remaining targets without any detectable activity could be attributed to one of the following reasons: (1) low expression level in *E. coli* or low solubility, (2) inactive enzyme or (3) different substrate preference. It is interesting to note that none of the targets from cluster 7 (*Massilia* spp.) have detectable lipase activity (FIG. 2A). It is revealed from the UniProt protein database that some of the proteins from this cluster share sequence homology with members of the palmitoyl protein thioesterase superfamily (PF02089). In higher organisms, these enzymes are involved in the hydrolysis of palmitate from cysteine residues during lysosomal degradation.²³ As the function of these enzymes in *Massilia* spp. is still unclear, it is likely that the substrates for these enzymes are thioesters but not triacylglycerides (TAGs).

To purify the active targets for further characterization, the enzymes were subcloned into p15mGG6 (SEQ ID NO: 117). p15mGG6 is a modified T7 expression vector that carries an N-terminal His Tag to facilitate purification by affinity chromatography. The constructs were transformed into XJb (DE3) (an autolysis strain of *E. coli*) cells to increase the throughput by eliminating manual cell lysis prior to protein purification. Each enzyme was purified using an automated protein purification system (Protemist DTII). The SDS-PAGE analysis of the purified enzymes revealed that at least 54 of the targets can be purified to greater than 90% homogeneity (FIG. 4). The concentration and yield of the purified proteins were estimated by measuring the absorbance at 280nm. For targets with low purity (SEQ ID NO: 2, SEQ ID NO: 19, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 39, SEQ ID NO: 45, SEQ ID NO: 61, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 85, SEQ ID NO: 101, SEQ ID NO: 107), a His Tag ELISA kit was used to estimate the concentration of the His-tagged proteins. The target with the highest yield (SEQ ID NO: 77) is isolated from *Clostridium sporogenes* (35 mg/L) while most of the other lipases have yield ranging from 0.01 – 30 mg/L (FIG. 2B and FIG. 10).

Activity and substrate specificity of T1.2-like lipases

The specific activity of each purified enzyme was determined using a microplate assay with *p*-nitrophenol and various substrates. *p*-nitrophenol is a pH indicator that results in an absorbance change at 405 nm upon protonation by the released fatty acids. The substrates selected for the assay (tributyrin, tricaprylin, vinyl butyrate and vinyl

decanoate) have all been shown to be compatible with the accurate measurement of lipase activity when used at a low concentration (less than 5 mM). This allows the rapid determination of substrate specificity and preference for each of the purified enzymes. Among the 72 active targets that were initially identified from the phenol red assay, 66 of the purified lipases were found to possess detectable activity against at least one of the substrates (FIG. 10 and Table 2a). The highest activity was detected against vinyl decanoate for SEQ ID NO: 92, SEQ ID NO: 93 and SEQ ID NO: 100 with a specific activity of at least 1.2×10^4 U/mg. SEQ ID NO: 55, SEQ ID NO: 109 and SEQ ID No. 35 has the highest activity against tributyrin, tricaprylin and vinyl butyrate, respectively (1.9×10^3 U/mg, 1.9×10^3 U/mg and 1.3×10^3 U/mg) (Table 2a).

To determine the specificity of the purified enzymes against substrates with various chain length, we compared the activity of each enzyme against short chain (tributyrin, vinyl butyrate) and medium chain (tricaprylin, vinyl decanoate) substrates (Tables 2a and 2b). The SSN of the substrate specificity mapping shows that most lipases isolated from *Geobacillus* spp. have a substrate preference towards medium chain fatty acid (FIG. 2C). Two of the major *Clostridium* clusters (4 and 5) also displayed a difference in their preference towards short and medium chain substrates. The *Clostridium* spp. from cluster 4 have a preference towards medium chain substrates while those from cluster 5 favor shorter chain substrates. The difference in substrate preference is likely to be resulted from the divergent evolution of *Clostridium* spp. through their adaption in different environments.

The specificity of the purified enzymes towards monoesters and triesters can also be determined by comparing the activity of each enzyme against vinyl butyrate/vinyl decanoate and tributyrin/tricaprylin (FIG. 2D). Most lipases have higher specificity towards the vinyl monoesters instead of triglycerides. This could be attributed to the reduced number of acyl chains present on monoesters which results in a better fit of the substrates in the active site of the enzymes. One exception is SEQ ID NO: 74, which displays a slight preference towards triglycerides over vinyl monoesters. The lack of a crystal structure for SEQ ID NO: 74 makes it difficult to rationalize the substrate specificity observed for this enzyme.

Effect of pH on the specific activity of T1.2-like lipases

Lipases that can survive at extreme pH are useful for various industrial applications. For example, acid-tolerant lipases can be used for the production of biodiesel from acid oil while alkaline lipases are useful for detergent formulations. Despite its accuracy, the pH-stat method by alkali titration cannot be used to rapidly measure the activity of lipases at various pH. Therefore, a broad range pH indicator assay from Camacho *et al.* was adopted to measure the hydrolytic activity of the purified enzymes. This technique allows the use of natural substrates for the reaction and can be easily adapted in a microplate format.

Among the lipases that prefer acidic conditions, one enzyme (SEQ ID NO: 30) was identified that works best at pH 5.0 and five enzymes (SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 46, SEQ ID NO: 85, SEQ ID NO: 113) were identified that work best at pH 6.0 (Tables 2 and 3b). From the SSN, SEQ ID NO: 30 is revealed to be a singleton isolated from *Dyella jiangningensis* (FIG. 2E). This is not surprising since *Dyella* spp. are commonly isolated from acidic habitats and enzymes from these organisms are expected to exhibit tolerance towards a lower pH range. The other lipases with moderate acid tolerance are mainly derived from *Clostridium* spp (Cluster 5). Among the four *Clostridium* spp that were identified, *Clostridium acidisoli* (SEQ ID NO: 46) was isolated from acidic peat-bog soil. *Clostridium saccharoperbutyl-acetonicum* (SEQ ID NO: 85) and *Clostridium felsineum* (SEQ ID NO: 40) were discovered to be capable of utilizing butyric and pectic acids as carbon source for fermentation, respectively. The extracellular lipases secreted by these microorganisms are likely to be capable of withstanding the acidic environment of their respective hosts. It is also interesting to note that acidic lipases were only identified from cluster 5 but not cluster 4, suggesting that only some *Clostridium* spp from cluster 5 have adapted for survival in acidic environments.

For alkaline tolerance, there are three enzymes (SEQ ID NO: 54, SEQ ID NO: 98, SEQ ID NO: 107) that work best at pH 8.2 and 10 enzymes (SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 59, SEQ ID NO: 68, SEQ ID NO: 73, SEQ ID NO: 83, SEQ ID NO: 86, SEQ ID NO: 97, SEQ ID NO: 101) that work best at pH 9.2 (FIG. 10 and Table 2b). SEQ ID NO: 68 and SEQ ID NO: 1 have the highest preference towards alkaline pH, with about 10 to 100-fold increase in activity when the pH was increased from

7.2 to 9.2 (260 U/mg to 3400 U/mg and 28 U/mg to 2900 U/mg, respectively). Most of the alkaline lipases that were identified are isolated from *Bacillus* and *Clostridium* spp, and many *Bacillus* spp are known to survive a high pH environment. Although a number of alkaline lipases were identified from the screen, no sequence-based correlation between the alkaline lipases from the SSN were able to be identified (FIG. 2E).

The assay used to determine the pH range of the purified enzymes provides a high throughput approach to rapidly determine the optimal pH of each enzyme. Although the assay has some limitations (reduced accuracy and limited pH range), it can be used to characterize many purified proteins within a short period of time. The data obtained suggests that despite the sequence similarity observed in some of these enzymes, their optimal pH profile can be significantly different. It is therefore unlikely that one can predict *a priori* if a lipase has the required properties for an application from the protein sequence alone.

Thermostability and methanol tolerance of T1.2-like lipases

Enzymes with high thermostability are often ideal for industrial applications as it increases the reusability of these enzymes and can decrease production cost. In addition, identifying a thermostable lipase is important as most industrial processes involving fats and lipids occur at elevated temperatures due to the high melting points of some feed stocks. To determine the thermostability of each purified enzyme, the melting temperature (T_m) of these proteins was measured using fluorescence spectroscopy. It has been shown that comparison between the T_m of each purified enzyme can provide a quick and quantitative method to assess the thermostability of these enzymes.

The T_m of each lipase was measured and shown on the SSN based on their temperature range in FIG. 2F. Not surprisingly, most of the enzymes isolated from the *Geobacillus* cluster (Cluster 3) have a T_m of at least 60 °C (Table 2c). Among these enzymes, SEQ ID NO: 96 (*Geobacillus stearothermophilus*) has the highest T_m at 87.0 °C. Outside cluster 3, SEQ ID NO: 70 (*Desulfosporosinus orientis*), SEQ ID NO: 5 (*Pelosinus fermentans*) and SEQ ID NO: 35 (*Thermosyntropho lipolytica*) are the only lipases that possess a T_m of at least 70.0 °C (Table 2c). The thermostability of SEQ ID NO: 35 is likely to be traced to the thermophilic nature of the host organism. It is interesting

to note that SEQ ID NO: 53 (*Clostridium botulinum*) has a high T_m (65.8 °C) even though most of the other lipases from the same cluster (cluster 4) have a lower T_m range (41.1 – 55.6 °C). These observations highlight the importance of screening for thermostable enzymes from mesophiles, instead of relying on enzymes from thermophilic organisms.

5 Although thermophilic enzymes are generally more thermostable than their mesophilic counterparts, they are often less active if the reaction has to be carried out at lower temperature. Most of the other lipases from the *Staphylococcus* spp. (Cluster 1), *Bacillus* spp. (Cluster 2) and *Clostridium* spp. (Cluster 3 and 4) clusters were discovered to have a lower T_m of less than 50 °C.

10 Since the enzymatic production of biodiesel typically requires the use of methanol at high concentrations in order to promote the transesterification of TAGs, the enzymes used for the reaction should exhibit tolerance towards high methanol concentrations. Structural analysis of proteins with enhanced thermostability and organic solvent tolerance have revealed that these characteristics are often mediated by intramolecular forces such as
15 the formation of hydrogen bonds, ionic interactions and hydrophobic interactions. However, it appears that there is a lack of direct correlation between the thermostability and organic solvent tolerance of lipases as some enzymes with high thermostability often display low tolerance towards organic solvent. Due to the sensitivity of the resorufin lipase substrate, it was used to screen against the lipases to determine their tolerance towards
20 increasing methanol concentrations.

Although most of the lipases were found to retain their activity in the presence of 40 % methanol, many of them have lower lipase activity in the presence of 10 – 30% methanol (FIG. 10 and Table 2c). This observation is likely due to the complexity of the reaction since methanol can act as both substrate and inhibitor of the reaction. At low
25 methanol concentrations, hydrolysis is the dominant reaction while transesterification takes over at higher methanol concentrations. Alternatively, some lipases are known to observe increased k_{cat} values in the presence of organic solvents due to changes in the environment of the active site. Nevertheless, it was possible to rank each enzyme according to its tolerance towards 40% methanol and the results are summarized in FIG.
30 2G.

From the screen, SEQ ID NO: 73 (uncultured bacterium) was found to possess the highest tolerance toward methanol, with a 104-fold increase in activity in the presence of 40% methanol. The least methanol-tolerant lipase is SEQ ID NO: 83 (*Staphylococcus massiliensis*), which retained only 28% of its activity in the presence of 10% methanol.

5 Lipases from the *Geobacillus* spp (cluster 3) were generally tolerant towards methanol, and the least tolerant enzyme from the cluster was able to retain 47% of its activity in the presence of 40% methanol. The results confirmed that beyond cluster 3, it is not possible to establish a good correlation between the thermostability and the methanol tolerance of lipases. The difference in methanol tolerance can also be observed in clusters with high
10 sequence similarity, suggesting that slight changes to amino acid sequences are sufficient to alter the tolerance of lipases towards methanol. An example is SEQ ID NO: 36 (*Clostridium cavendishii*) and SEQ ID NO: 85 (*Clostridium saccharoperbutylacetonicum*) from cluster 5, which share high sequence identity (71.6%) while having very different tolerance towards methanol.

15

Biodiesel production from T1.2-like lipases

Thin layer chromatography (TLC) is a cheap, reliable, and rapid method that can be used to detect the production of biodiesel from lipases. To determine if the purified enzymes can be used to produce biodiesel, a reaction containing triolein and methanol
20 was set up to detect methyl oleate (FAME) produced from the transesterification of triolein. After a 24-hour incubation and separation on TLC plate, significant amount of methyl oleate produced by 27 of the purified enzymes (identified with a white asterisk in FIG. 5) was detected including SEQ ID NO: 5, SEQ ID NO: 22, SEQ ID NO: 27, SEQ ID NO: 35, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO:
25 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 100, SEQ ID NO: 107, SEQ ID NO: 108, and SEQ ID NO: 109. The esterification reaction was also carried out using oleic acid and methanol as substrates, and six of the purified enzymes
30 were observed to possess esterification activity (identified with a white asterisk in FIG. 6) mainly SEQ ID NO: 57, SEQ ID NO: 74, SEQ ID NO: 90, SEQ ID NO: 92, SEQ ID NO:

93, and SEQ ID NO: 108. To demonstrate the ability of these enzymes to catalyze the complete transesterification of TAG, the reaction was also carried out for 3 days with fresh methanol and enzyme added to the reaction once a day. In addition to the commercial lipase (Eversa®, lipase from *Aspergillus oryzae*) from Novozyme, six of the bacterial lipases (SEQ ID NO: 27, SEQ ID NO: 53, SEQ ID NO: 55, SEQ ID NO: 60, SEQ ID NO: 73 and SEQ ID NO: 107) were found to be able to convert all the TAG into FAME within 3 days (FIG. 7). The conversion is around 90-95% FAME based on titration with NaOH. Interestingly, in a comparison of the activities of SEQ ID NO: 53 (T_m of 65.8 °C), SEQ ID NO: 90 (T_m of 72.3 °C), and the commercial Eversa® at 60 °C, FAME production was found to be highest with SEQ ID NO: 90, followed by SEQ ID NO: 53 and Eversa® (FIG. 8). SEQ ID NO: 90 shares a high sequence similarity (92%) with T1.2 (SEQ ID NO: 118) and has the advantage over Eversa® to be used in industrial processes that operate at higher temperatures. SEQ ID NO: 90 is more thermostable than Eversa® as Eversa® seems to be denatured at 60 °C. SEQ ID NO: 53 also shows slight activity at 60°C. SEQ ID NO: 90 is also able to retain its activity in the presence of 70% methanol (FIG. 9), highlighting the unique characteristics of SEQ ID NO: 90 which make it an ideal candidate for biodiesel production in industrial settings. The lipase for use in the production of FAME should preferably be to achieve at least 90% of the commercially available Eversa® lipase. For example, the lipase used should be able to achieve at least 80% yield of the FAME, and preferably at least 85%, more preferably at least 90%, and even more preferably at least 95%.

From the SSN, it was observed that lipases with high transesterification activity are localized in cluster 2, 3 and 4 while most lipases with high esterification activity are in cluster 3 (FIG. 2H). SEQ ID NO: 74 (*Desulfosporosinus acidiphilus*) is the only enzyme outside cluster 3 with esterification activity (cluster 2). Not surprisingly, most of the lipases that catalyze the transesterification of triolein also display a higher preference towards medium chain substrates and a higher tolerance towards methanol (FIGs. 2C and 2G). It is also interesting to note that although all the lipases from cluster 3 possess transesterification activity, only five of them were found to possess esterification activity. Given that lipases from cluster 3 have similar physicochemical properties and substrate specificity, it is likely that the ability of these enzymes to catalyze esterification is due to

the ability of the enzyme to bind oleic acid for productive catalysis (FIG. 10). As these enzymes have high sequence identities (more than 88%), the structures of these enzymes could provide clues on how subtle changes of amino acid residues might alter the type of reaction that can be catalyzed by these lipases. Lipases found to possess high esterification activity may be developed for the production of biodiesel from the byproducts of plant oil refining (palm oil fatty acid distillate), or the removal of residual free fatty acids from biodiesel.

A scalable workflow for the identification of candidate enzymes that may be further developed for use as green alternatives to existing chemical or other environmentally unsustainable industry practices has been described. The utility of this workflow in the identification of candidate lipases as green alternatives to produce biodiesel to replace fossil fuels was demonstrated. The workflow integrates available web-based tools with automation for scalability. The EFI-EST web-based tool was used to create a sequence similarity network of bacterial lipases based on a thermostable lipase from *Geobacillus* sp. More than 100 genes were targeted for synthesis, expression and screening in *E. coli*. The traditional approach of cloning, expression, and screening of enzymes one at a time is often labor-intensive and timing consuming. With recent developments in Golden Gate technology and the availability of automated colony pickers and liquid handlers, these enzymes can now be cloned, expressed, and purified for screening within a short period of time.

Despite the high sequence identity among some of the lipases that were selected, many were found to possess different substrate specificities and physicochemical properties. More structural studies are needed to better correlate the protein sequence of lipases with their function and physicochemical characteristics. This study also highlights that genome-mining can be used as a complementary approach to the engineering of enzymes for industrial applications. In addition to the 22 lipases that were identified to produce biodiesel, some of the lipases were found to possess optimal activity at acidic or alkaline pH. These enzymes can be further developed for use in conditions with extreme pH.

The workflow described herein may be further enhanced through integration with machine-learning algorithms. Machine-learning-based approaches have recently gained

popularity in enzyme discovery efforts as they may be used to recognize patterns found within a protein's sequence and generate accurate predictions of the protein from its orthologs. It has been successfully used to predict the solubility, thermostability and substrate specificity of many different enzyme superfamilies. Selection of targets from the SSN can be assisted by machine-learning algorithms to increase the number of soluble enzymes entering the pipeline. After several targets are isolated, machine-learning algorithms may be used to combine the useful traits (e.g. high expression yield, activity and thermostability) found in these enzymes and suggest sites for mutagenesis efforts, since it is unlikely to identify a 'perfect' enzyme with all the useful traits. The candidate enzymes may then be further engineered for commercial use in industrial applications.

With the onset of the Fourth Industrial Revolution, most industrial practices have begun to gear towards the use of automation and the adoption of green manufacturing practices. Such practices aim to improve energy efficiency, while resulting in lower levels of pollution and the decreased emission of greenhouse gases. Production of biodiesel with lipases is one example of industrial green manufacturing and can help to reduce a country's reliance on fossil fuels. The workflow described herein may also be applied to other industrial processes such as the bioleaching of valuable metals from spent batteries or the enzymatic degradation of plastic waste. Taken together, the scalable workflow may be adapted to identify important biocatalysts that may be used for the upcycling of waste products to achieve a circular bioeconomy.

Table 1: T1.2-like lipases that were expressed successfully and characterized.

SEQ ID NO:	Organism	Uniprot ID
1	<i>Bacillus manliponensis</i>	A0A073JVM2
2	<i>Bacillus gaemokensis</i>	A0A073K489
5	<i>Pelosinus fermentans DSM 17108</i>	A0A0A0YMQ9
19	<i>Deinococcus actinosclerus</i>	A0A0U4BXG4
21	<i>Staphylococcus simulans</i>	A0A133QFT9

22	<i>Geobacillus</i> sp. B4113_201601	A0A150MY45
23	<i>Staphylococcus pseudintermedius</i>	A0A1B1NXZ9
25	<i>Clostridium</i> sp. W14A	A0A1C0BQG4
26	<i>Clostridium taeniosporum</i>	A0A1D7XMR8
27	<i>Clostridium taeniosporum</i>	A0A1D7XMV6
30	<i>Marininema halotolerans</i>	A0A1G8VHQ4
31	<i>Clostridium estertheticum</i> subsp. <i>estertheticum</i>	A0A1I6S0B4
35	<i>Thermosyntropha lipolytica</i> DSM 11003	A0A1M5LUW5
36	<i>Clostridium cavendishii</i> DSM 21758	A0A1M6HQT9
38	<i>Paenibacillus macerans</i> (<i>Bacillus macerans</i>)	A0A1R1HF91
39	<i>Crenothrix polyspora</i>	A0A1R4HDF8
40	<i>Clostridium felsineum</i> DSM 794	A0A1S8LBI4
42	<i>Clostridium oryzae</i>	A0A1V4IRK2
43	<i>Firmicutes bacterium</i> ADurb.Bin262	A0A1V5TG20
44	<i>Firmicutes bacterium</i> ADurb.Bin262	A0A1V5TIK8
46	<i>Clostridium acidisoli</i> DSM 12555	A0A1W1XBE1
53	<i>Clostridium botulinum</i> (strain Hall / ATCC 3502 / NCTC 13319 / Type A)	A5I055
54	<i>Clostridium botulinum</i> C str. Eklund	B1B8C8
55	<i>Clostridium botulinum</i> (strain Eklund 17B / Type B)	B2TQP5
56	<i>Clostridium botulinum</i> (strain Eklund 17B / Type B)	B2TQP6
57	<i>Geobacillus</i> sp. SBS-4S	B9ZZP0
59	uncultured bacterium	C1ING4
60	<i>Bacillus thermoamylovorans</i>	C4T9I5

61	<i>Clostridium botulinum</i> E1 str. 'BoNT E Beluga'	C5UWM9
62	<i>Clostridium botulinum</i> D str. 1873	C5VNH6
64	<i>Aneurinibacillus thermoaerophilus</i>	D3XB96
66	<i>Deinococcus maricopensis</i> (strain DSM 21211 / LMG 22137 / NRRL B-23946)	E8U7U4
68	<i>Desmospora</i> sp. 8437	F5SFI7
70	<i>Desulfosporosinus orientis</i> (strain ATCC 19365) (<i>Desulfotomaculum orientis</i>)	G7WDF7
72	<i>Paenibacillus dendritiformis</i> C454	H3SPG0
73	uncultured bacterium	H9LD09
74	<i>Desulfosporosinus acidiphilus</i> (strain DSM 22704 / JCM 16185 / SJ4)	I4D483
77	<i>Clostridium sporogenes</i> (strain ATCC 15579)	J7T580
79	<i>Burkholderia</i> sp. SJ98	K8R3K1
80	<i>Staphylococcus massiliensis</i> S46	K9ARP2
82	<i>Staphylococcus massiliensis</i> S46	K9B136
83	<i>Staphylococcus massiliensis</i> S46	K9B5G6
85	<i>Clostridium saccharoperbutylacetonicum</i> N1-4(HMT)	M1MXZ7
86	<i>Bacillus</i> sp. GeD10	N1LN49
87	<i>Geobacillus stearothermophilus</i> (<i>Bacillus stearothermophilus</i>)	O66015
88	<i>Staphylococcus aureus</i> (strain Mu50 / ATCC 700699)	P65288
89	<i>Geobacillus thermocatenulatus</i>	Q59260
90	<i>Bacillus</i> sp. L2	Q5I4I3
91	<i>Geobacillus kaustophilus</i> (strain HTA426)	Q5KYG5
92	<i>Bacillus</i> sp.	Q5U780

93	<i>Geobacillus zalihae</i>	Q842J9
95	<i>Geobacillus stearothermophilus (Bacillus stearothermophilus)</i>	Q93A71
96	<i>Geobacillus stearothermophilus (Bacillus stearothermophilus)</i>	Q9L6D3
97	<i>Deinococcus radiodurans (strain ATCC 13939)</i>	Q9RXH9
98	<i>Firmicutes bacterium CAG:103</i>	R5ASH5
100	<i>Firmicutes bacterium CAG:555</i>	R5CTD5
101	<i>Ruminococcus sp. CAG:488</i>	R5XUW0
103	<i>Ruminococcus sp. CAG:177</i>	R6I3Z8
105	<i>Clostridium sp. CAG:413</i>	R6NFF0
107	<i>Clostridium sp. CAG:413</i>	R6NV36
108	<i>Geobacillus kaustophilus GBlys</i>	U2Y6D5
109	<i>Clostridium tetani 12124569</i>	U6EWE7
110	<i>Enterobacter cloacae S611</i>	V5A6K1
113	<i>Clostridium bornimense</i>	W6RTI9

Table 2a: Activity of selected lipases

SEQ ID NO:	Protein Yield	Substrate				Chain- length specificity
		Tributylin	Tricaprylin	Vinyl Butyrate	Vinyl Decanoate	
1	19.1	14	4	44	26	1.78
2	0.0	20	18	137	336	-0.79
5	7.1	46	31	56	554	-1.90
19	0.0	28	9	26	142	-0.56
21	10.0	251	421	159	2475	-3.26
22	10.5	589	776	493	10454	-3.33
23	4.2	7	0	17	23	1.64
25	0.1	25	16	30	154	-1.19
26	7.3	19	7	242	1069	-0.49
27	1.6	598	606	98	3326	-3.54
30	30.7	302	20	359	22	5.51
31	0.2	1	1	4	47	-2.46
35	15.9	925	184	1295	8082	-0.22
36	10.0	326	30	197	31	4.23
38	16.8	244	18	40	25	3.08
39	0.0	12	2	10	10	1.79
40	13.8	16	8	21	16	0.97
42	12.9	171	94	228	251	0.50
43	30.0	60	59	5	151	-3.39
44	5.7	45	0	0	27	0.51
45	0.0	1	2	26	5	0.96
46	0.6	431	4	28	178	2.83
53	6.3	504	765	90	8696	-4.99
54	23.8	1266	130	224	2155	0.01
55	2.1	1912	1377	158	4923	-3.11
56	4.1	25	0	198	920	1.68
57	16.3	881	1279	418	9565	-3.50
59	7.2	25	9	45	114	0.09
60	11.7	162	411	78	2010	-4.18
61	0.1	12	26	67	154	-1.61
62	5.8	407	68	77	938	-0.71
64	26.1	317	143	55	2999	-3.20
66	0.4	14	101	40	392	-4.26
67	0.0	6	6	27	9	1.10
68	20.1	2	0	7	12	0.15
70	21.4	471	269	255	2425	-1.69
72	0.5	266	284	327	840	-1.01
73	17.2	97	80	79	1149	-2.48

74	23.9	86	92	6	88	-2.75
77	35.0	26	9	40	60	0.66
79	11.6	103	246	49	426	-3.03
80	11.6	5	0	0	2	0.92
82	3.2	76	2	96	49	4.31
83	6.1	89	190	50	432	-2.91
85	0.1	16	0	21	17	2.98
86	21.0	22	3	2	24	-0.49
87	13.8	616	784	271	7273	-3.53
88	16.8	14	0	34	7	4.22
89	15.2	606	91	328	9391	-1.46
90	14.2	604	768	308	9809	-3.70
91	13.0	702	914	345	11492	-3.77
92	13.5	1048	1448	361	12740	-3.89
93	14.2	1128	1184	353	12481	-3.61
95	14.1	716	681	312	9478	-3.36
96	14.5	665	827	256	11011	-3.98
97	1.0	330	126	90	338	-0.36
98	22.5	117	36	129	344	0.20
100	19.9	488	757	481	12779	-3.72
101	0.6	38	31	33	91	-0.81
103	21.5	75	43	41	121	-0.53
105	13.0	15	12	12	32	-0.76
107	0.2	254	52	43	278	-0.28
108	22.6	815	870	562	10924	-3.03
109	10.8	300	1923	126	6699	-5.83
110	25.3	78	26	54	578	-1.27
113	14.9	44	22	20	72	-0.59

Table 2b: Activity of selected lipases

SEQ ID NO:	Triester/ Monoester specificity	pH				
		5.0	6.0	7.2	8.2	9.2
1	-3.02	0 ^b	8 ^b	28 ^b	1450 ^b	2907 ^b
2	-4.85	21 ^b	79 ^b	310 ^b	292 ^b	353 ^b
5	-3.08	35 ^b	508 ^b	864 ^b	758 ^b	771 ^b
19	-2.68	19 ^b	72 ^b	121 ^b	106 ^b	93 ^b
21	-1.31	0 ^b	74 ^b	2003 ^b	1372 ^b	1300 ^b
22	-2.42	125 ^b	828 ^b	2044 ^b	948 ^b	1118 ^b
23	-4.02	0 ^a	2 ^a	5 ^a	2 ^a	2 ^a
25	-2.45	13 ^a	20 ^a	13 ^a	22 ^a	32 ^a
26	-7.57	29 ^b	123 ^b	1000 ^b	694 ^b	427 ^b

27	0.11	105 ^b	456 ^b	3973 ^b	1648 ^b	1405 ^b
30	-0.27	34 ^a	25 ^a	10 ^a	18 ^a	8 ^a
31	-5.24	0 ^b	0 ^b	22 ^b	17 ^b	0 ^b
35	-4.12	184 ^b	1562 ^b	7064 ^b	4990 ^b	2644 ^b
36	0.47	89 ^a	52 ^a	101 ^a	39 ^a	12 ^a
38	1.48	4 ^a	13 ^a	43 ^a	9 ^a	0 ^a
39	-1.43	4 ^a	36 ^a	10 ^a	3 ^a	4 ^a
40	-0.97	8 ^b	81 ^b	77 ^b	23 ^b	0 ^b
42	-1.27	31 ^b	62 ^b	67 ^b	37 ^b	10 ^b
43	1.55	27 ^b	33 ^b	114 ^b	73 ^b	70 ^b
44	0.51	0 ^a	9 ^a	19 ^a	19 ^a	13 ^a
45	-4.17	1 ^a	0 ^a	5 ^a	2 ^a	2 ^a
46	-1.06	3 ^b	47 ^b	32 ^b	2 ^b	1 ^b
53	-0.71	204 ^b	2269 ^b	9425 ^b	7731 ^b	4678 ^b
54	-1.08	10 ^b	498 ^b	1042 ^b	1118 ^b	966 ^b
55	1.22	10 ^b	2258 ^b	5731 ^b	4583 ^b	2617 ^b
56	-8.89	0 ^b	55 ^b	626 ^b	446 ^b	383 ^b
57	-1.27	531 ^b	787 ^b	3786 ^b	2740 ^b	1646 ^b
59	-3.13	0 ^b	15 ^b	70 ^b	231 ^b	458 ^b
60	-0.86	6 ^b	327 ^b	1837 ^b	1433 ^b	1517 ^b
61	-3.50	22 ^b	62 ^b	290 ^b	111 ^b	105 ^b
62	-0.96	24 ^b	231 ^b	1125 ^b	531 ^b	248 ^b
64	-1.29	17 ^b	892 ^b	2388 ^b	2346 ^b	2012 ^b
66	-2.41	13 ^b	13 ^b	484 ^b	136 ^b	16 ^b
67	-1.91	70 ^b	669 ^b	2862 ^b	1608 ^b	1623 ^b
68	-3.74	13 ^b	51 ^b	259 ^b	497 ^b	3414 ^b
70	-1.59	18 ^b	747 ^b	2672 ^b	2125 ^b	2336 ^b
72	-1.29	34 ^b	460 ^b	1142 ^b	785 ^b	989 ^b
73	-2.46	4 ^b	323 ^b	1525 ^b	1157 ^b	1539 ^b
74	2.71	38 ^b	82 ^b	131 ^b	93 ^b	36 ^b
77	-2.33	0 ^b	11 ^b	86 ^b	69 ^b	39 ^b
79	0.19	71 ^b	28 ^b	192 ^b	45 ^b	9 ^b
80	0.92	2 ^a	0 ^a	5 ^a	2 ^a	0 ^a
82	-3.43	0 ^a	13 ^a	41 ^a	40 ^a	26 ^a
83	-0.24	0 ^b	44 ^b	279 ^b	281 ^b	287 ^b
85	-3.11	4 ^a	14 ^a	13 ^a	2 ^a	0 ^a
86	0.32	0 ^b	9 ^b	21 ^b	21 ^b	31 ^b
87	-1.41	722 ^b	3668 ^b	9009 ^b	6016 ^b	4129 ^b
88	-2.83	0 ^a	4 ^a	7 ^a	5 ^a	0 ^a
89	-4.02	751 ^b	1131 ^b	8538 ^b	5217 ^b	4646 ^b
90	-1.87	70 ^b	669 ^b	2862 ^b	1608 ^b	1623 ^b
91	-1.82	2172 ^b	3868 ^b	6867 ^b	5729 ^b	3962 ^b
92	-1.11	419 ^b	2668 ^b	5013 ^b	2638 ^b	3658 ^b
93	-1.19	580 ^b	3545 ^b	7431 ^b	4217 ^b	4459 ^b
95	-1.80	782 ^b	3513 ^b	8460 ^b	4921 ^b	3939 ^b

96	-1.63	38 ^b	1583 ^b	8973 ^b	4114 ^b	3345 ^b
97	0.31	12 ^b	10 ^b	455 ^b	515 ^b	853 ^b
98	-2.35	19 ^b	26 ^b	74 ^b	127 ^b	107 ^b
100	-2.81	497 ^b	6366 ^b	7247 ^b	6988 ^b	5230 ^b
101	-0.94	9 ^b	33 ^b	73 ^b	107 ^b	112 ^b
103	-0.43	26 ^b	54 ^b	259 ^b	206 ^b	174 ^b
105	-0.76	0 ^b	2 ^b	23 ^b	5 ^b	0 ^b
107	0.10	168 ^b	866 ^b	683 ^b	2912 ^b	999 ^b
108	-2.16	1550 ^b	4671 ^b	9327 ^b	4545 ^b	4018 ^b
109	-0.38	8 ^b	759 ^b	6239 ^b	4204 ^b	4140 ^b
110	-2.73	13 ^b	223 ^b	466 ^b	464 ^b	405 ^b
113	-0.40	0 ^b	99 ^b	81 ^b	32 ^b	0 ^b

^a Tributyrin used as substrate^b Vinyl decanoate used as substrate

Table 2c: Activity of selected lipases

SEQ ID NO:	T _m	Methanol Tolerance				
		0%	10%	20%	30%	40%
1	42.4	1.00	0.40	1.11	0.01	0.00
2	44.2	1.00	0.25	0.48	0.44	1.00
5	76.0	1.00	0.09	0.38	0.88	8.43
19	43.8	1.00	0.55	1.11	3.58	5.86
21	50.8	1.00	0.29	0.50	1.08	39.25
22	63.1	1.00	0.28	0.07	0.62	2.24
23	44.8	1.00	1.39	1.07	1.05	1.18
25	43.9	1.00	1.17	2.93	4.61	6.56
26	46.1	1.00	0.02	0.02	0.05	0.25
27	42.5	1.00	0.06	0.21	1.00	0.07
30	60.1	1.00	1.39	3.84	2.79	0.01
31	38.4	1.00	1.38	0.82	0.98	1.97
35	70.1	1.00	0.04	0.09	0.22	3.60
36	38.4	1.00	8.29	11.28	0.14	0.02
38	52.0	1.00	0.30	0.38	0.23	0.06
39	42.3	1.00	2.06	2.34	3.94	3.81
40	42.3	1.00	1.36	8.91	4.31	0.01
42	41.6	1.00	0.69	0.73	5.70	3.82
43	65.4	1.00	1.54	1.51	1.78	7.32
44	64.3	1.00	0.80	2.47	2.09	3.31
45	44.1	1.00	0.72	1.32	1.48	2.19
46	42.4	1.00	3.97	22.95	3.68	2.79
53	65.8	1.00	0.12	0.17	0.38	3.15
54	54.3	1.00	0.56	0.18	0.52	1.78

55	42.3	1.00	0.08	0.04	0.10	0.55
56	45.5	1.00	0.03	0.00	0.09	0.63
57	70.0	1.00	0.18	0.02	0.04	1.36
59	41.2	1.00	0.15	0.14	0.82	0.00
60	52.5	1.00	0.07	0.07	0.26	1.66
61	41.1	1.00	0.62	0.76	0.48	1.31
62	56.1	1.00	0.42	0.37	0.42	0.08
64	71.0	1.00	0.02	0.01	0.06	0.75
66	60.6	1.00	0.36	0.47	1.32	2.64
67	45.1	1.00	1.71	2.35	2.81	4.64
68	49.8	1.00	3.57	11.57	14.64	0.06
70	77.6	1.00	0.02	0.02	0.04	1.75
72	57.9	1.00	0.37	0.92	0.82	7.48
73	61.1	1.00	0.14	0.88	3.19	104.46
74	64.8	1.00	0.29	0.51	0.36	2.04
77	42.3	1.00	0.70	12.74	30.40	14.04
79	42.4	1.00	0.47	0.39	0.02	0.17
80	40.8	1.00	1.32	1.19	1.73	0.70
82	33.6	1.00	1.42	3.64	11.83	20.77
83	37.8	1.00	0.28	0.06	0.05	0.01
85	43.3	1.00	0.72	1.00	1.28	1.13
86	44.1	1.00	3.51	1.49	18.48	0.53
87	65.5	1.00	0.04	0.10	0.45	0.97
88	50.4	1.00	1.26	1.18	2.04	0.51
89	67.3	1.00	0.01	0.01	0.19	0.58
90	72.3	1.00	0.02	0.01	0.07	0.47
91	72.4	1.00	0.03	0.03	0.05	2.47
92	70.6	1.00	0.00	0.02	0.06	1.45
93	70.8	1.00	0.04	0.02	0.08	1.33
95	69.4	1.00	0.00	0.07	0.10	0.81
96	87.0	1.00	0.09	0.03	0.03	1.51
97	62.5	1.00	0.02	0.02	0.07	2.27
98	56.0	1.00	0.94	0.07	0.37	1.10
100	52.0	1.00	0.07	0.09	0.15	0.16
101	45.5	1.00	1.10	2.37	3.30	1.20
103	58.0	1.00	1.02	4.17	5.59	1.54
105	65.4	1.00	2.39	5.25	11.79	1.78
107	46.6	1.00	0.17	0.27	0.59	1.98
108	65.8	1.00	0.03	0.08	0.15	1.07
109	55.6	1.00	0.21	0.27	0.46	1.16
110	46.5	1.00	0.06	0.11	0.95	1.24
113	42.3	1.00	1.09	0.50	0.95	0.29

Whilst there has been described in the foregoing description preferred embodiments of the present invention, it will be understood by those skilled in the technology concerned that many variations or modifications in details of design or construction may be made without departing from the present invention.

CLAIMS

1. A method of producing a fatty acid ester, the method comprising
 - (a) providing a mixture comprising a triacylglyceride and/or a fatty acid, a short-chained alcohol, and an enzyme comprising a sequence selected from any one of
5 SEQ ID NO: 1 to 6, and 8 to 114; and
 - (b) incubating the mixture for a time and under conditions sufficient to allow for formation of the fatty acid ester.
2. The method according to claim 1, wherein the alcohol is selected from the group
10 consisting of methanol, ethanol, n-propanol, and i-propanol, preferably methanol.
3. The method according to any one of claims 1 or 2, wherein the alcohol is present from 10% v/v to 70% v/v.
- 15 4. The method according to claim 3, wherein the alcohol is present in at least 20% v/v, preferably at least 30% v/v, more preferably at least 40% v/v, even more preferably at least 50% v/v.
5. The method according to any one of claims 1 to 4, wherein the enzyme is selected
20 from the group consisting of: SEQ ID NO: 90, SEQ ID NO: 5, SEQ ID NO: 22, SEQ ID NO: 27, SEQ ID NO: 35, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 60, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID
25 NO: 100, SEQ ID NO: 107, SEQ ID NO: 108, and SEQ ID NO: 109.
6. The method according to any one of claims 1 to 5, wherein the triacylglyceride is present in the mixture and the triacylglyceride is a long chain triacylglyceride.
- 30 7. The method according to claim 6, wherein the triacylglyceride is triolein.

8. The method according to any one of claims 6 or 7, wherein the enzyme is selected from the group consisting of: SEQ ID NO: 90, SEQ ID NO: 53, SEQ ID NO: 27, SEQ ID NO: 55, SEQ ID NO: 60, SEQ ID NO: 73, and SEQ ID NO: 107.
- 5 9. The method according to any one of claims 1 to 5, wherein the fatty acid is present in the mixture and the fatty acid is a long chain fatty acid.
10. The method according to claim 9, wherein the fatty acid is oleic acid.
- 10 11. The method according to any one of claims 9 or 10, wherein the enzyme is selected from the group consisting of: SEQ ID NO: 90, SEQ ID NO: 57, SEQ ID NO: 74, SEQ ID NO: 92, SEQ ID NO: 93, and SEQ ID NO: 108.
12. The method according to any one of claims 1 to 5, wherein the triacylglyceride and/or fatty acid is palm oil fatty acid distillate.
- 15 13. The method according to any one of claims 1 to 12, wherein the enzyme comprises SEQ ID NO: 90.
- 20 14. The method according to any one of claims 1 to 13, wherein a yield of the fatty acid ester is at least 80%.
15. A method of hydrolyzing an ester, the method comprising providing a mixture comprising a triacylglyceride, water and an enzyme comprising a sequence selected from the group consisting of: SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID
- 25 30

NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 76, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 88, SEQ ID NO: 91, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 103, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110, and SEQ ID NO: 113.

16. A product obtained by the method according to any one of claims 1 to 15.
17. An isolated enzyme comprising a sequence selected from the group consisting of SEQ ID NO: 90, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 5, SEQ ID NO: 19, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 42, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 53, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56, SEQ ID NO: 57, SEQ ID NO: 59, SEQ ID NO: 60, SEQ ID NO: 61, SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 67, SEQ ID NO: 68, SEQ ID NO: 70, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 77, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 82, SEQ ID NO: 83, SEQ ID NO: 85, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 91, SEQ ID NO: 92, SEQ ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 96, SEQ ID NO: 97, SEQ ID NO: 98, SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 103, SEQ ID NO: 105, SEQ ID NO: 107, SEQ ID NO: 108, SEQ ID NO: 109, SEQ ID NO: 110 and SEQ ID NO: 113.
18. The enzyme according to claim 17, wherein the sequence is selected from the group consisting of SEQ ID NO: 30, SEQ ID NO: 39, SEQ ID NO: 40, SEQ ID NO: 46, SEQ ID NO: 85, and SEQ ID NO: 113.
19. The enzyme according to claim 17, wherein the sequence is selected from the group consisting of SEQ ID NO: 54, SEQ ID NO: 98, SEQ ID NO: 107, SEQ ID NO: 1,

SEQ ID NO: 2, SEQ ID NO: 25, SEQ ID NO: 59, SEQ ID NO: 68, SEQ ID NO: 73, SEQ ID NO: 83, SEQ ID NO: 86, SEQ ID NO: 97, and SEQ ID NO: 101.

20. The enzyme according to claim 17, wherein the sequence is selected from the group
5 consisting of SEQ ID NO: 96, SEQ ID NO: 70, SEQ ID NO: 5, SEQ ID NO: 35, SEQ
ID NO: 22, SEQ ID NO: 30, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 53, SEQ
ID NO: 57, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 73, SEQ ID NO: 74, SEQ
ID NO: 87, SEQ ID NO: 89, SEQ ID NO: 90, SEQ ID NO: 91, SEQ ID NO: 92, SEQ
ID NO: 93, SEQ ID NO: 95, SEQ ID NO: 97, SEQ ID NO: 105, and SEQ ID NO: 108.
10
21. The enzyme according to claim 17 comprising SEQ ID NO: 90.
22. A polynucleotide encoding the enzyme according to any one of claims 17 to 21.
- 15 23. A vector comprising the polynucleotide according to claim 22.
24. A host cell comprising the polynucleotide according to claim 22 or the vector
according to claim 23.