Saudi Journal of Biological Sciences 30 (2023) 103578

HOSTED BY

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Function of CAZymes encoded by highly abundant genes in rhizosphere microbiome of *Moringa oleifera*

Manal Tashkandi [a], Lina Baz [b,*]

[a] Department of Biochemistry, Faculty of Science, University of Jeddah, Jeddah, Saudi Arabia
[b] Department of Biochemistry, Faculty of Science, King AbdulAziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 24 October 2022
Revised 21 December 2022
Accepted 17 January 2023
Available online 1 February 2023

Keywords:
Morenga oleifera
CAZy
Glycoside hydrolases
Carbohydrates
Actinobacteria
Proteobacteria

ABSTRACT

Metagenomic analysis referring to CAZymes (Carbohydrate-Active enZymes) of CAZy classes encoded by the most abundant genes in rhizosphere versus bulk soil microbes of the wild plant *Moringa oleifera* was conducted. Results indicated that microbiome signatures and corresponding CAZy datasets differ between the two soil types. CAZy class glycoside hydrolases (GH) and its α-amylase family GH13 in rhizobiome were proven to be the most abundant among CAZy classes and families. The most abundant bacteria harboring these CAZymes include phylum Actinobacteria and its genus *Streptomyces* and phylum Proteobacteria and its genus *Microvirga*. These CAZymes participate in KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway "Starch and sucrose metabolism" and mainly use the "double displacement catalytic mechanism" in their reactions. We assume that microbiome of the wild plant *Moringa oleifera* is a good source of industrially important enzymes that act on starch hydrolysis and/or biosynthesis. In addition, metabolic engineering and integration of certain microbes of this microbiomes can also be used in improving growth of domestic plants and their ability to tolerate adverse environmental conditions.

1. Introduction

Moringa oleifera is a wild plant belonging to the family Moringaceae. This species is a large tree commonly known as the horseradish tree, which is native to several habitats including west region of Saudi Arabia (Al-Eisawi and Al-Ruzayza 2015). This wild plant has many health benefits (Selim et al., 2021) as it harbors several important bioactive (ex., minerals) and antioxidant (ex., beta-carotene) compounds (Gopalakrishnan et al., 2016). Edible seeds and/or leaves of *M. oleifera* may also be used in lowering human blood sugar and cholesterol levels and can also reduce inflammation related to chronic human diseases (Gopalakrishnan et al., 2016).

Utilization of next generation sequencing (NGS) in terms of metagenomic whole genome sequencing (mWGS) approach allows cataloging soil microbial genes that can be used in the detection of more accurate soil microbial composition and function (Vorholt et al., 2017). The soil microbes include bacteria, archaea, eukaryotic microorganisms and viruses (Odelade and Babalola 2019). mWGS also allows studying the environmental influence in shaping/reshaping microbiome signatures and the influence of plant root exudates in assemblage of beneficial microbial communities across different stages of plant development that most likely act in promoting plant growth (Raes et al., 2007). Based on the important medical characteristics of *M. oleifera*, the more the surrounding beneficial microbes that help the intact plant to grow better, the more the benefits this plant can gain.

Differential abundance of microbes in soil rhizosphere is mainly due to their differential response to varying chemical composition of the plant root exudates that seem to affect microbial growth dynamics, biomass, diversity, community assembly and metabolic potential (Pett-Ridge et al., 2021). It was recently reported that genomes of the microbial communities in rhizosphere region are highly abundant in genes encoding carbohydrate active enzymes (Levy et al., 2018). These enzymes are named Carbohydrate-Active enZymes (or CAZymes) as they act in building/degrading

* Corresponding author.
E-mail addresses: matashkandi@uj.edu.sa (M. Tashkandi), Lbaz@kau.edu.sa (L. Baz).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.sjbs.2023.103578
1319-562X/

soil carbohydrates (Lombard et al., 2014). CAZymes were assigned to CAZy classes (level 1) and families (level 2) and received enzyme classification (EC) codes (level EC), then, deposited in CAZy (https://www.cazy.org/) and CAZypedia (https://www.cazypedia.org/) databases. In addition, expression of rhizosphere bacterial genes at different environmental niches also changes in response to differences in physicochemical condition/type of the root rhizosphere (Nuccio et al., 2020). The high potential of rhizosphere genome translates into high rate of complex carbohydrate build-up/degradation that differ from one environmental niche to the other.

In the present study, we have searched CAZymes of rhizosphere soil microbes that are encoded by the most abundant microbial genes that allow cross-talking with roots of *M. oleifera* in order to reach insights as to how plant and microbes interact and mutually benefit from carbohydrates with different levels of complexities.

2. Materials and methods

2.1. Soil sample collection and DNA extraction

Soil samples were collected from coordinates 21°12′17.8″N 39°31′26.4″E of Mecca region of Saudi Arabia near the red sea coast as previously described (Al-Eisawi and Al-Ruzayza 2015). Collected bulk soil in three replicates was $\geq$ 10 m apart from collected rhizosphere soil. Trees selected for rhizosphere sample collection were single-grown and have similar sizes. Procedure and conditions of sample collection were previously described (Hurt et al., 2001) and concentration of isolated DNAs was adjusted to 10 ng/µl to meet whole genome sequencing requirements.

2.2. Whole genome shotgun sequencing and bioinformatics analysis

An amount of 30 µl of each DNA sample was shipped to Novogene Co. (Singapore) for whole metagenomic sequencing using Illumina HiSeq 2500 platform as described in the company report. Recovered sequences were deposited in the European Nucleotide Archive (ENA) (https://www.ebi.ac.uk/ena/browser/) under study no. ERP139990 and received accession nos. ERR10100770-74 and ERR10100781. Then, library preparation, dataset assembly and steps of quality control (QC) were performed as previously described (Karlsson et al., 2012, Mende et al., 2012, Oh et al., 2014). Novogene Co. approached to mix together unassembled clean reads of all samples that resemble low abundant genes to recover NOVO_MIX scaffolds. Scaffolds were used to generate scaftigs as described (Mende et al., 2012, Nielsen et al., 2014). Clean data of assembled ORFs and scaftigs were mapped using Soap 2.21, and predicted genes by MetaGeneMark (Nielsen et al., 2014) were dereplicated using Cluster Database at High Identity with Tolerance (CD-HIT) (Li and Godzik 2006, Fu et al., 2012). Then, we targeted non-redundant gene catalogues (nrGC) generated by greedy pairwise comparison (Li et al., 2014). These genes were annotated using MEGAN, then, functional abundance based on eggNOG database (version 4.0) was generated (Huson et al., 2011, Powell et al., 2014, Huson et al., 2016, Huerta-Cepas et al., 2017), and subsequent analyses including table clustering and PCoA were performed (Lozupone and Knight 2005, Lozupone et al., 2007, Lozupone et al., 2011). Deduced amino acid sequences of different annotated genes were mapped to the eggNOG database using Diamond (Buchfink et al., 2015) and integrated annotation information was subsequently utilized to map the recovered protein sequences against CAZy database (version 2014.11.25) (Lombard et al., 2014). CAZymes that have the highest gene abundance (a threshold of > 5000) were assigned to their families and manually drawn to their respective map pathway(s) of KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway database (https://www.genome.jp/kegg/pathway.html).

3. Validation of highly abundant genes encoding selected CAZymes via qPCR

Total RNAs of different samples were isolated using RNA PowerSoil® Total RNA isolation kit (Mo Bio, cat. no. 12866–25). First-strand cDNA was synthesized by mixing 2.5 ug RNA with 0.5 ug oligo (dT) primer, 4 uL first strand buffer (5x) and 1 uL Superscript II reverse transcriptase (200 U) (Invitrogen), then, volume reached 20 uL using sterile distilled water. Primers of five randomly-selected highly abundant genes encoding CAZymes of family GH13 were designed using Netprimer software (https://www.premierbiosoft.com/netprimer/index.html) (Table S1). Gene expression levels were quantified by qPCR with Maxima™ SYBR Green/ROX using Agilent Mx3000P System (Agilent technology, USA) as described (Bahieldin et al., 2015). Then, 16S rRNA gene of *Bacillus subtilis* (accession no. AB042061) was used in the reactions as the house-keeping gene.

4. Results

4.1. Fidelity testing of CAZy datasets

Fidelity of CAZy datasets encoded by non-redundant genes was tested by mapping the distance among samples within each of the two soil types surrounding *M. oleifera* at CAZy database levels 1, 2 and EC via principle coordinate analysis (PCoA) (Figure S1). The results indicated complete separation between microbiomes of the two soil types, where bulk soil microbiome samples were located at the negative side of the PCoA 1 (or PC1), while those of rhizobiome samples were located in the positive side of PC1 in terms of CAZy levels 1 and 2 (Fig. S1a-b). In terms of CAZy level EC, opposite results were reached where samples of rhizosphere soil were located at the negative side and *vice versa* for bulk soil samples (Fig. S1c). Thus, microbiome signatures and corresponding CAZy datasets differ in the two soil types.

4.2. Description of assembled sequences

ORFs/scaftigs described in Table S2 are the results of assembling raw reads of different samples. The table describes the queries after being aligned with analogue sequences in the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov/), namely subject ID, that shows match and mismatch in annotated ORFs and gap sizes (nt) and nucleotides assigning gene start and end points, if available. Numbers of 20,981 ORFs and 29,133 scaftigs were recovered from the two methods of gene assembly (Table S2). These numbers refer to aligned sequences with lengths between 30 and 1541 nt and identity of $\geq$ 50 %. Two hits in the NCBI were detected for a number of 2960 ORFs/scaftigs to reach the total of 53,076 assembled sequences to be analyzed further.

4.3. Differential abundance of genes encoding CAZymes

The results indicate that the total number of genes encoding the six different classes (or CAZy level 1) is 53,076 across the two soil types, where class glycoside hydrolases (GH) was the highest (23923 genes, ~45.0 %) followed by those of classes glycosyl transferases (GT) (16153 genes, ~30.0 %), carbohydrate-binding modules (CBM) (7159 genes, ~13.5 %), carbohydrate esterases (CE) (2925 genes, ~5.5 %), auxiliary activities (AA) (2194 genes, ~4.1 %), and polysaccharide lyases (PL) (722 genes, ~1.4 %) (Table S3 and Fig. 1a). Percentage of genes encoding CAZy classes

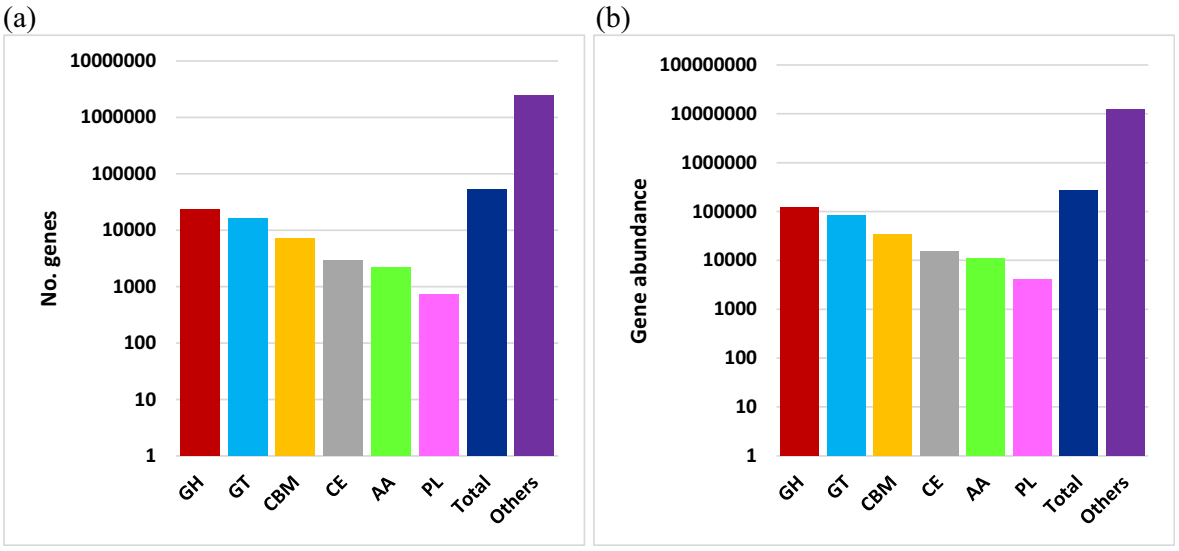

Fig. 1. Number (a) and abundance (b) of genes encoding different CAZy classes (e.g., AA, CBM, CE, GH, GT and PL) of CAZy database at level 1 across microbiomes of rhizosphere and surrounding bulk soils of *M. oleifera*.

out of the total assembled non-redundant genes in the microbiome was estimated to be ~ 2.2 % (Table S3).

4.3.1. CAZy database level 1

The results of gene abundance at CAZy database level 1 across soil types of *M. oleifera* perfectly align with those of number of genes, where class GH showed the highest abundance level followed by classes GT, CBM, CE, AA, then, PL (Table S4 and Fig. 1b). In other words, the higher the number of CAZyme encoding genes, the higher the abundance of these genes. In terms of gene abundance across CAZy classes among microbiome samples, the results indicated lower level in bulk soil microbiomes (227593 in average) compared with that in rhizosphere microbiome (305711 in average) (Table S4 and Fig. 2a). Relative gene abundance at CAZy level 1 within CAZy classes indicated no discrete differences between the two soil types except for classes GH, GT and PL, where gene abundance for the first and third classes was higher in the rhizobiome than that of bulk soil, while that of class GT showed opposite results (Table S4 and Fig. 2b). Therefore, we thought that class GH might require higher attention as number and abundance of encoding genes in rhizobiome are higher than those of other classes.

4.3.2. CAZy database level 2

Description of families encoding different classes of CAZy (e.g., AA, CBM, CE, GH, GT and PL) at level 2 across microbiomes of rhizosphere (R) and surrounding bulk soil (S) of *M. oleifera* along with gene (query) and subject IDs are shown in Table S5. The table also indicates the CAZyme members of the different families along with their EC numbers. Table S6 refers to abundance of CAZy families in microbiomes samples of rhizosphere (R) and surrounding bulk soil (S) of *M. oleifera*. Of which, we have selected the top 10 families in terms of gene abundance within CAZy classes AA, CBM, CE, GH, GT and PL of CAZy database (Fig. 3). Most of the families of different classes were highly abundant in rhizobiome, except for AA3 and

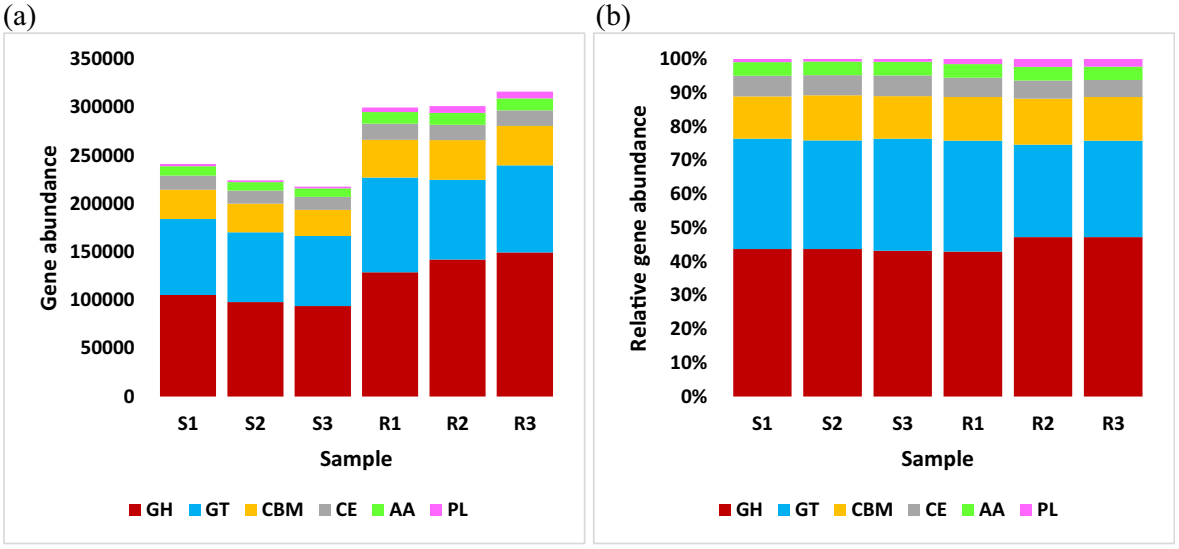

Fig. 2. Abundance (a) and relative abundance (b) of genes encoding different CAZy classes (e.g., AA, CBM, CE, GH, GT and PL) of CAZy database at level 1 in microbiome samples of rhizosphere (R) and surrounding bulk (S) soils of *M. oleifera*.

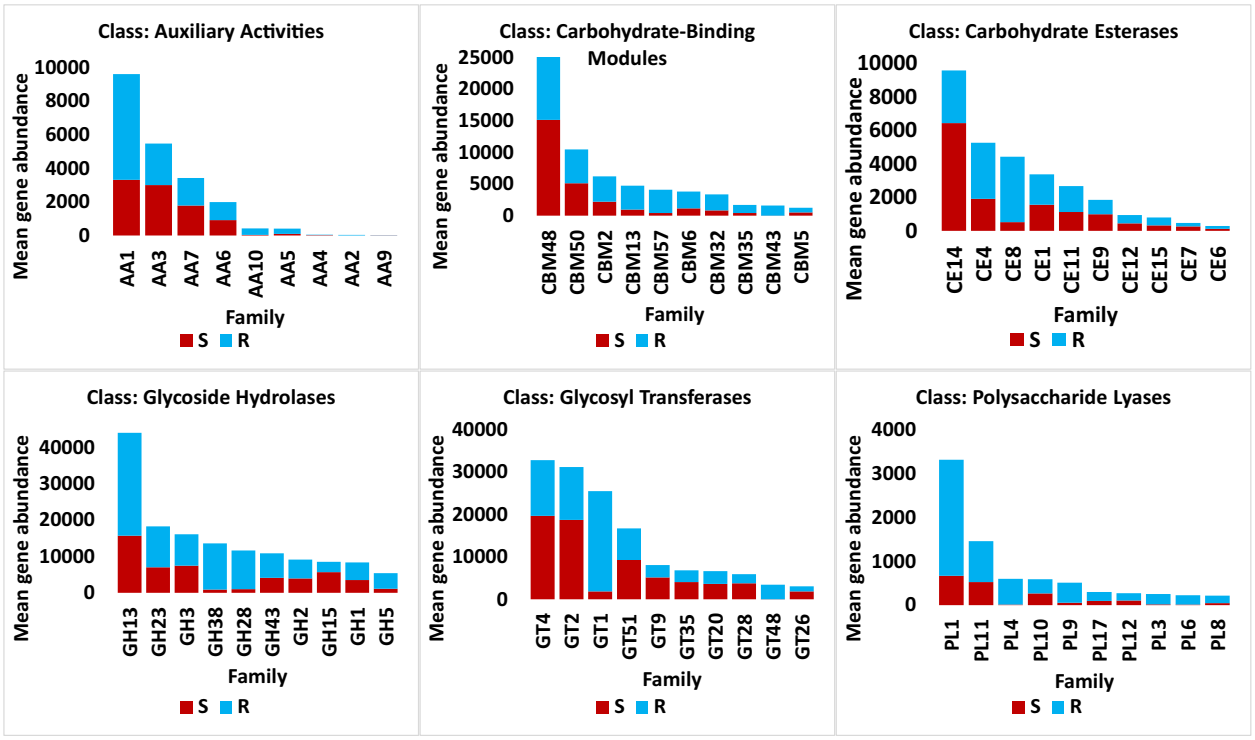

Fig. 3. The top 10 CAZy families in terms of gene abundance within each CAZy class (e.g., AA, CBM, CE, GH, GT or PL) of CAZy database at level 2 in microbiomes of rhizosphere (R) and surrounding bulk (S) soils of *M. oleifera*.

AA7 of class AA, CE14 of class CE, GH15 of class GH, and most GT families, e.g., GT4, GT2, GT51, GT9, GT35, GT 20 and GT28.

4.3.3. CAZy database level EC

In terms of gene abundance of CAZy database at level EC, the results perfectly align with those at level 2, where microbiome samples of the bulk soil showed lower level of gene abundance at CAZy database level EC compared with those of the rhizosphere (Figure S2). Table S7 indicates the microbes harboring the genes encoding CAZymes of different CAZy families of classes AA, CBM, CE, GH, GT and PL at CAZy database level EC along with encoding gene IDs (queries) in microbiomes of rhizosphere (R) and surrounding bulk soil (S) of *M. oleifera*. Abundances and IDs of these genes along with their encoded CAZymes are shown in Table S8. Expectedly based on data of Fig. 3, the highest gene abundance (>5000) across microbiomes of rhizosphere and surrounding bulk soils of *M. oleifera* was all scored for CAZy family GH13 with 26 CAZymes (Table S7). The latter CAZymes are partially shared with other GH families (Fig. 4). We have placed these CAZymes of CAZy family GH13 in five groups of which the first included family GH15 in addition to family GH13 (6217 genes), while the second included families GH4, GH31, GH97 and GH63 (6130 genes), the third included families GH57 and GH126 (5335 genes), the forth included GH133 (5306 genes), and the fifth included no CAZy families other than GH13 (5179 genes) (Fig. 4).

Referring to the bacteria harboring the 26 CAZymes, the calculated results shown in Table S7 and described in Fig. 5a indicate that phylum Actinobacteria dominates in terms of gene abundance (125920) across different CAZy classes followed by Proteobacteria (81180), Bacteroidetes (22605), Verrucomicrobia (19529), Gemmatimonadetes (6241) and Firmicutes (1610). Gene abundance of these phyla represent ~ 60 % of the abundance of all genes across microbiomes of rhizosphere and surrounding bulk soils of *M. oleifera*. Note that a large proportion of the left 40 % abundance refers to uncharacterized/unknown/unclassified microbes (Table S7).

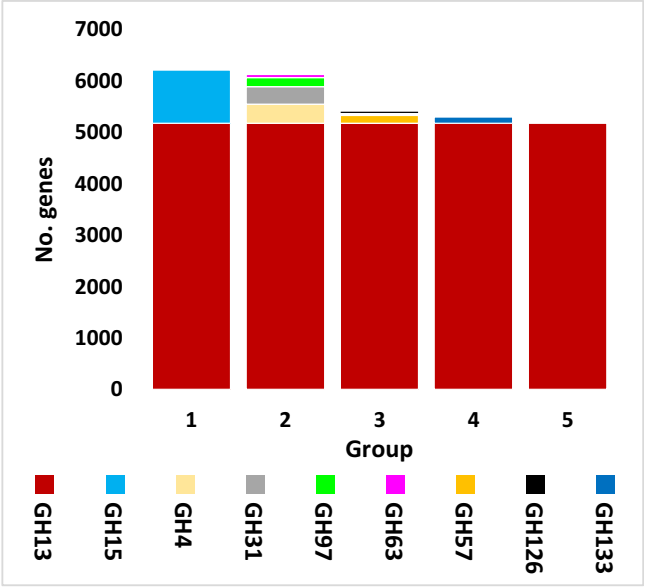

Fig. 4. Glycoside hydrolases (GHs) of CAZy family GH13 (26) with the highest gene abundance (>5000) that is partially shared with other GH families in four combinations across microbiomes of rhizosphere and surrounding bulk soils of M. oleifera. Group 1 = GH13/GH15, Group 2 = GH13/GH4/GH31/GH97/GH63, Group 3 = GH13/GH57/GH126, Group 4 = GH13/GH133, Group 5 = GH13 (for more details see Tables S7 and S8).

Phyla Actinobacteria and Proteobacteria seem to harbor all GH families containing the 26 CAZymes except for GH133 and GH126 for the first phylum, while GH126, only, for the second (Table 1). Phylum Firmicutes solely harbors GH126. At the genus level, genera *Streptomyces* and *Microvirga* harbor six (e.g., GH13, GH15, GH31, GH4, GH63, GH97) and four (e.g., GH13, GH15,

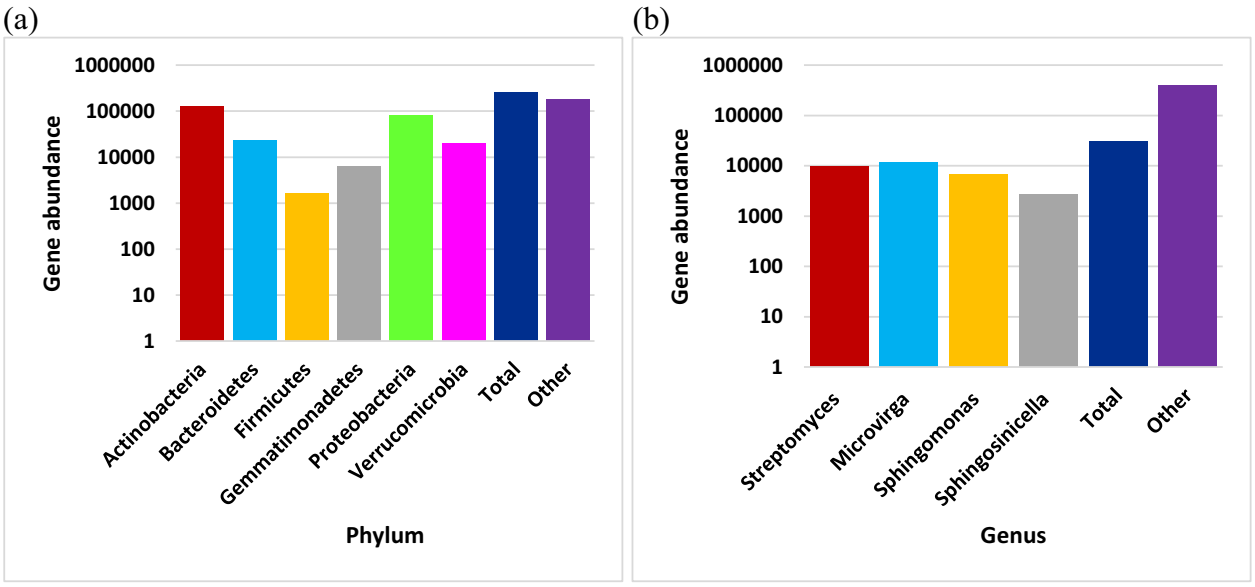

Fig. 5. Gene abundance of bacterial phyla harboring the 26 CAZy families (a) and of genera with the most abundant genes (b) across microbiomes of rhizosphere and surrounding bulk soils of M. oleifera. The term "Other" refers to other bacterial phyla or unknown/unclassified microbes.

Table 1
Bacterial phyla and their genera that harbor glycoside hydrolases (GHs) of CAZy family GH13 (26) with the highest gene abundance (>5000) that is partially shared with other GH families across microbiomes of rhizosphere and surrounding bulk soils of M. oleifera.

Phylum	Genus	CAZy family	Phylum	Genus	CAZy family
Actinobacteria	*Mycobacterium*	GH13, GH63	Actinobacteria	*Streptomyces*	GH13, GH15, GH31, GH4, GH63, GH97
	Mycolicibacterium	GH15	(Continued)	*Streptosporangium*	GH15
	Frankia	GH13		*Solirubrobacter*	GH57
	Blastococcus	GH13, GH15		*Thermoleophilum*	GH31
	Geodermatophilus	GH13, GH15	Bacteroidetes	*Niastella*	GH31, GH97
	Modestobacter	GH13		*Pontibacter*	GH97
	Kineococcus	GH13	Firmicutes	*Class: Bacilli*	GH126
	Quadrisphaera	GH13	Gemmatimonadetes	*Gemmatirosa*	GH13
	Cellulomonas	GH13, GH15	Proteobacteria	*Microvirga*	GH13, GH15, GH31, GH63
	Ornithinimicrobium	GH13		*Rubellimicrobium*	GH4
	Phycicoccus	GH31		*Belnapia*	GH13
	Diaminobutyricimonas	GH13		*Azospirillum*	GH133
	Arthrobacter	GH13, GH15, GH4		*Inquilinus*	GH4
	Kocuria	GH13, GH15		*Sphingomonas*	GH13, GH15, GH97
	Isoptericola	GH31		*Sphingosinicella*	GH15, GH31, GH97
	Kribbella	GH31		*Ramlibacter*	GH15
	Marmoricola	GH13		*Massilia*	GH31, GH97
	Nocardioides	GH13, GH15		*Methylobacillus*	GH57
	Amycolatopsis	GH57		*Myxococcus*	GH4
	Pseudonocardia	GH57	Verrucomicrobia	*Roseimicrobium*	GH133

GH31, GH63) GH families, respectively, out of the nine CAZy families containing the 26 CAZymes, while genera *Sphingomonas* and *Sphingosinicella* harbor only three GH families, e.g., GH13, GH15, GH97 for the first genus, while GH15, GH31, GH97 for the second (Table 1). In terms of gene abundance, genus *Microvirga* (11485) dominates followed by *Streptomyces* (9492), *Sphingomonas* (6840) and *Sphingosinicella* (2741) (Table S7 and Fig. 5b).

Gene abundance of the selected 26 CAZymes in microbiome samples of rhizosphere and surrounding bulk soils of *M. oleifera* shown in Table S9 and Fig. 6 also align with that of CAZy database level EC (Figure S2), where microbiomes of rhizosphere soil (R) were more abundant in genes encoding these CAZymes than those of bulk soil (S). EC hierarchy of these selected CAZymes that was retrieved from the Comprehensive Enzyme Information System of BRENDA Enzyme Database and BRITE of KEGG database across microbiomes of rhizosphere and surrounding bulk soils of *M. oleifera* is shown in Table 2. EC classes of these CAZymes involved "Transferases" (EC 2) with seven CAZymes, "Hydrolases" (EC 3) with as high as 16 CAZymes and "Isomerases" (EC 5) with three CAZymes. "Transferases" of these highly abundant GH CAZymes included EC subclass "Glycosyltransferases" (EC 2.4) with seven CAZymes that, in turn, included EC sub-subclasses "Hexosyltransferases" (EC 2.4.1) with six CAZymes and "Transferring glycosyl groups other than hexosyl and pentosyl" (EC 2.4.99) with one CAZyme. EC class "Hydrolases" included one EC subclass namely "Glycosylases" (EC 3.2), which included one EC sub-subclass namely "Glycosidases", i.e. enzymes that hydrolyze glycosyl compounds (EC 3.2.1). In terms of EC class "Isomerases", one EC subclass namely "Intramolecular transferases" (EC 5.4) and one EC sub-subclass namely "Transferring groups other than acyl, amino, hydroxyl and phosphorus" (EC 5.4.99) are included (Table 2).

CAZyme numbers in the five groups are 1, 1, 4, 1 and 19, respectively (Table 2). Interestingly, 16 of these CAZymes exist mainly in KEGG pathway "Starch and sucrose metabolism" (map00500) (Fig-

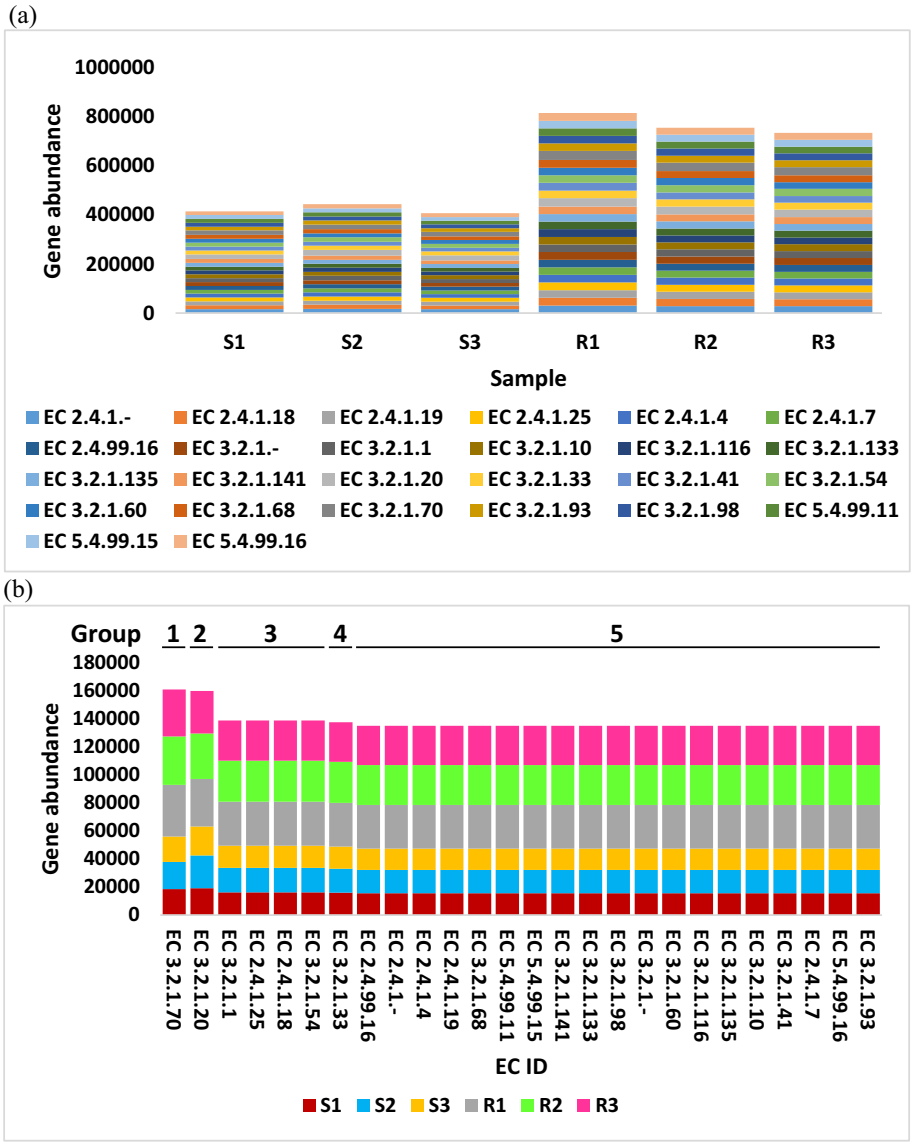

Fig. 6. CAZymes of glycoside hydrolases (GHs) of CAZy family GH13 (26) with the highest gene abundance (>5000) (a and b) that are partially shared with other GH families in four combinations (b) in microbiome samples of rhizosphere (R) and surrounding bulk (S) soils of M. oleifera. Group 1 = GH13/GH15, Group 2 = GH13/GH4/GH31/GH97/GH63, Group 3 = GH13/GH57/GH126, Group 4 = GH13/GH133, Group 5 = GH13 (for more details see Tables S8 and S9).

ure S3) with some of them exist in KEGG pathways "Metabolic pathways" and "Biosynthesis of secondary metabolites" (https://www.genome.jp/kegg/pathway.html). CAZymes with comparable functions to CAZymes with ECs 2.4.1.-, 3.2.1.-, 3.2.1.116, 3.2.1.135, 3.2.1.33, 3.2.1.60, 3.2.1.70 and 3.2.1.98 include CAZymes with ECs 2.4.1.7, 3.2.1.1, 3.2.1.98, 3.2.1.41, 3.2.1.116, 3.2.1.10 and 3.2.1.1/3.2.1.116, respectively (Table 2).

4.4. Validation of CAZy in silico datasets

CAZy datasets of microbiomes of the two soil types was done at the level of real time PCR (Figure S4). Reactions included five genes encoding highly enriched CAZymes with ECs of 2.4.1.18, 2.4.1.25, 2.4.99.16, 3.2.1.1 and 3.2.1.20. Chromosome accession no. of genes encoding these CAZymes is CP050522.1 referring to genome of *Streptomyces coelicolor* strain M1154, except for CAZyme with EC 3.2.1.1 whose accession no. of the encoding gene is M18244.1 of *Streptomyces limosus*. All studied CAZymes are members of CAZy family GH13 in addition to other families, e.g., GH57, GH126, GH4, GH31, GH97 and GH63 (Table 2). The results of real time PCR for selected *meta*-transcriptomic data in Figure S4 align with those for *in silico* genomic data in Fig. 6.

5. Discussion

Most of the prior plant metagenomics studies focus on crop plants although the domestication process during growth acts in reshaping structure of plant's native root rhizobiome. The latter possibly results in losing microbial taxa with crucial roles, e.g., nutrient acquisition, plant growth promotion, plant disease protection, etc. (Pérez-Jaramillo et al., 2017). Wild plants, on the other hand, grow in their native environment and have their natural root rhizobiome structures sustained (Bulgarelli et al., 2015). Besides, native rhizobiome allows researchers to study the succession of microbial community assembly and dynamics of microbial evolution in a given ecosystem (Pett-Ridge and Firestone 2017, Pett-Ridge et al., 2021). Therefore, we thought to study microbiomes of the wild plant *M. oleifera* growing naturally in the wild to know the native structure of these microbiomes and their natural interaction with plant root.

Table 2
EC hierarchy of glycoside hydrolases (GHs) retrieved from the Comprehensive Enzyme Information System of BRENDA Enzyme Database and BRITE of KEGG (Kyoto Encyclopedia of Genes and Genomes) database referring to CAZy family GH13 (26) with the highest gene abundance (>5000) that is partially shared with other GH families across microbiomes of rhizosphere and surrounding bulk soils of M. oleifera.

Class	Subclass	Sub-subclass	CAZyme EC	CAZyme name	GH family	Group
2 Transferases	2.4 Glycosyltransferases	2.4.1 Hexosyltransferases	EC 2.4.1.-[a]	6?-P-sucrose phosphorylase	GH13	5[j]
			EC 2.4.1.18*	1,4-alpha-glucan branching enzyme	GH13/GH57/GH126	3
			EC 2.4.1.19*	cyclomaltodextrin glucanotransferase	GH13	5
			EC 2.4.1.25*	4-alpha-glucanotransferase	GH13/GH57/GH126	3
			EC 2.4.1.4*	amylosucrase	GH13	5
			EC 2.4.1.7*	sucrose phosphorylase	GH13	5
		2.4.99 Transferring glycosyl groups other than hexosyl and pentosyl	EC 2.4.99.16*	alpha-1,4-glucan: phosphate, alpha-maltosyltransferase	GH13	5
3 Hydrolases	3.2 Glycosylases	3.2.1 Glycosidases, i.e. enzymes that hydrolyse O- and S-glycosyl compounds	EC 3.2.1.-[b]	maltopentaose-forming, alpha-amylase	GH13	5
			EC 3.2.1.1*	alpha-amylase	GH13/GH57/GH126	3
			EC 3.2.1.10*	oligo-alpha-glucosidase	GH13	5
			EC 3.2.1.116[c]	maltotriose-forming alpha-amylase	GH13	5
			EC 3.2.1.133*	maltogenic amylase	GH13	5
			EC 3.2.1.135[d]	neopullulanase	GH13	5
			EC 3.2.1.141*	malto-oligosyltrehalose	GH13	5
			EC 3.2.1.20*	trehalohydrolase	GH13/GH4/GH31/GH97/GH63	2 [k]
			EC 3.2.1.33[e]	alpha-glucosidase	GH13/GH133	4 [k]
			EC 3.2.1.41[f]	amylo-alpha-1,6-glucosidase	GH13	5
			EC 3.2.1.54*	pullulanase	GH13/GH57/GH126	3
			EC 3.2.1.60 [g]	cyclomaltodextrinase	GH13	5
			EC 3.2.1.68*	maltotetraose-forming alpha-amylase	GH13	5
			EC 3.2.1.70 [h]	isoamylase	GH13/GH15	1 [k]
			EC 3.2.1.93*	glucan 1,6-α-glucosidase	GH13	5
			EC 3.2.1.98 [g]	trehalose-6-phosphate hydrolase maltohexaose-forming alpha-amylase	GH13	5
5 Isomerases	5.4 Intramolecular transferases	5.4.99 Transferring groups other than acyl, amino, hydroxyl and phosphorus	EC 5.4.99.11[i]	isomaltulose synthase	GH13	5
			EC 5.4.99.15*	malto-oligosyltrehalose synthase	GH13	5
			EC 5.4.99.16*	trehalose synthase	GH13	5

KEGG pathway of most CAZymes is "Starch and sucrose metabolism" (ec00500) followed by "Metabolic pathways" (ec01100) and "Biosynthesis of secondary metabolites" (ec01110).
Total number of CAZymes of KEGG pathway "Starch and sucrose metabolism" is 76.

[a] *Cf. EC 2.4.1.7* [g]*Cf. EC 3.2.1.1 and* EC 3.2.1.116.
[b] Cf. EC 3.2.1.1 [h]Harobrs *similar activity as EC 3.2.1.10.*
[c] *Cf. EC 3.2.1.98* [i]No *hits or analogues in KEGG database.*
[d] Cf. EC 3.2.1.41 [j]*Group of one GH family (*e.g., *group 5).*
[e] *Acts on glycogen degradation, glycogen => glucose-6P and* Cf. *EC 2.4.1.25* [k]*Groups of one CAZyme each (*e.g., *groups 1, 2 and 4).*
[f] *Acts on glycogen degradation, glycogen => glucose-6P and modified in 2000 to 3.2.1.69 *CAZymes participating in pathway "Starch and sucrose metaolis.*

CAZy is a database that provides genomic, structural and biochemical information of Carbohydrate-Active enZymes (or CAZymes) that are separated into six classes and designated as (Lombard et al., 2014). Each of the CAZy classes was subclassified into subclasses and families with a range of discrete modules mainly based on amino acid sequence similarity, of which function/specificity characteristics of hundreds of these CAZymes remain to be deciphered. CAZymes are members of CAZy families that act on building or breaking down complex carbohydrates and/or glycoconjugates (Cantarel et al., 2009). The number of CAZyme protein families exceeds 300 separated in classes based on enzyme activities. These classes include glycoside hydrolases (GHs) that acts on hydrolysis and/or rearrangement of glycosidic bonds as exemplified in the KEGG pathway "Starch and sucrose metabolism" (Lombard et al., 2014). Such an updated CAZy database adds to our understanding of the nature, intensity and orientation of the breakdown/biosynthesis shuttle of complex carbohydrate in soil microbes of which decomposed organic carbon is the main source (Haiming et al., 2020).

In our bioinformatics analysis, the number of annotated non-redundant genes (or gene IDs) that encode CAZymes reflects the size of CAZy genome in a given microbiome, while gene abundance in CAZy database not only considers the number of genes but also the number of CAZy family members (or CAZymes) encoded by these genes (in the form of ORFs or scaftigs) that are eventually given query IDs in the microbiome. For example, the gene ID NOVO_MIX_2771553 in the present datasets comprises an ORF assembled from our *in silico* analysis. This gene or query ID has a certain identity percentage with an analogue gene in the NCBI that has a subject ID of WP_013705405.1 referring to CAZy family GH13. The latter family is made of 26 CAZymes (or family members) with different ECs. Thus, this gene ID solely refers to an abundance of 26 as long as it is assigned to family GH13 (Table S5). Then, genes referring to a given CAZy family will be multiplied by the number of family members to detect total gene abundance of a given microbiome sample. A given gene ID always refers only to one subject ID, while subject ID (ex., A5JTQ3.1, Table S5) of family GH3 can refer to a number of gene (ORF or scaftig) IDs (e.g., R3_258685, R3_135522, R3_132334, R1_201450, R1_144966, NOVO_MIX_2505990 and NOVO_MIX_172975, Table S5) with different percentages of sequence identity. Note that each CAZyme (ex., alpha-amylase with EC 3.2.1.1, Table 2) can have hits with more than one CAZy family (e.g., GH13/GH57/GH126, Table 2) of a given CAZy class (e.g., GH) as there are some similarities between different CAZy families of a given CAZy class in terms of structure, function and possibly the response of the encoding gene(s) to a given stimuli. As gene abundance is based on family members or CAZymes of a given microbiome sample, then, it can be described as functional abundance of the microbiome.

In the present study, the highest gene abundance (>5000) across microbiomes of rhizosphere and surrounding bulk soils of *M. oleifera* was scored for CAZy family GH13 with 26 CAZymes that are partially shared with other GH families (Table 2 and S7 and Fig. 4). Family GH13 is the largest family of glycoside hydrolases (MacGregor 1988). This family harbors majority of enzymes acting on starch metabolism (Stam et al., 2006). Some members of this family bear starch-binding modules (Janeček 1997). All members of this family received ECs of 3.2.1.X (carbohydrate-hydrolyzing enzymes), which is the most abundant (Stam et al., 2006), 2.4.1.X (carbohydrate-transferring enzymes) and 5.4.99.X (isomerases). Several members of these three enzyme categories adopt the "double displacement catalytic mechanism" in their reactions. These reactions proceed through both building and subsequent breaking down of a glycosyl-enzyme intermediate (Davies and Wilson 1999). A perfect example is the CAZyme 4-alpha-glucanotransferase (EC 2.4.1.25) or 4alphaGTase that seems to be involved in two opposite reactions during the transfer of 4-alpha-glucan (Fig. 7). These two reactions results in rearrangement of glucose units of maltodextrins (e.g., maltose) to yield a glucose unit and eventually a longer chain, e.g., amylose (Figure S5). The enzyme can also elongate short maltooligosaccharides (e.g., maltotriose) by linking two units of maltooligosaccharide via a 1,4-alpha-glucosidic bond (Yoon et al., 2017). This eventually results in the biosynthesis of amylose; a polysaccharide chains of α-1,4-linked D-glucose, with the disaccharide maltose used as the substrate (reaction 1) (orange arrow, Figure S3). In addition, 4alphaGTase is also involved in the production of glucose units as a breakdown reaction of maltose (reaction 2) (red arrow, Figure S3). Information in BRENDA database indicates that the two reactions of this mechanism occur concurrently, where CAZyme acts on transferring a segment of α-1,4-D-glucan of maltotriose (3 glucose units) to a new position in one of two acceptors, e.g., glucose or (1 → 4)-alpha-D-glucan. For more explanation of the mechanism, the site indicates that the enzyme can catalyze the reaction of maltononaose ($C_{54}H_{92}O_{46}$) and multotriose (trisaccharide) to produce maltoundecaose ($C_{66}H_{112}O_{56}$) (build-up reaction) and D-glucose (breakdown reaction). The mechanism in KEGG database indicates that the enzyme can also utilize maltose instead of mutotriose to produce amylose (see Figure S3). There is a possibility that abundance of the two final products (e.g., amylose and glucose) in the two reactions might differ as a result of the failure to conduct the build-up reaction. This speculation is based on one or more of the following reasons: (1) differential abundance of specific root exudates or signals favoring either reaction; (2) differences in environmental condition or pathogenesis; (3) the enzyme approaches conformational change post-translationally to block the build-up reaction; and/or (4) the occurrence of feedback inhibition of build-up reaction when enough amount of amylose was generated. Up to our knowledge, there is no prior research searched these possibilities. Thus, further experimentation is required to support or decline one or more of these claims

It was previously suggested that roots partially stimulate and/or inhibit growth of some specific rhizosphere bacterial and archaeal populations (Pett-Ridge et al., 2021). For example, relative abundance of several genera of the fast-growing bacterial phylum Proteobacteria, including genus *Microvirga* (Madigan et al., 2010), increase in the rhizosphere zone, while other genera of phyla Actinobacteria and Acidobacteria decrease (DeAngelis et al., 2009). However, other genera of phylum Actinobacteria, including genus *Streptomyces*, displayed high abundance in the rhizosphere zone (Nuccio et al., 2016, Shi et al., 2016). These data are in accordance with ours in terms of the high enrichment of GH CAZymes of KEGG pathway "Starch and sucrose metabolism" in genera *Microvirga* and *Streptomyces* (Table 1 and Figure S3). A recent report has shown that genus *Microvirga* harbors > 100 genes encoding enzymes that act in degrading and/or creating glycosidic bonds (Jiménez-Gómez et al., 2019). These activities are similar to those of family GH13 CAZymes in the present study that are involved in KEGG pathway "Starch and sucrose metabolism". Interestingly, Jiménez-Gómez and colleagues (Jiménez-Gómez et al., 2019) indicated that none of the detected genes encodes any members of family GH13 in disagreement with our results for the same microbe (Table 1). However, when we consider the factor of plant exudates, we might explain this contradicting results by the differential response of the genus due to the exposure to different exudates of two different plants, e.g., *Brassica napus* in prior research (Jiménez-Gómez et al., 2019), while *Moringa oleifera* in the present study. Another recent report indicated that a member of genus *Streptomyces* harbors 186 CAZyme domain sequences, where families of the CAZy class GH are the most abundant followed by those of GT (Králová et al., 2021). This makes the rhizosphere microbiome of *Moringa oleifera* a good source of these economically-important enzymes.

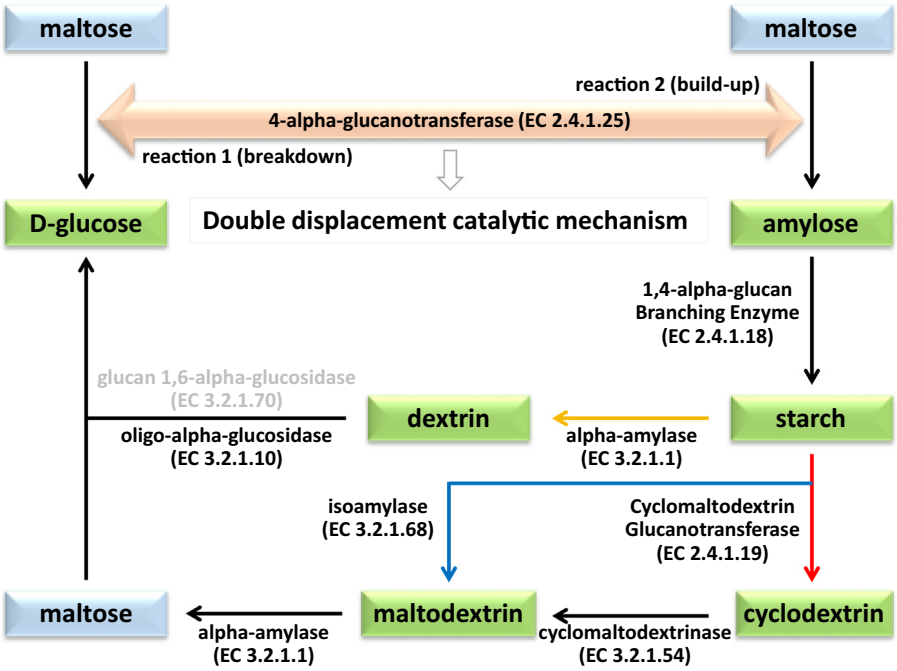

Fig. 7. CAZymes participating in starch breakdown and biosynthesis in the KEGG pathway "Starch and sucrose metabolism" that are highly enriched in rhizosphere soil of M. oleifera compared with those in the bulk soil. CAZyme with EC 2.4.1.25 conducts the "double displacement catalytic mechanism" by hydrolyzing maltose into two D-glucose units (breaking down) of which one unit is incorporated in biosynthesis of amylose (building up) by a 1,4-alpha-glucosidic bond. Four routes are shown for breaking down starch into D-glucose of which three routes involve enriched CAZymes in rhizosphere soil of M. oleifera (see Figure S3). Orange, blue and red arrows refer to the three routes. CAZyme glucan 1,6-alpha-glucosidase (EC 3.2.1.70) does not exist in the pathway but has similar activity to oligo-alpha-glucosidase (EC 3.2.1.10).

The KEGG pathway "Starch and sucrose metabolism" (KEGG level 3) belongs to "Carbohydrate metabolism" domain (KEGG level 2) of the KEGG category "Metabolism" (KEGG level 1). We speculate that the two major routes in this pathway in soil microbes surrounding rhizosphere of *M. oleifera* include starch biosynthesis as route 1 and starch hydrolysis as route 2. There is a chance that plant-stored starch can be provided exogenously via exudation (Pett-Ridge et al., 2021) in order for soil microbes to produce simpler carbohydrates required by the plant and surrounding rhizosphere microbiome (Nuccio et al., 2020). It seem that the major players in this pathway among the abundant CAZymes in the present study in the rhizosphere region strongly promote the intermediate steps towards starch biosynthesis. These CAZymes involve 4-alpha-glucanotransferase (EC 2.4.1.25) of GH group 3 and alpha-1,4-glucan:phosphate (or alpha-maltosyltransferase) (EC 2.4.99.16) of GH group 5 (see Table 2 and Fig. 4). The two CAZymes act, respectively, on converting maltose and α-maltose-1P to amylose; the unique substrate for starch biosynthesis (Figure S3). One more player in starch biosynthesis in the present study is amylosucrase (EC 2.4.1.4) that promotes biosynthesis of the polysaccharide amylose with the disaccharide sucrose as the substrate. Action of the latter enzyme is dependent on that of sucrose phosphorylase (EC 2.4.1.7) that makes biosynthesis of this sugar indirectly an important step towards starch biosynthesis. The other major player in this pathway, e.g., 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) of CAZyme group 3 (Fig. 4), acts on the final step of starch biosynthesis (Figure S3). This CAZyme promotes the conversion of amylose to starch. Soil complex carbohydrates, like starch, were reported to act for intra- and intercellular recognition and to significantly promote several biological functions (Lombard et al., 2014). Such complex compounds are considered as the most abundant with structurally the most diverse chemical structures in nature (Hart and Copeland 2010). Accordingly, number of enzymes acting in biodegrading soil complex compounds should be enormous, thus, deciphering function/activity and substrate specificity of these biodegrading enzymes require further attention.

In the present study, we also speculate that plant roots release specific exudates towards storing starch in the rhizobiome as an extra sink of carbohydrates required by the plant, while plant can release other specific exudates to promote hydrolysis of starch in the sink when the plant required simpler carbohydrates or sugars. The latter type of exudate pattern is much more dynamic in rhizosphere of *M. oleifera* where it promotes several steps in the pathway as it induces higher abundance of microbes harboring genes encoding certain starch-hydrolyzing CAZymes in this pathway (Figure S3). This dynamic multi-step approach results in the production of trehalose in two routes by the action of CAZymes isoamylase (EC 3.2.1.68), malto-oligosyltrehalose synthase (5.4.99.15) and malto-oligosyltrehalose trehalohydrolase (3.2.1.141) in one route, and cyclomaltodextrin glucanotransferase (2.4.1.19) and cyclomaltodextrinase (EC 3.2.1.54), malto-oligosyltrehalose synthase (5.4.99.15) and malto-oligosyltrehalose trehalohydrolase (3.2.1.141) in the second route (Figure S3). As indicated earlier, other important target carbohydrate is maltose. CAZymes participating in the hydrolysis of starch to maltose can be one of three, e.g., alpha-amylase (EC 3.2.1.1), beta-amylase (EC 3.2.1.2) or maltogenic amylase (3.2.1.133). The gene encoding the first CAZyme was proven *in silico* in the present study to be highly abundant (Fig. 6), thus, the enzyme can effectively act in hydrolyzing starch to maltose. The same enzyme in addition to oligo-alpha-glucosidase (EC 3.2.1.10) can also result in the production of D-glucose (Figure S3). The latter enzyme shares the same activity with the CAZyme encoded by the highly abundant gene in the rhizosphere microbiome of *M. oleifera* namely glucan 1,6-α-glucosidase (EC 3.2.1.70) (Stam et al., 2006). Thus, it is likely that the microbiome will accumulate excessive amounts of glucose due to the action of the latter enzyme (Fig. 7). Also, two other routes (with maltose as the substrate), including alpha-glucosidase (3.2.1.20) and 4-alpha-

glucanotransferase (EC 2.4.1.25), mediate the production of glucose (Figure S3). Overall, we speculate that excessive maltose, α-maltose-1P or the case of high availability of energy would promote plant root to release certain exudates to convert this simple disaccharide to starch as a storage step. In case plant requires to do metabolism to gain energy, then, it releases another type of exudates to promote the hydrolysis of starch, while certain microbes in rhizosphere microbiome serves the plant in either case (Fig. 7). The CAZymes oligo-alpha-glucosidase (EC 3.2.1.10) and alpha-glucosidase (EC 3.2.1.20) also act, respectively, in hydrolyzing isomaltose and maltose to produce glucose (Figure S3). The latter CAZyme can also hydrolyze sucrose to produce glucose.

Different combinations and abundances of simple and complex exudates were reported to result in specific responses of rhizobiome (Badri et al., 2013). For example, it was recently proposed that roots of *Avena barbata* plant have consistent successive patterns of root exudation during different developmental processes that are preferred by specific microbial metabolite substrates (Zhalnina et al., 2018, Pett-Ridge et al., 2021). Referring to temporal root exudation pattern, another recent report indicated that old roots that begun senescence, whose cells started to approach program cell death, secrete exudation patterns specific for distinct microbial populations required at this late growth stage (Pett-Ridge et al., 2021). Examples of these microbes involve family Streptomycetaceae and order Catenulisporales of Actinobacteria that approach high transcription levels of genes encoding CAZymes to breakdown plant cellulose and xylose; two processes that promote soil carbon turnover (Nuccio et al., 2020). Complications of studying exudation patterns include the complex nature of exudates that varies based on plant genotype, root maturity, and environmental condition (Sokol et al., 2019) and the large signal background that makes it difficult to characterize exudate chemical composition (Kuzyakov and Domanski 2000). Prior reports also indicate that progression of the phenological stages of plant growth results in spatiotemporal differences in root habitats, which promote rhizosphere microbial community to undergo a pattern of compositional succession meeting their required functions and life strategies (Shi et al., 2015, Shi et al., 2018). Root exudates also promote rhizosphere microbes to utilize and degrade existing organic materials and to colonize/stabilize carbon. The latter actions monitor/balance the level of soil carbon as carbon will be stabilized in the soil by incorporation into microbial cellular materials followed by subsequent association with cellular minerals, while the non-stabilized carbon will be lost to respiration (Pett-Ridge et al., 2021).

In general, we can conclude that class GH is the most abundant among CAZy classes in microbiomes surrounding roots of *M. oleifera*. These microbiomes can be a source of industrially important enzymes acting on starch hydrolysis/biosynthesis that are nowadays used in several industries like juice, paper and ethanol productions (Raveendran et al., 2018). Therefore, efforts to isolate microbes with high abundance of genes encoding CAZymes of family GH13 from the rhizosphere microbiome of *Moringa oleifera* will have good impact on these industries. We also deeply recommend detecting the relationship between root exudation pattern and function of rhizosphere soil microbes in terms of CAZy classes. This will allow engineering and integrating root microbiomes in breeding programs of domesticated plants to improve their growth by monitoring the availability of required nutrients and the ability to withstand adverse conditions.

6. Data availability statement

Supplemental data can be accessed at: https://drive.google.com/drive/folders/1UMKKBzjLDlQSdrSqxpvaqwWY9n6ClYqj?usp=sharing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.sjbs.2023.103578.

References

Al-Eisawi, D.M., Al-Ruzayza, S., 2015. The flora of holy mecca district, Saudi Arabia. Int. J. Biodiversity Conservation.

Badri, D.V., Chaparro, J.M., Zhang, R., Shen, Q., Vivanco, J.M., 2013. Application of natural blends of phytochemicals derived from the root exudates of arabidopsis to the soil reveal that phenolic-related compounds predominantly modulate the soil microbiome. J. Biol. Chem.

Bahieldin, A., Atef, A., Sabir, J.S., Gadalla, N.O., Edris, S., Alzohairy, A.M., Radhwan, N. A., Baeshen, M.N., Ramadan, A.M., Eissa, H.F., Hassan, S.M., Baeshen, N.A., Abuzinadah, O., Al-Kordy, M.A., El-Domyati, F.M., Jansen, R.K., 2015. Rna-seq analysis of the wild barley (h Spontaneum) leaf transcriptome under salt stress. C R Biol. https://doi.org/10.1016/j.crvi.2015.03.010.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using diamond. Nat. Methods.

Bulgarelli, D., Garrido-Oter, R., Munch, P.C., Weiman, A., Droge, J., Pan, Y., McHardy, A.C., Schulze-Lefert, P., 2015. Structure and function of the bacterial root microbiota in wild and domesticated barley. Cell Host Microbe. https://doi.org/10.1016/j.chom.2015.01.011.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The carbohydrate-active enzymes database (cazy): an expert resource for glycogenomics. Nucleic Acids Res. https://doi.org/10.1093/nar/gkn663.

Davies, G.J., Wilson, K.S., 1999. Trapped in the act of catalysis. Nat. Struct. Biol. https://doi.org/10.1038/8200.

DeAngelis, K.M., Brodie, E.L., DeSantis, T.Z., Andersen, G.L., Lindow, S.E., Firestone, M. K., 2009. Selective progressive response of soil microbial community to wild oat roots. ISME J. https://doi.org/10.1038/ismej.2008.103.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. Cd-hit: Accelerated for clustering the next-generation sequencing data. Bioinformatics. https://doi.org/10.1093/bioinformatics/bts565.

Gopalakrishnan, L., Doriya, K., Kumar, D.S., 2016. Moringa oleifera: a review on nutritive importance and its medicinal application. Food Sci. Human Wellness.

Haiming, T., Xiaoping, X., Chao, L., Xiaochen, P., Kaikai, C., Weiyan, L., Ke, W., 2020. Microbial carbon source utilization in rice rhizosphere and nonrhizosphere soils with short-term manure n input rate in paddy field. Sci. Rep. https://doi.org/10.1038/s41598-020-63639-8.

Hart, G.W., Copeland, R.J., 2010. Glycomics hits the big time. Cell. https://doi.org/10.1016/j.cell.2010.11.008.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., Von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggnog-mapper. Mol. Biol. Evol.

Hurt, R.A., Qiu, X., Wu, L., Roh, Y., Palumbo, A.V., Tiedje, J.M., Zhou, J., 2001. Simultaneous recovery of rna and DNA from soils and sediments. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.67.10.4495-4503.2001.

Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., Schuster, S.C., 2011. Integrative analysis of environmental sequences using megan4. Genome Res.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.-J., Tappu, R., 2016. Megan community edition-interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Comput. Biol.

Janeček, S.T., 1997. A-amylase family: Molecular biology and evolution. Prog. Biophys. Mol. Biol.

Jiménez-Gómez, A., Saati-Santamaría, Z., Igual, J.M., Rivas, R., Mateos, P.F., García-Fraile, P., 2019. Genome insights into the novel species microvirga brassicacearum, a rapeseed endophyte with biotechnological potential. Microorganisms.

Karlsson, F.H., Fåk, F., Nookaew, I., Tremaroli, V., Fagerberg, B., Petranovic, D., Bäckhed, F., Nielsen, J., 2012. Symptomatic atherosclerosis is associated with an altered gut metagenome. Nat. Commun.

Králová, S., Sandoval-Powers, M., Fawwal, D.V., Degnes, K.F., Lewin, A.S., Klinkenberg, G., Nguyen, G.-S., Liles, M.R., Wentzel, A., 2021. Streptomyces tardus sp. Nov.: A slow-growing actinobacterium producing candicidin, isolated from sediments of the trondheim fjord. Front. Microbiol.

Kuzyakov, Y., Domanski, G., 2000. Carbon input by plants into the soil. Rev. J. Plant Nutr Soil Sci. 163, 421–431.

Levy, A., Salas Gonzalez, I., Mittelviefhaus, M., Clingenpeel, S., Herrera Paredes, S., Miao, J., Wang, K., Devescovi, G., Stillman, K., Monteiro, F., 2018. Genomic features of bacterial adaptation to plants. Nat. Genet.

Li, W., Godzik, A., 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. https://doi.org/10.1093/bioinformatics/btl158.

Li, J., Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Arumugam, M., Kultima, J.R., Prifti, E., Nielsen, T., Juncker, A.S., Manichanh, C., Chen, B., Zhang, W., Levenez, F., Wang, J., Xu, X., Xiao, L., Liang, S., Zhang, D., Zhang, Z., Chen, W., Zhao, H., Al-Aama, J.Y., Edris, S., Yang, H., Wang, J., Hansen, T., Nielsen, H.B., Brunak, S., Kristiansen, K., Guarner, F., Pedersen, O., Dore, J., Ehrlich, S.D., Meta, H.I.T.C., Bork, P., Wang, J., Meta, H.I.T.C., 2014. An integrated catalog of reference genes in the human gut microbiome. Nat. Biotechnol. https://doi.org/10.1038/nbt.2942.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., Henrissat, B., 2014. The carbohydrate-active enzymes database (cazy) in 2013. Nucleic Acids Res. https://doi.org/10.1093/nar/gkt1178.

Lozupone, C.A., Hamady, M., Kelley, S.T., Knight, R., 2007. Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Appl. Environ. Microbiol.

Lozupone, C., Knight, R., 2005. Unifrac: a new phylogenetic method for comparing microbial communities. Appl. Environ. Microbiol. https://doi.org/10.1128/AEM.71.12.8228-8235.2005.

Lozupone, C., Lladser, M.E., Knights, D., Stombaugh, J., Knight, R., 2011. Unifrac: an effective distance metric for microbial community comparison. ISME J. https://doi.org/10.1038/ismej.2010.133.

MacGregor, E.A., 1988. Alpha-amylase structure and activity. J. Protein Chem. https://doi.org/10.1007/BF01024888.

Madigan, M.T., Clark, D.P., Stahl, D., Martinko, J.M., 2010. Brock biology of microorganisms. Benjamin Cummings.

Mende, D.R., Waller, A.S., Sunagawa, S., Järvelin, A.I., Chan, M.M., Arumugam, M., Raes, J., Bork, P., 2012. Assessment of metagenomic assembly using simulated next generation sequencing data. PLoS One.

Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., Plichta, D. R., Gautier, L., Pedersen, A.G., Le Chatelier, E., 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat. Biotechnol.

Nuccio, E.E., Anderson-Furgeson, J., Estera, K.Y., Pett-Ridge, J., De Valpine, P., Brodie, E.L., Firestone, M.K., 2016. Climate and edaphic controllers influence rhizosphere community assembly for a wild annual grass. Ecology. https://doi.org/10.1890/15-0882.1.

Nuccio, E.E., Starr, E., Karaoz, U., Brodie, E.L., Zhou, J., Tringe, S.G., Malmstrom, R.R., Woyke, T., Banfield, J.F., Firestone, M.K., 2020. Niche differentiation is spatially and temporally regulated in the rhizosphere. ISME J.

Odelade, K.A., Babalola, O.O., 2019. Bacteria, fungi and archaea domains in rhizospheric soil and their effects in enhancing agricultural productivity. Int. J. Environ. Res. Public Health.

Oh, J., Byrd, A.L., Deming, C., Conlan, S., Program, N.C.S., Kong, H.H., Segre, J.A., 2014. Biogeography and individuality shape function in the human skin metagenome. Nature. https://doi.org/10.1038/nature13786.

Pérez-Jaramillo, J.E., Carrión, V.J., Bosse, M., Ferrão, L.F., De Hollander, M., Garcia, A. A., Ramírez, C.A., Mendes, R., Raaijmakers, J.M., 2017. Linking rhizosphere microbiome composition of wild and domesticated phaseolus vulgaris to genotypic and root phenotypic traits. ISME J.

Pett-Ridge, J., Firestone, M.K., 2017. Using stable isotopes to explore root-microbe-mineral interactions in soil. Rhizosphere.

Pett-Ridge, J., Shi, S., Estera-Molina, K., Nuccio, E., Yuan, M., Rijkers, R., Swenson, T., Zhalnina, K., Northen, T., Zhou, J., 2021. Rhizosphere carbon turnover from cradle to grave: The role of microbe–plant interactions. Rhizosphere biology: Interactions between microbes and plants, Springer, pp. 51-73.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldon, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L.J., von Mering, C., Bork, P., 2014. Eggnog v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res. https://doi.org/10.1093/nar/gkt1253.

Raes, J., Foerstner, K.U., Bork, P., 2007. Get the most out of your metagenome: Computational analysis of environmental sequence data. Curr. Opinion Microbiol.

Raveendran, S., Parameswaran, B., Ummalyma, S.B., Abraham, A., Mathew, A.K., Madhavan, A., Rebello, S., Pandey, A., 2018. Applications of microbial enzymes in food industry. Food Technol. Biotechnol. https://doi.org/10.17113/ftb.56.01.18.5491.

Selim, S., Seleiman, M.F., Hassan, M.M., Saleh, A.A., Mousa, M.A., 2021. Impact of dietary supplementation with moringa oleifera leaves on performance, meat characteristics, oxidative stability, and fatty acid profile in growing rabbits. Animals (Basel). https://doi.org/10.3390/ani11020248.

Shi, S., Nuccio, E., Herman, D.J., Rijkers, R., Estera, K., Li, J., da Rocha, U.N., He, Z., Pett-Ridge, J., Brodie, E.L., Zhou, J., Firestone, M., 2015. Successional trajectories of rhizosphere bacterial communities over consecutive seasons. MBio. https://doi.org/10.1128/mBio.00746-15.

Shi, S., Nuccio, E.E., Shi, Z.J., He, Z., Zhou, J., Firestone, M.K., 2016. The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. Ecol. Lett. https://doi.org/10.1111/ele.12630.

Shi, S., Herman, D.J., He, Z., Pett-Ridge, J., Wu, L., Zhou, J., Firestone, M.K., 2018. Plant roots alter microbial functional genes supporting root litter decomposition. Soil Biol. Biochem.

Sokol, N.W., Kuebbing, S.E., Karlsen-Ayala, E., Bradford, M.A., 2019. Evidence for the primacy of living root inputs, not root or shoot litter, in forming soil organic carbon. New Phytol.

Stam, M.R., Danchin, E.G., Rancurel, C., Coutinho, P.M., Henrissat, B., 2006. Dividing the large glycoside hydrolase family 13 into subfamilies: Towards improved functional annotations of α-amylase-related proteins. Protein Eng. Des. Selection.

Vorholt, J.A., Vogel, C., Carlstrom, C.I., Muller, D.B., 2017. Establishing causality: opportunities of synthetic communities for plant microbiome research. Cell Host Microbe. https://doi.org/10.1016/j.chom.2017.07.004.

Yoon, S.-H., Oh, Y.-K., Kim, Y.-R., Park, J., Han, S.-I., Kim, Y.-W., 2017. Complex formation of a 4-α-glucanotransferase using starch as a biocatalyst for starch modification. Food Sci. Biotechnol.

Zhalnina, K., Louie, K.B., Hao, Z., Mansoori, N., da Rocha, U.N., Shi, S., Cho, H., Karaoz, U., Loque, D., Bowen, B.P., Firestone, M.K., Northen, T.R., Brodie, E.L., 2018. Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. Nat. Microbiol. https://doi.org/10.1038/s41564-018-0129-3.