Microbiomes of biohydrogen production from dark fermentation of industrial wastes: current trends, advanced tools and future outlook

Eka Latiffah Nadia Dzulkarnain1, Jemilatu Omuwa Audu1,2, Wan Rosmiza Zana Wan Dagang1 and Mohd Firdaus Abdul-Wahab1,3*

Abstract
Biohydrogen production through dark fermentation is very attractive as a solution to help mitigate the effects of climate change, via cleaner bioenergy production. Dark fermentation is a process where organic substrates are converted into bioenergy, driven by a complex community of microorganisms of different functional guilds. Understanding of the microbiomes underpinning the fermentation of organic matter and conversion to hydrogen, and the interactions among various distinct trophic groups during the process, is critical in order to assist in the process optimisations. Research in biohydrogen production via dark fermentation is currently advancing rapidly, and various microbiology and molecular biology tools have been used to investigate the microbiomes. We reviewed here the different systems used and the production capacity, together with the diversity of the microbiomes used in the dark fermentation of industrial wastes, with a special emphasis on palm oil mill effluent (POME). The current challenges associated with biohydrogen production were also included. Then, we summarised and discussed the different molecular biology tools employed to investigate the intricacy of the microbial ecology associated with biohydrogen production. Finally, we included a section on the future outlook of how microbiome-based technologies and knowledge can be used effectively in biohydrogen production systems, in order to maximise the production output.

Keywords: Biohydrogen microbiomes, Biohydrogen production, Dark fermentation, Palm oil mill effluent, Industrial wastes, Molecular biology tools

Introduction
Dark fermentation is a biological decomposition process reported to be one of the most promising approaches for the treatment of organic wastes. This is also the process commonly used in sustainable bioenergy production. A recent study by the World Bank in 2018 predicted that the global waste production will grow to 3.4 billion tonnes by 2050, with organic wastes generated from agricultural sectors representing more than 50% of the total waste composition (Kaza et al. 2018). This large amount of wastes has to be sustainably managed. For this purpose, dark fermentation can offer two simultaneous benefits of both waste treatment and sustainable bioenergy generation (Wang and Yin 2019). Methane is currently the commonly produced bioenergy from organic wastes, but hydrogen production is also gaining attention, as part of the hydrogen economy, to substitute the hydrogen produced from fossil fuels. Hydrogen has three times higher energy content (120 MJ/kg) than hydrocarbon fuels, and its combustion is clean and carbon free, producing only water as the by-product (Zhang et al. 2020). Dark
fermentation is more attractive than the other biological processes, due to the lower demand for light (unlike the photosynthetic routes), capable of high biohydrogen production rate, environmentally friendly, versatile substrate utilisation and less energy intensive (Ghimire et al. 2015; Mishra et al. 2019). In addition, the use of organic wastes as feedstocks in dark fermentative biohydrogen production is potentially cost competitive, since organic wastes are relatively abundant, renewable, cheap and highly biodegradable (Sharma et al. 2020).

Various renewable organic wastes such as sake lees, cassava, sago, glycerol, rice straw, vegetable waste, food waste, date seeds, sugarcane molasses, corn stover, alligator weed, oil palm sap and wheat straw have been explored as the potential substrate for dark fermentative biohydrogen production (Chen et al. 2021; Choiron et al. 2020; Li et al. 2020; Liu et al. 2013; Moreno-Andrade et al. 2015; Noparat et al. 2012; Oliveira et al. 2020; Panin et al. 2020; Pason et al. 2020; Rambabu et al. 2020; Saleem et al. 2020; Ulhiza et al. 2018; Zhang et al. 2011). Palm oil mill effluent (POME), a wastewater generated in large quantity during palm oil extraction process is another renewable organic waste of interest that is currently under intense investigations as biohydrogen production (Abdullah et al. 2020; Akhbari et al. 2021; Audu et al. 2019). Nonetheless, the co-existence of biohydrogen-producers such as methanogens and homoacetogens in the mixed culture, makes it a very biochemically complex environment. Despite the multiple studies carried out, there is still a gap in the understanding of the biological mechanisms of dark fermentation for biohydrogen production, including the specific microbial community and the trophic interactions (Cabrol et al. 2017; Das 2017). The methane-producing fermentation systems are more well characterised in this aspect.

Microbiomes are classically defined as the community consisting of microorganisms with distinct properties and metabolic functions, interacting with its environment which results in the formation of specific ecological niche (Whipps et al. 1988). The term “microbiome” was often used interchangeably with “microbiota”, but recently there has been efforts to distinguish these two. Berg et al. (2020) defined “microbiota” as the assemblage of living microorganisms (i.e. the bacteria, archaea, fungi, microalgae and the protists, excluding phages, viruses, plasmids, prions, viroids, and free DNA), while the “microbiomes” are the microbiota and their structural elements, metabolites/signal molecules, and the surrounding environmental conditions (Berg et al. 2020). Phages, viruses, plasmids, prions, viroids, and free DNA are part of the microbiomes. This review will refer to this updated definition.

Taxonomic classification of biogas microbiomes is often accomplished using sequence similarity searches against 16S ribosomal RNA (rRNA) gene reference databases, such as SILVA (Akhbari et al. 2021), Greengenes (Oliveira et al. 2020), Ribosomal Database Project (RDP) (Cho et al. 2018) or National Center for Biotechnology Information (NCBI) (Mazareli et al. 2020). However, the genome sequences of biogas-producing microorganisms are underrepresented in these reference databases, which hinder the reliable taxonomic classification for microbiomes present in the biogas production systems (Hassa et al. 2018). Functional roles of biogas microbiome are often inferred to related species in public genome database based on the 16S rRNA gene sequence similarity (Campanaro et al. 2016). Therefore, it is imperative to have a comprehensive reference database to improve the taxonomic annotation of biogas-producing microbiomes across the entire microbial databases. Metagenomics has been used in many biogas-producing studies to decipher the taxonomic diversity, metabolic functions and the physiology of biogas-producing microbiomes. This has led to the compilation of metagenome-assembled genomes (MAGs) belonging to the biogas-producing microbiomes in a repository, called the “Biogasmicrobiome” (https://biogasmicrobiome.env.dtu.dk/) (Campanaro et al. 2020). This database contains a collection of 1600 MAGs of bacterial and archaeal species that underpin various biogas production systems, substantially expanding the existing public genome databases (Campanaro et al. 2020). In addition, Microbial Database for Activated Sludge (MiDAS) Field Guide (https://www.midasfieldguide.org/guide/search) is an ecosystem-specific database for wastewater treatment systems which aims to facilitate collaborative research and compile information on the physiology and ecology of the key microorganisms present in activated sludge wastewater treatment systems (McIlroy et al. 2015). MiDAS 4 offers a comprehensive set of full-length amplicon sequence variant (ASV)-resolved 16S rRNA gene sequences which covers over 31,000 species, allowing researchers to dig into the microbiome compositions of activated sludge, anaerobic digesters and wastewater treatment systems at the genus to species level resolutions (Dueholm et al. 2021).
Dark fermentation for biohydrogen production is mediated by many different groups of microorganisms, to convert complex organic wastes into biohydrogen, volatile fatty acids and carbon dioxide (CO₂) (Hay et al. 2013). The efficiency and stability of dark fermentation system relies on the syntrophic activity of the microbial community belonging to different functional guilds, working in tight interaction (Cabrol et al. 2017). It has been reported that the understanding of the species composition, specific metabolic functions, and interspecies interactions are often more important than the species richness itself in maintaining the overall performance of the system (Cabrol et al. 2017). The rapid advancement of molecular tools has contributed to the major discoveries of the diversity and structure of the biohydrogen-producing consortia. In a mixed culture system, the microbiomes involved are phylogenetically diverse, with multiple contributions in the biohydrogen production and the breakdown of organic wastes (Cabrol et al. 2017).

This review summarises and evaluates the distinct microbial communities present in a biohydrogen production systems, and the molecular tools that have been used for microbiome analysis in biohydrogen production from industrial wastewater and POME. We also included a future outlook of how microbiome-based technologies and knowledge can be used effectively in biohydrogen production systems, in order to maximise the production output.

**Microbiomes in dark fermentative biohydrogen production**

The microorganisms present in dark fermentative biohydrogen production system include both the biohydrogen producers and non-producers. Biohydrogen producers possess the ability to convert complex organic substrates into biohydrogen in the absence of light. They can exist as a single strain or a community of various taxa. They can be found in a diverse environment, such as POME sludge (Jamali et al. 2019; Mahmood et al. 2019; Zainal et al. 2018), sludge from municipal wastewater plants (Chang et al. 2011; Viana et al. 2019), sludge from food waste (Li et al. 2018), cattle dung (Sen and Suttar 2012), pig manure (Wang et al. 2011) and marine sediments (Liu et al. 2018), many of which has been extensively studied. In general, *Clostridium* (Gram positive, spore former) and *Enterobacter* (Gram negative, non-spore former) are the most common biohydrogen-producing genera reported, for mesophilic conditions (Kumar et al. 2018). While under thermophilic and hyperthermophilic conditions, *Clostridium*, *Thermoanaerobacterium*, *Thermotoga* and *Caldicellulosiruptor* dominate (O-Thong 2017). Research on biohydrogen production using the lower temperature-adapted psychrophiles and psychrotrophs are still somewhat limited (Alvarado-Cuevas et al. 2015; Mohammed et al. 2018). In addition, other genera including *Bacillus*, *Ethanoligenens*, *Klebsiella*, *Citrobacter* and *Escherichia* also frequently reported as the biohydrogen producers. Non-biohydrogen producers on the other hand, could interfere with the overall biohydrogen yield, by either consuming the hydrogen produced, competing with the biohydrogen producers for substrates, or inhibit biohydrogen producers with their produced metabolites which eventually decrease the efficiency of the biohydrogen production system as a whole (Cabrol et al. 2017). Inoculum pre-treatment has become necessary in mixed culture systems in order to selectively enrich the biohydrogen producers and inactivate the hydrogen consumers.

The main biochemical pathways in dark fermentation overlaps with those of anaerobic digestion, where diverse microbial communities synergistically work together to ensure a stable degradation of organic substrates (Abendroth et al. 2015; Stolze et al. 2016). The pathways can be divided into four phases: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Fig. 1). In anaerobic digestion, hydrogen (H₂) is produced during acidogenesis and acetogenesis, by hydrolytic and fermentative bacteria. It is later consumed during methanogenesis, when methanogenic archaea use H₂ and CO₂ to produce methane (CH₄) (Hassa et al. 2018). Therefore, inhibition of methanogenesis is necessary to re-direct the pathway for hydrogen production, through the final step of dark fermentation. The initial hydrolysis starts when the complex substrates (polysaccharides, lipids, and proteins) are hydrolysed to monomers (sugars, amino acids, fatty acids) by the actions of extracellular hydrolytic enzymes such as cellulase, pectinase, lipase and protease. The microbial taxa responsible for the hydrolytic activities are mainly represented by two phyla, *Firmicutes* and *Bacteroidetes*, from the genera *Clostridium* and *Bacteroides*. The abundance of these phyla can be attributed to their degradative abilities, as the main degraders of cellulolytic materials (Abendroth et al. 2015). Members of these phyla are also known to be fast growers, utilising the hydrolysed products for growth through fermentation, and are usually present in the whole degradation process. They are also less sensitive to changes in environmental conditions (Li et al. 2017; Wang et al. 2018). The hydrolysis step is usually not a rate-limiting step, except with recalcitrant substrates such as lignocellulosic waste, which usually results in incomplete hydrolysis requiring a pre-treatment step (Menzel et al. 2020).

Next, in acidogenesis, the hydrolysed products are further metabolised to H₂, CO₂, and short-chain fatty acids (SCFA) (e.g. acetate, butyrate, formate, propionate, etc.) by acidogenic microbial communities. The predominant
phyla associated with this phase are *Bacteroidetes*, *Firmi-
cutes*, *Chloroflexi*, and *Proteobacteria* (Audu et al. 2021; Castellano-Hinojosa et al. 2018). Acidogenesis is usually a rapid process, accompanied by the accumulation of SCFA and the subsequent drop in pH. The microorganisms in the acidogenic phase consist of both facultative and obligate anaerobes and are often referred to as acido-
gens, or acid formers. The commonly reported genera participating in this phase are *Clostridium*, *Bacteroides*, *Bifidobacterium*, *Bacillus*, and *Streptococcus* (Gonzalez-
Martinez et al. 2016; Seon et al. 2014). In this phase, carbo-
hydrates (mostly glucose, the preferred substrates) are converted to pyruvate through the glycolytic pathway (Saravanan et al. 2021; Vardar-Schara et al. 2008). Under mesophilic condition, the H₂-yielding fermentation routes are the obligate anaerobic (*Clostridium* type) and facultative anaerobic (*Enterobacteria* type) fermentation route.

In the facultative anaerobic fermentation route, pyruvate is further converted to acetyl-CoA and formate, by pyruvate formate lyase (PFL), and H₂ is produced from formate by the formate hydrogen lyases enzyme complex. The strict anaerobic fermentation route involves the oxidation of pyruvate to acetyl-CoA and reduced ferredoxin (Fd) by pyruvate ferredoxin oxidoreduc-
tase (PFOR). H₂ is then released from the reduced Fd by the action of hydrogenase. Additional molecules of H₂ can also be produced from the oxidation of nicotinamide adenine dinucleotide (NADH) to reduced Fd by NADH:ferredoxin oxidoreductase (NFOR), followed by the subsequent release of H₂ from the reduced Fd by hydrogenase (Fig. 2). However, the activities of NFOR is usually inhibited under standard conditions and can only proceed when H₂ partial pressure is very low, as opposed to PFOR which is still active at standard H₂ par-
tial pressure (Kraemer and Bagley 2007). In addition to H₂, acetyl-CoA can also be further converted to non-gas-
eous products including SCFA (acetate, lactate, butyrate, propionate), alcohols (ethanol, butanol), and ketones (acetone). The overall theoretical maximum yield of H₂ from the reduced Fd pathway is 4 mol of H₂ per 1 mol of glucose, and 2 mol of H₂ per 1 mol of glucose from the formate pathway. The yield is influenced by the fermenta-
tion end products generated alongside H₂. Theoretically, based on the ‘Thauer limit’, the maximum yield of 4 mol of H₂ can be obtained with acetate as the fermentation end product, 2 mol with butyrate or propionate, and much lower yields when alcohols are the end products (Keskin et al. 2019; Vardar-Schara et al. 2008).

The intermediates products H₂, CO₂, and acetate can directly be utilised by methanogens for methane production, while other products such as butyrate, propionate, valerate require further transformation or conversion through syntrophic acetogenesis first (Lim et al. 2020). In the acetogenesis phase, unus-
able substrates are converted to acetate, CO₂, and H₂.
by hydrolytic and fermentative bacteria that do not possess hydrolytic activities. Acetogenesis is also the rate-limiting phase. In addition, the H₂ produced from acetogenesis is converted to CH₄ by the hydrogenotrophic methanogens (Venkiteshwaran et al. 2015). The oxidation of non-gaseous products of acidogenesis is based on the reverse electron transfer process, a thermodynamically unfavourable condition. The process requires energy input to drive the oxidation/reduction process involving multiple enzyme systems, such as formate dehydrogenases, ferredoxin:NAD oxidoreductases, hydrogenases, reactive quinone complexes, c-type cytochromes, etc. (Sieber et al. 2012). However, when oxidation is coupled with methane production, energy conversion is more feasible due to the diminishing effects of H₂ pressure created by the methanogenic activity (Sikora et al. 2017). The most commonly reported syntrophic acetogens in anaerobic digester are the propionate degraders belonging to the genera Pelotomaculum, Smithilela, and Syntrophobacter. While the oxidation of butyrate and other fatty acids is carried out by Syntrophus and Syntrophomonas (Venkiteshwaran et al. 2015). The acetogenesis phase is important because it ensures rapid and stable anaerobic digester operation by preventing methanogenic inhibition due to the high acid concentrations (Wang et al. 2018).

Methanogenesis is the final phase in anaerobic digestion, in which acetate, H₂ and CO₂ produced from the acidogenic and acetogenic phases are further transformed into biogas, in the form of CH₄ and CO₂. Methanogens are the main hydrogen consumers in the anaerobic environments, and for this reason they are usually suppressed in biohydrogen dark fermentation to maximise the hydrogen yield. Unlike the previous three phases which are dominated by fermentative bacteria, the methanogenesis phase is exclusively dominated by the methanogenic archaea. The methanogens are slow growers and sensitive to environmental changes. Methanogenesis can occur via three possible pathways based on the available substrate: acetoclastic, methylotrophic, or hydrogenotrophic methanogenesis. Acetotrophic/acetoclastic methanogens use acetate as substrate by catalysing its methane production, methylotrophic methanogens use methyl-based compounds, and hydrogenotrophic methanogens use CO₂ and H₂ for CH₄ production (Hassa et al. 2018; Lim et al. 2020; Menzel et al. 2020). The commonly observed methanogens associated with biogas production are from the genera Methanosaeta and Methanosarcina (acetotrophic

---

**Faculative anaerobic (Enterobacteria-type) fermentation**

**Strict anaerobic (Clostridium-type) fermentation**

Fig. 2 An overview of the metabolic pathways of acidogenesis
mammalans); and *Methanobacterium, Methanospirillum, Methanococcus*, and *Methanobrevibacter* (hydrogenotrophic methanogens) (Castellano-Hinojosa et al. 2018). The acetotrophic methanogens have been reported to be the most predominant type of methanogens in anaerobic digesters, and are responsible for about 70% of the methane generated (Gonzalez-Martinez et al. 2016). The genus *Methanoseta* for example, are obligate acetoclastic methanogens, characterised as slow growers, and only use acetate as the substrate. While *Methanosarcina* are facultative acetoclastic methanogens, have faster growth rate, and can utilise a wider substrate range in addition to acetate. Unlike *Methanoseta*, members of the *Methanosarcina* genus have a low affinity to acetate, which can account for its abundance in high acetate concentration condition by outgrowing the *Methanoseta* population. At low acetate concentration, *Methanoseta* have been reported to dominate the archaea community. Due to their high affinity to acetate, *Methanoseta* genus outcompete *Methanosarcina* population by utilising the available acetate in the environment (Conklin et al. 2006; Lim et al. 2020).

The intricacy and complexity of dark fermentation makes it a process "black box"; due to the variety of multiple metabolic activities and interactions within the microbial community, along with the limited biogas-producing microbial genomes in the reference databases. So far, biohydrogen yield obtained in practice are mostly up to 32%, hampered by the Thauer limit (Patel et al. 2018). In addition, biohydrogen production from dark fermentation of organic wastes seldom exceeded 2 mol H2/mol hexose (Wang and Yin 2019). A number of approaches have been explored to overcome the bottleneck of dark fermentative biohydrogen production, including reactor configurations, operational condition, inoculum and substrate types, pre-treatment strategies and integrating multiple biohydrogen production systems (Audu et al. 2020). The microbiomes of the systems are the integral part of these processes. Identifying the microorganisms and understanding the behaviour is crucial to dark fermentation robustness, as elaborated in sections "Industrial wastes as substrates" and "Palm oil mill effluent (POME) as substrate".

**Industrial wastes as substrates**

Dark fermentative biohydrogen production has been investigated using a diverse range of pure cultures (Table 1). *Clostridium butyricum* represents the most commonly studied pure culture under mesophilic condition. *C. butyricum* is known for its high biohydrogen yield regardless of the substrates complexity, which can range from simple carbohydrates such as glucose, xylose and sucrose, to complex biomass such as food waste (Kanchanasuta et al. 2017), glycerol (Kivistö et al. 2013; Yin and Wang 2017), microalgae (Ortigueira et al. 2015) and sugarcane bagasse (Plangklang et al. 2012). At present, the highest reported biohydrogen yield from the conversion of organic waste by *C. butyricum* was 3.0 mol H2/mol glycerol which is equal to 3.6 mol H2/mol hexose out of the theoretical stoichiometric yield (4 mol H2/mol hexose, the Thauer limit) when fermenting raw glycerol from biodiesel production process (Kivistö et al. 2013). In addition, other *Clostridia* species including *C. beijerinckii* (Rambabu et al. 2021), *C. pasteurianum* (Sarma et al. 2019), *C. acetobutylicum* (Azman et al. 2016) and *C. saccharoperbutylicum* (Dada et al. 2013) have also been investigated for biohydrogen production from organic wastes under mesophilic condition. Under thermophilic conditions, *C. thermocellum* has been reported to be the ideal strain. Rambabu et al. (2020) obtained 103.97 mmol H2/L using *C. thermocellum* to produce biohydrogen from date seeds waste through dark fermentation system operated at 50 °C and initial pH 7. Tian et al. (2015) also used *C. thermocellum* to ferment sugarcane bagasse at 55 °C and obtained 4.89 mmol H2/g medium added. Versatile substrate utilisation with high biohydrogen yields of 0.23–3.47 H2/mol hexose from *C. butyricum* and 0.52–3.0 mol H2/mol hexose from the other *Clostridium* species have made *Clostridia* popular for use in dark fermentation (Wang and Yin 2019). However, the strict anaerobic requirement of *Clostridia* complicates their practical applications.

Facultative anaerobes such as *Enterobacter, Klebsiella, Citrobacter, Escherichia* and *Bacillus* are the alternative candidates for biohydrogen dark fermentation. These species possess the ability to shift from aerobic respiration producing adenosine triphosphate (ATP) in the presence of oxygen, to fermentation in the absence of oxygen (Łukajtis et al. 2018). Nevertheless, facultative anaerobes generally produce lower biohydrogen yield than the strict anaerobes, e.g. *Clostridia*, but the high tolerance to oxygen makes them more attractive for practical applications. Pure cultures have commonly been used for investigations on biohydrogen production and the related metabolic activity. This allows the investigation into the metabolic pathways involved, and subsequently the feasible ways to enhance the biohydrogen production efficiency through metabolic engineering (Wang and Yin 2019). In addition, reliable biohydrogen yields by avoiding the production of undesired by-products, reproducible bioprocess and ease of genetic manipulations are the other benefits of using pure cultures (Kumar et al. 2018). However, aseptic condition is compulsory to handle pure cultures as they are highly susceptible to contaminations. They are also substrate selective, and developing pure
Table 1  Dark fermentative biohydrogen production from various substrates (including industrial wastes) using pure culture

| Microorganism          | Substrate       | Reactor type | Operating conditions | Biohydrogen yield | References |
|------------------------|-----------------|--------------|----------------------|-------------------|------------|
|                        |                 |              | Temperature, $T$ | pH | Biohydrogen yield | References |
| *Clostridium butyricum* | Glucose         | Batch        | 35 °C | Initial = 7.0, Operation = uncontrolled | 2.24 mol H$_2$/mol hexose | Yin and Wang (2017) |
| INET 1                 |                 |              |                     |                  |            |
| *Clostridium butyricum* | Xylose          | Batch        | 35 °C | Initial = 7.0, Operation = uncontrolled | 1.23 mol H$_2$/mol hexose | Yin and Wang (2017) |
| INET 1                 |                 |              |                     |                  |            |
| *Clostridium butyricum* | Sucrose         | Batch        | 35 °C | Initial = 7.0, Operation = uncontrolled | 1.44 mol H$_2$/mol hexose | Yin and Wang (2017) |
| INET 1                 |                 |              |                     |                  |            |
| *Clostridium butyricum* | Lactose         | Batch        | 35 °C | Initial = 7.0, Operation = uncontrolled | 1.83 mol H$_2$/mol hexose | Yin and Wang (2017) |
| INET 1                 |                 |              |                     |                  |            |
| *Clostridium butyricum* | Starch          | Batch        | 35 °C | Initial = 7.0, Operation = uncontrolled | 2.17 mol H$_2$/mol hexose | Yin and Wang (2017) |
| INET 1                 |                 |              |                     |                  |            |
| *Clostridium butyricum* | Glycerol        | Batch        | 35 °C | Initial = 7.0, Operation = uncontrolled | 362 mL H$_2$/g VS | Kanchanasuta et al. (2017) |
| INET 1                 |                 |              |                     |                  |            |
| *Clostridium butyricum* | Food waste      | CSTR         | 37 °C | Initial = 6.0, Operation = uncontrolled | 2.74 mol H$_2$/mol glucose | Ortigueira et al. (2015) |
| TISTR 1032             |                 |              |                     |                  |            |
| *Clostridium butyricum* | Microalgae      | Batch        | 37 °C | Initial = 5.2, Operation = 5.2 | 1.67 mol H$_2$/mol glucose | Puhulwella et al. (2014) |
| DSM 10,702             |                 |              |                     |                  |            |
| *Clostridium butyricum* | Glucose         | TBSBR        | 30 °C | Initial = 7.4, Operation = uncontrolled | 3.0 mol H$_2$/mol glycerol | Kivistö et al. (2013) |
| CWBI 1009              |                 |              |                     |                  |            |
| *Clostridium butyricum* | Glycerol        | Batch        | 37 °C | Initial = 7.0, Operation = uncontrolled | 0.67 mol H$_2$/mol glycerol | Yin and Wang (2017) |
| INET 1                 |                 |              |                     |                  |            |
| *Clostridium butyricum* | Glucose         | AnSBR        | 30 °C | Initial = 7.6, Operation = uncontrolled | 2.2 mol H$_2$/mol glucose | Beckers et al. (2013) |
| CWBI 1009              |                 |              |                     |                  |            |
| *Clostridium butyricum* | Glucose         | ABR          | 30 °C | Initial = 8.5, Operation = uncontrolled | 2.49 mol H$_2$/mol glucose | Laurent et al. (2012) |
| CWBI 1009              |                 |              |                     |                  |            |
| *Clostridium beijerinckii* | Sugarcane bagasse | Serum bottle | 37 °C | Initial = 6.5, Operation = 6.5 | 1.52 mol H$_2$/mol hexose$_{used}$ | Plangklang et al. (2012) |
| DSM 791                |                 |              |                     |                  |            |
| *Clostridium beijerinckii* | Rice mill wastewater | Serum bottle | 37 °C | Initial = 7.0, Operation = uncontrolled | 214.9 mL H$_2$/L | Rambabu et al. (2021) |
| DSM 791                |                 |              |                     |                  |            |
| *Clostridium beijerinckii* | Oil palm sap | Serum bottle | 30 °C | Initial = 7.0, Operation = uncontrolled | 141 mL H$_2$/g substrate | Noparat et al. (2012) |
| PS-3                   |                 |              |                     |                  |            |
| *Clostridium pasteurianum* | Glycerol | Serum bottle | 37 °C | Initial = 7.0, Operation = 7.0 | 1.10 mol H$_2$/mol glycerol | Sarma et al. (2019) |
| DSM 525                |                 |              |                     |                  |            |
| *Clostridium pasteurianum* | Glucose    | Serum bottle | 35 °C | Initial = 7.0, Operation = uncontrolled | 2.2 mol H$_2$/mol xylose | Hsieh et al. (2016) |
|                        |                 |              |                     |                  |            |
| *Clostridium BOH3*     | Fruit waste     | Serum bottle | 37 °C | Initial = 6.8, Operation = uncontrolled | 359.97 mL H$_2$/g TS utilised | Mahato et al. (2020) |
|                        |                 |              |                     |                  |            |
| *Clostridium thermocellum* | Date seeds | Serum bottle | 50 °C | Initial = 7.0, Operation = uncontrolled | 103.97 mmol H$_2$/L | Rambabu et al. (2020) |
| ATCC 27,405            |                 |              |                     |                  |            |
| *Clostridium acetobutylicum* | Rice bran | Batch        | 34 °C | Initial = 5.5, Operation = uncontrolled | 117.24 mL H$_2$/g sugar$_{consumed}$ | Azman et al. (2016) |
cultures to reach the optimal production period can be time consuming (Kumar et al. 2018).

Mixed cultures have also been widely used (Table 2). Inoculum pre-treatment is necessary in a mixed culture system to enhance biohydrogen production yield by suppressing the activity of competing species, such as the biohydrogen-consumers and substrates competitor. Different pre-treatment methods will result in different starting microbial community structures. Zhang et al. (2011) investigated the effects of five different inoculum pre-treatment methods on mixed culture of aerobic seed sludge to enhance biohydrogen production from corn stover hydrolysate. Inoculum with no pre-treatment is composed of complex microbial community mainly represented by Enterobacter sp. and Pantoea sp. Pre-treatment using heat achieved the highest biohydrogen yield,

Table 1 (continued)

| Microorganism                     | Substrate                      | Reactor type | Operating conditions | Biohydrogen yield | References                  |
|-----------------------------------|--------------------------------|--------------|----------------------|-------------------|-----------------------------|
| Clostridium thermocellum ATCC 27,405 | Sugarcane bagasse             | Serum bottle | Temperature, $T$ = 55 °C, pH = 6.6, Operation = uncon- trolled | 4.89 mmol H$_2$/g medium$_{added}$ | Tian et al. (2015)          |
| Clostridium saccharolyticum N1-4 | Rice bran                      | Batch        | Temperature, $T$ = 30 °C, pH = 6.0, Operation = uncon- trolled | 3.37 mol H$_2$/mol sugar$_{consumed}$ | Dada et al. (2013)          |
| Clostridium tyrobutyricum Fya102 | Glucose                        | CSTR         | Temperature, $T$ = 35 °C, pH = 6.0, Operation = uncon- trolled | 1.06 mmol H$_2$/mmol glucose | Whang et al. (2011)          |
| Enterobacter aerogenes ZJU1      | Aquatic weed                   | Batch        | Temperature, $T$ = 37 °C, pH = 6.0, Operation = uncon- trolled | 62.2 mL H$_2$/g VS | Song et al. (2020)          |
| Enterobacter aerogenes CDC 819–56 | Lactose                        | Batch        | Temperature, $T$ = 25.6 °C, pH = 7.2, Operation = uncon- trolled | 1.19 mol H$_2$/mol lactose | Alvarez-Guzmán et al. (2020) |
| Enterobacter aerogenes PTCC 1221 | Sago wastewater                | Serum bottle | Temperature, $T$ = 31 °C, pH = 6.8, Operation = uncon- trolled | 7.42 mmol H$_2$/mol glucose | Ulhiza et al. (2018)         |
| Enterobacter cloacae IIT-BT 08   | Rice straw                     | Serum bottle | Temperature, $T$ = 37 °C, pH = 7.5, Operation = uncon- trolled | 19.73 mL H$_2$/g straw | Asadi and Zilouei (2017)     |
| Bacillus cereus                  | Wheat straw                    | Batch        | Temperature, $T$ = 25.6 °C, pH = 7.5, Operation = uncon- trolled | 156.4 mL H$_2$/g VS | Saleem et al. (2020)         |
| Ethanoligenens harbinense B49   | Glucose                        | Serum bottle | Temperature, $T$ = 36 °C, pH = 6.5, Operation = uncon- trolled | 113.5 mmol H$_2$/L | Xu et al. (2016)             |
| Ethanoligenens harbinense YUAN-3 | Glucose                        | Batch        | Temperature, $T$ = 35 °C, pH = 6.0, Operation = uncon- trolled | 2.62 mol H$_2$/mol glucose | Zhang et al. (2015)          |
| Escherichia coli                | Glucose                        | Serum bottle | Temperature, $T$ = 37 °C, pH = 6.0, Operation = uncon- trolled | 2.0 mol H$_2$/mol glucose | Bisaihlon et al. (2006)      |
| Janthinobacterium agaricidamnosum | Glucose                      | Serological bottle | Temperature, $T$ = 25 °C, pH = 6.5, Operation = uncon- trolled | 0.86 mol H$_2$/mol glucose | Alvarado-Cuevas et al. (2015) |
| Polaromonas jejuensis             | Glucose                        | Serological bottle | Temperature, $T$ = 25 °C, pH = 6.5, Operation = uncon- trolled | 1.57 mol H$_2$/mol glucose | Alvarado-Cuevas et al. (2015) |
| Klebsiella pneumoniae            | Brewery wastewater            | AnBBR        | Temperature, $T$ = 35–36 °C, pH = 7.5, Operation = uncon- trolled | 0.80–1.67 mol H$_2$/mol glucose | Estevam et al. (2018)        |
| Klebsiella pneumoniae BLB01     | Glycerol                       | Batch        | Temperature, $T$ = 39 °C, pH = 9.0, Operation = uncon- trolled | 45.0 mol % | Costa et al. (2011)          |

ABR: anaerobic biodisc reactor; AnBBR: mechanically stirred anaerobic reactor; AnSBR: anaerobic sequenced-batch reactor; CSTR: continuous stirred tank reactor; TBSBR: trickling-bed sequenced-batch reactor; ND: no data; COD: chemical oxygen demand; TS: total solid; VS: volatile solid
| Inoculum source | Substrate | Inoculum pretreatment | Reactor type | Operating Conditions | % H₂ | H₂ yield | H₂ production rate | Dominant members | Technique for microbial community analysis | References |
|----------------|-----------|-----------------------|--------------|----------------------|------|----------|-------------------|-----------------|--------------------------------|------------|
| Wastewater treatment plant | Glycerol | IR | Batch | 36 °C | Initial = 7.0, Operation = uncontrolled | ND | 0.52 mol H₂/mol glycerol | ND | ND | ND | Chen et al. (2021) |
| Sewage treatment plant | Glucose | IR | Batch | 36 °C | Initial = 7.0, Operation = uncontrolled | ND | ND | ND | Oostridium sensu stricto 1 and Oostridium butyricum | Yin and Wang (2021) |
| Biogas slurry | Sake lees | HT, 150 °C, 40 min | Batch | 37 °C | Initial = 6.0, Operation = uncontrolled | ND | 112.07 mL H₂/g COD | ND | Pantosoma agglomerans, Oostridium acetobutylicum, Oostridium butyricum | Choiron et al. (2020) |
| Biogas digester of pig farm | Alligator weed | HT, 100 °C, 30 min | Batch | 37 °C | Initial = 7.0, Operation = uncontrolled | ND | 48.4 mL H₂/g VS | 563 mL H₂/g VS h | Oostridium sensu stricto 1, Anaerobacterium, Oostridium IV | Li et al. (2020) |
| Chicken manure | Cassava | HT, 95 °C, 15 min | Hungate tube | 36 °C | Initial = 6.0, Operation = uncontrolled | ND | ND | 1662 mL H₂/L h | Clostridiaceae, Porphyromonadaceae and Rimonococcaceae | Martinez-Burgos et al. (2020) |
| Vinasse effluent | Cassava | HT, 95 °C, 15 min | Hungate tube | 37 °C | Initial = 6.0, Operation = uncontrolled | ND | ND | 2182 mL H₂/L h | Clostridiaceae, Porphyromonadaceae and Rimonococcaceae | Martinez-Burgos et al. (2020) |
| Indigenous microbes | Banana wastes | ND | Batch | 37 °C | Initial = 7.5, Operation = uncontrolled | ND | 3.80 mL H₂ | ND | Lactobacillus and Clostridium | Mazareli et al. (2020) |
| Indigenous microbes | Coffee wastes | ND | Batch | 30 °C | Initial = 7.0, Operation = uncontrolled | ND | 240 mL H₂ | 3040 mL H₂/L day | Oostridium sp., Lactobacillus sp., Kazachstanina sp. and Saccharomyces sp. | Villa Montoya et al. (2020) |
| Inoculum source | Substrate | Inoculum pre-treatment | Reactor type | Operating Conditions | % H₂ | H₂ yield | H₂ production rate | Dominant members | Technique for microbial community analysis | References |
|-----------------|-----------|------------------------|--------------|----------------------|------|---------|-------------------|-----------------|------------------------------------------|------------|
| Indigenous microbes | Sugarcane molasses | ND | AnSTBR-A | Initial = 3.8, Operation = uncontrolled | 51 | 1.18 mol H₂/mol total carbohydrates | 88 mL H₂/L h | Thermoanaerobacterium | Amplicon sequencing — Ion Torrent | Oliveira et al. (2020) |
| Effluent of hydrogen fermenter | Vegetable waste | ND | Batch | Initial = 7.0, Operation = uncontrolled | 43.54 | 151.67 mL H₂/g VSadded | ND | ND | ND | Panin et al. (2020) |
| Soil sediment of mangroves | Cassava pulp | ND | Serum bottle | Initial = 7.0, Operation = uncontrolled | ND | 23 mL H₂/g substrate | ND | Clostridium thermo-thermophilum, Clostridium isatidis, Thermoanaerobacterium, Fonticella tunisiensis | Amplicon sequencing — Illumina | Pason et al. (2020) |
| Municipal wastewater treatment plant | Glycerol | ND | CSTR | Initial = 6.0, Operation = 6.5 | 49 | 403.6 mmol H₂/mol Glyc consumed | 494 mL H₂/L h | Clostridium intestinale | Amplicon sequencing — Illumina MiSeq | Paillet et al. (2019) |
| Sewage treatment plant | Glucose | HT, 100 °C, 15 min | Batch | Initial = 7.0, Operation = uncontrolled | ND | 1.4 mol H₂/mol glucose | ND | Clostridium sensu stricto 1 (C. butyricum and C. paraputrefaciens), Paraclostridium, Paeniclostridium and Romboutsia | Amplicon sequencing — Illumina MiSeq | Yang and Wang (2019) |
| Buffalo sludge and rumen | Buffalo waste | ND | Batch | Initial = 7.0, Operation = uncontrolled | 48.1 | 120.8 mL H₂/g VS | ND | Clostridia Incertae Sedis, Clostridium sensu stricto 1, Prevotella | Amplicon sequencing — Illumina MiSeq | Chiariotti and Crisì (2018) |
| Food waste treatment plant | Glucose | No pre-treatment | UASB | Initial = 4.9, Operation = uncontrolled | 43.49 | 0.89 mol H₂/mol hexose | 468 L H₂/L day | Enterobacter cloacae, Aeromonas hydrophila, Clostridium pasteurianum | Amplicon sequencing — Ion Torrent | Cho et al. (2018) |
| Inoculum source | Substrate | Inoculum pretreatment | Reactor type | Operating Conditions | % \( \text{H}_2 \) | \( \text{H}_2 \) yield | \( \text{H}_2 \) production rate | Dominant members | Technique for microbial community analysis | References |
|----------------|-----------|-----------------------|--------------|----------------------|-----------------|-------------------------|-------------------------------|----------------|-----------------------------------------------|------------|
| Mixed culture  | Tequila vinasse | ND | CSTR | 35 °C | Initial \( t = 6.5 \), Operation \( t = 5.8 \) | 70 | 124 NmL \( \text{H}_2/\text{g VS}_{\text{added}} \) | 159 NmL \( \text{H}_2/\text{L h} \) | Clostridium beijerinckii, Streptococcus sp. and Acetobacter kovaniensis | Amplicon sequencing—Illumina MiSeq | García-Depraect and León-Becer-rril (2018) |
| Soil and composting residue | Sugarcane bargasse | ND | Batch | 37 °C | Initial \( t = 6.0 \), Operation = uncontrolled | ND | 1.53 mmol \( \text{H}_2/\text{L} \) | ND | Clostridium, Bacteroides, Parabacteroides, Porphyromonas, Desulfitobacterium, Bacillus and Methanobacterium | Metagenomic—Illumina HiSeq | Soares et al. (2018) |
| Anaerobic sludge | Food waste | ND | Batch | 35 °C | Initial \( t = 6.3 \), Operation = 7.0 | ND | 2.3 mol \( \text{H}_2/\text{mol glucose}_{\text{added}} \) | 78 L \( \text{H}_2/\text{L day} \) | Clostridium butyricum, Enterococcus sp. and Enterobacter sp. | qPCR | Jia et al. (2017) |
| Brewery wastewater treatment plant | Glucose | HT, 90 °C, 30 min | FBR | 37 °C | Initial \( t = 5.5 \), Operation = 5.5 | ND | 0.58 mol \( \text{H}_2/\text{mol glycerol} \) | 880 mmoll \( \text{H}_2/\text{L day} \) | Clostridium sp., Pevotella sp. and Klebsiella sp. | CE-SSCP—Illumina MiSeq | Silva-Illanes et al. (2017) |
| Domestic wastewater treatment plant | Glycerol | Ae, 24 h | CSTR | 37 °C | Initial \( t = 5.5 \), Operation = 5.5 | 61 | 0.26 mol \( \text{H}_2/\text{mol glycerol} \) | 510 mmol \( \text{H}_2/\text{L day} \) | Clostridium sp., Enterococcus sp., Pevotella sp. and Snodgrassella sp. | CE-SSCP—Illumina MiSeq | Silva-Illanes et al. (2017) |
| Domestic wastewater treatment plant | Glycerol | Ae, 24 h | CSTR | 37 °C | Initial \( t = 6.0 \), Operation = 6.0 | 57 | 0.26 mol \( \text{H}_2/\text{mol glycerol} \) | 510 mmol \( \text{H}_2/\text{L day} \) | Clostridium sp., Enterococcus sp., Pevotella sp. and Snodgrassella sp. | CE-SSCP—Illumina MiSeq | Silva-Illanes et al. (2017) |
| Inoculum source | Substrate | Inoculum pretreatment | Reactor type | Operating Conditions | % H<sub>2</sub> | H<sub>2</sub> yield | H<sub>2</sub> production rate | Dominant members | Technique for microbial community analysis | References |
|-----------------|-----------|-----------------------|--------------|----------------------|------------|--------------|--------------------------|-------------------|---------------------------------|------------|
| Brewery industry | Food waste | HT, 103–105 °C, 60 min | ASBR         | 35 °C Initial = 5.5, Operation = 5.5 | 37.9       | 103.6 mL H<sub>2</sub>/g COD<sub>removed</sub> | 2264 mL H<sub>2</sub>/L h | Megasphaera, Veillonella | Pyrosequencing—Roche | Moreno-Andrade et al. (2015) |
| Fruit juice wastewater treatment plant | Food waste | HT, 100 °C, 60 min | Batch         | 37 °C Initial = 6.0, Operation = uncontrolled | ND         | 2.68 mol H<sub>2</sub>/mol hexose | ND | Clostridium frigidicarnis | Pyrosequencing—Roche | Laothanachareon et al. (2014) |
| Sugarcane stillage treatment plant | Sugarcane stillage | HT, 90 °C, 10 min | AFBR         | 55 °C Initial = 4.1–7.0, Operation = uncontrolled | ND         | 43.3–48.9 mol H<sub>2</sub>/g COD<sub>added</sub> | 149 L H<sub>2</sub>/L h | Megasphaera sp., Lactobacillus | PCR-DGGE—Cloning | Santos et al. (2014) |
| Garbage compost | Beer lees | No pre-treatment | Batch         | 37 °C Initial = 4.1–7.0, Operation = uncontrolled | ND         | 29.3 mL H<sub>2</sub>/g TS | ND | Clostridium roseum | PCR-DGGE | Bando et al. (2013) |
| Municipal wastewater treatment plant | Rice straw hydrolysate | HT, 95–100 °C, 60 min | CSABR | 37 °C Initial = 5.5, Operation = 5.5 | ND         | 0.69 mol H<sub>2</sub>/mol T-sugar | 10 L H<sub>2</sub>/L day | Clostridium pasteurianum | PCR-DGGE | Liu et al. (2013) |
| Beach | Sucrose | HT, 100 °C, 45 min | AGSBR | 35 °C Initial = 6.0, Operation = 6.0 | 37         | 1.04 mol H<sub>2</sub>/mol sucrose | 15.59 m<sup>3</sup> H<sub>2</sub>/m<sup>3</sup> day | Clostridium pasteurianum and Bifidobacteria sp. | PCR-DGGE | Lin et al. (2011) |
| Wastewater treatment plant | Corn stover hydrolysate | HT, 100 °C, 60 min | Serum bottle   | 37 °C Initial = 7.0, Operation = uncontrolled | ND         | 502 mL H<sub>2</sub>/L | 373 mL H<sub>2</sub>/h | Clostridium bifermens, Escherichia coli, Escherichia vulneris, Uncultured Escherichia sp. | PCR-DGGE—Cloning | Zhang et al. (2011) |
| Wastewater treatment plant | Corn stover hydrolysate | Ac, pH 3, 60 min | Serum bottle   | 37 °C Initial = 7.0, Operation = uncontrolled | ND         | 352.1 mL H<sub>2</sub>/L | 137 mL H<sub>2</sub>/h | Enterobacter aerogenes, Escherichia coli, Escherichia vulneris, Uncultured Escherichia sp. | PCR-DGGE—Cloning | Zhang et al. (2011) |
### Table 2 (continued)

| Inoculum source                  | Substrate Inoculum pretreatment | Reactor type | Operating Conditions | % \(H_2\) | \(H_2\) yield | \(H_2\) production rate | Dominant members | Technique for microbial community analysis | References |
|---------------------------------|---------------------------------|--------------|---------------------|------------|---------------|------------------------|----------------|--------------------------------------|------------|
| Wastewater treatment plant      | Corn stover hydrolysate (Ba, pH 12, 60 min) | Serum bottle | 37 °C Initial = 7.0, Operation = uncontrolled | ND         | 458.4 mL \(H_2\)/L | 123 mL \(H_2\)/h | Enterobacter aerogenes, Klebsiella, Pectobacterium sp. | PCR-DGGE—Cloning | Zhang et al. (2011) |
| Wastewater treatment plant      | Corn stover hydrolysate (Us, 15 min) | Serum bottle | 37 °C Initial = 7.0, Operation = uncontrolled | ND         | 295.9 mL \(H_2\)/L | 187 mL \(H_2\)/h | Enterobacter aerogenes, Pectobacterium sp. | PCR-DGGE—Cloning | Zhang et al. (2011) |
| Wastewater treatment plant      | Corn stover hydrolysate (Uv, 15 min) | Serum bottle | 37 °C Initial = 7.0, Operation = uncontrolled | ND         | 290.9 mL \(H_2\)/L | 589 mL \(H_2\)/h | Pectobacterium sp. | PCR-DGGE—Cloning | Zhang et al. (2011) |
| Sewage treatment plant          | Molasses (HT, 100 °C, 45 min) | CSTR | 35 °C Initial = 5.5, Operation = 5.5 | 47         | 2.1 mol \(H_2\)/mol hexose | 153 mmol \(H_2\)/L/day | Clostridium acetobutylicum and Clostridium pasteurianum | qPCR | Lay et al. (2010) |

HT: heat treatment; IR: ionising radiation; Ac: acid; Ae: aeration; Ba: base; ND: no data; Us: ultrasonication; Uv: ultraviolet; AFBR: anaerobic thermophilic fluidised bed reactor; AGSBR: agitated granular sludge bed reactor; AnSTBR-A: anaerobic structured-bed reactor; ASBR: anaerobic sequencing batch reactor; CSABR: continuously stirred anaerobic bioreactor; CSTR: continuous stirred tank reactor; FBR: fixed-bed reactor; UASB: up-flow anaerobic sludge blanket reactor; COD: chemical oxygen demand; Gly: glycerol; VS: volatile solid; CE-SSP: capillary electrophoresis single-stranded conformation; PCR-DGGE: polymerase chain reaction-denaturing gradient gel electrophoresis; qPCR: quantitative polymerase chain reaction
with the microbial community in the fermentation system dominated by \textit{C. bif ermentans}. Pre-treatment using base, acid, ultrasonic disruption and ultraviolet radiation favours facultative anaerobes, such as \textit{E. aerogenes}, \textit{Klebsiella}, \textit{Pectobacterium} and \textit{E. coli}. Heat pre-treatment is the commonly used inoculum pre-treatment method under mesophilic conditions (Table 2). It is usually selective to spore-forming species such as \textit{Clostridia}, and inhibits other non-spore formers. In general, the use of mixed culture in dark fermentation has been shown to be promising, and it offer high hydrogen evolution rate and yields (Pachapur et al. 2019). However, understanding the metabolic complexities and process kinetics taking place within undefined microbiome systems are challenging.

Artificial microbial consortia containing selected microorganisms with specific metabolic or ecological functions has been shown to overcome the limitations of wild type and undefined microorganisms (Ergal et al. 2020). Recently, precision design of an artificial microbial consortia consisting \textit{E. aerogenes} and \textit{C. acetobutylicum} yielded 5.6 mol H\textsubscript{2}/mol glucose. This was the highest biohydrogen yield reported so far, 40% beyond the Thauer limit (Ergal et al. 2020). The finding suggests that constructing a desired microbial consortium with well-studied biohydrogen-producing species will enable a comprehensive understanding of the microbial interactions, ease the control and balancing the effects of any perturbations. This will ultimately create a more efficient and robust engineered system.

**Palm oil mill effluent (POME) as substrate**

POME is the wastewater produced in large quantity during palm oil processing. It contains substantial amount of organic material, suspended solids, and oil and greases. Despite its nontoxic nature, POME is categorised as extremely high strength wastewater, which is 100 times more polluted than municipal sewage, and require effective treatment before discharge into the environment (Chia et al. 2020). Raw POME appears as thick brownish high colloidal suspension liquid mixture with a distinct offensive odour (Chia et al. 2020). It is characterised by high biological oxygen demand (BOD) (10, 250–80, 400 mg/L), high chemical oxygen demand (COD) (15,000–100,000 mg/L), high oil and grease content (130–18,000 mg/L), high suspended solids (5000–54,000 mg/L), high discharge temperature (50–90 °C) and is acidic (pH 3.4–6.9) (Audu et al. 2020). POME is rich with organic materials containing cellulose (11%), hemi-cellulose (7%) and lignin (42%) (O-Thong et al. 2012). Given the high organic matter properties, recent POME treatment methods are coupled with bioenergy production and other value-added products, such as solvents, biomethane and biohydrogen.

\textit{Clostridia} is the most commonly used genera for biohydrogen production from POME. \textit{C. butyricum} has been used in several studies as pure culture inoculum for mesophilic batch biohydrogen production from POME via dark fermentation (Table 3). Singh et al. (2013b) observed that biohydrogen yield increased 1.5- to 2-fold when using an acclimatised immobilised \textit{C. butyricum}. The immobilised cells recorded a biohydrogen yield of 5350 mL H\textsubscript{2}/L POME with maximum biohydrogen production rate of 510 mL H\textsubscript{2}/L POME/h. This species has also been reported to be the dominant biohydrogen producer in POME fermentation using mixed culture (Yossan et al. 2012) (Table 4). The effects of mesophilic and

### Table 3 Dark fermentative biohydrogen production from POME using pure culture

| Inoculum                   | Reactor type | Operating conditions | Biohydrogen yield | References         |
|---------------------------|--------------|----------------------|-------------------|-------------------|
| \textit{Clostridium beijerinckii} | Hungate tube | Initial = 7.0, Operation = uncontrolled | 4620 mL H\textsubscript{2}/L medium | Rosa et al. (2020) |
| \textit{Bacillus anthracis} | CSTR         | Initial = 6.5, Operation = uncontrolled | 236 mL H\textsubscript{2}/g COD | Mishra et al. (2017) |
| \textit{Escherichia coli} | Serum bottle | Initial = 8.5, Operation = uncontrolled | 0.66 mol H\textsubscript{2}/mol total monomeric sugars | Taifor et al. (2017) |
| \textit{Clostridium LS2}  | UASB         | Initial = 5.5, Operation = 5.5 | 380 mL H\textsubscript{2}/g COD | Singh et al. (2013b) |
| \textit{Clostridium butyricum} EB6 | Batch       | Initial = 5.5, Operation = 5.5 | 5350 mL H\textsubscript{2}/L POME | Singh et al. (2013c) |
| \textit{Clostridium butyricum} | Batch       | Initial = 7.0, Operation = 5.5 | 2.18 mol H\textsubscript{2}/mol total carbohydrate | Kamal et al. (2011) |
| \textit{Clostridium butyricum} EB6 | Batch       | Initial = 5.5, Operation = 5.5 | 3195 mL H\textsubscript{2}/L POME | Chong et al. (2009) |

CSTR: continuous stirred tank reactor; UASB: up-flow anaerobic sludge blanket reactor; COD: chemical oxygen demand; POME: palm oil mill effluent
| Inoculum                        | Inoculum pre-treatment | Reactor type | Operating conditions | % H2 | H2 yield | H2 production rate | Dominant microbes                                                                 | Technique for microbial community analysis | References                  |
|--------------------------------|------------------------|--------------|----------------------|------|----------|--------------------|----------------------------------------------------------------------------------|-------------------------------------------|------------------------------|
| POME anaerobic sludge          | HT, 80 °C, 50 min      | UASFF        | Initial = 52 - 5.8, Operation = 5.2 - 5.8 | 71.37| 800 mL H2/g COD consumed | 4.1 L H2/L day | Clostridium sensu stricto T and Lactobacillus Ostridia                        | Amplicon sequencing—Illumina MiSeq   | Akhbari et al. (2021)      |
| Anaerobic sludge from methane-producing anaerobic digester | HT, 90 °C, 60 min      | CSTR         | Initial = 55, Operation = 5.5 | 30 — 34| 249 mL H2 g COD | 22.22 mL H2/L h | Ostridium | Amplicon sequencing—Illumina MiSeq | Audu et al. (2021) |
| Thermophilic biohydrogen-producing sludge | HT, 80 °C, 60 min | Batch       | Initial = 60, Operation = uncontrolled | 38.77| 794.85 mL H2/L POME or 1.88 mol H2/mol sugar | ND | ND | Amplicon sequencing—Illumina MiSeq | Rosa et al. (2020) |
| Sugarcane cultivation soil     | ND                     | Hungate tube | Initial = 7, Operation = uncontrolled | ND | 1617 mL H2/L medium | ND | Sporolactobacillus and Clostridium | Amplicon sequencing—Illumina MiSeq | Rosa et al. (2020) |
| Vinasse pond                   | ND                     | Hungate tube | Initial = 7, Operation = uncontrolled | ND | 1550 mL H2/L medium | ND | Clostridium and Ruminococcus | Amplicon sequencing—Illumina MiSeq | Rosa et al. (2020) |
| POME sludge                    | HT, 80 °C, 60 min      | FBR          | Initial = 60, Operation = 6.0 | ND | 1.24 mol H2/mol sugar consumed | 5.2 mmol H2/L h | Thermoanaerobacterium thermosaccharolyticum sp. | PCR-DGGE                   | Jamali et al. (2019) |
| POME anaerobic sludge          | ND                     | Serum bottles | Initial = 6.5, Operation = uncontrolled | ND | 71 mL H2/g COD | 7.6 mL H2/g COD day | Thermoanaerobacterium sp., T. thermosaccharolyticum, T. aciditolerans, T. brockii, Clostridium increased over time | PCR-DGGE | Khongkliang et al. (2019) |
| POME sludge                    | HT                     | A3BR         | Initial = 60, Operation = uncontrolled | ND | 2.52 mol H2/mol sugar | 10.34 mmol H2/L h | Thermoanaerobacterium sp. | PCR-DGGE                   | Maaroff et al. (2019) |
| POME digested sludge           | HT, 80 °C, 60 min      | UASB         | Initial = 52, Operation = 5.2 | 52 | 2.45 mol H2/mol sugar consumed | 11.75 L H2/L POME day | Clostridium celerecrescens, Clostridium sp. and Proteobacteria | PCR-DGGE                   | Mahmod et al. (2019) |
### Table 4 (continued)

| Inoculum                          | Inoculum pre-treatment | Reactor type | Operating conditions | % H₂ | H₂ yield | H₂ production rate | Dominant microbes                              | Technique for microbial community analysis | References                  |
|-----------------------------------|------------------------|--------------|----------------------|------|----------|---------------------|-----------------------------------------------|------------------------------------------|------------------------------------------|
| Juice processing wastewater anaerobic sludge | HT, 105 °C, 30 min | Batch       | Initial = 60, Operation = uncontrolled | 55 °C | 23.7     | 77 mL H₂/g COD<sub>removed</sub> | ND                                           | ND                                      | ND                                      | Tanikkul et al. (2019a)               |
| Juice processing wastewater anaerobic sludge | HT, 105 °C, 30 min | Batch | Initial = 60, Operation = uncontrolled | 37 °C | 31       | 182 mL H₂/g COD or 7.96 mmol/g COD | 23.37 mL H₂/h | ND | ND                                      | Tanikkul et al. (2019b)               |
| POME anaerobic sludge            | HT, 90 °C, 60 min     | UASFF       | Initial = 50–5.2, Operation = uncontrolled | 37 °C | 57.11    | 1021.4 mL H₂/g COD<sub>consumed</sub> | 5.29 L H₂/L day | ND | ND                                      | Zainal et al. (2019)                  |
| POME anaerobic sludge            | HT, 100 °C, 60 min    | UASFF       | Initial = 55, Operation = uncontrolled | 50 °C | ND       | 28.47 mL H₂/g COD<sub>consumed</sub> | ND                                           | ND                                      | ND                                      | Zainal et al. (2018)                  |
| Sewage anaerobic sludge          | HT, 100 °C, 60 min    | UASFF       | Initial = 65, Operation = uncontrolled | 35 °C | 56.65    | 2.58 mmol H₂/g COD or 135.79 mL H₂/L POME | 11.32 mL H₂/L POME | ND | ND                                      | Garritano et al. (2017)               |
| POME sludge                      | HT, 80 °C, 60 min     | UASBR       | Initial = 60, Operation = 6.0 | 60 °C | ND       | 1.6 mol H₂/mol sugar | 61.5 mmol H₂/L day | Thermoanaerobacterium thermosaccharolyticum | 16S rRNA Identification | Jamali et al. (2017)                  |
| POME digested sludge             | HT, 80 °C, 60 min     | UASBR       | Initial = 58, Operation = uncontrolled | 60 °C | ND       | 1.24 mol H₂/mol glucose | 0.181 mmol H₂/L | ND | ND                                      | Mahmod et al. (2017)                  |
| POME digested sludge             | HT, 100 °C, 60 min    | UASFF       | Initial = 55, Operation = uncontrolled | 38 °C | 56.6     | ND       | 0.514 L H₂/g VSS | ND | ND                                      | Mohammadi et al. (2014)               |
| Immobilised POME sludge          | HT, 80 °C, 50 min     | UASB        | Initial = 55, Operation = 5.5 | 37 °C | 37.1     | ND       | 0.589 L H₂/L POME | ND | ND                                      | Singh et al. (2013a)                  |
Table 4 (continued)

| Inoculum                  | Inoculum pre-treatment | Reactor type | Operating conditions | % H₂ | H₂ yield | H₂ production rate | Dominant microbes                                                                 | Technique for microbial community analysis | References                  |
|---------------------------|------------------------|--------------|----------------------|------|----------|---------------------|----------------------------------------------------------------------------------|-------------------------------------------|-----------------------------|
| POME anaerobic sludge     | HT, 85 °C, 60 min      | ASBR         | 37 °C ND             | 50   | 940 mL H₂/g COD<sub> consumed</sub> | 6.7 L H₂/L day                  | Streptococcus macedonicus, Lactobacillus agilis and Clostridium butyricum GS56 | Conventional cultivation – 16S rDNA Identification | Badiei et al. (2012)        |
| Anaerobic sludge         | HT, 105 °C, 90 min     | Serum bottles| 44 °C Initial = 70, Operation = uncontrolled | ND   | 0.68 mmol H₂/g COD | ND                  | Clostridium spp. and Thermoaerobacterium spp.                                   | qPCR                                      | Leaño et al. (2012)         |
| Thermoanaerobacterium-rich sludge | ND                  | CSTR        | 60 °C Initial = 55, Operation = uncontrolled | ND   | 4.2 L H₂/L POME | ND                  | T. thermosaccharolyticum, T. aciditolerans                                      | PCR-DGGE                                  | Mamimin et al. (2012)       |
| POME digested sludge     | HT, 85 °C, 20 min      | Batch        | 36 °C Initial = 58, Operation = 5.8 | ND   | 1.32 L H₂/L POME | 0.144 L H₂/L h         | Clostridium paraputrificum, Weissella solii, C. butyricum, C. hydrogeniformans, C. beijerinckii and O. butyricum | PCR-DGGE                                  | Rasdi et al. (2012)         |
| POME anaerobic sludge     | HT, 90–95 °C, 30 min   | Serum bottles| 37 °C Initial = 60, Operation = uncontrolled | ND   | 27.09 mL H₂/g COD | 41.91 mL H₂/L h       | POME anaerobic sludge, Thermoaerobacterium thermosaccharolyticum, C. baratii and O. butyricum | PCR-DGGE                                  | Yossan et al. (2012)        |
| POME anaerobic sludge     | HT, 90–95 °C, 30 min   | Serum bottles| 55 °C Initial = 60, Operation = uncontrolled | ND   | 26.63 mL H₂/g COD | 49.34 mL H₂/L h       | Clostridium paraputrificum, C. butyricum, Thermoaerobacterium thermosaccharolyticum, C. beijerinckii and O. butyricum | PCR-DGGE                                  | Yossan et al. (2012)        |

POME: palm oil mill effluent; HT: heat treatment; ND: no data; ASBR: anaerobic sequencing batch reactor; CSTR: continuous stirred tank reactor; FBR: fluidised bed reactor; UASB: up-flow anaerobic sludge blanket reactor; UASFF: up-flow anaerobic sludge blanket fixed-film reactor; COD: chemical oxygen demand; PCR-DGGE: polymerase chain reaction denaturing gradient gel electrophoresis; qPCR: quantitative polymerase chain reaction.
thermophilic conditions were also investigated, using anaerobic sludge as inoculum. Higher biohydrogen yield was achieved from mesophilic fermentation (27.09 mL/g COD) while higher biohydrogen production rate was achieved under thermophilic condition (49.34 mL H₂/L POME/h). Microbial community analysis performed showed that Clostridia dominated all the biohydrogen production systems operated at 25, 37, 45 and 55 °C.

Different Clostridium species exhibit different metabolic activities, and their relative abundance vary depending on the operational conditions. Yossan et al. (2012) found that C. paraputrificum is the dominant member of the biohydrogen-producing community under all temperatures. In this study, C. butyricum was detected in the biohydrogen reactor operated at 37–55 °C, whereas C. beijerinckii and C. hydrogeniformans were only present at 37 °C. Biohydrogen production reactor at thermophilic condition was dominated by C. thermopalmarium, a non-celluolytic biohydrogen-producing bacteria (Yossan et al. 2012). In another study, C. sensu stricto contributed 800 mL H₂/g COD consumed of biohydrogen yield when treating POME with anaerobic sludge in up-flow anaerobic sludge blanket fixed-film (UASFF) reactor operated at 37 °C with the total abundance of 69.55% in the system (Akhhbari et al. 2021). C. celerecrescens was the dominant biohydrogen producer in up-flow anaerobic sludge blanket (UASB) reactor using POME substrate operated under thermophilic condition (Mahmod et al. 2019). Clostridia can also be the main biohydrogen producers even though they do not dominate the whole community. Badiei et al. (2012) performed microbial community analysis on the anaerobic sludge of an anaerobic sequencing batch reactor (ASBR) operating under mesophilic temperature. 940 mL H₂/g COD removed of biohydrogen was obtained in this system. Only 20% of the relative microbial abundance were represented by Clostridia. The community was dominated by Streptococcus (50% relative abundance) and Lactobacillus (30% relative abundance), and the biohydrogen yield was comparable with the yield obtained by Akhhbari et al. (2021) in a reactor dominated by Clostridium. This suggests that Clostridia does not have to dominate the system in order to obtain a high biohydrogen production yield. Deep metagenomics sequencing can help reveal the syntrophic relationship that may exist between Clostridia and the other genera not known to be the biohydrogen producers, and the connections between the different communities at different trophic levels in the reactor.

**Tools for biohydrogen microbiome analysis**

Biohydrogen production through dark fermentation from organic wastes, including POME, is a complex biochemical process, carried out by microbial communities with a range of relationships between them. Dark fermentation can be divided into four key stages which are hydrolysis, acidogenesis, acetogenesis and methanogenesis (section "Microbiomes in dark fermentative biohydrogen production"). Methanogenesis is often suppressed and undesired in biohydrogen production. These processes occur synergistically in a successive manner and each stage is facilitated by a distinct guild of microorganisms. A robust and efficient dark fermentation system requires a delicate balance of microbial population dynamics and metabolic activities among different guilds or trophic groups of the biohydrogen-producing microbiomes. Understanding of the microbial ecology of the dark fermentation process can help to improve the performance towards maximising biohydrogen production, and ensure that this process is economically feasible.

A range of techniques have been used in characterising the complex biohydrogen-producing microbial communities, from conventional cultivation-dependent approaches to cultivation-independent approaches. The advanced multi-omics technologies are also increasingly being use for this purpose. Cultivation-dependent method have contributed in the discovery of many key microbial species in biohydrogen-producing bioreactors from organic industrial waste and POME (Alvarado-Cuevas et al. 2015; Harun et al. 2012; Hsieh et al. 2016; Mishra et al. 2017; Noparat et al. 2012; Singh et al. 2014; Yin and Wang 2017; Zhang et al. 2015). While economical and a generally useful method to shed light on some key members, not many can be characterised this way, particularly when a system-based approach is required. Some key taxa also have their syntrophic partners belonging to different functional guilds (Lim et al. 2020). This method is further limited by species-specific morphological variations since some microorganisms share similar morphological, physiological or biochemical characteristics which makes the classification challenging (Lim et al. 2020). While cultivation method might be time consuming and labour intensive, it is the only technique to characterise a specific strain in detail. Current-omics technologies also require more reference genomes to evaluate the biohydrogen-producing microorganisms sequence data. Therefore, culture-dependent method will remain essential for studying the microbial diversity in biohydrogen-producing microbiomes. Recently, novel biohydrogen-producing bacteria, Clostridium sartagoforme NASGE 01 and Enterobacter cloacae NASGE 02 were isolated from sago industrial effluent using this method (Nizzy et al. 2020).

Advancement in molecular biology and DNA sequencing techniques has enabled various culture-independent methods to be used to study the microbiomes in biohydrogen-producing reactors. Denaturing gradient gel
electrophoresis (DGGE) and single-strand conformation polymorphism (SSCP) are among the microbiome fingerprinting techniques used to evaluate and compare different microbiomes in dark fermentation from organic wastes and POME. Both techniques involve polymerase chain reaction (PCR) amplification of a hypervariable region of the 16S rRNA gene and migration of the PCR product fragments on polyacrylamide gel that will provide different banding patterns, which reflect the structure of microbial communities and species abundance. Using PCR-DGGE, the genus *Megasphaera* sp. was identified as the main biohydrogen producer with 14% relative abundance, in thermophilic dark fermentation reactor of sugarcane stillage inoculated with granular sludge of a sugarcane stillage treatment plant. *Clostridia* were not detected in this system (Santos et al. 2014). While in dark fermentative biohydrogen production of beer lees inoculated with non-pre-treated garbage compost, using PCR-DGGE, *C. roseum* was found to be the prevalent biohydrogen producers in all high biohydrogen-producing batch fermentations, whereas *C. perfringens* and *C. sporogenes* were detected in low biohydrogen-producing batch fermentations (Bando et al. 2013). The presence of *Bifidobacterium* spp. and *Lactobacillus* spp. inhibited biohydrogen production through substrate competition with biohydrogen producers. Biohydrogen-producing species such as *C. butyricum* and *C. tyrobutyricum* were also found as the substrate competitors in biohydrogen fermenter dominated by *C. pasteurianum* (Lin et al. 2011). When POME was used as substrate, PCR-DGGE is still among the commonly used methods in studying the biohydrogen-producing microbiomes. The genus *Thermoanaerobacterium*, such as *T. thermosaccharolyticum*, was often reported as the main biohydrogen producers in thermophilic POME dark fermentation using POME anaerobic sludge as inoculum source (Jamali et al. 2011; Khongkliang et al. 2019; Maaroff et al. 2019).

The use of SSCP to investigate biohydrogen-producing microbial community structure is still limited. Using SSCP, operational pH of a continuous stirred tank reactor (CSTR) fed with glycerol was found to change the structure of the dominant microbial populations (Silva-Illanes et al. 2017). Hydraulic retention time (HRT) changed the metabolic pattern and the composition of subdominant microorganisms such as *Enterococcus, Prevotella*, *Sutterella*, *Pseudomonas* and *Acinetobacter*, ultimately affecting the ability of the consortium to produce biohydrogen. In general, DGGE and SSCP are not quantitative, more labour intensive, time consuming, prone to PCR biases and has low resolution in complex microbiome profiles (Kumar et al. 2018). Nevertheless, these microbiome fingerprinting techniques could remain useful for quick screening purposes, and to acquire a glimpse of biohydrogen-producing microbiomes from a large number of samples.

Quantitative PCR (qPCR) has also been used in studying several biohydrogen reactors using organic wastes (including POME) as substrates, to quantify the changes of specific microbial populations (Lay et al. 2010; Leaho et al. 2012; Pugazhendhi et al. 2017). In contrast to the common PCR which is qualitative, qPCR can accurately quantify the copy number of genes of interest in a sample by measuring the fluorescence of a specific probe used for amplification (Lim et al. 2020; Tolvanen and Karp 2011). This technique eliminates post-PCR target analysis, cheaper and offers a fast, accurate and simple approach for high-throughput analysis (Nurmi et al. 2002). Individual taxa or guilds in biohydrogen microbiomes can also be quantified using fluorescent in situ hybridisation (FISH) technique. In FISH, cells of interest is hybridised with a specific fluorogenic oligonucleotide probes and its relative abundance is then measured by quantifying the ratio of the hybridised cells to the total cell count using a fluorescence microscope. FISH probes Tbm1282, Ccs432 and Tbmthsac184 specific for detection of *Thermoanaerobacterium, Caldicellulosiruptor* and *T. thermosaccharolyticum* have been designed and used to assess the microbial composition in thermophilic and extreme thermophilic biohydrogen-producing reactors fed with POME, lignocellulosic hydrolysate and synthetic sugars (O-Thong et al. 2008). FISH overcomes the limitations of PCR-based molecular techniques. Nevertheless, cell hybridisation is time consuming, making FISH less suitable for high-throughput community structure investigation (Ravenschlag et al. 2001). Detection of novel microorganisms may also be challenging, the probe design and selection require some information on the community structure prior to the analysis (Lim et al. 2020).

High throughput next generation -omics technologies are increasingly being employed to better understand the complex microbiomes driving dark fermentative biohydrogen production. Amplicon sequencing, metagenomics and metaproteomics have all been employed in this context. Amplicon sequencing also known as metaprofiling is a culture-independent technique to profile the taxonomic diversity, structure and composition of a microbiome based on a marker gene (Escobar-Zepeda et al. 2015). 16S rRNA genes have been exclusively used as a marker gene for library preparation through PCR amplification in studies of microbial communities, including biogas-producing microbiomes (Sharpton 2014; Tonge et al. 2014). Using amplicon sequencing method, biohydrogen production using sake lees was found to be enhanced when the microbial community in the system changed from *Bacillus muralis* and *B. cereus* as the
dominant taxa, to *Pantoea agglomerans*, *C. acetobutylicum* and *C. butyricum* (Choiron et al. 2020). Besides, with amplicon sequencing, *Sporolactobacillus* was the dominant taxa with relative abundance 97% in the fermentation of POME using microbial consortia from sugarcane cultivation soil (Rosa et al. 2020). *Sporolactobacillus* is an anaerobic facultative bacterium producing lactic acid. Although its role in biohydrogen production is unknown, its metabolic by-products could be used as substrates for biohydrogen production by other microorganisms. Amplicon sequencing is commonly done on an Illumina MiSeq platform (Akhbari et al. 2021; Audu et al. 2021; Martinez-Burgos et al. 2020; Yang and Wang 2019) while Ion Torrent platform has also been used in several studies (Cho et al. 2018; Oliveira et al. 2020). A few studies have also attempted to predict the community functions from amplicon sequencing of biohydrogen microbiomes using bioinformatic tools, such as PICRUSt (Li et al. 2020; Yin and Wang 2021) mostly using other organic wastes. So far, this has not been reported for POME. Amplicon sequencing is the best and economical option to understand the microbial community members in general, but it has limitations and may result in biases (Lim et al. 2020), leading to the increasing applications of shotgun metagenomics.

 Shotgun metagenomics independently sequences total genomic DNA retrieved directly from a sample to produce reads that align to various genomic locations for the countless genomes present, including the non-microbes (Sharpton 2014). Metagenomic tools could unravel the vast taxonomic diversity, metabolic function potential and physiology of uncultivated microorganisms, including the novel and rare taxa, and previously unknown metabolic pathways (Vanwonterghem et al. 2014). A few studies have investigated the biohydrogen-producing microbiomes using metagenomics (Mazareli et al. 2020; Soares et al. 2018; Villa Montoya et al. 2020). Mazareli et al. (2020) used metagenomics to correlate taxonomic diversity of indigenous microbial biomass with the performance of biohydrogen reactor fed with banana wastes under mesophilic temperature. Using FMAP (Functional Mapping and Analysis Pipeline) for metagenomic and metatranscriptomic studies tool, *Clostridium* and *Lactobacillus* were the dominant indigenous acidogenic bacteria, and the main genes encoding key enzymes involved in the fermentation were found to be related to carbohydrate metabolism, acidogenesis and biohydrogen production enzymes such as glucose-6-phosphate dehydrogenase, fructokinase, lactate dehydrogenase and pyruvate ferredoxin oxidoreductase. Metagenomic study by Villa Montoya et al. (2020) reported domain Bacteria represented 97.2% relative abundance with the predominance of genera *Clostridium* (87.9% relative abundance) in the mesophilic biohydrogen-producing bioreactor fed with coffee wastes. Gene identifications showed that 8.3% of the genes were correspondence to anaerobic degradation enzymes mainly for the production of organic acids and alcohols and may be associated with the metabolic potential of *Clostridium* sp. In addition, 37 KEGG orthologues (KO) were identified to be associated with biohydrogen production, highlighting enzymes pyruvate-ferredoxin oxidoreductase, anaerobic carbon-monoxide dehydrogenase, formate dehydrogenase and ferredoxin hydrogenase. Genes related to these enzymes were mainly found in *Clostridium* sp. (Villa Montoya et al. 2020).

 Breakthroughs in next generation sequencing (NGS) technologies has also led to another subfield of -omic technologies, which is metaproteomics. Metaproteomics profiles enzymes and proteins in microbiomes, and can potentially link the function of a protein to a taxon and its metabolic activities (Chistoserdova 2009; Lim et al. 2020). Metaproteomics has been widely applied in studying anaerobic digester bioreactors and human gut microbiome, but its application is still limited in investigating the microbiomes of biohydrogen dark fermentation. Previously, metaproteomics was used to establish the relationship between phylogeny, function, and metabolic activity of biohydrogen and methane co-production microbiomes from food waste (Jia et al. 2017). A total of 651 bacterial proteins and 477 archaeal proteins were detected in the study, revealing the complexity and metabolic diversity during the biogas production process. The study also revealed that the key bacterial proteins from *Gammmaproteobacteria*, *Clostridia* and *Bacilli* related to biohydrogen production came from pyruvic acid decarboxylase and formic acid decomposition pathway in carbohydrate metabolisms.

 **Future outlook**

 Biogas (i.e. methane) microbiomes are more widely and intensively studied than biohydrogen, despite the fact that these two processes share many biochemical and metabolic routes. This is probably due the more advanced research and wider adoption of anaerobic digestion for biogas production as cleaner energy production technology in the society. It has been demonstrated that methane production is directly linked to the composition of the anaerobic digester microbiomes, in addition to the microbial metabolism, which is dependent on the environmental parameters of the reactor (Campanaro et al. 2020). This makes understanding of the microbial composition of a bioreactor and their behaviour a critical aspect in the quest for a feasible biohydrogen production via dark fermentation. Pugazhendhi et al. (2019) reviewed the microbiomes involved in the anaerobic hydrogen-producing
granules (HPG). Granulation increases the reaction efficiency of a fermenter, compared to using sludge. The dominant taxa in the microbial community of reactor systems employing HPG has been discussed, allowing the monitoring of the microbial species for easier control of the kinetic parameters, and contributes to the development of stable bioprocess system (Pugazhendhi et al. 2019). This suggests the importance of meta-analysis of hydrogen-producing microbial community from different reactor systems, and the correlation with their physicochemical parameters and reactor performance.

A summary of the molecular tools in analysing the biohydrogen-producing microbial community has recently been published (Kumar et al. 2020), describing the “targeted” molecular tools (e.g. FISH, qPCR) and the advantages of NGS in providing quicker and more comprehensive investigation. A combination of culture-dependent approach, targeted molecular tools and NGS, and multi-omics are definitely the way forward in providing a system-based understanding of the biohydrogen microbiomes. Multi-omics of this engineered reactor system can also benefit from the rapidly expanding experimental and computational tools for investigating human and environmental microbiomes, allowing for deeper understanding of the community structure and functions from the -omics data. This includes the advancements in co-occurrence network, genome-scale metabolic model, protein–protein interaction network, the metabolic-driven metabolomics network (Liu et al. 2020), and the integration of all the -omics data. This is in addition to the need for best practices for analysing the microbiomes towards a unified approach in the analysis of reactor systems.

Knowledge obtained from the -omics techniques can be used to engineer a desired community structure, towards maximising productivity of an engineered system, and balancing the effects of any perturbations. Tools for manipulating community structure in situ are also being investigated. CRISPR/Cas-related system has been used in a targeted genome editing of specific microorganisms within a complex microbial community (Rubin et al. 2020), paving the way for manipulation of microbiomes in many different applications, possibly in the biogas and biohydrogen-producing reactor system too. There is still a long way to go before this precise gene and genome manipulation system can be applied in a complex community like the anaerobic digester’s, but it is important to first have the full understanding of the microbial community and the relationships with the physicochemical parameters in controlling the production yield and rate.

Conclusion
Biohydrogen is a common by-product of many bacterial metabolic pathways during dark fermentation. Microbial communities involved in dark fermentation are phylogenetically and functionally diverse which contribute to biohydrogen production from the breakdown of complex organic substrates, such as POME and other industrial wastes. As a system which relies on microbial metabolisms, insights on the microbial members present in the reactor is important towards obtaining a robust and efficient biohydrogen production system. Numerous molecular tools for screening, quantification and identification of biohydrogen-producing microbial communities have been used to correlate the phylogeny, interspecies interactions and their function to dark fermentative biohydrogen process. Currently, DGGE and amplicon sequencing are widely used in the study of biohydrogen microbiomes. The use of -omics technologies in biohydrogen research are still relatively limited, compared to the more widely investigated anaerobic digester’s microbiomes for biomethane production. We believe similar advanced tools can be applied to biohydrogen-producing reactors too, with the prospect to unravel the limitless potential of the microbial members in the system.

Abbreviations
POME: Palm oil mill effluent; rRNA: Ribosomal RNA; RDP: Ribosomal Database Project; NCBI: National Center for Biotechnology Information; MAGs: Metagenome-assembled genomes; MiDAS: Microbial Database for Activated Sludge; ASV: Amplicon sequence variant; CO2: Carbon dioxide; H2: Hydrogen; CH4: Methane; SCFA: Short-chain fatty acid; PFL: Pyruvate formate lyase; Fd: Ferredoxin; PFOR: Pyruvate ferredoxin oxidoreductase; NADH: Nicotinamide adenine dinucleotide; NFOR: Nicotinamide adenine dinucleotide ferredoxin oxidoreductase; ATP: Adenosine triphosphate; BOD: Biological oxygen demand; COD: Chemical oxygen demand; UASFF: Up-flow anaerobic sludge blanket fixed-film; UASB: Up-flow anaerobic sludge blanket; ASBR: Anaerobic sequencing batch reactor; DGGE: Denaturing gradient gel electrophoresis; SSCP: Single-strand conformation polymorphism; PCR: Polymerase chain reaction; CSTR: Continuous stirred tank reactor; HRT: Hydraulic retention time; qPCR: Quantitative polymerase chain reaction; FISH: Fluorescent in situ hybridisation; FMAP: Functional Mapping and Analysis Pipeline; KOs: KEGG orthologues; NGS: Next generation sequencing; HPG: Hydrogen-producing granules.

Acknowledgements
Not applicable.

Authors’ contributions
ELND and MFA-W designed the content. ELND and JOA performed literature search and analysis. JOA designed the figures in the manuscript. ELND, JOA and MFA-W wrote the manuscript. MFA-W and WZWD provided expertise and revised the manuscript. All authors read and approved the final manuscript.

Funding
ELND would like to acknowledge The Ministry of Higher Education (MOHE) Malaysia for providing financial support through MyBrainSc scholarship. JOA would like to acknowledge the Nigerian Tertiary Education Trust Fund for the scholarship provided. MFA-W would like to acknowledge the funding received from UTM via Transdisciplinary Research Grant no. 05G24 and UTMShine Grant No. 09G86.
Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that no competing interests involved.

Author details
1Department of Biosciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia. 2Department of Science Laboratory Technology, Modibbo Adama University, PMB 2076, Yola, Adamawa, Nigeria. 3Taiwan-Malaysia Innovation Centre for Clean Water and Sustainable Energy (Wise Centre), Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia.

Received: 6 December 2021 Accepted: 14 February 2022 Published online: 05 March 2022

References
Abdullah MF et al (2020) Effect of carbon/nitrogen ratio and ferric ion on the production of biohydrogen from palm oil mill effluent (POME). Biocatal Agric Biotechnol 23:101445. https://doi.org/10.1016/j.bcab.2019.101445
Abendroth C et al (2015) Eubacteria and archaea communities in seven mesophilic anaerobic digester plants in Germany. Biotechnol Biofuels 8:1–10. https://doi.org/10.1186/s13068-015-0271-6
Akbari A et al (2021) Start-up study of biohydrogen production from palm oil mill effluent in a lab-scale up-flow anaerobic sludge blanket fixed-film reactor. Int J Hydrogen Energy. https://doi.org/10.1016/j.ijhydene.2020.12.125
Alvarado-Cuevas ZD et al (2015) Biohydrogen production using psychrophilic bacteria isolated from Antarctica. Int J Hydrogen Energy 40:7586–7592. https://doi.org/10.1016/j.ijhydene.2014.10.063
Alvarez-Guzmán CL et al (2020) Biohydrogen production from cheese whey powder by Enterobacter asburiae: effect of operating conditions on hydrogen yield and chemometric study of the fermentative metabolites. Energy Rep 6:1170–1180. https://doi.org/10.1016/j.egyr.2020.04.038
Asadi N, Zilouei H (2017) Optimization of organosolv pretreatment of rice straw for enhanced biohydrogen production using Enterobacter aerogenes. Biores Technol 227:335–344. https://doi.org/10.1016/j.biortech.2016.12.073
Audu JO et al (2020) Dark fermentation and bioelectrochemical systems for enhanced biohydrogen production from palm oil mill effluent: current progress, potentials, and future perspectives. In: Zakaria ZA, Boopathy R, ed) Biocatal approaches. Springer, Cham, pp 1–35
Audu JO et al (2021) Optimization of the operational parameters for mesophilic biohydrogen production from palm oil mill effluent using enriched mixed culture. Biomass Convers Biorefinery. https://doi.org/10.1007/s13399-021-01488-9
Azman NF et al (2016) Biohydrogen production from de-oiled rice bran as sustainable feedstock in fermentative process. Int J Hydrogen Energy 41:145–156. https://doi.org/10.1016/j.ijhydene.2015.10.018
Badiei M et al (2012) Microbial community analysis of mixed anaerobic microflora in suspended sludge of ASBR producing hydrogen from palm oil mill effluent. Int J Hydrogen Energy 37:3169–3176. https://doi.org/10.1016/j.ijhydene.2011.11.063
Bando Y et al (2013) A microbiological study of biohydrogen production from beer lees. Int J Hydrogen Energy 38:2709–2718. https://doi.org/10.1016/j.ijhydene.2012.11.142
Beckers L et al (2013) Improving effect of metal and oxide nanoparticles encapsulated in porous silica on fermentative biohydrogen production by Clostridium butyricum. Biores Technol 133:109–117. https://doi.org/10.1016/j.biortech.2012.12.168
Berg G et al (2020) Microbiome definition re-visited: old concepts and new challenges. Microbiome 8:1–22. https://doi.org/10.1186/s40168-020-0087-5
Bisallion A et al (2006) The effect of nutrient limitation on hydrogen production by batch cultures of Escherichia coli. Int J Hydrogen Energy 31:1504–1508. https://doi.org/10.1016/j.ijhydene.2006.06.016
Cabral L et al (2017) Microbial ecology of fermentative hydrogen producing bioprocesses: useful insights for driving the ecosystem function. FEMS Microb Rev 41:158–181. https://doi.org/10.1093/femsre/fuw043
Campanaro S et al (2016) Metagenomic analysis and functional characterization of the biogas microbiome using high throughput shotgun sequencing and a novel binning strategy. Biotechnol Biofuels 9:1–17. https://doi.org/10.1186/s13068-016-0441-1
Campanaro S et al (2020) New insights from the biogas microbiome by comprehensive genome-resolved metagenomics of nearly 1600 species originating from multiple anaerobic digesters. Biotechnol Biofuels 13:25. https://doi.org/10.1186/s13068-020-01679-y
Castellano-Hinojosa A et al (2018) New concepts in anaerobic digestion processes: recent advances and biological aspects. Appl Microb Biotechnol 102:5065–5076. https://doi.org/10.1007/s00253-018-8939-9
Chang S et al (2011) Evaluation of different pretreatment methods for preparing hydrogen-producing seed inocula from waste activated sludge. Renew Energy 36:1517–1522. https://doi.org/10.1016/j.renene.2010.11.023
Chen Y et al (2021) Comparison of fermentative hydrogen production from glycerol using immobilized and suspended mixed cultures. Int J Hydrogen Energy. https://doi.org/10.1016/j.ijhydene.2021.01.003
Chia WY et al (2020) Outlook on biorefinery potential of palm oil mill effluent for resource recovery. J Environ Chem Eng 8:104519. https://doi.org/10.1016/j.jece.2020.104519
Chiariotti A, Cristà A (2018) Bio-hydrogen production from buffalo waste with rumen inoculum and metagenomic characterization of bacterial and archaeal community. Front Sustain Food Syst. https://doi.org/10.3389/fsufs.2018.00013
Chistoserdova L (2009) Functional metagenomics: recent advances and future challenges. Biotechnol Genet Eng Rev 26:335–352. https://doi.org/10.1007/978-1-4419-06561/bger-26-335
Cho S-K et al (2018) Effects of low-strength ultrasonication on dark fermentative hydrogen production: start-up performance and microbial community analysis. Appl Energy 219:34–41. https://doi.org/10.1016/j.apenergy.2018.03.047
Choirion M et al (2020) Biohydrogen production improvement using hot compressed water pretreatment on sake brewery waste. Int J Hydrogen Energy 45:17220–17232. https://doi.org/10.1016/j.ijhydene.2020.04.199
Chong ML et al (2009) Biohydrogen production by Clostridium butyricum EB6 from palm oil mill effluent. Int J Hydrogen Energy 34:764–771. https://doi.org/10.1016/j.ijhydene.2008.10.095
Conklin A et al (2006) Growth kinetics and competition between Methanoarcina and Methanoseta in mesophilic anaerobic digestion. Water Environ Res 78:486–496. https://doi.org/10.2175/106143006x5393
Costa JB et al (2011) The optimization of biohydrogen production by bacteria using residual glycerol from biodiesel synthesis. J Environ Sci Health Part A 46:1461–1468. https://doi.org/10.1080/10934529.2011.69036
Dada O et al (2013) Biohydrogen production from rice bran using Clostridium saccharoperbutylacetonicum N1–4. Int J Hydrogen Energy 38:15063–15073. https://doi.org/10.1016/j.ijhydene.2013.07.048
Das D (2017) A road map on biohydrogen production from organic wastes. INAE Letters 2:153–160. https://doi.org/10.1007/s41403-017-0031-y
Dueholm MS et al (2021) MiDAS 4: a global catalogue of full-length 16S rRNA gene sequences and taxonomy for studies of bacterial communities in wastewater treatment plants. bioRxiv. https://doi.org/10.1101/2021.07.06.451231
Engel I et al (2020) Biohydrogen production beyond the Thauer limit by precision design of artificial microbial consortia. Commun Biol 3:443. https://doi.org/10.1038/s42003-020-01159-x
