Oecophyllibacter saccharovorans gen. nov. sp. nov., a bacterial symbiont of the weaver ant *Oecophylla smaragdina* with a plasmid-borne sole *rrn* operon

Kah-Ooi Chua[a], Wah-Seng See-Too[a], Jia-Yi Tan[a], Sze-Looi Song[b, c], Hoi-Sen Yong[a], Wai-Fong Yin[a], Kok-Gan Chan[a,d*]

[a] Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

[b] Institute of Ocean and Earth Sciences, University of Malaya, 50603 Kuala Lumpur, Malaysia

[c] China-ASEAN College of Marine Sciences, Xiamen University Malaysia, 43900 Sepang, Selangor, Malaysia

[d] International Genome Centre, Jiangsu University, Zhenjiang, China

[*] Corresponding author at: Institute of Biological Sciences, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
Email address: kokgan@um.edu.my

Keywords:

Acetobacteraceae, Insect, Whole genome sequencing, Phylogenomics, Plasmid-borne *rrn* operon

Abstract

In this study, acetic acid bacteria (AAB) strains Ha5^T, Ta1 and Jb2 that constitute the core microbiota of weaver ant *Oecophylla smaragdina* were isolated from multiple ant colonies and were distinguished as different strains by matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry and distinctive random-amplified polymorphic DNA (RAPD) fingerprints. These strains showed similar phenotypic characteristics and were considered a single species by multiple delineation indexes. 16S rRNA gene sequence-based phylogenetic analysis and phylogenomic analysis based on 96 core genes placed the strains in a distinct lineage in family *Acetobacteraceae*. Compared to *Acetobacteraceae* type members, these strains demonstrate average nucleotide identity (ANI), *in-silico* DNA-DNA hybridization (DDH) and average amino acid identity (AAI) values lower than proposed species-level cut-off values which indicated that they represent a novel genus of the family. Currently, strains Ha5^T, Ta1 and Jb2 possess the smallest genomes (1.92-1.95 Mb) with significantly lower gene and protein numbers among family *Acetobacteraceae*. Intriguingly, the sole *rrn* operon in their genomes is plasmid-borne instead of chromosomally located.

Furthermore, these strains harbour biosynthetic genes for various amino acids, cofactors and vitamins which supported a nutritional symbiotic interaction with the host *O. smaragdina*. Various phenotypic differences also distinguished the strains from closest relative genera in family *Acetobacteraceae*. Based on these results, these strains represent a novel species of a novel genus of family *Acetobacteraceae*, for which we propose the name *Oecophyllibacter saccharovorans* gen. nov. sp. nov., and strain Ha5^T as the type strain.

†The GenBank/EMBL/DDBJ accession numbers of the 16S rRNA gene sequences of strains Ha5^T, Ta1 and Jb2 are MG757796, MN540264 and MN540265, respectively. The Whole Genome Shotgun project of strain Ha5^T, Ta1 and Jb2 have been deposited at GenBank/EMBL/DDBJ under the accession numbers CP038143.1, SORY00000000 and SORZ00000000, respectively.

Introduction

The family *Acetobacteraceae* within the order Rhodospirillales consists of mostly acetic acid bacteria (AAB) that are Gram-negative and obligate aerobic bacteria [1]. These bacteria are known for their inability to oxidize carbohydrates and alcohols completely which leads to accumulation of partially oxidized metabolic products in the growth medium [2]. Their biochemical characteristics make them commercially important for manufacture of vinegar, foods and different chemical compounds [3].

In addition to AAB, the members of family *Acetobacteraceae* are widespread in the environment. At the time of writing, there are 39 validly published genera (http://www.bacterio.net/*Acetobacter*aceae.html) in the family that have been isolated from various sugary, alcoholic or acidic habitats including fermented food, alcoholic beverages [3], human patient [4], plants [5] and insects [6]. AAB from genera *Gluconobacter*, *Acetobacter* and *Asaia* have been cultivated from fruit fly *Drosophila melanogaster*, *Bactrocera oleae* and mosquito *Anopheles stephensi* respectively [7-9]. Notably, a novel AAB genus *Bombella* was recently isolated from crops of multiple bee species [10, 11].

It is believed that symbiotic interaction with bacteria is one of the key attributes to remarkable adaptability of insects to wide range of terrestrial habitats. In insects that survive on protein-poor diets such as honeydew and plant sap, the symbiotic bacteria provide amino acids as nutritional support and contribute to fitness of the host [12]. The use of high-throughput sequencing technology revealed a diverse group of AAB constituting insect-associated microbial community. These AAB are known to survive acidic environment in insect guts and tolerate sugar-rich diets of the host [13]. It is also noteworthy that long term symbiosis of these AAB with different insect hosts resulted in significantly reduced size and gene content of their genomes [14, 15].

The weaver ant *Oecophylla smaragdina* (Fabricius, 1775) is an obligate arboreal and omnivorous species widespread from Southeast Asia to northern Australia [16]. Their ability to build nest directly from leaves still attached on host trees enables them to survive on a large number of host plant species. In addition to feeding on sugar-rich plant exudates such as extrafloral nectar and honeydews, the highly aggressive major workers of *O. smaragdina* prey on a wide range of intruding arthropods including insect pests [17]. Their territorial behaviour has been exploited in biological control of many insect pests on tropical tree crops such as mango [18], cocoa [19] and cashew [20]. While its role as biological control agent continues to expand, the study about microbiota of *O. smaragdina* is still limited.

By 16S rRNA gene amplicon sequencing, our previous study has shown that several operational taxonomic units (OTUs) of the family *Acetobacteraceae* were consistently detected in microbiota of *O. smaragdina* [21]. These OTUs were assigned to genus *Neokomagataea* and were

categorized as core members of the *O. smaragdina* microbiome. In this study, we attempt to cultivate the *Acetobacteraceae* from *O. smaragdina* and report the taxonomic characterization of three novel strains using polyphasic approaches. In addition, we sequenced the genomes of the strains to elucidate their phylogenetic relationships and genome characteristics. Most interestingly, our findings revealed that the strains possess the smallest genomes (1.92-1.95 Mb) among *Acetobacteraceae* and a rare genome organization with the sole *rrn* operon located on a 6.5 kb plasmid rather than the chromosome.

Materials and methods

Ant sample collections, AAB isolation and cultivation

Major workers of *O. smaragdina* were caught in the field in June and July 2017 from 3 distantly separated locations in Petaling Jaya (Selangor, Malaysia). They were kept alive in sterile container and promptly transferred to laboratory in the University of Malaya (Kuala Lumpur, Malaysia). The ants were cold-anesthetized at 4°C for 5 min, surface sterilised with 70% v/v ethanol for 1 min and rinsed with sterile distilled water. The gasters of 5 ants were detached and homogenized in 500 µl 1× phosphate-buffered saline (PBS) (pH 7.4, 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4) with a sterile micro-pestle [22].

Three different pre-enrichment media were used for isolation of AAB from the homogenates. Enrichment medium A was composed of 2.0% (w/v) D-glucose, 0.5% (v/v) ethanol, 0.5% (w/v) peptone, 0.3% (w/v) yeast extract and 0.01% (w/v) cycloheximide in water, and pH of the medium was adjusted to pH 3.5 with hydrochloric acid (HCl) [5]. Enrichment medium B was composed of 2.0% (w/v) D-glucose, 0.5% (v/v) ethanol, 0.3% (v/v) acetic acid, 1.5% (w/v) peptone, 0.8% (w/v) yeast extract, 0.01% (w/v) cycloheximide in water, and adjusted to pH 3.5 with HCl [23]. Enrichment medium C was composed of 2.0% (w/v) D-sorbitol, 0.5% (w/v) peptone, 0.3% (w/v) yeast extract, and 0.01% (w/v) cycloheximide in water, and adjusted to pH 3.5 with HCl [24]. To avoid thermal decomposition due to autoclave-sterilization, D-glucose, D-sorbitol, ethanol, acetic acid and cycloheximide were first prepared in distilled water and filtered-sterilized by membrane-filter (pore size of 0.2µm) and added into autoclave-sterilized media cooled down to 50°C. Enrichment media inoculated with the homogenates were shaken at 220 r.p.m. and incubated at 28 °C for 7 days.

Enrichment media observed with microbial growth were then plated onto solid agar medium D containing 2.0% (w/v) D-glucose, 1.0% (v/v) glycerol, 0.5% (v/v) ethanol, 0.5% (w/v) peptone, 0.8% (w/v) yeast extract, 0.7% (w/v) calcium carbonate and 1.5% (w/v) agar in water [23] . Inoculated agar plates were incubated aerobically at 28 °C for 7 days. Colonies of distinctive morphologies on the agar were selected and pure cultures of isolates were obtained. The isolates were routinely maintained on LMG medium 404 (5% (w/v) D-glucose; 1% (w/v) yeast extract and 1.5% (w/v) agar) and frozen glycerol stocks (with 20% (v/v) glycerol) were maintained in – 80 °C.

Reference strains of closely related genera in family *Acetobacteraceae* were included in most analyses. The type strains *Bombella intestini* DSM 28636^T and *Saccharibacter floricola* DSM 15669^T were acquired from Leibniz Institute-Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) culture collection. *Neokomagataea thailandica* NBRC 106555^T, *N. tanensis* NBRC 106556^T and *Swingsia samuiensis* NBRC 107927^T were acquired from NITE Biological Resource Center (NBRC). Identity of all type strains was confirmed by 16S rRNA gene analysis and frozen glycerol stocks were maintained in – 80 °C.

Matrix-assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) Mass Spectrometry (MS) and Main Spectrum Profile (MSP) analysis of bacterial isolates

Isolates were cultured on LMG medium 404 for 2 days at 28 °C before being used for analysis of MSP using MALDI-TOF MS. The cultures were subjected to ethanol-formic acid protein extraction according to the MALDI Biotyper protocol (Bruker Daltonics GmbH, Bremen, Germany). Only 1 µl of supernatant generated was applied on wells of MSP 96 MALDI Target plate and dried at room temperature. The well was then flooded with 1 µl of HCCA (saturated α-cyano-4-hydroxycinnamic acid in 50% acetonitrile – 2.5% trifluoroacetic acid) matrix solution and let dried. The plate was then subjected to Microflex MALDI-TOF bench-top mass spectrometer (Bruker Daltonik GmbH, Leipzig, Germany) analysis using FlexControl software version 3.3. The generated spectra of the isolates was analysed using Bruker MALDI Biotyper Real Time Classification (RTC) (Version 3.1) by comparison to the best match in the Bruker database. Identity of each isolate was evaluated and scored. Similarity scorings of these mass spectra were used for construction of a dendrogram using MALDI Biotyper MSP creation method (Bruker Daltonics, Bremen, Germany) which clustered closely related isolates and illustrated graphical distance values between them [25].

DNA extraction

Genomic DNA (gDNA) extraction was performed using MasterPure DNA Purification Kit (EpiCenter, CA, USA) according to the manufacturer's protocol. Extracted gDNA was eluted in Elution Buffer. The quality and quantity of gDNA extracted were examined by NanoDrop 2000 UV-Vis spectrophotometer (Thermo Scientific, MA, USA) and Qubit 2.0 fluorometer (Life Technologies, MA, USA), respectively.

Phylogenetic analysis

From the cluster analysis of MALDI-TOF, representative bacterial isolates from each cluster were selected for 16S rRNA gene sequence analysis. 16S rRNA genes of the isolates were amplified with primers 27F (5' – AGAGTTTGGATCMTGGCTCAG – 3') and 1492R (5' – CGGTTACCTTGTTACGACTT – 3') [26]. Amplicons were sent for commercial Sanger sequencing (Apical Scientific, Selangor, Malaysia). The assembled sequences were analysed by NCBI BLAST tool and EzBioCloud server (http://www.ezbiocloud.net/eztaxon) [27] for identification. Alignment of 16S rRNA gene sequences with phylogenetically nearest genera and species were performed using ClustalW. Phylogenetic trees were constructed in MEGA7 software [28] by neighbour-joining and

maximum-likelihood methods. The robustness of tree topologies were estimated by bootstrap analysis (1000 replicates). The 16S rRNA gene sequences of strains Ha5^T (MG757796), Ta1 (MN540264) and Jb2 (MN540265) have been submitted to GenBank (http://www.ncbi.nlm.nih.gov).

Random Amplified Polymorphic DNA (RAPD) fingerprinting

RAPD for genotypic differentiation of the bacterial isolates in MALDI-TOF cluster 2 was performed as described by Krzewinski et al. [29] using RAPD primers 270 (5' – TGCGCGCGGG – 3') and 272 (5' – AGCGGGCCAA – 3'). The RAPD-PCR mixtures contains 100 ng of gDNA of each bacterial isolate as template, 0.45 µM of either primer 270 or primer 272, 200 µM of each dNTPs and 2.5 U of OneTaq™ DNA Polymerase (New England Biolabs) in total of 25 µl reaction volume. Amplification was carried out in a Veriti Thermal Cycler (Applied Biosystems) using the following conditions: (i) 1 cycle of 15 min at 95°C; (ii) 4 cycles of 5 min at 94°C, 5 min at 36°C, and 5 min at 72°C; and (iii) 30 cycles of 1 min at 94°C, 1 min at 36°C, and 1 min at 72°C, followed by (iv) a final extension step at 72°C for 10 min. RAPD amplicons were separated by 1% w/v agarose electrophoresis. RAPD fingerprints generated were compared in GelJ [30] and a cluster analysis using Unweighted Pair Group Method with Arithmetic mean (UPGMA) method was performed with similarity matrix calculated with the Pearson correlation.

Whole genome sequencing and genome analysis

Extracted gDNA of strain Ha5^T were prepared into barcoded and multiplexed 10kb SMRTbell libraries using SMRTbell Template Prep Kit 1.0-SPv3. Complete genomes of the strains were sequenced by single-molecule real-time sequencing (SMRT) in 1 Sequel SMRT Cell 1M v2 on the PacBio Sequel system (Pacific Biosciences, Menlo Park, CA). The generated subreads were assembled using Hierarchical Genome Assembly Process (HGAP) version 4 with optimized parameters [31] on SMRTLink version 5.1.0.26412. Overlapping regions at both ends of the assembled genomes were observed using Contiguity [32]. The genomes were assembled into circular contigs with complete closure using Minimus2 pipeline in the AMOS software package [33].

The gDNA from strains Ta1 and Jb2 were constructed into sequencing library using the Nextera DNA Library Prep Kit according to manufacturer's protocol and sequenced on Illumina HiSeq 2000 platform by NovogeneAIT Genomics (https://en.novogene.com). Quality filtering of the reads and *de novo* genome assemblies were performed using CLC Genomics Workbench version 7.0.4.

All assembled genomes were annotated using NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 2.10 [34] and Rapid Annotation using Subsystem Technology (RAST) version 3.0 [35]. The DNA G + C contents were determined from their whole genome sequences.

Calculations of Average Nucleotide Identity (ANI), *in silico* DNA-DNA hybridization (DDH), Average Amino acid Identity (AAI), Percentage of Conserved Protein (POCP) and phylogenomic analysis

To elucidate the evolutionary relationships of strains Ha5^T, Ta1, Jb2 and closely related genera in family *Acetobacteraceae*, whole genome sequences available in GenBank were acquired for ANI and *in silico* DDH analyses. The accession numbers of the genomes used in this study are shown in Supplementary Table S1. 16S rRNA gene identification was performed on the genomes using Eztaxon server for accurate species identification [27] and genome completeness was checked with BUSCO v.3.0.2 [36]. ANI values between genomes were determined using Orthologous Average nucleotide identity Tool (OAT) [37]. *In silico* DDH was calculated using Genome-to-Genome Distance Calculator (GGDC 2.0) (http://ggdc.dsmz.de/distcalc2.php) with the BLAST+ method [38] and the results based on formula 2 were obtained for further analysis. AAI between the genomes was determined using CompareM (https://github.com/dparks1134/CompareM) [39]. POCP between the genomes was calculated as reported by Qin et al. [40].

Phylogenomic analyses were performed on available genomes of family *Acetobacteraceae* members listed in Supplementary Table S1 by using panX for core-genome analysis [41]. Genomes with ambiguous 16S rRNA gene identity and BUSCO genome completeness lesser than 90% were excluded from analysis (Supplementary Table S1). With panX [41], the groups of homologous genes were identified by similarity search using DIAMOND and rRNA genes were compared by BLASTn, before creating clusters of putatively orthologous genes with Markov Clustering Algorithm (MCL). The core genes were aligned using MAFFT 1.3.7. The best-fit substitution model for each nucleotide sequence of the core-genes were determined using ModelTest-NG. The whole-genome phylogeny was constructed using RAxML-NG, by applying CAT modelling rate heterogeneity, and the robustness of branching was estimated with 100 bootstrap replicates.

Physiological and biochemical characterization

Phenotypic analyses necessary for taxonomic descriptions of new taxa in family *Acetobacteraceae* were performed on strains Ha5^T, Ta1, Jb2 and all acquired reference strains under similar conditions. All strains were observed under the Leica DM750 microscope (Leica Microsystems) after Gram-staining and cell morphology of strain Ha5^T cultured in LMG medium 404

broth for 24 hours was determined using scanning electron microscope (FEI Quanta FEG 650). Catalase and oxidase activities were assayed using 3% (v/v) hydrogen peroxide and 1% (w/v) *N*, *N*, *N*', *N*'-Tetramethyl-1, 4-phenylenediamine (bioMérieux) respectively. Motility was determined on 0.4% semi-solid LMG medium 404 slant [10, 42]. The strains were tested for temperature range (4, 10, 15, 20, 25, 30, 35, 37 and 40 °C) for growth on LMG 404 agar medium for 14 days. Tolerance to sodium chloride (NaCl) was determined by growing the strains in LMG medium 404 containing different concentration of NaCl (0, 1, 1.5, 2 and 2.5% (w/v)) [11].

Oxidations of lactate and acetate were tested in broth media containing 0.2% (w/v) yeast extract, 0.3% (w/v) peptone, either 0.2% (w/v) of sodium acetate or calcium lactate and 0.002% (w/v) bromothymol blue [43]. The production of 2-keto-D-gluconic acid and 5-keto-D-gluconic acid from D-glucose was detected using thin layer chromatography as described by Gosselé et al. [44]. Production of water-soluble brown pigment was examined on GYC agar containing 2% (w/v) glucose, 2% (w/v) yeast extract, 2% (w/v) calcium carbonate and 2% (w/v) agar [45]. The production of acetic acid from ethanol was tested on agar medium containing 1% (w/v) yeast extract, 2% (w/v) calcium carbonate, 1% (v/v) ethanol and 1.5% (w/v) agar [45]. Strains were cultured in broth with 1% (w/v) yeast extract and 3% (v/v) glycerol for 7 days at 28 °C and production of dihydroxyacetone from glycerol was tested by addition of Benedict's solution [46]. Ability of the strains to utilize ammonium as the sole nitrogen source was tested using Frateur's modified Hoyer medium containing either 3% (w/v) D-glucose, D-mannitol or 3% (v/v) ethanol [47].

Other biochemical tests included growth of strains in the presence of 30% (w/v) glucose and 1% (w/v) glucose (with modified LMG medium 404 broth), growth in the presence of 0.35% (v/v) acetic acid and in the presence of 1% (w/v) potassium nitrate (KNO_3) (both in LMG medium 404), growth on 1% (w/v) glutamate agar [48] and growth on mannitol agar [43]. Acid production from L-arabinose, D-arabinose, D-glucose, D-galactose, D-mannose, D-fructose, melibiose, sucrose, raffinose, maltose, methanol, D-mannitol, D-sorbitol, glycerol and ethanol was determined in broth medium containing 0.5% (w/v) yeast extract, 0.002% (w/v) bromocresol purple and 1% of the carbon sources [43].

Respiratory quinones were analysed by the Identification Service of the DSMZ, Braunschweig, Germany using the two-stage method as described by Tindall [49, 50]. The whole-cell fatty acid methyl esters (FAME) composition was determined using an Agilent Technologies 6890N gas chromatograph (GC). The strains were cultivated on LMG medium 404 agar under aerobic conditions at 28 °C for 48 h following the exact conditions applied on family *Acetobacteraceae* as described by Li et al. [10]. Cultures were extracted for fatty acids according to the standard protocol of the Microbial Identification System (Sherlock version 6.1; MIDI). The peaks profiles obtained were identified using the TSBA50 identification library version 5.0 (MIDI).

Plasmid and *rrn* operon analysis

Plasmids of strains Ha5^T, Ta1 and Jb2 were all detected with only one EcoRV cleavage site within the 23S rRNA gene from their complete genome sequences. Detection of *rrn* operon on plasmids was carried out by Southern hybridization analysis with the gDNA digested with restriction enzyme EcoRV-HF (NEB, cat. No. R3195S) [51]. To determine the presence of similar genome organization in closely related family *Acetobacteraceae* type members, the reference strains *B. intestini* DSM 28636^T, *S. floricola* DSM 15669^T, *N. thailandica* NBRC 106555^T, *N. tanensis* NBRC 106556^T and *S. samuiensis* NBRC 107927^T were included in the analysis. An initial sequence analysis confirmed that these reference strains contained restriction site(s) in their 23S rRNA gene but not 16S rRNA gene. Digestion of their gDNA with enzyme EcoRV-HF should not fragment 16S rRNA gene which could result in multiple detections of the same gene in Southern hybridization analysis.

Briefly, 1.5 µg of the digested gDNA was electrophoresed in a 1% w/v agarose gel and Southern blotted onto a Biodyne™ B nylon membrane (Thermo Scientific) by downward capillary transfer in alkaline buffer. A probe targeting 16S rRNA gene on *rrn* operon was amplified from gDNA of strain Ha5^T using the forward primer 5' – AGGTGATCCAGCCGCA – 3' and reverse primer 5' – CTTGAGAGTTTGATCCTGGC – 3' in a PCR reaction that contained 0.65 mM dTTP, 0.35 mM Biotin-11-dUTP (Thermo Scientific) and 1 mM of each of the other dNTPs. Probe hybridization and chromogenic detection of the biotin-labelled probe were performed using Biotin Chromogenic Detection Kit (Thermo Scientific) according to the manufacturer's protocol.

The copy number of plasmid in Ha5^T was determined by quantitative PCR (qPCR) as described by Anda et al. [51]. Ha5^T cultured in LMG medium 404 broth was collected at different time intervals for DNA extraction. The *rpsB* gene located on the chromosome and 16S rRNA (*rrs*) gene located on plasmid were selected for qPCR. Primer pair 5' – CCAACCTGCTGTTCGTGAT – 3' and 5' – GCAATGACGACGCTCTCC – 3' was used to amplify a 150-bp PCR product of *rpsB* gene while *rrs* gene was amplified by primer pair 5' – CCGACACTGAGAAGCCAAG – 3' and 5' – CCTTCGCAACTGGTGTTCT – 3' to generate a 147-bp PCR product. KAPA SYBR® FAST qPCR Kits (Kapa Biosystems) was employed for qPCR assay and each reaction contained 1X master mix, 0.2 ng of EcoRV-digested gDNA as template and 200 nM of the above primer pair in 20 µl volume. The qPCR was carried out in an Eco™ Real-Time PCR System (Illumina) using the following conditions: 1 cycle of initial denaturation for 10 min at 95 °C, followed by 45 cycles of 95 °C for 10 s, 60 °C for 10 s, and 72 °C for 15 s. A 10-fold serial dilution of the DNA fragments amplified by the above primers were included in qPCR reactions and used to construct the standard curves. The copy number of plasmid was calculated by dividing its copy number of *rrs* by the copy number of chromosomal *rpsB* [52].

Results and discussion

Bacterial isolation, DNA fingerprinting and cluster analysis

All isolates obtained were listed in Supplementary Table S2. Only enrichment media A and B were turbid with microbial growth but not enrichment medium C with D-sorbitol as the main carbon source. The purified isolates were subjected to MALDI-TOF analysis and the spectra generated were grouped into two distantly related MALDI-TOF MS clusters in dendrogram created with MALDI Biotyper MSP (Supplementary Fig. S1 and Supplementary Table S2). The isolates in MALDI-TOF MS cluster 1 were placed on the same clade with *Lactobacillus casei* DSM 20011^T, while cluster 2 consisted of isolates that were closest to *Gluconobacter oxydans* subspecies *oxydans* B540 UFL in Bruker database. The MALDI-TOF MSP of isolates in the same cluster was highly similar and thus, close in distance level in MSP dendrogram.

A preliminary analyses on partial 16S rRNA gene sequences of representative isolates in MALDI-TOF MS cluster 1 revealed that they shared highest similarity to *Lactobacillus sanfranciscensis* ATCC 27651^T (16S rRNA gene accession number: X76327, with similarity of 98.15%) in EzBioCloud 16S database. On the other hand, representative isolates from MALDI-TOF MS cluster 2 were closest in 16S rRNA gene sequence similarity with *N. tanensis* AH13^T (accession number: AB513363, with similarity of 94.15%) in EzBioCloud 16S database. The low 16S rRNA gene similarity of isolates in MALDI-TOF cluster 2 to the nearest neighbours suggested that these isolates should be classified into a new genus in family *Acetobacteraceae*.

RAPD analysis using primers 270 and 272 revealed highly similar DNA fingerprints among isolates from the same ant sample. Based on the cluster analysis using UPGMA method, isolates from same ant sample were clustered in the same clade, suggesting that these isolates represent a single strain (Supplementary Fig. S2). Different DNA fingerprints were obtained for isolates from different ant samples which formed separated clades in RAPD analysis. Strain Jb2 was the only isolate of the ant sample that belonged to MALDI-TOF cluster 2 and formed individual clade in cluster analysis. Strains Ha5^T, Ta1 and Jb2 were chosen as the representative strain from each cluster for further analysis.

Genome characteristics and gene annotation of the novel strains

Genomes of all novel strains were sequenced and deposited in GenBank (accession numbers in Supplementary Table S1). Complete genome sequencing of strain Ha5^T yielded a 1.95 Mb genome which consisted of a main circular chromosome of 1,938,936 bp (CP038143) and a much smaller replicon of 6,593 bp (CP038144) designated as pHa5 (Table 1). On the other hand, the total combined contig sizes of strains Ta1 and Jb2 genomes were 1.92 Mb and 1.94 Mb, respectively, based on their

draft genome sequences (Supplementary Table S3 and S4). Interestingly, a small circular replicon of 6,591 bp that was highly similar to pHa5 of strain Ha5^T was identified in both the genomes of strains Ta1 and Jb2 and they were designated as pTa1 (CM017016) and pJb2 (CM017012) (Supplementary Table S3 and S4) respectively.

The genome size of the novel strains were the smallest (1.92-1.95 Mb) among the compared genera with available genomes (Table 2, Supplementary Table S1). We believe that the small genome size is a result of genome reduction after participating in long term symbiotic relationship with the host ant species, *O. smaragdina*. Similar phenomenon of genome reduction had been reported in gut associated *Acetobacteraceae* of red carpenter ant [14] and *Asaia* endosymbionts of mosquitoes [15]. The genome of *B. intestini* DSM 28636^T isolated from bee gut is slightly larger (2.02 Mb) than the novel strains but has the lowest number of genes and protein coding genes in the family. We believe that this is due to gaps in its draft genome. Overall, strains Ha5^T, Ta1, Jb2 and *B. intestini* DSM 28636^T that were isolated from insects had the lowest number of genes as well as protein coding genes. The genomic DNA G + C contents of the novel strains were determined from their genome sequences and are in the range 61.56-61.65 mol% (Table 2, Supplementary Table S1).

To our surprise, no ribosomal RNA sequences was detected in the main chromosome of strain Ha5^T during assembly and annotation. Instead, further analysis revealed the presence of only one complete *rrn* operon in its genome which was located on the smaller replicon pHa5 (Table 1). The complete *rrn* operon consisted of genes for 16S rRNA, two genes coding for tRNA^{Ile} and tRNA^{Ala}, 23S and 5S rRNA, and a final tRNA^{Met} (Fig. 1). Similar *rrn* operon in the same gene order was observed on the small replicons in genomes of strains Ta1 and Jb2 during their draft genomes assembly and annotation. Upstream of the complete *rrn* operon was a replication initiation protein, RepE and a hypothetical protein located at the other end was observed (Fig. 1). Intriguingly, all the mentioned genes were compactly arranged on a 6.5 kb circular replicon with 99% sequence similarity in all three novel strains, indicating that the replicon is highly conserved among the strains.

From Southern hybridization analysis using biotinylated 16S rRNA amplicon as probe, a single band slightly larger than 6-kb was detected from EcoRV-digested gDNA of all three novel strains (Supplementary Fig. S3). The result proved the existence of 16S rRNA gene on a single locus and thus, the localization of *rrn* operon on a 6.5-kb circular replicon in the genomes of the novel strains (Supplementary Fig. S3). In addition to their small size, replicons pHa5, pTa1 and pJb2 were all having lower G + C contents (48.52-48.79 mol%) compared to the chromosomal DNA (Table 1, Supplementary Table S3 and S4). Moreover, the copy number of replicon pHa5 was determined to range from 15 to 25 per chromosome equivalent in a qPCR analysis (Supplementary Table S5). As these small replicons in the novel strains possessed the characteristics resembling a plasmid, we thus omitted the possibility that they were second chromosomes.

The localization of *rrn* operon on plasmid was first seen in *Bacillus megatarium* QM B1551 [53] and genus *Paracoccus* [54]. However, the mentioned bacteria have additional *rrn* operon on their chromosomes unlike Ha5^T, Ta1 and Jb2 that have their sole *rrn* operon on plasmid. To date, similar phenomenon had only been observed in the bacterial genus *Aureimonas* sp. associated with plants [51]. The group of *Aureimonas* sp. strains representing a close phylogenetic clade was reported with their main chromosomes lacking of *rrn* operon and the sole *rrn* operon in their genomes was located on small replicons (sizes ranged from 9.3-9.9 kb) designated as *rrn*-plasmids. The *rrn*-plasmids in Ha5^T, Ta1 and Jb2 have a smaller size (6.5 kb) than the mentioned *Aureimonas* sp. group and carried lesser number of genes. Nonetheless, these *rrn*-plasmids possessed all the characteristics similar to that of the *Aureimonas* sp. group including smaller size than main chromosomes, lower G + C contents, high copy number and lack of partitioning genes.

In the same Southern hybridization analysis, multiple bands were detected in the other closely related type strains including *B. intestini* DSM 28636^T that was detected with only a single copy of *rrn* operon in its draft genome (JATM00000000) (Supplementary Fig. S3). The result proved the presence of multiple 16S rRNA genes on more than one locus in their genomes and suggested that these type strains did not possess genome organization similar to the novel strains. As compared to the other bacteria, the possession of *rrn*-plasmid in high copy number and thus, more *rrn* operon copies in strains Ha5^T, Ta1 and Jb2 could be beneficial. A cell requires optimum number of *rrn* operon to meet the demand for ribosomes in protein synthesis and to support the growth rate of the cell [55]. *Escherichia coli* that is devoid of two out of seven *rrn* operons has showed decreased growth rate and fitness [55]. Hence, higher *rrn* operon copies in strains Ha5^T, Ta1 and Jb2 are essential for protein synthesis and rapid growth rate of the bacteria.

Besides, the number of *rrn* operons in bacteria does not correlate with the whole genome sizes but play essential functions in adaptation of bacterial hosts to resource availability. Study on soil bacteria that required less time to form colonies on high nutrient medium were discovered with higher number of *rrn* operons (about 5.5 copies) as compared to slow-growing bacteria (about 1.4 copies) [56]. *Bradyrhizobium diazoefficiens* [57] and *B. japonicum* are among the striking examples with large genome of 9.1 Mb but only a single copy of *rrn* operon and *B. japonicum* was reported with a cell division time of 9.6 hr [58]. Taking the small genome size and significantly lower number of genes and proteins into consideration, *rrn*-plasmids in strains Ha5^T, Ta1 and Jb2 could be providing selective advantage for rapid adaptation to environmental alterations, in addition to nutritional changes of the host *O. smaragdina*.

Despite their small genome sizes and significantly lower gene numbers, gene annotation revealed the presence of cofactors and vitamins including riboflavin, biotin, folate, pantothenate, thiamine, pyridoxine, protoheme and nicotinamide biosynthetic genes in the genomes of strains Ha5^T,

Ta1 and Jb2. These genes are crucial in nutritional contribution to the ant host *O. smaragdina*. Among the mentioned vitamins, riboflavin is an absolute dietary supplement for higher animals that do not possess the riboflavin biosynthetic pathway. As riboflavin is only synthesized by plants and most bacteria, *O. smaragdina* could be therefore relying on dietary riboflavin or provisioning by strains Ha5^T, Ta1 and Jb2 through biosynthesis.

In addition, genes involved in biosynthesis of biosynthetic genes for all non-essential amino acids and multiple essential amino acids (except for isoleucine, leucine and valine) were identified in the genomes of strains Ha5^T, Ta1 and Jb2. It is suggested the strains supply amino acids to the ant host that relies mostly on plant sap and honeydew which constitute an unbalanced diet lacking of amino acids [17]. However, the omnivorous *O. smaragdina* also prey on intruding insects which could be the source of nitrogen [18]. The lacking of biosynthetic genes in the strains may imply a less important role in supplying essential amino acids to the host. Nonetheless, bacterial symbionts are recognized for biosynthesis and supply of amino acids to their insect hosts. Notably, the endosymbiont *Blochmannia* sp. in ants of the genus *Camponotus* retains biosynthetic genes for amino acids including tyrosine even with a strongly reduced genome [12]. Overall, biosynthetic genes in the genomes indicated a nutritional basis of the symbiosis between the strains and the host and provided insight into their biological function as symbiont of *O. smaragdina*.

16S rRNA gene and genome-based phylogenetic analysis

Complete 16S rRNA gene sequences were extracted from all available whole genome sequences. The 16S rRNA gene sequence of strain Ha5^T shared 99.93% similarity with strains Ta1 and Jb2 and was only different by 1 bp (Supplementary Table S6). The strains shared highest 16S rRNA gene sequence similarities with two members of the genus *Neokomagataea*, *N. tanensis* NBRC 106556^T (94.56-94.63%) and *N. thailandica* NBRC 106555^T (94.03-94.10%) and *B. intestini* DSM 28636^T (94.09-94.16%). However, the three strains formed a cluster phylogenetically apart from genus *Neokomagataea* in phylogenetic trees calculated with ML and NJ methods. Instead, both phylogenetic trees grouped the novel strains in a distinct clade which was closer to a separated clade that included *S. floricola* DSM 15669^T and two members of *Bombella*, before placing the mentioned taxa in a larger clade that included genera *Neokomagataea* and *Swingsia samuiensis* NBRC 107927^T (Fig. 2 and Supplementary Fig. S4). In short, strains Ha5^T, Ta1 and Jb2 formed a distinct lineage in the 16S rRNA gene phylogenetic analysis of family *Acetobacteraceae* which suggests that they should be differentiated as a new genus.

Although 16S rRNA gene remains the widely used molecular marker in taxonomic and phylogenetic study, the approach omits the overall genetic differences especially among closely related species. Among *Acetobacteraceae* members, *Acetobacter* is known for having highly similar

16S rRNA gene sequences and requires other methods such as multilocus sequence typing (MLSA) for differentiation [59]. To obtain a robust construction of phylogeny among this family, we therefore decided to elucidate the phylogenetic relationship of strains Ha5^T, Ta1, Jb2 and family *Acetobacteraceae* members through phylogenomic analysis. A total of 75 whole genome sequences of strains acquired from this study or from GenBank (Supplementary Table S1) was included for phylogenomic analysis using panX. Genomes without 16S rRNA gene, ambiguous 16S rRNA gene identity and BUSCO genome completeness lower than 90% were excluded from analysis.

The phylogenomic analysis revealed the existence of 96 orthologous gene clusters which were identified as the core genes from all genomes included in this study. The ML phylogenetic tree based on these concatenated genes showed placement of all genera in family *Acetobacteraceae* on distinctive phylogenetic lineages (Fig. 3). Similar to phylogenetic trees based on 16S rRNA gene sequences, strains Ha5^T, Ta1 and Jb2 formed a distinct cluster in the phylogenomic tree and were placed in the same clade with *S. floricola* DSM 15669^T and *B. intestini* DSM 28636^T. These analyses showed that all nodes were supported by bootstrap values greater than 50 which indicated confident robustness of the result. The phylogenomic tree corroborated the phylogenetic distinction of the three novel strains from all members of the family *Acetobacteraceae* and indicated that they should be classified as a novel genus in the family.

In addition, the observation was confirmed through ANI, *in silico* DDH and AAI calculations. Strains Ha5^T, Ta1 and Jb2 shared ANI values in the range of 97.50-97.69% (Supplementary Fig. S5), *in silico* DDH values from 77.4% to 79.3% (Supplementary Fig. S6), and AAI values in the range of 97.93-98.08% (Fig. 4). These values are higher than the cut-off values (95-96% for ANI; >70% for *in silico* DDH; >95% for AAI) proposed for species boundary and thus showed that these strains belonged to the same species [39, 60].

Furthermore, the ANI, *in silico* DDH and AAI values between the novel strains and other members of family *Acetobacteraceae* were far below the proposed cut-off values for species delineation (≤70.67% for ANI; ≤37.1% for *in silico* DDH; ≤66.01% for AAI) (Supplementary Fig. S5, S6, Fig. 4). It is also noteworthy that the ≤66.01% pairwise AAI values between strains Ha5^T, Ta1, Jb2 and type members of family *Acetobacter*aceae were within the reported range of genus-level pairwise AAI difference (65% to 72% AAI values) [39]. Through these indices, it was substantiated that strains Ha5^T, Ta1, Jb2 did not belong to any of the current species in family *Acetobacter*aceae and should be delineated as a single species within a distinct genus in the family. Overall, these findings enhanced the robustness of the phylogeny of strains Ha5^T, Ta1 and Jb2 and further supported that the strains represent a novel genus of the family *Acetobacteraceae*.

On the other hand, we had looked into POCP that was suggested as an alternative prokaryotic genus delineation index recently [40]. Based on the proposed formula for POCP calculation, strains

Ha5^T, Ta1 and Jb2 shared high POCP values in the range of 97.24-97.61%. However, the novel strains also shared higher than 50% POCP values with many type members of family *Acetobacteraceae*, including *Bombella*, *Neokomagataea*, *Swingsia*, *Saccharibacter*, *Gluconobacter*, *Acetobacter* and *Asaia* (Supplementary Fig. S7). The high POCP values could be attributed to the low number of protein coding genes in the novel strains compared to the other genera. With reference to the suggested cut-off threshold, two strains with a POCP value of higher than 50% between them can be defined as species grouped in a prokaryotic genus [40]. The result contradicted with ANI, *in silico* DDH and AAI that proved strains Ha5^T, Ta1 and Jb2 did not belong to any of the existing genera of family *Acetobacteraceae*. Besides, the suggested cut-off threshold did not apply to many of the other existing genera of family *Acetobacteraceae* too as POCP values higher than 50% had been observed between members of multiple genera (Supplementary Fig. S7). Our findings showed that POCP may not be reliable for genus-level delineation of family *Acetobacteraceae*.

Morphological, physiological and biochemical characterizations

Strains Ha5^T, Ta1 and Jb2 formed beige-coloured, shiny, circular, convex, smooth and opaque small colonies on LMG medium 404 agar after incubation for 2 days at 28 °C. The cells are Gram-stain negative, occurred rod in shape when observed under the light microscope. From SEM micrograph, these rod-shaped cells had a size of 0.45 – 0.5 µm in width and 0.8 – 1.0 µm in length (Fig. 5). No flagella were observed and all strains were non-motile on semi-solid agar.

Phenotypic and physiological tests performed on strains Ha5^T, Ta1 and Jb2 showed similar results and their physiological differentiation with all reference strains are listed in Table 2. Similar to the closest members in family *Acetobacteraceae*, strains Ha5^T, Ta1 and Jb2 were tested negative for oxidase activity but positive for catalase activity. They did not carry out oxidation of lactate and acetate. Strains Ha5^T, Ta1 and Jb2 showed glucose requirement for growth and were able to grow in media with 1% and 30% (w/v) glucose (Table 2). While growth was observed in temperature from 15 to 37 °C, optimum growth occurred aerobically at 28-30 °C. No growth occurred at 10 °C and 40 °C. All three strains showed growth in medium containing NaCl up to 1.5% (w/v) but optimum growth occurred at 0%. The inability to produce acid from D-sorbitol was consistent with absence of growth in enrichment C that contained D-sorbitol as the main carbon source. Phenotypic features that differentiate strain Ha5^T and closest members of family *Acetobacteraceae* included the production of 5-keto-D-gluconate from D-glucose, production of water soluble brown pigment, production of dihydroxyacetone from glycerol, growth in the presence of 0.35% v/v acetic acid, growth in the presence of 1% KNO_3 (w/v) and acid production from various carbon sources (Table 2).

Analysis of respiratory quinones revealed ubiquinone-10 (Q-10) as the major ubiquinone in strain Ha5^T. This is consistent with all members of family *Acetobacteraceae* except genus

Acetobacter that is having ubiquinone-9 (Q-9) as the major ubiquinone [61, 62]. The major cellular fatty acid profile of the Ha5^T contained $C_{18:1}$ $\omega 7c$ (36.11%), $C_{16:0}$ (22.31%) and $C_{19:0}$ $\omega 8c$ cyclo (18.43%) which accounted for 76.85% of the total fatty acids (Table 3). Strain Ha5^T has similar overall major fatty acid profile to those of strains Ta1 and Jb2 except differences in respective proportions of some fatty acid components (Table 3). Strain Ha5^T and all reference strains have $C_{18:1}$ $\omega 7c$ and $C_{16:0}$ as the major fatty acids and differences are observed mostly with the low abundance fatty acids (Table 3).

Conclusions

With reference to the results from phenotypic and chemotaxonomic characterization, phylogenetic analyses based on 16S rRNA gene sequences and phylogenomic analysis as well as whole-genome comparisons, strains Ha5^T, Ta1 and Jb2 were observed with highly similar characteristics and are discernible from all neighbouring species. Taken together, the strains represent a novel species of a novel genus within family *Acetobacteraceae*, for which we proposed the name *Oecophyllibacter saccharovorans* gen. nov. and sp. nov. with strain Ha5^T (= DSM106907^T = LMG30589^T = NBRC 113643^T = KCTC62952^T) as the type strain. Strains Ta1 (= NBRC 113642 = KCTC62953) and Jb2 (= NBRC 113641 = KCTC62954) represent additional strains of the species which could be differentiated from the type strain by genomic fingerprinting using RAPD, slight differences in 16S rRNA gene sequences and all genomic comparisons employed in this study. Currently, these strains possess the smallest genome sizes in family *Acetobacteraceae* that could be attributed to genome reduction. A rare genome organization with the sole *rrn* operon in the genome located on a plasmid is observed in all three strains. As one of the few family *Acetobacteraceae* members that are associated with insects, the ecological role of strains Ha5^T, Ta1 and Jb2 in the host weaver ant *O. smaragdina* clearly deserves further study. Descriptions of *Oecophyllibacter* gen. nov. and *Oecophyllibacter saccharovorans* sp. nov. are given in Table 4.

Acknowledgements

The authors are grateful to Professor Aharon Oren for his advice on etymology and nomenclature of the strains. KOC thanks MyBrain15 Postgraduate Scholarship Programme for the scholarship (MyPhD, KPT(B)900909146137) awarded. WSST thanks the Bright Sparks Program of the University of Malaya for scholarship awarded. This work was supported by University of Malaya Research Grants (FRGS grant FP022-2018A), University of Malaya High Impact Research Grants (UM-MOHE HIR Grant UM.C/625/1/HIR/MOHE/ CHAN/14/1, Grant No. H-50001-A000027; UM-MOHE HIR Grant UM.C/625/1/HIR/MOHE/CHAN/01, Grant No. A-000001-50001) awarded to KGC and Postgraduate Research (PPP) Grant (Grant No. PG089-2015B) awarded to KOC.

Declarations of interest: none.

References

[1] Gillis, M., De Ley, J. (1980) Intra-and intergeneric similarities of the ribosomal ribonucleic acid cistrons of *Acetobacter* and *Gluconobacter*. Int. J. Syst. Evol. Microbiol. 30, 7-27.
[2] Kersters, K., Lisdiyanti, P., Komagata, K., Swings, J. (2006) The family *Acetobacter*aceae: the genera *Acetobacter*, Acidomonas, *Asaia*, *Gluconacetobacter*, *Gluconobacter*, and *Kozakia*. The Prokaryotes: Volume 5: Proteobacteria: Alpha and Beta Subclasses. 163-200.
[3] Gomes, R.J., Borges, M.d.F., Rosa, M.d.F., Castro-Gómez, R.J.H., Spinosa, W.A. (2018) Acetic acid bacteria in the food industry: systematics, characteristics and applications. Food Technol. Biotechnol 56, 139-151.
[4] Greenberg, D.E., Porcella, S.F., Stock, F., Wong, A., Conville, P.S., Murray, P.R., Holland, S.M., Zelazny, A.M. (2006) *Granulibacter bethesdensis* gen. nov., sp. nov., a distinctive pathogenic acetic acid bacterium in the family *Acetobacter*aceae. Int. J. Syst. Evol. Microbiol. 56, 2609-2616.
[5] Yukphan, P., Malimas, T., Muramatsu, Y., Potacharoen, W., Tanasupawat, S., Nakagawa, Y., Tanticharoen, M., Yamada, Y. (2011) *Neokomagataea* gen. nov., with descriptions of *Neokomagataea* thailandica sp. nov. and *Neokomagataea* tanensis sp. nov., osmotolerant acetic acid bacteria of the α-Proteobacteria. Bioscience, biotechnology, and biochemistry 75, 419-426.
[6] Crotti, E., Rizzi, A., Chouaia, B., Ricci, I., Favia, G., Alma, A., Sacchi, L., Bourtzis, K., Mandrioli, M., Cherif, A. (2010) Acetic acid bacteria, newly emerging symbionts of insects. Appl. Environ. Microbiol. 76, 6963-6970.
[7] Roh, S.W., Nam, Y.-D., Chang, H.-W., Kim, K.-H., Kim, M.-S., Ryu, J.-H., Kim, S.-H., Lee, W.-J., Bae, J.-W. (2008) Phylogenetic characterization of two novel commensal bacteria involved with innate immune homeostasis in *Drosophila melanogaster*. Appl. Environ. Microbiol. 74, 6171-6177.
[8] Kounatidis, I., Crotti, E., Sapountzis, P., Sacchi, L., Rizzi, A., Chouaia, B., Bandi, C., Alma, A., Daffonchio, D., Mavragani-Tsipidou, P. (2009) *Acetobacter tropicalis* is a major symbiont of the olive fruit fly (*Bactrocera oleae*). Appl. Environ. Microbiol. 75, 3281-3288.
[9] Favia, G., Ricci, I., Damiani, C., Raddadi, N., Crotti, E., Marzorati, M., Rizzi, A., Urso, R., Brusetti, L., Borin, S. (2007) Bacteria of the genus *Asaia* stably associate with *Anopheles stephensi*, an Asian malarial mosquito vector. Proc. Natl. Acad. Sci. 104, 9047-9051.
[10] Li, L., Praet, J., Borremans, W., Nunes, O.C., Manaia, C.M., Cleenwerck, I., Meeus, I., Smagghe, G., De Vuyst, L., Vandamme, P. (2015) *Bombella intestini* gen. nov., sp. nov., an acetic acid bacterium isolated from bumble bee crop. Int. J. Syst. Evol. Microbiol. 65, 267-273.
[11] Yun, J.-H., Lee, J.-Y., Hyun, D.-W., Jung, M.-J., Bae, J.-W. (2017) *Bombella apis* sp. nov., an acetic acid bacterium isolated from the midgut of a honey bee. Int. J. Syst. Evol. Microbiol. 67, 2184-2188.
[12] Feldhaar, H., Straka, J., Krischke, M., Berthold, K., Stoll, S., Mueller, M.J., Gross, R. (2007) Nutritional upgrading for omnivorous carpenter ants by the endosymbiont *Blochmannia*. BMC Biol. 5, 48.

[13] Crotti, E., Chouaia, B., Alma, A., Favia, G., Bandi, C., Bourtzis, K., Daffonchio, D. (2016) Acetic acid bacteria as symbionts of insects, In: Acetic Acid Bacteria, Springer, pp. 121-142.
[14] Brown, B.P., Wernegreen, J.J. (2019) Genomic erosion and extensive horizontal gene transfer in gut-associated *Acetobacter*aceae. BMC Genomics 20, 472.
[15] Alonso, D.P., Mancini, M.V., Damiani, C., Cappelli, A., Ricci, I., Alvarez, M.V.N., Bandi, C., Ribolla, P.E.M., Favia, G. (2018) Genome reduction in the mosquito symbiont *Asaia*. Genome Biol. Evol. 11, 1-10.
[16] Pimid, M., Hassan, A., Tahir, N.A., Thevan, K. (2014) Colony Structure of the Weaver Ant, *Oecophylla smaragdina* (Fabricius)(Hymenoptera: Formicidae). Sociobiol. 59, 1-10.
[17] Holldobler, B., Wilson, E. (1990) The ants. Cambridge, US, In: Harvard University Press.
[18] Peng, R., Christian, K. (2004) The weaver ant, *Oecophylla smaragdina* (Hymenoptera: Formicidae), an effective biological control agent of the red-banded thrips, *Selenothrips rubrocinctus* (Thysanoptera: Thripidae) in mango crops in the Northern Territory of Australia. Int. J. Pest Manage. 50, 107-114.
[19] Way, M., Khoo, K. (1989) Relationships between *Helopeltis theobromae* damage and ants with special reference to Malaysian cocoa smallholdings. J. Plant Prot. 6, 1-11.
[20] Peng, R., Christian, K., Gibb, K. (1995) The effect of the green ant, *Oecophylla smaragdina* (Hymenoptera: Formicidae), on insect pests of cashew trees in Australia. B. Entomol. Res. 85, 279-284.
[21] Chua, K.-O., Song, S.-L., Yong, H.-S., See-Too, W.-S., Yin, W.-F., Chan, K.-G. (2018) Microbial community composition reveals spatial variation and distinctive core microbiome of the weaver ant *Oecophylla smaragdina* in Malaysia. Scientific Reports 8, 10777.
[22] Lin, J.Y., Russell, J.A., Sanders, J.G., Wertz, J.T. (2016) *Cephaloticoccus* gen. nov., a new genus of '*Verrucomicrobia*' containing two novel species isolated from *Cephalotes* ant guts. Int. J. Syst. Evol. Microbiol. 66, 3034-3040.
[23] Yamada, Y., Hosono, R., Lisdyanti, P., Widyastuti, Y., Saono, S., Uchimura, T., Komagata, K. (1999) Identification of acetic acid bacteria isolated from Indonesian sources, especially of isolates classified in the genus *Gluconobacter*. J. Gen. Appl. Microbiol. 45, 23-28.
[24] Yamada, Y., Katsura, K., Kawasaki, H., Widyastuti, Y., Saono, S., Seki, T., Uchimura, T., Komagata, K. (2000) *Asaia bogorensis* gen. nov., sp. nov., an unusual acetic acid bacterium in the alpha-Proteobacteria. Int. J. Syst. Evol. Microbiol. 50, 823-829.
[25] Goh, S.-Y., Khan, S.A., Tee, K.K., Kasim, N.H.A., Yin, W.-F., Chan, K.-G. (2016) Quorum sensing activity of *Citrobacter amalonaticus* L8A, a bacterium isolated from dental plaque. Scientific Reports 6, 20702.
[26] Lane, D.J. (1991) 16S/23S rRNA sequencing, In: E.S.a.M. Goodfellow (Ed.) Nucleic acid techniques in bacterial systematics., Wiley & Sons, Chichester, United Kingdom., pp. p. 115–175.
[27] Kim, O.-S., Cho, Y.-J., Lee, K., Yoon, S.-H., Kim, M., Na, H., Park, S.-C., Jeon, Y.S., Lee, J.-H., Yi, H. (2012) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. Int. J. Syst. Evol. Microbiol. 62, 716-721.
[28] Kumar, S., Stecher, G., Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870-1874.
[29] Krzewinski, J.W., Nguyen, C.D., Foster, J.M., Burns, J.L. (2001) Use of random amplified polymorphic DNA PCR to examine epidemiology of *Stenotrophomonas maltophilia* and *Achromobacter* (*Alcaligenes*) *xylosoxidans* from patients with cystic fibrosis. J. Clin. Microbiol. 39, 3597-3602.
[30] Heras, J., Domínguez, C., Mata, E., Pascual, V., Lozano, C., Torres, C., Zarazaga, M. (2015) GelJ–a tool for analyzing DNA fingerprint gel images. BMC Bioinformatics 16, 270.
[31] Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10, 563-569.

[32] Sullivan, M.J., Zakour, N.L.B., Forde, B.M., Stanton-Cook, M., Beatson, S.A. (2015) Contiguity: contig adjacency graph construction and visualisation, In: PeerJ PrePrints.
[33] Sommer, D.D., Delcher, A.L., Salzberg, S.L., Pop, M. (2007) Minimus: a fast, lightweight genome assembler. BMC Bioinformatics 8, 64.
[34] Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J. (2016) NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44(14):6614-6624.
[35] Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206-D214.
[36] Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M. (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212.
[37] Lee, I., Kim, Y.O., Park, S.-C., Chun, J. (2016) OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int. J. Syst. Evol. Microbiol. 66, 1100-1103.
[38] Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14, 60.
[39] Konstantinidis, K.T., Tiedje, J.M.J.J.o.b. (2005) Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187, 6258-6264.
[40] Qin, Q.-L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., Oren, A., Zhang, Y.-Z. (2014) A proposed genus boundary for the prokaryotes based on genomic insights. J. Bacteriol. 196, 2210-2215.
[41] Ding, W., Baumdicker, F., Neher, R.A. (2017) panX: pan-genome analysis and exploration. Nucleic Acids Res. 46, e5-e5.
[42] Tittsler, R.P., Sandholzer, L.A. (1936) The use of semi-solid agar for the detection of bacterial motility. J. Bacteriol. 31, 575.
[43] Asai, T., Iizuka, H., Komagata, K. (1964) The flagellation and taxonomy of genera *Gluconobacter* and *Acetobacter* with reference to the existence of intermediate strains. J. Gen. Appl. Microbiol. 10, 95-126.
[44] Gosselé, F., Swings, J., De Ley, J. (1980) A rapid, simple and simultaneous detection of 2-keto-, 5-keto-and 2, 5-diketogluconic acids by thin-layer chromatography in culture media of acetic acid bacteria. Zentralblatt für Bakteriologie: I. Abt. Originale C: Allgemeine, angewandte und ökologische Mikrobiologie 1, 178-181.
[45] Shimwell, J., Carr, J.G., Rhodes, M.E. (1960) Differentiation of *Acetomonas* and *Pseudomonas*. Microbiol. 23, 283-286.
[46] Aydin, Y.A., Aksoy, N.D. (2010) Utilization of vinegar for isolation of cellulose producing acetic acid bacteria, In: AIP Conference Proceedings. AIP, pp. 340-348.
[47] De Ley, J. (1984) Family VI. *Acetobacter*aceae. Bergey's Manual of Systematic Bacteriology 1, 267-278.
[48] Jojima, Y., Mihara, Y., Suzuki, S., Yokozeki, K., Yamanaka, S., Fudou, R. (2004) *Saccharibacter floricola* gen. nov., sp. nov., a novel osmophilic acetic acid bacterium isolated from pollen. Int. J. Syst. Evol. Microbiol. 54, 2263-2267.
[49] Tindall, B. (1990) Lipid composition of *Halobacterium lacusprofundi*. FEMS Microbiol. Letters 66, 199-202.
[50] Tindall, B. (1990) A comparative study of the lipid composition of *Halobacterium saccharovorum* from various sources. Syst. Appl. Microbiol. 13, 128-130.
[51] Anda, M., Ohtsubo, Y., Okubo, T., Sugawara, M., Nagata, Y., Tsuda, M., Minamisawa, K., Mitsui, H. (2015) Bacterial clade with the ribosomal RNA operon on a small plasmid rather than the chromosome. Proc. Natl. Acad. Sci. 112, 14343-14347.

[52] Zhong, C., Peng, D., Ye, W., Chai, L., Qi, J., Yu, Z., Ruan, L., Sun, M. (2011) Determination of plasmid copy number reveals the total plasmid DNA amount is greater than the chromosomal DNA amount in *Bacillus thuringiensis* YBT-1520. PloS one 6, e16025.
[53] Kunnimalaiyaan, M., Stevenson, D.M., Zhou, Y., Vary, P.S. (2001) Analysis of the replicon region and identification of an rRNA operon on pBM400 of *Bacillus megaterium* QM B1551. Mol. Microbiol. 39, 1010-1021.
[54] Battermann, A., Disse-Krömker, C., Dreiseikelmann, B. (2003) A functional plasmid-borne rrn operon in soil isolates belonging to the genus *Paracoccus*. Microbiol. 149, 3587-3593.
[55] Asai, T., Condon, C., Voulgaris, J., Zaporojets, D., Shen, B., Al-Omar, M., Squires, C., Squires, C.L. (1999) Construction and initial characterization of *Escherichia coli* strains with few or no intact chromosomal rRNA operons. J. Bacteriol. 181, 3803-3809.
[56] Klappenbach, J.A., Dunbar, J.M., Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66, 1328-1333.
[57] Sugawara, M., Tsukui, T., Kaneko, T., Ohtsubo, Y., Sato, S., Nagata, Y., Tsuda, M., Mitsui, H., Minamisawa, K. (2017) Complete genome sequence of *Bradyrhizobium diazoefficiens* USDA 122, a nitrogen-fixing soybean symbiont. Genome Announc. 5, e01743-01716.
[58] Keyser, H.H., Bohlool, B.B., Hu, T., Weber, D.F. (1982) Fast-growing rhizobia isolated from root nodules of soybean. Science 215, 1631-1632.
[59] Kim, K.H., Cho, G.Y., Chun, B.H., Weckx, S., Moon, J.Y., Yeo, S.-H., Jeon, C.O. (2018) *Acetobacter oryzifermentans* sp. nov., isolated from Korean traditional vinegar and reclassification of the type strains of *Acetobacter pasteurianus* subsp. ascendens (Henneberg 1898) and *Acetobacter pasteurianus* subsp. *paradoxus* (Frateur 1950) as *Acetobacter ascendens* sp. nov., comb. nov. Syst. Appl. Microbiol. 41, 324-332.
[60] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. 106, 19126-19131.
[61] Yamada, Y., Yukphan, P. (2008) Genera and species in acetic acid bacteria. Int. J. Food Microbiol. 125, 15-24.
[62] Yamada, Y., Aida, K., Uemura, T. (1968) Distribution of ubiquinone 10 and 9 in acetic acid bacteria and its relation to the classification of genera *Gluconobacter* and *Acetobacter*, especially of so-called intermediate strains. Agric. Biol. Chem. 32, 786-788.

Table 1. Genome properties of Ha5^{T}.

Replicon	Chromosome	pHa5	Value (Total)	% total
Accession number	CP038143	CP038144	-	-
Size (bp)	1,938,936	6,593	1,945,529	100
DNA G + C (bp)	1,194,450	3,217	1,197,667	61.56
DNA G + C content (mol%)	61.60	48.79	61.56	61.56
Total genes	1,681	8	1,689	100
Protein coding genes	1,616	2	1,617	95.73
RNA genes	46	6	52	3.08
rRNA genes	0	3	3	0.18
tRNA genes	42	3	45	2.66

Table 2. Differential characteristics between novel strains Ha5^T, Ta1, Jb2 and type strains of closely related genera of family *Acetobacteraceae*.

Characteristics	Ha5 ^{T}	Ta1	Jb2	*N. thailandica* NBRC106555^T	*N. tanensis* NBRC106556^T	*S. samuiensis* NBRC107927^T	*B. intestini* LMG 28161^T	*S. floricola* DSM15669^T
Isolation source	Insect (Ant, *Oecophylla smaragdina*)	Insect (Ant, *Oecophylla smaragdina*)	Insect (Ant, *Oecophylla smaragdina*)	Plant (Flower, *Lantana camera*)	Plant (Flower, *Senna alata*)	Plant (Flower, *Allamanda cathartica*)	Insect (Bee, *Bombus lapidarius*)	Plant (Pollen)
Cell morphology	Rods	Rods	Rods	Rods	Rods	Rods	Straight rods	Straight rods
Oxidase	–	–	–	–	–	–	–	–
Catalase	+	+	+	+	+	+	+	+
Motility	–	–	–	–	–	–	–	–
Major respiratory quinones	Q-10	ND	ND	Q-10	Q-10	Q-10	Q-10	Q-10
Temperature growth range (optimum)	15-37 (28-30)	15-37 (28-30)	15-37 (28-30)	15-30 (28)	15-30 (28)	15-37 (28-30)	15-40 (28-30)	15-30 (28-30)
NaCl% (w/v) (optimum)	0-1.5 (0)	0-1.5 (0)	0-1.5 (0)	0-0.5 (0)	0-1 (0)	0-1.5 (0)	0-1 (0)	0-0.5 (0)
Oxidation of								
Acetate	–	–	–	–	–	–	–	–
Lactate	–	–	–	–	–	–	–	–
Production of keto-D-gluconates from D-glucose								
2-keto-D-gluconate	+	+	+	+	+	+	+	+
5-keto-D-gluconate	–	–	–	+	+	+	–	+
Production of water-soluble brown pigment	–	–	–	–	–	+	–	–
Production of acetic acid from ethanol	–	–	–	w	w	w	–	w
Production of dihydroxyacetone from glycerol	w	w	w	–	–	+	w	–
Assimilation of ammonium nitrogen on Frateur-modified Hoyer medium with:								
D-Glucose	–	–	–	–	–	–	–	–
D-Mannitol	–	–	–	–	–	+	–	–
Ethanol	–	–	–	–	–	–	–	–
Growth in the presence of 0.35% (v/v) acetic acid	+	+	+	–	–	+	+	–
Growth in the presence of 1% (w/v) D-glucose	+	+	+	+	+	+	+	–
Growth in the presence of 30% (w/v) D-glucose	+	+	+	+	+	+	+	+

Growth on LMG medium 404 at 37°C	+	+	+	−	−	+	+	−
Growth on 1% (w/v) glutamate agar	−	−	−	+	+	+	−	−
Growth on Mannitol agar	−	−	−	+	+	+	+	−
Growth in the presence of 1% KNO_3 (w/v)	+	+	+	−	−	+	+	+
Methanol as sole carbon source	−	−	−	−	−	−	−	−
Acid production from								
L-Arabinose	+	+	+	+	+	+	+	+
D-Arabinose	−	−	−	−	−	−	−	−
D-Glucose	+	+	+	+	+	+	+	+
D-Galactose	+	+	+	+	+	+	+	−
D-Mannose	+	+	+	+	+	+	+	−
D-Fructose	+	+	+	+	+	+	+	−
Melibiose	+	+	+	+	+	−	−	−
Sucrose	−	−	−	+	+	+	−	+
Raffinose	−	−	−	−	+	−	−	−
Maltose	−	−	−	−	−	−	+	−
Methanol	−	−	−	−	−	−	−	−
D-Mannitol	+	+	+	+	+	+	+	−
D-Sorbitol	−	−	−	−	−	−	−	−
Glycerol	−	−	−	−	−	−	−	−
Ethanol	+	+	+	−	−	−	−	−
Genome size (Mb)	1.95	1.92	1.94	2.49	2.58	2.19	2.02	2.38
G+C content (mol%)	61.6	61.7	61.6	52.4	52.0	45.1	54.9	51.2
No. of genes	1689	1673	1680	2317	2351	2098	1632	2264
Protein coding	1617	1601	1612	2156	2155	1988	1574	2133

ND, Not determined; w, weak; +, positive reaction; −, negative reaction

Table 3. Fatty acid composition of strains Ha5^T, Ta1, Jb2 and type strains of closely related genera in family *Acetobacteraceae*.

Fatty acid	1	2	3	4	5	6	7	8
Saturated fatty acids								
$C_{12:0}$	0.65	–	–	–	–	–	–	–
$C_{14:0}$	5.61	3.37	3.55	0.60	–	0.40	9.30	2.80
$C_{16:0}$	22.31	17.77	16.92	13.9	10.4	19.8	12.7	24.4
$C_{18:0}$	2.52	2.41	1.92	8.70	4.20	5.50	–	1.10
Hydroxy fatty acids								
$C_{14:0}$ 2-OH	4.52	5.42	6.23	–	–	–	8.50	–
$C_{16:0}$ 2-OH	5.29	5.84	5.98	3.20	5.70	10.0	1.10	6.40
$C_{16:0}$ 3-OH	1.84	1.79	2.06	2.40	4.10	2.30	2.80	2.90
$C_{18:0}$ 3-OH	–	–	–	4.60	3.50	2.20	–	–
$C_{18:1}$ 2-OH	–	–	–	12.7	17.2	0.80	–	–
Unsaturated fatty acids								
$C_{18:1}$ $\omega 7c$	36.11	42.78	45.66	50.2	49.2	52.8	43.1	29.8
$C_{20:2}$ $\omega 6,9c$	1.08	0.92	–	–	–	–	–	–
Cyclopropane acids								
$C_{19:0}$ cyclo $\omega 8c$	18.43	17.85	16.91	–	–	3.70	–	28.4
Methyl fatty acids								
11-methyl $C_{18:1}$ $\omega 7c$	0.95	0.97	–	–	–	–	–	–
10-methyl $C_{19:0}$	–	–	–	–	1.30	–	–	–
Summed features[a]								
Summed feature 2	–	–	–	–	–	0.80	–	–
Summed feature 3	0.69	0.87	0.77	1.40	–	–	–	–

Strains 1-3 were investigated in this study. 1, Ha5^T; 2, Ta1; 3, Jb2; 4, *N. tanensis* NBRC 106556^T (data from Ref. [5]); 5, *N. thailandica* LMG 27021^T [10]; 6, *S. samuiensis* NBRC107927^T [10]; 7, *B. intestini* LMG 28161^T [10]; 8, *S. floricola* LMG 23170^T [10]. All strains were cultivated on LMG medium 404 agar under aerobic conditions at 28 °C for 48 h for the analysis. The values are shown in percentage of the total fatty acids. –, not detected

[a] Summed features are groups of two or more fatty acids that cannot be separated by GLC with MIDI system. Summed feature 2 contains $C_{12:0}$ aldehyde, $C_{14:0}$ 3-OH and/or iso-$C_{16:1}$. Summed feature 3 contains $C_{16:1}$ $\omega 7c$ and/or iso-$C_{15:0}$ 2-OH.

Table 4. Descriptions of *Oecophyllibacter* gen. nov. and *Oecophyllibacter saccharovorans* sp. nov.

Genus name	*Oecophyllibacter*	
Species name		*Oecophyllibacter saccharovorans*
Genus status	**gen. nov.**	
Genus etymology	**Oe.co.phyl.li.bac.ter. N.L. fem. n. *Oecophylla* a weaver ant genus; N.L. masc. n. *bacter* a rod; N.L. masc. n. *Oecophyllibacter*, a rod from the weaver ant *Oecophylla***	
Type species of the genus	***Oecophyllibacter saccharovorans***	
Specific epithet	–	*saccharovorans*
Species status	–	sp. nov.
Species etymology	–	*Oecophyllibacter saccharovorans* (sac.cha.ro.vo'rans. Gr. fem. n. *sakchar* sugar; L. pres. part. *vorans* eating; N.L. part adj. *saccharovorans* sugar eating)
Description of the new taxon and diagnostic traits	Gram-stain negative, rods forming, aerobic, non-motile. Catalase positive and oxidase negative. Major respiratory quinone is ubiquinone Q-10. Predominant cellular fatty acids are $C_{18:1}$ $\omega 7c$, $C_{16:0}$, $C_{19:0}$ $\omega 8c$ cyclo, $C_{14:0}$ and $C_{16:0}$ 2-OH. Chemoorganotrophic metabolism. The genus is affiliated to the family *Acetobacter*aceae, order Rhodospirillales, class Alphaproteobacteria	Cells are rod-shaped that occur singly, 0.45 – 0.5 µm in width and 0.8 – 1.0 µm in length after 24 hr incubation in LMG medium 404 broth. Colonies are beige-colored, small (approximately 1-3 mm), circular, convex, shiny and smooth after 2 days incubation at 28 °C on LMG medium 404. Growth occurs at 15-37 °C (optimum 28-30 °C) and with 0-1.5% (w/v) NaCl (optimum 0% (w/v)). Show glucose requirement for growth and able to grow in 1% and 30% glucose (w/v) media with 1% yeast extract. Growth occurs in the presence of 0.35% (v/v) acetic acid and 1% KNO_3 (w/v). Produce 2-keto-D-gluconic acid but not 5-keto-D-gluconic acid from D-glucose. Do not oxidize ethanol to acetic acid. Do not oxidize lactate and acetate to carbon dioxide and water. No growth with ammonium as the sole nitrogen source. Do not produce

		water-soluble brown pigment. Weak in production of dihydroxyacetone from glycerol. No growth on 1% glutamate agar. Do not utilize methanol as sole carbon source. Produce acids from L-arabinose, D-glucose, D-galactose, D-mannose, D-fructose, melibiose, D-mannitol and ethanol, but not from D-arabinose, sucrose, raffinose, maltose, methanol, D-sorbitol and glycerol.
Country of origin	–	Malaysia
Region of origin	–	Kuala Lumpur
Date of isolation	–	20/6/2017
Source of isolation	–	Weaver ant *Oecophylla smaragdina*
Sampling date	–	14/6/2017
Latitude	–	3°07'21.9"N
Longitude	–	101°39'22.7"E
16S rRNA gene accession nr.	–	MG757796
Genome accession number	–	CP038143.1
Genome status	–	Complete
Genome size (bp)	–	1,945,529
GC mol%	–	61.56
Number of strains in study	–	3
Source of isolation of non-type strains	–	Weaver ant *Oecophylla smaragdina* from different colonies and geographical locations
Information related to the Nagoya Protocol	–	Not applicable
Designation of the Type Strain	–	Ha5^T
Strain Collection Numbers	–	DSM106907^T = LMG30589^T = NBRC 113643^T = KCTC62952^T

Fig. 1. Map of pHa5 in strain Ha5^T carrying the sole and complete *rrn* operon in its genome. The tRNAs were shown in black. Restriction site of EcoRV is marked with a red "E".

Fig. 2. Maximum likelihood tree showing the phylogenetic relationship of strains Ha5^T, Ta1, Jb2 and their closest relatives in the family *Acetobacteraceae*. The tree was constructed based on 16S rRNA gene sequences. Number displayed at nodes are bootstrap values, based on 1000 replications. Numbers in brackets represent the number of type strains included in the respective clusters. The tree was rooted with genera *Stella* and *Zavarzinia* as outgroups.

Fig. 3. Phylogenomic trees of 96 concatenated core gene sequences revealed from whole genomes of family *Acetobacteraceae* showing the relationship of strains Ha5^T, Ta1, Jb2 and their closest relatives in the family *Acetobacteraceae*. The tree was constructed using maximum likelihood method. The values at the nodes indicate bootstrap values expressed as percentages of 100 replications while the bar length indicates 0.20 substitutions per site.

Fig. 4. Average amino acid identity (AAI) of strains Ha5^T, Ta1, Jb2 and their closest relatives in the family *Acetobacteraceae*. AAI values lesser than 95% are colored in blue, indicating that the strains are of different genera. Color intensity fades as the AAI values approach 95%. AAI values equal or greater than 95% are colored in red, indicating that these strains belong to the same genus. Color intensity strengthens as the AAI values approach 100%.

Fig. 5. Cellular morphology of strain Ha5^T. (a) Gram-stained cells of Ha5^T under light microscope. Bar, 10 µm. (b) Scanning electron micrograph (SEM) showing the rod-shaped cells of strain Ha5^T. Bar, 2 µm.

Fig. 1

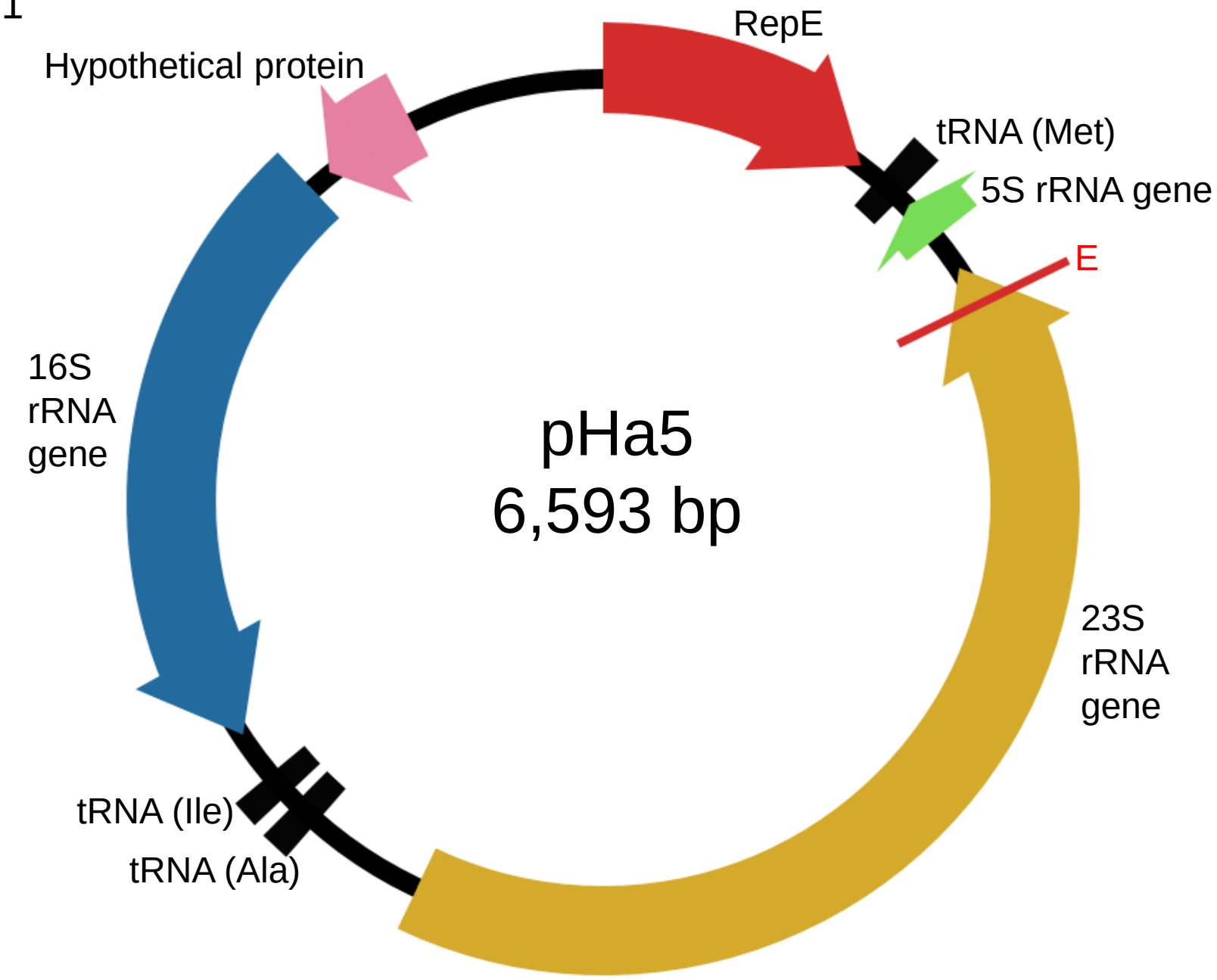

RepE
Hypothetical protein
tRNA (Met)
5S rRNA gene
E
16S
rRNA
gene
pHa5
6,593 bp
23S
rRNA
gene
tRNA (Ile)
tRNA (Ala)

Fig. 2

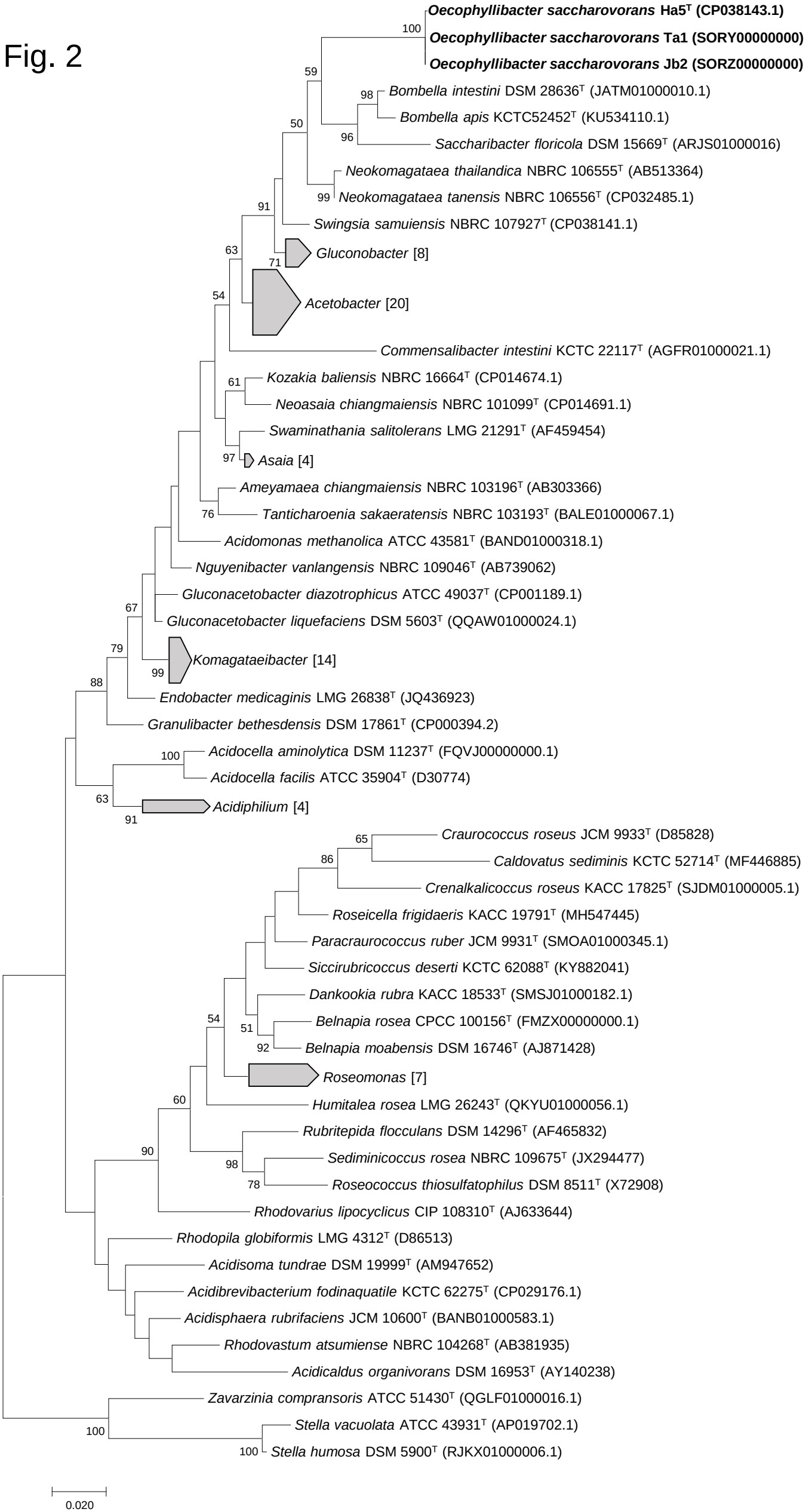

Oecophyllibacter saccharovorans Ha5T (CP038143.1)
Oecophyllibacter saccharovorans Ta1 (SORY00000000)
Oecophyllibacter saccharovorans Jb2 (SORZ00000000)
Bombella intestini DSM 28636T (JATM01000010.1)
Bombella apis KCTC52452T (KU534110.1)
Saccharibacter floricola DSM 15669T (ARJS01000016)
Neokomagataea thailandica NBRC 106555T (AB513364)
Neokomagataea tanensis NBRC 106556T (CP032485.1)
Swingsia samuiensis NBRC 107927T (CP038141.1)
Gluconobacter [8]
Acetobacter [20]
Commensalibacter intestini KCTC 22117T (AGFR01000021.1)
Kozakia baliensis NBRC 16664T (CP014674.1)
Neoasaia chiangmaiensis NBRC 101099T (CP014691.1)
Swaminathania salitolerans LMG 21291T (AF459454)
Asaia [4]
Ameyamaea chiangmaiensis NBRC 103196T (AB303366)
Tanticharoenia sakaeratensis NBRC 103193T (BALE01000067.1)
Acidomonas methanolica ATCC 43581T (BAND01000318.1)
Nguyenibacter vanlangensis NBRC 109046T (AB739062)
Gluconacetobacter diazotrophicus ATCC 49037T (CP001189.1)
Gluconacetobacter liquefaciens DSM 5603T (QQAW01000024.1)
Komagataeibacter [14]
Endobacter medicaginis LMG 26838T (JQ436923)
Granulibacter bethesdensis DSM 17861T (CP000394.2)
Acidocella aminolytica DSM 11237T (FQVJ00000000.1)
Acidocella facilis ATCC 35904T (D30774)
Acidiphilium [4]
Craurococcus roseus JCM 9933T (D85828)
Caldovatus sediminis KCTC 52714T (MF446885)
Crenalkalicoccus roseus KACC 17825T (SJDM01000005.1)
Roseicella frigidaeris KACC 19791T (MH547445)
Paracraurococcus ruber JCM 9931T (SMOA01000345.1)
Siccirubricoccus deserti KCTC 62088T (KY882041)
Dankookia rubra KACC 18533T (SMSJ01000182.1)
Belnapia rosea CPCC 100156T (FMZX00000000.1)
Belnapia moabensis DSM 16746T (AJ871428)
Roseomonas [7]
Humitalea rosea LMG 26243T (QKYU01000056.1)
Rubritepida flocculans DSM 14296T (AF465832)
Sediminicoccus rosea NBRC 109675T (JX294477)
Roseococcus thiosulfatophilus DSM 8511T (X72908)
Rhodovarius lipocyclicus CIP 108310T (AJ633644)
Rhodopila globiformis LMG 4312T (D86513)
Acidisoma tundrae DSM 19999T (AM947652)
Acidibrevibacterium fodinaquatile KCTC 62275T (CP029176.1)
Acidisphaera rubrifaciens JCM 10600T (BANB01000583.1)
Rhodovastum atsumiense NBRC 104268T (AB381935)
Acidicaldus organivorans DSM 16953T (AY140238)
Zavarzinia compransoris ATCC 51430T (QGLF01000016.1)
Stella vacuolata ATCC 43931T (AP019702.1)
Stella humosa DSM 5900T (RJKX01000006.1)
0.020

Fig. 3

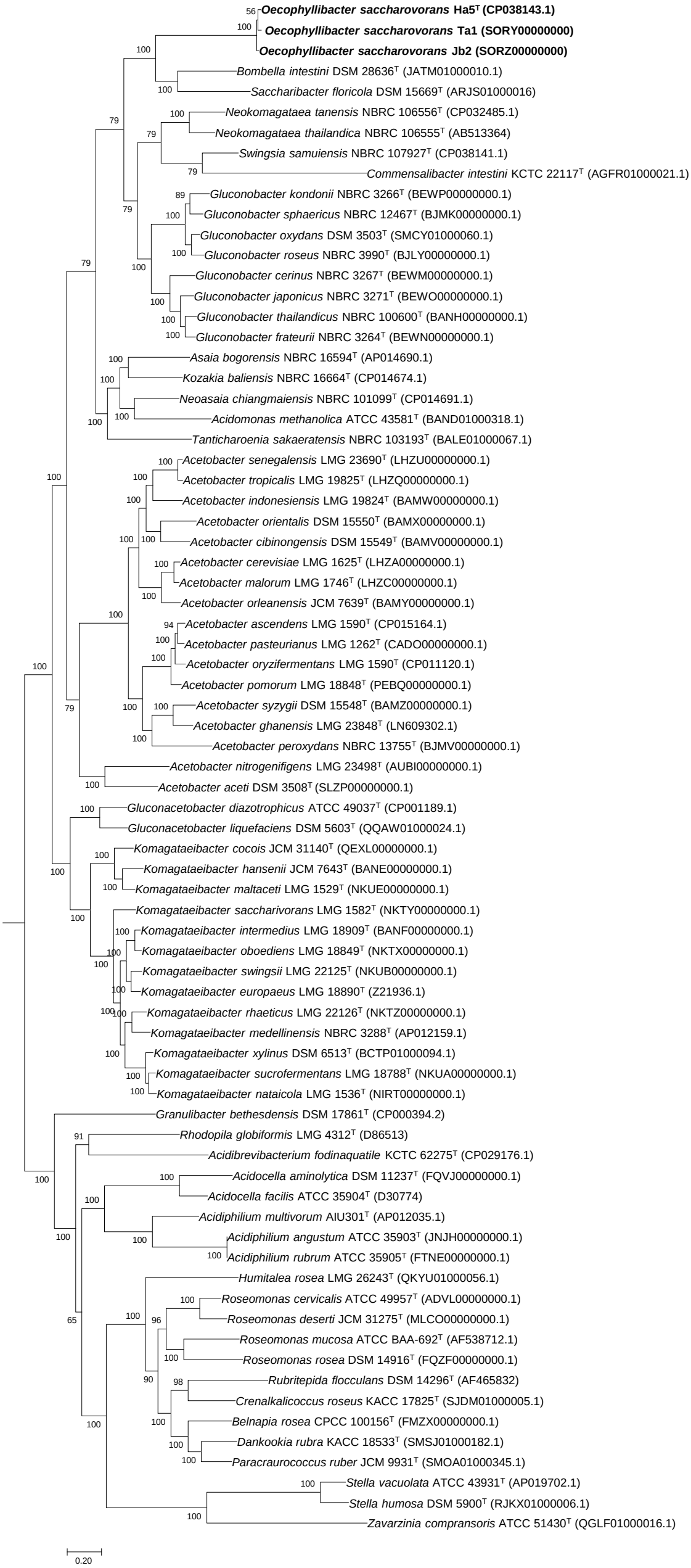
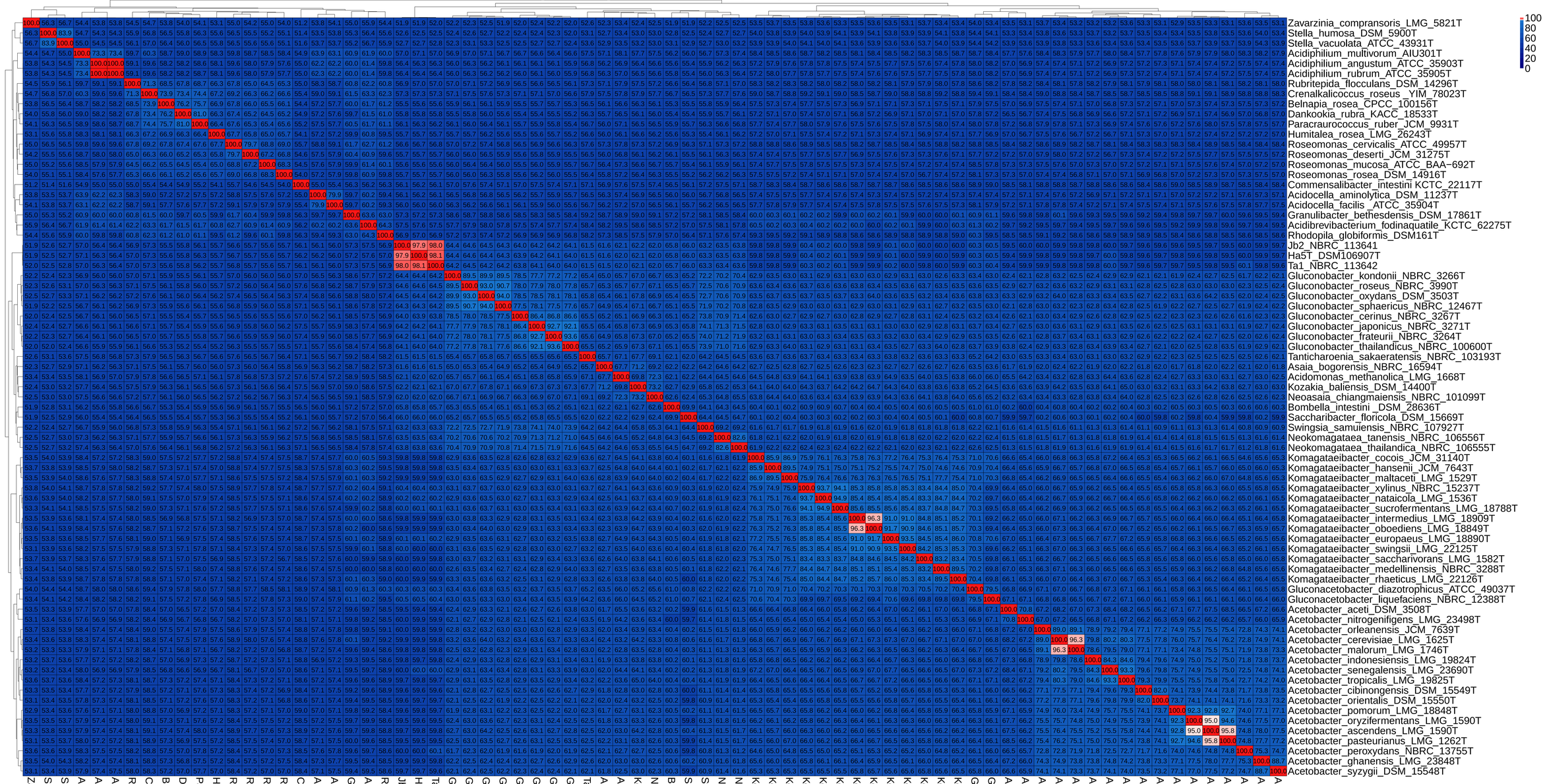

Oecophyllibacter saccharovorans Ha5T (CP038143.1)
Oecophyllibacter saccharovorans Ta1 (SORY00000000)
Oecophyllibacter saccharovorans Jb2 (SORZ00000000)
Bombella intestini DSM 28636T (JATM01000010.1)
Saccharibacter floricola DSM 15669T (ARJS01000016)
Neokomagataea tanensis NBRC 106556T (CP032485.1)
Neokomagataea thailandica NBRC 106555T (AB513364)
Swingsia samuiensis NBRC 107927T (CP038141.1)
Commensalibacter intestini KCTC 22117T (AGFR01000021.1)
Gluconobacter kondonii NBRC 3266T (BEWP00000000.1)
Gluconobacter sphaericus NBRC 12467T (BJMK00000000.1)
Gluconobacter oxydans DSM 3503T (SMCY01000060.1)
Gluconobacter roseus NBRC 3990T (BJLY00000000.1)
Gluconobacter cerinus NBRC 3267T (BEWM00000000.1)
Gluconobacter japonicus NBRC 3271T (BEWO00000000.1)
Gluconobacter thailandicus NBRC 100600T (BANH00000000.1)
Gluconobacter frateurii NBRC 3264T (BEWN00000000.1)
Asaia bogorensis NBRC 16594T (AP014690.1)
Kozakia baliensis NBRC 16664T (CP014674.1)
Neoasaia chiangmaiensis NBRC 101099T (CP014691.1)
Acidomonas methanolica ATCC 43581T (BAND01000318.1)
Tanticharoenia sakaeratensis NBRC 103193T (BALE01000067.1)
Acetobacter senegalensis LMG 23690T (LHZU00000000.1)
Acetobacter tropicalis LMG 19825T (LHZQ00000000.1)
Acetobacter indonesiensis LMG 19824T (BAMW00000000.1)
Acetobacter orientalis DSM 15550T (BAMX00000000.1)
Acetobacter cibinongensis DSM 15549T (BAMV00000000.1)
Acetobacter cerevisiae LMG 1625T (LHZA00000000.1)
Acetobacter malorum LMG 1746T (LHZC00000000.1)
Acetobacter orleanensis JCM 7639T (BAMY00000000.1)
Acetobacter ascendens LMG 1590T (CP015164.1)
Acetobacter pasteurianus LMG 1262T (CADO00000000.1)
Acetobacter oryzifermentans LMG 1590T (CP011120.1)
Acetobacter pomorum LMG 18848T (PEBQ00000000.1)
Acetobacter syzygii DSM 15548T (BAMZ00000000.1)
Acetobacter ghanensis LMG 23848T (LN609302.1)
Acetobacter peroxydans NBRC 13755T (BJMV00000000.1)
Acetobacter nitrogenifigens LMG 23498T (AUBI00000000.1)
Acetobacter aceti DSM 3508T (SLZP00000000.1)
Gluconacetobacter diazotrophicus ATCC 49037T (CP001189.1)
Gluconacetobacter liquefaciens DSM 5603T (QQAW01000024.1)
Komagataeibacter cocois JCM 31140T (QEXL00000000.1)
Komagataeibacter hansenii JCM 7643T (BANE00000000.1)
Komagataeibacter maltaceti LMG 1529T (NKUE00000000.1)
Komagataeibacter saccharivorans LMG 1582T (NKTY00000000.1)
Komagataeibacter intermedius LMG 18909T (BANF00000000.1)
Komagataeibacter oboediens LMG 18849T (NKTX00000000.1)
Komagataeibacter swingsii LMG 22125T (NKUB00000000.1)
Komagataeibacter europaeus LMG 18890T (Z21936.1)
Komagataeibacter rhaeticus LMG 22126T (NKTZ00000000.1)
Komagataeibacter medellinensis NBRC 3288T (AP012159.1)
Komagataeibacter xylinus DSM 6513T (BCTP01000094.1)
Komagataeibacter sucrofermentans LMG 18788T (NKUA00000000.1)
Komagataeibacter nataicola LMG 1536T (NIRT00000000.1)
Granulibacter bethesdensis DSM 17861T (CP000394.2)
Rhodopila globiformis LMG 4312T (D86513)
Acidibrevibacterium fodinaquatile KCTC 62275T (CP029176.1)
Acidocella aminolytica DSM 11237T (FQVJ00000000.1)
Acidocella facilis ATCC 35904T (D30774)
Acidiphilium multivorum AIU301T (AP012035.1)
Acidiphilium angustum ATCC 35903T (JNJH00000000.1)
Acidiphilium rubrum ATCC 35905T (FTNE00000000.1)
Humitalea rosea LMG 26243T (QKYU01000056.1)
Roseomonas cervicalis ATCC 49957T (ADVL00000000.1)
Roseomonas deserti JCM 31275T (MLCO00000000.1)
Roseomonas mucosa ATCC BAA-692T (AF538712.1)
Roseomonas rosea DSM 14916T (FQZF00000000.1)
Rubritepida flocculans DSM 14296T (AF465832)
Crenalkalicoccus roseus KACC 17825T (SJDM01000005.1)
Belnapia rosea CPCC 100156T (FMZX00000000.1)
Dankookia rubra KACC 18533T (SMSJ01000182.1)
Paracraurococcus ruber JCM 9931T (SMOA01000345.1)
Stella vacuolata ATCC 43931T (AP019702.1)
Stella humosa DSM 5900T (RJKX01000006.1)
Zavarzinia compransoris ATCC 51430T (QGLF01000016.1)
0.20

Fig. 4. Average Amino Acid Identity (AAI)



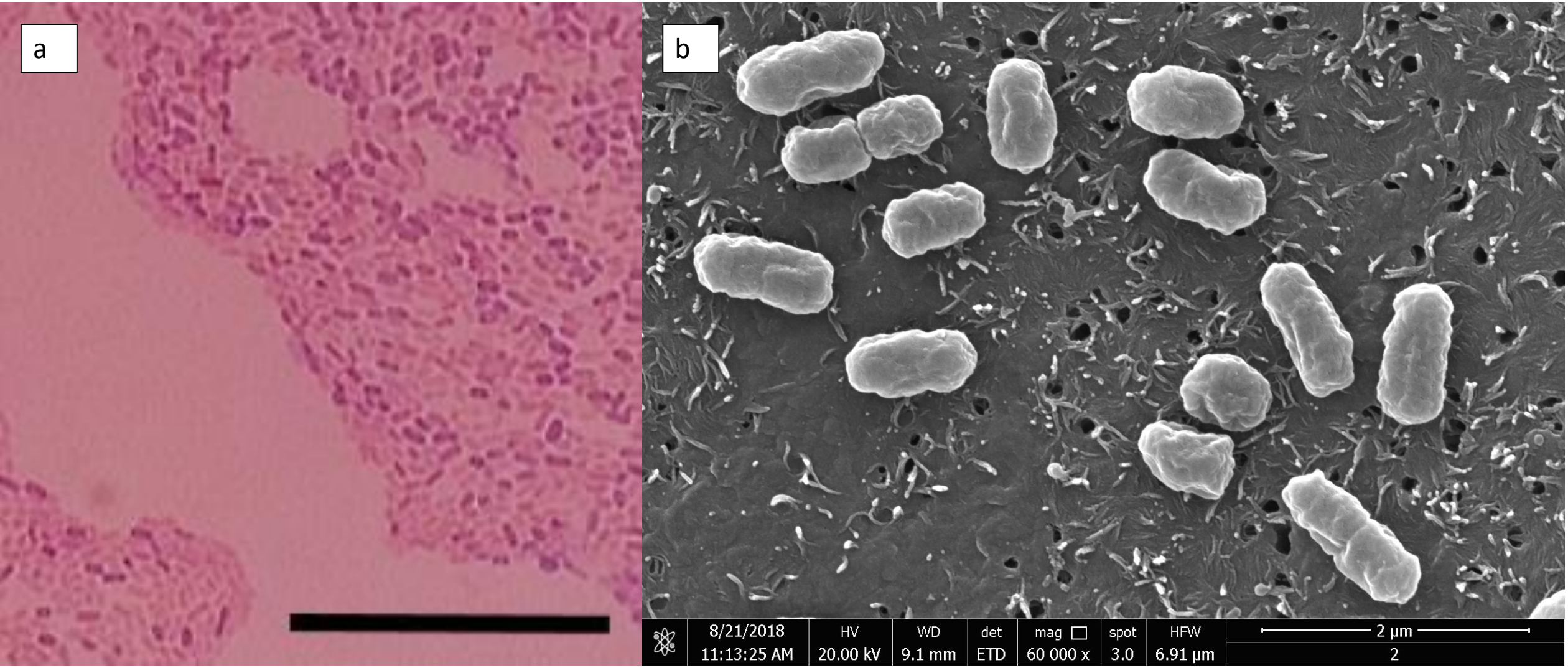

a
b
8/21/2018
11:13:25 AM
HV
20.00 kV
WD
9.1 mm
det
ETD
mag
60 000 x
spot
3.0
HFW
6.91 µm
2 µm
2

Fig. 5