Journal of Academia Vol. 9, Issue 1 (2021) 97 – 105

PHYLOGENETIC ANALYSIS OF *Komagataeibacter* sp.: A CELLULOSE-PRODUCER BACTERIA BASED ON 16S rRNA GENE SEQUENCES

Nur Aisyah Atikah Alizan, Sarah S. Zakaria*

School of Biology, Faculty of Applied Sciences, Universiti Teknologi MARA, Cawangan Negeri Sembilan, Kampus Kuala Pilah, 72000 Kuala Pilah, Negeri Sembilan, Malaysia

**Corresponding author: shazwani@uitm.edu.my*

Abstract

Bacteria of the genus *Komagataeibacter* are described to be the most noteworthy for having several of its species being efficient and strong cellulose producers. The 16S ribosomal RNA (rRNA) gene analysis is often used for the identification and taxonomic classification of these bacteria species. In order to observe the phylogenetic relationship among *Komagataeibacter* sp., twelve sequences of the 16S rRNA gene with three sequences each for species namely *Komagataeibacter europaeus*, *Komagataeibacter hansenii*, *Komagataeibacter intermedius* and *Komagataeibacter xylinus* were retrieved from NCBI GenBank database. The sequences were aligned and analysed using PAUP, OrthoANI and BLAST, followed by the phylogenetic tree construction using a Maximum Likelihood method. The parsimony character diagnostic analysis showed very few numbers of parsimony-informative characters present in the aligned sequences which is only 1.5% of the total characters. The inferred phylogenetic relationships demonstrated the unexpected positioning of *K. xylinus* (GQ240638: *Gluconacetobacter xylinus* strain) and *K. xylinus* (KC11853: *G. xylinus* strain) into the clades of *K. europaeus* and *K. hansenii* respectively. The also very low bootstrap values of the branch points linking the *K. europaeus* species indicated low support for the produced topologies. The findings of this study indicate that more phylogenies information can be attained by increasing the taxon sampling. In addition, more robust molecular data are needed to infer the phylogenetic relationships between the *Komagataeibacter* species more accurately.

Keywords: Phylogenetics, 16S rRNA, *Acetobacter*, Bacterial cellulose, *Komagataeibacter*

Article History:- Received: 21 October 2020; Accepted: 11 March 2021; Published: 30 April 2021

Introduction

Bacterial cellulose (BC) is a polysaccharide produced by bacteria from genera such as *Acetobacter* and *Komagataeibacter*. The polymer consists of repeating units of glucose that are linked by β (1–4) glycosidic bonds (Rehm, 2010). The exceptional physicochemical and mechanical properties of BC make it an attractive alternative for plant-produced cellulose. They possess properties such as high purity, high tensile strength, high water-holding capacity, high crystallinity as well as being biodegradable (Reiniati *et al.*, 2017). BC is preferred over the plant cellulose due to its purity from any substances that are typically found in cellulose that was extracted from plants, such as hemicellulose, lignin and pectin (Cheng *et al.*, 2009).

The bacteria *Acetobacter* sp. and *Komagataeibacter* sp. are known to be efficient in cellulose production and are currently common for their use in industrial vinegar fermentation as they are able to oxidise ethanol and can resist high concentrations of ethanol and acetic acid (Barja *et al.*, 2016; Ryngajłło *et al.*, 2019). Cellulose-producing bacteria can be isolated from various habitats that are rich in sugar and alcohol such as fruits, flowers as well as fermented products (Naloka *et al.*, 2018). In order to identify these isolates, molecular analysis from 16S ribosomal RNA gene sequencing is described to be a reliable and widely used method as reported in previous related studies. The 16S rRNA analysis had been shown

to be an appropriate technique to be used for the taxonomic classification of microorganisms (Yang *et al.*, 2016). Additionally, the size of the 16S rRNA gene being approximately 1,500 bp is large enough to contain information that are statistically relevant possible to be used for informatics purposes (Janda and Abbott, 2007). From this analysis, phylogenetic trees can be reconstructed to observe the evolutionary relationships among the cellulose-producing bacteria. This study focuses on analysing the phylogenetic relationships of *Komagataeibacter* species and the reliability of its classification based on the phylogenetic marker 16S rRNA gene.

The identification of BC-producing species through morphological, structural and biochemical characterisations may often be affected by various factors such as the culture conditions used during the synthesis of BC (Marsh *et al.*, 2014). The procedures for conducting morphological and biochemical tests are also tedious which may result in the misidentification of bacterial species. Other than that, the classification of species on the species-level is difficult if conducted only based on morphological and biochemical characters. To conduct phenotypic identifications of microorganisms, a database of accurate morphological and biochemical description is needed. However, in some instances, the morphology or phenotypic observations obtained may not match the available database and thus judgment of the most probable species must be made. These judgements could vary among individuals conducting the study even with the help of technology such as computer programs and thus render unreliable (Clarridge, 2004).

Due to the many benefits of BC, proper and reliable identification and classification of cellulose-producing bacterial species are very important in the efforts of discovering the most prominent cellulose producer. As this study involves the comparison of 16S rRNA gene sequences and the analysis of evolutionary relationships between the species, the results of this study could offer a reliable method for identification and classification of the cellulose-producing bacteria, *Komagataeibacter* sp. From multiple sequence alignment, the conserved regions and sequence similarities identified within the 16S rRNA gene would also be useful for future references and studies in determining evolutionary relationships of this species.

Methods

Data retrieval

A total of twelve sequences of the 16S rRNA gene of *Komagataeibacter* sp. were chosen and retrieved from the NCBI GenBank database website (https://www.ncbi.nlm.nih.gov/genbank/). Four *Komagataeibacter* species were selected for the sequence analysis namely *K. europaeus* (Z21936; FN429075; Y15269), *K. hansenii* (AM999342; NR_112227; NR_043112), *K. intermedius* (AJ316550; NR_113394; NR_118180) and *K. xylinus* (GQ240638; KC118537; AB161453). The length of the sequence chosen was between the range of 1300 to 1500 base pairs long. For each species, three sequences of the 16S rRNA gene were obtained for comparison purposes. The sequences were downloaded from the website and stored in FASTA format that is compatible to be used in many bioinformatics tools. The reconstructed phylogenetic trees were saved in the PNG format for results presentation.

Multiple sequence alignment and evaluations

Once the nucleotide sequences of 16S rRNA gene of each *Komagataeibacter* species were compiled, multiple sequence alignment was conducted using the ClustalW function of the software MEGA X (Kumar *et al.*, 2018). Through this method, significant similarities in the sequences could be detected and sorted accordingly. Additional quantitative analysis on nucleotide composition homogeneity between taxa was undertaken using simple chi-square test in PAUP* 4.0b10, a widely used phylogenetic software package written by Wilgenbusch and Swofford (2003) to infer evolutionary trees. The analysis was conducted with maximum parsimony as the optimality criterion method. Analysis for average nucleotide identity (ANI) of the twelve *Komagataeibacter* species were also conducted using the software OrthoANI to determine the genetic relatedness between the 16S rRNA gene sequences (Lee *et al.*, 2016).

Reconstruction of a phylogenetic tree

The aligned sequences of the 16S rRNA gene of the *Komagataeibacter* species were then used to reconstruct a phylogenetic tree. The tree was reconstructed using Maximum Likelihood, a method most often used in publications (Marič *et al.*, 2020; Matsutani *et al.*, 2015). The confidence values of individual branches in the reconstructed phylogenetic tree were calculated using the bootstrap analysis based on 1,000 replications.

Result and Discussion

Twelve sequences of the 16S rRNA gene of the selected *Komagataeibacter* species that are 1215 to 1482 bp long retrieved from the NCBI GenBank database were aligned by ClustalW using the MEGA X software. The quantitative analysis on nucleotide composition homogeneity analysed in PAUP* was conducted under equal weights with gaps treated as missing and maximum parsimony as the optimality criterion method. Maximum parsimony is a character-state method commonly used in molecular phylogenetics that relies on the use of optimality criterion. According to Swofford and Sullivan (2003), this criterion-based method is believed to be advantageous as the basis for preferring a tree over another is programmed to be mathematically precise compared to algorithmic methods. The character-state approach is also claimed to be superior compared to distance method due to the use of raw data. The underlying idea of parsimony analysis is that the tree which minimises the amount of evolutionary change (the total number of character-state changes) required to explain the data is considered the best (Swofford and Sullivan, 2003). Ordered characters are characters of which the order has been determined and transformation between any two adjacent states of this kind would cost the same number of steps, while transformation between non-adjacent states costs the sum of the steps between their implied adjacent states. Unordered characters on the other hand are characters that can transform into any other state with equal cost without having to pass through intermediate states (Kitching *et al.*, 1998).

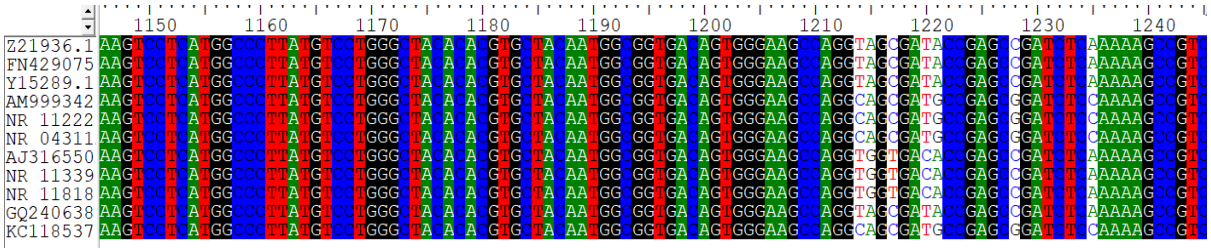

Figure 1 Multiple sequence alignment (MSA) of partial 16S rRNA gene sequences from twelve *Komagataeibacter* species. The conserved residues have been shaded and the non-identical residues were unshaded such as position site 1214th to 1215th. *K. europaeus* (Z21936: *A. europaeus* strain); *K. europaeus* (FN429075: *G. europaeus* strain); *K. europaeus* (Y15289: *A. europaeus* strain); *K. hansenii* (AM999342: *G. kombuchae* strain); *K. hansenii* (NR_112227); *K. hansenii* (NR_043112); *K. intermedius* (AJ316550: *G. intermedius* strain); *K. intermedius* (NR_113394); *K. intermedius* (NR_118180); *K. xylinus* (GQ240638: *G. xylinus* strain); *K. xylinus* (KC118537: *G. xylinus* strain) and *K. xylinus* (AB161453: *G. xylinus* strain).

The result of the PAUP* analysis indicated that all characters were unordered. The alignment of the twelve taxa (Figure 1) showed that of 1495 total characters, 23 were parsimony-informative, 661 characters were variable and parsimony-uninformative and a total of 822 were invariable characters. The parsimony-informative characters are only 1.5% of the total characters. The parsimony-informative sites refer to those that consist of at least two nucleotide or amino acid types with at least two of them occurring a minimum frequency of two (Syed Ibrahim *et al.*, 2017). The alignment of the sequences would reveal parsimony informative sites which is important in analysing phylogenies as according to Fitch (1977), only the parsimony-informative sites are useful in producing parsimonious trees (Papathanassopoulou and Lorentzos, 2014). As such, the lack of parsimony-informative sites in the sequences being studied may have contributed to the low resolution of the phylogenetic tree and provides little information on relationships of the species. Conversely, small numbers of informative sites result in trees being constructed shorter than it is and infers false relationships due to homoplasies being misconstrued as historical signals (Ribeiro *et al.*, 2012).

Phylogenetic trees convey a lot of information about evolutionary relationships from which the origins, evolution, as well as the possible changes in the functional or structural properties of studied genes, could be effectively inferred from. The phylogenetic tree reconstructed based on the 16S rRNA gene sequences of twelve *Komagataeibacter* species retrieved from the NCBI GenBank database by using Mega X software is shown in Figure 2.

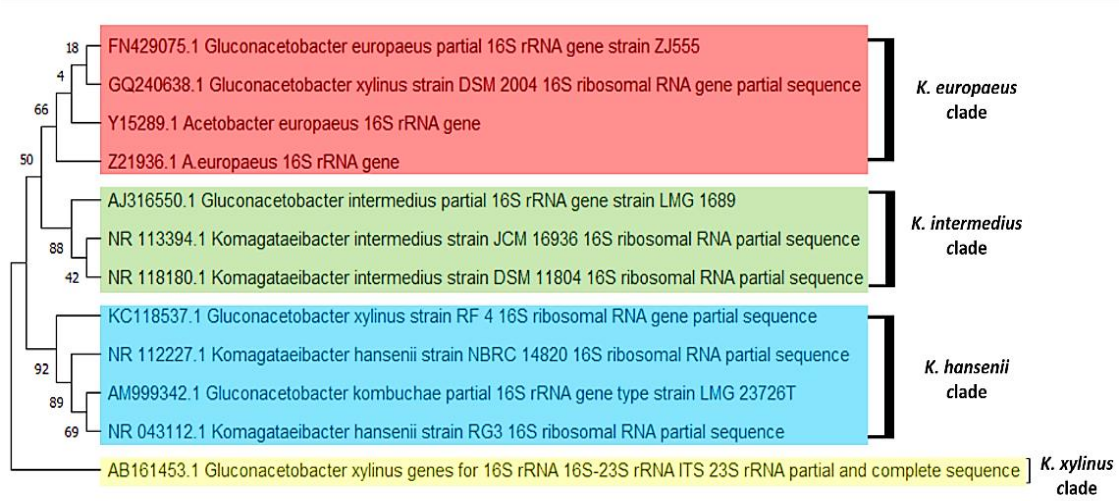

Figure 2. Maximum Likelihood tree reconstructed from 16S rRNA gene sequences showing four distinct clades: the *K. europaeus, K. intermedius, K. hansenii* and *K. xylinus* clades with *Gluconacetobacter xylinus* strain (AB161453: a *K. xylinus* species group) as outgroup species.

The yielded tree displays four clades, the *K. europaeus*, *K. intermedius*, *K. hansenii* and *K. xylinus* (Figure 2). From the phylogenetic tree, *G. xylinus* strain (AB161453, yellow highlighted), a *K. xylinus* species group, branched off nearest to the phylogenetic tree base. Sequence similarity values between the twelve species sequences were calculated using the BLAST program. The sequence similarity between *G. xylinus* strain (AB161453; yellow highlighted) with other species sequences were all 100% except with the *K. hansenii* (NR_043112) which obtained 96.88% identity. From the three selected sequences of *K. xylinus*, *K. xylinus* (GQ240638: *G. xylinus* strain) was unexpectedly clustered together with the *K. europaeus* species clade (in red box), while the *G. xylinus* strain (KC118537) was placed close together with the cluster of *K. hansenii* species clade (in blue box). The sequences similarity calculated between the *G. xylinus* strain (GQ240638) and *K. europaeus* species group, *G. europaeus* strain (FN429075), strains *Acetobacter europeaus* (Y15289) and *Acetobacter europaeus* (Z21936) are 99.92%, 99.92% and 99.85% respectively. Based on these high percentages and the position of *K. xylinus* strain (GQ240638: *G. xylinus* strain) in the phylogenetic tree, this could indicate that the species is closely related to *K. europaeus* species as it meets the threshold of 98-99%.

As for the *K. xylinus* (KC118537: *G. xylinus* strain) that had been clustered in the same clade as *K. hansenii* species, the sequence similarity obtained between the *K. xylinus* (KC118537: *G. xylinus* strain) and the three *K. hansenii* species sequences: *K. hansenii* (NR_112227), *K. hansenii* (NR_043112) and *K. hansenii* (AM999342: *G. kombuchae* strain) are 99.54%, 99.47% and 99.54% respectively. This result is thus indicative that the *K. xylinus* (KC118537: *G. xylinus* strain) is closely related to the *Komagataeibacter hansenii* species. The phylogenetic analysis using BLAST program also revealed that the sequence similarity between the two *K. xylinus* (KC118537: *G. xylinus* strain) and *K. xylinus* (GQ240638: *G. xylinus* strain) is 98.66% which is lower than the sequence identity obtained comparing those two *K. xylinus* strains (KC118537 and GQ240638: *G. xylinus* strains) with the *K. hansenii* and *K. europaeus* species, respectively, further supporting the phylogenetic tree branching.

The species grouped under the same clade of the phylogenetic tree can be defined as operational taxonomic unit (OTU) clusters as it meets the 97% identity threshold. Individual species located in different OTUs are described to be impossible to belong to the same bacterial species. Another

interesting finding from the phylogenetic analysis results showed that the *K. Intermedius* strain (AJ316550: *G. intermedius* strain) had a sequence similarity of 100% indicating homology with the base positioned *K. xylinus* (AB161453: *G. xylinus* strain).

Further ANI analysis is conducted using OrthoANI program (Lee *et al.*, 2016) to measure the similarity between the twelve taxa of 16S rRNA gene sequences. This analysis is considered the best approach and is chosen because a large number of genes is included in its calculation, including both slow and fast-evolving genes which minimise the effect of variable evolutionary rates or horizontal gene transfer events (Konstantinidis and Tiedje, 2005; Rashid *et al.*, 2015). Several previous studies have shown good correlation with values obtained through DNA-DNA hybridisation where the standard > 70% DDH cut-off value for delineating species corresponds to > 95-96% of ANI values (Tortoli *et al.*, 2019). The ANI values between the taxa calculated in this study are shown in Table 1.

Table 1. OrthoANI values between the 16S rRNA gene sequences of twelve *Komagataeibacter* species (%).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	100											
2	99.90	100										
3	99.80	100	100									
4	98.80	98.89	98.90	100								
5	98.80	98.90	98.90	99.80	100							
6	98.59	98.80	98.69	99.80	99.6	100						
7	99.41	99.49	99.41	98.60	98.60	98.39	100					
8	99.60	99.70	99.70	98.82	98.82	98.62	99.80	100				
9	99.58	99.69	99.68	98.76	98.76	98.76	99.79	100	100			
10	99.80	99.90	99.90	98.81	98.81	98.70	99.39	99.60	99.59	100		
11	98.74	98.89	98.85	99.38	99.38	99.28	98.53	98.77	98.62	98.78	100	
12	0	0	0	0	0	0	0	0	0	0	0	100

Taxa: 1, *K. europaeus* (Z21936: *A. europaeus* strain); 2, *K. europaeus* (FN429075: *G. europaeus* strain); 3, *K. europaeus* (Y15289: *A. europaeus* strain); 4, *K. hansenii* (AM999342: *G. kombuchae* strain); 5, *K. hansenii* (NR_112227); 6, *K. hansenii* (NR_043112); 7, *K. intermedius* (AJ316550: *G. intermedius* strain); 8, *K. intermedius* (NR_113394); 9, *K. intermedius* (NR_118180); 10, *K. xylinus* (GQ240638: *G. xylinus* strain); 11, *K. xylinus* (KC118537: *G. xylinus* strain); 12, *K. xylinus* (AB161453: *G. xylinus* strain). The species *G. kombuchae* was reclassified as a later heterotypic synonym of *K. hansenii* (Semjonovs *et al.*, 2017).

The ANI values calculated between the taxa range from 98 to 100% for all except the values for *K. xylinus* (AB161453**:** *G. xylinus* strain) and other *Komagataeibacter* taxa which revealed unexpected values of 0. Additionally, the ANI values between *K. xylinus* (KC118537: *G. xylinus* strain) and the *K. hansenii* species with the accession number of AM999342, NR_112227 and NR_043112 are 99.38%, 99.38% and 99.20% respectively. This high percentage further supports the inferred relationship between these species that had been displayed in the topology of the phylogenetic tree. According to Jain *et al.* (2018), 95% ANI cut-off is the most commonly used standard for species demarcation.

The robustness of the reconstructed phylogenetic tree is tested by bootstrapping in order to determine how reliable the data is in supporting the phylogenetic hypothesis. Bootstrap is the most popular and widely used method to assess branch support (Morrison, 2006). The calculated bootstrap values are displayed at the nodes of the phylogenetic tree, obtaining six out of nine bootstrap values that are higher than 50%. The three remaining bootstrap values are however very low. The branch point separating the strains *K. intermedius* DSM 11804 (NR_118180) and *K. intermedius* JCM 16936 (NR_113394) obtained bootstrap value of 42%. The bootstrap analysis also revealed that the node linking the *A. europaeus*

(Y15289) with *G. europaeus* strain (FN429075) and *G. xylinus* strain (GQ240638) had very low support value of 4% and the node linking *G. europaeus* strain (FN429075) and *G. xylinus* strain (GQ240638) with each other obtained 18%. The bootstrap value obtained for the branch point connecting *K. xylinus* (KC118537: *G. xylinus* strain) with the *K. hansenii* species is 92% which is the highest bootstrap value of the reconstructed tree. This indicates that the relationship between *K. xylinus* (KC118537: *G. xylinus* strain) and the *K. hansenii* inferred in this phylogenetic tree could be reliable.

As the calculation of bootstrap value is used for testing reliability of phylogenetic trees, the low bootstrap value would mean that the branch point or node concerned was built with low support of only a few characters and if the characters were removed at random, the matrix would lead to the node being differently reconstructed (Jill Harrison and Langdale, 2006). The obtained findings from the bootstrap analysis could be justified based on several factors. The low bootstrap value could indicate that the groups are not tightly coupled phylogenetically, and the members of the low bootstrap value branch could jump into other branches on every bootstrapping replication. Such low values are usually obtained when the information on a concerned branch is very limited in the alignment thus very few character sites could support the branch (Alkindy *et al.*, 2015). In the case of this study, low bootstrap values may be due to the insufficient number of informative sites that are available amongst the characters aligned which are only 23 (1.5%) of the total characters. The focus of this study was to analyse the 16S rRNA gene at species level in which the sufficient number of samples for species level studies are less than three samples (Martens *et al.*, 2008). That said, the number of collected 16S rRNA gene sequences of *Komagataeibacter* totalling up to only 12 sequences suggests that the low taxon sampling may have contributed to the relatively low branching point bootstrap values. According to Mariadassou *et al.* (2012), larger taxon sampling leads to the construction of more accurate phylogenies. This, however, depends on the position of the added taxon in the phylogeny where if an addition of a taxon causes long branches to break then it is shown to improve the tree stability while the addition of additional long branches could cause the stability and accuracy of inferred phylogeny to be hindered. Similarly, the addition of an outgroup can also cause the topologies of a tree to be disrupted (Shavit *et al.*, 2007).

Although the 16S rRNA gene being studied is assumed to be reliable due to the ribosomal component being species-specific and the gene is not influenced by horizontal gene transfer (HGT), several studies have now provided evidence of such horizontal exchange of 16S rRNA gene fragments occurring (Kitahara *et al.*, 2012; Sato and Miyazaki, 2017). The secondary structure disruption of 16S rRNAs found by Laios *et al.* (2004) explained by the high susceptibility of the gene for point mutations had made it reasonable to think that 16S rRNA genes had evolved with the help of HGT as the major driving force, although Tian *et al.* (2015) described the frequency of such event happening for 16S rRNA genes to be rare. Many studies had discovered the 16S rRNA derived phylogenies to be often conflicting with other reliable molecular and phenotyping data, which raises questions of the validity of prokaryotic systematics developed using 16S rRNA gene (Rastogi *et al.*, 2019). These findings which may not be similar in our case are however significant and worth mentioning in the effort of justifying the results obtained from the analyses conducted in this study.

Another possible factor for the low bootstrap analysis value obtained could also be the occurrence of long-branch attraction (LBA). Under certain conditions, a long-branch attraction, which refers to a systematic error in which distantly related lineages are incorrectly inferred to be closely related, could occur. This phenomenon causes phylogenetic methods to become misled by distinct branch length differences. LBA arises when two or more rapidly evolving lineages are concerned (Bybee, 2008). When there are sufficient molecular or morphological changes occurring in a lineage, it appears similar to another long-branched lineage or an outgroup taxon and is branched together only by the basis of both undergoing a large number of changes instead of actually being related by descent (Simpson, 2010; Soltis and Doyle, 2012). Although it is expected that the occurrence of LBA is reduced with the use of reliable methods such as Maximum Likelihood, the relatively rapid and diverse evolution rate of bacteria makes the possibility of the LBA phenomenon to have occurred worth considering.

Conclusion

In conclusion the region of 16S rRNA gene sequences of twelve Komagataeibacter species in this study suggested to be a conserved region of this gene as the analyses show high similarity among nucleotides. In addition, the phylogenetic relationships of only twelve Komagateibacter species using Maximum Likelihood method resulted in four distinguished species clades, the *K. europaeus*, *K. intermedius*, *K. hansenii* and *K. xylinus*. However, *Gluconacetobacter xylinus* (GQ240638) and *Gluconacetobacter xylinus (*KC11853) was unexpectedly positioned into the clades of *K. europaeus* and *K. hansenii*, respectively. Both strains *Gluconacetobacter xylinus* (GQ240638 and KC11853) were previously assumed to be identified as *Komagataeibacter xylinus* and were predicted to be clustered together with the other *K. xylinus* strains studied. To infer more robust phylogenetic relationships of *Komagataeibacter* sp., a high taxon sampling should be considered by retrieving more sequences from all sixteen species of *Komagataeibacter* sp. as well as species from closely related genera such as *Acetobacter* and *Gluconacetobacter*. This approach will increase the phylogenetic information and improve the resolution of phylogenetic relationships.

Acknowledgement

Authors would like to thank UiTM Negeri Sembilan Kuala Pilah Campus for the facilities and cooperation. We also express our appreciation to Ann Nurrizka Abd Hamid and Nur Amira Aminuddin for the assistance and support in the completion of this manuscript.

References

Alkindy, B., Al-Nuaimi, B., Guyeux, C., Couchot, J.-F., Salomon, M., Alsrraj, R., & Philippe, L. (2016). Binary article swarm optimization versus hybrid genetic algorithm for inferring well supported phylogenetic trees. *Computational Intelligence Methods for Bioinformatics and Biostatistics*, 165–179. https://doi.org/10.1007/978-3-319-44332-4_13

Barja, F., Andrés-Barrao, C., Ortega Pérez, R., Cabello, E. M., & Chappuis, M.-L. (2016). Physiology of *Komagataeibacter spp.* during acetic acid fermentation. *Acetic Acid Bacteria*, 201–221. https://doi.org/10.1007/978-4-431-55933-7_9

Bybee, S. M. (2008). Phylogenetics, evolution and systematics of Holodonata with special focus on wing structure evolution: morphological, molecular and fossil evidence. Doctoral dissertation. University of Florida.

Cheng, K.-C., Catchmark, J. M., & Demirci, A. (2009). Effect of different additives on bacterial cellulose production by *Acetobacter xylinum* and analysis of material property. *Cellulose*, *16*(6), 1033–1045. https://doi.org/10.1007/s10570-009-9346-5

Clarridge, J. E. (2004). Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clinical Microbiology Reviews*, *17*(4), 840–862. https://doi.org/10.1128/cmr.17.4.840-862.2004

Fitch, W. M. (1977). On the problem of discovering the most parsimonious tree. *The American Naturalist*, *111*(978), 223–257. https://doi.org/10.1086/283157

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 1-8. https://doi.org/10.1038/s41467-018-07641-9.

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, *45*(9), 2761–2764. https://doi.org/10.1128/jcm.01228-07

Jill Harrison, C., & Langdale, J. A. (2006). A step by step guide to phylogeny reconstruction. *The Plant Journal*, *45*(4), 561–572. https://doi.org/10.1111/j.1365-313x.2005.02611.x

Kitahara, K., Yasutake, Y., & Miyazaki, K. (2012). Mutational robustness of 16S ribosomal RNA, shown by experimental horizontal gene transfer in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *109*(47), 19220–19225. https://doi.org/10.1073/pnas.1213609109

Kitching, I. J., Forey, P. L., Humphries, C. J., & Williams, D. M. (1998). *Cladistics: The theory and practice of parsimony analysis (Oxford Science Publications)* (2nd ed.). Oxford University Press.

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences, 102*(7), 2567–2572. https://doi.org/10.1073/pnas.0409727102

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution, 35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Laios, E., Waddington, M., Saraiya, A. A., Baker, K. A., O'Connor, E., Pamarathy, D., & Cunningham, P. R. (2004). Combinatorial genetic technology for the development of new anti-infectives. *Archives of Pathology & Laboratory Medicine, 128*(12), 1351–1359. https://doi.org/10.5858/2004-128-1351-cgtftd

Lee, I., Ouk Kim, Y., Park, S.-C., & Chun, J. (2016). OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *International Journal of Systematic and Evolutionary Microbiology, 66*(2), 1100–1103. https://doi.org/10.1099/ijsem.0.000760

Mariadassou, M., Bar-Hen, A., & Kishino, H. (2012). Taxon influence index: Assessing taxon-induced incongruities in phylogenetic inference. *Systematic Biology, 61*(2), 337–345. https://doi.org/10.1093/sysbio/syr129

Marič, L., Cleenwerck, I., Accetto, T. ž., Vandamme, P., & Trček, J. (2020). Description of *Komagataeibacter melaceti* sp. nov. and *Komagataeibacter melomenusus* sp. nov. isolated from apple cider vinegar. *Microorganisms, 8*(8), 1178. https://doi.org/10.3390/microorganisms8081178

Marsh, A. J., O'Sullivan, O., Hill, C., Ross, R. P., & Cotter, P. D. (2014). Sequence-based analysis of the bacterial and fungal compositions of multiple kombucha (tea fungus) samples. *Food Microbiology, 38*, 171–178. https://doi.org/10.1016/j.fm.2013.09.003

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P., & Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: A case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *International Journal of Systematic and Evolutionary Microbiology, 58*(1), 200–214. https://doi.org/10.1099/ijs.0.65392-0

Matsutani, M., Ito, K., Azuma, Y., Ogino, H., Shirai, M., Yakushi, T., & Matsushita, K. (2015). Adaptive mutation related to cellulose producibility in *Komagataeibacter medellinensis* (*Gluconacetobacter xylinus*) NBRC 3288. *Applied Microbiology and Biotechnology, 99*(17), 7229–7240. https://doi.org/10.1007/s00253-015-6598-x

Morrison, D. A. (2006). Phylogenetic analyses of parasites in the new millennium. *Advances in Parasitology*, 1–124. https://doi.org/10.1016/s0065-308x(06)63001-7

Naloka, K., Yukphan, P., Matsushita, K., & Theeragool, G. (2018). Molecular taxonomy and characterization of thermotolerant *Komagataeibacter* species for bacterial nanocellulose production at high temperatures. *Chiang Mai Journal of Science, 45*(4), 1610-1622.

Papathanassopoulou, A. D., & Lorentzos, N. A. (2014). Parsimony-Informative Characters. *In 9th Conf. Hellenic Society for Computational Biology and Bioinformatics*. pp. 10-12.

Rashid, M. H.-, Young, J. P. W., Everall, I., Clercx, P., Willems, A., Santhosh Braun, M., & Wink, M. (2015). Average nucleotide identity of genome sequences supports the description of *Rhizobium lentis* sp. nov., *Rhizobium bangladeshense* sp. nov. and *Rhizobium binae* sp. nov. from lentil (*Lens culinaris*) nodules. *International Journal of Systematic and Evolutionary Microbiology, 65*(Pt_9), 3037–3045. https://doi.org/10.1099/ijs.0.000373

Rastogi, A., Gautam, S., Kumar, M., & Tomar, R. S. (2019). Ribosomal gene based comparative phylogenies for the genus *Mycobacterium*: An in-silicoapproach. *Journal of Scientific Research, 63*, 89-103.

Rehm, B. H. A. (2010). Bacterial polymers: Biosynthesis, modifications and applications. *Nature Reviews Microbiology, 8*(8), 578–592. https://doi.org/10.1038/nrmicro2354

Reiniati, I., Hrymak, A. N., & Margaritis, A. (2016). Recent developments in the production and applications of bacterial cellulose fibers and nanocrystals. *Critical Reviews in Biotechnology, 37(*4), 510–524. https://doi.org/10.1080/07388551.2016.1189871

Ribeiro, P. L., Rapini, A., Silva, U. C. S., & Berg, C. (2012). Using multiple analytical methods to improve phylogenetic hypotheses in *Minaria* (Apocynaceae). *Molecular Phylogenetics and Evolution, 65*(3), 915–925. https://doi.org/10.1016/j.ympev.2012.08.019

Ryngajłło, M. ł., Jacek, P., Cielecka, I., Kalinowska, H., & Bielecki, S. ł. (2019). Effect of ethanol supplementation on the transcriptional landscape of bionanocellulose producer *Komagataeibacter xylinus* E25. *Applied Microbiology and Biotechnology, 103*(16), 6673–6688. https://doi.org/10.1007/s00253-019-09904-x

Sato, M., & Miyazaki, K. (2017). Phylogenetic network analysis revealed the occurrence of horizontal gene transfer of 16S rRNA in the genus *Enterobacter*. *Frontiers in Microbiology, 8*, 1–10. https://doi.org/10.3389/fmicb.2017.02225

Semjonovs, P., Ruklisha, M., Paegle, L., Saka, M., Treimane, R., Skute, M., Rozenberga, L., Vikele, L., Sabovics, M., & Cleenwerck, I. (2017). Cellulose synthesis by *Komagataeibacter rhaeticus* strain P 1463 isolated from Kombucha. *Applied Microbiology and Biotechnology, 101(*3), 1003–1012. https://doi.org/10.1007/s00253-016-7761-8

Shavit, L., Penny, D., Hendy, M. D., & Holland, B. R. (2007). The problem of rooting rapid radiations. *Molecular Biology and Evolution, 24*(11), 2400–2411. https://doi.org/10.1093/molbev/msm178

Simpson, M. G. (2010). *Plant Systematics* (2nd ed.). San Diego, California, USA: Academic Press: pp. 35 – 36.

Soltis, P., & Doyle, J. J. (1998). *Molecular Systematics of Plants II: DNA Sequencing (v. 2)* (1998th ed.). Springer Science & Business Media New York: pp. 140 – 141.

Swofford, D. L., & Sullivan, J. (2003). Phylogeny inference based on parsimony and other methods using PAUP. In *The Phylogenetic Handbook, A Practical Approach to DNA and Protein Phylogeny*. Cambridge, England: Cambridge Univ. Press: pp. 182 –206.

Syed Ibrahim, K., Gurusubramanian, G., Zothansanga, Yadav, R. P., Senthil Kumar, N., Pandian, S. K., Borah, P., & Mohan, S. (2017). Nucleotide analysis. *Bioinformatics - A Student's Companion*, 1–116. https://doi.org/10.1007/978-981-10-1857-2_1

Tian, R.-M., Cai, L., Zhang, W.-P., Cao, H.-L., & Qian, P.-Y. (2015). Rare events of intragenus and intraspecies horizontal transfer of the 16S rRNA gene. *Genome Biology and Evolution, 7*(8), 2310–2320. https://doi.org/10.1093/gbe/evv143

Tortoli, E., Meehan, C. J., Grottola, A., Fregni Serpini, G., Fabio, A., Trovato, A., Pecorari, M., & Cirillo, D. M. (2019). Genome-based taxonomic revision detects a number of synonymous taxa in the genus *Mycobacterium*. *Infection, Genetics and Evolution, 75*, 1 – 17. https://doi.org/10.1016/j.meegid.2019.103983

Wilgenbusch, J. C., & Swofford, D. (2003). Inferring evolutionary trees with PAUP. *Current Protocols in Bioinformatics*, (1), 6-4. https://doi.org/10.1002/0471250953.bi0604s00

Yang, B., Wang, Y., & Qian, P.-Y. (2016). Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinformatics, 17*(1), 1–9. https://doi.org/10.1186/s12859-016-0992-y