




RESEARCH PAPER



The evolution and competitive strategies of *Akkermansia muciniphila* in gut

Ji-Sun Kim ^a, Se Won Kang^a, Ju Huck Lee^a, Seung-Hwan Park ^a, and Jung-Sook Lee ^{a,b}

^aKorean Collection for Type Cultures, Korea Research Institute of Bioscience and Biotechnology, Jeongseup-si, Republic of Korea; ^bDepartment of Environmental Biotechnology, University of Science and Technology, Yuseong-gu, Republic of Korea

ABSTRACT

Akkermansia muciniphila is a commensal bacterium using mucin as its sole carbon and nitrogen source. *A. muciniphila* is a promising candidate for next-generation probiotics to prevent inflammatory and metabolic disorders, including diabetes and obesity, and to increase the response to cancer immunotherapy. In this study, a comparative pan-genome analysis was conducted to investigate the genomic diversity and evolutionary relationships between complete genomes of 27 *A. muciniphila* strains, including KGMB strains isolated from healthy Koreans. The analysis showed that *A. muciniphila* strains formed two clades of group A and B in a phylogenetic tree constructed using 1,219 orthologous single-copy core genes. Interestingly, group A comprised of strains from human feces in Korea, whereas most of group B comprised strains from human feces in Europe and China, and from mouse feces. As group A and B branched, mucin hydrolysis played an important role in the stability of the core genome and drove evolution in the direction of defense against invading pathogens, survival in, and colonization in the mucus layer. In addition, WapA and anSME, which function in competition and post-translational modification of sulfatase, respectively, have been a particularly important selective pressure in the evolution of group A. KGMB strains in group A with anSME gene showed sulfatase activity, but KCTC 15667^T in group B without anSME did not. Our findings revealed that KGMB strains evolved to gain an edge in the competition with other gut bacteria by increasing the utilization of sulfated mucin, which will allow it to become highly colonized in the gut environment.

ARTICLE HISTORY

Received 12 July 2021
Revised 20 December 2021
Accepted 28 December 2021

KEYWORDS

Akkermansia muciniphila; evolution; environmental adaptation; comparative genomic analysis

Introduction


The human gastrointestinal (GI) tract is colonized by billions of commensal microbes, which constitute a complex and diverse community known as the gut microbiota.^{1,2} Recently, it has become evident that the intestinal microbiota plays an essential role in human well-being.³ The composition of microbial communities colonizing the GI tract differs according to the prevailing environmental conditions in the gut. Factors such as nutrition, transit time, host secretions, and pH shape the gut microbiota.⁴

Through competition, certain bacteria have evolved mechanisms to metabolize complex glycans in the mucus layer. These mucosa-associated microbiota form a distinct population in the gut and are affected by the proximity of the epithelial layer and the nutrients present in the mucus layer.^{5–8} The mucus layer in the colon can be divided into an inner mucus layer tightly attached to the epithelium

and a loosely adherent outer mucus layer. The inner layer is devoid of bacteria in healthy individuals, while the outer mucus layer is colonized with an abundance of commensal bacteria, particularly, mucin-degrading (mucinolytic) bacteria such as *Akkermansia muciniphila*, *Bifidobacterium bifidum*, *Bacteroides fragilis*, *Bacteroides thetaiotaomicron* and *Ruminococcus gnavus*.^{7,9,10} One of the key players in this community is *A. muciniphila*, which has a great impact on host physiology and microbiome composition.^{11–13}

A. muciniphila is a Gram-negative bacteria, and most strains have been isolated from human fecal samples.¹⁴ *A. muciniphila* is easily detected in meta-omic studies, since it is the only intestinal isolate of the deeply rooted *Verrucomicrobia* phylum.¹⁵ It accounts for 1–3% of the total fecal microbiota since early life,^{16,17} and is present abundantly in the colonic mucosal layer.^{18,19} *A. muciniphila* type strain ATCC BAA-835^T encodes various mucin-

CONTACT Jung-Sook Lee  jslee@kribb.re.kr  Korean Collection for Type Cultures, Korea Research Institute of Bioscience and Biotechnology, Jeongseup-si 56212 Republic of Korea

 Supplemental data for this article can be accessed [here](#).

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

degrading enzymes in its relatively small 2.6 Mb genome.²⁰ *A. muciniphila* produces metabolites such as short-chain fatty acids (SCFAs), which may play an important role in the metabolic health or inflammatory status of the host.²¹ The relative abundance of *A. muciniphila* in the gut is highly responsive to changes in the gut environment and health, such as age, degree of obesity, and polypharmacy.^{22–25} It was reported that the availability of mucin in the residual environment affects the *A. muciniphila* transcriptome and proteome.^{23,26} Numerous genes and proteins encoding glycosyl hydrolases and fucosidase, required for mucin degradation, are upregulated during *A. muciniphila* growth on mucin, compared to growth on glucose.²⁶ Environmental variation was also observed in the outer membrane proteome of *A. muciniphila*, where the abundance of one-third of the outer membrane proteins was different between bacterial cells grown on mucin and on glucose.²⁷ Further studies demonstrated that Amuc_1100, a specific outer membrane protein, activated Toll-like receptor 2 (TLR2) pathways and protected the integrity of the intestinal epithelium.¹² These findings highlight the importance of detailed characterization of gut microbiota to understand the mechanisms underlying the reported benefits to the host. In the gut, the abundance of *A. muciniphila* was negatively correlated with numerous diseases, including inflammatory bowel diseases (IBD),^{28,29} cancer,³⁰ diabetes,¹² and obesity.^{31,32} Further mechanistic studies have revealed the anti-inflammatory role of *A. muciniphila* in the gut environment.³³ Currently, the most convincing evidence of its beneficial effect on health comes from studies linking *A. muciniphila* to metabolic disorders, such as diabetes and obesity. However, the exact signaling mechanisms by which *A. muciniphila* interacts with the host and its effect on the overall microbial community in the gut require further investigation. Furthermore, how *A. muciniphila* strains have evolved in a variety of organisms, and the factors that drive their evolution in response to the dynamic mucosal environment, remain unknown.

In this study, we sequenced four additional *A. muciniphila* strains isolated from the feces of healthy Koreans and a type strain KCTC 15667^T obtained from the KCTC culture collection using single-molecule real-time (SMRT) sequencing

technology.^{34,35} Using the dataset of complete chromosomal sequences of the 27 *A. muciniphila* strains, we estimated the sizes of pan- and core-genomes and functional features, and evaluated population diversity by phylogenetic analysis. In addition, we attempted to assess selection pressure and selection functions in the diversification across the single-copy core genomes. Furthermore, the genetic organization and evolution of glycoside hydrolase genes were investigated to study their evolution at the genomic level.

Results

The isolation and whole-genome sequencing of Akkermansia muciniphila strains

Currently, we are carrying out the project of Korean Gut Microbiome Banking (KGMB), which is the research on the isolation and acquisition of gut microbiota using culturomics from healthy Korean feces, the analysis of gut microbial population through metagenomic analysis, and the regulation of gut homeostasis by isolated gut microbiota. For this study, four *Akkermansia muciniphila* strains (hence referred to as KGMB strains) were isolated from the fecal samples of healthy Koreans (Table 1). Whole genomes of KGMB isolates and *A. muciniphila* type strain KCTC 15667^T, which was obtained from the KCTC culture collection, were sequenced, and their complete genome sequences were obtained. There was a slight difference in genome size and CDS numbers between the two *A. muciniphila* type strains, KCTC 15667^T sequenced in this study and ATCC BAA-835^T obtained from GenBank. Therefore, in this study, the whole-genome sequences of both type strains were analyzed. Average nucleotide identity (ANI) was calculated to compare the genome distances between the KGMB strains and the type strains. It was found that ANI values between KGMB strains were 99.99–100%, indicating that KGMB strains were remarkably close to each other (Table 2). In addition, ANI values between type strain KCTC 15667^T and KGMB strains were 97.53–97.54%, indicating that KGMB strains are further away from the *A. muciniphila* type strain despite being of the same species. Notably, the proposed and generally accepted species boundary for ANI values is 95–96%.³⁶

Table 1. Strain information sequenced in this study.

Strains	No. of contigs	Genome size (bp)	GC ratio	Topology	Country	Host
KGMB01988	1	2,844,056	55.23	Circular	Republic of Korea	<i>Homo sapiens</i>
KGMB01989	1	2,844,036	55.23	Circular	Republic of Korea	<i>Homo sapiens</i>
KGMB01990	1	2,844,062	55.23	Circular	Republic of Korea	<i>Homo sapiens</i>
KGMB02009	1	2,844,059	55.23	Circular	Republic of Korea	<i>Homo sapiens</i>
KCTC 15667 ^T	1	2,664,051	55.76	Circular	Netherlands	<i>Homo sapiens</i>

Sequence comparison between KGMB strains

Multiple genome alignments were performed to identify the structural differences in the genome. Genome synteny also showed no significant differences between the KGMB strains. However, it was found that there are length variations in the homopolymeric polyguanine (poly G) region in the promoter of fumarate hydratase between type strain KCTC 15667^T and KGMB strains (Figure S1). KGMB strains had a greater number of homopolymeric guanosine repeats, 22–29 mer Gs, compared to the type strain with 18-mer Gs. Fumarate hydratase, also known as fumarase, converts fumaric acid to L-malic acid in the tricarboxylic acid (TCA) cycle, and is a conserved protein in all organisms, from bacteria to humans, with respect to its sequence, structure, and enzymatic activity.^{37,38} Although the intergenic region (297 bp) of fumarase was identical between the type strain KCTC 15667^T and KGMB strains, differences in the number of poly G repeats in the promoter may cause physiological differences between them.

General features of *Akkermansia muciniphila* genomes

Since the ANI value between the KGMB strains was 99.99–100% and their genome sequences were almost identical, the complete genome of KGMB01988 was selected among the KGMB strains for comparative analysis between the *A. muciniphila* genomes. Twenty-five complete sequences of *A. muciniphila* were obtained from GenBank (www.ncbi.nlm.nih.gov/genome/browser). Comparative analysis was performed using only *A. muciniphila* strains with complete genomic sequence to obtain accurate results. Phylogenetic analysis was performed using *A. glycaniphila* APyt^T, which is the closest to *A. muciniphila*, as an outgroup. Most of

A. muciniphila strains used in this analysis were isolated from the feces of *Homo sapiens*, but *A. muciniphila* YL44, YL44_sDMDMm2, and ‘139’ were isolated from mice (Table 3). The genome sizes ranged between 2.66 Mb and 2.84 Mb, and the GC content ranged from 55.23–55.76%. Interestingly, KGMB01988 had the largest genome size after strain CBA5201, among the *A. muciniphila* strains used in this study. None of the 27 strains contained additional amplicons besides a chromosome and had an average of 2,262 CDSs, as predicted using the prodigal program (Table 3). To analyze the genomic distance between strain KGMB01988 and *A. muciniphila* reference strains, dDDH (digital DNA-DNA hybridization) and ANI values were calculated (Table 4). As a result, the ANI and dDDH values for pair-wise comparisons were in the range of 97.44–99.90% and 77.70–98.90%, respectively. These results revealed that genetic divergence exists between *A. muciniphila*.

Comparative analysis of *Akkermansia muciniphila* genomes

A comparative pan-genome analysis was conducted to study the genomic diversity and evolutionary relationships among 27 *A. muciniphila* strains. Orthologous gene clusters were analyzed using the Markov cluster (MCL) algorithm.³⁹ The pan-genome contained 3,811 orthologous groups (OGs). Members of gene families in these strains were divided into three categories (core, accessory, and unique) based on their appearance in different genomes. Among the 3,811 OGs, 1,749 OGs were conserved in all 27 strains and represented the universal core gene sets (core genome). The remaining 1,255 OGs corresponded to the accessory genome and were present in more than one, but not all 27 genomes. Finally, 807 OGs existed in only one genome, comprising unique genes.

Table 2. ANI values between strains sequenced in this study.

Strains	KGMB01989	KGMB01990	KGMB02009	KGMB01988	KCTC 15667 ^T
KGMB01989	-	99.99	99.99	99.99	97.53
KGMB01990	99.99	-	100	100	97.53
KGMB02009	99.99	100	-	100	97.53
KGMB01988	99.99	100	-	-	97.53
KCTC 15667 ^T	97.54	97.54	97.54	97.53	-

Table 3. Complete genome sequences of *A. muciniphila* analyzed in this study.

	Species	Strains	Tag	Genome assembly	Country	Host	Genome size	G + C ratio	No. of CDS
1	<i>A. muciniphila</i>	EB-AMDK-3	MYE3	GCA_003716935.1	Republic of Korea	<i>Homo sapiens</i>	2,663,833	55.76	2,217
2	<i>A. muciniphila</i>	EB-AMDK-4	MYE4	GCA_003716955.1	Republic of Korea	<i>Homo sapiens</i>	2,664,010	55.76	2,166
3	<i>A. muciniphila</i>	KCTC 15667 ^T	KCTC 15667	in this study GCA_017504145.1	Netherlands	<i>Homo sapiens</i>	2,664,051	55.76	2,149
4	<i>A. muciniphila</i>	ATCC BAA-835 ^T	ATCC BAA-835	GCA_000020225.1	Netherlands	<i>Homo sapiens</i>	2,664,102	55.76	2,150
5	<i>A. muciniphila</i>	YL44_sDMDMm2	MYL2	GCA_002201495.1	Switzerland	<i>Mus musculus</i>	2,737,357	55.66	2,242
6	<i>A. muciniphila</i>	YL44	MYL4	GCA_001688765.2	Switzerland	<i>Mus musculus</i>	2,745,278	55.66	2,254
7	<i>A. muciniphila</i>	EB-AMDK-7	MYE7	GCA_004015245.1	Republic of Korea	<i>Homo sapiens</i>	2,799,431	55.30	2,305
8	<i>A. muciniphila</i>	139	M139	GCA_004319565.1	China	<i>Mus musculus</i>	2,801,917	55.74	2,315
9	<i>A. muciniphila</i>	EB-AMDK-16	MY16	GCA_004015205.1	Republic of Korea	<i>Homo sapiens</i>	2,770,073	55.30	2,270
10	<i>A. muciniphila</i>	EB-AMDK-15	MY15	GCA_004015305.1	Republic of Korea	<i>Homo sapiens</i>	2,770,098	55.30	2,274
11	<i>A. muciniphila</i>	EB-AMDK-18	MY18	GCA_004015085.1	Republic of Korea	<i>Homo sapiens</i>	2,770,124	55.30	2,287
12	<i>A. muciniphila</i>	EB-AMDK-17	MY17	GCA_004015225.1	Republic of Korea	<i>Homo sapiens</i>	2,770,146	55.30	2,269
13	<i>A. muciniphila</i>	EB-AMDK-1	MYE1	GCA_003716915.1	Republic of Korea	<i>Homo sapiens</i>	2,772,237	55.39	2,261
14	<i>A. muciniphila</i>	H2	MHA2	GCA_004101765.1	Belgium	<i>Homo sapiens</i>	2,819,944	55.32	2,293
15	<i>A. muciniphila</i>	CBA5201	MC01	GCA_004104435.1	Republic of Korea	<i>Homo sapiens</i>	2,860,407	55.32	2,348
16	<i>A. muciniphila</i>	EB-AMDK-21	MY21	GCA_004015345.1	Republic of Korea	<i>Homo sapiens</i>	2,724,154	55.32	2,243
17	<i>A. muciniphila</i>	EB-AMDK-22	MY22	GCA_004015125.1	Republic of Korea	<i>Homo sapiens</i>	2,724,161	55.32	2,252
18	<i>A. muciniphila</i>	EB-AMDK-20	MY20	GCA_004015325.1	Republic of Korea	<i>Homo sapiens</i>	2,724,186	55.32	2,220
19	<i>A. muciniphila</i>	EB-AMDK-19	MY19	GCA_004015105.1	Republic of Korea	<i>Homo sapiens</i>	2,724,248	55.32	2,211
20	<i>A. muciniphila</i>	EB-AMDK-10	MYE0	GCA_004015005.1	Republic of Korea	<i>Homo sapiens</i>	2,763,834	55.25	2,363
21	<i>A. muciniphila</i>	EB-AMDK-13	MY13	GCA_004015285.1	Republic of Korea	<i>Homo sapiens</i>	2,763,965	55.25	2,329
22	<i>A. muciniphila</i>	EB-AMDK-14	MY14	GCA_004015065.1	Republic of Korea	<i>Homo sapiens</i>	2,764,188	55.25	2,267
23	<i>A. muciniphila</i>	EB-AMDK-2	MYE2	GCA_003716975.1	Republic of Korea	<i>Homo sapiens</i>	2,764,211	55.25	2,251
24	<i>A. muciniphila</i>	EB-AMDK-12	MY12	GCA_004015045.1	Republic of Korea	<i>Homo sapiens</i>	2,764,297	55.26	2,238
25	<i>A. muciniphila</i>	EB-AMDK-11	MY11	GCA_004015025.1	Republic of Korea	<i>Homo sapiens</i>	2,764,311	55.26	2,243
26	<i>A. muciniphila</i>	EB-AMDK-8	MYE8	GCA_004015265.1	Republic of Korea	<i>Homo sapiens</i>	2,824,041	55.39	2,330
27	<i>A. muciniphila</i>	KGMB01988	KGMB01988	in this study GCA_017570525.1	Republic of Korea	<i>Homo sapiens</i>	2,844,056	55.23	2,315
	Outgroup <i>A. glycaniphila</i>	APyt ^T	GAPY	GCA_900097105.1	Netherlands	<i>Malayopython reticulatus</i>	3,074,078	57.65	2,497

Table 4. Genome to genome distance with strain KGMB01988.

Query genome	Strain	ANI value	dDDH value	Distance	Probability that dDDH > 70%	G + C difference
KGMB01988	EB-AMDK-3	97.64	78.90	0.0246	89.71	0.53
	EB-AMDK-4	97.64	78.90	0.0246	89.73	0.53
	ATCC BAA-835 ^T	97.66	78.90	0.0246	89.73	0.53
	KCTC 15667 ^T	97.63	78.90	0.0246	89.73	0.53
	YL44_sDMDMm2	97.55	78.40	0.0252	89.30	0.43
	YL44	97.63	78.40	0.0253	89.24	0.43
	EB-AMDK-7	97.46	77.70	0.0262	88.57	0.07
	139	97.44	78.20	0.0255	89.10	0.51
	EB-AMDK-16	98.06	82.30	0.0207	92.19	0.07
	EB-AMDK-15	98.04	82.30	0.0207	92.21	0.07
	EB-AMDK-18	98.03	82.30	0.0207	92.21	0.07
	EB-AMDK-17	98.06	82.30	0.0206	92.22	0.07
	EB-AMDK-1	98.26	84.00	0.0187	93.21	0.16
	H2	98.08	82.10	0.0209	92.07	0.09
	CBA5201	98.12	83.00	0.0199	92.63	0.09
	EB-AMDK-21	99.89	98.90	0.0019	98.04	0.09
	EB-AMDK-22	99.89	98.80	0.0020	98.03	0.09
	EB-AMDK-20	99.88	98.90	0.0019	98.04	0.09
	EB-AMDK-19	99.90	98.90	0.0019	98.04	0.09
	EB-AMDK-10	98.97	90.80	0.0112	96.07	0.02
	EB-AMDK-13	99.00	90.80	0.0112	96.07	0.02
	EB-AMDK-14	99.03	90.90	0.0111	96.10	0.02
	EB-AMDK-2	99.00	91.00	0.0111	96.11	0.02
	EB-AMDK-12	99.03	91.00	0.0111	96.11	0.03
	EB-AMDK-11	99.03	90.90	0.0111	96.10	0.03
	EB-AMDK-8	99.10	92.30	0.0096	96.52	0.16

The number of non-redundant strain-specific genes across different genomes varied from 2 to 108. *A. muciniphila* CBA5201 strain harboring the largest genome size had the largest number of unique gene families, 108, while *A. muciniphila* type strain KCTC 15667^T possessed the smallest number of unique gene families. The large proportion of specific genes suggested that the *A. muciniphila* strains harbored a high level of genomic diversity, indicating their ability to survive in various gut environments. Cumulative curves were generated using the PanGP program,⁴⁰ a tool for quickly analyzing bacterial pan-genome profiles. The openness of the pan-genome was calculated based on Heap's law model.^{41,42} When $\gamma > 0$, it means that pan-genomes are open state. The size of the pan-genome increased unboundedly with the increase of new genomes, including 3,811 non-redundant genes, which indicated that the *A. muciniphila* pan-genome was still "open" (Figure 1). This open pan-genome showed great potential for discovering novel genes with every *A. muciniphila* strain sequenced. In contrast to the pan-genome, the size of the core genome

appeared to reach a steady-state approximation with 1,749 non-redundant genes. In addition, 45.9% of the pan-genome was found to be conserved, while the remaining 54.1% varied across the strains, indicating that the pan-genome exhibited a high level of genome variability.

Phylogenetic analysis of *Akkermansia muciniphila*

To gain insights into the similarity and distance between *A. muciniphila* genomes, 4,598 OGs were identified from 28 strains of the genus *Akkermansia*, including the outgroup *A. glycaniphila* APyt^T. Of these, 1,219 orthologous single-copy core genes were used to construct the phylogenetic tree. The concatenated amino acid sequences encoded by 1,219 single-copy core genes were aligned using the MUSCLE algorithm. To choose the most suitable model for creating a phylogenetic tree, ProtTest v3.2 was used. As a result, a pan-genome phylogenetic tree was constructed using the LG-I-G-F model and RAxML program with bootstrap 100 replications (Figure 2). In the phylogenetic tree, all

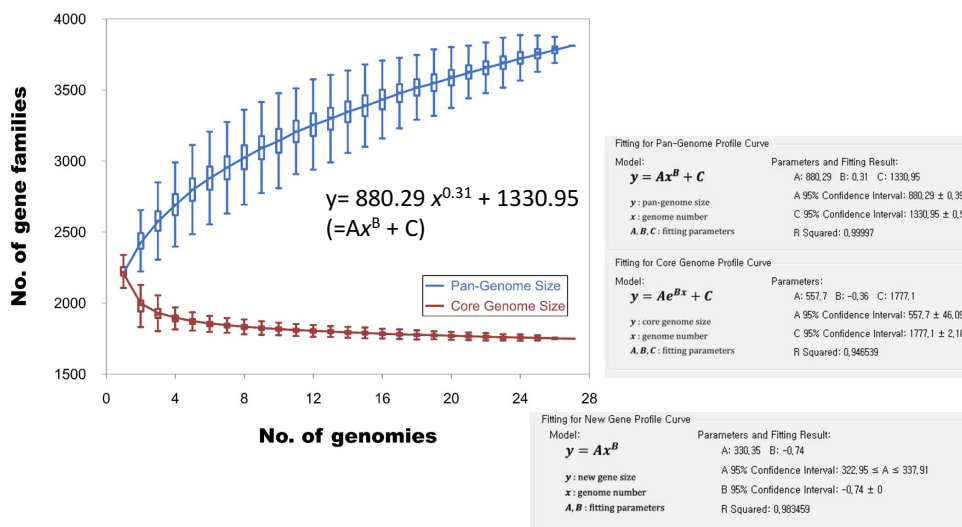


Figure 1. Curves for size of pan-genome and core-gene sets from completely sequenced *Akkermansia muciniphila* strain genomes. The strains shared 1,749 gene families. Pan-genome of 27 *A. muciniphila* strains consists of 3,811 gene families. Estimation of openness based on Heaps' law model showed that the *A. muciniphila* pan genome is open with a parameter (γ) of 0.78.

bootstrap values in the internal node were 100, indicating that this tree was very rigid. As shown in Figure 2, *A. muciniphila* strains were grouped into two distinct clades. The first clade (group A) was composed of strains isolated in Korea, including KGMB strains, whereas the second clade (group B) consisted of strains isolated from Europe, China, and mice, although three strains isolated from Korea (EB-AMDK-3, -4 and -7) were included. The type strain ATCC BAA-835^T (= KCTC 15667^T) isolated from a healthy adult in Netherlands was included in group B. The comparison of the genome distance and the phylogeny revealed that strain KGMB01988 showed dDDH values greater than 80% with group A members, while showing values below 70–80% with strains belonging to group B. Furthermore, there were some differences between groups A and B. The G + C ratio of group A members was $55.3 \pm 0.05\%$ on average, whereas that of group B members was $55.7 \pm 0.16\%$. And the genome size of group B members was 2,663,833–2,801,917 bp and that of group A members was 2,724,154–2,860,407 bp, indicating that the genomes of group A were larger than those of group B.

To date, there are insufficient general guidelines for defining subspecies using genomic data. However, Chen *et al.*³⁶ suggested that the ANI

value between subspecies and other species should be lower than the species-level cutoff value, the ANI value between subspecies should be higher than the species-level cutoff, and strains belonging to different subspecies should be genomically coherent and form distinguishable clades in the phylogenetic tree. Based on the above genome relatedness and phylogenetic tree, we suggest that *A. muciniphila* strains should be divided into two subspecies of group A, including KGMB strains and group B, including type strain ATCC BAA-835^T (= KCTC 15667^T).

Functional profiling of pan-genome

To gain insights into functional diversity in *A. muciniphila* strains, the pan-genome of the 27 *A. muciniphila* genomes was analyzed using COG functional category: 'cellular processes and signaling', 'metabolism', 'information storage and processing' and 'poorly characterized'. Genes with predicted functions and unknown functions were more abundant in the pan-genome (Figure 3 and Figure S2). Most of the core genome was involved in 'cell motility' (N), 'carbohydrate transport and metabolism' (G), 'nucleotide transport and metabolism' (F), and 'translation, ribosomal structure and biogenesis' (J). Most of accessory and unique genomes were concerned with 'defense mechanisms' (V) and

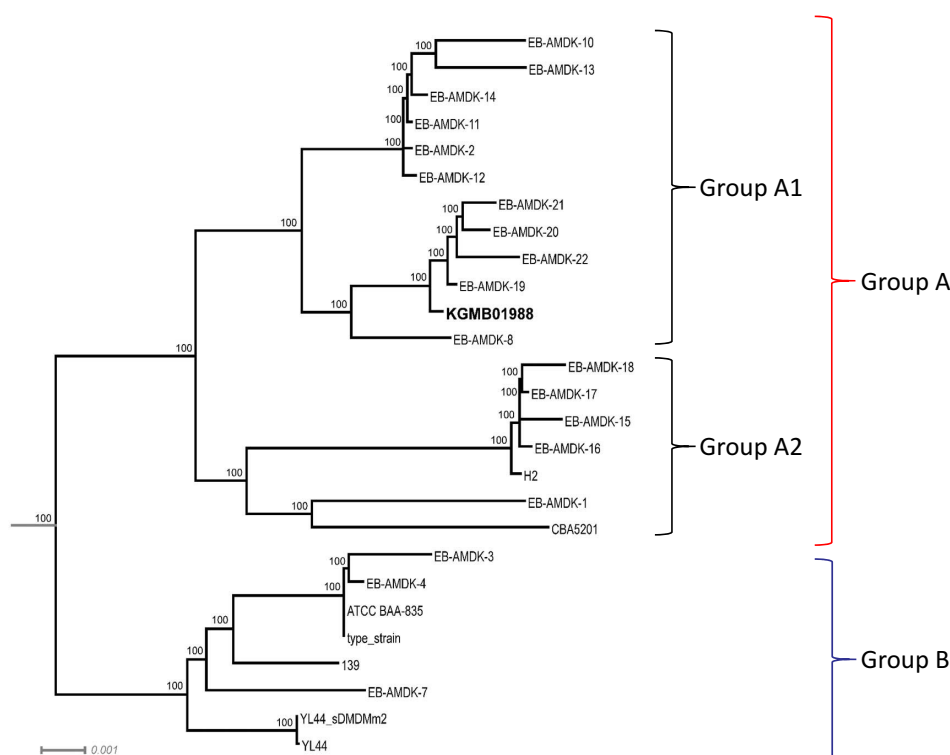


Figure 2. Phylogenetic tree of *Akkermansia muciniphila* group based on the concatenation of the amino acid alignments deduced from 1,219 core genes with maximum likelihood approach. Numbers above branches show maximum-likelihood bootstrap supports from 100 non-parametric replicates. The tree was rooted by *A. glycaniphila* APyT^T as an outgroup. The scale represents the number of substitutions per site.

‘replication, recombination and repair’ (L), indicating that defense, replication and repair systems were acquired to adapt to environment causing evolution. The mucus layer in the gut consists of an outer gel-forming layer that provides a habitat for bacteria, and an inner layer, which is devoid of bacteria. Its major components, mucins, are a source of nutrients for intestinal bacteria because they are composed of amino acids and oligosaccharides. Some gut bacteria possess the enzymatic machinery necessary for the breakdown of the mucin oligosaccharide chains, which in turn release fucose, galactose, N-acetylglucosamine, N-acetylgalactosamine, sialic acid, and sulfate. And then, breakdowned products, disaccharides and small oligosaccharides can be further metabolized by the resident microbiota. In the genus *Akkermansia*, carbohydrate metabolism is one of the most important metabolic activities because it hydrolyzes mucins and is utilized as the sole carbon source. Therefore, we analyzed the genes involved in the mucin-degrading pathway and carbon metabolism using the CAZy database

(Table 5). Fifty-four glycoside hydrolases (GH) genes were identified, which belonged to 26 GH families, in pan-genome. These GH genes were more abundant in the core genome than in the accessory and unique genomes. Of the 54 GH genes, 47 belonged to 24 GH families and were present in the core genome of 27 *A. muciniphila* strains. In addition, genes of family GH20 were the most abundant with numbers of 10–12, followed by genes of family GH2 with numbers of 5–7, which are expected to function as galactosidase.

The family GH84 gene with N-acetyl β -glucosaminidase activity was found in all strains except for the strain EB-AMDK-3. In addition, the family GH84 genes had two copies in all group A members, except for the EB-AMDK-21 strain. In group B, strains EB-AMDK-7, YL44 and ‘139’, the deepest branch in the evolutionary tree, contained two copies of the family GH84 gene, while group B members, considered to have recently diverged lineages, had a GH84 gene of 0 or 1. In contrast, the GH18 gene was not found in group A, but only in group B. Some sialidases (GH33, neuraminidase)

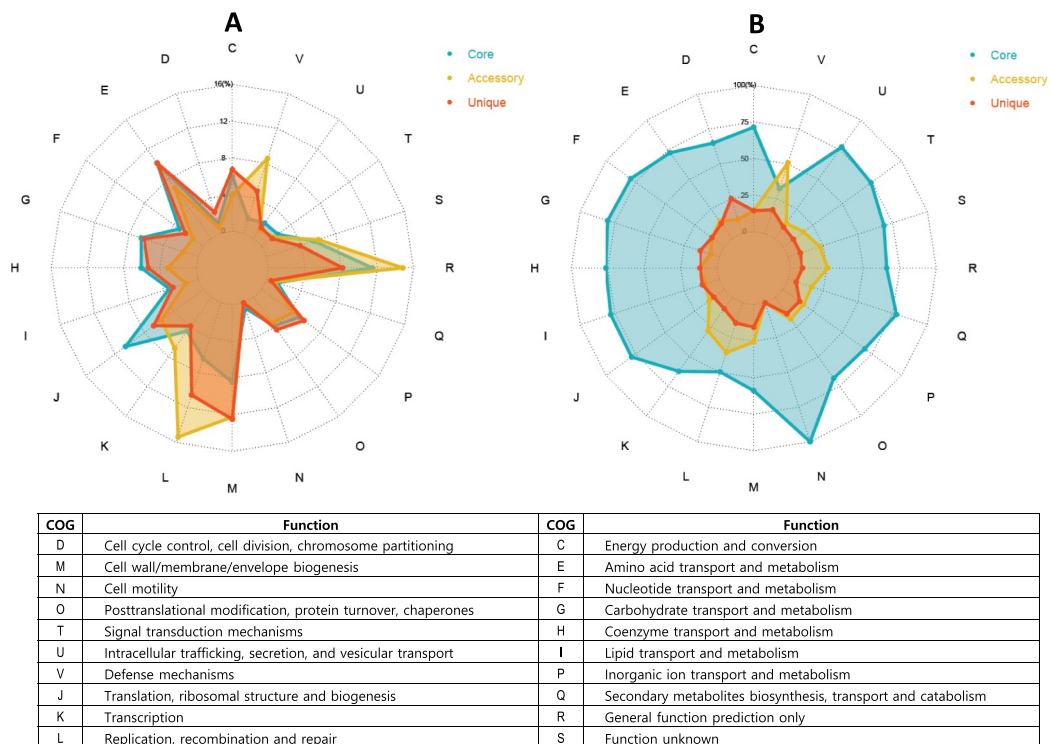


Figure 3. Comparison of functional categories of *Akkermansia muciniphila*. The ordinate axis represents the percentages of genes in each functional category. The pan-genome from the 27 *A. muciniphila* genomes was analyzed using COG functional category: ‘cellular processes and signaling’, ‘metabolism’, ‘information storage and processing’ and ‘poorly characterized’. A. The radar plots of totalContents is the ratio of Core/Accessory/Unique genes in each COG category. B. The radar plots of eachCategory is the ratio of orthologs corresponding to each COG category in Core/Accessory/Unique genes. The radar plots were drawn using the ‘fmsb’ R package.

and α -L-fucosidase (GH29) have different numbers depending on the strain (Table 5). Essential genes related to the degradation of mucin and fucose were maintained in all *A. muciniphila* strains, indicating that it is pivotal to hydrolyzing mucin, to be used as a sole carbon source.

Gene gain and loss analysis – evolutionary events

Bacterial evolution happens by frequent gene gain and loss within gene families. Although identical genus or species share a common set of core genes, individuals within the genus or species may have different subsets of genes.⁴³ This subset of genes can be the key to the bacterial ability to survive in certain habitats, increasing the fitness of bacteria within habitats leading to evolution. To obtain deeper insight into the evolution of gene families on

phylogeny, gene expansion and contraction at each branch of the strains were investigated. The evolutionary pathway was analyzed based on the phylogenetic tree, and the gene counts for each gene family were generated from the MCL. Diversification of branches was associated with a large number of CDSs of gene families across the *A. muciniphila* tree. As shown in Figures 4, 2,021 CDSs were more likely to have been inherited from the most recent common ancestor (MRCA) of *A. muciniphila* strains, and the number of CDSs was increased by evolutionary events. The increase in the CDS number was associated with gene duplication and horizontal gene transfer but might include gene fragmentation and multiplication of repeat proteins such as transposase. At the point divided into the two groups, A and B, it seems that the ancestors of group A evolved to have more

Table 5. Glycoside hydrolase in *Akkermansia muciniphila* strains. Values that are significantly different in KGMB01988 compared to members of other groups are highlighted in bold.

Glycosylhydro lasefamily	ATCC BAA- 835 ^T	KCTC 15667 ^T	Group B													Group A													UniProtID										
			EB- AMDK- 3	EB- AMDK- 4	EB- AMDK- 7	EB- AMDK- YL44 Mm2)	YL44 5201	EB- AMDK- 1	H2	EB- AMDK- 15	EB- AMDK- 16	EB- AMDK- 17	EB- AMDK- 18	EB- AMDK- 8	EB- AMDK- 19	EB- AMDK- 20	EB- AMDK- 21	EB- AMDK- 22	EB- AMDK- 2	EB- AMDK- 10	EB- AMDK- 11	EB- AMDK- 12	EB- AMDK- 13	EB- AMDK- 14															
GH 18	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	B2UPU3			
GH 29	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A0A1H6L863			
GH 33	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A0A1C7PBY9			
GH 43_24	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	A0A3R5QZ58			
GH 84	1	1	2	0	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	B2ULA9			
GH 2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	B2UQC2			
GH 2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A3G3PEN8			
GH 2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8ILY3			
GH 2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8HGL1			
GH 2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UM41		
GH 2	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	A0A1H6L742			
GH 3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UPPO		
GH 13_5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8I8L1		
GH 13_8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UJML3		
GH 13_38	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UM13		
GH 16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UPN9		
GH 16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	R6ZJ9		
GH 16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8HLJ8		
GH 20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	B2UN02		
GH 20	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	B2UP57		
GH 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UN22		
GH 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UPR7		
GH 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A354E9T1	
GH 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UNM1		
GH 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8JUN6		
GH 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UNC4		
GH 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8JNZ9	
GH 27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2URC7	
GH 29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UQE4	
GH 29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8HG39	
GH 29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	R6I8W4	
GH 31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UN73	
GH 31	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UQU9	
GH 33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UN42	
GH 33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UPI5	
GH 33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2ULI1	
GH 35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	R7EZK5	
GH 35	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	A0A2N8IUG9	
GH 36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	B2UQF3

(Continued)

Table 6. Gene families gained at the ancestral branch of group A.

Before	After	No. of changes	Annotation
8	13	5	tRNA3(Ser)-specific nuclease WapA
0	1	1	Sulfite reductase (NADPH) flavoprotein alpha-component
0	1	1	Sulfate adenylyltransferase subunit 1
0	1	1	Sulfate adenylyltransferase subunit 2
0	1	1	Thioredoxin-dependent 5'-adenylylsulfate reductase
0	1	1	Modification methylase DpnIIb
0	1	1	Type I restriction enzyme EcoKI M protein
0	1	1	Type I restriction enzyme EcoR124II R protein
0	1	1	5-Methylcytosine-specific restriction enzyme B
0	1	1	Putative glycosyltransferase
0	1	1	Putative glycosyltransferase EpsJ
0	1	1	Bicarbonate transport system permease protein CmpB
0	1	1	Bicarbonate transport ATP-binding protein CmpD
0	1	1	Putative aliphatic sulfonates-binding protein
0	1	1	O-acetylserine sulfhydrylase
0	1	1	Outer membrane protein beta-barrel domain
0	1	1	Maltose O-acetyltransferase
0	1	1	Teichuronic acid biosynthesis protein TuaB
0	1	1	Putative teichuronic acid biosynthesis glycosyltransferase TuaC
0	1	1	Tyrosine recombinase XerD
0	1	1	Serine/threonine-protein kinase PrkC
0	1	1	Regulatory protein SoxS
0	1	1	Ornithine cyclodeaminase

genes. Interestingly, strains belonging to group A had 2,211–2,363 CDSs, whereas group B members had 2,149–2,315 CDSs, indicating that evolution to group B led to fewer gene gains than group A. In particular, *A. muciniphila* type strains, KCTC 15667^T and ATCC BAA-835^T, had approximately 2,150 CDSs, which is the lowest CDS number in the *A. muciniphila* strains analyzed. Figure 5 shows the minimum gene gain/loss events at each internal and external branch. The number of expanded CDSs in group A was predominant in the branch that divided groups A and B. Furthermore, expansions outnumbered contractions on all branches, indicating that the gene gain function plays an important role in dynamic evolution.

In most bacterial genomes, transposable elements are usually responsible for a high level of gene turnover (outlier events). However, *wapA*, a gene encoding wall-associated protein A, was recognized as an outlier in *A. muciniphila* strains. Interestingly, it occurred in the internal branch evolving from the MRCA to group A, which means that the addition of *wapA* gene may be an important driving force of the evolution of group A. Rearrangement hotspot (Rhs) and related YD-peptide repeat proteins are widely distributed in

bacteria and eukaryotes. It has been reported that Rhs (Rearrangement hotspot) proteins are found in gram-negative bacteria and *WapA* proteins are present in gram-positive bacteria.^{44,45} Although *WapA* proteins have not been assigned a definitive function, it has recently been reported that these proteins are involved in intercellular competition via contact-dependent growth inhibition.^{44,45}

TopGO analysis showed that sulfite reductase, sulfate adenylyltransferase, and adenylylsulfate reductase involved in hydrogen sulfide biosynthesis were acquired during evolution into group A (Table 6 and Figure 6). This indicates that sulfur metabolism may have been a particularly important selective pressure in the evolution into group A. Furthermore, the gain of genes involved in sulfur metabolism was found in only human-originated strains, while it was not observed in non-human-originated strains such as YL44, YL44 sDMDMm2, and '139'. These results suggest that genes related to sulfur metabolism may be required for *A. muciniphila* to inhabit and adapt to the human gut.

Gain/loss events of genes related to nucleic acid modification, such as DNA methyltransferases and the restriction enzymes occur

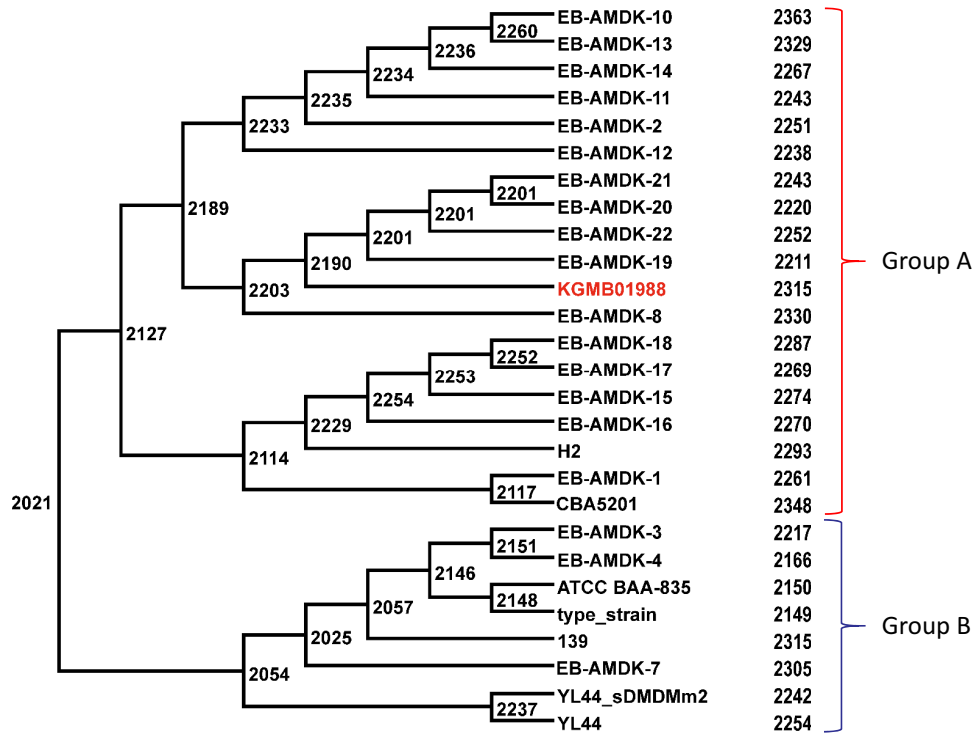


Figure 4. Analysis of ancestral genes in evolutionary path. Numbers adjacent to internal nodes indicate the number of estimated ancestral genes (protein coding genes). Right panel indicates the number of CDSs of *Akkermansia muciniphila* strains.

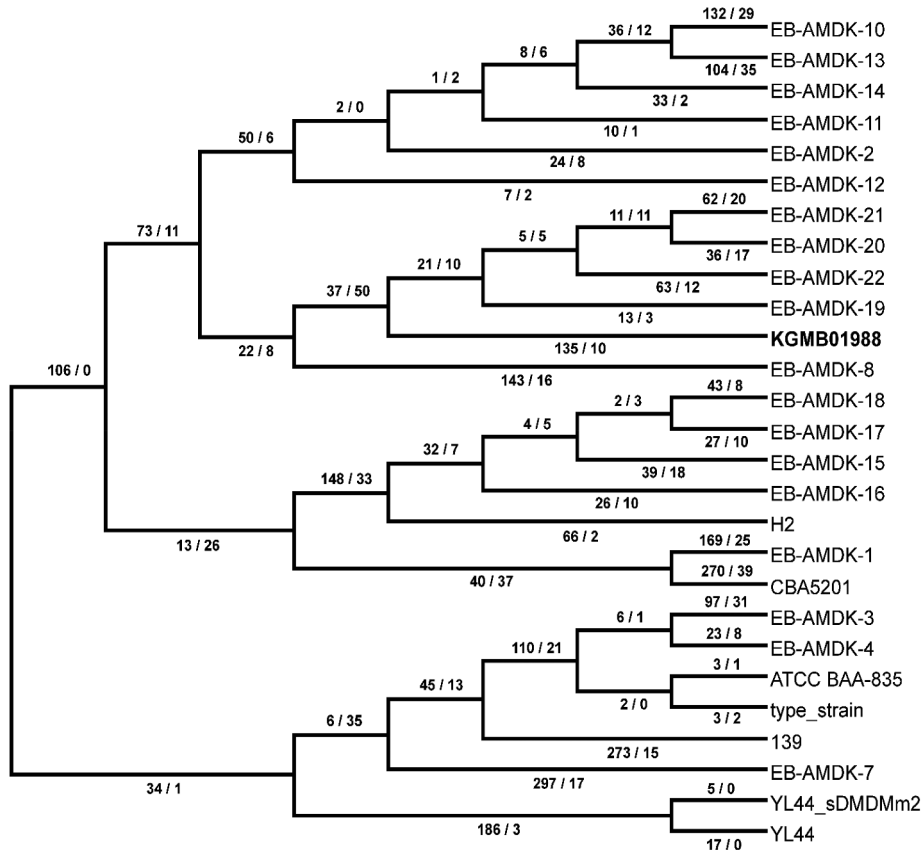


Figure 5. Minimal gene gain and loss events under the best fit model (GD-FR-ML). Numbers on the branches denote the minimum number of gains and losses in that order.

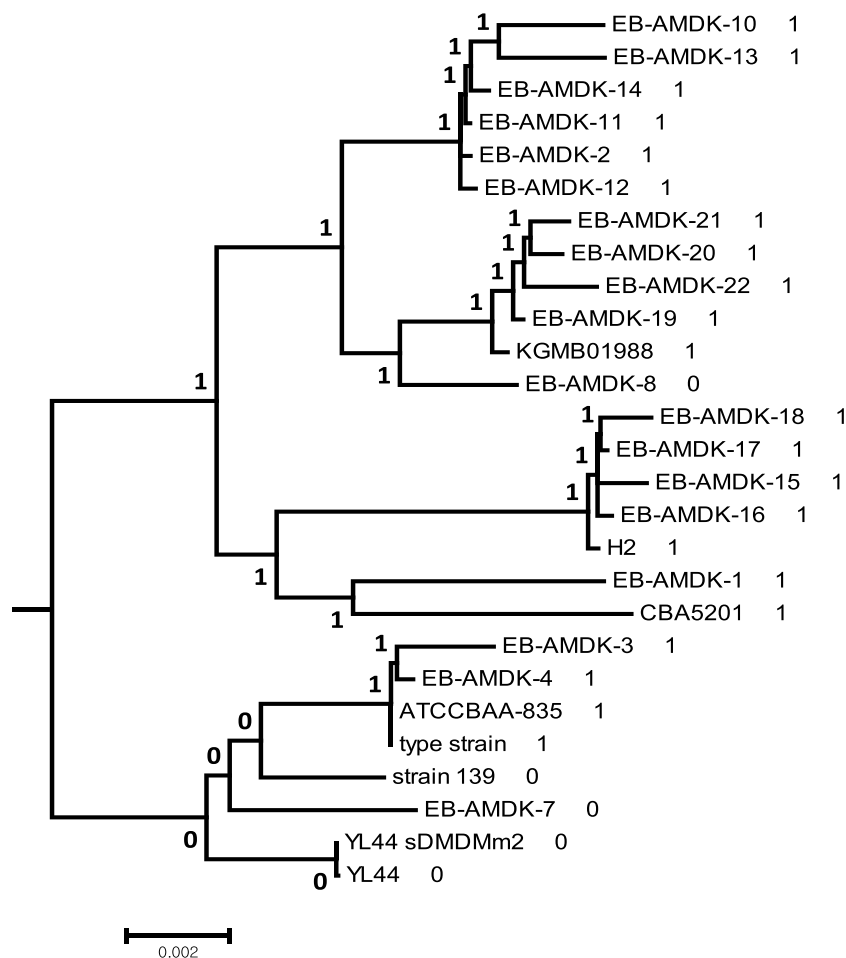


Figure 6. Gene counts of sulfate adenylyltransferase involved in sulfur metabolism in phylogenetic tree.

frequently in the process of evolution to group A members, such as strain KGMB01988 (Table 6). The location of genes gained in evolution from the MRCA of strain KGMB01988 was displayed on the genome map, and the distribution of genes that were obtained was confirmed (Figure 7). Recently introduced genes have been clustered in some regions of the genome, whereas previously introduced genes have been scattered in the genome. Based on the GC skew or GC ratio in acquired genes, most of the acquired genes were determined to be introduced through horizontal gene transfer (HGT). In addition, recently introduced gene groups contained bacteriophage-related proteins and/or mobile elements, while the lack of mobile elements around genes introduced long ago may be due to loss during the evolution (Figure 7).

Comparison of type strain *KCTC 15667^T* and strain *KGMB01988*

We tried to determine the differences between the type strain and strain KGMB01988. Strain KGMB01988 had a relatively higher number of genes encoding Rhs repeat-associated core domain-containing proteins (WapA) than those in the type strain (Table 7 and Table S1). These genes were longer than average genes and were comprised of multiple domains. Furthermore, genes encoding the family GH84 had two copies in strain KGMB01988 and most of the *A. muciniphila* strains, whereas the type strain KCTC 15667^T had only one copy (Table 5). Interestingly, group A members can be divided into two clades, referred to as groups A1 and A2, depending on the presence of anSME (anaerobic sulfatase maturation enzyme; Figure 2, Table 7 and

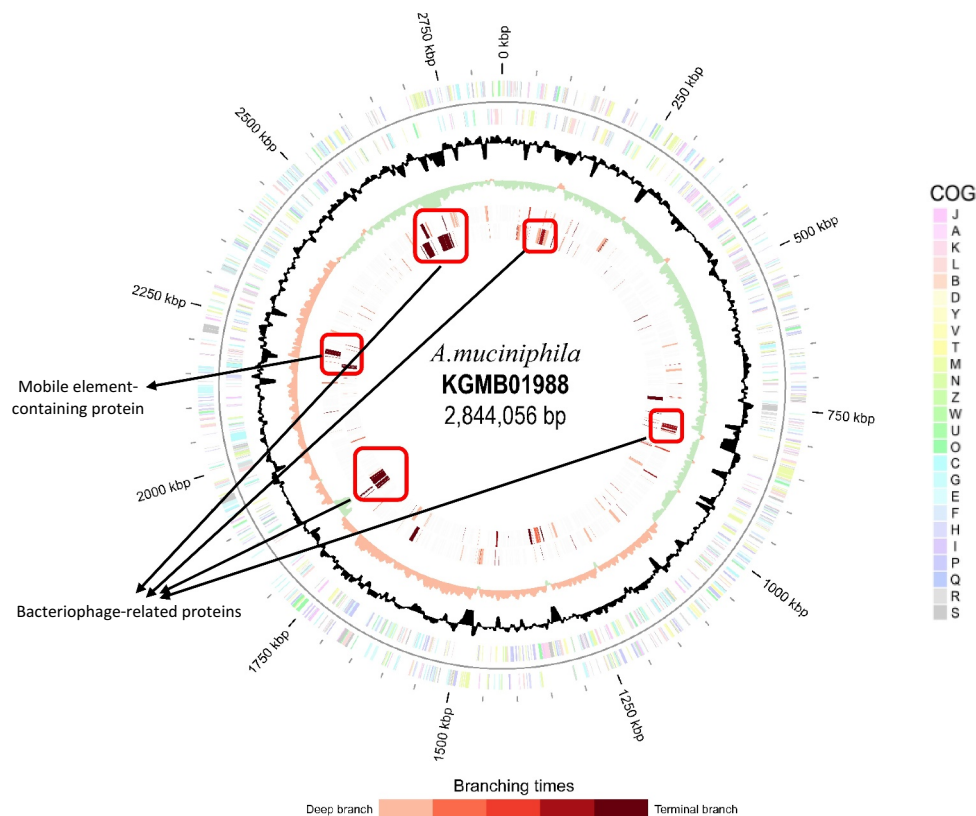


Figure 7. Graphical circular map of the chromosome of KGMB01988. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), GC content (black), GC skew (light green/Orange), and gained gene families on forward strand and negative strand after speciation (color by branching times).

Table S1). Group A1 members, including strain KGMB01988, had anSME, while group A2 and group B members, including type strain, did not, which was similar to the presence of sulfur metabolism genes. Since anSME accounts for the maturation of sulfatase under anaerobic conditions,⁴⁶ it is speculated that group A2 and B members have no sulfatase activity. The expression of anSME gene and sulfatase activities in *A. muciniphila* strains were also confirmed (Figure S3). Instead, it was confirmed that the type strain had additional sialidase (GH33, UniPlot ID A0A1C7PBY9) and α -L-fucosidase (GH29, UniPlot ID A0A1H6L863), which may cause a difference in mucin degradation compared to strain KGMB01988 (Table 5 and Table S2). Group B, including the type strain, contained copper oxidase, while group A members, including KGMB01988, did not (Table 7). However, it was expected to overcome heavy metal or copper stress by the copper efflux system because all strains of *A. muciniphila* were included. There were also differences in the types of glycosyltransferases and

acetyltransferases, and the number of genes encoding PEP-CTERM sorting domain-containing proteins (Table 7 and Table S1), which can be distinguished from the type strain by serotype through differences in EPS or LPS. In addition, it was observed that the number of glycosyltransferases and acetyltransferases also differed between the type strain KCTC 15667^T and KGMB01988.

Discussion

In vivo, *A. muciniphila* plays a crucial role in maintaining the integrity of the mucus layer, thereby decreasing intestinal permeability and subsequently reducing the penetration of gut-derived proinflammatory lipopolysaccharides. Accumulating evidence has uncovered the benefits of *A. muciniphila* to the host.⁴⁷ *A. muciniphila* treatment has been reported to restore high-fat diet-induced metabolic disorders in animal models by regulating adipose tissue metabolism, reducing insulin resistance, and maintaining glucose homeostasis.⁴⁸

The mucus layer is a niche colonized by a specific mucosal community that can degrade sugars and the protein backbone comprising mucin.⁴⁹ The mucus layer consists of an outer gel-forming layer that provides a habitat for bacteria, and an inner layer devoid of bacteria. MUC2 is the most abundant gel-forming mucin secreted in the intestine and is constructed of a PTS backbone with O-linked glycans. However, the diverse structure and linkages within the glycan chains make it difficult for most bacteria to access their amino acids and monosaccharides. Certain specialists, such as *A. muciniphila*, *B. bifidum* and *B. thetaiotaomicron*, have specific enzymes that degrade and use these mucins.^{49,50} In particular, sialidases and fucosidase play an important role in mucin degradation, since most of the terminal ends of these oligosaccharide chains are sialic acids or fucose. Sialidases and fucosidase are not commonly encoded in the metagenome of the microbiota. These terminal sugars are thought to prevent comprehensive microbial utilization of mucin.¹⁰ The other sugars in the oligosaccharide chain are more easily degraded by members of the microbiota. The mucinolytic activity of *A. muciniphila* leads to the accumulation of acetate and monosaccharides in the medium, and are able to ferment butyrate by cross-feeding butyrogenic bacteria, such as *Anaerostipes caccae*, *Eubacterium hallii*, and *Faecalibacterium prausnitzii*.

Most functional studies of *A. muciniphila* have been performed using only the type strain ATCC BAA-835^T. The genome of *A. muciniphila* ATCC BAA-835^T, comprising one circular chromosome of 2.66 Mb, was first sequenced in 2008 and announced in 2011.²⁰ This genome showed distinct phylogenetic features in contrast with other genomes of the *Verrucomicrobia* phylum, as it shared only about 29% of genes with its closest relatives, varying largely in GC content and genome size, indicating a unique and conservative evolutionary status of this bacterium.²⁰ Recently, strain-specific physiological properties of *A. muciniphila* were reported,³³ which were supported by genomic analysis of 39 *A. muciniphila* strains.⁵¹ Guo *et al.* reported that 39 *A. muciniphila* strains isolated from the gut of mammals revealed notable genomic diversity, which implies functional specificity.⁵¹ Therefore, it is particularly important to

investigate and dissect the genome content and strain-specific properties of *A. muciniphila* strains in order to select more promising candidates as probiotics. As of now, there are 153 assembled *A. muciniphila* genomes that were isolated from humans from various countries, and mice, chickens, and pigs, in GenBank/EMBL/DDBJ. In this study, only the complete genome sequences of the 27 *A. muciniphila* strains, including four KGMB strains isolated from healthy Koreans, were compared (Table 1–3). The genome sizes varied between 2.66 and 2.86 Mb and GC contents ranged from 55.23–55.76%. Strain KGMB01988 had the largest genome size after strain CBA5201 among *A. muciniphila* strains used in this study. The pan-genome of 27 *A. muciniphila* was open, indicating that the size of the pan-genome increased unboundedly with the increase in new genomes (Figure 1). In a broad sense, the size and expansion of pangenomes indicate a species's ability to adapt and evolve to specific environment. As a natural ecological habitat, the host GI tract constantly exerts a variety of long-term selection pressures to the intestinal bacteria that inhabit the colonic environment. Bacteria actively respond and interact with ecological niche and direct environment through genome adaptations such as gene acquisition and gene loss that occur naturally both *in vitro*⁵² and *in vivo*.^{53,54} It suggests the potential that isolated gut bacteria may undergo the molecular evolution and adaptive changes during long-term cultivation under specified growth conditions.⁵⁵

Comparative genomic analysis showed that *A. muciniphila* strains formed two clades in a genomic phylogenetic tree constructed using 1,219 orthologous single-copy core genes (Figure 2). All members of group A were isolated from Koreans, while most of group B contained strains isolated from human feces in Europe, China, and Korea, and from mouse. The comparison of the genome distance and the phylogeny revealed that strain KGMB01988 showed ANI and dDDH values of 98.03–99.90% and 82.10–98.90% in group A members, but of 97.44–97.66% and 77.70–78.90% in strains belonging to Group B (Table 4). According to Chen *et al.*,³⁶ it is suggested that *A. muciniphila* strains will be divided into two subspecies of group A, including KGMB

strains, and Group B, including type strain ATCC BAA-835^T (= KCTC 15667^T). Analysis of gene gain/loss events showed that gene gain plays an important role in dynamic evolution. In addition, as groups A and B were branched, mucin hydrolysis plays an important role in the stability of the core genome and drives evolution in the direction of defense, survival, and colonization in the mucus layer. To gain insights into the functional diversity of *A. muciniphila* strains, the pan-genome from the 27 *A. muciniphila* genomes was analyzed using the COG functional category (Figure 3). As the results suggest, the core genomes contained ‘cell motility’ (N), ‘carbohydrate transport and metabolism’ (G), ‘nucleotide transport and metabolism’ (F), and ‘translation, ribosomal structure and biogenesis’ (J). However, most accessory and unique genomes are involved in ‘defense mechanisms’ (V) and ‘replication, recombination, and repair’ (L). Enrichments in accessory and unique genes under specific environmental conditions may imply adaptation to the particular site or host. Analysis using the CAZy database also demonstrated that 24 of 26 GH families were present in the core genome of 27 *A. muciniphila* strains. However, there were differences in the numbers of families in GH18 and GH84 between groups A and B. To obtain deeper insight into the evolution of gene families on phylogeny, gene expansion and contraction at each branch of the strains were analyzed. The 2,021 CDSs were more likely to have been inherited from the MRCA of *A. muciniphila* strains, and the number of CDSs was increased by evolutionary events. Rhs protein (WapA) and sulfatase activity are considered to be particularly important selective pressures in the evolution of group A. Although a definitive function of WapA proteins has not been assigned, it has recently been reported that these proteins are involved in inter-cellular competition by contact-dependent growth inhibition.^{44,45} Interestingly, group A1 members, including KGMB01988 strain, possessed anSME, whereas group A2 and B members did not. This tendency is also observed in genes involved in sulfur metabolism. To be active, sulfatases must undergo a critical post-translational modification catalyzed in anaerobic bacteria by the radical AdoMet enzyme anSME.^{46,56} Therefore, it is assumed that groups A2 and B have no sulfatase

activity. However, it was confirmed that group B members additionally possessed sialidase and α -L-fucosidase genes compared to group A members, which may lead to differences in sulfated mucin degradation patterns between the type strain and strain KGMB01988. To confirm our genomic analysis, we determined anSME expression and sulfatase activity in *A. muciniphila* strains. The expression of anSME was determined by PCR using cDNA as template as described in Materials and Method. As the results, full size of anSME gene was present in only KGMB *A. muciniphila* strains, but not in type strain KCTC 15667^T (Figure S3A). Furthermore, as expected, no sulfatase activity was observed in type strain KCTC 15667^T without anSME (Figure S3B).

To date, although there have been reports of several gut bacteria that degrade mucin, the bacterial enzyme that initiates the breakdown of highly complex O-glycans found in mucins remains unclear. In the colon, these O-glycans are heavily sulfated, but specific sulfatases that are active on colonic mucins have not been identified. Previous reports have shown that *B. thetaiotaomicron* has a strong ability to grow on highly sulfated mucin oligosaccharides from colonic tissue and has active sulfatases capable of removing sulfates under all conditions known to cause sulfation in mucins.⁵⁶ Surprisingly, anSME deletion in *B. thetaiotaomicron* results in loss of sulfatase activity and impaired ability to use sulfated polysaccharides as a carbon source, resulting in drastically reduced competitive colonization in an animal model. These findings suggest that *A. muciniphila* KGMB strains may compete with other gut bacteria in the gut environment, as WapA and anSME act as drivers of evolution. Our findings provide insights into how *A. muciniphila* strains evolve to adapt to the gut environment.

Here we conducted the comparative pan-genome analysis using only complete genomes of *A. muciniphila* strains isolated from feces of Koreans, Chinese, Europeans, and mouse. Unexpectedly, *A. muciniphila* strains are divided into two groups of Korean isolates (group A) and non-Korean isolates (group B), based on the genomic relatedness and phylogenetic tree. To obtain deeper insight into the evolution from MRCA to two groups, gene gain and loss events

at each branch on a phylogeny were investigated. WapA and sulfatase activity are considered to be particularly important selective pressures in the evolution of group A including KGMB strains. Therefore, it is supposed that KGMB strains evolved to gain an edge in the competition with other gut bacteria via contact-dependent growth inhibition owing to WapA. In addition, KGMB strains utilize sulfated mucin owing to anSME presence, leading to become highly colonized in the gut.

However, the reason why the *wapA* gene has become an important driving force in the evolution of *A. muciniphila* isolated from Koreans remains unknown. The high sulfatase activity in Korean-originated strains is probably due to the dietary habits of Koreans who eat many algae, such as kelp and seaweed, which contain a large amount of sulfated carbohydrates. It has been reported that fucoidan, a sulfated carbohydrate from algae, increases the *Akkermansia* population in the mouse gut.⁵⁷ Based on this study, we plan to analyze the differences in the utilization of the sulfated carbohydrates between *A. muciniphila* strains, and to investigate the roles of sulfatase activity and WapA in human gut. Sulfatase activity is a key step in bacterial mucin degradation and has been reported to be associated with IBD and other diseases. Therefore, further studies will be needed on the prevention and treatment effects of KGMB *A. muciniphila* strains against IBD and various metabolic syndrome, since they are expected to show excellent colonization in the gut. We also hope that these findings will help researchers investigate the role of *A. muciniphila* in ecology and evolution, as well as the strain-specific probiotic potential of *A. muciniphila*.

Materials and methods

Isolation and culture of bacterial strains

To study gut microbiome of healthy Koreans, fresh stool samples were collected in Bundang Seoul National Hospital, Republic of Korea. Subjects were selected based on various health indicators (blood test, body mass index, antibiotic use,

smoking, alcohol use, drug use, and Bristol stool chart). The fecal samples were suspended in saline solution, serially diluted, and spread onto trypsin soy agar supplemented 5% horse blood (TSAB) plates. After an incubation of 3–5 d at 37°C in an anaerobic chamber filled with 90% N₂, 5% CO₂, and 5% H₂, single, white, and translucent colonies were isolated and transferred onto fresh TSA plates. To identify the bacterial strains, 16S rRNA gene sequencing was performed using the following universal bacterial primers: 27 F (5'-AGA GTT TGA TCC TGG CTC AG-3'), 1492 R (5'-TAC GGC TAC CTT GTT ACG ACT T-3'), 518 F (5'-CCA GCA GCC GCG GTA ATA CG-3') and 800 R (5'-TAC CAG GGT ATC TAA TCC-3'). The amplified PCR products were sequenced by Macrogen Inc., Korea.

Genomic DNA extraction and whole-genome sequencing

Genomic DNA was extracted from cells grown on TSAB as described previously.⁵⁸ Whole-genome sequencing of the *A. muciniphila* strains was performed using PacBio RS II single-molecule real-time (SMRT) sequencing technology (Pacific Biosciences). A standard PacBio library with an average of 20 kb inserts were prepared and were sequenced. De novo assembly was conducted using the hierarchical genome-assembly process (HGAP) pipeline of the SMRT Analysis v2.3.0. In order to correct sequencing errors that can occur at both ends of a contig, the SMRT resequencing protocol was performed with assembly that the first half of the contig was switched with the second half. As the result of assembly, *A. muciniphila* KGMB strains and type strain KCTC 15667^T had complete circular genome sequences as described in Table 1.

The expression of anSME gene in *A. muciniphila* strains was determined. Total RNA was extracted using Trizol reagent (Invitrogen). cDNA was synthesized from total RNA by using SuperScript III reverse transcriptase (Invitrogen). The synthesized cDNA was used as a template for amplification of full

sequence of anSME gene (1,412 bp) using the following primers: Forward: 5'-TACATATGAA TACTATTCTTCTCCCA-3', Reverse: 5'-TACTCGAG ATGAATCCAAGAATTCAT-3'.

Datasets

The genomic features, geographical origin, and isolation site characteristics of the genomic sequences used in this study are provided in Tables 1–3. Genomes and protein sequences were downloaded from the National Center for Biotechnology Information (NCBI) database, representing 24 different strains. The protein coding sequences (CDS) of each genome were predicted using Prodigal v.2.6.3. Orthologs were identified using the OrthoMCL program with an inflation value of 1.8.⁵⁹ The pan-genome profiles of the species *Akkermansia muciniphila* were evaluated and visualized using PanGP v.1.0.1.⁴⁰ The openness of the pan-genome was estimated using the R package micropan based on Heaps' law model.⁴¹

Construction of phylogenetic tree

Duplicated genes from core gene sets were excluded for the construction of the phylogenetic tree. Amino acid sequences of each ortholog were aligned with MUSCLE v3.8.31,⁶⁰ and aligned positions with >50% gaps were removed using Gblocks v0.91.⁶¹ The final gene alignments were concatenated using FASconCAT.⁶² The phylogenies based on the maximum likelihood approach were inferred with RAxML v8.2.4,⁶³ using the PROTGAMMAILGF model selected by ProtTest v3.4.⁶⁴ The trees were visualized using Dendroscope v3.2.2.⁶⁵

Genome similarity measures

The average nucleotide identity (ANI) is a widely accepted genomic method for species delineation. An ANI-based all-vs-all matrix and the resulting clustering tree were constructed using ANI calculator (Ezbiocloud website). Digital-DNA/DNA hybridization (dDDH) was calculated for pairs of genomes using the genome-to-genome distance calculator (GGDC),⁶⁶ by GGDC Formula 2 that is the most effective for incomplete genomes.⁶⁷

Gene gain and loss

The gain and loss events, and turnover rates of gene families by maximum likelihood were analyzed using the Gain-Death (GD) stochastic model in the BadiRate software with the phylogenetic maximum likelihood tree that we constructed.⁶⁸ We fitted two different branch models, a global-rates model and a free-rates model, to our data. The goodness of fit of these models was assessed using likelihood ratios. Functional assignment of gene families (containing orthologs and singleton ORFans) was conducted using BLAST searches with UniProt⁶⁹ and CAZy.⁷⁰ Many gene families were not functionally annotated (56.70% of gene families were annotated as hypothetical proteins). Only 1,995 gene families were mapped using Gene Ontology. Gene ontology (GO) enrichment analysis of gain or loss genes was conducted using the R package topGO.⁷¹

Measurement of sulfatase activity

vA. muciniphila strains were cultured in BHI medium containing 2.5 mg/ml mucin from porcine stomach type III and cultured in anaerobic condition at 37°C for 60 h. The bacterial cells resuspended in PBS were homogenized and centrifuged at 8,000 rpm for 10 min. The sulfatase activity was measured using Sulfatase Activity Assay Kit (Sigma). Sulfatase from *H. pomatia* (Sigma) was used as positive control.

Availability of Data and Materials

The genome sequencing data have been deposited in the GenBank/EMBL/DDBJ database of the National Center for Biotechnology Information (NCBI) under the Bioproject accession number PRJDB7416. The accession numbers of BioSample and genome are SAMN18309824 and CP071888 for KGMB01988, SAMN18309825 and CP071887 for KGMB 01989, SAMN18309826 and CP071886 for KGMB01990, SAMN18309827 and CP071885 for KGMB02009 and SAMN00138213 and CP71807 for KCTC15667^T, respectively. The data generated or analyzed during this study are included in this article and its supplemental information files.

Ethics statement

This study was conducted in accordance with the IRB regulations of the Korea Research Institute of Bioscience and Biotechnology (KRIBB) in Korea (P01-201702-31-007).

Acknowledgments

The authors would like to thank Dr Byung Kwon Kim for his advice and support in analyzing the data, and Dr Dong-Ho Lee and Dr Hyuk Yoon at Bundang Seoul National Hospital for providing the fecal samples for this study.

Abbreviations

KCTC	Korean Collection for Type Cultures
TSA	tryptic soy agar; ANI, average nucleotide identity
GH	glycoside hydrolases
CDS	coding sequences of proteins

Author contributions

J-SK performed all the experiments and data analysis, and wrote the manuscript. SWK, JHL, S-HP, and J-SL guided the experimental design and data interpretation. J-SL edited the manuscript and supervised this study. All authors read and approved the final manuscript.


Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (NRF-2016M3A9F3947962) and Korea Research Institute of Bioscience and Biotechnology (KRIBB) Research Initiative Program (KGM5232221).

ORCID

Ji-Sun Kim  <http://orcid.org/0000-0003-4736-6010>
 Seung-Hwan Park  <http://orcid.org/0000-0003-4613-4680>
 Jung-Sook Lee  <http://orcid.org/0000-0003-0300-8836>

References

1. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010;464(7285):59–65. doi:10.1038/nature08821.
2. Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes*. 2012;3:289–306. doi:10.4161/gmic.19897.
3. Den Besten G, van Eunen K, Groen AK, Venema K, Reijngoud DJ, Bakker BM. The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res*. 2013;54:2325–2340. doi:10.1194/jlr.R036012.
4. Gerritsen J, Smidt H, Rijkers GT, de Vos WM. Intestinal microbiota in human health and disease: the impact of probiotics. *Genes Nutr*. 2011;6:209–240. doi:10.1007/s12263-011-0229-7.
5. Belkaid Y, Hand TW. Role of the microbiota in immunity and inflammation. *Cell*. 2014;157:121–141. doi:10.1016/j.cell.2014.03.011.
6. Duerr CU, Hornef MW. The mammalian intestinal epithelium as integral player in the establishment and maintenance of host-microbial homeostasis. *Semin Immunol*. 2012;24:25–35. doi:10.1016/j.smim.2011.11.002.
7. Ouwerkerk JP, de Vos WM, Belzer C. Glycobiome: bacteria and mucus at the epithelial interface. *Best Pract Res Clin Gastroenterol*. 2013;27:25–38. doi:10.1016/j.bpg.2013.03.001.
8. Koropatkin NM, Cameron EA, Martens EC. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol*. 2012;10:323–335. doi:10.1038/nrmicro2746.
9. Paone P, Cani PD. Mucus barrier, mucins and gut microbiota: the expected slimy partners? *Gut*. 2020;69:2232–2243. doi:10.1136/gutjnl-2020-322260.
10. Tailford LE, Crost EH, Kavanaugh D, Juge N. Mucin glycan foraging in the human gut microbiome. *Front Genet*. 2015;6:81. doi:10.3389/fgene.2015.00081.
11. Everard A, Belzer C, Geurts L, Ouwerkerk JP, Druart C, Bindels LB, Guiot Y, Derrien M, Muccioli GG, Delzenne NM, et al. Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. *Proc Natl Acad Sci USA*. 2013;110(22):9066–9071. doi:10.1073/pnas.1219451110.
12. Plovier H, Everard A, Druart C, Depommier C, Van Hul M, Geurts L, Chilloux J, Ottman N, Duparc T, Lichtenstein L, et al. A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. *Nat Med*. 2017;23:107–113. doi:10.1038/nm.4236.
13. Belzer C, Chia L, Aalvink S, Chamlagain BVP, de Vos WM, Knol J, de Vos WM. Microbial metabolic networks at the Mucus Layer lead to

- diet-independent butyrate and vitamin B12 production by intestinal symbionts. *mBio*. 2017;8(5): e00770–e00817. doi:10.1128/mBio.00770-17.
14. Derrien M, Vaughan EE, Plugge CM, de Vos WM. *Akkermansia muciniphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. *Int J Syst Evol Microbiol*. 2004;54:1469–1476. doi:10.1099/ijs.0.02873-0.
 15. Rooijers K, Kolmeder C, Juste C, Doré J, de Been M, Boeren S, Galan P, Beauvallet C, de Vos WM, Schaap PJ. An iterative workflow for mining the human intestinal metaproteome. *BMC Genomics*. 2011;12:6. doi:10.1186/1471-2164-12-6.
 16. Collado MC, Derrien M, Isolauri E, de Vos WM, Salminen S. Intestinal integrity and *Akkermansia muciniphila*, a mucin-degrading member of the intestinal microbiota present in infants, adults, and the elderly. *Appl Environ Microbiol*. 2007;73:7767. doi:10.1128/AEM.01477-07.
 17. Derrien M, Collado MC, Ben-Amor K, Salminen S, de Vos WM. The mucin degrader *Akkermansia muciniphila* is an abundant resident of the human intestinal tract. *Appl Environ Microbiol*. 2008;74:1646–1648. doi:10.1128/AEM.01226-07.
 18. Png CW, Lindén SK, Gilshenan KS, Zoetendal EG, McSweeney CS, Sly LI, McGuckin MA, Florin TH. Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. *Am J Gastroenterol*. 2010;105:2420–2428. doi:10.1038/ajg.2010.281.
 19. Lyra A, Forssten S, Rolny P, Wettergren Y, Lahtinen SJ, Salli K, Cedgård L, Odin E, Gustavsson B, Ouwehand AC. Comparison of bacterial quantities in left and right colon biopsies and faeces. *World J Gastroenterol*. 2012;18:4404–4411. doi:10.3748/wjg.v18.i32.4404.
 20. van Passel MW, Kant R, Zoetendal EG, Plugge CM, Derrien M, Malfatti SA, Chain PS, Woyke T, Palva A, de Vos WM, et al. The genome of *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. *PLoS One*. 2011;6:e16876. doi:10.1371/journal.pone.0016876.
 21. Puertollano E, Kolida S, Yaqoob P. Biological significance of short-chain fatty acid metabolism by the intestinal microbiome. *Curr Opin Clin Nutr Metab Care*. 2014;17:139–144. doi:10.1097/MCO.0000000000000025.
 22. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, Pudlo NA, Kitamoto S, Terrapon N, Muller A, et al. A dietary fiber-deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell*. 2016;167:1339–1353. doi:10.1016/j.cell.2016.10.043.
 23. Mack I, Cuntz U, Grämer C, Niedermaier S, Pohl C, Schwartz A, Zimmermann K, Zipfel S, Enck P, Penders J. Weight gain in anorexia nervosa does not ameliorate the faecal microbiota, branched chain fatty acid profiles, and gastrointestinal complaints. *Sci Rep*. 2016;6:26752. doi:10.1038/srep26752.
 24. Louis S, Tappu RM, Damms-Machado A, Huson DH, Bischoff SC. Characterization of the gut microbial community of obese patients following a weightloss intervention using whole metagenome shotgun sequencing. *PLoS One*. 2016;11:e0149564. doi:10.1371/journal.pone.0149564.
 25. Everard A, Lazarevic V, Derrien M, Girard M, Muccioli GG, Neyrinck AM, Possemiers S, Van Holle A, François P, de Vos WM, et al. Responses of gut microbiota and glucose and lipid metabolism to prebiotics in genetic obese and diet-induced leptin-resistant mice. *Diabetes*. 2011;60:2775–2786. doi:10.2337/db11-0227.
 26. Ottman N, Davids M, Suarez-Diez M, Boeren S, Schaap PJ, Dos Santos VAP M, Smidt H, Belzer C, de Vos WM. Genome-scale model and omics analysis of metabolic capacities of *Akkermansia muciniphila* reveal a preferential mucin-degrading lifestyle. *Appl Environ Microbiol*. 2017;83:e01014–e01017. doi:10.1128/AEM.01014-17.
 27. Ottman N, Huuskonen L, Reunanen J, Boeren S, Klievink J, Smidt H, Belzer C, de Vos WM. Characterization of outer membrane proteome of *Akkermansia muciniphila* reveals sets of novel proteins exposed to the human intestine. *Front Microbiol*. 2016;7:1157. doi:10.3389/fmicb.2016.01157.
 28. Lopez-Siles M, Enrich-Capó N, Aldeguer X, Sabat-Mir M, Duncan SH, Garcia-Gil LJ, Martinez-Medina M. Alterations in the abundance and co-occurrence of *Akkermansia muciniphila* and *Faecalibacterium prausnitzii* in the colonic mucosa of inflammatory bowel disease subjects. *Front Cell Infect Microbiol*. 2018;8:281. doi:10.3389/fcimb.2018.00281.
 29. Morgan XC, Kabackchiev B, Waldron L, Tyler AD, Tickle TL, Milgrom R, Stempak JM, Gevers D, Xavier RJ, Silverberg MS, et al. Associations between host gene expression, the mucosal microbiome, and clinical outcome in the pelvic pouch of patients with inflammatory bowel disease. *Genome Biol*. 2015;16:67. doi:10.1186/s13059-015-0637-x.
 30. Naito Y, Uchiyama K, Takagi T. A next-generation beneficial microbe: *akkermansia muciniphila*. *Naito Y, Uchiyama K, Takagi T. J Clin Biochem Nutr*. 2018;63:33–35. doi:10.3164/jcbtn.18-57.
 31. Cani PD, de Vos WM. Next-generation beneficial microbes: the case of *Akkermansia muciniphila*. *Front Microbiol*. 2017;8:1765. doi:10.3389/fmicb.2017.01765.
 32. Dao MC, Everard A, Aron-Wisniewsky J, Sokolovska N, Prifti E, Verger EO, Kayser BD, Levenez F, Chilloux J, Hoyle L, et al. *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut*. 2016;65:426–436. doi:10.1136/gutjnl-2014-308778.

33. Zhai R, Xue X, Zhang L, Yang X, Zhao L, Zhang C. Strain-specific anti-inflammatory properties of two *Akkermansia muciniphila* strains on chronic colitis in mice. *Front Cell Infect Microbiol.* 2019;9:239. doi:10.3389/fcimb.2019.00239.
34. Gupta PK. Single-molecule DNA sequencing technologies for future genomics research. *Trends Biotechnol.* 2008;26:602–611. doi:10.1016/j.tibtech.2008.07.003.
35. McCarthy A. Third generation DNA sequencing: pacific Biosciences' single molecule real time technology. *Chem Biol.* 2010;17:675–676. doi:10.1016/j.chembiol.2010.07.004.
36. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, Da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol.* 2018;68:461–466. doi:10.1099/ijsem.0.002516.
37. Mann PJ, Woolf B. The action of salts on fumarase. *Biochem J.* 1930;24:427–434. doi:10.1042/bj0240427.
38. Woods SA, Schwartzbach SD, Guest JR. Two biochemically distinct classes of fumarase in *Escherichia coli*. *Biochim Biophys Acta.* 1988;954:14–26. doi:10.1016/0167-4838(88)90050-7.
39. Enright AJ, Van Dongen S, Quzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 2002;30:1575–1584. doi:10.1093/nar/30.7.1575.
40. Zhao Y, Jia X, Yang J, Ling Y, Zhang Z, Yu J, Wu J, Xiao J. PanGP: a tool for quickly analyzing bacterial pan-genome profile. *Bioinformatics.* 2014;30:1297–1299. doi:10.1093/bioinformatics/btu017.
41. Snipen L, Liland KH. micropan: an R-package for microbial pan-genomics. *BMC Bioinform.* 2015;16:79. doi:10.1186/s12859-015-0517-0.
42. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol.* 2008;11:472–477. doi:10.1016/j.mib.2008.09.006.
43. Li N, Wang K, Williams HN, Sun J, Ding C, Leng X, Dong K. Analysis of gene gain and loss in the evolution of predatory bacteria. *Gene.* 2017;598:63–70. doi:10.1016/j.gene.2016.10.039.
44. Koskiniemi S, Lamoureux JG, Nikolakakis KC, T'kint de Roodenbeke C, Kaplan MD, Low DA, Hayes CS. Rhs proteins from diverse bacteria mediate intercellular competition. *Proc Natl Acad Sci USA.* 2013;110:7032–7037. doi:10.1073/pnas.1300627110.
45. Koskiniemi S, Garza-Sánchez F, Sandegren L, Webb JS, Braaten BA, Poole SJ, Andersson DI, Hayes CS, Low DA. Selection of orphan Rhs toxin expression in evolved *Salmonella enterica* serovar Typhimurium. *PLoS Genet.* 2014;10:e1004255. doi:10.1371/journal.pgen.1004255.
46. Benjdia A, Martens EC, Gordon JI, Berteau O. Sulfatases and a radical S-adenosyl-L-methionine (AdoMet) enzyme are key for mucosal foraging and fitness of the prominent human gut symbiont *Bacteroides Thetaiotaomicron*. *J Biol Chem.* 2011;286(29):25973–25982. doi:10.1074/jbc.M111.228841.
47. Zhang T, Li Q, Cheng L, Buch H, Zhang F. *Akkermansia muciniphila* is a promising probiotic. *Microb Biotechnol.* 2019;12:1109–1125. doi:10.1111/1751-7915.13410.
48. Zhai Q, Feng S, Arhan N, Chen W. A next generation probiotic *Akkermansia Muciniphila*. *Crit Rev Food Sci Nutr.* 2019;59:3227–3236. doi:10.1080/10408398.2018.1517725.
49. Backhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. *Science.* 2005;307:1915e20. doi:10.1126/science.1104816.
50. Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science.* 2003;299:2074e6. doi:10.1126/science.1080029.
51. Guo X, Li S, Zhang J, Wu F, Li X, Wu D, Zhang M, Ou Z, Jie Z, Yan Q, et al. Genome sequencing of 39 *Akkermansia muciniphila* isolates reveals its population structure, genomic and functional diversity, and global distribution in mammalian gut microbiotas. *BMC Genomics.* 2017;18:800. doi:10.1186/s12864-017-4195-3.
52. Barroso-Batista J, Demengeot J, Gordo I. Adaptive immunity increases the pace and predictability of evolutionary change in commensal gut bacteria. *Nat Commun.* 2015;6:8945. doi:10.1038/ncomms9945.
53. Lawrence D, Fiegna F, Behrends V, Bundy JG, Phillimore AB, Bell T, Barraclough TG. Species interactions alter evolutionary responses to a novel environment. *PLoS Biol.* 2012;10:e1001330. doi:10.1371/journal.pbio.1001330.
54. Li W, Yao G, Cai H, Bai M, Kwok L-Y, Sun Z. Comparative genomics of in vitro and in vivo evolution of probiotics reveals energy restriction not the main evolution driving force in short term. *Genomics.* 2021;113:3373–3380. doi:10.1016/j.ygeno.2021.07.022.
55. Hao P, Zheng H, Yu Y, Ding G, Gu W, Chen S, Yu Z, Ren S, Oda M, Konno T, et al. Complete sequencing and pan-genomic analysis of *Lactobacillus delbrueckii* subsp. *bulgaricus* reveal its genetic basis for industrial yogurt production. *PLoS One.* 2011;6:e15964. doi:10.1371/journal.pone.0015964.
56. Luis A, Jin C, Pereira GV, Glowacki R, Gugel S, Singh S, Byrne D, Pudlo N, London J, Baslé A, et al. A single bacterial sulfatase is required for metabolism of colonic mucin O-glycans and intestinal colonization by a symbiotic human gut bacterium. doi:10.1101/2020.11.20.392076. 2020. fhal-03025006f.
57. Shang Q, Song G, Zhang M, Shi J, Xu C, Hao J, Li, G, Yu, G. Dietary fucoidan improves metabolic syndrome in association with increased *Akkermansia* population in the gut microbiota of high-fat diet-fed mice. *J Funct Foods.* 2017;28:138–146. doi:10.1016/j.jff.2016.11.002.

58. Kim JS, Lee KC, Suh MK, Han KI, Eom MK, Lee JH, Park SH, Kang SW, Park JE, Oh BS, et al. *Mediterraneibacter butyricigenes* sp. nov., a butyrate-producing bacterium isolated from human faeces. *J Microbiol.* 2019;57(1):38–44. doi:10.1007/s12275-019-8550-8.
59. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178–2189. doi:10.1101/gr.1224503.
60. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–1797. doi:10.1093/nar/gkh340.
61. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol.* 2000;17:540–552. doi:10.1093/oxfordjournals.molbev.a026334.
62. Kück P, Meusemann K. FASconCAT: convenient handling of data matrices. *Mol Phylogenet Evol.* 2010;56:1115–1118. doi:10.1016/j.ympev.2010.04.024.
63. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–1313. doi:10.1093/bioinformatics/btu033.
64. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics.* 2005;21:2104–2105. doi:10.1093/bioinformatics/bti263.
65. Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinform.* 2007;8:460. doi:10.1186/1471-2105-8-460.
66. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* 2013;14:60. doi:10.1186/1471-2105-14-60.
67. Auch AF, Von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci.* 2010;2:117–134. doi:10.4056/sigs.531120.
68. Librado P, Vieira FG, Rozas J. BadiRate: estimating family turnover rates by likelihood-based methods. *Bioinformatics.* 2012;28:279–281. doi:10.1093/bioinformatics/btr623.
69. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, et al. The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.* 2006;34:D187–D191. doi:10.1093/nar/gkj161.
70. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 2014;42:D490–D495. doi:10.1093/nar/gkt1178.
71. Alexa A, Rahnenfuhrer J. topGO: enrichment analysis for gene ontology. 2019. R package version 2.36.0.