1   **Supplementary Figures**

2

3   **Pangenomic analysis identifies correlations between *Akkermansia***

4   **species and subspecies and human health outcomes**

5

6   **Katherine D. Mueller[1,2], M. Emilia Panzetta[1,2], Lauren Davey[3], Jessica R.**

7   **McCann[4], John F. Rawls[2,4], Gilberto E. Flores[5], Raphael H. Valdivia[1,2]**

8

9   [1]Department of Integrative Immunobiology, Duke University, Durham, NC 27710,
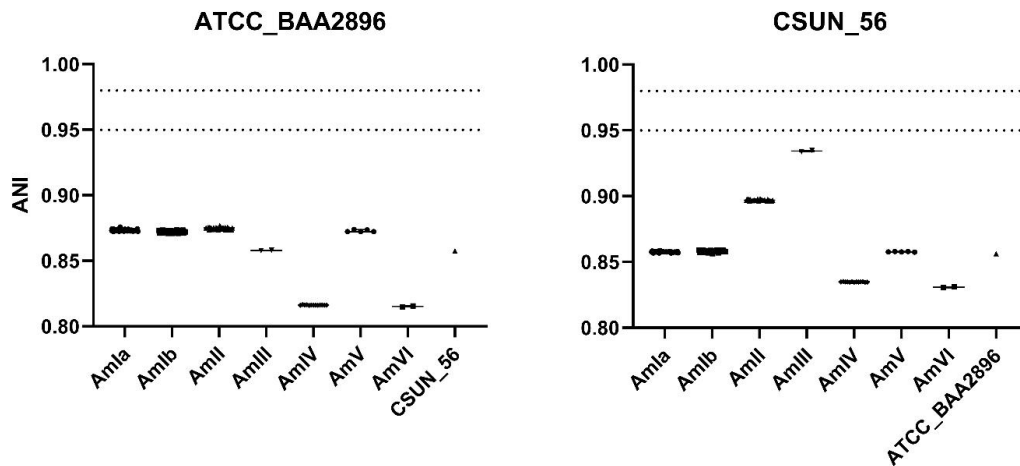
10   USA.

11   [2]Duke Microbiome Center, Duke University School of Medicine, Durham, NC 27710,

12   USA.

13   [3]Department of Biochemistry and Microbiology, University of Victoria, Victoria V8P

14   5C2, British Columbia, Canada.

15   [4]Department of Molecular Genetics and Microbiology, Duke University, Durham, NC
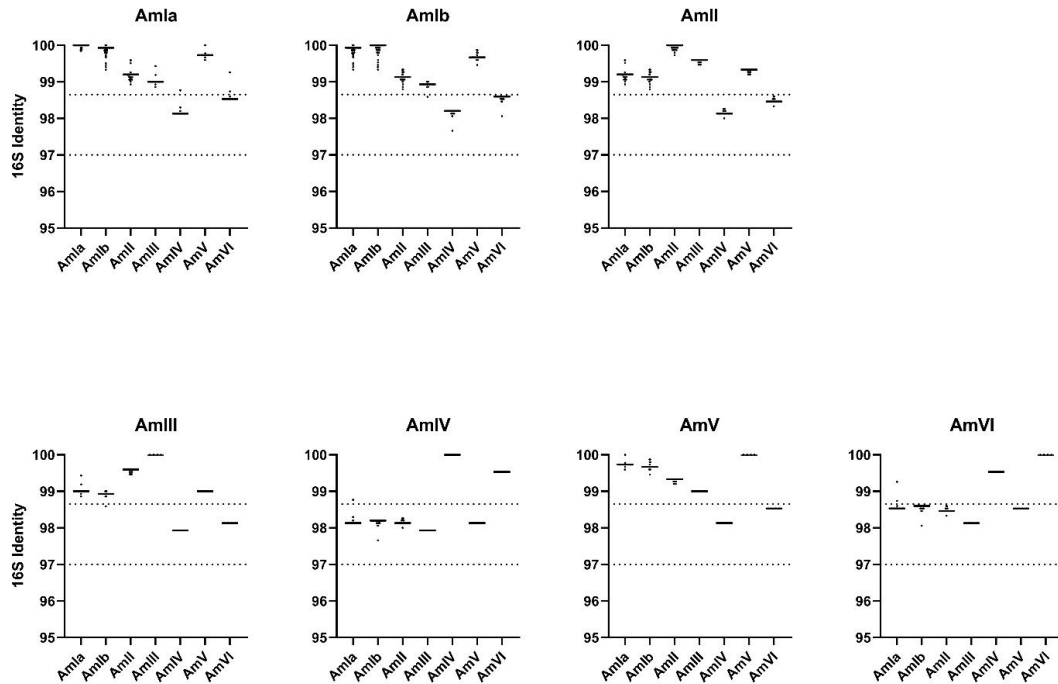
16   27710, USA.

17   [5]Department of Biology, California State University, Northridge, CA 91330, USA.

18

19   **Correspondence to:** Dr. Raphael H. Valdivia, Department of Integrative

20   Immunobiology, Duke University, 3110 MSRB3, 3 Genome Court, Durham, NC

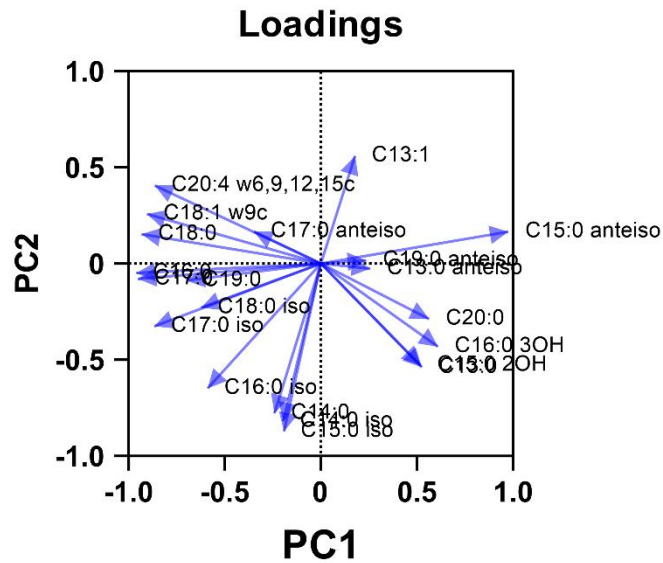21   27710, USA. E-mail: raphael.valdivia@duke.edu

ATCC_BAA2896

CSUN_56

22
23   **Supplementary Figure 1. Analysis of average nucleotide identity suggests that**

24   **two strains previously classified as AmIII may represent two novel species.**
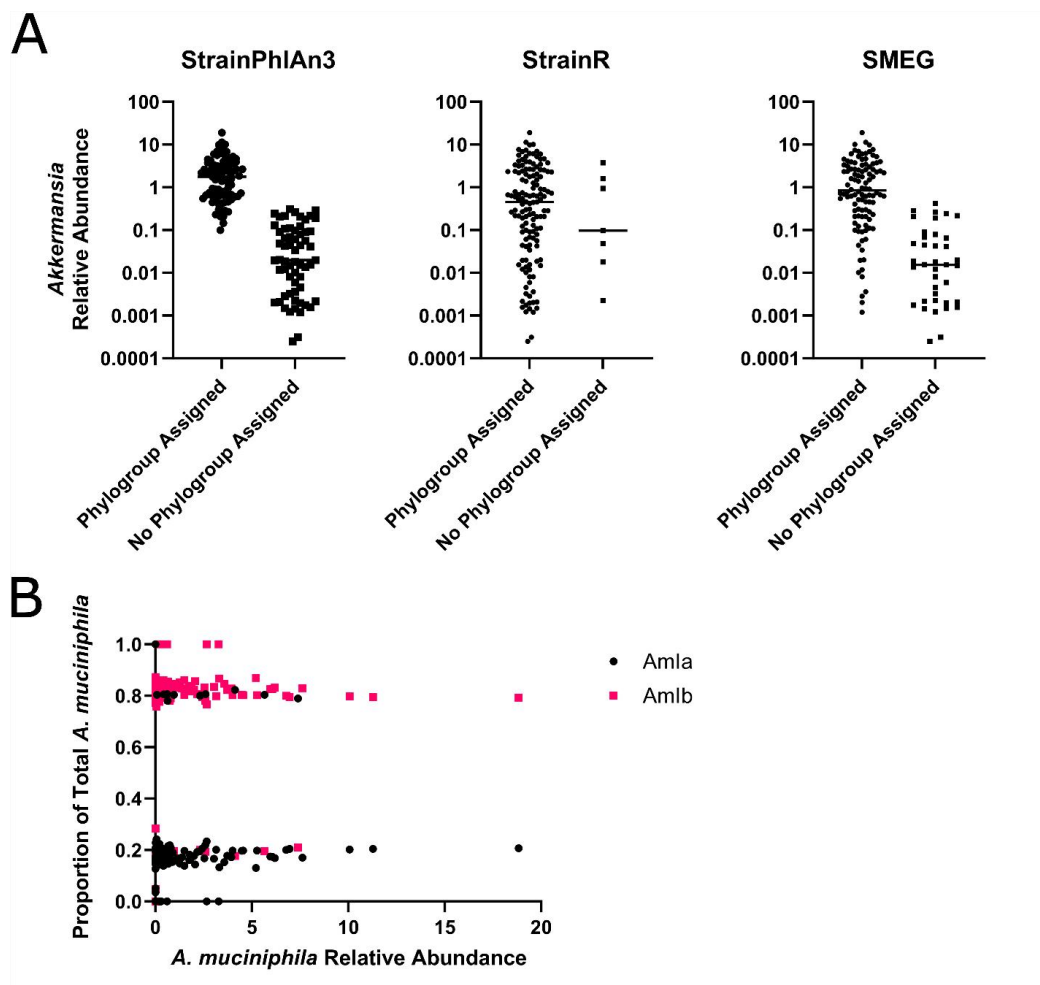
25   Average nucleotide identity (ANI) based on whole-genome comparisons between

26   ATCC BAA2896, CSUN-56, and all other *Akkermansia* phylogroups, as calculated

27   using pyANI in Anvi'o. Differences in average ANI below 95% were used to denote

28   new species among phylogroups, and an average ANI between 95% and 98% was

29   defined to assign new sub-phylogroups within those new species (dotted lines on each

30   subplot). From this, it may be determined that ATCC BAA2896 and CSUN-56 should

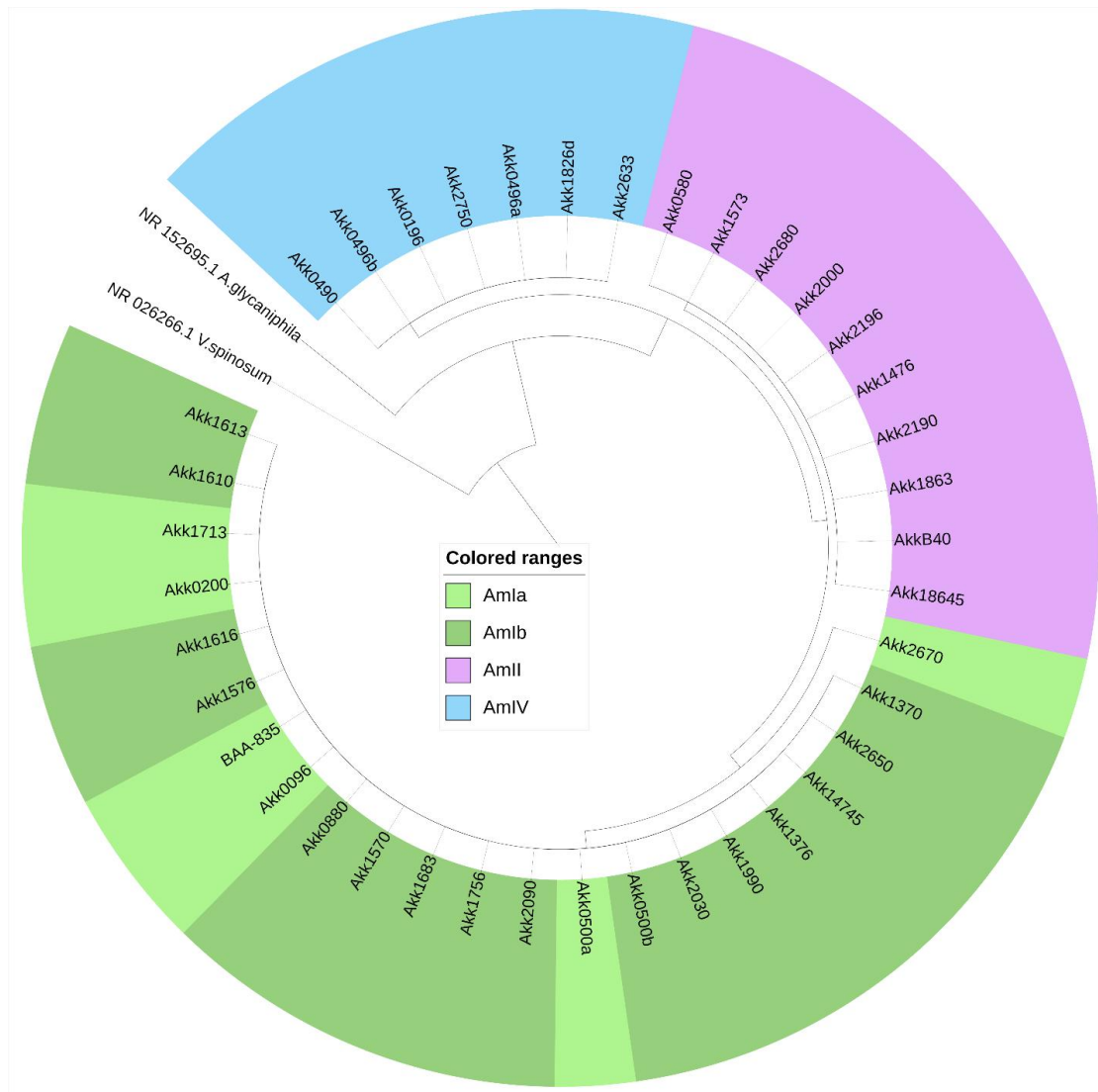31   each be considered single representatives of new species.

32

Supplementary Figure 2. The average 16S rRNA gene sequence identity between phylogroups does not meet commonly used thresholds for delineating species.
Full-length 16S rRNA gene sequences were extracted from genome sequences within Anvi'o. Clustal Omega was then used to compare the average sequence identity between phylogroups. The dotted lines indicate 97% and 98.65% identity, which are commonly used thresholds to delineate species.

Loadings

Supplementary Figure 3. Loadings from a principal components analysis suggest that the factors driving separation across PC2 include C14:0 iso and C15:0 iso. A principal components analysis was performed to compare fatty acid composition across species. The biplot [Figure 2A] revealed that *A. massiliensis* could be clearly separated from other *Akkermansia* species along PC2. Loadings indicate that the most informative fatty acid features along PC2 include C14:0 iso and C15:0 iso.

Supplementary Figure 4. Each of the strain-finding programs used in the identification of *Akkermansia* species from metagenomic sequencing data is imperfect. (A) StrainPhlAn3, StrainR, and SMEG were each run on the entire POMMS dataset to determine how many samples for an *Akkermansia* species could be identified. StrainPhlAn3 frequently did not detect *Akkermansia* in samples where the relative abundance was less than 0.5% of total bacterial sequences. StrainR and SMEG each identified *Akkermansia* species and phylogroups in the same low-abundance samples with greater rates of success; **(B)** The proportion of AmIa and AmIb present in samples, as calculated using StrainR, was determined according to the total relative abundance of AmI in metagenomic samples. Black indicates AmIa and pink indicates AmIb. Note that the proportion of these AmI phylogroups does not change in association with relative abundance.

Supplementary Figure 5. Sub-phylogroups of *A. muciniphila* are not resolvable by their V3-V4 16S rRNA sequences. The V3-V4 region of the 16S rRNA gene sequences of 39 *Akkermansia* isolates were aligned using Clustal Omega, and the resulting phylogenetic tree was used for visualization. *A. glycaniphila* and *V. spinosum* were used as outgroups. Individual strains have been highlighted by species, showing clear clustering of *A. massiliensis* and *A. biwaensis*. However, the AmIa and AmIb isolates of *A. muciniphila* are not able to be differentiated by the V3-V4 region alone.