

1 **Isolates from colonic spirochaetosis in humans show high genomic divergence and carry**  
2 **potential pathogenic features but are not detected by 16S amplicon sequencing using**  
3 **standard primers for the human microbiota**

4

5 Kaisa Thorell <sup>a,b</sup>, Linn Inganäs <sup>a,c</sup>, Annette Backhans <sup>d</sup>, Lars Agréus <sup>c</sup>, Åke Öst <sup>e</sup>, Marjorie  
6 Walker <sup>f</sup>, Nicholas J Talley <sup>f</sup>, Lars Kjellström <sup>g</sup>, Anna Andreasson <sup>c,h</sup>, Lars Engstrand <sup>a,i</sup>

7

8 <sup>a</sup> Center for Translational Microbiome Research, Department of Microbiology, Cell and  
9 Tumor biology, Karolinska Institutet, Stockholm, Sweden

10 <sup>b</sup> Department for Infectious Diseases, Sahlgrenska Academy, University of Gothenburg,  
11 Gothenburg, Sweden

12 <sup>c</sup> Division for Family Medicine and General Practice, Department for Neurobiology, Care  
13 Sciences and Society, Karolinska Institutet, Huddinge, Sweden

14 <sup>d</sup> Department of Clinical Sciences, Swedish University of Agriculture, Uppsala, Sweden

15 <sup>e</sup> Pathology, Aleris Medilab, Taby, Sweden.

16 <sup>f</sup> University of Newcastle, New South Wales, Australia

17 <sup>g</sup> Gastromottagningen City, Stockholm, Sweden

18 <sup>h</sup> Stress Research Institute, Stockholm University, Stockholm, Sweden

19 <sup>i</sup> Clinical Genomics, Science for Life Laboratory, Stockholm, Sweden

20 Corresponding authors: [kaisa.thorell@ki.se](mailto:kaisa.thorell@ki.se), [lars.engstrand@ki.se](mailto:lars.engstrand@ki.se)

21

1 **Abstract**

2 Colonic spirochaetosis, diagnosed based on the striking appearance in histological sections,  
3 still has an obscure clinical relevance and only few bacterial isolates from this condition have  
4 been characterized to date. In a randomized, population-based study in Stockholm, Sweden,  
5 745 healthy individuals underwent colonoscopy with biopsy sampling. In these individuals,  
6 17 (2.3 %) had colonic spirochaetosis, which was associated with eosinophilic infiltration and  
7 a three-fold increased risk for irritable bowel syndrome (IBS). We aimed to culture the  
8 bacteria and perform whole genome sequencing of the isolates from this unique  
9 representative population sample. From 14 out of 17 individuals with spirochaetosis we  
10 successfully isolated, cultured and performed whole genome sequencing of in total 17  
11 isolates including the *Brachyspira aalborgi* type strain 513A<sup>T</sup>. Also, 16S analysis of the  
12 mucosa-associated microbiota was performed in the cases and non-spirochaetosis controls.

13 This is the first report of whole genome analysis of clinical isolates from individuals with  
14 colonic spirochaetosis. We found one isolate to be of the species *Brachyspira pilosicoli* and  
15 all remaining isolates were of the species *Brachyspira aalborgi*. Besides displaying extensive  
16 genetic heterogeneity, the isolates harboured several mucin-degrading enzymes and other  
17 virulence-associated genes that could confer a pathogenic potential in the human colon. We  
18 also showed that 16S amplicon sequencing using standard primers for human microbiota  
19 studies fail to detect *Brachyspira* due to primer incompatibility. This failure to detect colonic  
20 spirochaetosis should be taken into consideration in project design and interpretation of  
21 gastrointestinal tract microbiota in population-based and clinical settings.

22

## 1 **Introduction:**

2 Colonic spirochaetosis (CS) was first described in 1967, when spirochaetes adherent to the  
3 surface of colonic epithelium were identified by electron microscopy of rectal biopsies in a  
4 seminal paper by Harland and Lee who coined the term intestinal spirochaetosis <sup>1</sup>. The  
5 characteristic histological appearance of the bacteria has been considered to be  
6 pathognomonic for diagnosis, but the biology and origin of intestinal spirochaetes in humans  
7 is still poorly understood.

8 The nosology for spirochaetes in the human intestine has varied from *Borrelia* and *Serpulina*,  
9 to the present classification *Brachyspira*, and spirochaetal bacteria have been identified in  
10 the intestines of several animals including monkeys, dogs, chickens, rodents and pigs <sup>2</sup>. In  
11 animals with colonic spirochaetosis, a spectrum of medical conditions is well described.  
12 While rodents present with asymptomatic excretion, in swine the spirochaetes can cause  
13 pathologic changes leading to diarrhoea, malnutrition, and declining growth rates, resulting  
14 in high economic losses <sup>3</sup>. Primates occupy a rather intermediate position, where even  
15 though spirochaetal organisms can colonize the colonic mucosa, the animals rarely present  
16 with enteric symptoms <sup>4</sup>. In humans, at least two spirochaete species, *Brachyspira pilosicoli*  
17 and *Brachyspira aalborgi*, are associated with spirochaetosis <sup>5</sup>, and rare findings of  
18 concomitant infection by *B. aalborgi* and *B. pilosicoli* have been described <sup>6,7</sup>. These  
19 spirochaetes are both fastidious and rather slow-growing anaerobes <sup>8</sup>, and successful culture  
20 of spirochaetes from stool or biopsy material is rarely reported in human studies. However,  
21 the type species of the genus *Brachyspira* (*B. aalborgi*) was originally isolated from a patient  
22 with diarrhoea <sup>9</sup> and more recently another strain isolated from a colonic biopsy in a patient  
23 with blood and mucus in stool was reported <sup>10</sup>. However, no *B. aalborgi* and only four *B.*  
24 *pilosicoli* isolates have been genomically characterized on the whole genome level to date <sup>11</sup>.

25 Even though there has been progress in detecting and identifying spirochaetosis in humans,  
26 studies based on histopathological diagnosis without detailed symptom correlation has left  
27 the question as to whether it represents a disease process unanswered <sup>12,13</sup>. However, more  
28 recent studies have shown an association between spirochaetosis and the irritable bowel  
29 syndrome (IBS), and identified a unique colonic pathology characterized by increased  
30 eosinophils in this condition <sup>14,15</sup>. As IBS affects one in 10 people worldwide, impairs quality

1 of life and is highly costly, the observation that colonic spirochaetes are associated with IBS-  
2 diarrhoea is of major interest <sup>16</sup>.

3 In a unique representative random population sample that underwent a study colonoscopy  
4 in Sweden, we aimed to culture and perform whole genome sequencing of colonic  
5 spirochaetosis isolates obtained from colonic biopsies. To characterize isolates from human  
6 spirochaetosis, we in this study cultured spirochaetal isolates from individuals with  
7 microscopically determined spirochaetosis from within the randomized, population-based  
8 colonoscopy study PopCol that was performed in Stockholm, Sweden in 2001-2006 <sup>17</sup>. A total  
9 of 745 healthy adults underwent colonoscopy with biopsy sampling of four sites of the colon  
10 and of the terminal ileum. Out of these individuals, 17 presented with spirochaetosis, an  
11 observation that was shown to be associated with eosinophilic infiltration in the tissue and a  
12 three-fold increased risk for IBS <sup>14</sup>. Since the spirochaetes were found solely in the colon  
13 (and not in the terminal ileum), we will hereafter use the term colonic spirochaetosis (CS) <sup>3</sup>.  
14 To further characterize the spirochaete bacteria and their effect on the colonic microbiota,  
15 we performed whole- genome sequencing of the bacterial isolates together with type strain  
16 *513A<sup>T</sup>* [ATCC 43994] <sup>9</sup>, and reference strain W1 <sup>10</sup>, as well as performed 16S amplicon  
17 sequencing for microbiota profiling of colonic biopsies in the individuals with colonic  
18 spirochaetosis and unaffected controls. We here present the first whole-genome sequences  
19 of isolates from human colonic spirochaetosis, which display extensive genetic heterogeneity  
20 also between members of the same species.

## 21 22 **Materials and methods**

### 23 ***Population***

24 The population-based colonoscopy (PopCol) study has been described in detail elsewhere <sup>17</sup>.  
25 Briefly, a random sample of the adult population in two adjacent parishes in Södermalm,  
26 Stockholm, were sent a validated abdominal symptom questionnaire (the ASQ) <sup>18</sup>. Of 3556  
27 recipients, 2293 responded to the questionnaire and of these, 1643 persons were reached  
28 by phone with an invitation to an interview with a colonoscopist and a subsequent  
29 colonoscopy. A total of 1244 participated in the interview of which 745 also underwent  
30 colonoscopy. The subjects who underwent colonoscopy were similar to the total population

1 enrolled<sup>17</sup>. In 17 of the individuals spirochaetes were observed in Haematoxylin and Eosin-  
2 stained tissue sections, which was subsequently confirmed by immunohistochemistry using  
3 a polyclonal rabbit antiserum<sup>10,14</sup>. The study was approved by the local Ethics committee  
4 (No 394/01, Forskningskommitté Syd), and all participants gave their informed consent.  
5 Irritable bowel syndrome (IBS) and bowel habit subtypes (diarrhoea, mixed diarrhoea and  
6 constipation, and unclassifiable) was defined by applying Rome III criteria<sup>19</sup>, and has been  
7 described in detail in Walker et al. 2015<sup>14</sup>.

8

### 9 ***Collection of biopsies***

10 The colonoscopies were performed by 7 experienced endoscopists. Biopsies were taken at 4  
11 levels in the colon (caecum, transverse, sigmoid and rectum) and in terminal ileum. For more  
12 details of biopsy collection, see Kjellström et al.<sup>17</sup>.

13 For both the isolation of bacteria and the 16S amplicon sequencing, sigmoid colon biopsies  
14 were used. For the microbiota analysis four age- and sex matched controls without  
15 spirochaetosis per CS case were randomly selected from the participants with normal  
16 pathology (both macroscopic and microscopic) from the PopCol study.

17

### 18 ***Isolation and biochemical characterisation of spirochaetes***

19 Biopsies from the 17 individuals with confirmed spirochetosis were homogenised in freezing  
20 media and 50 µl was cultivated anaerobically on selective medium for spirochaete isolation  
21 and subculturing as previously described<sup>10</sup>. Pure growth of spirochaetal bacteria was  
22 confirmed with a phase-contrast microscope. From three of the individuals we could not  
23 isolate any spirochaetes by our methods. From one individual two isolates were selected  
24 since they were displaying different morphology on the plate. *B. aalborgi* type strain 513A<sup>T</sup>  
25 [ATCC 43994 / NCTC 11492]<sup>9</sup>, and the Swedish clinical *B. aalborgi* isolate W1<sup>10</sup>, from a strain  
26 collection at the National Veterinary Institute, Uppsala, were also cultured and used for  
27 comparison. See Table 1 for a summary of the individuals and isolates.

28 The spirochaetes were tested for indole production, their ability to hydrolyse sodium  
29 hippurate, beta-haemolysis capacity, and cellular  $\alpha$ -galactosidase and  $\beta$ -glucosidase activity,

1 as has been described previously<sup>10</sup>

2

### 3 ***DNA extraction and sequencing library preparation of spirochaetes***

4 The spirochaetes were harvested and DNA was extracted from the pellet using the DNeasy  
5 Blood & Tissue kit (Qiagen, Hilden, Germany) with proteinase K treatment at 56 C° for  
6 enhanced lysis.

7 Library preparation was performed using the TruSeq Nano DNA library preparation kit  
8 (Illumina) using the 550 bp insert size protocol and the libraries were subsequently  
9 sequenced on the MiSeq platform, v2 chemistry, 300 bp paired end reads. The average  
10 coverage of the genomes was >300-fold and the average insert size 900 bp. For more  
11 detailed sequencing and assembly statistics, see Table S1.

12

### 13 **Bioinformatics analysis of whole genome sequences**

#### 14 ***De novo genome assembly and annotation***

15 Reads were quality assessed and trimmed using bbdutk (bbmap v. 37.77)<sup>20</sup> and *de novo*  
16 assembled using SPAdes v 3.11.1<sup>21</sup> assembly, using the --careful option and kmer options  
17 21, 33, 55, 77, 99, and 127, and the assemblies were filtered to remove contigs of low  
18 coverage or a length under 500 bp. The draft genome of strain 513A<sup>T</sup> was ordered using the  
19 Mauve<sup>22</sup> Order Contig function to the incomplete draft of *Brachyspira aalborgi* 513A  
20 available at the MetaHIT consortium repository  
21 (<https://www.sanger.ac.uk/resources/downloads/bacteria/metahit/>) and the other draft  
22 genomes were subsequently ordered to the 513A<sup>T</sup> draft. The ordered drafts were annotated  
23 using the *prokka* annotation pipeline v 1.12<sup>23</sup>.

24

#### 25 ***Species Circumscription and Phylogeny***

26 To investigate the species designation of the CS isolates, phylogenetic analysis of 16S rRNA  
27 sequences was performed by comparing the 16S sequences of the human CS strains to the  
28 Ref and RefNR *Brachyspira* 16S sequences retrieved from the SILVA database

1 (<https://www.arb-silva.de/>). The sequences were aligned using muscle<sup>28</sup> and a Maximum  
2 Likelihood tree was constructed using PhyML<sup>29</sup>, both software integrated in Seaview v. 4.5.4  
3<sup>30</sup>.

4 The NADH oxidase (*nox*) gene has often been used for phylogenetic comparison of  
5 spirochaetes<sup>31</sup> and was analysed by combining publically available *Brachyspira nox*  
6 sequences from GenBank (Table S3) and the *nox* sequences from our newly sequenced  
7 strains. The nucleotide sequences were aligned using muscle and a phylogenetic tree was  
8 reconstructed using PhyML as described above.

9

10 To study the genetic variability between the CS strains and hitherto described genomes  
11 within the *Brachyspira* genus listed in Table S1, an average nucleotide identity (ANI) analysis  
12 was performed using MUMmer alignment<sup>32</sup> implemented in the *pyani* software v. 0.2.7  
13 (<https://github.com/widowquinn/pyani>).

14 *B. sp.* CAG-484 ([GCA\\_000431315.1](https://genbank.ncbi.nlm.nih.gov/GenBank/FASTA/000431315.1)), a metagenomic assembly from the metaHIT consortium  
15 was excluded from these analyses since it was highly divergent from the others and seemed  
16 to be misclassified to the *Brachyspira* genus.

17 Digital DNA-DNA hybridization (dDDH) analysis was performed using the Genome-to-  
18 Genome Distance Calculator (GGDC) web tool v. 2.1<sup>33</sup>.

19 To compare the genetic content of the isolates, the Roary pan-genome pipeline<sup>34</sup> was used,  
20 with an identity cut-off of 70% on protein level. This analysis was performed both including  
21 all the above-mentioned *Brachyspira* genome references and then separately for the *B.*  
22 *pilosicoli* genomes, and only including the *B. aalborgi* isolates.

23 A more detailed analysis of strains associated with human colonic spirochaetosis was also  
24 performed by retrieving the 16S sequences described in Mikosza et al<sup>35</sup>, Pettersson et al.<sup>36</sup>  
25 and Westerman et al<sup>37</sup> (Table S2). The sequences were aligned and a PhyML tree was  
26 constructed as described above.

27

28 ***Comparative genomics and functional genomics analysis***

1 To further investigate the functional characteristics of the CS genomes, we further  
2 scrutinized the Roary pan-genome data by classifying the representative sequences of each  
3 core and accessory gene cluster according to KEGG orthology using the BlastKOALA online  
4 tool<sup>38</sup>. We also classified all *B. aalborgi* reference strain 513A<sup>T</sup> genes into Clusters of  
5 Orthologous Groups (COG) categories<sup>39</sup> to facilitate functional interpretation. To further  
6 functionally characterise the genomes we applied the macsyfinder *TXSScan* tool<sup>40</sup> to identify  
7 bacterial secretion systems (T1SS-T6SS, T9SS, flagella, type IV pili and Tad pili) and the  
8 prophage prediction tool ProPhet<sup>41</sup>.

9

## 10 **Colonic microbiota composition in individuals with CS compared to controls**

### 11 ***DNA extraction***

12 The homogenised sigmoid colon biopsies were transferred from their original freezing media  
13 into DNA/RNA Shield lysis buffer prior to extraction with ZymoBIOMICS (Zymo Research  
14 Corp, Irvine, CA).

15 The biopsies were submitted to bead beating with Matrix E (MP Biomedicals, Santa Ana, CA,  
16 USA) in a 96 FastPrep shaker (MP Biomedicals, Santa Ana, CA, USA) for 2-6 minutes (the  
17 beating proceeds until the sample is visually homogeneous) and samples were then spun  
18 down to remove beads from the solution. The supernatants were then incubated in  
19 lysozyme buffer (20 mM Tris-CL, 2 mM sodium-EDTA, lysozyme to 100 g/mL; Sigma, St. Louis,  
20 MO, USA) at 37°C for 45' to 1h at 1000 RPM. Following this, samples were again spun down  
21 and the supernatant transferred to a new plate, to eliminate larger particles, and then  
22 incubated with proteinase K at 55°C at 250 RPM for 30 minutes. Finally, the samples were  
23 cleaned through several washing and magnetic bead pelleting steps according to the  
24 instructions of the manufacturer (Genomic DNA MagPrep kit, Zymo Research Corp, Irvine,  
25 CA, USA) and the DNA eluted from the magnetic beads with 70 µL of Elution Buffer (10 mM  
26 Tris-Cl, pH 8.5; Qiagen, Venlo, Netherlands). The DNA extraction was automated on the  
27 FreedomEVO robot (TECAN Trading AG, Männendorf, Switzerland).

28 A blank negative control and a positive mock control (ZymoBIOMICS Mock Community  
29 Standard, Zymo Research Corp, Irvine, CA, USA) were included in each extraction round.



1

## 2 ***DNA amplification and sequencing***

3 Prior to amplification samples were normalized and a total 170 ng DNA was used to amplify  
4 the V3-V4 region of the 16S rRNA gene using primer pair 341F/805R<sup>42</sup>.

5 For the 1-step PCR procedure, amplification was carried out by a high fidelity proofreading  
6 polymerase for a total of 25 cycles. For amplification of the sequencing libraries, forward  
7 primer 5'-CAAGCAGAAGACGGCATAACGAGAT-N<sub>8</sub>-

8 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC-3' and reverse  
9 primer 5'AATGATACGGCGACCACCGAGATC-N<sub>8</sub>-

10 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3', where N<sub>8</sub>

11 represents an identifying 8-mer (barcode) and the last 21 and 19 bases in each construct are  
12 the sequence specific forward and reverse primers, respectively, were used.

13 Samples were then pooled to equimolar amounts and sequenced in parallel to whole  
14 bacterial genomes on the MiSeq instrument (Illumina Inc., San Diego, CA, USA). All controls  
15 from the extraction phase, as well as a negative (blank) PCR control were also prepared and  
16 sequenced.

17

## 18 ***Sequence correction and taxonomic assignment***

19 The processing of 16S amplicon reads was performed as described previously<sup>43</sup>. Briefly,  
20 Cutadapt<sup>44</sup> was used to eliminate all sequences not containing the amplification primers,  
21 remove the primer sequences, all bases with a Phred score below 15 and all reads with less  
22 than 120 bp left after trimming. The resulting reads were merged with usearch v.9.0.2132<sup>45</sup>  
23 and reads failing to merge or producing merging products shorter than 380 bp, longer than  
24 520 bp or with more than three expected errors were discarded. All unique full-length  
25 samples occurring at a frequency higher than 10<sup>-6</sup> in the dataset were submitted to unoise  
26<sup>46</sup>. All the merged reads were then mapped back to the accepted centroids and assigned to  
27 the OTU with the highest identity, at a minimum of 98%. In case of equally high identity  
28 matches, the most abundant centroid is selected. Taxonomy was assigned using SINA

1 v.1.2.13<sup>47</sup>. Scripts for performing this analysis are available at  
2 [https://github.com/ctmrbio/Amplicon\\_workflows](https://github.com/ctmrbio/Amplicon_workflows).

3

#### 4 ***Statistical analysis***

5 For statistical analyses, all samples with fewer than 5000 total Operational Taxonomic Units  
6 (OTU) counts were discarded, as were OTUs not present in at least 3 samples. The resulting  
7 filtered OTU counts were parsed to group counts on the different taxonomical levels and  
8 visualised using the R software packages ggplot<sup>48</sup> and pheatmap<sup>49</sup>. Shannon and Chao1  
9 alpha diversity statistics and Bray-Curtis similarity were calculated using the vegan R package  
10<sup>50</sup>. Differences in diversity statistics between Spirochetosis cases and controls were tested  
11 using student's t-test.

12

#### 13 ***Verification of 16S primer performance***

14 Due to the low total counts of reads from the biopsy amplicon sequencing assigned to the  
15 *Brachyspira* genus we performed a control experiment to verify the amplification  
16 performance of the 341F/805R primer combination on mixes of bacterial DNA. The mixes  
17 were prepared of 75% ZymoBIOMICS™ Microbial Community DNA Standard and 25% DNA  
18 from five of the different *Brachyspira* isolates, selected from different clades of the 16S rRNA  
19 phylogenetic tree (Figure 2), namely *B. aalborgi* 513A<sup>T</sup>, W1, PC3939II and PC3714II, and *B.*  
20 *pilosicoli* PC5538III. The standard is a defined mixture of 5 Gram-positive and 3 Gram-  
21 negative bacteria plus 2 yeast species with wide GC range (15%-85%) for evaluation and  
22 optimisation of microbiomics workflows. A pure ZymoBIOMICS™ Microbial Community DNA  
23 standard sample and a blank sample were also included as positive and negative PCR  
24 control.

25

#### 26 ***Analysis of Human Microbiome Project data***

27 To search for the presence of spirochaetes in the publically available microbiota data from  
28 the Human Microbiome Project (HMP) we downloaded 16S amplicon data from 324 stool

1 samples and metagenomic assemblies from 179 stool samples (Table S5). The 16S samples  
2 were analysed with the methods described above while the assembled shotgun  
3 metagenome data was screened for the presence of spirochetes using MASH<sup>51</sup>. Briefly, the  
4 colonic spirochetosis *Brachyspira aalborgi* and *pilosicoli* genomes previously described in this  
5 study were combined and a MASH sketch was created from these, to which all the  
6 downloaded assemblies were compared. The assemblies showing signs to contain  
7 spirochaete sequences were aligned to the combined genomes using blastn<sup>52</sup> and the  
8 contigs having hits with an e-value of lower than  $e^{-9}$  were extracted for further investigation.

9

## 10 **Results**

### 11 ***Isolation of Spirochaetes***

12 In the present study, spirochaetes were successfully isolated from frozen biopsies from 14  
13 out of the 17 individuals (Table 1). All three individuals from whom we were not able to  
14 retrieve viable bacteria were in the IBS group. From one of the individuals, colonies of  
15 different morphologies were identified and two isolates were propagated, PC5587-p and  
16 PC5587-u. In addition to the clinical isolates, *B. aalborgi* type strain 513A<sup>T</sup> [ATCC  
17 43994/NCTC 11492]<sup>9</sup>, and the Swedish *B. aalborgi* reference isolate W1<sup>10</sup> were sequenced.

18

### 19 ***Morphological and biochemical characterization of the isolates***

20 Originally identified in Haematoxylin and Eosin-stained sections and confirmed with  
21 immunohistochemistry we could observe the spirochaetes to completely cover the colonic  
22 epithelial membranes in all individuals. This could also be visualized using Warthin-Starry  
23 silver stain and representative pictures of the histological manifestation are shown in Figure  
24 1. The spirochetes were present at all the colon levels studied except for in rectal biopsies  
25 where we could not detect the spirochaetes in 4 out of 17 individuals. In the tissue sections  
26 the bacteria were of similar morphology in all individuals with medium-long spiral shaped  
27 bacteria seemingly adherent to the colon mucosa. In culture, the growth on the fastidious  
28 anaerobic agar plates was mostly weak with small colonies and occasional swarming. The  
29 isolates, in contrast to what was observed in the tissue sections, showed a considerable

1 variation in phenotype, both in size, growth rate, cell size, and biochemical reactions (Table  
2 2).

3 Out of the 14 spirochaetal isolates that were investigated, one isolate shared reaction  
4 pattern with the reference strain W1 and five isolates shared reaction patterns with the type  
5 strain 513A<sup>T</sup> (Table 2). Eight strains showed different reaction patterns including four indole  
6 positive isolates. Two isolates, including the reference strain W1 had a positive but weak  
7 hippurate hydrolysis. When compared to the results from the phylogenetic analysis there  
8 was no obvious correlation.

9

### 10 **Genome assembly and phylogenetic classification**

11 Whole genome *de novo* assembly of the 17 isolates revealed that the bacteria had genomes  
12 of approximately the same size, on average 2.67 Mbp with the smallest being 513A<sup>T</sup> with a  
13 genome of 2.50 Mbp (Table 3). The PC5538III culture turned out to contain two isolates that  
14 could be separated in the assembly based on coverage over the contigs. These were termed  
15 PC5538III-hc and PC5538III-lc for high- and low coverage, respectively. The GC content  
16 varied between 27.6% in PC5538III-hc to 28.3%. Most of the genomes assembled nicely with  
17 an average of 31 (8-113) contigs per genome. For more detailed assembly statistics, see  
18 Table S4.

19

20 To identify the species of the isolates, ribosomal RNA sequences were extracted and  
21 searched against the SILVA rRNA database, identifying PC5538III-hc to be *Brachyspira*  
22 *pilosicoli* and all other isolates to belong to *Brachyspira aalborgi* (Figure 2A). Another  
23 approach commonly used for the phylogenetic comparison between *Brachyspira* strains is to  
24 use the NADPH oxidase (*nox*) gene<sup>31</sup>. We therefore combined the publicly available  
25 *Brachyspira nox* sequences (Table S3) with the *nox* sequences from our strains. The  
26 phylogenetic tree can be seen in Figure 2B and confirms the classification of all but one of  
27 the strains to *B. aalborgi*, grouping together with the previously sequenced *nox* gene of *B.*  
28 *aalborgi* ATCC 43994 [513A<sup>T</sup>] and the last, PC5538III-hc to *B. pilosicoli*.

29

1 To investigate the diversity between the *Brachyspira* isolates we also compared the  
2 genomes to publicly available whole genome sequences within the *Brachyspira* genus, both  
3 using average nucleotide identity (Fig. S1A, Table S6, digital DNA-DNA hybridization (dDDH)  
4 (Fig. S1B), and based on gene content in a pan-genome analysis (Fig. S1C), confirming the  
5 species identification. Also, we could confirm the previously unclassified *Brachyspira* sp.  
6 CAG-700, derived from a human gut metagenome<sup>53</sup> to belong to the *B. aalborgi* species and  
7 it was therefore included in further comparative analyses of the species.

8

9 Since there are no previous studies reporting whole-genome sequences of isolates from  
10 colonic spirochaetosis we made a comparison on 16S rRNA level with published rRNA  
11 sequences from three previous studies on human intestinal spirochaetosis; Pettersson et al.  
12 2000<sup>36</sup>, Mikosza et al. 2004<sup>35</sup> and Westerman et al. 2012<sup>37</sup>, (Table S2). The CS isolates were  
13 shown to belong to different clusters of the *B. aalborgi* tree (Fig. 3). None of the isolates fell  
14 within the clade proposed to be called *Brachyspira hominis*<sup>37</sup>, while isolate PC3714II  
15 grouped away from the others, forming its own branch between *B. aalborgi* and *B. pilosicoli*.  
16 Also, the analysis showed that all our isolates clustered with previously described cluster 1  
17 isolates<sup>35,36</sup>.

18

### 19 ***The Brachyspira aalborgi* species is highly divergent**

20 Closer investigation and comparative genomics of the 15 *B. aalborgi* CS genomes show a  
21 relatively high heterogeneity between isolates. The average nucleotide identity (ANI) ranged  
22 between 97.07% and 99.93% (average 97.49%), with an average of 86.09% (76.4 - 100%)  
23 aligned percentage, see Fig. S1A, Table S6. This could be compared with the divergence  
24 between the 4 sequences available from *B. pilosicoli* isolates (98.14 - 98.82 %, average 98.32  
25 %) and for the 24 publicly available whole-genome sequenced *B. hyodysenteriae* (98.92 -  
26 99.96 %, average 99.13 %). Also, the dDDH values between the *B. aalborgi* strains were  
27 sometimes below the 70% percent that is the suggested threshold for a species, on average  
28 73.6% (68.4 - 99.6%). The PC3714II isolate, which grouped away from the other *B. aalborgi*  
29 in the 16S analysis, did not show any deviant characteristics neither based on the *nox*  
30 analysis, ANI or dDDH.

1

## 2 **Genome content and comparative genomics**

3 Annotation of the new genomes showed a predicted number of coding sequences (CDS)  
4 between 2294 and 2554 per genome (Table 3). Out of these, for on average 49 % (46 - 52 %)  
5 of the predicted CDS, no annotations of blast e-value  $< 1e^{-9}$  was found. To investigate the  
6 difference in functional potential between the isolates we performed pan-genome analysis  
7 using the Roary tool. The core genome size (number of genes present in >95% of the  
8 isolates) among the *B. aalborgi* isolates was 1498 out of the total pan genome of 4691  
9 genes. This also revealed a large heterogeneity between strains with three clear clusters of  
10 strains with different gene content (Figure 4).

11 We also performed a pan-genome analysis including the previously described genomes of  
12 the *Brachyspira* genus. Using an amino acid identity cut-off of 70%, we identified 749 core  
13 genes that were common for over 99% of the genomes, out of a total pan genome of 8504  
14 gene clusters. Out of the accessory genes 2026 were unique for strains of *B. aalborgi*, out of  
15 which 418 were common for all the 18 genomes. In the same comparison 1003 genes were  
16 only found in *B. pilosicoli* strains, out of which 425 were common for all the 5 *pilosicoli*  
17 genomes included in the comparison.

18 Using tools to identify bacterial secretion systems and CRISPR-Cas systems showed all  
19 isolates to have genes encoding for complete flagella systems but no other complete  
20 secretion systems. Using the prophage prediction tool ProPhet revealed no putative  
21 prophage sequences in any of our newly sequenced genomes.

22

## 23 **Host interaction factors**

24 Further scrutiny of the *B. aalborgi* pan genome revealed several genes encoding for sialidase  
25 in the genomes, with a different number of genes in the different strains. Most extreme is  
26 the reference strain 513A<sup>T</sup> where no sialidase gene was annotated, to PC3714II and  
27 PC4226IV that harboured 13 genes encoding for sialidase. Most of the predicted sialidase  
28 genes were annotated as sialidase *nedA* by homology to UniProtKB Q02834, a sialidase from  
29 *Micromonospora viridifaciens*, but also to UniProtKB P15698, a sialidase from

1 *Paeniclostridium sordellii* and sialidase *nanH*, UniProtKB P10481 from *Clostridium perfringens*  
2 (Fig. S2). Interestingly, many of these genes were unique for *B. aalborgi* in the pan-genome  
3 analysis described above.

4 We also identified another gene encoding for a potential mucin-degrading enzyme in the  
5 core genome, namely the protease YdcP, which by COG was classified as a Collagenase-like  
6 protease, PrtC family. This gene was present in all other *Brachyspira* genomes studied, as  
7 well as another noteworthy core gene in *aalborgi*, encoding for TolC, an outer membrane  
8 protein required for the export of virulence proteins and toxic compounds without a  
9 periplasmic intermediate<sup>54</sup>.

10

### 11 ***Spirochaetosis is not detected using standard primers for GI tract microbiota analysis***

12 To verify the presence of *Brachyspira* bacteria in the colonic microbiota using molecular  
13 techniques we used 16S amplicon sequencing where the 16S ribosomal RNA was amplified  
14 from DNA extracted from the sigmoid biopsies. The method and primer combination used  
15 are among the most commonly used to study the human gastrointestinal microbiota and  
16 amplifies the V3-V4 region of the 16S rRNA. Examining the classification of the reads it was  
17 noteworthy that the level of OTUs classified to the *Brachyspira* genus and even to the  
18 Spirochaete phylum was very low (0 - 6.6 % of the reads, average 0.7 %) in the individuals  
19 with CS and completely undetected in the age- and sex matched controls (Fig. S3).

20

21 To investigate the amplification capacity we first looked into the potential for primer  
22 annealing bioinformatically and observed that the reverse primer had a mismatch in the  
23 non-degenerate positions close the 3' end of the primer, which we suspected would lead to  
24 poor amplification efficiency (Fig. S4). To verify this experimentally we made mixes of  
25 genomic DNA from five of the *Brachyspira* isolates at a fixed DNA-to-DNA ratio (1:3) with a  
26 Microbial mock of known composition respectively to evaluate the performance of the  
27 primers. The results gave at hand that 0.05-0.7% of the amplified reads were classified as  
28 *Brachyspira*, much below the expected 25%. The higher number (0.7%) was seen for the *B.*  
29 *pilosicoli* strain PC5538III (Fig. S5).

1

2 Despite the failure of the primers to amplify *Brachyspira* 16S rRNA we evaluated whether CS  
3 status verified by histopathology was associated with alterations in the microbial community  
4 composition and diversity. However, this did not reveal any significant differences between  
5 the groups, as illustrated by the lack of segregation between the groups using principal  
6 component analysis (Fig. S6A), or in terms of Shannon diversity ( $p = 0.07$ ) or Chao1 index ( $p =$   
7  $0.41$ ) (Fig. S6B and S6C).

8

### 9 ***Looking for CS in the Human Microbiome Project data***

10 To look for spirochaetosis in a publicly available dataset from another population-based  
11 microbiota study we downloaded the 16S amplicon data from stool samples from 324  
12 individuals from the Human Microbiome Project (HMP)<sup>55</sup>. These samples had been  
13 amplified using two different primer combinations compared to our dataset, one amplifying  
14 the V1-V3 region of the 16S rRNA and the other the V3-V5 region. We also downloaded  
15 assembled shotgun metagenome sequences from 179 individuals to see if we could find any  
16 signs of *Brachyspira* sequences in that data. We found two shotgun metagenome assemblies  
17 containing *Brachyspira* bacteria and those two samples also had 16S amplicon reads  
18 classified to the *Brachyspira* genus, however the levels were almost undetectable, 0.03 %  
19 (V3-V5) and 0.04 % (V1-V3) *Brachyspira* reads, respectively, (Fig. S7). These assemblies came  
20 from the same individual, albeit from different stool samples. Further investigation of the  
21 primer pairs used for the two amplification protocols used show that, alike 341F/805R, both  
22 these primer combinations are likely to be very inefficient in amplifying *Brachyspira* 16S  
23 rRNA due to primer mismatches (Fig. S4).

24

### 25 **Discussion**

26 Human colonic spirochaetosis (CS) is a condition of striking histological appearance with  
27 bacteria adhering to the colonic epithelium<sup>1,9,56</sup>. Despite this, the clinical impact of CS is not  
28 established and successful culture of spirochaetes from human material is rarely reported.  
29 Nor have any isolates from human colonic spirochaetosis (CS) been characterised at the



1 whole genome level. In this study we report isolates and whole genome sequences from 14  
2 individuals with CS, together with the *Brachyspira aalborgi* type strain 513A<sup>T</sup> [ATCC 43994]  
3 and previously described clinical isolate W1<sup>10</sup>.

4  
5 From 17 individuals with histological CS, isolation of spirochaete colonies from three of them  
6 failed. Interestingly, all three of these individuals were diagnosed with IBS based on Rome III  
7 criteria, and both of the individuals with IBS diarrhoea were among these. In total we only  
8 obtained viable isolates from 3 out of the 6 individuals with CS *and* IBS. Comparing the 16S  
9 rRNA sequences of the sequenced isolates with previously published studies of samples from  
10 human CS, isolates belonging to the cluster 2-4 in the phylogenetic tree are missing in our  
11 material. This could be because none of the individuals in our study carried these kinds of  
12 bacteria but could also be due to that only cluster 1 isolates are cultivable, as assumed by  
13 others<sup>35</sup>. However isolates in cluster 1 also seem to be more common when other methods  
14 are used<sup>57</sup>.

15  
16 Trying to circumscribe the species of the isolates we were limited by the fact that no *B.*  
17 *aalborgi* isolates were hitherto whole genome sequenced, and we therefore also included  
18 the *B. aalborgi* type strain 513A<sup>T</sup> in the sequencing. We then compared the 16S rRNA  
19 sequence of the newly sequenced isolates to publically available *Brachyspira* 16S sequences  
20 from the SILVA database as well as tried to delineate the phylogeny based on the NADH  
21 oxidase, *nox*, gene (Figure 2A and B)<sup>31</sup>. This allowed us to classify 14 of the patient isolates  
22 to be *Brachyspira aalborgi* and one to be a *Brachyspira pilosicoli*, which was further  
23 confirmed with better resolution using ANI and dDDH comparison to publicly available  
24 *Brachyspira* whole-genome sequences (Fig. S1).

25  
26 Using the comparative metrics described above, we could also see that the sequence  
27 diversity within *B. aalborgi* is high compared to the other *Brachyspira* species where several  
28 whole-genome sequences are available, despite that the CS isolates were all from a limited  
29 geographical area consisting of two adjacent parishes in Stockholm, Sweden. The proposed  
30 and generally accepted species boundary for ANI and dDDH values are 95-96 % and 70 %, respectively.

1 respectively <sup>60</sup> and the lowest ANI values within the *aalborgi* genomes were 97.07 %.  
2 However, the lowest dDDH value was 68.4 %, and several strains showed this level of  
3 similarity to the type strain. This means that several isolates are on the borderline of  
4 qualifying as novel species but there is also noteworthy continuum of decreasing similarity  
5 values, and more isolates from human colonic spirochaetosis would be needed to correctly  
6 delineate further differentiation within the *Brachyspira* genus. Also, the biochemical tests  
7 did not show any correlation with the ANI, dDDH or accessory genome-based groupings of  
8 the isolates suggesting that these tests, which have been proven useful in characterization of  
9 several other *Brachyspira* species cannot be used for species designation of *B. aalborgi*  
10 isolates. A high sequence diversity and signs of extensive recombination have also been  
11 described in *B. pilosicoli* both using MLST <sup>61</sup>, and in comparisons between the three  
12 previously sequenced *pilosicoli* genomes <sup>11</sup>, but with the whole-genome sequences available  
13 to this date, this variation seems to be lower than within the *B. aalborgi* species.

14

15 To investigate whether the genomes encoded for proteins that could participate in host-  
16 bacterial interaction and/or explain the capacity of the spirochaetes to penetrate both the  
17 loose and inner, attached mucus layers in the colon, we searched the genomic content of  
18 the CS isolates for genes involved in mucus metabolism. We found an abundance of  
19 sialidases in the *B. aalborgi* genomes, ranging from 2-13 genes encoding for sialidases per  
20 isolate, with the one exception being the type strain 513A<sup>T</sup> that interestingly did not harbour  
21 any sialidase genes. Sialidases are proteins that hydrolyse  $\alpha$ 2,3-,  $\alpha$ 2,6-, and  $\alpha$ 2,8- glycosidic  
22 linkages of terminal sialic acid residues in oligosaccharides, glycoproteins and glycolipids <sup>62</sup>.  
23 Sialidated sugars are abundant in mucins and other glycoproteins on the intestinal  
24 epithelium where there is a gradient of increasing terminal sialic acids from the ileum to the  
25 rectum in humans <sup>63</sup>. The purpose of this cleavage can be to release sialic acids for use as  
26 carbon and energy sources, whereas in pathogenic microorganisms, sialidases have been  
27 suggested to be pathogenic factors where the bacteria use the sialic acid to evade the  
28 immune system <sup>64</sup>. No sialidase/sulfatase have been found in *B. hyodysenteriae* but in *B.*  
29 *pilosicoli* and *aalborgi* <sup>65</sup>. Interestingly, the large majority of the sialidase genes identified in  
30 the *B. aalborgi* genomes did not have homologues in the other *Brachyspira* species  
31 according to the pan-genome analysis. This multitude of highly homologous genes also

1 posed problems for the *de novo* assembly and several of the annotated sialidase genes were  
2 truncated due to contig breaks.

3

4 Another potential virulence factor in *Brachyspira aalborgi* is the collagenase PrtC. This  
5 protease was originally described in *Porphyromonas gingivalis* and has also been found in  
6 other human pathogens such as *Helicobacter pylori*, *Salmonella enterica*, *Escherichia coli*,  
7 and *Bacillus subtilis*<sup>66</sup>. Bacterial collagenases have been highlighted in pathogenicity due to  
8 their ability to cleave extracellular matrix components and thereby facilitate colonization  
9 and invasion<sup>67</sup> and PrtC has also been suggested to act as an adhesin to collagenous  
10 structures<sup>68</sup>.

11 Another gene encoding what seems to be a core protein in *Brachyspira* is TolC, located  
12 between the genes encoding for the AcrAB multidrug efflux pump system subunits AcrB and  
13 AcrA (membrane-fusion protein)<sup>69</sup>. The AcrAB–TolC efflux pump, which spans the inner and  
14 outer membranes of the bacterium, is able to transport a broad range of structurally  
15 unrelated small molecules/drugs out of certain Gram-negative bacteria, and this AcrAB-TolC  
16 complex have been shown to confer antibiotic resistance and survival in the gastrointestinal  
17 tract<sup>70</sup> and have also been described in *B. pilosicoli*<sup>11</sup>.

18

19 Detection of spirochaetosis by histology requires invasive methods such as colonoscopy and  
20 collection of biopsied for histology assessment. To see if the spirochaetes could be detected  
21 by standard 16S amplicon studies of the microbiota we performed such an analysis using a  
22 primer combination commonly used in population-based studies of the human microbiota.  
23 To our surprise, the level of reads classified to the *Brachyspira* genus was very low despite  
24 being amplified from a tissue locations adjacent to those where an abundance of  
25 spirochaetes could be observed under the microscope. To identify the reason for this  
26 detection failure we compared the primers used in our study and the HMP study, from  
27 which we also analysed samples, and which is also a large population-based study. The  
28 results are shown in Fig. S4 and verified that none of the three primer combinations can  
29 efficiently amplify *Brachyspira* 16S rRNA, in our study due to a mismatch in the 805R primer.  
30 The HMP studies either used V3-V5 357F + 926R or V1-V3 27F + 534R<sup>71</sup>, for which V1-V3

1 amplification of *Brachyspira* is hampered due to a mismatch in 27F and for the V3-V5  
2 amplification a mismatch in 926R. Considering that spirochaetosis based on observations in  
3 the PopCol study increase the risk for having IBS three-fold<sup>14</sup>, this failure to detect  
4 *Brachyspira* using standard protocols for microbiota studies should be taken seriously since  
5 this potentially important etiological agent in the pathology of IBS will be missed.

6

7 The current study had a number of strengths being population-based with good  
8 representation of the general population<sup>17</sup>. There were also weaknesses. We only obtained  
9 viable isolates from few individuals with IBS and considering this together with the large  
10 genetic heterogeneity between isolates we did not have statistical power to further  
11 investigate associations between bacterial genetic traits and lower gastrointestinal  
12 symptoms, including inflammatory parameters such as eosinophil infiltration. However, the  
13 notion of the poor success rate in culturing IBS isolates could also be important and should  
14 be investigated further. Also, studying more isolates and genomic sequences in the future  
15 from human CS in general would allow us to make more thorough inferences on functional  
16 genomics and pathogenic potential. Albeit looking similar morphologically in the tissue  
17 sections, when culturing the isolates we observed differences in phenotype, both in cell size,  
18 shape and growth rate. Previous studies have shown that there might exist a considerable  
19 complexity in the spirochaete population in each individual<sup>36</sup> but our approach of purifying  
20 and sequencing one isolate per individual did not allow us to investigate this. However, from  
21 one of the individuals we had two morphologically different isolates identified and from one  
22 the genome sequences turned out to consist of two species, suggesting there is more to be  
23 investigated in this aspect.

24

25 In conclusion, we report for the first time the whole genome analysis of a collection of  
26 human clinical isolates from individuals with colonic spirochaetosis where we show that  
27 genomic diversity between isolates is high. From one individual we could retrieve whole-  
28 genome sequences from two distinct species, confirming the observation of mixed  
29 infections. The sequencing of the *B. aalborgi* type strain also allowed us to make genome-  
30 based species circumscription with high resolution compared to previous 16S and *nox*-based

1 investigations. We found several genes in the spirochaetosis genomes that suggest intimate  
2 host-microbe interaction, and that could confer a pathogenic potential to spirochaetes in the  
3 human colon, such as a multitude of genes encoding for mucin-degrading proteins. The  
4 association of these bacteria with for example irritable bowel syndrome and the notion that  
5 they go undetected in major studies of the gut microbiota raise questions as to whether they  
6 represent an underestimated etiological agent in functional bowel disorders, an issue that  
7 should be considered when designing future microbiota analyses.

8

### 9 **Acknowledgements:**

10 KT was supported by the Carl Lenhoff foundation for Spirochaetosis research. AA was  
11 supported by Ruth and Richard Julin Foundation and Nanna Svartz Foundation. The study  
12 was also supported by grants from the Söderberg Foundation and Medical Research  
13 Foundation to LE.

14

### 15 **Author contributions:**

16 KT performed the whole genome sequencing and data analysis, bioinformatics analyses,  
17 interpretation of the results, and wrote the manuscript. LI performed the biochemical  
18 characterisation of the spirochaetes. AB supervised the experimental work with the bacteria.  
19 ÅÖ and MMW performed the histopathology. LE, LA and NJT designed the PopCol study, and  
20 LA was principal investigator and responsible for the management of the study. LK led the  
21 data collection. AA characterised cases based on anamnesis and questionnaire data and  
22 participated in drafting the manuscript. LE was responsible for the management of the  
23 culturing and sequencing part of the study.

24 All authors critically revised the manuscript for intellectual content and approved the final  
25 version of the manuscript including the author list.

26

### 27 **Data availability:**

28 The genomes are submitted to GenBank under BioProject PRJNA513011.

## 1 References

- 2 1 Harland, W. A. & Lee, F. D. Intestinal spirochaetosis. *British medical journal* **3**, 718-719  
3 (1967).
- 4 2 Hampson, D. J. The Spirochete *Brachyspira pilosicoli*, Enteric Pathogen of Animals and  
5 Humans. *Clinical microbiology reviews* **31**, doi:10.1128/CMR.00087-17 (2018).
- 6 3 Duhamel, G. E. Comparative pathology and pathogenesis of naturally acquired and  
7 experimentally induced colonic spirochetosis. *Animal health research reviews* **2**, 3-17 (2001).
- 8 4 Rodgers, F. G., Rodgers, C., Shelton, A. P. & Hawkey, C. J. Proposed pathogenic mechanism  
9 for the diarrhea associated with human intestinal spirochetes. *Am J Clin Pathol* **86**, 679-682  
10 (1986).
- 11 5 Trott, D. J. *et al.* Identification and characterization of *Serpulina pilosicoli* isolates recovered  
12 from the blood of critically ill patients. *J Clin Microbiol* **35**, 482-485 (1997).
- 13 6 Helbling, R., Osterheld, M. C., Vaudaux, B., Jaton, K. & Nydegger, A. Intestinal Spirochetosis  
14 mimicking inflammatory bowel disease in children. *BMC Pediatr* **12**, 163, doi:1471-2431-12-  
15 163 [pii]  
16 10.1186/1471-2431-12-163 (2012).
- 17 7 Kanazawa, T., Takezawa, T., Iwamoto, J. & Adachi, Y. Simultaneous isolation of two species,  
18 *Brachyspira pilosicoli* and *Brachyspira aalborgi*, from a patient with ulcerative colitis. *The*  
19 *Journal of veterinary medical science* **80**, 1416-1419, doi:10.1292/jvms.18-0220 (2018).
- 20 8 Mikosza, A. S., La, T., de Boer, W. B. & Hampson, D. J. Comparative prevalences of  
21 *Brachyspira aalborgi* and *Brachyspira (Serpulina) pilosicoli* as etiologic agents of histologically  
22 identified intestinal spirochetosis in Australia. *J Clin Microbiol* **39**, 347-350,  
23 doi:10.1128/JCM.39.1.347-350.2001 (2001).
- 24 9 Hovind-Hougen, K. *et al.* Intestinal spirochetosis: morphological characterization and  
25 cultivation of the spirochete *Brachyspira aalborgi* gen. nov., sp. nov. *Journal of clinical*  
26 *microbiology* **16**, 1127-1136 (1982).
- 27 10 Kraaz, W., Pettersson, B., Thunberg, U., Engstrand, L. & Fellstrom, C. *Brachyspira aalborgi*  
28 infection diagnosed by culture and 16S ribosomal DNA sequencing using human colonic  
29 biopsy specimens. *J Clin Microbiol* **38**, 3555-3560 (2000).
- 30 11 Mappley, L. J. *et al.* Comparative genomics of *Brachyspira pilosicoli* strains: genome  
31 rearrangements, reductions and correlation of genetic complement with phenotypic diversity.  
32 *BMC genomics* **13**, 454, doi:10.1186/1471-2164-13-454 (2012).
- 33 12 Weisheit, B., Bethke, B. & Stolte, M. Human intestinal spirochetosis: analysis of the  
34 symptoms of 209 patients. *Scand J Gastroenterol* **42**, 1422-1427,  
35 doi:10.1080/00365520701245629 (2007).
- 36 13 Tsinganou, E. & Gebbers, J. O. Human intestinal spirochetosis--a review. *Ger Med Sci* **8**,  
37 Doc01, doi:10.3205/000090 (2010).
- 38 14 Walker, M. M. *et al.* Colonic spirochetosis is associated with colonic eosinophilia and irritable  
39 bowel syndrome in a general population in Sweden. *Hum Pathol* **46**, 277-283,  
40 doi:10.1016/j.humpath.2014.10.026 (2015).
- 41 15 Goodsall, T. M. *et al.* Unique pathology of colonic spirochaetosis characterised by mucosal  
42 eosinophilia is linked to diarrhoea and IBS. *Gut* **66**, 978-979, doi:10.1136/gutjnl-2016-312405  
43 (2017).
- 44 16 Ford, A. C., Lacy, B. E. & Talley, N. J. Irritable Bowel Syndrome. *The New England journal of*  
45 *medicine* **376**, 2566-2578, doi:10.1056/NEJMra1607547 (2017).
- 46 17 Kjellstrom, L. *et al.* A randomly selected population sample undergoing colonoscopy:  
47 prevalence of the irritable bowel syndrome and the impact of selection factors. *Eur J*  
48 *Gastroenterol Hepatol* **26**, 268-275, doi:10.1097/MEG.0000000000000024 (2014).
- 49 18 Agreus, L., Svardsudd, K., Nyren, O. & Tibblin, G. Reproducibility and validity of a postal  
50 questionnaire. The abdominal symptom study. *Scand J Prim Health Care* **11**, 252-262 (1993).



- 1 19 Longstreth, G. F. *et al.* Functional bowel disorders. *Gastroenterology* **130**, 1480-1491,  
2 doi:10.1053/j.gastro.2005.11.061 (2006).
- 3 20 Bushnell, B. BBMap short read aligner. *University of California, Berkeley, California*. URL  
4 <http://sourceforge.net/projects/bbmap> (2016).
- 5 21 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-  
6 cell sequencing. *Journal of computational biology : a journal of computational molecular cell*  
7 *biology* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).
- 8 22 Rissman, A. I. *et al.* Reordering contigs of draft genomes using the Mauve aligner.  
9 *Bioinformatics* **25**, 2071-2073, doi:10.1093/bioinformatics/btp356 (2009).
- 10 23 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069,  
11 doi:10.1093/bioinformatics/btu153 (2014).
- 12 24 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site  
13 identification. *BMC Bioinformatics* **11**, 119, doi:1471-2105-11-119 [pii]  
14 10.1186/1471-2105-11-119 (2010).
- 15 25 barrnap v. 0.6 (<https://github.com/tseemann/barrnap>, 2015).
- 16 26 Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in  
17 nucleotide sequences. *Nucleic acids research* **32**, 11-16, doi:10.1093/nar/gkh152 (2004).
- 18 27 Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal  
19 peptides from transmembrane regions. *Nature methods* **8**, 785-786,  
20 doi:10.1038/nmeth.1701 (2011).
- 21 28 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput.  
22 *Nucleic acids research* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).
- 23 29 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood  
24 phylogenies: assessing the performance of PhyML 3.0. *Systematic biology* **59**, 307-321,  
25 doi:10.1093/sysbio/syq010 (2010).
- 26 30 Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user  
27 interface for sequence alignment and phylogenetic tree building. *Molecular biology and*  
28 *evolution* **27**, 221-224, doi:10.1093/molbev/msp259 (2010).
- 29 31 Fellstrom, C., Rasback, T., Johansson, K. E., Olofsson, T. & Aspan, A. Identification and genetic  
30 fingerprinting of *Brachyspira* species. *Journal of microbiological methods* **72**, 133-140,  
31 doi:10.1016/j.mimet.2007.11.015 (2008).
- 32 32 Richter, M. & Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic  
33 species definition. *Proceedings of the National Academy of Sciences of the United States of*  
34 *America* **106**, 19126-19131, doi:10.1073/pnas.0906412106 (2009).
- 35 33 Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. P. & Goker, M. Genome sequence-based species  
36 delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*  
37 **14**, 60, doi:10.1186/1471-2105-14-60 (2013).
- 38 34 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**,  
39 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).
- 40 35 Mikosza, A. S., Munshi, M. A. & Hampson, D. J. Analysis of genetic variation in *Brachyspira*  
41 *aalborgi* and related spirochaetes determined by partial sequencing of the 16S rRNA and  
42 NADH oxidase genes. *Journal of medical microbiology* **53**, 333-339,  
43 doi:10.1099/jmm.0.05430-0 (2004).
- 44 36 Pettersson, B. *et al.* Phylogenetic evidence for novel and genetically different intestinal  
45 spirochetes resembling *Brachyspira aalborgi* in the mucosa of the human colon as revealed  
46 by 16S rDNA analysis. *Systematic and applied microbiology* **23**, 355-363, doi:10.1016/S0723-  
47 2020(00)80065-X (2000).
- 48 37 Westerman, L. J. *et al.* Development of a real-time PCR for identification of *brachyspira*  
49 species in human colonic biopsies. *PloS one* **7**, e52281, doi:10.1371/journal.pone.0052281  
50 (2012).

- 1 38 Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG Tools for  
2 Functional Characterization of Genome and Metagenome Sequences. *Journal of molecular*  
3 *biology* **428**, 726-731, doi:10.1016/j.jmb.2015.11.006 (2016).
- 4 39 Tatusov, R. L., Galperin, M. Y., Natale, D. A. & Koonin, E. V. The COG database: a tool for  
5 genome-scale analysis of protein functions and evolution. *Nucleic acids research* **28**, 33-36  
6 (2000).
- 7 40 Abby, S. S. & Rocha, E. P. C. Identification of Protein Secretion Systems in Bacterial Genomes  
8 Using MacSyFinder. *Methods in molecular biology* **1615**, 1-21, doi:10.1007/978-1-4939-7033-  
9 9\_1 (2017).
- 10 41 Reis-Cunha, J. L., Bartholomeu, D. C., Earl, A. M., Birren, B. W. & Cerqueira, G. C. ProphET,  
11 Prophage Estimation Tool: a standalone prophage sequence prediction tool with self-  
12 updating reference database. *bioRxiv* (2017).
- 13 42 Herlemann, D. P. *et al.* Transitions in bacterial communities along the 2000 km salinity  
14 gradient of the Baltic Sea. *The ISME journal* **5**, 1571-1579, doi:10.1038/ismej.2011.41 (2011).
- 15 43 Hugerth, L. W. *et al.* A comprehensive automated pipeline for human microbiome sampling,  
16 16S rRNA gene sequencing and bioinformatics processing. *bioRxiv*, doi:10.1101/286526  
17 (2018).
- 18 44 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.  
19 *2011* **17**, 3, doi:10.14806/ej.17.1.200 (2011).
- 20 45 Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**,  
21 2460-2461, doi:10.1093/bioinformatics/btq461 (2010).
- 22 46 Edgar, R. C. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon  
23 sequencing. *bioRxiv*, doi:10.1101/081257 (2016).
- 24 47 Pruesse, E., Peplies, J. & Glockner, F. O. SINA: accurate high-throughput multiple sequence  
25 alignment of ribosomal RNA genes. *Bioinformatics* **28**, 1823-1829,  
26 doi:10.1093/bioinformatics/bts252 (2012).
- 27 48 Wickham, H. *Ggplot2 : elegant graphics for data analysis.* (Springer, 2009).
- 28 49 pheatmap: Pretty Heatmaps v. 1.0.8 (2012).
- 29 50 vegan: Community Ecology Package v. 2.4-4 (2017).
- 30 51 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash.  
31 *Genome biology* **17**, 132, doi:10.1186/s13059-016-0997-x (2016).
- 32 52 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421,  
33 doi:10.1186/1471-2105-10-421 (2009).
- 34 53 Nielsen, H. B. *et al.* Identification and assembly of genomes and genetic elements in complex  
35 metagenomic samples without using reference genomes. *Nature biotechnology* **32**, 822-828,  
36 doi:10.1038/nbt.2939 (2014).
- 37 54 Zgurskaya, H. I., Krishnamoorthy, G., Ntrel, A. & Lu, S. Mechanism and Function of the Outer  
38 Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria. *Frontiers*  
39 *in microbiology* **2**, 189, doi:10.3389/fmicb.2011.00189 (2011).
- 40 55 Human Microbiome Project, C. Structure, function and diversity of the healthy human  
41 microbiome. *Nature* **486**, 207-214, doi:10.1038/nature11234 (2012).
- 42 56 Lee, F. D. *et al.* Intestinal spirochaetosis. *Gut* **12**, 126-133 (1971).
- 43 57 Rojas, P. *et al.* Distribution and phylogeny of Brachyspira spp. in human intestinal  
44 spirochetosis revealed by FISH and 16S rRNA-gene analysis. *Anaerobe* **47**, 25-32,  
45 doi:10.1016/j.anaerobe.2017.03.012 (2017).
- 46 58 Fellström, C., Pettersson, B., Uhlén, M., Gunnarsson, A. & Johansson, K. E. Phylogeny of  
47 Serpulina based on sequence analyses of the 16S rRNA gene and comparison with a scheme  
48 involving biochemical classification. *Res Vet Sci* **59**, 5-9 (1995).
- 49 59 Fossil, M. *et al.* Neither hippurate-negative Brachyspira pilosicoli nor Brachyspira pilosicoli  
50 type strain caused diarrhoea in early-weaned pigs by experimental infection. *Acta veterinaria*  
51 *Scandinavica* **46**, 257-267 (2005).



- 1 60 Chun, J. *et al.* Proposed minimal standards for the use of genome data for the taxonomy of  
2 prokaryotes. *International journal of systematic and evolutionary microbiology* **68**, 461-466,  
3 doi:10.1099/ijsem.0.002516 (2018).
- 4 61 Trott, D. J., Mikosza, A. S., Combs, B. G., Oxberry, S. L. & Hampson, D. J. Population genetic  
5 analysis of *Serpulina pilosicoli* and its molecular epidemiology in villages in the eastern  
6 Highlands of Papua New Guinea. *International journal of systematic bacteriology* **48 Pt 3**,  
7 659-668, doi:10.1099/00207713-48-3-659 (1998).
- 8 62 Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. Diversity of microbial sialic acid  
9 metabolism. *Microbiology and molecular biology reviews : MMBR* **68**, 132-153 (2004).
- 10 63 Juge, N., Tailford, L. & Owen, C. D. Sialidases from gut bacteria: a mini-review. *Biochemical*  
11 *Society transactions* **44**, 166-175, doi:10.1042/BST20150226 (2016).
- 12 64 Severi, E., Hood, D. W. & Thomas, G. H. Sialic acid utilization by bacterial pathogens.  
13 *Microbiology* **153**, 2817-2822, doi:10.1099/mic.0.2007/009480-0 (2007).
- 14 65 Wanchanthuek, P. *et al.* The complete genome sequence of the pathogenic intestinal  
15 spirochete *Brachyspira pilosicoli* and comparison with other *Brachyspira* genomes. *PloS one*  
16 **5**, e11455, doi:10.1371/journal.pone.0011455 (2010).
- 17 66 Zhang, Y. Z., Ran, L. Y., Li, C. Y. & Chen, X. L. Diversity, Structures, and Collagen-Degrading  
18 Mechanisms of Bacterial Collagenolytic Proteases. *Applied and environmental microbiology*  
19 **81**, 6098-6107, doi:10.1128/AEM.00883-15 (2015).
- 20 67 Duarte, A. S. *et al.* *Aeromonas piscicola* AH-3 expresses an extracellular collagenase with  
21 cytotoxic properties. *Letters in applied microbiology* **60**, 288-297, doi:10.1111/lam.12373  
22 (2015).
- 23 68 Kato, T., Takahashi, N. & Kuramitsu, H. K. Sequence analysis and characterization of the  
24 *Porphyromonas gingivalis* prtC gene, which expresses a novel collagenase activity. *Journal of*  
25 *bacteriology* **174**, 3889-3895 (1992).
- 26 69 Tikhonova, E. B., Yamada, Y. & Zgurskaya, H. I. Sequential mechanism of assembly of  
27 multidrug efflux pump AcrAB-TolC. *Chemistry & biology* **18**, 454-463,  
28 doi:10.1016/j.chembiol.2011.02.011 (2011).
- 29 70 Weston, N., Sharma, P., Ricci, V. & Piddock, L. J. V. Regulation of the AcrAB-TolC efflux pump  
30 in Enterobacteriaceae. *Research in microbiology* **169**, 425-431,  
31 doi:10.1016/j.resmic.2017.10.005 (2018).
- 32 71 Watanabe, K., Kodama, Y. & Harayama, S. Design and evaluation of PCR primers to amplify  
33 bacterial 16S ribosomal DNA fragments used for community fingerprinting. *Journal of*  
34 *microbiological methods* **44**, 253-262 (2001).

35

36

## 1 **Figures legends**

2 **Figure 1.** Histological sections showing, in the same formalin-fixed, paraffin embedded  
3 biopsy, representative pictures of the spirochaetosis using A) Haematoxylin and Eosin  
4 staining, B) Immunohistochemistry using polyclonal spirochaete-specific antibodies, C)  
5 Warthin-Starry Silver staining.

6  
7 **Figure 2.** Phylogenetic trees based on A) the 16S ribosomal RNA sequences, and B) the NADH  
8 oxidase (NOX) protein sequence respectively. The isolates marked in red were sequenced  
9 within this study. The sequences marked in green come from the whole-genome sequences  
10 described in Table S1. The taxa in bold are the type strains for their respective species. In  
11 panel B clades containing only sequences from one species are collapsed for better  
12 readability.

13  
14 **Figure 3.** Maximum likelihood tree of 16S rRNA sequences from human intestinal  
15 spirochaetosis. Isolates in red are sequenced in this study, the ones in blue are from <sup>36</sup>,  
16 isolates in green from <sup>35</sup>, and those in orange from <sup>37</sup>. The phylogenetic clusters are  
17 annotated according to the terminology used in the three papers respectively.

18  
19 **Figure 4.** Heatmap showing hierarchical clustering of the accessory genome in *Brachyspira*  
20 *aalborgi*, i.e. genes present in < 99% of the genomes, together with the results from the  
21 biochemical characterization of isolates. Presence of an orthologous group is shown in blue  
22 and absence in white. Included genomes are the 16 newly sequenced *B. aalborgi* together  
23 with *B. sp.* CAG-700.

24  
25 **Figure S1.** Comparative genomics within the *Brachyspira* genus. A) Average Nucleotide  
26 Identity (ANI) comparison. B) Digital DNA-DNA Hybridisation (dDDH) comparison. C) Tree  
27 based on accessory genome analysis (presence/absence of accessory genes).

28

1 **Figure S2.** Maximum Likelihood tree of the genes annotated to encode for Sialidase in the *B.*  
2 *aalborgi* genomes. Leaves are coloured according to the source of annotation inference and  
3 the table is summarizing the number of sialidase genes per isolate.

4

5 **Figure S3.** 16S amplicon-based microbiota analysis in spirochaetosis cases and unaffected  
6 controls. A) Relative phylum composition in cases and controls. B) Proportion of reads  
7 assigned to the *Brachyspira* genus, cases and controls.

8

9 **Figure S4.** Primer mismatches in the different primer combinations compared to *Brachyspira*  
10 16S rRNA sequences. Magnified areas show WebLogo representation of the *Brachyspira* 16S  
11 alignment in the primer target regions, together with the primers in question. Boxes mark  
12 the bases with mismatches.

13

14 **Figure S5.** Verification of primer failure in samples of known compositions. All samples were  
15 containing a 1:3 proportion of pure *Brachyspira* DNA and a mock community. A) Phylum  
16 composition in the different samples. B) Proportion of reads classified to the *Brachyspira*  
17 genus depending on the *Brachyspira* species spiked in. The left-most column represents the  
18 expected distribution.

19

20 **Figure S6.** 16S amplicon-based microbiota analysis in spirochaetosis cases and unaffected  
21 controls. A) Principle component analysis of OTU composition in cases and controls. B)  
22 Shannon & Chao1 diversity

23

24 **Figure S7.** Proportion of reads assigned to the *Brachyspira* genus in 324 individuals from the  
25 Human Microbiome Project (HMP).

26

**Table 1. Subject and isolate information**

Patient	Isolate	Sex	Age	IBS
1	n.c.	M	37	IBS-D
2	PC2022III	M	58	-
3	PC2777IV	M	45	-
4	PC3053II	F	46	-
5	PC3517II	F	58	-
6	PC3714II	M	60	-
7	PC390II	M	34	-
8	PC3939II	M	66	-
9	PC3997IV	F	31	IBS-M
10	n.c.	F	27	IBS-U
11	PC4226IV	M	67	-
12	PC4580III	F	69	-
13	PC4597II	F	36	-
14	n.c.	M	49	IBS-D
15	PC5099IV	F	55	IBS-U
16	PC5538III	M	44	IBS-M
17	PC5587-p PC5587-u	F	57	-
Type strain	513A <sup>T</sup>	-	-	-
Ref. strain	W1			-

n.c. = not culturable. IBS-D = IBS-diarrhea, IBS-M = IBS mixed, IBS-U = IBS undefined

**Table 2. Biochemical analyses**

		Biochemical test				
	Isolate	Haemolysis	Indole production	Hippurate hydrolysis	$\alpha$ -gal <sup>a</sup>	$\beta$ -glu <sup>b</sup>
Clinical Isolates	PC4226IV	Weak	-	-	-	+
	PC3053II	None	+	-	-	-
	PC4580III	None	-	-	-	-
	PC3997IV	Weak	-	-	-	-
	PC2777IV	Weak	-	-	-	-
	PC3939II	Weak	+	-	-	+
	PC3517II	Weak	+	-	-	+
	PC3714II	None	-	-	-	+
	PC5099IV	None	-	-	-	-
	PC5538III	None	+	-	-	-
	PC4597II	None	-	-	-	(+)
	PC390II	None	-	-	-	(+)
	PC2022II	None	-	(+)	-	-
	PC5587II-p	None	-	-	-	-
Type strain	<i>513A<sup>T</sup></i>	None	-	-	-	-
Ref. strain	W1	None	-	(+)	-	-

<sup>a</sup>  $\alpha$ -gal = alpha-galactosidase activity.

<sup>b</sup>  $\beta$ -glu = beta-glucosidase activity.

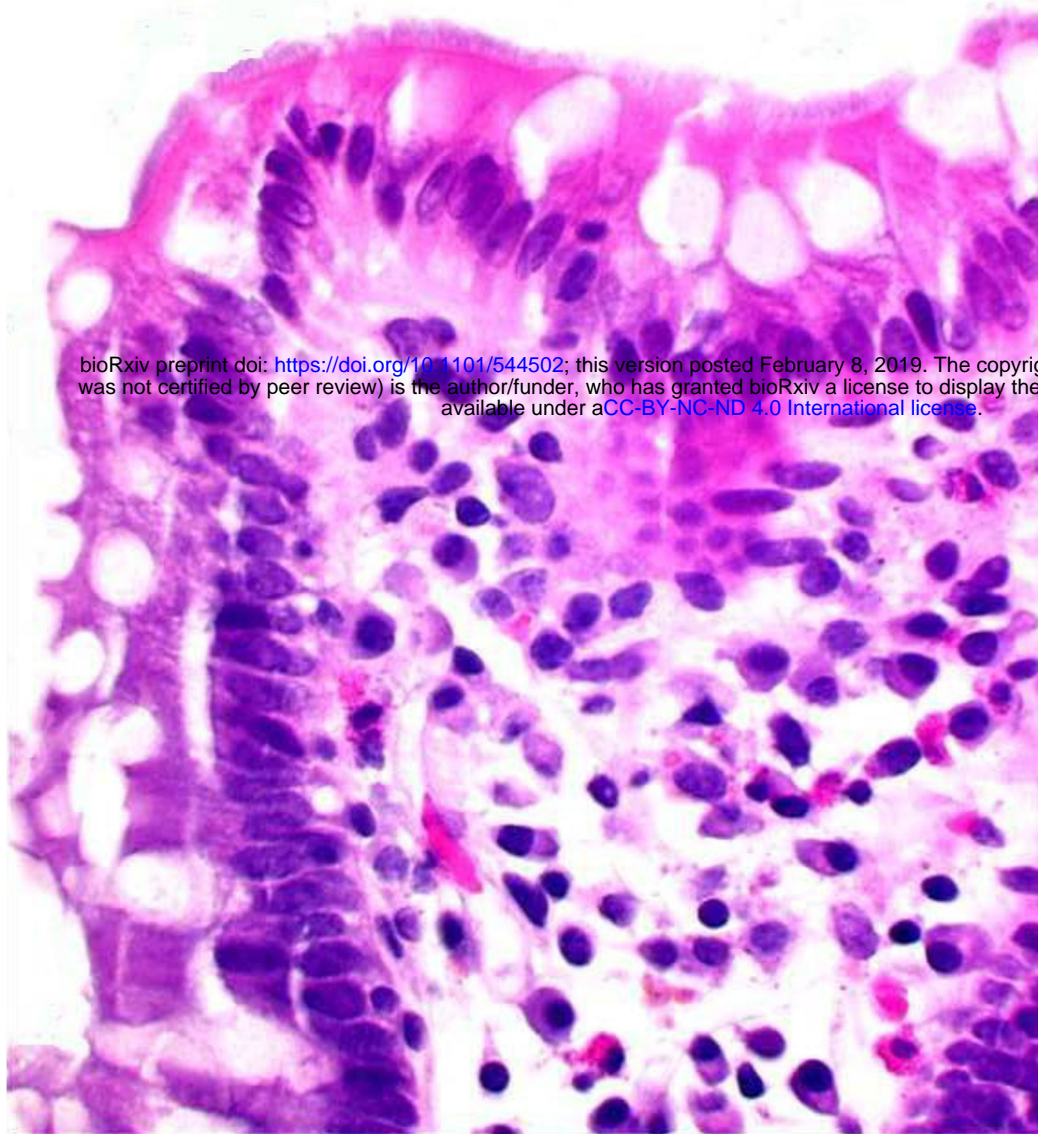
**Table 3. Genome characteristics**

Isolate	Number of contigs	Genome size (bp)	%GC	CDS	rRNA	tRNA	Signal peptides	Repeat regions
<i>Brachyspira aalborgi</i> PC2022III	15	2609344	28.3	2368	3	36	132	1
<i>Brachyspira aalborgi</i> PC2777IV	30	2655229	28.1	2449	3	36	142	2
<i>Brachyspira aalborgi</i> PC3053II	34	2667320	28.2	2465	3	36	136	1
<i>Brachyspira aalborgi</i> PC3517II	35	2645532	28.2	2430	3	36	138	1
<i>Brachyspira aalborgi</i> PC3714II	24	2743156	28.1	2512	3	36	132	1
<i>Brachyspira aalborgi</i> PC390II	32	2660153	28.1	2479	3	35	132	1
<i>Brachyspira aalborgi</i> PC3939II	24	2589393	28.2	2343	3	36	123	2
<i>Brachyspira aalborgi</i> PC3997IV	33	2649511	28.1	2441	3	36	139	1
<i>Brachyspira aalborgi</i> PC4226IV	40	2715122	28.1	2549	3	35	137	5
<i>Brachyspira aalborgi</i> PC4580III	10	2591031	28.3	2361	3	37	129	1
<i>Brachyspira aalborgi</i> PC4597II	28	2670785	28.1	2495	3	35	137	1
<i>Brachyspira aalborgi</i> PC5099IV	21	2714460	28.1	2554	3	35	139	2
<i>Brachyspira aalborgi</i> PC5587-p	33	2665010	28.1	2479	3	35	132	1
<i>Brachyspira aalborgi</i> PC5587-u	32	2662929	28.1	2483	3	35	133	1
<i>Brachyspira pilosicoli</i> PC5538III-hc	113	2755401	27.6	2507	3	34	169	1
<i>Brachyspira aalborgi</i> PC5538III-lc	32	2746864	28.1	2500	3	36	139	1
<i>Brachyspira aalborgi</i> 513A <sup>T</sup>	8	2504147	28.1	2294	3	36	121	1
<i>Brachyspira aalborgi</i> W1	15	2666344	28.2	2413	3	35	128	1

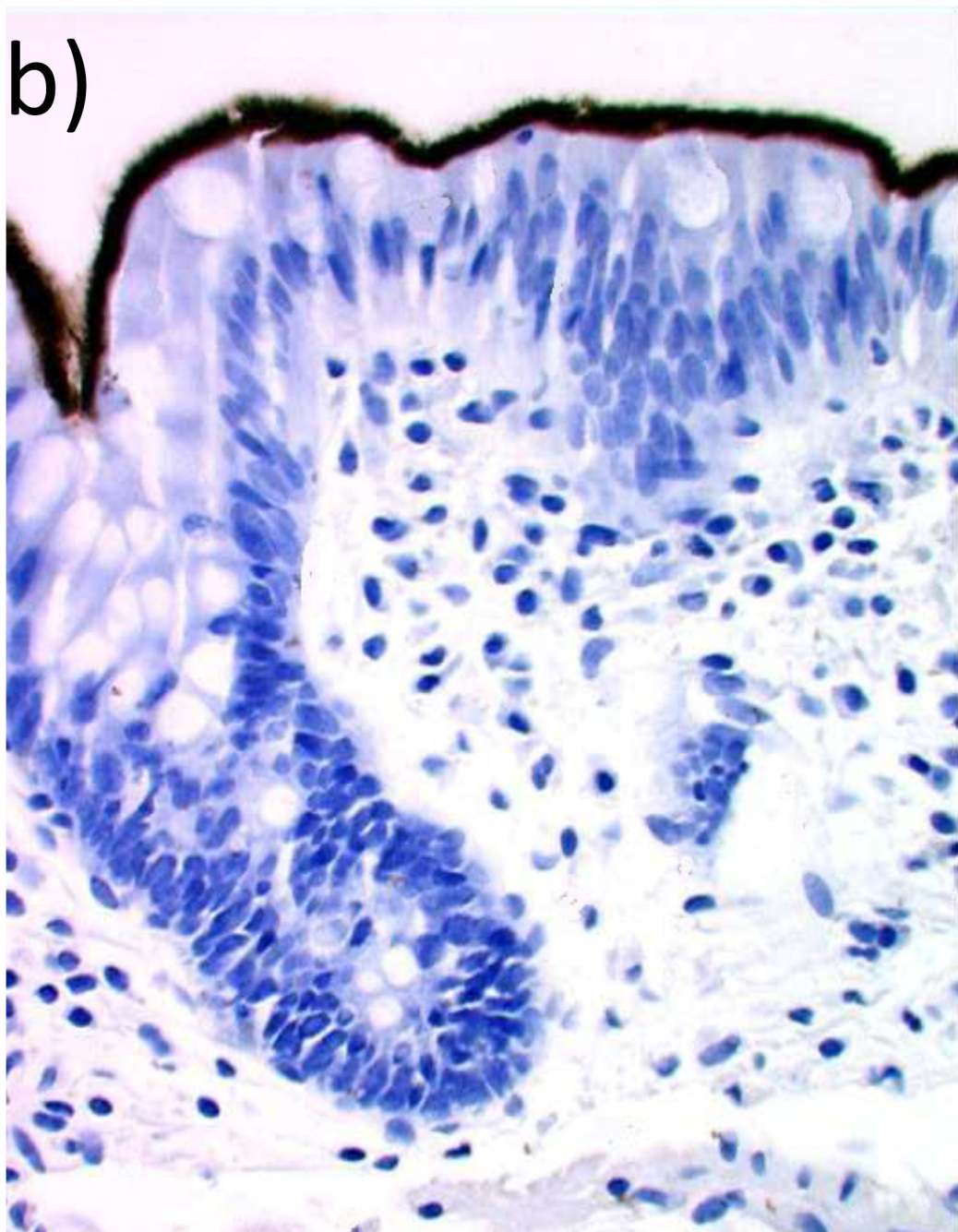
CDS = coding sequences, rRNA = ribosomal RNA, tRNA = transfer RNA



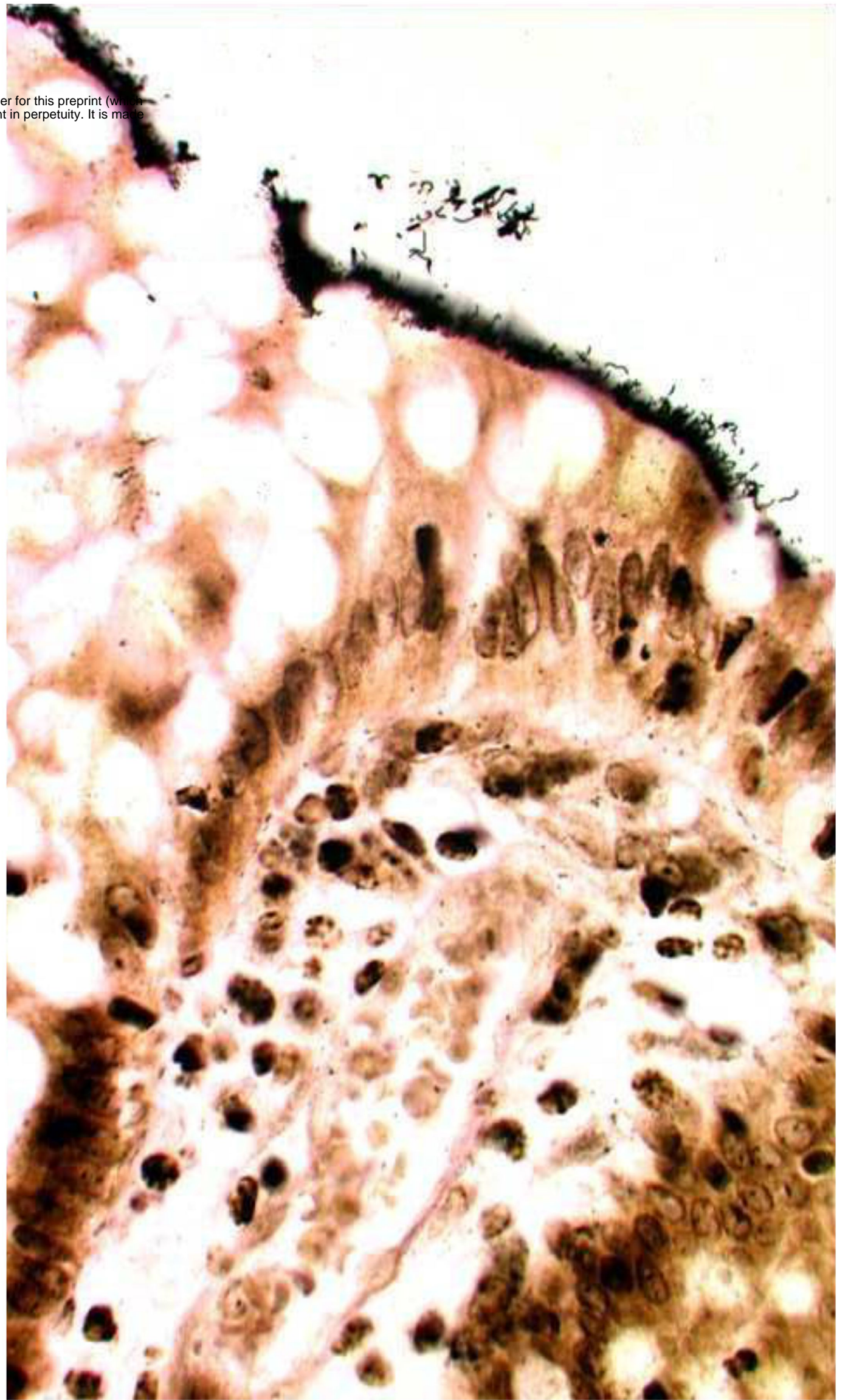
a)



b)



c)









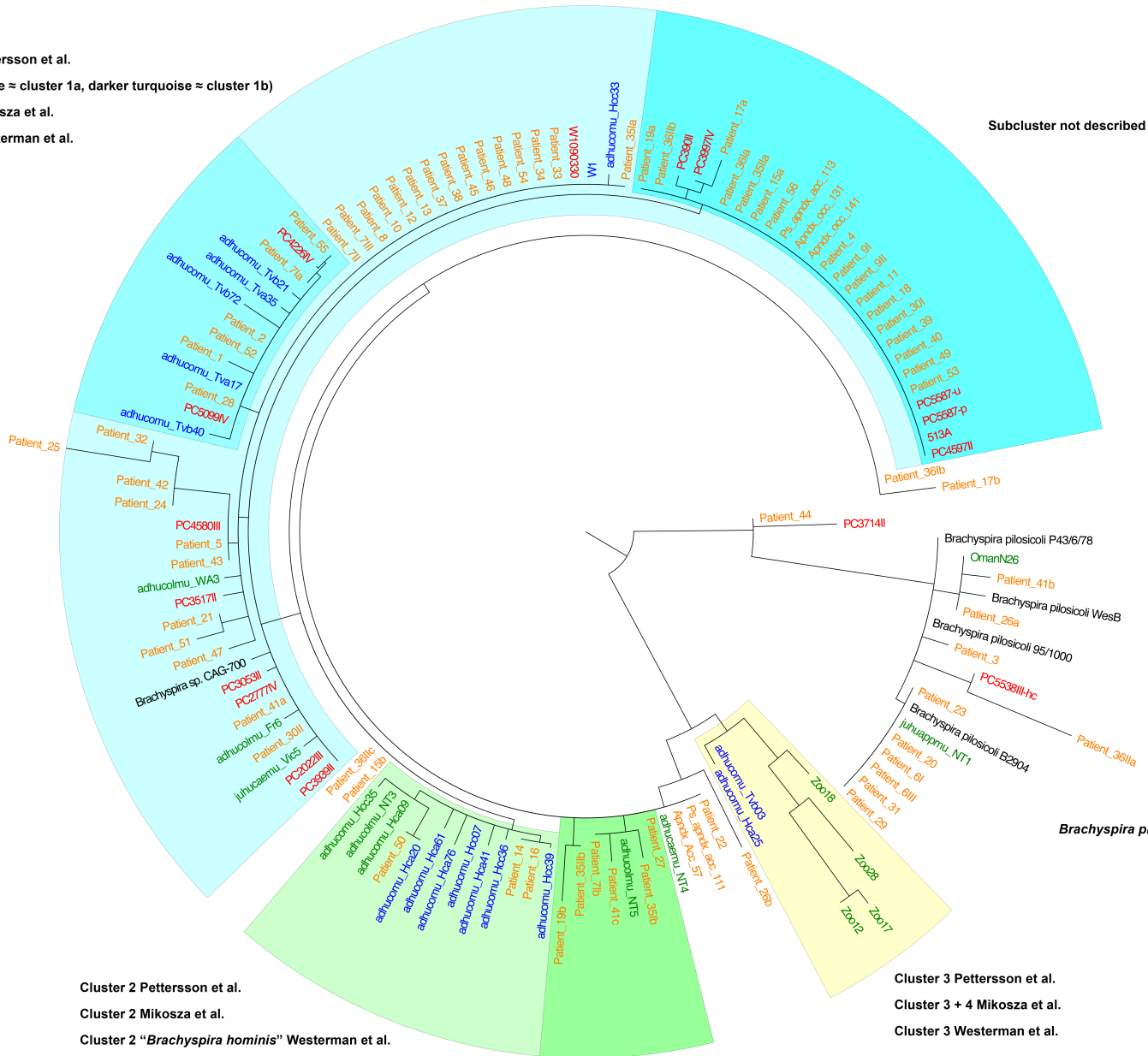
Cluster 1 Pettersson et al.

(light turquoise ≈ cluster 1a, darker turquoise ≈ cluster 1b)

Cluster 1 Mikosza et al.

Cluster 1 Westerman et al.

Subcluster not described previously



Cluster 2 Pettersson et al.

Cluster 2 Mikosza et al.

Cluster 2 "*Brachyspira hominis*" Westerman et al.

Cluster 3 Pettersson et al.

Cluster 3 + 4 Mikosza et al.

Cluster 3 Westerman et al.

N.A. Pettersson et al.

Part of cluster 2 Mikosza et al.

Part of cluster 2 Westerman et al.

Strains from this study

Sequences from Petterson et al. 2000

Sequences from Mikosza et al. 2004

Sequences from Westerman et al. 2012

0.004

## Accessory genome *Brachyspira aalborgi*

