ARTICLE

Check for updates

A *Sarcina* bacterium linked to lethal disease in sanctuary chimpanzees in Sierra Leone

Leah A. Owens [1], Barbara Colitti[2], Ismail Hirji [3], Andrea Pizarro[3], Jenny E. Jaffe [4,5], Sophie Moittié [6,7], Kimberly A. Bishop-Lilly[8], Luis A. Estrella[8], Logan J. Voegtly [8,9], Jens H. Kuhn [10], Garret Suen [11], Courtney L. Deblois[11], Christopher D. Dunn[1], Carles Juan-Sallés[12] & Tony L. Goldberg [1]✉

Human and animal infections with bacteria of the genus *Sarcina* (family *Clostridiaceae*) are associated with gastric dilation and emphysematous gastritis. However, the potential roles of sarcinae as commensals or pathogens remain unclear. Here, we investigate a lethal disease of unknown etiology that affects sanctuary chimpanzees (*Pan troglodytes verus*) in Sierra Leone. The disease, which we have named "epizootic neurologic and gastroenteric syndrome" (ENGS), is characterized by neurologic and gastrointestinal signs and results in death of the animals, even after medical treatment. Using a case-control study design, we show that ENGS is strongly associated with *Sarcina* infection. The microorganism is distinct from *Sarcina ventriculi* and other known members of its genus, based on bacterial morphology and growth characteristics. Whole-genome sequencing confirms this distinction and reveals the presence of genetic features that may account for the unusual virulence of the bacterium. Therefore, we propose that this organism be considered the representative of a new species, named "*Candidatus* Sarcina troglodytae". Our results suggest that a heretofore unrecognized complex of related sarcinae likely exists, some of which may be highly virulent. However, the potential role of "*Ca.* S. troglodytae" in the etiology of ENGS, alone or in combination with other factors, remains a topic for future research.

[1] Department of Pathobiological Sciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI, USA. [2] Department of Veterinary Science, University of Torino, Torino, Italy. [3] Tacugama Chimpanzee Sanctuary, Freetown, Sierra Leone. [4] Tai Chimpanzee Project, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [5] Epidemiology of Highly Pathogenic Microorganisms, Robert Koch Institute, Berlin, Germany. [6] School of Veterinary Medicine and Sciences, University of Nottingham Sutton Bonington Campus, Sutton Bonington, Leicestershire, UK. [7] Twycross Zoo, Atherstone, UK. [8] Genomics and Bioinformatics Department, Biological Defense Research Directorate, Naval Medical Research Center, Fort Detrick, MD, USA. [9] Leidos, Reston, VI, USA. [10] Integrated Research Facility at Fort Detrick, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Fort Detrick, MD, USA. [11] Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA. [12] Noah's Path, Elche, Spain. ✉email: tony.goldberg@wisc.edu

Emerging pathogens pose a substantial risk to animal and human health[1,2]. Pathogens can emerge due to the acquisition of virulence factors through genetic mutation and horizontal gene transfer[3–5] and due to ecological processes that alter their epidemiology/epizootiology and host range[6–8]. These processes are accelerating due to factors such as agricultural intensification[9], demographic shifts[10], ecosystem perturbations[11], geophysical processes[12], and global environmental changes[13]. Furthermore, improved diagnostic methods facilitate detection and characterization of new pathogens and the genetic features that contribute to their emergent phenotypes[14–16].

Emerging pathogens of non-human primates are especially salient examples of this phenomenon because of the high potential for such pathogens to infect humans, who are genetically similar hosts[17,18]. For example, one of the main causative agents of human malaria, *Plasmodium falciparum*, once thought to have co-evolved with humans, actually arose from a recent zoonotic transmission from a western gorilla (*Gorilla gorilla* (Savage, 1847))[19–21]. Nowhere are such risks more evident than in zoological and sanctuary settings, where captive and semi-captive primates come into frequent close contact with people[22,23]. For example, contact with New World primates led to simian foamy virus transmission to primate workers in Brazil[24] and monkeypox virus transmission occurred in staff at a primate sanctuary following a monkeypox outbreak in sanctuary chimpanzees (*Pan troglodytes* (Blumenbach, 1775)) in Cameroon[25].

Since 2005, western chimpanzees (*Pan troglodytes verus* Schwarz, 1934; "chimpanzees" hereafter) in Sierra Leone's Tacugama Chimpanzee Sanctuary (TCS, in Western Area Peninsula National Park) have suffered from a lethal disease of unknown etiology. Characteristic signs are neurologic (weakness, ataxia, seizures) and gastrointestinal (abdominal distension, anorexia, vomiting), resulting in death even after aggressive medical treatment by staff veterinarians. To date, a total of 56 individuals at this facility have died of this condition, which we have named "epizootic neurologic and gastroenteric syndrome" (ENGS), constituting a medical emergency in this population, which averages 93 chimpanzees at any given time. TCS is the largest repository of Sierra Leone's chimpanzee genetic diversity, a training site for conservationists throughout western Africa, an educational/ecotourism destination important for the local economy, and the only home for displaced or orphaned chimpanzees in the country.

Despite ongoing efforts by veterinary staff and international collaborators, the etiology of ENGS has remained elusive. Encephalomyocarditis virus (EMCV; *Picornaviridae*: *Cardiovirus*) infection and toxicity from certain plants (*Dichapetalum toxicarium* (G. Don) Baill./*D. heudelotii* (Planch. ex Oliv.) Baill.) were both suspected but, after investigation, deemed unlikely to be causal. In such circumstances where infection has been suspected but known agents have not been identified, diagnostic approaches based on metagenomics have proven useful[26–28]. We therefore undertook a case-control epizootiological investigation to identify potential pathogens of all major types (viruses, bacteria, and eukaryotes) using metagenomics and traditional methods in the TCS chimpanzees, to detect associations between particular microbes and ENGS.

Here we report the finding of a *Sarcina* genus bacterium in 13 of 19 ENGS cases but no controls. We also report the occurrence of gross and histopathological lesions in affected chimpanzees consistent with the most severe forms of *Sarcina* infection reported in humans and animals. By studying the morphology and growth characteristics of the bacterium, and by sequencing the complete genome of an isolate, we identify features that distinguish it from all previously described members of its genus. In particular, we show that the bacterium possesses genes encoding biochemical pathways potentially contributing to enhanced virulence, including an encoded urea degradation biochemical pathway, consistent with the clinical signs observed in chimpanzees. We conclude that the genus *Sarcina* likely contains an overlooked complex of species ranging from benign commensals to frank pathogens. In light of these findings, the importance of sarcinae in human and animal clinical disease should be re-evaluated.

Results

Epizootiology, clinical signs, and pathology. From 2005 to 2018, 56 resident chimpanzees of TCS died of ENGS. In 32 of 56 cases, affected individuals displayed signs including anorexia, neuromuscular weakness, ataxia, seizures, vomiting, and abdominal distension (Fig. 1; "Clinical signs" group). Signs persisted for a median of 6 days (range: 1–90 days) prior to recovery or death (Supplementary Table 1). In all recovered cases, clinical disease subsequently recurred and resulted in death. In the remaining 24 cases, individuals were discovered post-mortem with no premonitory signs noted by care staff or developed signs which progressed to death in 12 h or less (Fig. 1; "Sudden death" group). Despite these disparate manifestations, all clinical presentations were associated and clearly recognizable as the same "mystery disease" (described as "unmistakable" by veterinarians). We therefore chose the term "syndrome" to reflect the heterogenous nature of the clinical presentations and the suspicion of a common etiopathogenesis. ENGS represented the highest cause of mortality in this population, affecting 33.7% of chimpanzees and accounting for 63.6% of deaths during this time period (Fig. 2a), with a case fatality rate of 100% and a seasonal distribution peaking in March (Fig. 2b). The etiological agent of ENGS does not appear to be transmitted directly, as there were few instances of cases clustering in time and space (Supplementary Table 1).

Frequent lesions included acute shock (congestion involving multiple organs), neutrophilic margination in the microcirculation, moderate to marked gastric dilation, pulmonary edema, acute aspiration of digestive contents, and acute hemorrhage in the thymus, pancreas, or both. In total, post-mortem evaluation documentation was available for 28 chimpanzees, but in only 17 of these was the gastrointestinal tract assessed. Of these 17 patients, 14 had gross evidence of acute gastric dilation, 1 did not show such evidence, and 2 were inconclusive (Fig. 1). One of the chimpanzees with acute gastric dilation also had massive hemorrhagic diathesis (Fig. 3a) and multiple gas-filled lesions in the cecal wall (emphysematous typhlocolitis; Fig. 3b). Microscopically, these gas-filled lesions were surrounded by infiltrates of macrophages, eosinophils, and multinucleate giant cells (Fig. 3c).

Samples. We selected 95 archived samples from 32 chimpanzees for analysis (Supplementary Data 1). These samples comprised 19 individuals (7 males and 12 females) that had died from ENGS (cases), representing a subset of the 56 total cases that occurred since the epizootic began in 2005 (Fig. 2), and 14 healthy individuals (7 males and 7 females) sampled during routine veterinary health checks or, in 2 instances, sampled post-mortem when cause of death was known and clearly unrelated to ENGS (e.g., from trauma; controls). In one instance, matched samples were available from a healthy chimpanzee who subsequently became ill and died from ENGS (hence this individual was first a control and then a case). The chimpanzees in this study ranged in age from 5 to 27 years (median age = 12 years) and were sampled between 14 March 2013 and 11 July 2016, although not all cases within this time period were sampled or available for study. Among ENGS cases included, clinical signs were similar to those
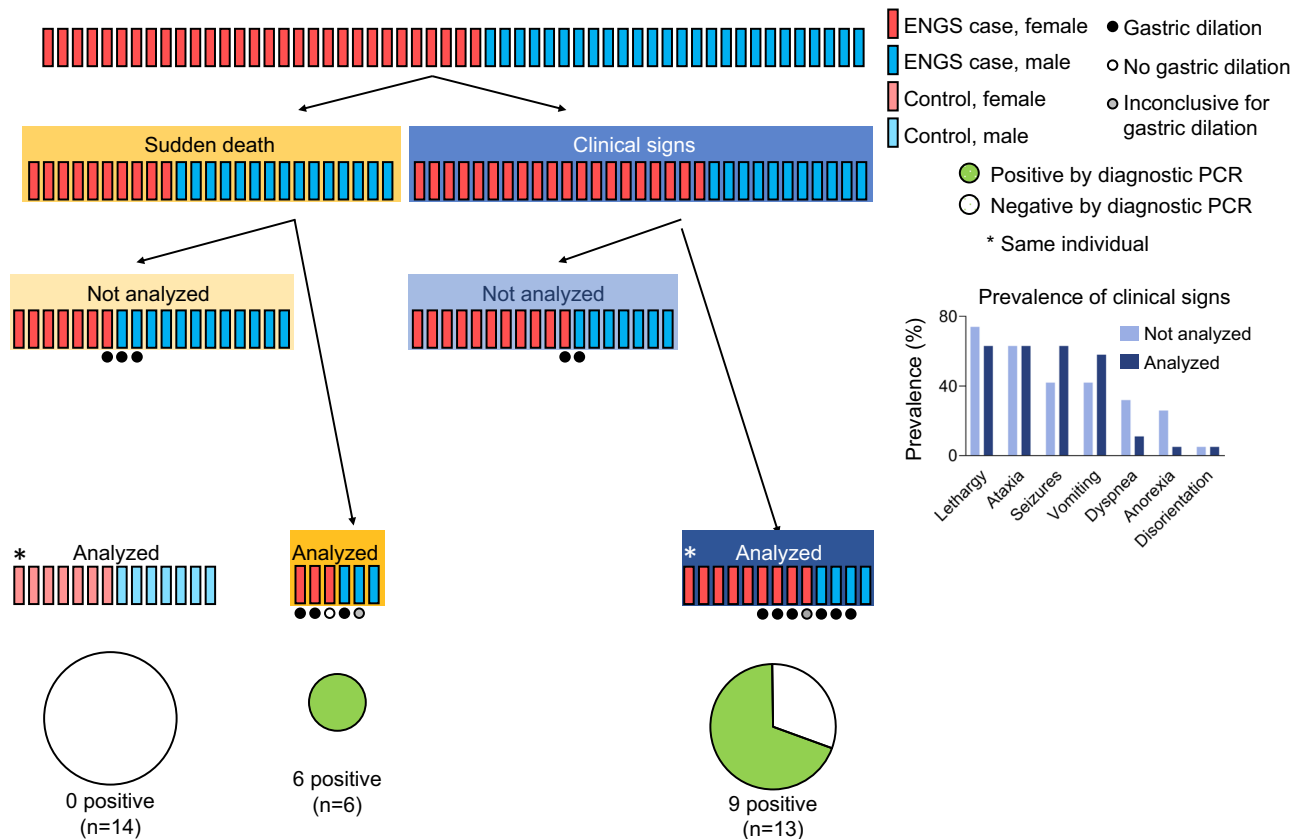
Fig. 1 Graphic representation of epizootic neurologic and gastroenteric syndrome (ENGS) cases and study samples. Individual chimpanzees are represented by rectangles and colors denote sex and case/control status (red, female ENGS case, $n = 30$; blue, male ENGS case, $n = 26$; pink, female ENGS control, $n = 7$; light blue, male ENGS control, $n = 7$). Post-mortem findings regarding gastric dilation are shown for those cases where documentation was available ($n = 17$; black circle, gastric dilation, $n = 14$; open circle, no gastric dilation, $n = 1$; filled gray circle, inconclusive for gastric dilation, $n = 2$). "Analyzed" indicates an individual from whom we obtained at least one sample used in this study ($n = 32$) and the asterisk indicates the single individual from whom we obtained samples pre- and post-ENGS. Pie charts are scaled to sample size with green representing individuals positive by diagnostic PCR and white representing individuals negative by diagnostic PCR.

of cases that were not available for inclusion (Fig. 1), the most common of which were ataxia ($n = 12$), seizures ($n = 12$), vomiting ($n = 10$), and abdominal distention ($n = 9$).

Parasitology. Microscopic examinations of fecal samples were performed on site for 30 chimpanzees (17 ENGS cases and 13 controls) from 2005 to 2018 using standard direct and sedimentation methods[29], comprising 155 analyses with a median 5 analyses per chimpanzee. Nine records indicated no detectable parasites and 146 records indicated ≥1 parasite, with a median parasite richness of 2. Parasites identified included *Entamoeba* spp., *Troglodytella abrassarti*, *Balantidium coli*, *Ascaris* spp., *Enterobius* spp., *Strongyloides* spp., *Trichostrongylus* spp., *Trichuris* spp., *Taenia* spp., *Schistosoma* spp., and unspecified flagellated protozoa and hookworms (Supplementary Fig. 1), all of which are common in this population of chimpanzees and in apparently healthy animals in other captive and wild settings[29–31]. Fortuitously, three ENGS cases had undergone fecal parasitological examinations immediately prior to or just after the time of death, and only representatives of these same typical/commensal organisms were identified: *T. abrassarti*, *Entamoeba hartmanni*, *B. coli*, and *Enterobius* spp.

Eukaryotic metabarcoding for parasite identification using Earth Microbiome Project (EMP) protocols[32] and previously published primers[33,34] for 12 ENGS cases and 6 controls (24 samples; Supplementary Data 2) yielded a total of 2,955,014 raw reads (1,477,507 paired reads) with good overall sequencing quality (~21% of reads removed during quality filtering; Supplementary Table 2). Due to the pan-eukaryotic nature of the primers, and despite the use of a mammal-blocking primer, the majority of reads were identified as host (~83%; Supplementary Table 2). This identification was not surprising due to the sample types (host tissues) and because of previously published similar findings using the same primers and protocols[35]. We processed data from all samples through all filtering steps, after which we excluded those samples that represented <0.5% of the total filtered data set, resulting in removal of 7 of 24 samples (Supplementary Table 2). From the remaining reads (range 681–31,131 per sample) we identified seven operational taxonomic units (OTUs) representing three parasitic and four environmental eukaryotic organisms (Supplementary Fig. 2). No parasites thus identified were case-associated (i.e. found at statistically significantly different prevalence in case versus control groups using a Fisher's exact test (two-tailed); Table 1).

Virology. Metagenomics for virus discovery conducted using previously published methods[36–38] on 12 ENGS cases and 6 controls (24 samples; Supplementary Data 2) generated a total of 151,206,140 reads (mean per sample 6,300,256, standard deviation 3,738,306; average length 161, standard deviation 27). After trimming and filtering on length and quality, 71.3% of sequences remained (mean per sample 4,495,233, standard deviation 2,792,589; average length 113.3, standard deviation 16.3) which were then assembled into 952 contiguous sequences (contigs
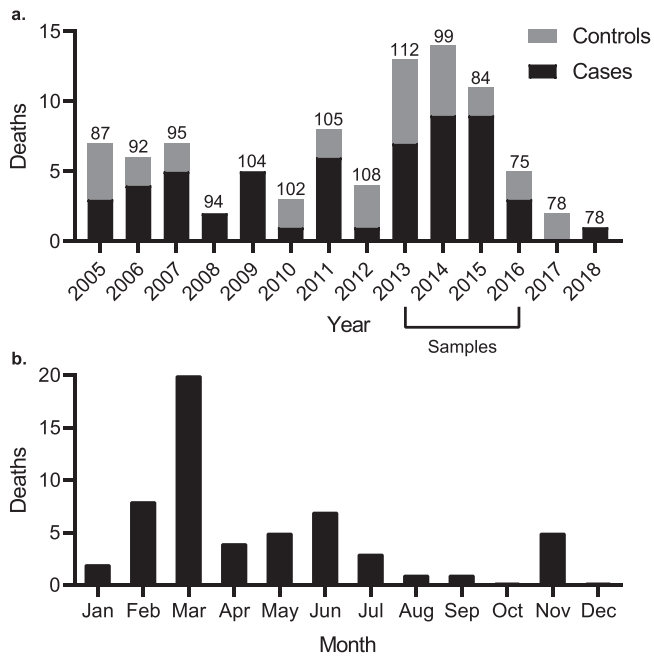
Fig. 2 Deaths attributed to epizootic neurologic and gastroenteric syndrome (ENGS) at Tacugama Chimpanzee Sanctuary from 2005 through 2018. a Annual chimpanzee deaths from ENGS (black, overall total = 56) and other causes (gray, overall total = 32). Numbers above the bars indicate the yearly total chimpanzee population and the horizontal bracket below the x-axis denotes the period during which we obtained samples for analysis. **b** Summed totals of ENGS fatalities by month over 2005–2018 (n = 56). Source data are provided as a Source Data file.

hereafter; mean per sample 39.7, standard deviation 45.5; average length per contig 969.5, standard deviation 690.8) at an average sequence depth of 63.4 (standard deviation 172.4, minimum 3.5, maximum 851). Overall, 21.0% of reads assembled into contigs and 79.0% did not. Analyses of sequence data at the individual read level confirmed the results of the analysis of contigs (i.e. identified the same viruses) and did not identify any additional viruses. Eleven viruses were thus identified, each of which was identical to or closely related to a known virus (Supplementary Table 3), and no viruses were case-associated (Table 1). One control animal was infected with a rhinovirus C (*Picornaviridae*: *Enterovirus*) subsequent to an outbreak of respiratory illness in the population. This pathogen was previously documented as a cause of epizootic respiratory disease in wild chimpanzees[37], but the clinical features of this rhinovirus C infection (characterized by upper respiratory signs) are not consistent with ENGS.

Bacterial metabarcoding. PCR amplification of the 16S rDNA V4 region was attempted on all 96 samples (19 cases and 14 controls; Supplementary Data 1) which yielded amplicons in 10 of 19 ENGS cases (35 total samples) but none of the controls (0 samples). In total, 1,131,561 raw sequences were generated for all 35 samples of which 787,263 were high quality after filtering in mothur. Good's coverage estimation was calculated for all samples and only those samples that had a value >0.99 were retained for downstream analysis. As a result, 23 samples were considered for this analysis (Supplementary Data 2), which totaled 774,339 high-quality sequences with an average of 33,667 ± 17,122 standard deviation per sample. These sequences were binned at 97% similarity into 2592 OTUs. The OTU counts were normalized to 5900 sequences per sample, and these normalized sequences were used for all further analyses.
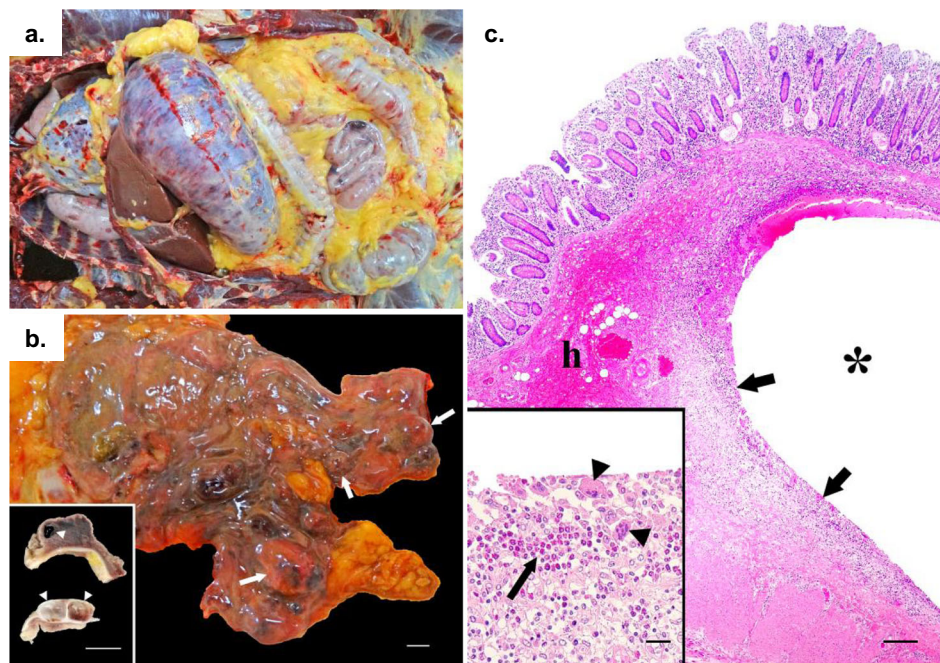

Fig. 3 Gross and histopathologic images of a chimpanzee that died of epizootic neurologic and gastroenteric syndrome (ENGS). Photographic images from an adult male chimpanzee who died of ENGS showing moderate to severe gastric dilation (**a**), hemorrhagic diathesis (**b**), and emphysematous typhlocolitis (**c**; arrows point to gas-filled pockets within the cecum wall, which has reddened areas [in inset, white arrowheads point to gas bubbles in the cut surfaces of the formalin-fixed cecum]; scale bars = 1 cm). On histology (**c**), gas-filled space (*) in the cecal submucosa, surrounded by inflammatory infiltrates (arrows) and hemorrhage (h) were visualized by hematoxylin and eosin staining (inset depicts inflammatory infiltrates, which include eosinophils [arrow] and multinucleate giant cells [arrowheads]; scale bars = 300 μm [main image] and 30 μm [inset]; micrograph is a representative image of three similar sections obtained from the same individual).

Table 1 Case/Control statistical summary.

Organism ID	Organism type	Diagnostic	Prevalence in cases (%)	Prevalence in controls (%)	P[a]	Odds ratio	95% CI low	95% CI high
Blastocystis	Parasite	18S MB	50.00	33.33	0.633	2.000	0.244	16.363
Oesophagostomum	Parasite	18S MB	0.00	16.67	0.375	0.175	0.006	5.041
Troglodytella	Parasite	18S MB	30.00	33.33	>0.999	0.857	0.098	7.510
GB virus C	Virus	Virus Seq	70.59	100.00	0.273	0.175	0.008	3.678
Rhinovirus C	Virus	Virus Seq	0.00	16.67	0.261	0.105	0.004	2.959
Gemykibivirus 2	Virus	Virus Seq	41.18	83.33	0.155	0.140	0.013	1.474
Chimpanzee parvovirus	Virus	Virus Seq	5.88	0.00	>0.999	1.182	0.042	32.915
Human picobirnavirus 4	Virus	Virus Seq	23.53	0.00	0.309	4.333	0.202	93.159
Macaque picobirnavirus 24	Virus	Virus Seq	11.76	16.67	>0.999	0.667	0.049	9.022
Chimpanzee anellovirus	Virus	Virus Seq	11.76	50.00	0.089	0.133	0.015	1.176
Torque teno virus 4	Virus	Virus Seq	23.53	16.67	>0.999	1.538	0.137	17.335
Torque teno virus 23	Virus	Virus Seq	35.29	33.33	>0.999	1.091	0.153	7.802
Torque teno virus 14	Virus	Virus Seq	23.53	33.33	>0.999	0.615	0.081	4.704
Torque teno virus 16	Virus	Virus Seq	17.65	33.33	0.576	0.429	0.052	3.522
"Ca. S. troglodytae"	Bacterium	PCR	68.42	0.00	0.0001	56.077	2.866	1097.182

Virus Seq Virome shotgun sequencing, *18S MB* 18S metabarcoding, *CI* Confidence interval around odds ratio.
[a]Fisher's Exact test, 2-tailed.

Analysis of these samples showed that 9 of 23 samples contained a large proportion of sequences (>5% of total reads and up to 97.4% in one sample) belonging to a single OTU belonging to an unknown member of the bacterial family *Clostridiaceae* (*Clostridia*: *Clostridiales*) (Fig. 4a). This OTU most closely resembled *Clostridium perfringens* in the Greengenes database[39,40]. However, *C. perfringens* diagnostic PCR using published protocols[41–43] failed to yield amplicons in any instances, including in tissues found positive by 16S sequencing.

Re-examination of the representative sequence from this OTU against the National Center for Biotechnology Information's (NCBI's) GenBank (GenBank hereafter) non-redundant database excluding uncultured organisms identified a putative match (97.2% nucleotide identity) to *Clostridium* (*Sarcina*) *ventriculi* from feces of Japanese macaques (*Macaca fuscata* Blyth, 1875; GenBank accession numbers LC101491 and LC101492). Re-examination of all samples by including the *Clostridium* (*Sarcina*) *ventriculi* sequence from GenBank in the Greengenes database demonstrated this organism to be present in all 23 ENGS case samples (Supplementary Data 2). Notably, the organism was present not only in gastrointestinal contents but also in internal organs such as brain, liver, and spleen, sometimes at high abundance (Fig. 4b). The nomenclature of the genus *Sarcina* is contested (and sometimes the genus name *Clostridium* is substituted) because *Sarcina* is phylogenetically situated within the "cluster I" group of *Clostridia*[44], considered the "true" *Clostridia*, although these organisms are polyphyletic. A proposal was made to change the name *Sarcina* to *Clostridium*, but was not approved because the name *Sarcina* predates the name *Clostridium* and therefore has priority[45].

Diagnostic PCR. Oligonucleotide primers specific to the 16S rDNA gene of the unknown organism were successfully developed. PCR with these primers yielded amplicons of the predicted length (289 base pairs [bp]) in 13 of 19 ENGS cases (68.4%) but 0 of 13 controls (Supplementary Fig. 3) which was statistically significant (odds ratio = 56.1; 95% CI 2.87–1097.2; Fisher's exact $P = 0.0001$, two-tailed; Table 1). For one individual, blood samples were available both before and after clinical illness and death

from ENGS; the pre-illness blood sample (collected in February 2016) was PCR-negative whereas the post-mortem blood sample (collected in July 2016) was PCR-positive. Sanger sequences of all amplicons were identical to each other and to the representative sequence generated by metabarcoding, except for one sample with a single nucleotide polymorphism (C → T transition) at position 51 of the diagnostic fragment. Sequences of the diagnostic fragment were also identical to published 16S rDNA sequences in GenBank for *S. ventriculi* (AF110272) and *Clostridium ventriculi* DSM286 (NR026146), with the exception of the one variant sequence (1 nucleotide mismatch to the aforementioned published sequences).

Bacterial isolation and characterization. We attempted to culture the bacterium using 44 combinations of cell preparations and culture conditions (Supplementary Table 4), 2 of which resulted in growth of colonies that resembled sarcinae. Specifically, wet mounts of colonies grown on egg yolk agar plates and *Sarcina ventriculi* growth medium (SVGM) plates revealed refractile, cuboid cells in packets, a morphology that is distinctive of members of the genus *Sarcina* (Fig. 5a, right panel). These colonies were derived from the liver of one individual (1 colony on egg yolk agar plates) and the brain of another (many colonies on SVGM plates; Supplementary Data 2). We confirmed the identity of every colony using diagnostic PCR and Sanger sequencing (see above).

We were repeatedly able to isolate the organism by plating brain tissue onto SVGM plates, but the organism ceased to remain viable after 2–3 passages and did not grow in any of the seven liquid media tested (Supplementary Table 4). These results are consistent with previous studies reporting great difficulty in isolating and propagating sarcinae[46,47]. Furthermore, we were unable to recover live organisms after freezing colonies placed in 10% or 20% glycerol under various conditions. In contrast, we successfully grew the type strain *S. ventriculi* "Goodsir" (American Type Culture Collection [ATCC] 29068) under the same conditions with ease, including propagating the strain in solid and liquid media, freezing the strain in 10% glycerol, and subsequently recovering the bacterium. On SVGM media, our
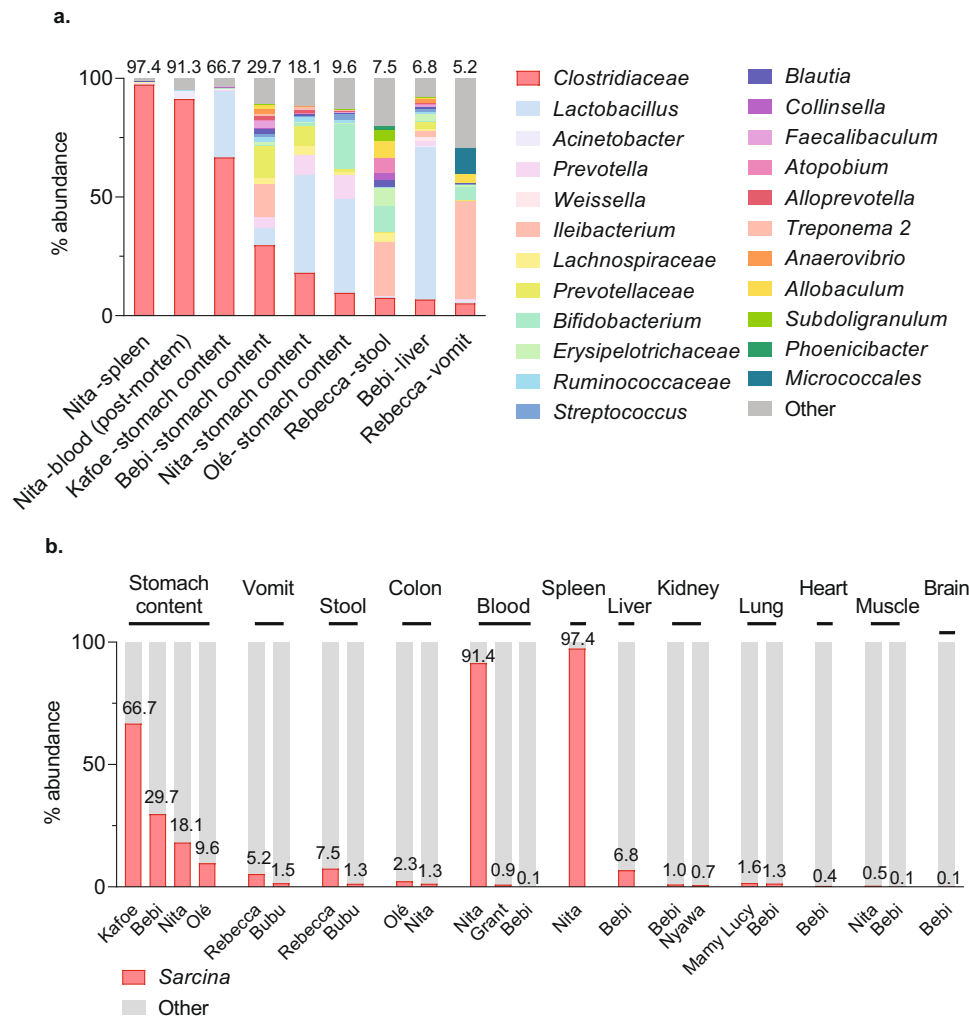
Fig. 4 Bacterial 16S rDNA microbiome analysis of epizootic neurologic and gastroenteric syndrome (ENGS) case samples. a The percentage of total reads ($n = 5900$ per sample) from 9 ENGS case samples at genus-level OTU with % reads mapped to the *Clostridiacea* OTU shown on top of each bar. **b** Percent abundance of reads from 23 ENGS case samples classified to genus-level OTUs *Sarcina* (% above each bar) and other, arranged by tissue type.

isolate (JB1) grew more slowly than *S. ventriculi* "Goodsir" (~3–4 days until colonies were visible for JB1, versus 24 h for *S. ventriculi* "Goodsir"). The JB1 isolate displayed morphology (Fig. 5a), Gram's staining characteristics (Fig. 5b), and methylene blue staining characteristics (Fig. 5c) similar to *S. ventriculi* "Goodsir", as both were Gram-positive with a darkly staining outer layer. However, JB1 cells were statistically significantly larger than those of *S. ventriculi* "Goodsir" (mean diameters of 4.29 μm versus 2.83 μm, respectively, Mann–Whitney *U P* = 0.0006, two-tailed; Fig. 5d). The cellular diameter of JB1 falls within the published range for *S. maxima* (4–4.5 μm)[48], but methylene blue staining showed a cellulose-containing cell wall for isolate JB1 which is not characteristic of *S. maxima* (Fig. 5c). In addition, the flattened cellular morphology and large packet size of JB1 cells resemble *S. ventriculi* and not *S. maxima*[49–51].

Archived histologic preparations of tissues collected from ENGS cases during postmortem examination and stained with hematoxylin and eosin clearly revealed sarcinae, visible as packets of darkly staining basophilic cells in gastric contents of the chimpanzee with hemorrhagic diathesis, gastric dilation, and emphysematous gastritis, and in the pulmonary alveoli of another chimpanzee (Fig. 6a). A wet mount direct smear of homogenized brain tissue from the aforementioned ENGS case also demonstrated the presence of packets of sarcinae (Fig. 6b).

16S rDNA phylogeny. Alignment of 16S rDNA sequences from the "cluster I" group of *Clostridia*[44], including the new organism (isolate JB1), yielded a final alignment length of 1585 positions. A maximum-likelihood phylogeny built from this alignment shows isolate JB1 to represent a sister taxon to *S. ventriculi*, forming a clade with *S. maxima*, *Eubacterium tarantellae*, and *C. perfringens* (Fig. 7). Bacteria of 13 other recognized species pairs included in the analysis had a lower phylogenetic distance between them than the distance between isolate JB1 and *S. ventriculi* (Supplementary Table 5), lending support to the designation of the bacterium as a representative of a distinct species. To reflect the discovery of this bacterium in chimpanzees (*Pan troglodytes* spp.), we designated it "*Candidatus* S. troglodytae". We propose the *Candidatus* designation in this instance because we were unable to generate a culture suitable for deposition in the requisite two publicly accessible culture repositories in two different countries[52].

Whole-genome sequencing, assembly, and annotation. To generate sufficient material for whole-genome sequencing, we repeated bacterial isolation from the brain tissue described above using identical methods and allowed colonies to grow to a large size on SVGM plates. We then harvested a single, large colony, confirmed its identity using microscopy and PCR/sequencing,
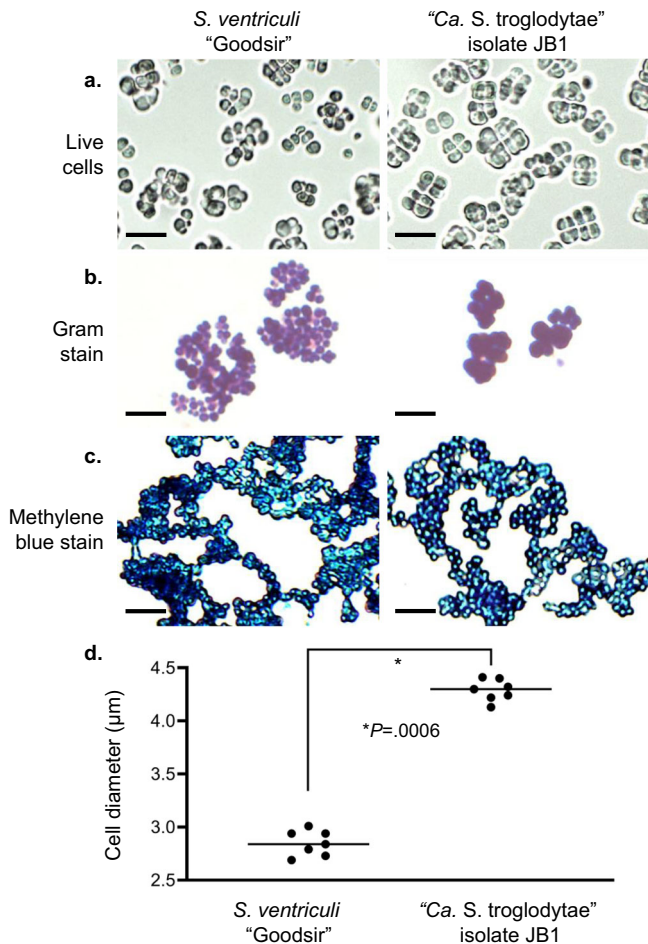
Fig. 5 Comparative morphology of type strain *Sarcina ventriculi* "Goodsir" ATCC 29068 and "*Ca*. S. troglodytae" isolate JB1. a–c Cells were imaged live (**a**) or heat-fixed and stained with Gram stain (**b**) or methylene blue stain (**c**). Micrographs are from a single experiment and are representative of three independent experiments with similar results. Scale bars = 10 µm. **d** Live cell diameters were compared, with lines representing median diameters among seven distinct bacterial colonies. Source data are provided as a Source Data file. *Calculated using a Mann–Whitney *U* test, two-tailed.

and extracted DNA from this colony (isolate JB2). We performed whole-genome sequencing using a hybrid approach of mate-pair and shotgun sequencing (Supplementary Table 6) followed by de novo genomic assembly using SPAdes[53] and in silico genome closure. The resulting full, high-quality "*Ca*. S. troglodytae" assembly consists of a circular chromosome of 2,435,860 base pairs resolved into one single contig and 10 plasmids (totaling 205,993 base pairs, range: 4.6–78.9 kb; Supplementary Table 7).

We annotated the genome with PATRIC[54] and confirmed that the organism is closely related, but not identical, to *S. ventriculi* (98.5% nucleotide similarity; Supplementary Table 8)[55,56]. Total GC content in our organism was 27.6%, which is similar to that of *S. ventriculi* (27.7%). The organism shares 96.5% of its open-reading frames (ORFs) with *S. ventriculi*, with notable differences in sugar pathways and capsule biosynthesis (Fig. 8). The genome of the JB2 strain contains DNA elements encoding metabolic pathways with the potential for formation of bacterial endospores in addition to anaerobic fermentation pathways, including alcohol fermentation, sulfur reduction, and nitrogen reduction. Interestingly, the "*Ca*. S. troglodytae" genome, but not *S. ventriculi*, possesses ORFs encoding for urea degradation

enzymes, including the urease sub-units alpha, beta, and gamma, and ORFs whose products are predicted to be urease accessory proteins UreE, UreF, and UreG, which are needed for urease maturation. Ureases are nickel-containing enzymes, found in a variety of bacteria, which catalyze the breakdown of urea (a ubiquitous metabolic byproduct in most animals) to ammonia and carbon dioxide[57]. Bacterial and fungal ureases play a key role in gastrointestinal tract colonization and in chronic human diseases such as gastritis and peptic ulcers[58]. Additionally, in the extrachromosomal sequences, we found a 34 kb plasmidial prophage containing ORFs for an ATP-binding cassette (ABC) transporter. ABC transporters use ATP to move specific substrates across a cellular membrane and can function either as an importer (e.g., for uptake of nutrients) or an exporter (e.g., to efflux toxic molecules, including xenobiotic compounds such as drugs)[59]. ABC importers have been associated with increased bacterial survival during colonization of hosts[60] and exporters with bacterial drug-resistance[61], both of which may enhance pathogenesis of an organism.

Sarcinae are not known to produce toxins[62]. However, because of the unusual neurologic disease associated with ENGS, we scanned the genome sequence of "*Ca*. S. troglodytae" for ORFs with sequence homology to known virulence genes using ShortBRED[63] and a customized version of the Virulence Factor Database (VFDB)[64], but found no evidence of such genes. Using the Comprehensive Antibiotic Resistance Database (CARD)[65], we identified two antibiotic resistance ORFs, *OXA-241* on the chromosome and *salA* on plasmid 1, which confer resistance to carbapenems (*OXA-241*) and lincosamides and streptogramins (*salA*)[66,67].

Summary description of the provisional species. "*Candidatus* Sarcina troglodytae" is a proposed member of the established genus *Sarcina*, most closely related to *S. ventriculi*, as determined by full-length 16S rDNA phylogenetic analysis. It is an uncultivated, Gram-positive coccus with a tetrad structure and slightly flattened cell morphology and may be identified using the PCR primers: Tacu-Sarc_Diag_F: 5′-TGAAAGGCATCTTTTAACAATCAAAG-3′ and TacuSarc_Diag_R: 5′-TACCGTCATTATCGTCCCTAAA-3′ or the full genome sequence (accessions CP051754–CP051764). We isolated "*Ca*. S. troglodytae" in an anaerobic environment and at mesophilic temperature (37 °C) but were unable to maintain a viable culture for deposition in at least two publicly accessible culture collections, hence the *Candidatus* status. Samples described here are derived from the brain, liver, and lung tissues of sanctuary western chimpanzees (*Pan troglodytes verus*) diagnosed with ENGS.

Discussion

The genus *Sarcina* within the *Clostridiaceae* is poorly studied in comparison to the highly studied toxigenic clostridia. In 1842, Goodsir described the type species, *S. ventriculi*, in the stomach contents of a human patient with recurrent vomiting[68]. Subsequent studies have provided evidence that bacteria morphologically consistent with *S. ventriculi* cause abdominal pain, nausea, anorexia, vomiting, hematemesis, dysphagia, diarrhea, and generalized weakness in people[69], with esophagitis[70] and duodenitis[71] noted surgically or as a post-mortem finding[72]. Morphologically indistinguishable bacteria assumed to be *S. ventriculi* have also been associated with abomasal bloat in young pre-ruminant animals[73–75], characterized by sudden onset of anorexia, abdominal discomfort, lethargy, dehydration, and shock culminating in high lethality (75–100%) despite treatment[76]. Gastric dilation in monogastric animals (horses, dogs, and cats) has also been linked to putative *S. ventriculi* infection[77,78].
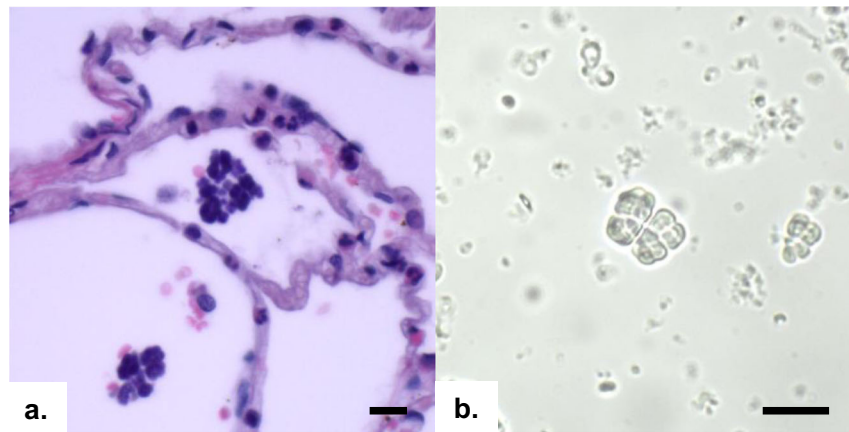
Fig. 6 Characteristic cuboidal packets of *Sarcina*-like organisms in tissues of ENGS-affected chimpanzees. a Basophilic packets of cells in a tetrad formation can be seen amongst and within alveoli of hematoxylin and eosin-stained lung tissue of one individual ("Jumu"). **b** Unstained brain tissue homogenate from another individual ("Joko") contains highly refractile, cuboid packets of cells. Scale bars = 10 μm. Micrographs are representative images of one of at least three sections (**a**) or smears (**b**) obtained from the same individual with similar results.

Our results demonstrate a statistically significant association between a bacterium, "*Ca*. S. troglodytae", and ENGS, a protracted lethal epizootic syndrome in sanctuary chimpanzees in Sierra Leone. We designed a case-control epizootiological study using a case definition that encompassed the range of clinical presentations associated with the syndrome, which included both sudden death and gastrointestinal and neurologic signs prior to death. This case definition will likely become more refined as the syndrome is studied further. By applying metabarcoding and metagenomics, we did not find differences between case and control groups with respect to infection with any parasite or virus and therefore deemed these types of organisms unlikely to be causes of ENGS. However, bacterial metabarcoding and a subsequent PCR revealed infection with "*Ca*. S. troglodytae" in 68.4% of ENGS cases but no controls. In one instance, a chimpanzee was PCR-negative for "*Ca*. S. troglodytae" when healthy but subsequently became PCR-positive after succumbing to ENGS.

Sarcinae are notoriously difficult to culture, particularly from non-environmental sources such as animal tissues[46,47]. Despite being studied since the 1800s, sarcinae have been isolated successfully from only a handful of animal or human sources[75,79,80] (see Supplementary Data 3 for review). Prior to this study, only one photomicrograph of unfixed sarcinae cells in their native morphology was published[81]. Although we were able to isolate JB1, it did not survive repeated passages or freezing, distinguishing it from its closest relative, *S. ventriculi*, as does its larger cell size in culture and slower growth. Flattened cell morphology and a cellulose-containing cell wall distinguish "*Ca*. S. troglodytae" from *S. maxima*[51], its next closest relative, despite overlapping cell size. Phylogenetic analysis based on 16S rDNA demonstrates the difference between "*Ca*. S. troglodytae" and *S. ventriculi* to be greater than the difference between bacteria of 13 other recognized species pairs within the clostridial rDNA group I[82]. Whole-genome sequencing and genetic characterization revealed 69 ORFs that were not found in the genome of *S. ventriculi* "Goodsir." For these reasons, we propose that this organism be considered the representative of a new species within the genus *Sarcina*.

Bacteria within the family *Clostridiaceae* include organisms linked to life-threatening diseases, as well as benign commensals and environmental bacteria[62]. Patterns of virulence/toxigenicity do not correspond to phylogeny, and pathogenicity cannot be predicted based on 16S rDNA sequence grouping alone[83]. Whole genome sequencing of "*Ca*. S. troglodytae" revealed no toxin

ORFs similar to those present in toxigenic clostridia[84]. The pathogenic effects of "*Ca*. S. troglodytae" on chimpanzees may therefore be caused by mechanisms other than toxicity. For example sarcinae have an unusual yeast-like metabolism[85] that is active over a wide pH range[86], allowing bacteria to produce carbon dioxide gas and ethanol prolifically, both of which can cause disease in the gastrointestinal tract and the central nervous system[87,88].

We also found that the genome sequence of "*Ca*. S. troglodytae" contains ORFs encoding for a predicted urease. Although urease expression is associated with normal microbial flora in some instances, ureases are better known as a key virulence factors in pathogenic bacteria such as *C. perfringens*, *Helicobacter pylori*, and *Klebsiella pneumoniae*[89] and are associated with diseases including ammonia encephalopathy, hepatic encephalopathy, hepatic coma, and gastroduodenal infections[57]. Because, in other bacteria, ureases have established roles in infection and persistence in the host[90], stimulation of host inflammatory reactions[91], cytotoxic effects on host cells[92], and damage to extracellular matrix[93] and tight junctions[94], the presence of a urease biochemical pathway in "*Ca*. S. troglodytae" could help explain the bacterium's pathogenesis and dissemination outside of the gastrointestinal tract. For example, urease activity in the yeast *Cryptococcus neoformans* is responsible for central nervous system invasion; unlike the wild-type organism, mutants lacking this enzyme do not disseminate to the brain and cause meningoencephalitis[95]. Moreover, the major product of urea degradation, ammonia, could enhance the ability of "*Ca*. S. troglodytae" to cause neurologic signs, because ammonia is highly neurocytotoxic in vivo[96].

In some cases of *Sarcina* infection in humans, symptoms are preceded by evidence of delayed gastric emptying[69,70,97–99]. With ENGS, however, affected chimpanzees appeared healthy prior to the onset of signs. It is therefore noteworthy that several studies have shown colonization of *Sarcina* and lesions in the absence of delayed gastric emptying[71,100,101]. For example, a recent publication concerning a lethal case of human emphysematous gastritis highlights several similarities to ENGS, including lack of gastroparesis, afebrile and normotensive presentation, gastrointestinal and neurologic signs, and rapid death[102]. The occurrence of acute gastric dilation and emphysematous lesions in the digestive tract of one ENGS case included in our study recalls cases of *Sarcina* infection in humans and other animal species, which include acute gastric dilation and emphysematous gastritis.
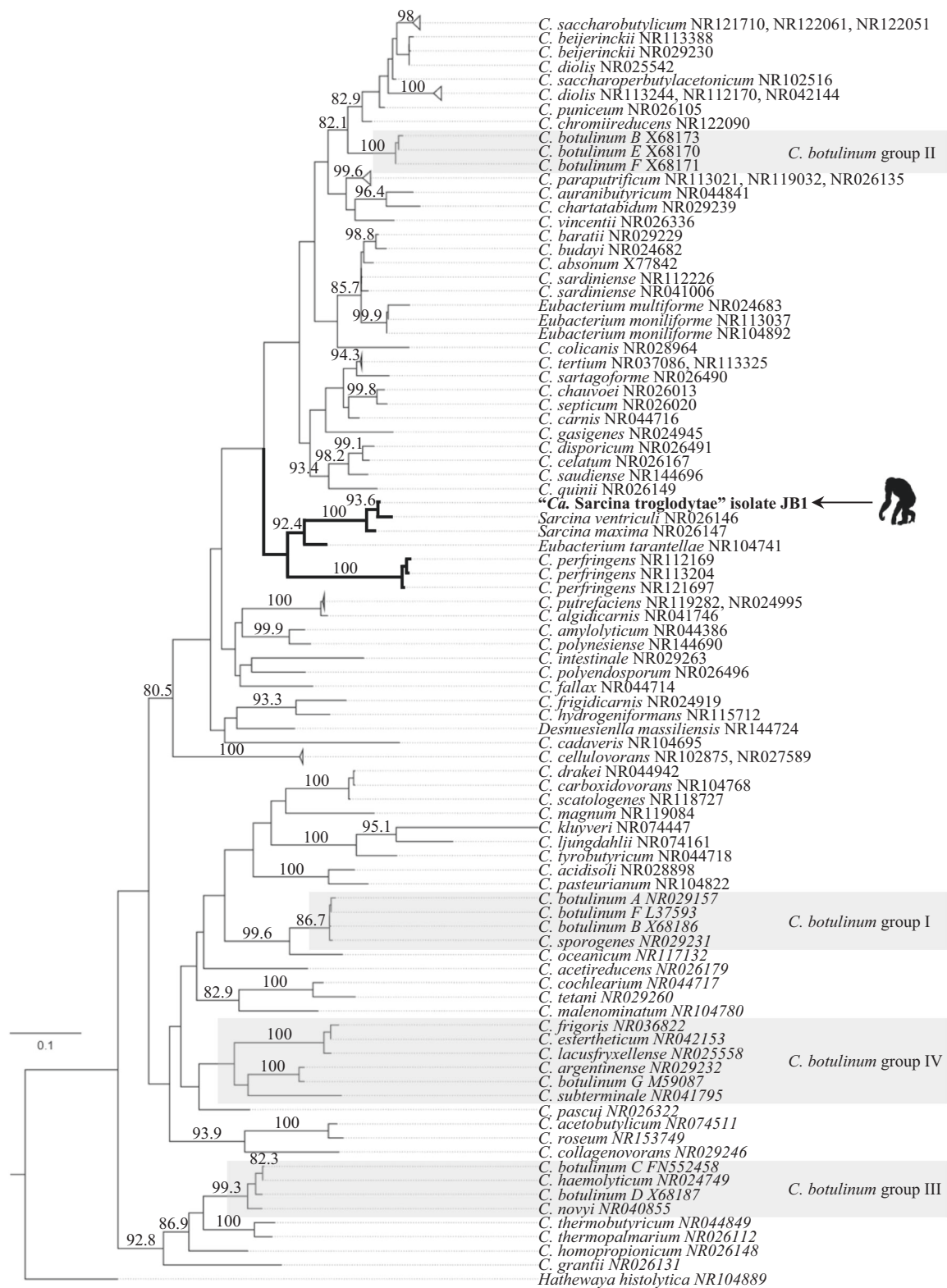
Fig. 7 Maximum-likelihood 16S rDNA gene phylogeny of the *Clostridiaceae*. The phylogeny is based on the complete 16S rDNA sequence of "*Ca.* Sarcina troglodytae" isolate JB1 (arrow and silhouette) and 98 other *Clostridia sensu stricto*, with *Hathewaya histolytica* as the outgroup. Gray boxes indicate *Clostridium botulinum* groups[62]. Numbers above the branches are bootstrap values (%) based on 1000 bootstrap replicates (only values ≥75% are shown). Scale bar indicates nucleotide substitutions per site.

A more consistent gross and histopathologic evaluation of affected chimpanzees in the future may reveal a higher proportion of affected chimpanzees because acute gastric dilation and emphysematous gastrointestinal lesions may be misinterpreted as autolysis or overlooked grossly. For example, a chimpanzee in this population who died of ENGS subsequent to the analyses presented here clearly showed emphysematous lesions throughout the gastrointestinal tract. Although each of the clinical
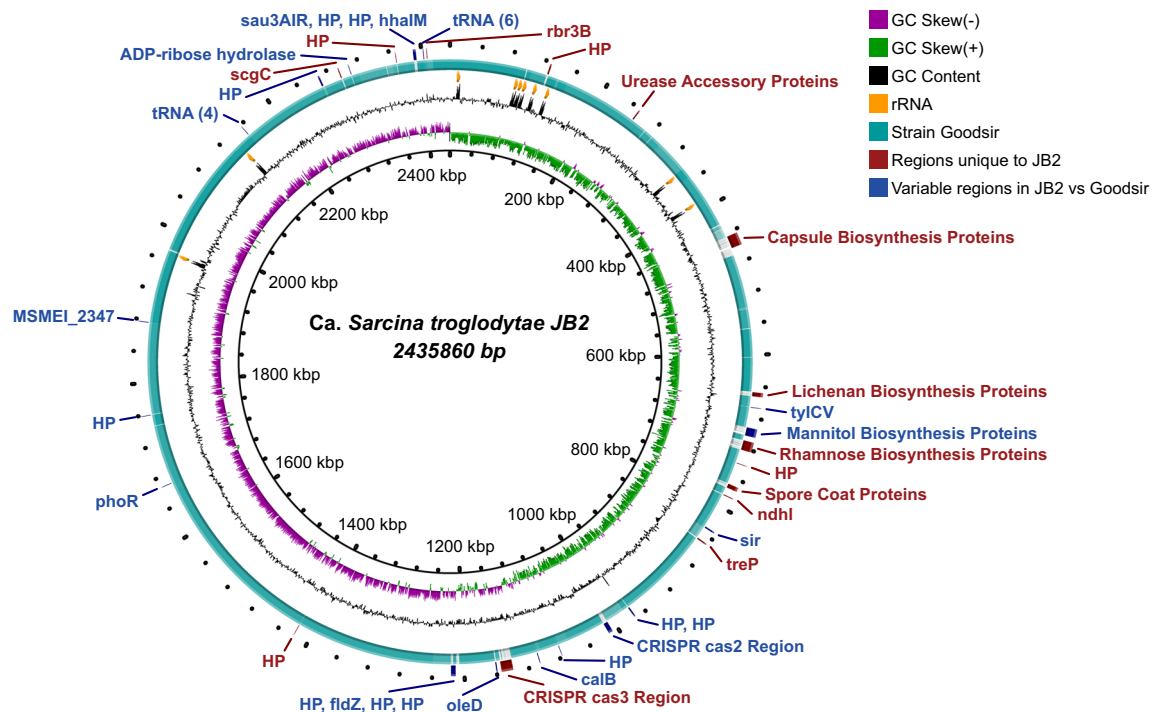
Fig. 8 Whole chromosome comparison of "*Ca.* Sarcina troglodytae" isolate JB2 compared to the type strain *S. ventriculi* "Goodsir". Red: regions which are unique to JB2. Blue: different ORF sets in the same relative region in JB2 and Goodsir. Teal: 90–100% identity gradient between JB2 and Goodsir (Goodsir chromosome sequence is based on a manually scaffolded genome).

characteristics of ENGS (abdominal distention, nausea, vomiting, anorexia, diarrhea, and neurologic deficits) are common to diverse diseases, they are all consistent with emphysematous gastroenteritis, as is acute lethality, which likely results from irreversible hemodynamic instability and resulting systemic shock in emphysematous gastritis cases[103].

Notably, we found "*Ca.* S. troglodytae" not only in the gastrointestinal tracts of affected individuals, but also in internal organs, including the brain. For DNA extraction and culture, all tissues were maintained on dry ice, carefully sectioned using sterile technique, and subsampled from the innermost area while still frozen, leading us to conclude that our findings are not likely due to environmental contamination and instead reflect true infection. In human cases of *S. ventriculi* infection, there is precedent for bacteremia, likely originating from gastrointestinal translocation[104,105], but to our knowledge presence of viable sarcinae in the central nervous system has not been previously reported. Central nervous system colonization may therefore be an overlooked clinical feature of severe *Sarcina* infection, or "*Ca.* S. troglodytae*" may be a particularly virulent bacterium within the genus *Sarcina*. We advocate that prior and future human and animal cases of severe disease associated with sarcinae, particularly those cases without clear predisposing factors, be revisited, as they could represent other heretofore unrecognized presentations of infection with *Sarcina* bacteria. We also speculate that many documented cases of infection that were assumed to be caused by *S. ventriculi* based on morphology alone may actually have been caused by infection with taxonomically distinct sarcinae. If so, the genus *Sarcina* may contain a complex of morphologically cryptic species varying from benign environmental bacteria to lethal pathogens.

Many questions regarding ENGS and "*Ca.* S. troglodytae" remain unexplained. For example, epizootiologically, ENGS incidence peaks in March each year. As with other disease-associated clostridia, sarcinae form environmentally stable spores[106] and may be ubiquitous in soil[79,99,107], but environmental factors may

contribute to germination of spores and overgrowth. Seasonal changes known to be important in chimpanzees, such as habitat use, diet (including exposure to plant or arthropod toxins), and physiological condition[108], may increase infection risk at certain times of the year. Alternatively, "*Ca.* S. troglodytae" may be reintroduced seasonally, for instance by migratory animals[109], perhaps with seasonal weather patterns facilitating its establishment[110]. The potential role of "*Ca.* S. troglodytae" in the etiology of ENGS, alone or in combination with other factors, remains a topic for future research.

Due to lack of infrastructure at TCS and limitations on sample shipments to the U.S., our analysis could only include samples obtained between 2013 and 2016, even though ENGS was first noted in 2005. Ideally we would have obtained the same tissues, especially from the gastrointestinal tract, from all cases, in addition to complete medical records and post-mortem examination notes including gross and histological findings. Unfortunately, this was not possible due to the resource-limited setting and resulting opportunistic sampling. Fortunately, the veterinary staff at TCS have begun standardizing sample and record collection for future cases.

Practical and ethical considerations preclude collecting invasive samples from sanctuary chimpanzees, so control samples were limited to serum and feces collected at annual health checks and samples collected post-mortem from individuals determined to have died from causes other that ENGS (e.g., accidental death), which are rare at TCS. Furthermore, the evidence presented here supports an association between "*Ca.* S. troglodytae" and ENGS which, due to ethical considerations, cannot be further investigated by experimentation on chimpanzees. Likewise, infection trials in an animal model (e.g., laboratory mice) would require a pure culture and, as of yet, we are unable to maintain a culture of this bacterium, hence the *Candidatus* designation[52].

We also note that clinically similar cases have not been reported in other captive or wild populations of chimpanzees or other primates. Moreover, despite over 10 years of illness among

the TCS chimpanzees, human cases have not been reported, even among personnel with close daily contact with affected individuals. The genetic and physiological similarities between humans and chimpanzees are often cited as predisposing them to cross-species pathogen exchange[17,111]. It is therefore surprising that no human disease similar to ENGS has been reported to date. Should "*Ca.* S. troglodytae" indeed affect chimpanzees but not humans, it would represent a rare example of such a pathogen[112]. However, we cannot rule out physiological stressors, diet-related factors, environmental conditions, or other pathogens as predisposing factors that differ between humans and chimpanzees. For example, in the case of *C. perfringens*-associated enteritis in humans, changes in gastric and intestinal pH, altered nutritional status, and concurrent infection, particularly with intestinal viruses and parasites, can drastically alter clinical outcomes[113].

To our knowledge, only 44 cases of *Sarcina* infection in humans have been reported in the peer-reviewed literature since the beginning of the 1900s (Supplementary Data 3), and currently no standard treatment for such infections is available. Of published cases treated with at least one specifically mentioned antibiotic (19 of 44), the most common regimen was a combination of oral ciprofloxacin and metronidazole (11 of 19) with a proton-pump inhibitor (8 of 11) or antacid (2 of 11). With the exception of one case involving other complications, treatment was successful when follow-up was noted (9 of 10). Four published cases detail dosages of the antibiotics, all in adult males, and most dosages (3 of 4) were identical: 250 mg metronidazole three times daily and 250 mg ciprofloxacin twice daily for a course of 7 days. Recently, *S. ventriculi* cultured from human blood was shown to be susceptible to other antibiotics including penicillin (minimal inhibitory concentration [MIC] = 0.25 mg/l), amoxicillin (MIC = 0.50 mg/l), amoxicillin–clavulanic acid, piperacillin–tazobactam, imipenem, clindamycin, levofloxacin, rifampicin, vancomycin, and linezolid[104].

Treatment of emphysematous gastritis is similarly unstandardized and includes hemodynamic stabilization with intravenous fluids, broad spectrum intravenous antibiotics effective against Gram-negative and anaerobic bacteria, including meropenem[114,115], cefuroxime and metronidazole[88], nafcillin and cefoxitin[116], and surgery in some cases, but is associated with 60% lethality[103]. That we found evidence of two antibiotic resistance genes in our "*Ca.* S. troglodytae" isolate, a chromosomal *OXA-241*-like gene involved in carbapenem resistance and a plasmid-associated *salA*-like gene linked to lincosamide/streptogramin resistance[67], is noteworthy, as these findings may influence ENGS treatment decisions. Overgrowth of sarcinae in the stomach appears to predispose patients to clinical disease[72]; therefore, probiotics, particularly those containing acidophilic organisms, may prove useful for the treatment or prevention of "*Ca.* S. troglodytae" infections in chimpanzees. For example, probiotics have proven useful for the prevention of *C. difficile*-related disease in humans[117,118]. Finally, autogenous vaccines have proven useful for the prevention of *C. perfringens*-related disease in animals[119,120]. Such an approach could prove useful for the prevention of ENGS if in vitro growth conditions for "*Ca.* S. troglodytae" can be determined.

Since 2011, case studies and reviews concerning *S. ventriculi* and human disease have increased in the medical literature from 0 articles from 1900–2000, to 2 articles from 2000–2010 and 33 articles from 2011—November 2019 (Supplementary Data 3). Increased recent attention to *Sarcina* despite establishment of the genus in 1842 may be coincidental. Alternatively, it may indicate a nascent trend of bacterial emergence[121,122]. The physiological and environmental drivers of *Sarcina* acquisition and subsequent disease progression merit greater attention than they have heretofore received, as does the genetic diversity of the genus. In 34 of 44 published cases, diagnosis of *Sarcina* infection was based on

morphology and/or Gram staining alone with no other diagnostics for confirmation (Supplementary Data 3). Cases of clinical disease associated with *Sarcina* infections should be re-evaluated in light of the possibility that the bacteria identified may represent a complex of cryptic species and strains, some of which are benign but others of which may be highly virulent.

Methods

Ethics statement. TCS located in Western Area National Park, Sierra Leone, is a non-governmental organization that operates under the purview and with the permission of the Ministry of Agriculture, Forestry, and Food Security. All animals originated from Sierra Leone and were confiscated or handed over to TCS under the authority of the Ministry. TCS does not remove any animals from the wild but works to rescue chimpanzees that have been removed from the wild illegally. The care and sampling of resident chimpanzees is officially sanctioned by the Government of Sierra Leone, and samples were shipped to the USA with the official permission of the Government of Sierra Leone under Convention on International Trade in Endangered Species of Wild Fauna and Flora permit number 17US19807C/9.

The presented study was retrospective, did not involve collection of any samples solely for the purpose of this research, and utilized surplus samples collected by TCS veterinarians during routine veterinary procedures and post-mortem examination, which are standard at the sanctuary for any fatality, in compliance with the "Pan African Sanctuary Alliance Primate Veterinary Healthcare Manual"[123] and the policies of TCS.

Clinical data and samples. We obtained clinical data from veterinary records for chimpanzees who had died of all causes from 2005 through 2018. These data were compiled by year and by month to make epizootic curves. We then used these data to select samples from the TCS freezer archive according to a case-control study design. Due to resource limitations, samples were collected opportunistically (as opposed to systematically) and archived samples were only available from a subset of cases that occurred from 2013 to 2016. Samples had been collected by staff veterinarians during routine health checks or during post-mortem examination, and samples were fresh-frozen (at −20 or −80 °C) upon collection in whirl packs or test tubes (Supplementary Data 1) and stored long-term at −80 °C. Samples were shipped frozen on dry ice to the United States, stored at −80 °C upon arrival, and kept frozen through processing. To obtain sub-samples of solid tissues and avoid contamination from the external surfaces of organs, we cut frozen tissues with a sterile razor blade and extracted tissue plugs from the newly exposed area with a sterile 6-mm biopsy punch.

Parasitology. Microscopy for parasite identification was performed at TCS from 2005 to 2018 following standard veterinary protocols[29]. Briefly, freshly voided fecal samples were collected from individuals and macroscopic features were noted. A direct smear was then made by mixing fecal material with saline and observing the mixture under a light microscope at ×100 and ×400 total magnification, with an additional formalin–ether (10% formalin and ethyl acetate) sedimentation performed as warranted. Slides were read by trained and experienced staff veterinarians. Data on the occurrence of parasites thus identified were compiled from 155 such analyses conducted from 2005 through 2018, representing 17 ENGS-affected chimpanzees (cases) and 13 apparently healthy chimpanzees (controls).

Molecular parasitology using metabarcoding was performed with methods modified from the EMP[32]. DNA was extracted from tissue samples (blood, plasma, serum, lung, and brain) using the DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to manufacturer's instructions and eluted in 50 μl of buffer AE (10 mM Tris–HCl, 0.5 mM ethylenediaminetetraacetic acid). Tissue samples were considered appropriate for this analysis based on published literature showing that infections, including with eukaryotes, can be detected in such samples, even when the tissues analyzed are not the anatomic sites of infection[124,125]. Primers were used to amplify the V9 region of the 18S rDNA gene and were based on published pan-eukaryotic sequences[126,127] (see Supplementary Table 9 for all primers used in this study). These sequences were modified, replacing the individual barcodes with overhang sequences compatible with the Nextera system (Illumina, San Diego, CA, USA)[33–35]. The primers used were EMP_Next_F: 5′-TC GTCGGCAGCGTCAGATGTGTATAAGAGACAGGTACACACCGCCCGTC-3′ and EMP_Next_R: 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGT GATCCTTCTGCAGGTTCACCTAC-3′ (IDT, Newark, NJ, USA). To reduce host signal, we used the EMP mammal blocking primer EMP_Mmmal_Block: 5′-GC CCGTCGCTACTACCGATTGGII IIITTAGTGAGGCCCT-[C3 Spacer]-3′ (IDT). PCR reactions were carried out in 25 μl volumes containing 0.3 μM of each primer, 1.6 μM of mammal blocking primer, 12.5 μl 2× HotStart ReadyMix (KAPA Biosystems, Wilmington, MA, USA), and 25 ng template DNA on a C-1000 thermocycler (BioRad, Hercules, CA, USA) with the following cycling conditions: 95 °C for 3 min; 24 cycles of 98 °C for 20 s, 67 °C for 15 s, 62 °C for 30 s, 72 °C for 15 s; and 72 °C for 1 min.

PCR products were purified using the DNA Clean and Concentrator Kit (Zymo Research, Irvine, CA, USA) and eluted in 25 μl of provided elution buffer. From the

25 µl, 5 µl was then used as a template in a 25-µl PCR mix with the Nextera XT Index Kit v2 (Illumina) and limited-cycle PCR for indexing using an annealing temperature of 55 °C with 12 cycles. Products were separated on a 1.5% agarose gel and visualized to confirm band lengths of ~260 bp. Amplicons were then excised from gels and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research) and eluted in 20 µl of water. Products were quantified using a Qubit fluorometer (Thermo-Fisher Scientific Inc., Waltham, MA, USA). Libraries were sequenced on a MiSeq instrument using paired-end 300 ×300 cycle chemistry (Illumina).

Raw reads were processed with the QIIME v.1.9.1 pipeline[128]. Forward and reverse reads were assembled into paired contigs using the command multiple_join_paired_ends.py and quality filtered using the command multiple_split_libraries_fastq.py with default parameters, except for setting the Phred threshold to 30 or higher (-q 29) and minimum length to 100 bp (-l 100). Chimeras were identified with Usearch v.6.1[129] and removed. Reads were then assigned to OTUs using the QIIME protocol for open reference OTU picking with the command pick_open_reference_otus.py and the default UCLUST tool (v.0.2.0)[129], and taxonomy was assigned to OTUs using default settings with the command assign_taxonomy.py against the SILVA database v. 132[130]. Still-undetermined OTUs were assigned using BLAST within QIIME (-m blast) against the full GenBank database[131] and non-target sequences were then removed by filtering. We processed all data from through all filtering steps after which we removed those samples that represented <0.5% of the total filtered data set from further analyses.

Virology. Samples were homogenized by bead beating (for solid tissues), clarified by centrifugation, and treated with nucleases (0.02 U/µl DNAse I, 1 U/µl RNAse T1, 0.04 U/µl RNAse I, 25 ng/µl RNAse A)[36–38]. Viral RNA was isolated using a QIAamp MinElute virus spin kit (Qiagen), omitting carrier RNA. Extracted nucleic acids were then converted to double-stranded cDNA using the SuperScript double-stranded cDNA Synthesis Kit (Invitrogen, Carlsbad, CA, USA) and random hexamers and purified using Ampure XP beads (Beckman Coulter, Brea, CA, USA). Approximately 1 ng of DNA was prepared as a library for pair-ended sequencing on a MiSeq instrument (MiSeq Reagent kit v3, 150 cycle) using the Nextera XT DNA Library Prep Kit (Illumina). Sequence data were analyzed using CLC Genomics Workbench version 11.0 (Qiagen). In brief, we trimmed low-quality bases (Phred quality score <30), discarded short reads (<75 bp), and subjected the remaining reads to de novo assembly using the CLC assembler with automatic word and bubble size selection and a minimum contig length of 500. We then analyzed contigs for nucleotide-level (blastn) and protein-level (blastx) similarity to known viruses in GenBank. For blastx we applied the BLASTX algorithm with the BLOSUM62 matrix to sequences translated into all six frames. We also analyzed all sequence data at the individual read level by mapping reads to viruses in the GenBank database using the CLC mapping tool at low stringency (length fraction of 0.5, similarity fraction of 0.6).

Bacterial metabarcoding. Genomic DNA was extracted from solid tissue and blood samples using a DNeasy Blood and Tissue Kit and from fecal, vomit, and stomach content samples using a DNeasy PowerSoil DNA Isolation Kit (Qiagen) according to manufacturer's instructions. The V4 region of the bacterial 16S rRNA gene was amplified using the following primers, which contain the Illumina sequencing adapters, an 8-nt indexed barcode sequence, a 10-nt pad sequence, a 2-nt linker, and the 16S rRNA V4-specific primer: 16SV4_F: 5′-AATGATACGGC GACCACCGAGATCTACACNNNNNNNNNTATGGTAATTGTGTGCCAGCMG CCGCGGTAA-3′ and 16S_V4_R: 5′-GGACTACHVGGGTWTCTAATCCAGT CAGTCAGNNNNNNNNNCAAGCAGAAGACGGCATACGAGAT-3′[132]. PCR reactions were carried out in 25 µl volumes containing 0.3 µM of each primer, 12.5 µl 2× HotStart ReadyMix (KAPA Biosystems), 6.5 µl water, and 25 ng template DNA with the following cycling conditions: 95 °C for 3 min; 30 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s; and 72 °C for 5 min. PCR products were then electrophoresed on 1% low melt agarose gels (National Diagnostics, Atlanta, GA, USA), excised, purified using a ZR-96 Zymoclean Gel DNA Recovery Kit (Zymo Research), and quantified using a Qubit® Fluorometer (Thermo Fisher Scientific). Equimolar amounts of the barcoded V4 amplicons were pooled and sequenced using a MiSeq 2 × 250 bp v2 kit (Illumina) using custom sequencing primers with 10% PhiX control DNA.

All sequences were demultiplexed on the Illumina MiSeq and were processed and analyzed using mothur v.1.42.0.[132] Poor quality sequences were removed after paired end sequences were combined into contigs. Sequences were aligned against the SILVA 16S rRNA gene reference alignment database to screen for alignment to the correct region. Preclustering was performed to reduce error and chimeras were detected and removed using UCHIME v.4.2[133]. The SILVA database was used to classify bacterial sequences while sequences classifying to mitochondria, cyanobacteria, Eukarya, Archaea, or Fungi were removed along with singletons to streamline analysis. Percent abundance graphs were created in GraphPad Prism v8.4.3 (GraphPad Software, San Diego, CA, USA).

Diagnostic PCR. PCR primers were designed to the V2–V3 region of the "*Ca*. S. troglodytae" 16S rRNA gene: TacuSarc_Diag_F: 5′-TGAAAGGCATCTTTTAAC AATCAAAG-3′ (T_m = 52.8 °C) and TacuSarc_Diag_R: 5′-TACCGTCATTATCG TCCCTAAA-3′ (T_m = 53 °C) (IDT). PCR reactions were carried out in 25 µl

volumes containing 0.2 µM of each primer, 12.5 µl 2× HotStar Master Mix (Qiagen), 10 µl water, and 25 ng template DNA on a C-1000 thermocycler (BioRad, Hercules, CA, USA) with the following cycling conditions: 95 °C for 15 min; 29 cycles of 94 °C for 30 s, 48 °C for 30 s, 72 °C for 30 s; and 72 °C for 10 min. PCR products (289 bp expected length) were then electrophoresed on 1.5% low-melt agarose gels with ethidium bromide and 1 kb plus DNA length standards (BioRad), visualized under UV light, and photographed using a GelDoc XR imager (BioRad). Amplicons were then excised and purified as described above and Sanger sequenced on ABI 3730xl DNA Analyzers (Applied Biosystems, Foster City, CA, USA) at the University of Wisconsin-Madison Biotechnology Center.

Bacterial isolation and characterization. Liquid samples were pipetted directly onto sterile agar plates and solid tissues (deep cut sections collected with a 6-mm biopsy punch to avoid external contamination) were placed in sterile Petri dishes and minced with two sterile blades until homogenized. 200 µl of pre-reduced thioglycollate medium (Hardy Diagnostics, Santa Maria, CA, USA) was added and the mixture was streaked by inoculating loop onto a 100 mm × 15 mm plate, placed immediately into an Anaerogen Compact anaerobic pouch (Oxoid Limited, Hampshire, UK), sealed, and incubated at 37 °C. For liquid growth media, a sterile 18-gauge needle with 1-ml syringe was used to inoculate stoppered tubes that were incubated anaerobically at 37 °C. Cultures were screened by PCR, and cells were directly visualized and grown at least 10 days before they were deemed negative for growth of the bacterium of interest.

For comparison, the type strain *S. ventriculi* "Goodsir" (ATCC 19633 or ATCC 29068) was obtained from the ATCC (American Type Culture Collection, Manassas, VA, USA) and grown according to ATCC guidelines.

Bacterial imaging. Live bacterial cells diluted in sterile water or phosphate buffered saline were placed on glass slides, examined with light microscopy, and imaged immediately. For Gram staining, heat-fixed slides were flooded with crystal violet solution for 1 min, rinsed with water, flooded with iodine solution for 1 min, rinsed with water, flooded with decolorizer solution for 1–5 s, rinsed with water, counterstained with five drops of safranin solution for 30 s, rinsed with water, and air dried. For methylene blue staining, heat-fixed slides were flooded with 1% aqueous solution of methylene blue for 1 min at room temperature, then washed with distilled water and then air dried. All slides were visualized and photographed at ×400 on a Panthera U microscope with a Moticam 5.0 camera (Motic, British Columbia, Canada). For cell size measurements, strains JB1 and ATCC 29068 "Goodsir" were plated and grown on SVGM plates under identical conditions for 72 h. Bacterial cells from seven distinct colonies were harvested, blinded to the investigator, and examined as follows: live cells in phosphate buffered saline from five non-overlapping visual fields were captured as above and single-cell diameters were quantified using the circle (3-point) measurement tool in the Images Plus software suite v.2.0 (Motic). Cellular diameters were compared using a Mann–Whitney *U* test (two-tailed).

Tissues for histopathology were collected during post-mortem examination by staff veterinarians and immediately fixed in 4% paraformaldehyde at least overnight, then later dehydrated in alcohol, embedded in paraffin wax, cut in 6-µm sections, stained with hematoxylin and eosin, visualized under a light microscope, and photographed. For direct visualization of brain tissue, a 6-mm biopsy punch (Integra LifeSciences, Plainsboro, NJ, USA) was taken from the interior of the cerebrum, minced with sterile blades, smeared onto a clean glass slide, and immediately imaged as described above.

16S rDNA phylogeny. The full 16S rDNA sequence from "*Ca*. S. troglodytae" isolate JB1 (1508 bp) was queried against the NCBI 16S ribosomal RNA sequence (*Bacteria* and *Archaea*) database using megablast[131] with default parameters, and the top 50 results as of 17 July 2019 were downloaded from NCBI's RefSeq or GenBank (all *e*-values were 0). For taxonomic completeness, 73 sequences comprising "cluster I" *Clostridia*[134] as previously published[135] were also retrieved from GenBank, and duplicates were removed. The type organism from the closest known clade (*Hathewaya histolytica*, located in "cluster II" of the clostridia as defined by Collins et al. 1994[134]) was included as an outgroup[135]. The resulting 98 sequences, plus the "*Ca*. S. troglodytae" sequence, were aligned using MUSCLE3.8.31[136] (final alignment length 1585 positions). To quantify nucleotide-level distances among sequences, a pairwise distance matrix was calculated using MEGA7 v.7.0.96[137] with pairwise deletion and 1000 bootstrap replicates to estimate standard errors. The phylogenetic position of "*Ca*. S. troglodytae" was then inferred with PhyML v. 1.8.1[138] using the general time reversible (GTR) substitution model as determined by Smart Model Selection[139], and 1000 bootstrapped data sets were used to estimate statistical confidences of clades.

Whole-genome sequencing, assembly, and annotation. A large, single colony of cells morphologically consistent with "*Ca*. S. troglodytae" (Fig. 5a) which tested positive by diagnostic PCR was grown for 7 days on an SVGM plate at 37 °C in an anaerobic pouch. The entire colony ("isolate JB2") was transferred into a sterile 1.5-ml tube and genomic DNA was extracted using the Wizard Genomic DNA purification kit (Promega) according to the manufacturer's instructions. Mate-pair sequencing libraries were constructed using 1 µg of resulting DNA and the Nextera

gel-free protocol and quantified using NEBNext qPCR (New England Biolabs, Ipswich, MA, USA). All libraries were loaded in equimolar amounts, multiplexed, and sequenced on an Illumina MiSeq using the 2 × 300 bp V3 chemistry.

The mate-pair reads were processed with Nxtrim v.0.4.4[140], then processed with bbduk v.38.72[141] using Phred = 20, length >50 base pairs, and subsampled to 100× coverage with bbnorm v.3.8.9[142]. We used bowtie2 v.2.3.5[143] to remove sample-to-sample bleed through. All reads were assembled with SPAdes v.3.10.1[53] using -careful and -mp arguments and manually closed with Bandage v.0.8.1[144], CLC Workbench 12.0 and EDGE Bioinformatics v.2.3.1[145]. The PATRIC server v.3.6.2[146] with default settings was used for all ORF annotations, CLC Genomics Workbench 12.0 was used for variant analysis, Mauve v.2.4.0[147] was used for genome comparisons, ISFinder[148] was used for insertion sequence identification, the Resistance Gene Identifier (RGI v.4.2.2) from the Comprehensive Antibiotic Resistance Database (CARD v.3.0.0)[65] was used for antibiotic resistance gene identification, a customized database from the Virulence Factor Database (VFD)[64] was used for virulence factor identification with ShortBRED v.0.9.5[63], and average nucleotide identity (ANI) calculator (Enveomics v.0.1.3)[55] was used to generate the average nucleotide identities.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Sequence data that support the findings of this study have been deposited in the National Center for Biotechnology Information (NCBI) GenBank database with the accession codes CP051754, CP051755, CP051756, CP051757, CP051758, CP051759, CP051760, CP051761, CP051762, CP015763, CP015764 (Bacterial genome assembly) and MT350347, MT350348, MT350349, MT350350, MT350351, MT350352, MT350353, MT350354, MT350355, MT350356, MT350357 (viral replicase genes) and in the NCBI Sequence Read Archive with BioProject accession codes PRJNA648419 (16S and 18S metabarcoding reads) and PRJNA625238 (Bacterial genome sequencing reads). Source data are provided with this paper.

References

1. Brower, J. L. The threat and response to infectious diseases (Revised). *Microb. Ecol.* **76**, 19–36 (2018).
2. Cunningham, A. A., Daszak, P. & Wood, J. L. N. One Health, emerging infectious diseases and wildlife: two decades of progress? *Philos. Trans. R. Soc. B* **372**, doi:ARTN 2016016710.1098/rstb.2016.0167 (2017).
3. Lakhundi, S. & Zhang, K. Methicillin-resistant *Staphylococcus aureus*: molecular characterization, evolution, and epidemiology. *Clin. Microbiol. Rev.* **31**, https://doi.org/10.1128/CMR.00020-18 (2018).
4. Gomez-Simmonds, A. & Uhlemann, A. C. Clinical implications of genomic adaptation and evolution of carbapenem-resistant. *J. Infect. Dis.* **215**, S18–S27 (2017).
5. Harrison, L. H., Simonsen, V. & Waldman, E. A. Emergence and disappearance of a virulent clone of *Haemophilus influenzae* biogroup aegyptius, cause of Brazilian purpuric fever. *Clin. Microbiol. Rev.* **21**, 594–605 (2008).
6. Plowright, R. K. et al. Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510 (2017).
7. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).
8. Rogalski, M. A., Gowler, C. D., Shaw, C. L., Hufbauer, R. A. & Duffy, M. A. Human drivers of ecological and evolutionary dynamics in emerging and disappearing infectious disease systems. *Philos. Trans. R. Soc. Lond. B* **372**, https://doi.org/10.1098/rstb.2016.0043 (2017).
9. Rohrl, J. R. et al. Emerging human infectious diseases and the links to global food production. *Nat. Sustain.* **2**, 445–456 (2019).
10. Johnson, P. T., de Roode, J. C. & Fenton, A. Why infectious disease research needs community ecology. *Science* **349**, 1259504 (2015).
11. Hernandez, H. & Martinez, L. R. Relationship of environmental disturbances and the infectious potential of fungi. *Microbiology* **164**, 233–241 (2018).
12. McMahon, B. J., Morand, S. & Gray, J. S. Ecosystem change and zoonoses in the Anthropocene. *Zoonoses Public Health* **65**, 755–765 (2018).
13. Hoberg, E. P. & Brooks, D. R. Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. *Philos. Trans. R. Soc. Lond. B* **370**, https://doi.org/10.1098/rstb.2013.0553 (2015).
14. Chiu, C. Y. & Miller, S. A. Clinical metagenomics. *Nat. Rev. Genet.* **20**, 341–355 (2019).
15. Radoshevich, L. & Cossart, P. *Listeria monocytogenes*: towards a complete picture of its physiology and pathogenesis. *Nat. Rev. Microbiol.* **16**, 32–46 (2018).
16. Hoffmann, C. et al. Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. *Nature* **548**, 82–86 (2017).
17. Devaux, C. A., Mediannikov, O., Medkour, H. & Raoult, D. Infectious disease risk across the growing human-non human primate interface: a review of the evidence. *Front. Public Health* **7**, 305 (2019).
18. Sharp, P. M., Rayner, J. C. & Hahn, B. H. Evolution. Great apes and zoonoses. *Science* **340**, 284–286 (2013).
19. Plenderleith, L. J. et al. Ancient introgression between two ape malaria parasite species. *Genome Biol. Evol.* **11**, 3269–3274 (2019).
20. Loy, D. E. et al. Out of Africa: origins and evolution of the human malaria parasites *Plasmodium falciparum* and *Plasmodium vivax*. *Int. J. Parasitol.* **47**, 87–97 (2017).
21. Sundararaman, S. A. et al. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. *Nat. Commun.* **7**, 11078 (2016).
22. Dunay, E., Apakupakul, K., Leard, S., Palmer, J. L. & Deem, S. L. Pathogen transmission from humans to great apes is a growing threat to primate conservation. *EcoHealth* **15**, 148–162 (2018).
23. Muehlenbein, M. P. Primates on display: potential disease consequences beyond bushmeat. *Am. J. Phys. Anthropol.* **162**, 32–43 (2017).
24. Muniz, C. P. et al. Zoonotic infection of Brazilian primate workers with New World simian foamy virus. *PLoS ONE* **12**, e0184502 (2017).
25. Guagliardo, S. A. J. et al. Asymptomatic orthopoxvirus circulation in humans in the wake of a monkeypox outbreak among chimpanzees in Cameroon. *Am. J. Trop. Med. Hyg.* https://doi.org/10.4269/ajtmh.19-0467 (2019).
26. Murkey, J. A. et al. Hepatitis E virus-associated meningoencephalitis in a lung transplant recipient diagnosed by clinical metagenomic sequencing. *Open Forum Infect. Dis.* **4**, ofx121 (2017).
27. Wilson, M. R. et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N. Engl. J. Med.* **370**, 2408–2417 (2014).
28. Wilson, M. R. et al. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. *N. Engl. J. Med.* **380**, 2327–2340 (2019).
29. Modrý, D., P., Barbora, Petrželková, K. Hasegawa, H. *Parasites of Apes an Atlas of Coproscopic Diagnostics*, Vol. 78 (Edition Chimaira, 2018).
30. Howells, M. E., Pruetz, J. & Gillespie, T. R. Patterns of gastro-intestinal parasites and commensals as an index of population and ecosystem health: the case of sympatric western chimpanzees (*Pan troglodytes verus*) and guinea baboons (*Papio hamadryas papio*) at Fongoli, Senegal. *Am. J. Primatol.* **73**, 173–179 (2011).
31. McLennan, M. R., Hasegawa, H., Bardi, M. & Huffman, M. A. Gastrointestinal parasite infections and self-medication in wild chimpanzees surviving in degraded forest fragments within an agricultural landscape mosaic in Uganda. *PLoS ONE* **12**, e0180431 (2017).
32. Thompson, L. R. et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**, 457–+ (2017).
33. Lappan, R. et al. Meta-taxonomic analysis of prokaryotic and eukaryotic gut flora in stool samples from visceral leishmaniasis cases and endemic controls in Bihar State India. *PLoS Negl. Trop. Dis.* **13**, e0007444 (2019).
34. Waraniak, J. M., Marsh, T. L. & Scribner, K. T. 18S rRNA metabarcoding diet analysis of a predatory fish community across seasonal changes in prey availability. *Ecol. Evol.* **9**, 1410–1430 (2019).
35. Belda, E. et al. Preferential suppression of *Anopheles gambiae* host sequences allows detection of the mosquito eukaryotic microbiome. *Sci. Rep.* **7**, 3241 (2017).
36. Goldberg, T. L., Bennett, A. J., Kityo, R., Kuhn, J. H. & Chapman, C. A. Kanyawara virus: a novel rhabdovirus infecting newly discovered nycteribiid bat flies infesting previously unknown pteropodid bats in Uganda. *Sci. Rep.* **7**, 5287 (2017).
37. Scully, E. J. et al. Lethal respiratory disease associated with human rhinovirus C in wild chimpanzees, Uganda, 2013. *Emerg. Infect. Dis.* **24**, 267–274 (2018).
38. Toohey-Kurth, K., Sibley, S. D. & Goldberg, T. L. Metagenomic assessment of adventitious viruses in commercial bovine sera. *Biologicals* **47**, 64–68 (2017).
39. DeSantis, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**, 5069–5072 (2006).
40. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
41. Rinttilä, T., Kassinen, A., Malinen, E., Krogius, L. & Palva, A. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. *J. Appl. Microbiol.* **97**, 1166–1177 (2004).
42. Wang, R. F., Cao, W. W., Franklin, W., Campbell, W. & Cerniglia, C. E. A 16S rDNA-based PCR method for rapid and specific detection of *Clostridium perfringens* in food. *Mol. Cell. Probes* **8**, 131–137 (1994).

43. Wise, M. G. & Siragusa, G. R. Quantitative detection of *Clostridium perfringens* in the broiler fowl gastrointestinal tract by real-time PCR. *Appl. Environ. Microbiol.* **71**, 3911–3916 (2005).

44. Johnson, J. L. & Francis, B. S. Taxonomy of the *Clostridia*: ribosomal ribonucleic acid homologies among the species. *J. Gen. Microbiol.* **88**, 229–244 (1975).

45. Tindall, B. J. Priority of the genus name *Clostridium Prazmowski* 1880 (Approved Lists 1980) vs *Sarcina Goodsir* 1842 (Approved Lists 1980) and the creation of the illegitimate combinations *Clostridium maximum* (Lindner 1888) Lawson and Rainey 2016 and *Clostridium ventriculi* (Goodsir 1842) Lawson and Rainey 2016 that may not be used. *Int. J. Syst. Evol. Microbiol.* **66**, 4890–4894 (2016).

46. Edwards, G. T. et al. *Sarcina*-like bacteria associated with bloat in young lambs and calves. *Vet. Rec.* **163**, 391–393 (2008).

47. Panciera, R. J., Boileau, M. J. & Step, D. L. Tympany, acidosis, and mural emphysema of the stomach in calves: report of cases and experimental induction. *J. Vet. Diagn. Invest.* **19**, 392–395 (2007).

48. Canale-Parola, E. Biology of the sugar-fermenting Sarcinae. *Bacteriol. Rev.* **34**, 82–97 (1970).

49. Canale-Parola, E., Mandel, M. & Kupper, D. G. The classification of sarcinae. *Arch. Mikrobiol.* **58**, 30–34 (1967).

50. Canale-Parola, E. & Wolfe, R. S. Synthesis of cellulose by *Sarcina ventriculi*. *Biochim. Biophys. Acta* **82**, 403–405 (1964).

51. Holt, S. C. & Canale-Parola, E. Fine structure of *Sarcina maxima* and *Sarcina ventriculi*. *J. Bacteriol.* **93**, 399–410 (1967).

52. International code of nomenclature of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **69**, S1–S111 (2019).

53. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

54. Wattam, A. R. et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **45**, D535–D542 (2017).

55. Rodriguez-R, L. M. & Konstantinidis, K. T. *The Enveomics Collection: A Toolbox For Specialized Analyses of Microbial Genomes and Metagenomes*. Report No. 2167-9843 (PeerJ Preprints, 2016).

56. Goris, J. et al. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).

57. Burne, R. A. & Chen, Y. Y. Bacterial ureases in infectious diseases. *Microbes Infect.* **2**, 533–542 (2000).

58. Konieczna, I. et al. Bacterial urease and its role in long-lasting human diseases. *Curr. Protein Pept. Sci.* **13**, 789–806 (2012).

59. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10**, 218–227 (2009).

60. Tanaka, K. J., Song, S., Mason, K. & Pinkett, H. W. Selective substrate uptake: the role of ATP-binding cassette (ABC) importers in pathogenesis. *Biochim. Biophys. Acta Biomembr.* **1860**, 868–877, (2018).

61. Lubelski, J., Konings, W. N. & Driessen, A. J. Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiol. Mol. Biol. Rev.* **71**, 463–476 (2007).

62. Hatheway, C. L. Toxigenic clostridia. *Clin. Microbiol. Rev.* **3**, 66–98 (1990).

63. Kaminski, J. et al. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLoS Comput. Biol.* **11**, e1004557 (2015).

64. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: hierarchical and refined dataset for big data analysis–10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).

65. Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).

66. Higgins, P. G., Zander, E. & Seifert, H. Identification of a novel insertion sequence element associated with carbapenem resistance and the development of fluoroquinolone resistance in *Acinetobacter radioresistens*. *J. Antimicrob. Chemother.* **68**, 720–722 (2013).

67. Hot, C., Berthet, N. & Chesneau, O. Characterization of sal(A), a novel gene responsible for lincosamide and streptogramin A resistance in *Staphylococcus sciuri*. *Antimicrob. Agents Chemother.* **58**, 3335–3341 (2014).

68. Goodsir, J. & Wilson, G. History of a case in which a fluid periodically ejected from the stomach contained vagetable organisms of an undescribed form. *Edinb. Med. Surg. J.* **51**, 430–443 (1842).

69. Laass, M. W. et al. Emphysematous gastritis caused by *Sarcina ventriculi*. *Gastrointest. Endosc.* **72**, 1101–1103 (2010).

70. de Meij, T. G. J., van Wijk, M. P., Mookhoek, A. & Budding, A. E. Ulcerative gastritis and esophagitis in two children with *Sarcina ventriculi* infection. *Front. Med.* **4**, 145 (2017).

71. Canan, O., Ozkale, M. & Kayaseicuk, F. Duodenitis caused by *Sarcina ventriculi* in a case with Celiac disease and selective IgA deficiency. *Cukurova Med. J.* **42**, 766–768 (2017).

72. Gaspar, B. L. The significance of *Sarcina* in routine surgical pathology practice. *APMIS* **124**, 436–443 (2016).

73. DeBey, B. M., Blanchard, P. C. & Durfee, P. T. Abomasal bloat associated with *Sarcina*-like bacteria in goat kids. *J. Am. Vet. Med. Assoc.* **209**, 1468–1469 (1996).

74. Van Kruiningen, H. J. et al. Clostridial abomasal disease in Connecticut dairy calves. *Can. Vet. J.* **50**, 857–860 (2009).

75. Vatn, S., Tranulis, M. A. & Hofshagen, M. *Sarcina*-like bacteria, *Clostridium fallax* and *Clostridium sordellii* in lambs with abomasal bloat, haemorrhage and ulcers. *J. Comp. Pathol.* **122**, 193–200 (2000).

76. Marshall, T. S. Abomasal ulceration and tympany of calves. *Vet. Clin. North Am. Food Anim. Pract.* **25**, 209–220 (2009).

77. Im, J. Y., Sokol, S. & Duhamel, G. E. Gastric dilatation associated with gastric colonization with *Sarcina*-like bacteria in a cat with chronic enteritis. *J. Am. Anim. Hosp. Assoc.* **53**, 321–325 (2017).

78. Vatn, S., Gunnes, G., Nybø, K. & Juul, H. M. Possible involvement of *Sarcina ventriculi* in canine and equine acute gastric dilatation. *Acta Vet. Scand.* **41**, 333–337 (2000).

79. Browne, H. P. et al. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* **533**, 543 (2016).

80. Tolentino, L. E., Kallichanda, N., Javier, B., Yoshimori, R. & French, S. W. A case report of gastric perforation and peritonitis associated with opportunistic infection by *Sarcina ventriculi*. *Lab. Med.* **34**, 535–537 (2003).

81. Vatn, S. *Abomasal Disease in Lambs: Aetiological and Predisposing Factors Associated with Bloat, Haemorrhage and Ulcers* (S. Vatn, 1999).

82. Lawson, P. A., Llop-Perez, P., Hutson, R. A., Hippe, H. & Collins, M. D. Towards a phylogeny of the clostridia based on 16S rRNA sequences. *FEMS Microbiol. Lett.* **113**, 87–92 (1993).

83. Popoff, M. R. & Bouvet, P. Genetic characteristics of toxigenic *Clostridia* and toxin gene evolution. *Toxicon* **75**, 63–89 (2013).

84. Carter, G. P., Cheung, J. K., Larcombe, S. & Lyras, D. Regulation of toxin production in the pathogenic clostridia. *Mol. Microbiol.* **91**, 221–231 (2014).

85. Stephenson, M. P. & Dawes, E. A. Pyruvic acid and formic acid metabolism in *Sarcina ventriculi* and the role of ferredoxin. *J. Gen. Microbiol.* **69**, 331–343 (1971).

86. Goodwin, S. & Zeikus, J. G. Physiological adaptations of anaerobic bacteria to low pH: metabolic control of proton motive force in *Sarcina ventriculi*. *J. Bacteriol.* **169**, 2150–2157 (1987).

87. Adinoff, B., Bone, G. H. & Linnoila, M. Acute ethanol poisoning and the ethanol withdrawal syndrome. *Med. Toxicol. Advers. Drug Exp.* **3**, 172–196 (1988).

88. Al-Jundi, W. & Shebl, A. Emphysematous gastritis: case report and literature review. *Int. J. Surg.* **6**, e63–e66 (2008).

89. Mora, D. & Arioli, S. Microbial urease in health and disease. *PLoS Pathog.* **10**, e1004472 (2014).

90. Schoep, T. D. et al. Surface properties of *Helicobacter pylori* urease complex are essential for persistence. *PLoS ONE* **5**, e15042 (2010).

91. Scopel-Guerra, A. et al. The impact of *Helicobacter pylori* urease upon platelets and consequent contributions to inflammation. *Front. Microbiol.* **8**, 2447 (2017).

92. Dunn, B. E. & Phadnis, S. H. Structure, function and localization of *Helicobacter pylori* urease. *Yale. J. Biol. Med.* **71**, 63–73 (1998).

93. Dubreuil, J. D., Giudice, G. D. & Rappuoli, R. *Helicobacter pylori* interactions with host serum and extracellular matrix proteins: potential role in the infectious process. *Microbiol. Mol. Biol. Rev.* **66**, 617–629 (2002). table of contents.

94. Wroblewski, L. E. et al. *Helicobacter pylori* dysregulation of gastric epithelial tight junctions by urease-mediated myosin II activation. *Gastroenterology* **136**, 236–246 (2009).

95. Olszewski, M. A. et al. Urease expression by *Cryptococcus neoformans* promotes microvascular sequestration, thereby enhancing central nervous system invasion. *Am. J. Pathol.* **164**, 1761–1771 (2004).

96. Norenberg, M. D., Rama Rao, K. V. & Jayakumar, A. R. Signaling factors in the mechanism of ammonia neurotoxicity. *Metab. Brain Dis.* **24**, 103–117 (2009).

97. DiMaio, M. A., Longacre, P. W. & Gastric, T. A. *Sarcina* organisms in a patient with cystic fibrosis. *Hum. Pathol. Case Rep.* **1**, 45–48 (2014).

98. Gulati, R., Khalid, S., Tafoya, M. A. & McCarthy, D. Nausea and vomiting in a diabetic patient with delayed gastric emptying: do not delay diagnosis. *Dig. Dis. Sci.* **64**, 681–684 (2019).

99. Smit, J. The biology of the fermenting Sarcinae. *J. Pathol. Bacteriol.* **36**, 455–468 (1933).

100. Alvin, M. & Al Jalbout, N. Emphysematous gastritis secondary to *Sarcina ventriculi*. *BMJ Case Rep.* **2018**, https://doi.org/10.1136/bcr-2018-224233 (2018).

101. Ratuapli, S. K., Lam-Himlin, D. M. & Heigh, R. I. *Sarcina ventriculi* of the stomach: a case report. *World J. Gastroenterol.* **19**, 2282–2285 (2013).

102. Singh, K. Emphysematous gastritis associated with *Sarcina ventriculi. Case Rep. Gastroenterol.* **13**, 207–213 (2019).

103. van Mook, W. N., van der Geest, S., Goessens, M. L., Schoon, E. J. & Ramsay, G. Gas within the wall of the stomach due to emphysematous gastritis: case report and review. *Eur. J. Gastroenterol. Hepatol.* **14**, 1155–1160 (2002).

104. Bortolotti, P. et al. *Clostridium ventriculi* bacteremia following acute colonic pseudo-obstruction: a case report. *Anaerobe* **59**, 32–34 (2019).

105. Tuuminen, T., Suomala, P. & Vuorinen, S. *Sarcina ventriculi* in blood: the first documented report since 1872. *BMC Infect. Dis.* **13**, 169 (2013).

106. Lowe, S. E., Pankratz, H. S. & Zeikus, J. G. Influence of pH extremes on sporulation and ultrastructure of *Sarcina ventriculi. J. Bacteriol.* **171**, 3775–3781 (1989).

107. Rings, D. M. Clostridial disease associated with neurologic signs: tetanus, botulism, and enterotoxemia. *Vet. Clin. North Am. Food Anim. Pract.* **20**, 379–391 (2004). vii–viii.

108. Chancellor, R. L., Rundus, A. S. & Nyandwi, S. Chimpanzee seed dispersal in a montane forest fragment in Rwanda. *Am. J. Primatol.* **79**, 1–8 (2017).

109. Bandelj, P. et al. Prevalence and molecular characterization of *Clostridium difficile* isolated from European Barn Swallows (*Hirundo rustica*) during migration. *BMC Vet. Res.* **10**, 40 (2014).

110. Lipp, E. K., Farrah, S. A. & Rose, J. B. Assessment and impact of microbial fecal pollution and human enteric pathogens in a coastal community. *Mar. Pollut. Bull.* **42**, 286–293 (2001).

111. Calvignac-Spencer, S., Leendertz, S. A. J., Gillespie, T. R. & Leendertz, F. H. Wild great apes as sentinels and sources of infectious disease. *Clin. Microbiol. Infect.* **18**, 521–527 (2012).

112. Sakuma, R. & Takeuchi, H. SIV replication in human cells. *Front. Microbiol.* **3**, 162 (2012).

113. Allaart, J. G., van Asten, A. J. & Grone, A. Predisposing factors and prevention of *Clostridium perfringens*-associated enteritis. *Comp. Immunol. Microbiol. Infect. Dis.* **36**, 449–464 (2013).

114. Huang, H. et al. *Clostridium difficile* infections in a Shanghai hospital: antimicrobial resistance, toxin profiles and ribotypes. *Int. J. Antimicrob. Agents* **33**, 339–342 (2009).

115. Loi, T. H., See, J. Y., Diddapur, R. K. & Issac, J. R. Emphysematous gastritis: a case report and a review of literature. *Ann. Acad. Med. Singap.* **36**, 72–73 (2007).

116. Moosvi, A. R., Saravolatz, L. D., Wong, D. H. & Simms, S. M. Emphysematous gastritis: case report and review. *Rev. Infect. Dis.* **12**, 848–855 (1990).

117. Goldenberg, J. Z. et al. Probiotics for the prevention of *Clostridium difficile*-associated diarrhea in adults and children. *Cochrane Database Syst. Rev.* **12**, CD006095 (2017).

118. Johnston, B. C. et al. Probiotics for the prevention of *Clostridium difficile*-associated diarrhea: a systematic review and meta-analysis. *Ann. Intern. Med.* **157**, 878–888 (2012).

119. Gohari, I. M. et al. Characterization of *Clostridium perfringens* in the feces of adult horses and foals with acute enterocolitis. *Can. J. Vet. Res.* **78**, 1–7 (2014).

120. Springer, S. et al. [Occurrence and control of the *Clostridium perfringens* type A associated diarrhea of the suckling pigs with special consideration of the immunoprophylaxis]. *Tierarztl. Prax. Ausg. G Grosstiere Nutztier.* **40**, 375–382 (2012).

121. Bushnell, G., Mitrani-Gold, F. & Mundy, L. M. Emergence of New Delhi metallo-beta-lactamase type 1-producing enterobacteriaceae and non-enterobacteriaceae: global case detection and bacterial surveillance. *Int. J. Infect. Dis.* **17**, e325–e333 (2013).

122. Shurin, P. A. et al. Emergence of beta-lactamase-producing strains of *Branhamella catarrhalis* as important agents of acute otitis media. *Pediatr. Infect. Dis.* **2**, 34–38 (1983).

123. Unwin, S., Cress, D., Colin, C., Bailey, W. & Boardman, W. Section 3. Veterinary protocols and procedures. (Pan African Sanctuary Alliance (PASA), Portland, OR, 2009).

124. Blauwkamp, T. A. et al. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat. Microbiol.* **4**, 663–674 (2019).

125. Hong, D. K. et al. Liquid biopsy for infectious diseases: sequencing of cell-free plasma to detect pathogen DNA in patients with invasive fungal disease. *Diagn. Microbiol. Infect. Dis.* **92**, 210–213 (2018).

126. Amaral-Zettler, L. A., McCliment, E. A., Ducklow, H. W. & Huse, S. M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* **4**, e6372 (2009).

127. Stoeck, T. et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31 (2010).

128. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).

129. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).

130. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).

131. Morgulis, A. et al. Database indexing for production MegaBLAST searches. *Bioinformatics* **24**, 1757–1764 (2008).

132. Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. & Schloss, P. D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **79**, 5112–5120 (2013).

133. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).

134. Collins, M. D. et al. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int. J. Syst. Bacteriol.* **44**, 812–826 (1994).

135. Lawson, P. A. & Rainey, F. A. Proposal to restrict the genus *Clostridium* Prazmowski to *Clostridium butyricum* and related species. *Int. J. Syst. Evol. Microbiol.* **66**, 1009–1016 (2016).

136. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).

137. Kumar, R. et al. PCV: an alignment free method for finding homologous nucleotide sequences and its application in phylogenetic study. *Interdiscip. Sci.* **9**, 173–183 (2017).

138. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

139. Lefort, V., Longueville, J. E. & Gascuel, O. SMS: smart model selection in PhyML. *Mol. Biol. Evol.* **34**, 2422–2424 (2017).

140. O'Connell, J. et al. NxTrim: optimized trimming of Illumina mate pair reads. *Bioinformatics* **31**, 2035–2037 (2015).

141. Joshi, N. A. & Fass, J. N. *Sickle: A Sliding-window, Adaptive, Quality-based Trimming Tool for FastQ files v. Version 1.33* https://github.com/najoshi/sickle (2011).

142. Hornick, L. BBNorm Guide https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbnorm-guide/ (2019).

143. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

144. Wick, R. R., Schultz, M. B., Zobel, J. & Holt, K. E. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics* **31**, 3350–3352 (2015).

145. Li, P. E. et al. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. *Nucleic Acids Res.* **45**, 67–80 (2017).

146. Aziz, R. K. et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).

147. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* **5**, e11147 (2010).

148. Siguier, P., Pérochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* **34**, D32–D36 (2006).

Acknowledgements

Author contributions

T.L.G. conceived the study; L.A.O., T.L.G., and I.H. designed the study; I.H., A.P., J.E.J., and S.M. coordinated sample collection and transport; L.A.O., B.C., K.A.B.-L., L.J.V., C.L.D., C.D.D., and C.J.-S. performed laboratory work and collected data; L.A.O., K.A.B.-L., L.A.E., L.J.V., J.H.K., G.S., C.L.D., analyzed data; L.A.O. and T.L.G. wrote the manuscript; all authors made substantive intellectual contributions, revised the manuscript, and approved the final draft.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-21012-x.

Correspondence and requests for materials should be addressed to T.L.G.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.