Bifidobacterium castoris strains isolated from wild mice show evidence of frequent host switching and diverse carbohydrate metabolism potential

Kujawska, Magdalena; Raulo, Aura; Millar, Molly; Warren, Fred; Baltrūnaitė, Laima; Knowles, Sarah C. L.; Hall, Lindsay J.

DOI:
10.1038/s43705-022-00102-x

ISME

www.nature.com/ismecomms

ARTICLE OPEN

Check for updates

Bifidobacterium castoris strains isolated from wild mice show evidence of frequent host switching and diverse carbohydrate metabolism potential

Magdalena Kujawska[1,2], Aura Raulo[3], Molly Millar[4], Fred Warren[4], Laima Baltrūnaitė[5], Sarah C. L. Knowles[3,6] and Lindsay J. Hall[1,2,7]✉

Members of the gut microbiota genus *Bifidobacterium* are widely distributed human and animal symbionts believed to exert beneficial effects on their hosts. However, in-depth genomic analyses of animal-associated species and strains are somewhat lacking, particularly in wild animal populations. Here, to examine patterns of host specificity and carbohydrate metabolism capacity, we sequenced whole genomes of *Bifidobacterium* isolated from wild-caught small mammals from two European countries (UK and Lithuania). Members of *Bifidobacterium castoris*, *Bifidobacterium animalis* and *Bifodobacterium pseudolongum* were detected in wild mice (*Apodemus sylvaticus*, *Apodemus agrarius* and *Apodemus flavicollis*), but not voles or shrews. *B. castoris* constituted the most commonly recovered *Bifidobacterium* (78% of all isolates), with the majority of strains only detected in a single population, although populations frequently harboured multiple co-circulating strains. Phylogenetic analysis revealed that the mouse-associated *B. castoris* clades were not specific to a particular location or host species, and their distribution across the host phylogeny was consistent with regular host shifts rather than host-microbe codiversification. Functional analysis, including *in vitro* growth assays, suggested that mouse-derived *B. castoris* strains encoded an extensive arsenal of carbohydrate-active enzymes, including putative novel glycosyl hydrolases such as chitosanases, along with genes encoding putative exopolysaccharides, some of which may have been acquired via horizontal gene transfer. Overall, these results provide a rare genome-level analysis of host specificity and genomic capacity among important gut symbionts of wild animals, and reveal that *Bifidobacterium* has a labile relationship with its host over evolutionary time scales.

ISME Communications; https://doi.org/10.1038/s43705-022-00102-x

INTRODUCTION

Species and strains belonging to the bacterial genus *Bifidobacterium* are prominent members of the gut microbiota in many animals, and are universally distributed among animals exhibiting parental care, including humans and non-human mammals, birds and social insects [1]. *Bifidobacterium* species that colonise the human gut, especially those associated with early life stages, have received much attention in recent years due to their ability to confer health benefits on their host, including supporting development of the wider gut microbial ecosystem through the production of short-chain fatty acids, colonisation resistance against pathogens and immune modulation. These beneficial properties have been linked to their carbohydrate metabolism and exopolysaccharide (EPS) biosynthesis capabilities [2, 3].

At the time of the writing (January 2021), over 80 *Bifidobacterium* species and subspecies have been identified in multiple animal hosts, with around 2400 genome assemblies available via the NCBI Genome database [4]. However, the majority of these sequences come from human-associated species, namely; *Bifidobacterium longum*, *Bifidobacterium breve*, *Bifidobacterium bifidum*, and *Bifidobacterium pseudocatenulatum*. Consequently, strains belonging to these species, and single strains that represent type strains (some of which were originally isolated from captive animals), comprise the majority of genomes available, and are thus the best studied with respect to comparative genomics [1, 5–10]. Therefore, our current knowledge of the diversity and evolution of *Bifidobacterium* within the mammalian gut remains limited.

Reports of congruent phylogenies between mammals and gut microbial taxa have suggested that animal hosts and their symbionts may diversify and speciate together [11–14]. Codiversification patterns have been linked to convergent acquisition of function by different bacterial phylogenetic clades, with horizontal gene transfer (HGT) and gene loss proposed as potential mechanisms involved in the process [15, 16]. Furthermore, increased host specificity has been suggested to be linked to reduced transmission capacity due to anaerobic and non-spore forming bacterial lifestyles [17, 18]. As gut microbes can affect host

[1]Gut Microbes & Health, Quadram Institute Biosciences, Norwich Research Park, Norwich, UK. [2]Intestinal Microbiome, ZIEL – Institute for Food & Health, Technical University of Munich, Freising, Germany. [3]Department of Zoology, University of Oxford, Mansfield Road, Oxford, UK. [4]Food Innovation and Health, Quadram Institute Biosciences, Norwich Research Park, Norwich, UK. [5]Nature Research Centre, Vilnius, Lithuania. [6]Department of Pathobiology and Population Sciences, The Royal Veterinary College, Hawkshead Lane, Hatfield, Herfordshire, UK. [7]Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK. ✉email: Lindsay.Hall@quadram.ac.uk

Received: 4 June 2021 Revised: 30 January 2022 Accepted: 9 February 2022
Published online: 25 February 2022

phenotype in a variety of ways, for example by modulating host energy acquisition and immune development [19, 20], understanding and documenting the distribution and diversification patterns of key bacterial species within hosts constitutes an important step towards understanding host-microbe interactions over evolutionary time [21].

In *Bifidobacteriaceae* specifically, phylogenetic congruence between four great ape species and their *Bifidobacteriaceae* gyrB lineages suggested a significant degree of codiversification, albeit with some host switching [11]. Furthermore, the analysis of co-phylogenetic relationships between 24 primate hosts and 23 primate-associated bifidobacterial species revealed the existence of phylogenetic congruence between *Bifidobacterium* typically associated with human hosts (*B. adolescentis*, *B. bifidum*, *B. breve*, *B. catenulatum*, *B. dentium*, *B. longum* spp. and *B. pseudocatenulatum*) and the members of Hominidae family (*Gorilla gorilla*, *Homo sapiens* and *Pan troglodytes*) [22]. In contrast, insights into genotypic and phenotypic properties of animal-associated *Bifidobacterium animalis* and *Bifidobacterium pseudolongum* revealed that strains belonging to these species were found in animal hosts spanning both mammals and birds, and thus appear to be generalist at the species level, rather than having host-specific niches [23, 24]. The widespread distribution of different *Bifidobacterium* species at higher taxonomic ranks of the mammalian tree of life (family and order) has also been suggested based on the analysis of the short internally transcribed spacer (ITS) rRNA sequences [25]. However, since robust strain-level information on the majority of animal-associated *Bifidobacterium* species is lacking, it remains unclear to what extent the proposed generalism holds true at higher taxonomic resolution, and whether highly resolved strains are in fact host specific. This is particularly important when exploring potential adaptation of bacterial symbionts to their hosts in wild populations (rather than model organisms or captive animals), because the adaptive relationship is dependent on the environmental niche, to which the host (and potentially gut microbe) have adapted over evolutionary time scales.

It remains unclear which processes govern the evolution of mammalian gut microbial symbionts [26, 27]. Recently, Groussin et al. [28] proposed that reciprocal and specific functional dependencies between mammalian hosts and single bacterial clades are not strong enough for coevolution to occur and drive cospeciation events. Instead, the authors proposed that allopatric speciation, which implies geographic isolation of the host species and subsequent limited symbiont dispersal and diversification, may lead to host-symbiont cospeciation and phylogenetic congruence patterns. According to this model, host adaptation to new conditions following an allopatric event can result in an altered intestinal environment, to which symbiotic bacteria can then quickly adapt (e.g., different glycan composition of plant-derived substrates in herbivore hosts) [28]. While studies using methods of broad taxonomic resolution (e.g. down to genus- or species level) provide some information on evolutionary relationships between hosts and their gut microbes, studies involving high (strain-level, or genomic) taxonomic resolution in natural systems are lacking, yet may reveal cryptic diversity and patterns of host specificity and host-microbe evolution not previously appreciated, and shed light on the functional role that particular bacterial symbionts play in the wider gut microbiota of wild animals.

Wild small mammals provide an excellent study system in which to explore host–gut microbe relationships in more detail, as they are geographically widespread, diverse, easily trapped and possess a rich gut microbiota. To bridge the current knowledge gap on the distribution of *Bifidobacterium* in wild animals, we chose wild rodents as a target host group in which to profile *Bifidobacterium* species and strain diversity, and determine potential functional adaptation to different host species and geographical regions. To this end, we surveyed wild mice, voles and shrews from multiple populations in two geographically distinct parts of Europe, performed *Bifidobacterium* isolations, and subsequently investigated a collection of derived *Bifidobacterium* genomes. Phylogenomic and functional genomic analysis indicated enrichment for *Bifidobacterium castoris* and particular carbohydrate metabolism and host modulatory properties.

RESULTS

Isolation, sequencing and phylogenomic analysis of recovered *B. castoris* isolates

Between December 2015 and December 2018, we collected and processed 220 faecal samples from 9 species of small mammals (mice, voles and shrews) caught at 14 sites across two European countries—Lithuania and the UK. In the UK, samples were collected from two mouse species (*Apodemus sylvaticus* and *Apodemus flavicollis*) in both Wytham Woods (Oxfordshire, $n = 78$) and Silwood Park (Berkshire, $n = 14$). In Lithuania, 54 samples from mice (*Apodemus agrarius*, *A. flavicollis*), 61 samples from voles (*Microtus agrestis*, *Microtus arvalis, Microtus oeconomus, Myodes glareolus*) and 13 samples from shrews (*Sorex araneus*, *Sorex minutus*, *Neomys fodiens*) were obtained across 12 trapping sites (Supplementary Table S1).

Given the small amount of faecal material obtained, we purposely focused our experimental efforts on isolating *Bifidobacterium* from collected samples, which resulted in the recovery of 51 *Bifidobacterium* isolates from a total of 32 individuals belonging to three wild mouse species—*A. flavicollis, A. sylvaticus*, and *A. agrarius*. *Bifidobacterium* was isolated from 21.9% of mouse samples screened, including 27.0% *A. flavicollis* samples, 26.3% *A. sylvaticus* samples and 6.9% *A. agrarius* samples, respectively. We were not successful in recovering *Bifidobacterium* from voles or shrews. The probability of isolating *Bifidobacterium* varied strongly across host families (Pearson's χ^2 ($df = 2$) $= 18.98$, $P < 0.001$). Whole-genome sequencing of recovered isolates yielded a mean of 265-fold coverage for samples sequenced on HiSeq (minimum 172-fold, maximum 300-fold) and 225-fold for samples sequenced on MiSeq (minimum 130-fold, maximum 325-fold). One sequence did not assemble correctly and was removed from further analysis. Based on the literature, we defined sequences exhibiting the average nucleotide identity (ANI) value > 99.9% as identical [29]. Using this threshold, we excluded further 17 duplicate genomes representing identical isolates from the same individuals sequenced multiple times. This resulted in the final dataset comprised of 33 *Bifidobacterium* genomes representing isolates recovered from 31 individual hosts. In total, we identified 26 isolates as *B. castoris*, 4 isolates as *B. animalis* and a further 3 isolates as *B. pseudolongum* (Supplementary Tables S1 and S2).

The assembled draft genome sizes for mouse-associated *B. castoris* ranged from 2.27 Mb to 2.39 MB, possessing an average G + C% content of 65.53% and a number of contigs ranging from 9 to 56. The number of predicted ORF in each genome ranged from 1832 to 1980. Genome size and gene number were therefore lower in comparison to the only current genome sequence available (January, 2021) for the type strain *B. castoris* 2020B^T (GenBank accession: GCA_003952025.1), isolated from a captive beaver (*Castor fiber*) in Italy, whose genome size was 2.50 Mb, with 2053 ORFs and an average G + C% content of 65.41% [30]. The sizes of draft genomes for *B. animalis* and *B. pseudolongum* ranged from 2.15 to 2.19 Mb (1808 to 1849 ORFs) and 2.03 to 2.06 Mb (1705 to 1725 ORFs), respectively. *B. animalis* strains had an average G + C% content of 60.00%, while this value was at 63.34% for *B. pseudolongum*. These findings are in line with previous reports for members of these species isolated from rodents [23, 24].

In terms of the distribution of recovered *Bifidobacterium* isolates across host species, we isolated *B. castoris* and *B. animalis* from

A. sylvaticus ($n = 19$, UK). Only *B. animalis* isolates were recovered from *A. flavicollis* in the UK ($n = 2$), whereas *B. castoris* and *B. pseudolongum* were successfully isolated from the same host species in Lithuania ($n = 8$). Furthermore, we recovered *B. castoris* from *A. agrarius* ($n = 2$, Lithuania). Based on the previously set ANI threshold (ANI > 99.9%) [29], we identified five *B. castoris* and two *B. animalis* strains in *A. sylvaticus*, five *B. castoris*, two *B. animalis* and three *B. pseudolongum* in *A. flavicollis*, and two *B. castoris* strains in *A. agrarius*. Overall, our isolation efforts resulted in the recovery of 12 *B. castoris*, 4 *B. animalis* and 3 *B. pseudolongum* strains. On average, we recovered one unique *Bifidobacterium* strain per individual, except for one *A. sylvaticus* individual from Wytham (X0418EBC072), from whom both *B. castoris* and *B. animalis* were isolated. However, a much larger isolation and sequencing effort per sample would be required to assess the prevalence and the abundance of multiple species and strains in individual hosts.

Interestingly, all newly sequenced isolates belonged to the previously established *B. pseudolongum* phylogenetic group [31]. Recent taxonomic analyses of the genus *Bifidobacterium* indicated that this phylogenetic group was very diverse in terms of ecological niches represented by host species, and encompassed strains isolated from animals as diverse as chickens, geese, dogs, oxen, pigs, rabbits, hamsters and rats [31]. Since the relatedness of organisms can effectively be predicted based on their shared gene content [32, 33], we constructed a maximum-likelihood phylogenetic tree using single-copy core genes ($n = 610$) to assess relationships between our isolates and representative members of the *B. pseudolongum* group ($n = 112$), with a particular focus on strains isolated from rodents (Fig. 1 and Supplementary Table S1).

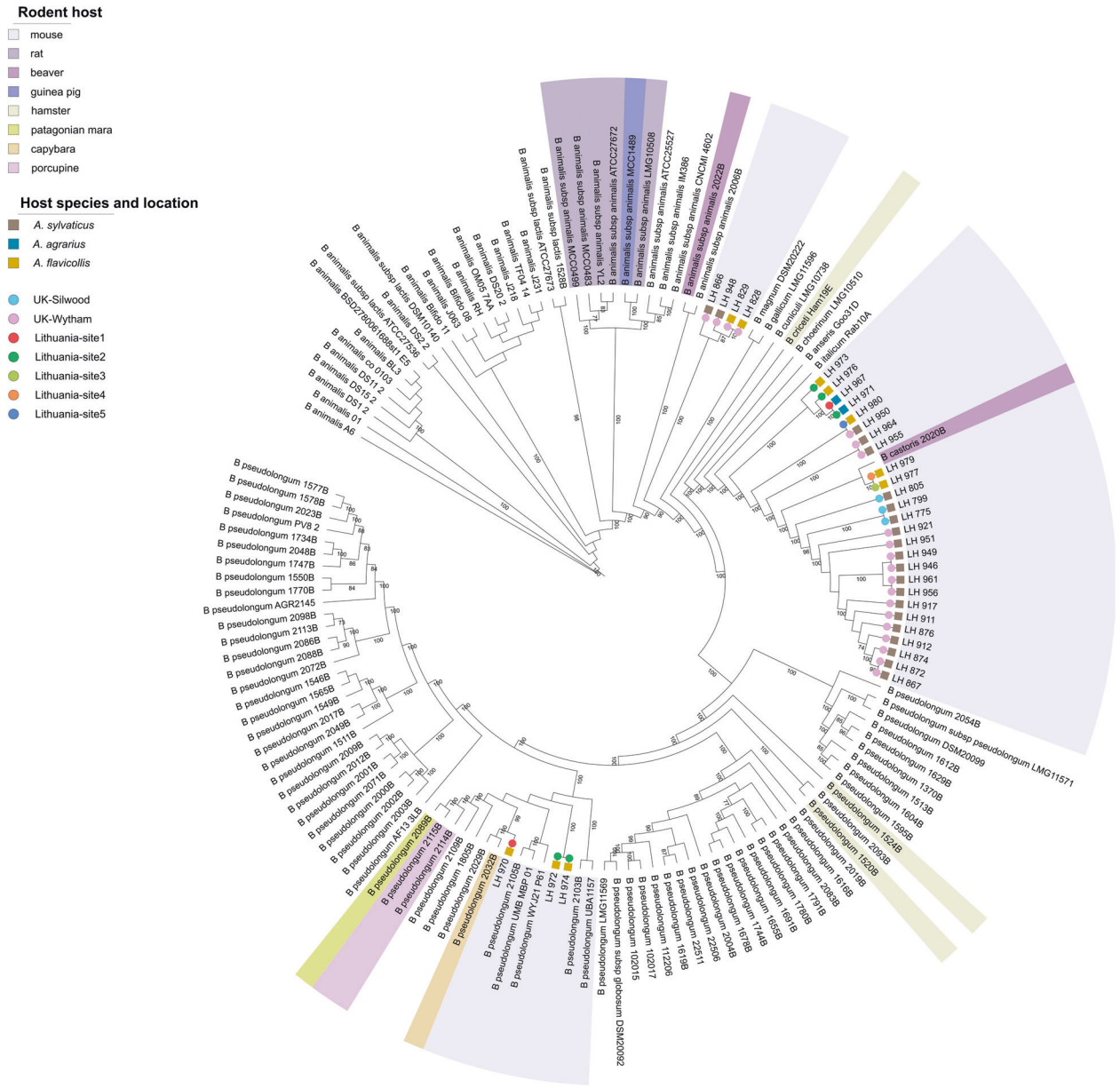

Fig. 1 Cladogram of *B. pseudolongum* phylogenetic group, including 112 publicly available representative strains and the 33 isolates recovered in this study. Maximum-likelihood phylogeny was based on protein sequences of single-copy core genes ($n = 610$), employing the 'WAG' general matrix model with 1000 bootstrap iterations. Bootstrap values above 70% are displayed on tree branches. Strains isolated from rodent hosts are marked with coloured background. Coloured symbols on the branches depict respective host species (square) and trapping sites (circle) for isolates recovered in this study.

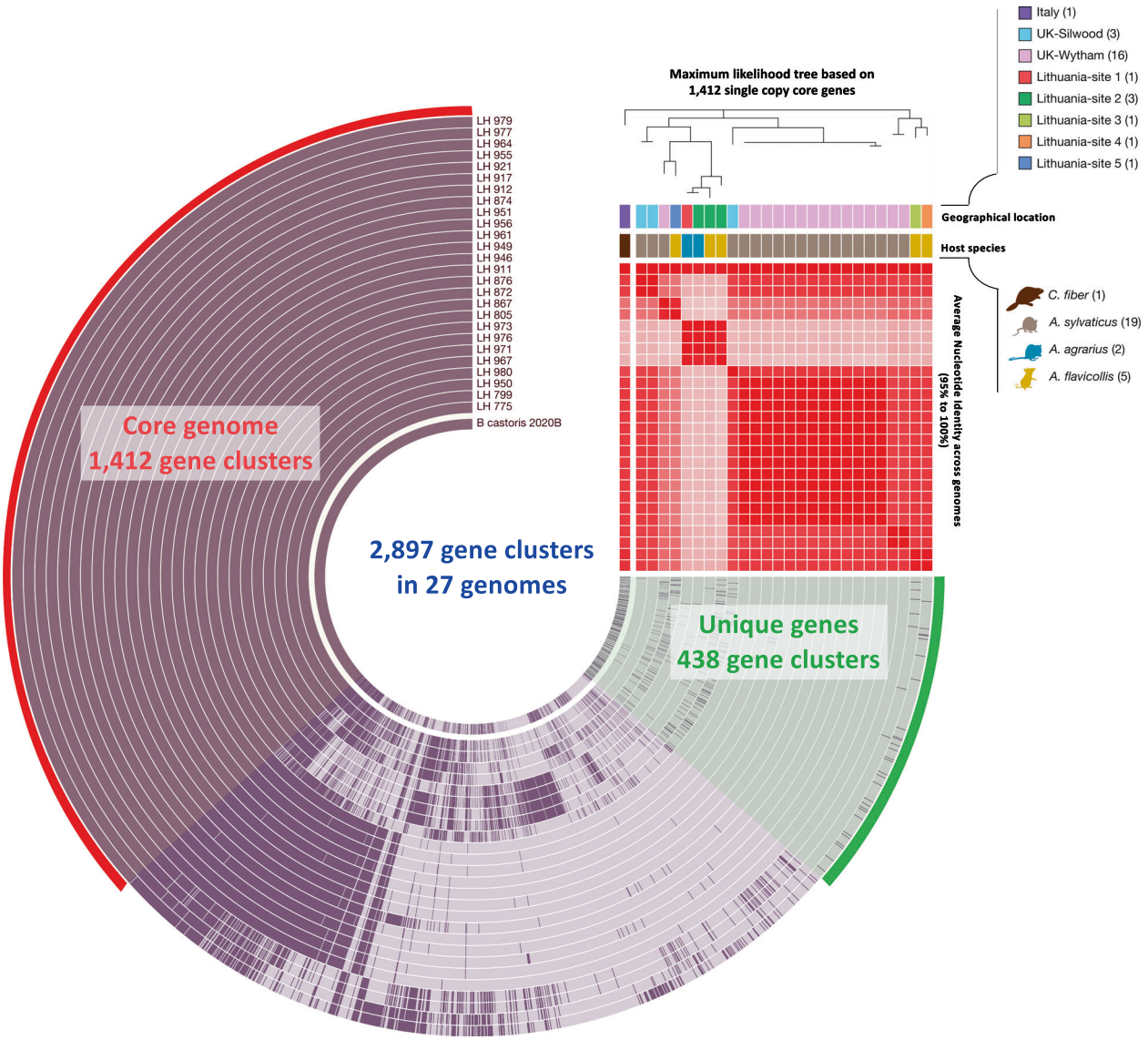

Fig. 2 The pangenome of *B. castoris*. Pangenomic analysis of 27 genomes of *Bifidobacterium castoris* revealed 1412 (48.7% of all clusters) core gene clusters, and 438 (15.1%) strain-specific unique gene clusters among 2897 total gene clusters, along with their distribution and average nucleotide identities (ANI > 95%).

Despite the limited number of rodent-associated *Bifidobacterium* genomes available for this analysis ($n = 20$), results indicate some clustering of strains according to host phylogeny. For example, while *B. animalis* isolates recovered from mice tend to cluster separately from those from rats, *B. pseudolongum* isolates from a porcupine and a patagonian mara cluster together. This observation is in line with previous genomic analysis of animal-associated *B. pseudolongum* isolates, which indicated that different animal hosts harbour specific clusters of members of this taxon [24].

Since *B. castoris* constituted 78% of all *Bifidobacterium* isolates recovered in this study, and this species is the least well-characterised, we focused further detailed genomic analyses on this species. The pangenomic analysis of these isolates alongside the type strain previously isolated from a beaver revealed a total of 2897 gene clusters (Fig. 2). Based on the distribution of gene clusters in the pangenome, we identified 1412 gene clusters that constituted the core genome shared by all isolates (48.7% of all clusters), while 438 clusters (15.1% of all clusters) were unique genes (Supplementary Table S3). Using protein sequences for single-copy core genes of the pangenome we constructed a *B. castoris* phylogeny based on maximum-likelihood estimation (Figs. 2 and 3a).

Examination of the *B. castoris* phylogenetic tree (Fig. 3a), revealed that the 12 mouse-isolated strains fell into three major clusters. One cluster is more distant from the other two and contains strains from both *A. sylvaticus* and *A. flavicollis*. The second cluster seems to be *A. sylvaticus*-specific and appears to contain 2 main strains, while the third main cluster contains strains isolated from all three *Apodemus* species, including one *A. sylvaticus*-specific subclade, and another subclade that contains two clusters of strains that each colonise two different host species. Overall, these observations are not consistent with a strong pattern of cospeciation. Therefore, we sought to test for co-phylogenetic signal between *Bifidobacterium* isolates and their hosts [34]. The host tree was constructed using concatenated sequences for part of the cytochrome b (cytb) gene and the mitochondrial 12S rRNA gene (Fig. 3b) [35]. Cophylogeny was first tested for using the global ParaFit statistic (H0: *B. castoris* and its hosts have independent phylogenetic structure). The result (permutational $P = 0.3586$ after 9999 permutations) was not

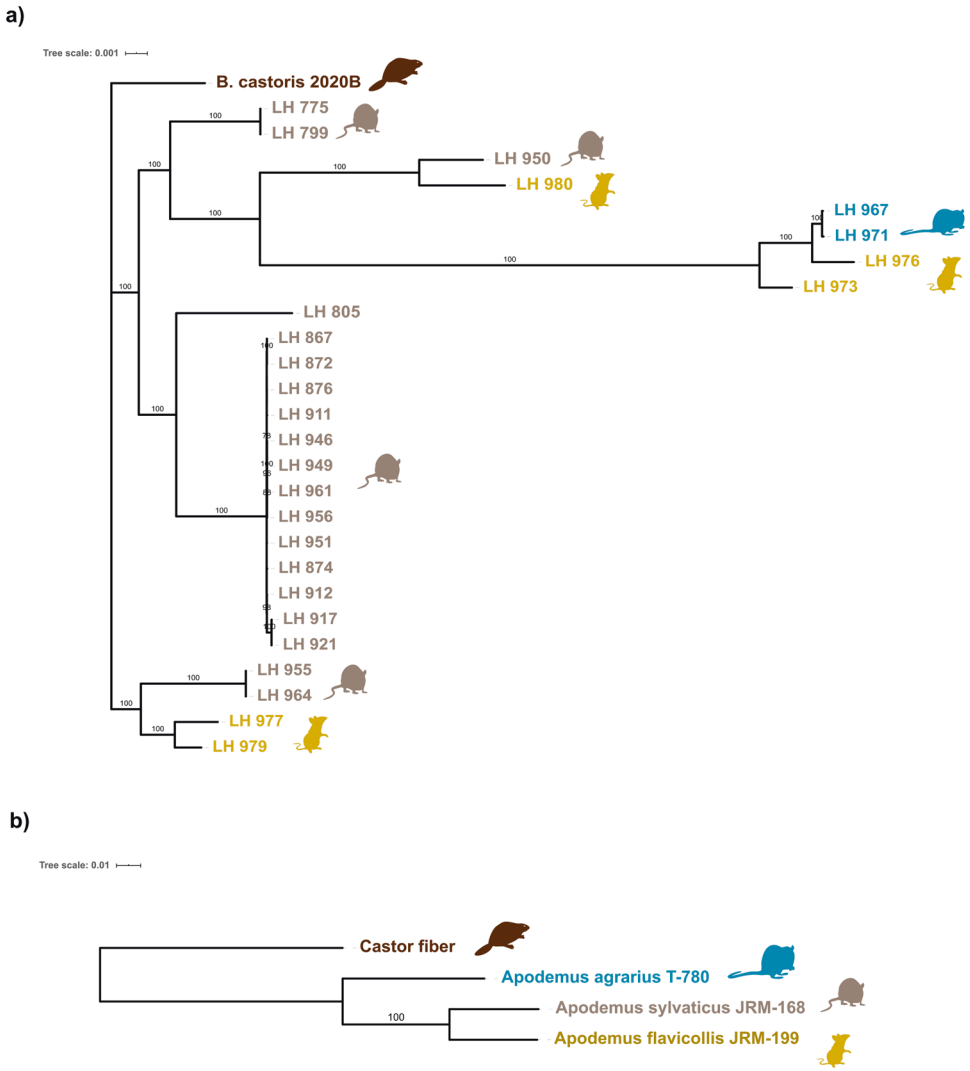

Fig. 3 Phylogenetic analyses of *B. castoris* and their hosts. Phylogeny of 27 *Bifidobacterium castoris* isolates (**a**) and their rodent hosts (**b**). Maximum-likelihood trees were constructed using single copy core genes employing 'WAG' model and 1000 bootstrap iterations for *B. castoris* and concatenated 12S rRNA and partial cytochrome b genes employing 'GTR'model with 1000 bootstrap iterations for host species. Bootstrap values above 70% are displayed on tree branches.

significant, failing to reject the null hypothesis, and thus providing no evidence that the phylogeny of *B. castoris* and its hosts are correlated. Further ParaFit test of associations between individual *Bifidobacterium* isolates and their respective hosts did not reveal any significant links ($P > 0.05$, Table 1 and Supplementary Table S4), suggesting the absence of co-phylogenetic patterns.

With five *B. castoris* strains isolated from *A. sylvaticus*, five from *A. flavicollis*, and two from *A. agrarius*, we next sought to determine their distribution across geographical locations. Overall, a total of 5 *B. castoris* strains were recovered from the two UK sites: three distinct strains from samples collected in Wytham, with one strain more common than the others (detected 13 times), and two strains from samples collected at Silwood Park. In Lithuania, only one site showed evidence of more than one strain (two strains recovered from *A. flavicollis* and one strain isolated from *A. agrarius* at site2). No strains were recovered from more than one murine host species nor multiple locations. For the only strain that we detected fairly regularly (Strain 3 LH_867-LH_961, found 13 times in *A. sylvaticus* in Wytham), we tested whether prevalence varied by host species or geographical location. This strain's occurrence differed significantly across host species and countries (which are confounded), as it was only detected in UK wood mice (Fisher's exact tests for both host species and country, $p = 0.005$). Altogether, these results suggest that *B. castoris* strains display a certain degree of species- and site-specificity.

Characterisation of the glycobiome of *B. castoris*

To determine functional differences between *B. castoris* isolates, we next functionally annotated ORFs of each genome based on orthology assignment using eggNOG-mapper. This analysis resulted in the classification of an average of 82.46% genes per genome into COG categories, and reflected the saccharolytic lifestyle of *B. castoris*, with carbohydrate transport and metabolism identified as the second most abundant category (after unknown function) constituting 9.98% of functionally annotated genes. This value is slightly higher compared to previous findings for the pangenome of animal-associated *B. pseudolongum* taxon (9%) and within the range reported for other bifidobacteria [24, 36] (Supplementary Table S5). Members of *Bifidobacterium* have been shown to synthesise and digest a wide range of carbohydrates through an extensive arsenal of carbohydrate-active enzymes (CAZymes) [37, 38]. We thus sought to investigate the genetic repertoire predicted to be involved in carbohydrate metabolism and biosynthesis in *B. castoris*. *In silico* analyses performed using

Table 1. Results of cophylogeny test using the global ParaFit statistic (H0: *B. castoris* and its hosts have independent phylogenetic structure) and the ParaFit test of associations between individual *Bifidobacterium* isolates and their respective hosts.

Global test: ParaFitGlobal = 8.835461e-05, *p* value = 0.3586 (9999 permutations)						
Test of individual host-parasite links (9999 permutations):						
	Host	***B. castoris* isolate**	**F1.stat**	**p.F1**	**F2.stat**	**p.F2**
[26,]	*A. agrarius*	LH_967	2.66E−05	0.3073	9.07E−03	0.3068
[27,]	*A. agrarius*	LH_971	2.66E−05	0.3064	9.08E−03	0.3064
[21,]	*A. flavicollis*	LH_980	−6.22E−07	0.7765	−2.12E−04	0.7765
[22,]	*A. flavicollis*	LH_976	−1.76E−05	0.808	−5.99E−03	0.8093
[23,]	*A. flavicollis*	LH_973	−1.52E−05	0.7834	−5.17E−03	0.7849
[24,]	*A. flavicollis*	LH_977	−2.70E−08	0.5686	-9.20E−06	0.5686
[25,]	*A. flavicollis*	LH_979	7.89E−09	0.5594	2.69E−06	0.5594
[2,]	*A. sylvaticus*	LH_775	1.14E−06	0.4388	3.90E−04	0.4365
[3,]	*A. sylvaticus*	LH_799	1.14E−06	0.4455	3.90E−04	0.4446
[4,]	*A. sylvaticus*	LH_950	−3.04E−06	0.8379	-1.04E−03	0.8376
[5,]	*A. sylvaticus*	LH_805	6.00E−06	0.2779	2.04E−03	0.2764
[6,]	*A. sylvaticus*	LH_867	6.98E−06	0.2134	2.38E−03	0.2133
[7,]	*A. sylvaticus*	LH_872	6.99E−06	0.2109	2.38E−03	0.2114
[8,]	*A. sylvaticus*	LH_876	6.98E−06	0.2105	2.38E−03	0.211
[9,]	*A. sylvaticus*	LH_911	6.99E−06	0.2112	2.38E−03	0.2117
[10,]	*A. sylvaticus*	LH_946	6.99E−06	0.212	2.38E−03	0.2125
[11,]	*A. sylvaticus*	LH_949	6.99E−06	0.2096	2.38E−03	0.2097
[12,]	*A. sylvaticus*	LH_961	6.99E−06	0.2187	2.38E−03	0.2189
[13,]	*A. sylvaticus*	LH_956	6.99E−06	0.2164	2.38E−03	0.2166
[14,]	*A. sylvaticus*	LH_951	6.99E−06	0.2112	2.38E−03	0.2116
[15,]	*A. sylvaticus*	LH_874	7.00E−06	0.2105	2.39E−03	0.2102
[16,]	*A. sylvaticus*	LH_912	6.99E−06	0.2121	2.38E−03	0.2131
[17,]	*A. sylvaticus*	LH_917	7.21E−06	0.2088	2.46E−03	0.2093
[18,]	*A. sylvaticus*	LH_921	7.21E−06	0.214	2.46E−03	0.2135
[19,]	*A. sylvaticus*	LH_955	3.41E−06	0.4128	1.16E−03	0.4104
[20,]	*A. sylvaticus*	LH_964	3.41E−06	0.4074	1.16E−03	0.4049
[1,]	*B. castoris*	B_castoris_2020B	2.75E−06	0.3883	9.36E−04	0.3854

dbCAN2 identified three classes of enzymes, namely glycosyl hydrolases (GHs), glycosyl transferases (GTs) and carbohydrate esterases (CEs), as well as enzyme-associated carbohydrate-binding modules (CBMs) (Fig. 4a and Supplementary Table S6). On average, *B. castoris* genomes harboured 86.52 ± 3.15 CAZymes. Previous reports on CAZyme abundances in strains isolated from different hosts and environments showed that, on average, *Bifidobacterium* isolated from rodents had less than 50 CAZyme genes in their genomes, a number comparable with strains isolated from dairy and wastewater [39]. Interestingly, the abundance of CAZymes similar to that of *B. castoris* was reported for strains isolated from non-human primates (84 ± 10 CAZymes) [39].

Glycosyl hydrolases are key enzymes in carbohydrate metabolism that catalyse the hydrolysis of glycosidic bonds between two or more carbohydrates or between a carbohydrate and non-carbohydrate moiety [40]. We identified a total of 25 different GH families in *B. castoris* isolates containing an average of 49.48±1.99 GH genes per genome (2.62 ± 0.16% of ORFs and 57.68 ± 0.01% of predicted glycobiome). The predominant GH family, with 14.15 ± 1.29 GH genes per genome (mean ± sd), constituting on average 29.98 ± 1.88% of the GH repertoire, was GH13 whose members include enzymes acting on a very wide range of carbohydrates containing α-glucoside linkages, e.g. starches and related substrates, trehalose, raffinose, stachyose and melibiose [9, 38, 41]. These results were consistent with previous studies of the type strains representative of the genus *Bifidobacterium*, as well as strains belonging to *B. pseudolongum* species, which identified this particular GH family as the most commonly detected, in particular in the genomes of strains isolated from mammals [9, 24]. Families GH31 (a diverse group of enzymes with α-glucosidase and α-xylosidase activities) and GH36 (enzymes metabolising α-galacto-oligosaccharides present in various plants, i.e. melibiose, raffinose, stachyose) [42, 43] followed, with 3.63±0.68 and 3.18±0.39 GH genes per genome, respectively.

It has been well established that the GH repertoires of *Bifidobacterium* are species- and often strain specific [9, 23, 24, 44]. Therefore, we expected that the diverging *B. castoris* strains would also harbour diverging GH profiles. Indeed, at every geographical location, the isolates identified as unique strains also displayed individual GH profiles (Fig. 4a, b, and Supplementary Table S6). However, we noted one instance of an inconsistent prediction of the number of genes belonging to GH13 family between the two isolates identified as Strain 1 from Silwood (13 vs 15 GH genes in LH_775 and LH_799, respectively), with identical values for all other identified GH families. This result may possibly be explained by the differences in the ORF calling and subsequent annotation between the two genomes, resulting from differences in the number of contigs in their assemblies.

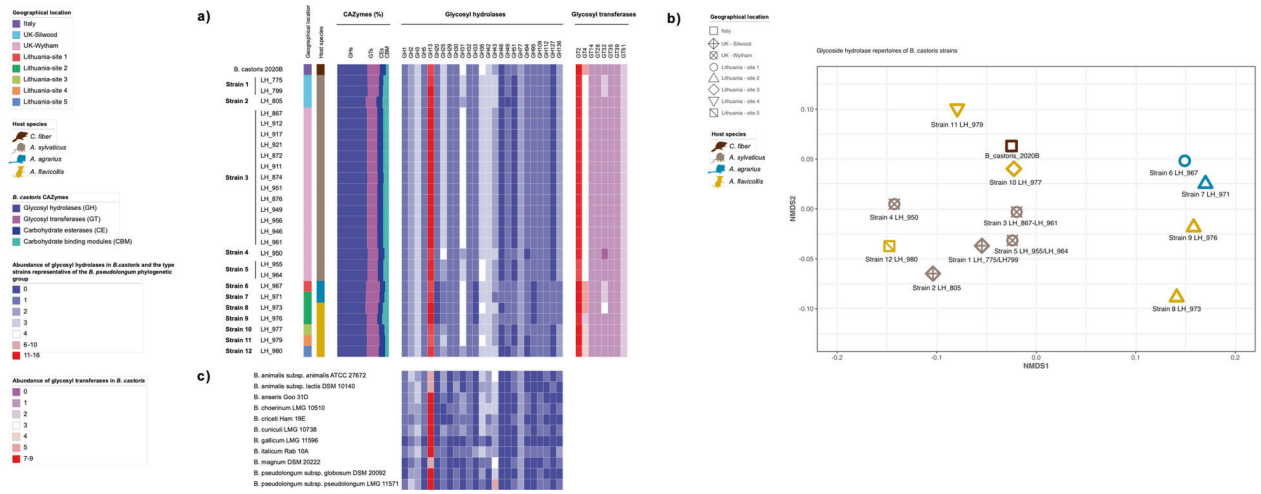

Fig. 4 Characterisation of the glycobiome of *B. castoris*. Predicted glycobiome of *B. castoris* (**a**), clustering of GH profiles of *B. castoris* strains using non-metric multidimensional scaling (NMDS) (**b**) and the abundance of glycosyl hydrolases in *Bifidobacterium* type strains representative of the *B. pseudolongum* phylogenetic group (**c**). **a** Coloured strips indicate geographical location and host species. The bar plot (CAZymes %) shows proportional representation of four CAZyme classes constituting the predicted *B. castoris* glycobiome. The two heat maps present abundances of genes predicted to belong to specific glycosyl hydrolase (GH - blue) and glycosyl transferase (GT- purple) families. **b** The heat maps show abundances of genes predicted to belong to the same glycosyl hydrolase families as those identified in *B. castoris*.

The comparison between GH profiles of *B. castoris* (Fig. 4a and Supplementary Table S6) and those of 9 *Bifidobacterium* type strains representative of the members of the *B. pseudolongum* phylogenetic group (Fig. 4c and Supplementary Table S6) revealed that *B. castoris* species alone appears to possess GH49 family, which contains dextranases acting on dextran and pullulan [45]. The majority of *B. castoris* strains isolated from hosts across all trapping sites appear to be lacking families GH51 and GH127, but possess several copies of genes predicted to encode enzymes belonging to GH43 family. These GH families predominantly contain α- and β-L-arabinofuranosidases that hydrolyse the glycosidic bond between L-arabinofuranoside side chains of hemicelluloses such as arabinoxylan, arabinogalactan and L-arabinan, naturally present in cereal grains [46, 47]. The exception are 4 strains isolated from *A. agrarius* and *A. flavicollis* from site1 and site2 in Lithuania, 3 out of which appear to additionally possess predicted GH46 chitosanases acting on chitin-derived substrates (i.e. mushrooms, fungi and insects) [48] and lack GH20, GH33 and GH95. The latter GH families, only detected in *B. castoris* and its closest relative *B. italicum*, contain enzymes previously associated with degradation of host-derived carbohydrates in human-associated *Bifidobacterium*: lacto-*N*-biosidases, exo-sialidases and α-L-fucosidases, reported to be involved in metabolism of specific oligosaccharides, including human milk oligosaccharides present in maternal breast milk and intestinal glycoconjugates [49–52]. Furthermore, the results of the multilevel pattern analysis identified the presence of GH46 chitosanases as the factor driving the differences in GH profiles between strains isolated from *A. agrarius* and the remaining *B. castoris* strains (association index = 0.8819171, $P < 0.05$) (Supplementary Table S7). However, only two strains were recovered from this host species, and this small sample size could explain these results. Overall, these findings highlight a predominance of genes encoding GH families predicted to be responsible for the breakdown of plant-derived polysaccharides in the genomes of *B. castoris* species, and shed light on potential evolutionary adaptations of *B. castoris* strains to the host diet, however with only one non-mouse strain available for the analysis, interpretations can only be tentative.

To validate the results of our in silico analysis, we selected normal maize starch (NMS) and chitosan as substrates for *in vitro* growth experiments involving representatives of the 12 *B. castoris* strains identified. All tested isolates could apparently metabolise NMS, however no growth was observed with chitosan (Supplementary Fig. 1).

In silico assessment of the EPS production potential of *B. castoris* isolates

The GT class of enzymes catalyse the formation of glycosides involved in the biosynthesis of oligosaccharides, polysaccharides and glycoconjugates [53] and have previously been associated with production of EPS in different bacterial species [54]. A total of 8 GT families were predicted in *B. castoris* genomes, with 18.70 ± 1.61 GT genes on average (21.41 ± 0.02% of predicted glycobiome). GT2 family was predominant in all analysed strains, with an average of 8.15 ± 0.53 GT genes per genome. Carbohydrate esterases, whose function is to release acyl or alkyl groups attached by ester linkage to carbohydrates [55], and carbohydrate-binding modules, which have no hydrolytic activity, but bind to carbohydrate ligands and enhance the catalytic efficiency of carbohydrate-active enzymes [55], constituted 10.66 ± 0.01% and 10.24 ± 0.02% of predicted glycobiome, with 9.33 ± 1.00 and 9.00 ± 1.75 genes per genome, respectively (Fig. 4a and Supplementary Table S6).

Given the prediction of GT genes in the *B. castoris* glycobiome, we next examined our collection of genomes for the presence of genes potentially involved in EPS biosynthesis, as previous studies have indicated that EPS may support gut colonisation and stimulate host immune responses [3]. Recently, relatively conserved genomic regions predicted to contain genes involved in EPS production have been identified in *Bifidobacterium* type strains, including members of *B. pseudolongum* phylogenetic group (gene clusters *eps*3 and *eps*4). For our search, we selected amino acid sequences of *eps* gene clusters from *B. animalis* subsp. *lactis* BI12 (*eps*3: BI12_1287 – BI12_1328) and *B. pseudolongum* subsp. *globosum* LMG 11569^T (*eps*4: BPSG_1548 – BPSG_1565) as references [3].

This analysis identified homologues of several conserved *eps*-key genes in all analysed *B. castoris* genomes (Fig. 5a and Supplementary Table S8), including those predicted to encode the priming glycosyl transferase (pGTF), which catalyses the first step in EPS biosynthesis, as well as enzymes putatively involved in rhamnose biosynthesis and the transport of the formed EPS-unit

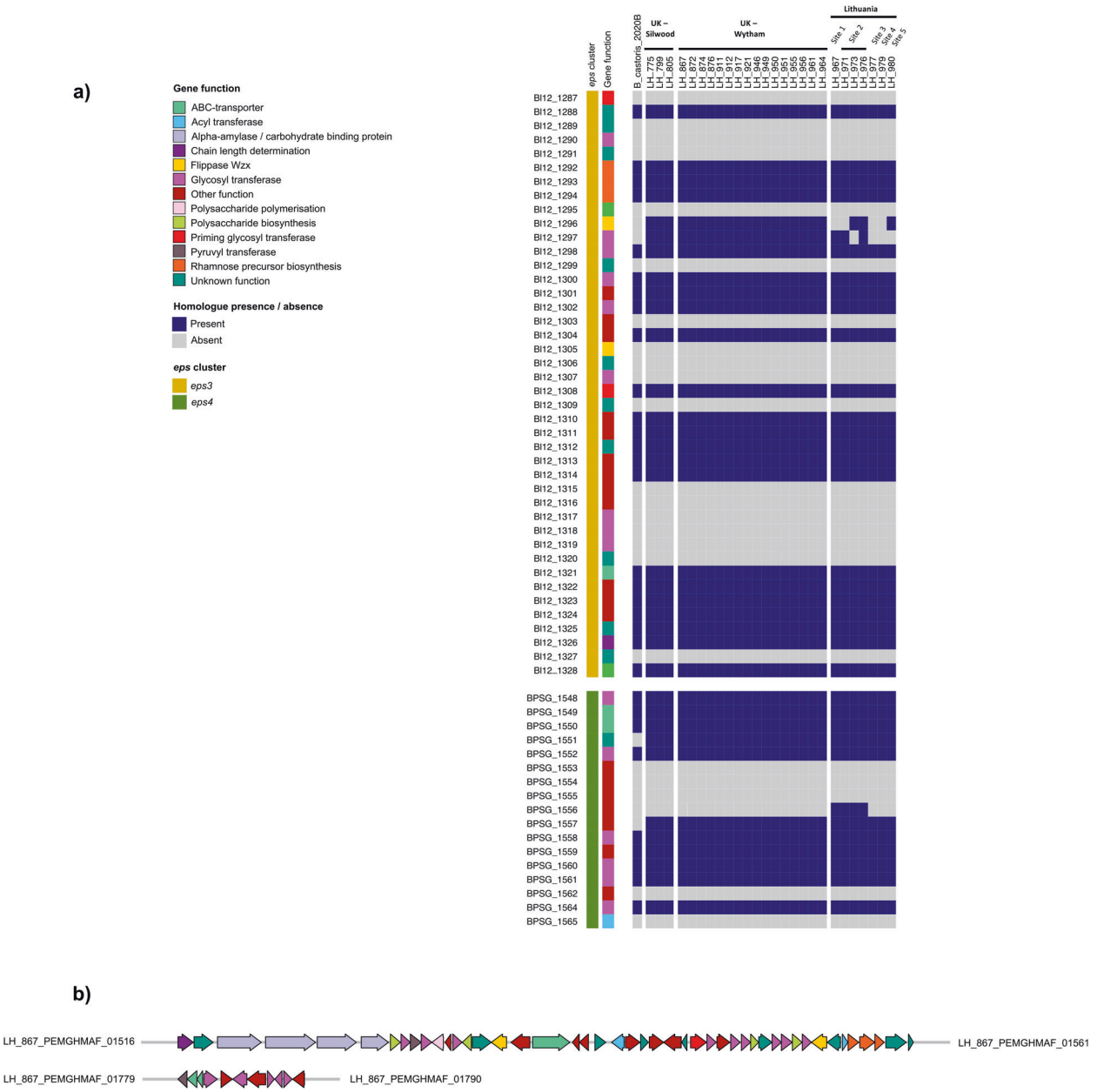

Fig. 5 Identification of putative *eps* loci in *B. castoris* genomes. Panel (**a**) shows amino acid homology between *B. castoris* sequences and previously identified *Bifidobacterium* eps-key components. Sequences of *B. animalis* subsp. *lactis* BI12 (accession number CP004053.1, eps3: BI12_1287 – BI12_1328) and *B. pseudolongum* subsp. *globosum* LMG 11569^{T} (accession number JGZG01000015.1, eps4: BPSG_1548 – BPSG_1565) were used as reference. Panel (**b**) depicts a proposed map of putative *B. castoris eps* loci constructed for isolate LH_867, representative of identified Strain 3. Gene functions were predicted based on BLASTP searches against the NCBI non-redundant protein database and results generated with dbCAN2.

across the cytoplasmic membrane (either an ABC-type transporter or a "flippase"-like protein). These results confirm that *B. castoris* strains harbour putative *eps*-key genes and suggest potential ability for this species to produce EPS. Further searches against the NCBI database coupled with additional analysis of results obtained from dbCAN2 allowed for the construction of gene maps of the putative *B. castoris eps* gene clusters (Fig. 5b).

Previously, potential horizontal transfer of *eps* clusters in *Bifidobacterium* has been suggested [3]. Thus, given the identification of putative *eps*-key genes in our collection of genomes, we next sought to determine the role HGT has played in the evolution of *B. castoris* strains. For this purpose, we used the SIGI-HMM tool implemented in software IslandViewer4. This analysis revealed that on average, 6.43 ± 1.15 % of ORFs in *B. castoris* genomes were predicted to be horizontally acquired (range 59–191 genes per genome). This value is lower than that previously reported for *B. pseudolongum* species (average of 11.1%), and similar to those reported for *B. animalis* subsp. *lactis* (5.1%) and *B. animalis* subsp. *animalis* (4.6%) [23, 24]. Cross-referencing with the results of the functional analysis revealed that the eggNOG annotation was available for 49.22 ± 5.69% of putative horizontally acquired genes per genome, on average (Supplementary Fig. 2 and Supplementary Table S9). The highest proportion of annotated HGT genes in each genome were those of unknown function (38.46 ± 4.36% on

average), followed by genes involved in replication, recombination and repair (an average of 19.25 ± 4.43%). This category encompassed CRISPR-Cas-associated proteins, transposases and DNA methylases and methyltransferases. Further analysis of the HGT predictions revealed that genes involved in the cell wall/membrane/envelope biosynthesis constituted on average, 11.35 ± 4.35% of annotated putative HGT genes per genome. This group contained genes neighbouring those identified by our BLAST+ analysis as putative *eps* genes, suggesting they might also be part of *B. castoris eps* clusters (Supplementary Tables S8 and S9).

Moreover, genes identified as involved in carbohydrate transport and metabolism constituted on average 3.34 ± 2.10% of predicted annotated horizontally acquired genes per genome. Further analysis of the results for this group revealed that 5 strains from across all locations (the two strains from Silwood Park, one strain from Wytham Woods (LH_950) and two strains from Lithuania (LH_973 and LH_980)) might have acquired a GH43 family member annotated as α-L-arabinofuranosidase through an HGT event. Additionally, strains LH_955 and LH_964 from Wytham Woods, as well as strains LH_973 and LH_979 from Lithuania were predicted to horizontally acquire a GH36 family α-galactosidase (Supplementary Table S9). Overall, these results suggest that HGT may have contributed to the evolution of *B. castoris* strains and their glycobiome, however experimental validation would be essential to confirm the functional importance of these events.

DISCUSSION

This study is the first to explore strain-level genomic signatures of the beneficial bacterial symbiont *Bifidobacterium* in wild rodent populations within an evolutionary and ecological framework. Isolation, whole-genome sequencing and in-depth phylogenetic and functional genomic analysis indicates that *B. castoris* appears to be a resident microbiota member of wild rodents belonging to genus *Apodemus*, irrespective of geographical location or host species, with presence of key carbohydrate degradation clusters revealing potential diet-host-microbe evolutionary adaptions.

Our isolation efforts enabled recovery of *B. castoris*, *B. animalis* and *B. pseudolongum* from three species of wild mouse in the genus *Apodemus* (*A. sylvaticus*, *A. flavicollis* and *A. agrarius*) from two European countries, UK and Lithuania. Interestingly, despite testing roughly equal numbers of mice and voles captured at the same sites in Lithuania, we did not isolate *Bifidobacterium* from voles. Our inability to detect *Bifidobacterium* in voles could reflect either its absence in these hosts, presence at low abundance precluding isolation, or potential low affinity of resident strains with the implemented isolation methodology. Patterns of *Bifidobacterium* diversity in small mammals could be profiled in future studies through amplicon or metagenomic approaches, such as 16S rRNA amplicon sequencing or shotgun metagenomics. Although, a recent study that reconstructed metagenome-assembled genomes (MAGs) from faecal metagenomics data from 20 bank voles (*M. glareolus*) also indicated an absence of *Bifidobacterium* [56]. However, given the very particular trapping location (Chernobyl Exclusion Zone in Ukraine), it is difficult to speculate on how representative these findings are of vole populations in areas not contaminated with radiation.

B. castoris distribution across the host phylogeny is consistent with host shifts

Since host specificity represents an avenue of evolutionary development that allows a microbe to colonise a specific host, studying the patterns of host specificity is important for the understanding of the complexities of coevolution between two organisms [57]. While the widespread distribution of different *Bifidobacterium* species in mammals has previously been postulated, some species seem to show higher host specificity than others. For example, *Bifidobacterium tissieri* and *Bifidobacterium saeculare* have so far only been isolated from primates (based on bifidobacterial ITS rRNA sequences) [25]. In contrast, *B. animalis* and *B. pseudolongum* had previously been isolated from a number of animal hosts, including mammals and birds, and at the species-level are considered host generalists [23, 24]. *B. castoris* had previously only been isolated from beavers, but here we show it is also regularly detected in wild mice, which suggests that this species might in fact colonise a broad range of rodents. However additional isolation efforts would be required to confirm its true host range.

Previous research has shown a strong pattern of cophylogeny between some mammalian hosts and *Bifidobacteriaeceae*. Moeller et al. [11] showed tight congruence between the phylogenies of *Bifidobacteriaeceae* and their hominid hosts, providing support for codiversification. Similarly, a more recent genome-level co-phylogenetic analysis of primate-associated *Bifidobacterium* and their hosts revealed the existence of phylogenetic congruence between the type strains of *Bifidobacterium* typically associated with human hosts and the three members of *Hominidae* family (*G. gorilla*, *H. sapiens* and *P. troglodytes*) [22]. Here, we test whether a similar pattern holds true based on the analysis of genomic sequences of isolates from a single *Bifidobacterium* species (*B. castoris*) and their wild murine hosts. Moreover, as the species of *Apodemus* are closely related, this provides a different evolutionary perceptive when compared to studies focusing on more distantly related primate species [35, 58]. In contrast to previous work [11, 22], we found no statistical congruence between the *B. castoris* and host phylogenies. *B. castoris* isolates did not show clear phylogenetic clustering by host species or geographic region. Despite this, all *B. castoris* strains identified were only found in a single mouse species, suggesting strains may be host species specific, but more extensive sampling would be needed to test this definitively. Our data are more consistent with regular host shifts by *B. castoris* strains, and if codiversification has in fact occurred its signature has been eroded by subsequent host shifts.

An interesting question is what might drive the contrasting host-*Bifidobacterium* evolutionary patterns in hominids and mice. We speculate that differences in host evolutionary history and ecology may be important. One possibility is that the level of host-microbe codiversification could differ between members of these two mammalian families. For example, if dietary divergence and associated microbial selection were to be stronger among speciating hominids than *Apodemus*, this could drive stronger codiversification in hominids than mice. Previous analysis of the dentition of fossil apes (over 15 my old) indicated that they were primarily frugivorous, and to this day herbivorous—and largely fruit-based—diet dominates among the lesser apes and hominids [59]. Murine rodents diversified in Europe around 10 mya from the generalist *Progonomys* that later evolved into lineages related to *Apodemus* [60]. Based on the analysis of the dental pattern of the *Apodemus* mice, it has been suggested that this taxon displayed a relative morphological stability consistent with stabilising selection and retained a primitive, largely granivorous diet over its evolutionary time. The more diversified primate diet, compared to that of *Apodemus* species, could perhaps explain the higher *Bifidobacterium-Hominidae* specificity.

Alternatively, even if hominids and mice experienced similar codiversification with bifidobacteria initially, variation in the potential for subsequent host shifts may be important. While speciation was likely to have been allopatric for both groups [11, 61, 62], allopatry may not have persisted among nascent *Apodemus* species for very long, with higher post-speciation contact among *Apodemus* species allowing more frequent *Bifidobacterium* transfer and host shifts. Hominid species show very strongly bounded present-day geographical ranges [63], whereas *Apodemus* species in Europe have very broad overlapping ranges, and different species (e.g. *A. flavicollis* and *A. agrarius* at

our Lithuanian sites) can often be caught in adjacent traps. Thus, earlier and more extensive contact between *Apodemus* mouse species may have allowed more cross-species transmission and host shifts of *B. castoris* strains over evolutionary time than could have occurred among hominid species. Future work testing the generality of codiversification across hosts groups with different speciation patterns would be highly informative to understand how the biogeographic and temporal patterns of speciation may affect evolutionary patterns in host-symbiont relationships [28].

B. castoris genomes encode an extensive arsenal of carbohydrate-active enzymes

Characterisation of the *B. castoris* glycobiome provided insight into the strain-specific genetic repertoire predicted to be involved in carbohydrate metabolism and synthesis. The data indicated that *B. castoris* is predominantly enriched in GH families implicated in the degradation of plant-derived carbohydrates. This finding is consistent with the largely plant-based (granivorous) diet of *Apodemus* mice, including as found at several of our sites in UK and Lithuania [64–67]. Indeed, our in vitro experiments confirmed that all strains could metabolise NMS, native to maize grains.

Previously, several bifidobacterial arabinofuranosidases that belong to families GH43, GH51 and GH127 and act on arabinose-substituted polysaccharides have been structurally and functionally identified in *Bifidobacterium* [68–71]. Our analysis of the GH profiles in *B. castoris* suggested that most strains isolated from mouse hosts lack family GH51 and GH127. Moreover, the data indicate that all *B. castoris* genomes harbour 2–3 copies of GH43 arabinofuranosidases and that some of these copies seemed to have been acquired via an HGT event. These findings suggest there might be an evolutionary advantage for *B. castoris* in possessing GH43 arabinofuranosidases over those belonging to families GH51 and GH127, which may be linked to the composition of the host diet. However, it is difficult to speculate on biological significance of these results without the supporting experimental data. It has previously been shown that arabinofuranosidases characterised in *B. adolescentis* belonging to different families (GH43 vs. GH51) display variation in substrate specificity, while those belonging to the same family (GH43) can act in a synergistic fashion, which may link to enhanced energy harvest from dietary components associated with low bioavailability [70].

Interestingly, three Lithuanian strains isolated from *A. agrarius* and *A. flavicollis* from two distinct trapping sites (site1 and site2) possess genes encoding predicted chitosanases (GH46). The presence of such genes may reflect nutrient availability or dietary preferences of their animal hosts, though no data on the diet of mice analysed in this study were available. Nonetheless, stomach contents analysis has previously detected fungi as a dietary item of *A. flavicollis* in Lithuania at sites close to those studied here [64] as well as in one of our UK sampling sites [65], and spores of putatively edible fungi (with macroscopic fruiting bodies) were frequently detected in the faeces *Apodemus* spp. in other parts of Lithuania [72], suggesting mycophagy in *Apodemus* spp. may not be uncommon. The potential of *Bifidobacterium* to degrade chitin-derived substrates is currently not very well understood. Previously, studies looking into functional effects of chito-oligosaccharides on *Bifidobacterium* produced inconsistent results [73–76]. We observed no *B. castoris* growth in medium supplemented with chitosan, however given its alkaline nature and the viscosity of solutions prepared with this substrate [77, 78], further work on assay optimisation would be required to assess its bioavailability and metabolism.

B. castoris isolates show potential for EPS production

The identification of genes predicted to encode GTs in genomes of *B. castoris* prompted questions about potential ability of members of this species to produce EPS. This bifidobacteria trait is associated with immune modulation and longer-term persistence in the (laboratory) murine gastrointestinal tract, suggesting EPS may also play a key role in microbe-host interactions in wild mice populations [79–81]. Recent genomic studies on *Bifidobacterium* type strains have described high levels of inter-species variation with respect to the number, function and organisation of genes in *Bifidobacterium eps* clusters [3, 54]. The results of our BLASTP search for the homologues of *eps*-key genes previously identified in *Bifidobacterium* type strains most closely related to *B. castoris* 2020B^T—including pGTFs, transporter enzymes and carbohydrate precursor biosynthesis enzymes—revealed their presence in *B. castoris* isolates, suggesting this species may be able to synthesise EPS. Furthermore, the analysis of predicted HGT events in our isolates identified additional genes of unknown function neighbouring the *eps*-key genes that may be part of a distinct *B. castoris eps* cluster. These findings support previous suggestions on possible role of HGT in acquisition of complete or partial *eps* clusters by *Bifidobacterium* [3]. In line with these observations, we identified homologues of protein members of clusters *eps3* and *eps4* in *B. castoris*, including enzymes involved in rhamnose biosynthesis in cluster *eps3* which may link to additional biological properties of these polymers [82, 83]. However, additional studies are required to assess the functionality of the putative EPS biosynthesis machinery in *B. castoris* species.

CONCLUSION

It is well recognised that members of *Bifidobacterium* exert beneficial health effects on their host, however current knowledge of their diversity, distribution across the host phylogeny, and metabolic capability in non-human hosts, especially in wild animal populations, is limited. This research provides novel insights into the host-microbe evolutionary relationships and genomic features of *B. castoris* isolated from geographically distinct wild mouse populations. Our initial observations on strain-specific carbohydrate metabolism repertoires and the presence of *eps* genes require further investigation to understand how *Bifidobacterium* adapts, persists and interacts with the animal host and explain the functionality of mechanisms underlying bifidobacterial metabolic activity. This could be achieved through a combination of experimental and *in silico* methods, including additional isolation experiments and whole-genome sequencing, combined with community analyses based on metagenomic approaches, as well as carbohydrate metabolism and transcriptomics assays.

MATERIALS AND METHODS

Faecal sample collection

In the UK, live rodent trapping was carried out at two sites of mixed deciduous woodland approximately 50 km apart: Wytham Woods (51° 46′ N,1° 20′ W) and Nash's Copse, Silwood Park (51° 24′ N, 0° 38′ W). All animals were live-trapped using a standard protocol across both sites, using small Sherman traps baited with peanuts and apple and provisioned with bedding, set at dusk and collected at dawn the following day. All newly captured individuals were marked with subcutaneous PIT-tags for permanent identification, and all captures were weighed and various morphometric measurements taken. Faecal samples were collected using sterilised tweezers from the base of Sherman traps into sterile tubes. Samples from Silwood were frozen within 8 h of collection at −80 °C and sent frozen to the Quadram Institute (Norwich, UK) for *Bifidobacterium* culturing; samples from Wytham were posted on the day of sampling at room temperature to the same address for culturing. Upon reception of samples, they were immediately frozen at −80 °C. To ensure no cross-contamination and identification of samples to specific individuals, any traps that showed signs of rodent presence (captures and trigger failures) were washed thoroughly in a bleach solution and autoclaved between uses.

In Lithuania, small mammals were trapped in October 2017 and between May–November in 2018 using live- and snap-traps at twelve locations: Site 1: 54.92878, 25.33333; Site 2: 55.02814, 25.27380; Site 3: 54.92866, 25.31524, Site 4: 55.05938, 25.35643, Site 5: 54.96362, 25.35640; Site 6: 54.93026, 25.24138; Site 7: 54.99276, 25.24909, Site 8: 55.02700,

25.35867; Site 9: 55.06756, 25.29782; Site 10: 54.76482, 25.31283; Site 11: 54.76322, 25.35052; Site 12: 54.93650, 25.28442. Traps baited with bread soaked in sunflower oil (in case of live-traps, apple and bedding were also added) were set in the evening and retrieved in the morning. Small mammals trapped with snap-traps were placed in separate bags and transported to the lab on ice. Live-trapped animals were transported to the lab and humanely killed by cervical dislocation. Species, sex, age, reproduction status of small mammals were identified. Content of distal part of colon (~20–30 mm) was removed, placed in Eppendorf tube and stored at −80 °C. Frozen samples were sent to the Quadram Institute (Norwich, UK) for *Bifidobacterium* culturing.

The distance between trapping sites in both the UK and Lithuania was far enough for the animals not to move between them. All studied species have small home ranges—*Apodemus* spp. have the widest range and rarely move more than 0.25 km [84].

Strain isolation

Depending on the number of available faecal pellets, samples were re-suspended in either 450 or 900 µl of sterile phosphate buffer saline (Sigma-Aldrich, UK) and used to produce tenfold serial dilutions (neat - 10^{-4}). The samples were then vortexed for 30 s and mixed using on a shaker at 1600 rpm. An aliquot of each dilution (100 µl) was plated onto Brain Heart Infusion (BHI) (Oxoid, UK) agar supplemented with mupirocin (50 mg/l) (Sigma-Aldrich, UK), L-cysteine hydrochloride monohydrate (50 mg/l) (Sigma-Aldrich, UK) and sodium iodoacetate (7.5 mg/l) (Sigma-Aldrich, UK) and incubated in an anaerobic cabinet for 48–72 h (Baker Ruskinn, UK). Three colonies from each dilution were randomly selected and streaked to purity on BHI agar supplemented with L-cysteine hydrochloride monohydrate (50 mg/l). Pure cultures were stored in cryogenic tubes at −80 °C.

DNA extraction, whole-genome sequencing, assembly and annotation

DNA for whole-genome sequencing was extracted from pure bacterial cultures using the phenol-chloroform method as described previously [36], and subjected to multiplex Illumina library preparation protocol and sequencing on Illumina HiSeq 2500 platform at the Wellcome Trust Sanger Institute (Hinxton, UK) or Illumina MiSeq at the Quadram Institute Bioscience (Norwich, UK) (Supplementary Table S1). Sequencing reads were screened for contamination using Kraken v1.1 (MiniKraken) [85] and pre-processed with fastp v0.20 [86]. SPAdes v3.11 with "--careful" option [87] was used to produce assemblies, after which contigs below 500 bp were filtered out. Additionally, publicly available genome assembly of *Bifidobacterium* castoris 2020B^T type strain (accession number NZ_QXGI00000000.1) was retrieved from NCBI Genome database [4] and all genomes were annotated with Prokka v1.13 [88]. The draft genomes of 50 *Bifidobacterium* isolates sequenced here have been deposited to GOLD database at https://img.jgi.doe.gov, GOLD Study ID: Gs0153956.

Pangenomic and phylogenomic analysis

Anvi'o version 6.1 [89] was used to generate *B. castoris* pangenome and single copy core gene data for other analyses. Briefly, we created a text file containing required information on our collection of genomes and used this file to generate genomes storage database. We then computed the pangenome using the script anvi-pan-genome with parameters "--minbit 0.5 --mcl-inflation 10 --use-ncbi-blast". We next identified the single copy core genes in the pangenome and recovered their aligned amino acid sequences using with anvi-get-sequences-for-gene clusters with parameters "--min-num-genomes-gene-cluster-occurs, --max-num-genes-from-each-genome, --concatenate-gene clusters". The resulting output was cleaned from poorly aligned positions using trimAl v1.4.1 (gaps in more than 50% of the genes) [90]. IQ-TREE v1.6.1 [91] employing the 'WAG' general matrix model with 1000 bootstrap iterations [92] was used to infer the maximum-likelihood trees from protein sequences. The host tree for cophylogentic analysis was constructed using concatenated sequences for 12S rRNA and partial ctyb genes employing the 'GTR' model with 1000 bootstrap iterations. ParaFit [34] in the 'ape' package of R [34, 93] was used for distance-based comparisons, with the code included in the Supplementary Table S4. Python3 module pyANI v0.2.10 with default BLASTN+ settings was used to calculate the ANI [94]. Species delineation cut-off was set at 95% identity [95]. Isolates showing identity values above 99.9% were considered identical [29]. We used the script anvi-import-misc-data to import the results the anvi'o pangenome database and visualised the output with anvi-display-pan.

Comparative genomics

Functional categories (COG categories) were assigned to genes using EggNOG-mapper v0.99.3, based on the EggNOG database (bacteria) [96] and the abundance of genes involved in carbohydrate metabolism was calculated. Standalone version of dbCAN2 (v2.0.1) was used for glycobiome prediction [97]. 'Vegan' and 'indicspecies' packages implemented in R were used for the plotting of the *B. castoris* GH profiles using non-metric multidimensional scaling and for the multilevel pattern analysis with the point biserial correlation coefficient, respectively [98, 99], with the code included in the Supplementary Table S4. BLAST+ v2.9.0 (*e*-value of 1e−5 and 50% identity over 50% sequence coverage) [100] was used to screen *B. castoris* genomes for the presence of homologues of *eps* genes from *B. animalis* subsp. *lactis* BI12 (accession number CP004053.1, *eps*3: BI12_1287 – BI12_1328) and *B. pseudolongum* subsp. *globosum* LMG 11569^T (accession number JGZG01000015.1, *eps*4: BPSG_1548 – BPSG_1565). Further, amino acid sequences neighbouring the homologues of key-*eps3* and key-*eps4* genes were screened against the online version of the NCBI non-redundant protein database (Release 209, taxid 1678) using the BLASTP algorithm employing BLOSUM62 matrix. SIGI-HMM [101] tool implemented in Islandviewer4 [102] was employed to predict HGT events.

Carbohydrate metabolism experiments

B. castoris isolates representative of the 12 strains identified through genomic analysis, as well as *B. longum* NCIMB 8809 (negative control) were grown in unsupplemented mMRS and mMRS supplemented with either 0.5% (w/v) NMS or chitosan (Sigma-Aldrich, USA) and 0.05% cysteine-HCl (Sigma-Aldrich, USA) at pH 6.0 for 48 h at 37 °C (Baker Ruskinn, UK). Bacteria were introduced into each media type in triplicate with 1 in 100 inoculation. To track bacterial growth in each media type, 10 µl samples from one replicate (at $T = 0$ h) or all three replicates (at $T = 48$ h) of each media type and strain were removed for tenfold serial dilution (10^{-2}–10^{-7}) and plating on MRS plates. After 48 h anaerobic incubation as above, colonies formed were counted and colony forming units per ml (CFU/ml) were calculated.

REFERENCES

1. Turroni F, van Sinderen D, Ventura M. Genomics and ecological overview of the genus *Bifidobacterium*. Int J Food Microbiol. 2011;149:37–44.
2. O'Callaghan A, van Sinderen D. Bifidobacteria and their role as members of the human gut microbiota. Front Microbiol. 2016;7:925.
3. Ferrario C, Milani C, Mancabelli L, Lugli GA, Duranti S, Mangifesta M, et al. Modulation of the eps-ome transcription of bifidobacteria through simulation of human intestinal environment. FEMS Microbiol Ecol. 2016;92:fiw056.
4. Sayers EW, Beck J, Brister JR, Bolton EE, Canese K, Comeau DC, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2020;48:D9–D16.
5. Bottacini F, Medini D, Pavesi A, Turroni F, Foroni E, Riley D, et al. Comparative genomics of the genus *Bifidobacterium*. Microbiology. 2010;156:3243–54.
6. Turroni F, Duranti S, Bottacini F, Guglielmetti S, Van Sinderen D, Ventura M. *Bifidobacterium bifidum* as an example of a specialized human gut commensal. Front Microbiol. 2014;5:437.
7. Bottacini F, Motherway MO, Kuczynski J, O'Connell KJ, Serafini F, Duranti S, et al. Comparative genomics of the *Bifidobacterium breve* taxon. BMC Genomics. 2014;15:170.
8. Milani C, Lugli GA, Duranti S, Turroni F, Bottacini F, Mangifesta M, et al. Genomic encyclopedia of type strains of the genus *Bifidobacterium*. Appl Environ Microbiol. 2014;80:6290–302.
9. Milani C, Lugli GA, Duranti S, Turroni F, Mancabelli L, Ferrario C, et al. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. Sci Rep. 2015;5:15782.
10. Milani C, Turroni F, Duranti S, Lugli GA, Mancabelli L, Ferrario C, et al. Genomics of the genus Bifidobacterium reveals species-specific adaptation to the glycan-rich gut environment. Appl Environ Microbiol. 2016;82:980–91.
11. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, et al. Cospeciation of gut microbiota with hominids. Science. 2016;353:380–2.
12. Groussin M, Mazel F, Sanders JG, Smillie CS, Lavergne S, Thuiller W, et al. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nat Commun. 2017;8:14319.
13. Gaulke CA, Arnold HK, Humphreys IR, Kembel SW, O'Dwyer JP, Sharpton TJ. Ecophylogenetics clarifies the evolutionary association between mammals and their gut microbiota. mBio. 2018;9:e01348–18.

14. Youngblut ND, Reischer GH, Walters W, Schuster N, Walzer C, Stalder G, et al. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nat Commun. 2019;10:2200.
15. Lozupone CA, Hamady M, Cantarel BL, Coutinho PM, Henrissat B, Gordon JI, et al. The convergence of carbohydrate active gene repertoires in human gut microbes. Proc Natl Acad Sci USA. 2008;105:15076–81.
16. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA. 2006;103:15611–6.
17. Moeller AH, Suzuki TA, Phifer-Rixey M, Nachman MW. Transmission modes of the mammalian gut microbiota. Science. 2018;362:453–7.
18. Browne HP, Almeida A, Kumar N, Vervier K, Adoum AT, Viciani E, et al. Host adaptation in gut Firmicutes is associated with sporulation loss and altered colonisation patterns. 2020. https://www.biorxiv.org/content/10.1101/2020.09.09.289504v1.
19. Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, et al. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science. 2011;332:970–4.
20. Suzuki TA. Links between natural variation in the microbiome and host fitness in wild mammals. Integr Comp Biol. 2017;57:756–69.
21. McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, Douglas AE, et al. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci USA. 2013;110:3229–36.
22. Lugli GA, Alessandri G, Milani C, Mancabelli L, Ruiz L, Fontana F, et al. Evolutionary development and co-phylogeny of primate-associated bifidobacteria. Environ Microbiol. 2020;22:3375–93.
23. Lugli GA, Mancino W, Milani C, Duranti S, Mancabelli L, Napoli S, et al. Dissecting the evolutionary development of the species *Bifidobacterium* animalis through comparative genomics analyses. Appl Environ Microbiol. 2019;85:e02806–18.
24. Lugli GA, Duranti S, Albert K, Mancabelli L, Napoli S, Viappiani A, et al. Unveiling genomic diversity among members of the species *Bifidobacterium pseudolongum*, a widely distributed gut commensal of the animal kingdom. Appl Environ Microbiol. 2019;85:e03065–18.
25. Milani C, Mangifesta M, Mancabelli L, Lugli GA, James K, Duranti S, et al. Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life. ISME J. 2017;11:2834–47.
26. Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. The evolution of the host microbiome as an ecosystem on a leash. Nature. 2017;548:43–51.
27. van Vliet S, Doebeli M. The role of multilevel selection in host microbiome evolution. Proc Natl Acad Sci USA. 2019;116:20591–7.
28. Groussin M, Mazel F, Alm EJ. Co-evolution and co-speciation of host-gut bacteria systems. Cell Host Microbe. 2020;28:12–22.
29. Olm MR, Brown CT, Brooks B, Firek B, Baker R, Burstein D, et al. Identical bacterial populations colonize premature infant gut, skin, and oral microbiomes and exhibit different in situ growth rates. Genome Res. 2017;27:601–12.
30. Duranti S, Lugli GA, Napoli S, Anzalone R, Milani C, Mancabelli L, et al. Characterization of the phylogenetic diversity of five novel species belonging to the genus *Bifidobacterium*: *Bifidobacterium castoris* sp. nov., *Bifidobacterium callimiconis* sp. nov., *Bifidobacterium goeldii* sp. nov., *Bifidobacterium samirii* sp. nov. and *Bifidobacterium dolichotidis* sp. nov. Int J Syst Evol Microbiol. 2019;69:1288–98.
31. Lugli GA, Milani C, Duranti S, Mancabelli L, Mangifesta M, Turroni F, et al. Tracking the taxonomy of the genus *Bifidobacterium* based on a phylogenomic approach. Appl Environ Microbiol. 2017;84:e02249–17.
32. Snel B, Bork P, Huynen MA. Genome phylogeny based on gene content. Nat Genet. 1999;21:108–10.
33. Dutilh BE, Huynen MA, Bruno WJ, Snel B. The consistent phylogenetic signal in genome trees revealed by reducing the impact of noise. J Mol Evol. 2004;58:527–39.
34. Legendre P, Desdevises Y, Bazin E. A statistical test for host-parasite coevolution. Syst Biol. 2002;51:217–34.
35. Michaux JR, Chevret P, Filippucci MG, Macholan M. Phylogeny of the genus Apodemus with a special emphasis on the subgenus Sylvaemus using the nuclear IRBP gene and two mitochondrial markers: cytochrome b and 12S rRNA. Mol Phylogenet Evol. 2002;23:123–36.
36. Lawson MAE, O'Neill IJ, Kujawska M, Gowrinadh Javvadi S, Wijeyesekera A, Flegg Z, et al. Breast milk-derived human milk oligosaccharides promote Bifidobacterium interactions within a single ecosystem. ISME J. 2020;14:635–48.
37. Van Den Broek LAM, Voragen AGJ. *Bifidobacterium* glycoside hydrolases and (potential) prebiotics. Innov Food Sci Emerg Technol 2008;9:401–7.
38. Pokusaeva K, Fitzgerald GF, van Sinderen D. Carbohydrate metabolism in Bifidobacteria. Genes Nutr. 2011;6:285–306.
39. Rodriguez CI, Martiny JBH. Evolutionary relationships among bifidobacteria and their hosts and environments. BMC Genomics. 2020;21:26.
40. Henrissat B, Davies GJ. Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. Plant Physiol. 2000;124:1515–9.
41. Stam MR, Danchin EG, Rancurel C, Coutinho PM, Henrissat B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Eng Des Sel. 2006;19:555–62.
42. Miyazaki T, Ishizaki Y, Ichikawa M, Nishikawa A, Tonozuka T. Structural and biochemical characterization of novel bacterial alpha-galactosidases belonging to glycoside hydrolase family 31. Biochem J. 2015;469:145–58.
43. Hachem MA, Fredslund F, Andersen JM, Jonsgaard Larsen R, Majumder A, Ejby M, et al. Raffinose family oligosaccharide utilisation by probiotic bacteria: insight into substrate recognition, molecular architecture and diversity of GH36 α-galactosidases. Biocatal Biotransform. 2012;30:316–25.
44. Kujawska M, La Rosa SL, Roger LC, Pope PB, Hoyles L, McCartney AL, et al. Succession of *Bifidobacterium longum* strains in response to a changing early life nutritional environment reveals dietary substrate adaptations. iScience. 2020;23:101368.
45. Liu H, Ren W, Ly M, Li H, Wang S. Characterization of an alkaline GH49 dextranase from marine bacterium Arthrobacter oxydans KQ11 and its application in the preparation of isomalto-oligosaccharide. Mar Drugs. 2019;17:479.
46. Michlmayr H, Hell J, Lorenz C, Bohmdorfer S, Rosenau T, Kneifel W. Arabinoxylan oligosaccharide hydrolysis by family 43 and 51 glycosidases from *Lactobacillus brevis* DSM 20054. Appl Environ Microbiol. 2013;79:6747–54.
47. Fujita K, Takashi Y, Obuchi E, Kitahara K, Suganuma T. Characterization of a novel beta-L-arabinofuranosidase in *Bifidobacterium longum*: functional elucidation of a DUF1680 protein family member. J Biol Chem. 2014;289:5240–9.
48. Viens P, Lacombe-Harvey ME, Brzezinski R. Chitosanases from Family 46 of glycoside hydrolases: from proteins to phenotypes. Mar Drugs. 2015;13:6566–87.
49. Sela DA, Garrido D, Lerno L, Wu S, Tan K, Eom HJ, et al. *Bifidobacterium longum* subsp. infantis ATCC 15697 alpha-fucosidases are active on fucosylated human milk oligosaccharides. Appl Environ Microbiol. 2012;78:795–803.
50. Garrido D, Ruiz-Moyano S, Kirmiz N, Davis JC, Totten SM, Lemay DG, et al. A novel gene cluster allows preferential utilization of fucosylated milk oligosaccharides in *Bifidobacterium longum* subsp longum SC596. Sci Rep-Uk. 2016;6:35045.
51. Kitaoka M. Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides. Adv Nutr. 2012;3:422S–9S.
52. Kiyohara M, Tanigawa K, Chaiwangsri T, Katayama T, Ashida H, Yamamoto K. An exo-alpha-sialidase from bifidobacteria involved in the degradation of sialyllo-igosaccharides in human milk and intestinal glycoconjugates. Glycobiology. 2011;21:437–47.
53. Breton C, Snajdrova L, Jeanneau C, Koca J, Imberty A. Structures and mechanisms of glycosyltransferases. Glycobiology. 2006;16:29R–37R.
54. Hidalgo-Cantabrana C, Sanchez B, Milani C, Ventura M, Margolles A, Ruas-Madiedo P. Genomic overview and biological functions of exopolysaccharide biosynthesis in *Bifidobacterium* spp. Appl Environ Microbiol. 2014;80:9–18.
55. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The carbohydrate-active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37:D233–8.
56. Lavrinienko A, Tukalenko E, Mousseau TA, Thompson LR, Knight R, Mappes T, et al. Two hundred and fifty-four metagenome-assembled bacterial genomes from the bank vole gut microbiota. Sci Data. 2020;7:312.
57. Baumler A, Fang FC. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med. 2013;3:a010041.
58. Glazko GV, Nei M. Estimation of divergence times for major lineages of primate species. Mol Biol Evol. 2003;20:424–34.
59. Milton K. The critical role played by animal source foods in human (Homo) evolution. J Nutr. 2003;133:3886S–92S.
60. Renaud S, Michaux J, Schmidt DN, Aguilar JP, Mein P, Auffray JC. Morphological evolution, ecological diversification and climate change in rodents. Proc Biol Sci. 2005;272:609–17.
61. Michaux JR, Libois R, Filipucci M-G. So close and so different: comparative phylogeography of two small mammal species, the Yellow-necked fieldmouse (*Apodemus flavicollis*) and the Woodmouse (*Apodemus sylvaticus*) in the Western Palearctic region. Heredity. 2005;94:52–63.
62. Ge D, Feijó A, Cheng J, Lu L, Liu R, Abramov AV, et al. Evolutionary history of field mice (Murinae: Apodemus), with emphasis on morphological variation among species in China and description of a new species. Zool J Linn Soc. 2019;187:5188–534.
63. Moeller AH, Peeters M, Ndjango JB, Li Y, Hahn BH, Ochman H. Sympatric chimpanzees and gorillas harbor convergent gut microbial communities. Genome Res. 2013;23:1715–20.
64. Knowles SCL, Eccles RM, Baltrunaite L. Species identity dominates over environment in shaping the microbiota of small mammals. Ecol Lett. 2019;22:826–37.
65. Watts CHS. The foods eaten by wood mice (*Apodemus sylvaticus*) and bank voles (Clethrionomys glareolus) in Wytham Woods, Berkshire. J Anim Ecol. 1968;37:25–41.
66. Abt KF, Bock WF. Seasonal variations of diet composition in farmland field mice Apodemus spp. and bank voles Clethrionomys glareolus. Acta Theriol. 1998;43:379–89.
67. Rogers LM, Gorman ML. The diet of the wood mouse *Apodemus sylvaticus* on set-aside land. J Zool. 1995;235:77–83.

68. Van Laere KM, Beldman G, Voragen AG. A new arabinofuranohydrolase from *Bifidobacterium* adolescentis able to remove arabinosyl residues from double-substituted xylose units in arabinoxylan. Appl Microbiol Biotechnol. 1997;47:231–5.
69. Margolles A, de los Reyes-Gavilan CG. Purification and functional characterization of a novel alpha-L-arabinofuranosidase from *Bifidobacterium longum* B667. Appl Environ Microbiol. 2003;69:5096–103.
70. Lagaert S, Pollet A, Delcour JA, Lavigne R, Courtin CM, Volckaert G. Substrate specificity of three recombinant alpha-L-arabinofuranosidases from *Bifidobacterium adolescentis* and their divergent action on arabinoxylan and arabinoxylan oligosaccharides. Biochem Biophys Res Commun. 2010;402:644–50.
71. Ito T, Saikawa K, Kim S, Fujita K, Ishiwata A, Kaeothip S, et al. Crystal structure of glycoside hydrolase family 127 beta-l-arabinofuranosidase from *Bifidobacterium longum*. Biochem Biophys Res Commun. 2014;447:32–7.
72. Kataržytė M, Kutorga E. Small mammal mycophagy in hemiboreal forest communities of Lithuania. Central Eur J Biol. 2011;6:446–56.
73. Lee HW, Park YS, Jung JS, Shin WS. Chitosan oligosaccharides, dp 2-8, have prebiotic effect on the *Bifidobacterium bifidium* and *Lactobacillus* sp. Anaerobe. 2002;8:319–24.
74. Vernazza CL, Gibson GR, Rastall RA. In vitro fermentation of chitosan derivatives by mixed cultures of human faecal bacteria. Carbohyd Polym. 2005;60:539–45.
75. Yang CM, Ferket PR, Hong QH, Zhou J, Cao GT, Zhou L, et al. Effect of chito-oligosaccharide on growth performance, intestinal barrier function, intestinal morphology and cecal microflora in weaned pigs. J Anim Sci. 2012;90:2671–6.
76. Zhang C, Jiao S, Wang ZA, Du Y. Exploring effects of chitosan oligosaccharides on mice gut microbiota in in vitro fermentation and animal model. Front Microbiol. 2018;9:2388.
77. Wu J, Zhang L. Dissolution behavior and conformation change of chitosan in concentrated chitosan hydrochloric acid solution and comparison with dilute and semidilute solutions. Int J Biol Macromol. 2019;121:1101–8.
78. Costa CN, Teixeira VG, Delpech MC, Souza JV, Costa MA. Viscometric study of chitosan solutions in acetic acid/sodium acetate and acetic acid/sodium chloride. Carbohydr Polym. 2015;133:245–50.
79. Kiu R, Treveil A, Harnisch LC, Caim S, Leclaire C, van Sinderen D, et al. *Bifidobacterium breve* UCC2003 induces a distinct global transcriptomic program in neonatal murine intestinal epithelial cells. iScience. 2020;23:101336.
80. Hughes KR, Harnisch LC, Alcon-Giner C, Mitra S, Wright CJ, Ketskemety J, et al. *Bifidobacterium breve* reduces apoptotic epithelial cell shedding in an exopolysaccharide and MyD88-dependent manner. Open Biol. 2017;7:160155.
81. Fanning S, Hall LJ, Cronin M, Zomer A, MacSharry J, Goulding D, et al. Bifidobacterial surface-exopolysaccharide facilitates commensal-host interaction through immune modulation and pathogen protection. Proc Natl Acad Sci USA. 2012;109:2108–13.
82. Roca C, Alves VD, Freitas F, Reis MA. Exopolysaccharides enriched in rare sugars: bacterial sources, production, and applications. Front Microbiol. 2015;6:288.
83. Balzaretti S, Taverniti V, Guglielmetti S, Fiore W, Minuzzo M, Ngo HN, et al. A novel rhamnose-rich hetero-exopolysaccharide isolated from *Lactobacillus paracasei* DG activates THP-1 human monocytic cells. Appl Environ Microbiol. 2017;83:e02702–16.
84. Stradiotto A, Cagnacci F, Delahay R, Tioli S, Nieder L, Rizzoli A. Spatial organization of the yellow-necked mouse: effects of density and resource availability. J Mammal. 2009;90:704–14.
85. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46.
86. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–i90.
87. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
88. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
89. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 2015;3:e1319.
90. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
91. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
92. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol. 2001;18:691–9.
93. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 2019;35:526–8.
94. Pritchard L, Glover RH, Humphris S, Elphinstone JG, Toth IK. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal Methods. 2016;8:12–24.
95. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol. 2018;68:461–6.
96. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017;34:2115–22.
97. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46:W95–W101.
98. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGillin D, et al. vegan: community ecology package. R package version 25-6. 2019. https://CRAN.R-project.org/package=vegan.
99. De Caceres M, Legendre P, Moretti M. Improving indicator species analysis by combining groups of sites. Oikos. 2010;119:1674–84.
100. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.
101. Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, et al. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics. 2006;7:142.
102. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res. 2017;45:W30–W5.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Florent Mazel for insightful discussions and providing feedback on the study, and Dr. Hannah Harris for her help with the in vitro growth curves. LJH is supported by Wellcome Trust Investigator Awards 100974/C/13/Z and 220876/Z/20/Z; the Biotechnology and Biological Sciences Research Council (BBSRC), Institute Strategic Programme Gut Microbes and Health BB/R012490/1, and its constituent projects BBS/E/F/000PR10353 and BBS/E/F/000PR10356, and a BBSRC Norwich Research Park Bioscience Doctoral Training grant no. BB/M011216/1 (supervisors LJH and FW, students MK and MM). SCLK was funded by a NERC fellowship (NE/L011867/1) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 851550), and AR by a Clarendon Scholarship from the University of Oxford. LB was funded by grant S-MIP-17-86 from the Research Council of Lithuania. FW thanks the BBSRC Institute Strategic Programme (ISP) Food Innovation and Health BB/R012512/1 and its constituent projects (BBS/E/F/000PR10343, BS/E/F/000PR10346). The funding bodies did not contribute to the design of the study, collection, analysis, and interpretation of data or in writing the manuscript.

AUTHOR CONTRIBUTIONS

LJH and MK designed the overall study. AR, LB and SCLK provided the faecal samples from wild mice populations and associated metadata. MK isolated the strains, extracted the DNA and prepared it for WGS, performed all genomic analysis and visualisation. MM and FW helped design the growth assays with MM performing these experiments. MK and LJH analysed the data, with input and discussion from SCLK and AR, and drafted the manuscript. AR, LB and SCLK provided further edits and co-wrote the final version. All authors read and approved the final manuscript.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

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/s43705-022-00102-x.

Correspondence and requests for materials should be addressed to Lindsay J. Hall.

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.