Horizontal gene transfer shapes pathogenic bacteria in multiple sclerosis

Takashi Yamamura (yamamura@ncnp.go.jp)
National Center of Neurology and Psychiatry

Daiki Takewaki
National Center of Neurology and Psychiatry

Yuya Kiguchi
RIKEN Center for Integrative Medical Sciences

Hiroaki Masuoka
RIKEN Center for Integrative Medical Sciences

Mallahalli Manu
National Center of Neurology and Psychiatry

Ben Raveney
National Center of Neurology and Psychiatry https://orcid.org/0000-0003-2641-5733

Seiko Narushima
RIKEN Center for Integrative Medical Sciences

Rina Kurokawa
RIKEN Center for Integrative Medical Sciences

Yusuke Ogata
Laboratory for Microbiome Sciences, RIKEN Center for Integrative Medical Sciences
https://orcid.org/0000-0002-5993-4356

Yukio Kimura
National Center of Neurology and Psychiatry

Noriko Sato
National Center of Neurology and Psychiatry

Yusuke Ozawa
National Center of Neurology and Psychiatry

Sosuke Yagishita
National Center of Neurology and Psychiatry

Toshiyuki Araki
National Center of Neurology and Psychiatry https://orcid.org/0000-0003-3625-2042

Sachiko Miyake
Juntendo University School of Medicine

Wakiro Sato
National Institute of Neuroscience, National Center of Neurology and Psychiatry

Wataru Suda
RIKEN Center for Integrative Medical Sciences

Biological Sciences - Article

Keywords:

Posted Date: December 8th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3716024/v1

Additional Declarations: Yes there is potential Competing Interest. D.Takewaki received support for research from Yakult Bio-Science Foundation. W. Sato received support for clinical trials from Chugai Pharmaceutical Co., Ltd. and Biogen Japan Ltd. T. Yamamura served on the scientific advisory board of Biogen Japan Ltd. and received support for research and clinical trials from Sanofi K.K., Chugai Pharmaceutical Co., Ltd., UCB Japan Co., Ltd., Biogen Japan Ltd., Chiome Bioscience Inc., Novartis Pharma K.K., and Mebix, Inc.

Horizontal gene transfer shapes pathogenic bacteria in multiple sclerosis

Authors & Affiliations

Daiki Takewaki[1,2,3,9], Yuya Kiguchi[3,4,9], Hiroaki Masuoka[3], Mallahalli Manu[1,2], Ben J E Raveney[1,2], Seiko Narushima[5], Rina Kurokawa[3], Yusuke Ogata[3], Yukio Kimura[2,6], Noriko Sato[2,6], Yusuke Ozawa[7], Sosuke Yagishita[7], Toshiyuki Araki[7], Sachiko Miyake[8], Wakiro Sato[1,2], Wataru Suda[3,10*], Takashi Yamamura[1,2,10*]

1. Department of Immunology, National Institute of Neuroscience, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8502, Japan
2. Multiple Sclerosis Center, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8502, Japan
3. Laboratory for Symbiotic Microbiome Sciences, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
4. Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba 277-8568, Japan
5. Laboratory for Mucosal Immunity, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
6. Department of Radiology, National Center of Neurology and Psychiatry Hospital, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo 187-8502, Japan
7. Department of Peripheral Nervous System Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, 4-1-1 Ogawa-Higashi, Kodaira, Tokyo, 187-8502, Japan
8. Department of Immunology, Juntendo University School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan
9. These authors contributed equally: Daiki Takewaki and Yuya Kiguchi.
10. These authors jointly supervised this work: Wataru Suda and Takashi Yamamura.

* **Email:** yamamura@ncnp.go.jp; wataru.suda@riken.jp

SUMMARY

Multiple sclerosis (MS) is an autoimmune demyelinating disease influenced by environmental factors. Except during relapses, baseline neurological status is generally stable in the early stage, whereas progressive deterioration may occur silently. The progressive disease form (secondary progressive MS; SPMS) characterised by both neuroinflammation and neurodegeneration differs significantly from the non-progressive form in microbiome profiles[1 2 3]. After confirming an increased abundance of gut bacterium "*Tyzzerella nexilis"* in SPMS, the role of *T.nexilis* in progressive MS was studied. The strain-level analysis based on long-read metagenomics identified a distinct cluster of *T.nexilis* highly enriched in SPMS. *T.nexilis* strains in this novel cluster were characterised by an incredible number of mobile genetic elements (MGEs) and the absence of defence systems against MGEs. Mono-colonisation with this MGEs-enriched *T.nexilis* strain made germ-free mice more susceptible to induction of experimental autoimmune encephalomyelitis. The pathogenicity of this strain was mediated by TLR5 stimulation by flagella encoded on MGEs. Moreover, this *T.nexilis* strain was thought to have potentials of causing neurodegeneration, because of its ability to produce reduced sulphur compounds encoded on MGEs. Such a horizontal gene transfer, causing functional diversity beyond existing bacterial taxonomy, may have causal implications in chronic disorders influenced by gut microbiome.

MAIN TEXT

Most patients with multiple sclerosis (MS) initially experience intermittent acute inflammation that causes relapses (relapsing-remitting MS; RRMS), while some patients with RRMS shift to secondary progressive MS (SPMS). SPMS is generally defined by the presence of progression independent of relapse activity attributable to cytotoxic T lymphocytes[4,5], activated glial cells[6], and enhanced oxidative stress[7] leading to chronic neuroinflammation and neurodegeneration.

Comprehensive microbial analysis has revealed dysbiosis of the gut microbiome in RRMS as reported in various countries[1 2 3 8 9 10 11]. Meanwhile, studies in the rodent model, experimental autoimmune encephalomyelitis (EAE), showed that commensal microbiome affected various immune cell populations including T and B lymphocytes and microglial cells, thereby regulating the central nervous system (CNS) autoimmune disorder[12 13 14 15 16]. Recent reports demonstrated that gut microbiome profiles significantly differ between RRMS and SPMS[1 2 3]. However, its biological meaning remains unclear. The purpose of this research is to reveal the characteristics of causative bacteria underlying MS progression and clarify the mechanism of exacerbating neuronal inflammation and degeneration.

Identification of gut bacteria associated with MS progression

First, we investigated the composition of gut microbiomes of MS using original short-read metagenomic data, followed by marker gene-based operational taxonomic unit (mOTU) analysis (**Extended data Table 1**). To identify specific bacteria characteristic of SPMS, we employed MaAsLin2[17]. In the exploratory comparison between RRMS and SPMS, 38 mOTUs were significantly increased or decreased in SPMS. Among them, *Tyzzerella nexilis* (id03689) belonging to Firmicutes was the most enriched species in SPMS based on the coefficient value of MaAsLin2 (**Fig.1a and Extended data Table 2**). In the comparison between patients with SPMS and healthy control (HC) participants, 40 mOTUs were highly enriched or depleted in SPMS (|coefficient| >1 and p <0.05 shown in **Fig.1b**). Among them, *Tyzzerella nexilis* (id03689) was the most enriched species in SPMS (**Fig.1b and Extended data Fig.1a**). Next, to explore if enrichment of *T.nexilis* (id03689) in progressive MS is present in other regions or countries, we conducted mOTU analysis using the large-scale metagenomic data from the international MS microbiome study (iMSMS) consortium[3].

To compare the microbial taxa with the same analytic procedure using mOTUs, we analysed 1,139 samples (435 RRMS, 133 progressive MS, and 571 HC). Comparisons of iMSMS samples between progressive MS and HC revealed a significant enrichment of *T.nexilis* (id03689) in the three progressive MS groups from Western countries (**Fig.1b and Extended data Fig.1b-d**). However, no such differences were observed between RRMS and HC in most countries (**Extended data Fig.1f**). We also conducted mOTU analysis using the public data of inflammatory bowel disease (IBD)[18], rheumatoid arthritis[19], colorectal cancer[20], and Parkinson's disease[21]. These diseases did not show enrichment of *T.nexilis* (id03689) (**Extended data Fig.1f**).

To validate the possible effects of immunotherapy, we compared the abundance of 40 mOTUs between the untreated progressive MS and HC in the iMSMS data. The abundance of *T.nexilis* (id03689) was significantly increased in progressive MS even after removing the treated patients (**Fig.1b and Extended data Fig.1e**), excluding the effects of immunotherapy. We subsequently conducted a correlation analysis between gut microbial abundance and expanded disability status scale (EDSS) score, a clinical measure of neurological impairment, in both our original and iMSMS cohorts. While the abundance of *T.nexilis* (id03689) significantly correlated with the EDSS score (**Fig.1b**), the abundance of *T.nexilis* (id03689) tended to increase during the transition phase (EDSS 3 to 4.5) in both cohorts (**Fig.1c**). As transition from RRMS to SPMS generally occurs at EDSS 3 to 4.5, we suspected that the increase of *T.nexilis* may trigger progressive MS, while it decreases at later points. In contrast, the other four mOTUs were positively correlated with EDSS score without bias for the transition phase (**Extended data Fig.2**). Then we also conducted correlation analysis between the abundance of 40 mOTUs and four quantitative brain magnetic resonance imaging (MRI) parameters. The brain volume was significantly lower, and the lesion volume was significantly higher in SPMS than in RRMS (**Extended data Fig.3a**), and the abundance of *T.nexilis* (id03689) negatively correlated with the total brain volume and grey matter volume of these patients (**Extended data Fig.3b**).

We further analysed the non-redundant metagenome-assembled genome (MAG) by using our original metagenomic data. Based on the taxonomical assignment, two MAGs were annotated to *T.nexilis*. We named one of them *T.nexilis*_strain A MAG and the other *T.nexilis*_strain B MAG. Then we evaluated the number of mapped reads per million to the strain-specific marker genes

between *T.nexilis*_strain A and B MAGs (**Extended data Table 3**). Notably, strain B MAG was significantly enriched in SPMS compared with HC and RRMS, whereas the abundance of strain A MAG was not significantly different between the three groups (**Fig.1d**). The prevalence of strain B MAG was much higher in SPMS (33%) than in HC (14%) and RRMS (13%) (**Fig.1e**).

Discovery of *Tyzzerella nexilis* strains encoding numerous mobile genetic elements

For high-resolution genomic analysis, we tried to obtain the high-quality genomes of *T.nexilis* from selected five "*T.nexilis* (id03689)-high" patients. We successfully isolated two *T.nexilis* strains from the faecal samples of RRMS_031 and SPMS_002. We performed long-read sequencing using PacBio HiFi reads and obtained two complete genomes by *de novo* assembly. Additionally, we performed long-read metagenomics for the other three samples and obtained metagenomic HiFi reads with 19.3 Gb and 8.3 kb read lengths on average (**Extended data Table 1**). We conducted the long-read metagenomic assembly for the samples from SPMS_015 and RRMS_004 and obtained the three complete *T.nexilis* genomes (**Extended data Fig.4a**). For the sample from RRMS_041, we performed a combinational analysis of a long-read metagenomic assembly, binning, and reference-guided assembly because we could not construct complete genomes through the simple metagenomic assembly (**Extended data Fig.4b**). All six genomes were closed circular and satisfied the completeness >99% and contamination <2% (**Fig.2a**). Based on the phylogenetic tree of *T.nexilis* genomes, we found two distinct clusters in *T.nexilis*. One included strain A MAG and the other included strain B MAG (**Fig.2b**). Among the seven circular genomes, the type strain (*T.nexilis*_DSM1787) and three genomes from patients with MS were clustered to the strain A lineage (**Fig.2b**) and had >98.7% average nucleotide identity (ANI) with strain A MAG (**Fig.2a**). Conversely, the other three genomes from patients with MS were clustered to the strain B lineage (**Fig.2b**) and had >99.6% ANI with strain B MAG (**Fig.2a**). We named these novel genomes as A1, A2, A3, B1, B2, and B3, respectively. Dot plot for genomic comparison also suggested that strain B1, B2, and B3 had clearly different genomic structures from strain A1, A2, and A3 (**Extended data Fig.5**). Further, we performed comparative genomics of *T.nexilis* strains. The genome size and number of total genes were much larger in strain B1, B2, and B3. The number of virulent genes assigned to the virulence factor database was also larger in strain B1, B2, and B3. Notably, the

number of mobile genetic elements (MGEs) that are closely associated with the bacterial horizontal gene transfer, including insertion sequences (IS), prophages, and integrative and conjugative elements (ICE), was much larger in strain B1, B2, and B3 (**Fig.2c**). Then, we analysed various defence systems that protect bacteria from MGEs, among which CRISPR-Cas9 (CAS_Class1) works against various types of MGEs[22] as well as restriction modification (RM). Notably, the number of defence systems was much smaller in strain B1, B2, and B3 (**Fig.2c**). Moreover, an incredibly large number of MGEs found in the strain B1, B2, and B3 were outstanding among the 727 representative Firmicutes genomes in the GenBank (**Fig.2d and Extended data Table 4**). The genomes of strain B1, B2, and B3 have an average of nine times larger ISs, 18 times larger ICEs, and three times larger prophages than the average of the Firmicutes genomes. In the process of the comparison between two strain B1 genomes from short-read MAG and isolation-based complete genome, we realised that a large number of MGEs were not detected in the short-read MAG (**Extended data Fig.6**), which suggested the superiority of long-read metagenomics over short-read metagenomics. This is why we have not used fragmented genomes for the comparison of *T.nexilis* strains.

Functional significance of *Tyzzerella nexilis* strains

Subsequently, we conducted *in vivo* bacterial transfer experiments using two isolated *T.nexilis* strains belonging to distinct clusters (**Fig.3a**). Germ-free mice were administered *T.nexilis*_strain A1 or B1 in culture media one and two weeks before sensitisation to myelin oligodendrocyte glycoprotein (MOG) peptide fragment 35-55 (MOG_{35-55}). Colonisation with these bacteria was confirmed by quantitative PCR (qPCR) using universal primers for 16S rRNA genes one week after the last bacterial administration (**Extended data Fig.7a**). Strain B1-colonised mice showed an increased severity of EAE symptoms compared with germ-free mice, whereas colonisation with strain A1 showed the restricted effects (**Fig.3b**). The frequency of T helper 17 (Th17) cells among $CD4^+$ T cells in the small-intestinal lamina propria (SILP), large-intestinal lamina propria (LILP), and CNS was much higher in strain A1 or B1 mono-colonised mice than in germ-free mice. Moreover, the frequency of Th17 cells in strain B1 mono-colonised mice was significantly higher than in strain A1 mono-colonised mice in the LILP and CNS (**Fig.3c, d**). The frequency of regulatory T cells in the

SILP, LILP, and CNS was not significantly different between strain A1 and B1 mono-colonised mice (**Extended data Fig.7b, c**).

To elucidate the mechanism with which strain B1 caused severe EAE symptoms compared to strain A1, we compared the complete genomes of these two strains. We annotated open reading frames (ORFs) of each genome based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database and compared the differences of the number of KEGG orthologies (KOs) in each pathway. Among them, we analysed the four most enriched pathways in strain B1 (**Fig. 3e**). Initially, we focused on the pathways "flagellar assembly" and "bacterial chemotaxis," both related to flagellar functions. As flagellin is a selective agonist of toll-like receptor 5 (TLR5), we addressed whether each *T.nexilis* strain had a differential ability to stimulate TLR5. Using TLR5-reporter cells, we confirmed that strain B1, but not strain A1, stimulated TLR5, suggesting that strain B1 possesses flagella (**Fig.3f**). Consistently, a flagella-like structure in strain B1was confirmed using scanning electron microscopic (SEM) images (**Fig.3g**). Then, we co-cultured specific pathogen-free mice-derived $CD11c^{high}$ $CD11b^{high}$ intestinal lamina propria dendritic cells (LPDCs) (a TLR5-expressing LPDC population)[23] with each *T. nexilis* strain in the presence or absence of a TLR5 antagonist. Co-culture of the LPDCs with strain B1 promoted the secretion of IL-6 in the supernatant compared to that with strain A1. However, the supplementation with a TLR5 antagonist almost completely suppressed the secretion of IL-6 from the LPDCs. These findings suggested that *T.nexilis*_strain B1 can stimulate TLR5 expressed by intestinal LPDCs and subsequently induce the secretion of IL-6 (**Fig.3h**). Next, we focused on another flagellar function, which is promoting adherence to intestinal epithelial cells (ECs)[24]. To verify the bacterial adherence on the intestinal ECs, we prepared strain A1 or B1 mono-colonised mice. After cleaning up the intestinal contents, we observed the surface of the colon. Although no bacteria were detected on the colonic ECs of strain A1 mono-colonised mice despite sufficient colonisation (**Extended data Fig.7a**), many bacteria were found to be adherent to the ECs of the proximal colon in strain B1 mono-colonised mice (**Fig.3i**). Then, we examined the effects of strain B1 adhesion on colonic EC gene expression profiles using RNA-sequencing. Gene ontology (GO) analysis revealed that various biological process terms possibly related to immune activation and defence response against invasive bacteria such as immune system process (GO: 0002376), immune response (GO: 0006955), and defence response (GO: 0006952) were highly enriched in the

colonic ECs of strain B1 mono-colonised mice compared with those of strain A1 mono-colonised mice (**Extended data Fig.7d**). In consistence with this result, the expression of α-defensin genes was upregulated in the strain B1 mono-colonised mice (**Extended data Fig.7e**). Among various upregulated genes, we particularly focused on the serum amyloid A (SAA) genes because prior studies showed that SAA1 secreted from the intestinal ECs upon bacterial adherence promotes Th17 cell differentiation[13 25]. By conducting RNA-sequencing and qPCR analyses, we observed a significant increase in *SAA1* expression in the colonic ECs of strain B1 mono-colonised mice (**Fig.3j and Extended data Fig.7f**). To verify the biological significance of the enrichment of IL-6 and SAA1 in the local environment, we conducted an *in vitro* assay for the activation of naive $CD4^+$ T cells. We observed that a combination of IL-6 and SAA1 robustly promoted the secretion of IL-17 and GM-CSF from the activated T cells compared with IL-6 and TGF-β (the combination for non-pathogenic Th17 cell differentiation) under T cell stimulations of different strength via anti-CD3/CD28 antibodies (**Fig.3k, l**). In a previous paper, the combination of IL-6 and SAA1 was reported to promote a higher expression of genes related to pathogenic Th17 cells, including *Tbx21*, *Gzmb*, and *IL-23R* in activated T cells, as compared with the combination of IL-6 and TGF-β and subsequently exacerbated CNS autoimmunity[26]. Collectively, we postulate that the strain B1-flagella probably accelerate neuronal inflammation via the combination of TLR5 stimulation and promotion of adherence to colonic ECs. Then, we evaluated the abundance of strain B1-flagellar genes among HC, RRMS, and SPMS. We focused on *fliC* genes because *fliC* encodes a filament of flagella that contains the epitope of the TLR5 ligand. The number of mapped reads per million to *fliC* genes of strain B1 was significantly increased in SPMS, suggesting that the strain B1-flagella are enriched in SPMS (**Fig.3m**). The pathway "cationic antimicrobial peptide (CAMP) resistance" was also enriched in strain B1. In this pathway, 4 KOs included in strain B1 were mapped to the module *dltABCD* operon (M00725), which is associated with increased resistance to CAMPs such as α-defensin[27] (**Extended data Fig.7g**). High expression of α-defensin genes in colonic ECs of strain B1-monocolonised mice implied the CAMP resistance in strain B1 (**Extended data Fig.7e**). Finally, we focused on the pathway "sulphur metabolism." In this pathway, 4 and 1 KOs included in strain B1 (not in strain A1) were mapped to the module assimilatory sulphate reduction (M00176) and dissimilatory sulphate reduction (M00596), respectively (**Fig.3n**). To confirm the effects of strain B1

on sulphur metabolism in the gut, we analysed the concentration of 30 sulphur metabolites in the faeces of germ-free and strain A1 or B1 mono-colonised mice. The ratio of glutathione (reductive form) to glutathione disulphide (oxidative form), a quantitative indicator of redox states[28], was significantly higher in strain B1 than in strain A1 (**Fig. 3o**). We also measured the production of hydrogen sulphide (H_2S) in the gas phase of sealed liquid culture media of strain A1 and B1 because H_2S is a terminal product of sulphate reduction. The production of H_2S was higher in strain B1 than in strain A1 (**Fig.3p**).

Mechanisms for the acquisition of the potentially pathogenic genes

The prominent accumulation of MGEs in *T.nexilis*_strain B1 urged us to explore the origin of these potentially pathogenic genes. The comparison of seven *T.nexilis* genomes revealed that 11 KOs in "flagella assembly" and 10 KOs in "sulphur metabolism" were present in the genomes of cluster B but absent in those of cluster A, although the average similarity of 16S rRNA genes between cluster A and B was more than 99.2% (**Extended data Fig.8 and Extended data Table 5**). Initially, we focused on several genes related to sulphur metabolism. In the specific region of the strain B1 genome, four sulphur metabolic genes were aligned to phage proteins with >75% amino acid identity on average (**Fig.4a**), and a 25kb region that included sulphur metabolic genes was aligned to an isolated *Ruminococcus gnavus* strain (strain AF33-12[29]) with >99% nucleotide similarity (**Fig.4a**). Although we could not find obvious characteristics of known MGEs in this genomic region encoding a sulphur metabolic gene cluster, these results suggest that horizontal transmission had occurred between *T. nexilis* and *R.gnavus*, a prevalent gut species with pathogenic potentials whose enrichment was reported in various diseases[30].

Then, we explored flagellar genes highly similar to those of strain B1 in public bacterial genomes obtained from a variety of environments[31 32 33]. However, we could not find any relevant results, which prompted us to devise a different strategy. We attempted to recapitulate the evolutionary history of a *fliC* gene based on the genomes of various bacterial species with flagellar genes. Specifically, we collected human gut bacterial genomes encoding *fliC* from the HumGut database and constructed the phylogenetic tree of *fliC* including the four genes of *T.nexilis*_strain B1. The *fliC* genes of strain B1 were clustered within the Firmicutes (**Extended data Fig.9**). The phylogenetic

tree of Firmicutes genomes based on the taxonomic marker genes consisted of five clusters at an order level and had *T.nexilis*_strain B1 in a cluster of the Lachnospirales (**Fig.4b**). The phylogenetic tree of *fliC* genes also consisted of several clusters generally consistent with classification at the order-level; however, the *fliC* genes of *T.nexilis*_strain B1 were located within a cluster of the Oscillospirales (**Fig.4c**). These conflicts between bacterial taxonomy and phylogeny of *fliC* suggested that the strain B1-*fliC* genes were acquired by horizontal gene transfer. Moreover, the phylogenetic tree of *fliC* indicates that the strain B1-*fliC* was initially derived from Oscillospirales. Finally, to clarify the genomic structure of this flagella-horizontal gene transfer region in the *T.nexilis*_strain B1, we compared the two bacterial genomes of strain B1 and B2 because the genome of strain B2 does not include any flagellar genes (**Extended data Fig.8**) although strain B1 and B2 are phylogenetically close with >99.6% 16S rRNA gene similarity. In the comparison of these two genomes, we could determine the edge of the flagella-horizontal gene transfer region which does not exist in the strain B2 genome and is flanked by two direct terminal repeats and the IS family 91 (IS91) (**Fig.4d**). These results suggest that this flagella-horizontal gene transfer region spanning one IS91 to the other IS91 was transferred as a single unit called as composite transposon[34]. Interestingly, typical ICE signatures such as the presence of ICE marker genes, alteration of the GC content and skew, and the direct terminal repeat at the candidate boundaries of the ICE were found inside the structure of the composite transposon (**Fig.4d**). Moreover, 18 out of 22 ICE-related genes were aligned to ICE-derived genes from *Clostridioides difficile* with 90% amino acid identity on average (**Extended data Table 6**). As physical contact between bacteria is necessary for ICE acquisition, these results imply the past event of conjugative interaction between *T.nexilis*_strain B1 and *Clostridioides difficile*.

Discussion

The gut microbiome altered in progressive MS appears to influence the neuronal inflammation and degeneration[1 2 3]. To understand the correlation of altered microbiome components to the brain disorder, we sought a pathogenic bacterium triggering or boosting MS progression. Comprehensive metagenomic analysis revealed that "*Tyzzerella nexilis"* was the most significantly enriched species in SPMS compared to RRMS and HC. Notably, it was an exceptional bacterium whose abundance was significantly correlated with neurological disability and showed an inverse correlation with the brain volume of the patients. Its enrichment in patients with progressive MS was also confirmed in Western cohorts during the critical phase of EDSS 3 to 4.5, corresponding to the border between RRMS and SPMS[3]. The specific increase of *T.nexilis* prior to the definitive development of SPMS might indicate a causative role in MS progression. Recently, the immunogenicity of this species has gained much attention because *T.nexilis* stimulated a wide variety of T cell receptor repertoire against various bacterial antigens in contrast to other commensal species[35].

While analysing various *T.nexilis* strains, we identified a novel cluster of *T.nexili*s strains that was specifically enriched in SPMS. Mono-colonisation with *T.nexilis*_strain B1 from this cluster rendered germ-free mice more susceptible to EAE induction. Genome comparison between strain A1 and B1, belonging to distinct clusters, revealed numerous differences. In particular, specific genes associated with sulphate reduction and flagella formation were specifically associated with strain B1. In the present study, sulphate reduction was accelerated in the gut of strain B1 mono-colonised mice, and the liquid culture of strain B1 produced abundant H_2S, a terminal product of sulphate reduction. Although the pathological role of H_2S in neurodegenerative disorders remains controversial[36], excessive H_2S is linked to neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS)[36 37 38]. The concentration of H_2S in the cerebrospinal fluid was maintained at a high level in patients with sporadic ALS[37], and the disease-enhancing effect of H_2S as a glial-released inflammatory factor was confirmed in several studies using an animal model of ALS[37 38]. Moreover, the enrichment of sulphate-reducing bacteria in the gut of patients with Parkinson's disease and the active involvement of gut-derived H_2S in the pathogenesis of this disease were reported[39].

Considering various similarities between these neurodegenerative disorders and progressive MS, the

promotion of sulphate reduction accompanied by H_2S production in the gut might exacerbate neurodegeneration.

On the other hand, the strain B1-derived flagella would provoke over-production of IL-6 by $CD11c^{high}CD11b^{high}$ LPDCs via TLR5 stimulation. We also obtained morphological evidence that strain B1 adheres to the gut epithelium across the mucus barrier. This is probably due to the presence of the flagella, as flagellated motile bacteria can swim into an inner mucus layer where most of the other commensal bacteria cannot exist[24]. The close attachment of strain B1 to the gut epithelium provokes the secretion of SAA1. Combined effects of TLR5 signals and SAA1 were thought to induce pathogenic Th17 cells and exacerbate neuroinflammation[26]. In the present study, we analysed the specific mechanism for bystander activation of pathogenic T cells rather than that for antigen-specific T cell responses because of the following limitations: true CNS autoantigens have not been fully elucidated in patients with MS and major histocompatibility complex diversity causes inconsistent antigen-specific T cell reactivities between individuals (in contrast to EAE). However, considering the previous findings indicating a cross-reaction of T cells between MOG peptides and commensal bacteria-derived proteins in an EAE model[13], a similar mechanism may exist in patients with MS, which warrants further investigation.

Remarkably, the genes related to the flagella in strain B1 were thought to be horizontally-transferred, based on the detailed genome analysis between strain B1 and B2, where each strain underwent its own evolution under the common characteristics of vulnerability to MGEs, resulting in genomic diversities including with or without flagellar genes. On the other hand, these strains are enriched in various virulent genes as a common feature of strains included in this novel cluster of *T.nexilis*. Initially, we focused on strain B1-flagellar genes because we confirmed its pathogenicity and enrichment in SPMS (**Fig.3**). However, this was only an example and we soon noticed that the genomes of strain B1, B2, and B3 contained exceptionally high numbers of horizontally-transferred genes, having pathogenic implications. As known, frequent exchanges of MGEs between bacteria promote the spread of functional traits such as antibiotic resistance and exotoxins[40 41]. However, bacterial horizontal gene transfer in chronic human diseases has rarely been investigated in the context of pathogenic implications. Notably, anti-MGEs defence systems in strain B1, B2, and B3 were greatly reduced, indicating their relationships with horizontal gene transfer. We speculated that

the MGEs-enriched strains were selected in the potentially harsh environment by reducing their anti-MGEs defence systems to acquire advantageous auxiliary genes via horizontal gene transfer.

In the present study, we observed an enrichment of several genes related to sulphate reduction in strain B1, B2, and B3 with the deviation of faecal sulphur metabolites towards potentially reductive states. Similarities of these genes between *T.nexilis*_strain B1 and *R.gnavus*_strain AF33-12 suggested that these genes are encoded on MGEs. Moreover, four sulphur related genes included in strain B1, B2, and B3 (but not in strain A1, A2, and A3) were similar to those of phage genomes, which is consistent with the previous finding showing that many microbial genes related to sulphur metabolism are encoded on prophages[42]. Considering that several sulphate-reducing bacteria have resistance to oxidative stress[43], this feature may affect the survival of strain B in an oxidative environment. Further, strain B1, B2, and B3, but not strain A1, A2, and A3, had several resistant genes for CAMP (*dltABCD* operon). CAMP is a critical contributor to host defence against invasive bacteria. This property might indicate that these bacteria could survive in an environment where the activation of innate immunity occurs, and other susceptible bacteria cannot survive. During the process of acquiring MGEs, the bacteria obtain genes for them to adapt to the environment; however, the acquired genes may potentially promote chronic disorders such as MS. The elucidation of the evolutionary history and ecological niche of *T.nexilis* will help us to understand the precise mechanism of how pathogenic bacteria are shaped in the process of evolution.

Although our patient cohort was relatively small, the significant increase of *T.nexilis* (id03689) in untreated progressive MS, which may include both SPMS and primary progressive MS (PPMS), was confirmed by the metagenomic analysis using the large-scale validation cohort in Western countries[3]. As we could not recruit a sufficient number of patients with PPMS due to a low prevalence of the patients in Japan[44], it remains to be seen if PPMS is also characterised by increased numbers of *T.nexilis*. As *T.nexilis*_strain B is a novel cluster that we had never isolated, we encountered many difficulties in adding a genetic modification to strain B1, such as targeting flagellar genes. However, sequential results based on the comparison of several strains belonging to the same species suggested the significant role of strain B1-flagella in the context of chronic neuroinflammation.

In summary, we demonstrated that horizontal transfer of functional genes may determine the pathogenicity of gut bacteria underlying chronic disorders such as MS (**Extended data Fig.10**).

The lack of defence systems and enrichment of MGEs in disease-associated bacteria might be a promising therapeutic target in various diseases influenced by the gut microbiome.

FIGURE LEGENDS

Fig.1. Identification of gut bacteria associated with MS progression

(a) The coefficient value based on MaAsLin2 and -$\log_{10}$ *p*-value obtained from the comparison between the 62 patients with RRMS and 15 patients with SPMS are shown in the volcano plot. (b) Global microbiome alterations in progressive MS compared with HC. SPMS-enriched or depleted ($p < 0.05$ and |coefficient| >1) mOTUs are shown in the left column. The differences of these SPMS-enriched or depleted mOTUs in the comparison between the progressive MS and HC groups in each country using the validation cohort of international MS microbiome study (iMSMS) are shown. A coefficient value based on MaAsLin2 in the comparison between the disease and healthy groups is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. The coefficient values based on MaAsLin2 in the correlation between the relative abundance of each mOTU and an expanded disability scale (EDSS) score in the original and iMSMS cohorts are depicted from the lowest (purple) to the highest (orange) according to the scale shown on the right. (c) The relative abundance of *Tyzzerella nexilis* (id03689) at each EDSS score separately in the original (n = 77) and iMSMS (n = 568) cohorts of patients with MS. The non-linear regression curves for the average of the relative abundance in each EDSS group are shown in red. (d-e) Number of mapped reads per million (RPM) to the 11 strain-specific marker genes between *T.nexilis*_strain A and B MAGs (d) and their prevalence (e) among the 29 HCs, 62 patients with RRMS, and 15 patients with SPMS. Data are presented as the mean ± S.D. ns $p > 0.05$, *$p < 0.05$; MaAsLin2.

Fig.2. Genome comparison among various *Tyzzerella nexilis* strains

(a) Schematic representation of the strategy for construction of complete *T.nexilis* genomes. Five faecal samples derived from patients with RRMS and SPMS were selected as sources of *T.nexilis* strains. Closed circular genomes were obtained from two isolated strains (*T.nexilis*_strain A1 and B1) and from four non-isolated strains (*T.nexilis*_strain A2, A3, B2, and B3). The structure of the six closed circular *T.nexilis* genomes is shown. The average nucleotide identity (ANI) between *T.nexilis*_strain A MAG (or strain B MAG) and the obtained *T.nexilis* genomes is shown. (b) Phylogenetic tree of the short-read MAGs, seven closed circular genomes, and publicly available 12 genomes of *T.nexilis* based on 120 bacterial marker genes. (c) Comparisons of the various genomic

characteristics among various *T.nexilis* genomes [number of total genes, number of virulent genes in VFDB, number of insertion sequence (IS)/ prophage/ integrative and conjugative elements (ICE), and number of defence system]. The presence or absence of anti-MGE defence systems including Abortive infection (Abi), CRISPR-Cas9, DISARM, Gao, Hachiman, Restriction modification (RM), and Wadjet is also shown. The z-score based on the number of genes is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. (d) Comparisons of the number of mobile genetic elements among the 727 representative Firmicutes genomes in the GenBank and seven *T.nexilis* strains.

Fig.3. Functional characteristics of *Tyzzerella nexilis* strains

(a) Schematic of the experimental flow. (b) EAE score of germ-free (GF) mice inoculated with *T.nexilis*_strain A1 or B1. Combined results of two independent experiments are shown (GF, n=13; strain A1 and B1, n=11). (c) Representative FACS plots (gated on $CD3^+CD4^+CD8^-$). (d) Frequency of Th17 cells in the small intestine (n=5 mice), large intestine (n=5 mice), and central nervous system (GF, n=13; strain A1 and B1, n=11). (e) Differences of gene contents between strain A1 and B1 based on KEGG pathways. (f) Co-culture of heat-inactivated strains with TLR5 reporter HEK cells. (g) SEM images of cultured strains. A flagella-like structure is indicated by an arrow. (h) Co-culture of heat-inactivated strains with $CD11c^{high}CD11b^{high}$ intestinal dendritic cells. (i) SEM images on the surface of the colon in the mono-colonised mice. (j) Gene expression of *SAA1* on the colonic epithelial cells was assessed by qPCR. (k, l) The assay for Th17 cell differentiation. The concentrations of IL-17 (k) and GM-CSF (l) were assessed by ELISA. (m) Number of mapped reads per million (RPM) to *fliC* genes of *T.nexilis*_strain B1 among the 29 HCs, 62 RRMS, and 15 SPMS patients. (n) The presence or absence of KEGG orthologies (KOs) included in strain B1 but not in strain A1 in each module within sulphur metabolism (map00920). (o) The ratio of glutathione to glutathione disulphide in the faeces of GF and strain A1 or B1 mono-colonised mice (n=3 mice). (p) The production of hydrogen sulphide (H_2S) in YCFA liquid culture media was assessed by lead acetate paper. Data are presented as the mean ± S.D. ns $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; Kruskal-Wallis test with Dunn`s multiple comparisons test (b, m), two-sided unpaired t-test (d, k, l), one-way ANOVA with Tukey's multiple comparisons test (f, h, j, o).

Fig.4. Acquisition mechanism of potentially pathogenic genes

(a) Genomic comparison between *Tyzzerella nexilis*_strain B1 and *Ruminococcus gnavus*_strain AF33-12. The ORF annotations of sulphur metabolic genes and phage-aligned sulphur metabolic genes are shown coloured in red and blue, respectively. Grey connections indicate conserved regions between two genomes. (b) Phylogenetic tree of the Firmicutes genomes encoding *fliC* genes from the *T.nexilis*_strain B1 and HumGut based on 120 bacterial marker genes. Each colour shows the order of each genome. (c) Phylogenetic tree of the *fliC* genes from *T.nexilis*_strain B1 and HumGut database in five orders in Firmicutes. Each branch colour shows the order of the genomes encoding *fliC* genes. (d) Genomic comparison between *T.nexilis*_strain B1 (with flagellar genes) and B2 (without flagellar genes). The ORF annotations of flagellar genes, ICE-related genes, and ICE signature genes are shown coloured in red, blue, and brown, respectively. Insertion sequence (IS) regions are shown as blue boxes. The name of each IS indicates a family of IS and IS new means a novel family of IS. Grey connections indicate conserved regions between two genomes. The direct terminal repeat sequences and the coordinate at the edge of ICE and composite transposon including flagellar genes are shown with dashed lines of blue and black, respectively.

MAIN REFERENCES

1 Takewaki, D. *et al.* Alterations of the gut ecological and functional microenvironment in different stages of multiple sclerosis. *Proc Natl Acad Sci U S A* **117**, 22402–22412, doi:10.1073/pnas.2011703117 (2020).

2 Cox, L. M. *et al.* Gut Microbiome in Progressive Multiple Sclerosis. *Ann Neurol*, doi:10.1002/ana.26084 (2021).

3 Gut microbiome of multiple sclerosis patients and paired household healthy controls reveal associations with disease risk and course. *Cell* **185**, 3467–3486.e3416, doi:10.1016/j.cell.2022.08.021 (2022).

4 Raveney, B. J. *et al.* Eomesodermin–expressing T–helper cells are essential for chronic neuroinflammation. *Nat Commun* **6**, 8437, doi:10.1038/ncomms9437 (2015).

5 Raveney, B. J. E. *et al.* Involvement of cytotoxic Eomes–expressing CD4(+) T cells in secondary progressive multiple sclerosis. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2021818118 (2021).

6 Absinta, M. *et al.* A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis. *Nature* **597**, 709–714, doi:10.1038/s41586–021–03892–7 (2021).

7 Lassmann, H., van Horssen, J. & Mahad, D. Progressive multiple sclerosis: pathology and pathogenesis. *Nature reviews. Neurology* **8**, 647–656, doi:10.1038/nrneurol.2012.168 (2012).

8 Miyake, S. *et al.* Dysbiosis in the Gut Microbiota of Patients with Multiple Sclerosis, with a Striking Depletion of Species Belonging to Clostridia XIVa and IV Clusters. *PLoS One* **10**, e0137429, doi:10.1371/journal.pone.0137429 (2015).

9 Jangi, S. *et al.* Alterations of the human gut microbiome in multiple sclerosis. *Nat Commun* **7**, 12015, doi:10.1038/ncomms12015 (2016).

10 Berer, K. *et al.* Gut microbiota from multiple sclerosis patients enables spontaneous autoimmune encephalomyelitis in mice. *Proc Natl Acad Sci U S A* **114**, 10719–10724, doi:10.1073/pnas.1711233114 (2017).

11 Cekanaviciute, E. *et al.* Gut bacteria from multiple sclerosis patients modulate human T cells and exacerbate symptoms in mouse models. *Proc Natl Acad Sci U S A* **114**, 10713–10718, doi:10.1073/pnas.1711235114 (2017).

12 Yokote, H. *et al.* NKT cell–dependent amelioration of a mouse model of multiple sclerosis by altering gut flora. *Am J Pathol* **173**, 1714–1723, doi:10.2353/ajpath.2008.080622 (2008).

13 Miyauchi, E. *et al.* Gut microorganisms act together to exacerbate inflammation in spinal cords. *Nature* **585**, 102–106, doi:10.1038/s41586–020–2634–9 (2020).

14 Kadowaki, A. *et al.* Gut environment–induced intraepithelial autoreactive CD4(+) T cells suppress central nervous system autoimmunity via LAG–3. *Nat Commun* **7**, 11639, doi:10.1038/ncomms11639 (2016).

15 Rojas, O. L. *et al.* Recirculating Intestinal IgA–Producing Cells Regulate Neuroinflammation via IL–10. *Cell* **176**, 610–624.e618, doi:10.1016/j.cell.2018.11.035 (2019).

16 Hosang, L. *et al.* The lung microbiome regulates brain autoimmunity. *Nature* **603**, 138–144, doi:10.1038/s41586–022–04427–4 (2022).

17 Mallick, H. *et al.* Multivariable association discovery in population–scale meta–omics studies. *PLoS Comput Biol* **17**, e1009442, doi:10.1371/journal.pcbi.1009442 (2021).

18 Lloyd–Price, J. *et al.* Multi–omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662, doi:10.1038/s41586–019–1237–9 (2019).

19 Zhang, X. *et al.* The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nature medicine* **21**, 895–905, doi:10.1038/nm.3914 (2015).

20 Thomas, A. M. *et al.* Metagenomic analysis of colorectal cancer datasets identifies cross–cohort microbial diagnostic signatures and a link with choline degradation. *Nature medicine* **25**, 667–678, doi:10.1038/s41591–019–0405–7 (2019).

21 Wallen, Z. D. *et al.* Metagenomics of Parkinsons disease implicates the gut microbiome in multiple disease mechanisms. *Nat Commun* **13**, 6958, doi:10.1038/s41467–022–34667–x (2022).

491 22 Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer
492 in staphylococci by targeting DNA. *Science (New York, N.Y.)* **322**, 1843–1845,
493 doi:10.1126/science.1165771 (2008).
494 23 Uematsu, S. *et al.* Regulation of humoral and cellular gut immunity by lamina propria
495 dendritic cells expressing Toll–like receptor 5. *Nat Immunol* **9**, 769–776,
496 doi:10.1038/ni.1622 (2008).
497 24 Haiko, J. & Westerlund–Wikstrom, B. The role of the bacterial flagellum in adhesion and
498 virulence. *Biology (Basel)* **2**, 1242–1267, doi:10.3390/biology2041242 (2013).
499 25 Atarashi, K. *et al.* Th17 Cell Induction by Adhesion of Microbes to Intestinal Epithelial
500 Cells. *Cell* **163**, 367–380, doi:10.1016/j.cell.2015.08.058 (2015).
501 26 Lee, J. Y. *et al.* Serum Amyloid A Proteins Induce Pathogenic Th17 Cells and Promote
502 Inflammatory Disease. *Cell* **180**, 79–91 e16, doi:10.1016/j.cell.2019.11.026 (2020).
503 27 Peschel, A. *et al.* Inactivation of the dlt operon in Staphylococcus aureus confers
504 sensitivity to defensins, protegrins, and other antimicrobial peptides. *J Biol Chem* **274**,
505 8405–8410, doi:10.1074/jbc.274.13.8405 (1999).
506 28 Yamada, K. *et al.* Characterization of sulfur–compound metabolism underlying wax–
507 ester fermentation in Euglena gracilis. *Scientific reports* **9**, 853, doi:10.1038/s41598–
508 018–36600–z (2019).
509 29 Zou, Y. *et al.* 1,520 reference genomes from cultivated human gut bacteria enable
510 functional microbiome analyses. *Nat Biotechnol* **37**, 179–185, doi:10.1038/s41587–
511 018–0008–8 (2019).
512 30 Crost, E. H., Coletto, E., Bell, A. & Juge, N. Ruminococcus gnavus: friend or foe for human
513 health. *FEMS Microbiol Rev* **47**, doi:10.1093/femsre/fuad014 (2023).
514 31 OLeary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status,
515 taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–745,
516 doi:10.1093/nar/gkv1189 (2016).
517 32 Fullam, A. *et al.* proGenomes3: approaching one million accurately and consistently
518 annotated high–quality prokaryotic genomes. *Nucleic Acids Res* **51**, D760–d766,
519 doi:10.1093/nar/gkac1078 (2023).

33 Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell* **176**, 649–662.e620, doi:10.1016/j.cell.2019.01.001 (2019).

34 Ross, K. *et al.* TnCentral: a Prokaryotic Transposable Element Database and Web Portal for Transposon Analysis. *mBio* **12**, e0206021, doi:10.1128/mBio.02060–21 (2021).

35 Nagashima, K. *et al.* Mapping the T cell repertoire to a complex gut bacterial community. *Nature*, doi:10.1038/s41586–023–06431–8 (2023).

36 Tripathi, S. J., Chakraborty, S., Miller, E., Pieper, A. A. & Paul, B. D. Hydrogen sulfide signalling in neurodegenerative diseases. *Br J Pharmacol*, doi:10.1111/bph.16170 (2023).

37 Davoli, A. *et al.* Evidence of hydrogen sulfide involvement in amyotrophic lateral sclerosis. *Ann Neurol* **77**, 697–709, doi:10.1002/ana.24372 (2015).

38 Spalloni, A. *et al.* Impact of Pharmacological Inhibition of Hydrogen Sulphide Production in the SOD1G93A–ALS Mouse Model. *Int J Mol Sci* **20**, doi:10.3390/ijms20102550 (2019).

39 Huynh, V. A., Takala, T. M., Murros, K. E., Diwedi, B. & Saris, P. E. J. Desulfovibrio bacteria enhance alpha–synuclein aggregation in a Caenorhabditis elegans model of Parkinsons disease. *Front Cell Infect Microbiol* **13**, 1181315, doi:10.3389/fcimb.2023.1181315 (2023).

40 Moura de Sousa, J., Lourenço, M. & Gordo, I. Horizontal gene transfer among host–associated microbes. *Cell Host Microbe* **31**, 513–527, doi:10.1016/j.chom.2023.03.017 (2023).

41 Ma, Y. *et al.* Epsilon toxin–producing Clostridium perfringens colonize the multiple sclerosis gut microbiome overcoming CNS immune privilege. *J Clin Invest* **133**, doi:10.1172/jci163239 (2023).

42 Johansen, J. *et al.* Centenarians have a diverse gut virome with the potential to modulate metabolism and promote healthy lifespan. *Nature microbiology* **8**, 1064–1078, doi:10.1038/s41564–023–01370–6 (2023).

43 Dolla, A., Fournier, M. & Dermoun, Z. Oxygen defense in sulfate-reducing bacteria. *J Biotechnol* **126**, 87-100, doi:10.1016/j.jbiotec.2006.03.041 (2006).

44 Houzen, H., Kano, T., Kondo, K., Takahashi, T. & Niino, M. The prevalence and incidence of multiple sclerosis over the past 20 years in northern Japan. *Mult Scler Relat Disord* **73**, 104696, doi:10.1016/j.msard.2023.104696 (2023).

METHODS

Ethics approval and consent to participate

This study was approved by the National Center of Neurology and Psychiatry (NCNP) Ethics Committee (A2016-132 and A2021-116) and the Research Ethics Committee of RIKEN Center for Integrative Medical Sciences (RIKEN-Y-2022-030 and RIKEN-Y-2022-096). Signed informed consent was obtained from all participants who provided specimens.

Participants

In total, 77 patients with MS and 29 HC participants were included. See **Extended data Table 7 and Extended data Table 8** for clinical phenotypes of all the participants. Patients were recruited through a MS clinic at the NCNP (Tokyo, Japan). Inclusion criteria required that participants in the MS groups fulfilled the McDonald's criteria for diagnosis[45] and had MS-specific brain MRI lesions. Exclusion criteria for MS and control participants included the presence of infectious diseases and the usage of antibiotics during the collection of faecal samples. Among the 77 patients with MS, RRMS was defined based on a relapsing-remitting clinical course, whereas SPMS was diagnosed retrospectively by an attending physician based on the establishment of a sustained period of worsening neurological impairments[46]. No patients had an active relapse at the time of study enrolment. The clinical severity of the patients was evaluated using the EDSS score which is a measure of neurological impairment based on clinical assessment of MS[47]. Samples from 77 patients with MS were previously subjected to the metagenomic analysis[1].

MRI analysis

Imaging was performed on a 3-T MR system (Philips, Best, The Netherlands, or Siemens, Munich, Germany). We measured total brain volume, grey matter volume, and white matter volume from 3D-T1WI using FreeSurfer 6.0 (http://surfer.nmr.mgh.harvard.edu/), and total T2 lesion volume from 3D-T1WI and 3D-FLAIR using LST toolbox version 3.0.0 (https://www.statistical-modelling.de/lst.html). This analysis included 60 of 77 patients with MS who received brain MRI within 1 year before or after the faecal sampling.

Faecal DNA preparation

In accordance with a previously described method[48], freshly collected human faecal samples were transported at 4 °C to the laboratory in a plastic bag containing a disposable oxygen-absorbing and carbon dioxide-generating agent in which anaerobes sensitive to oxygen can survive. In the laboratory, the faecal samples were suspended in phosphate-buffered saline containing 20% glycerol, immediately frozen using liquid nitrogen and stored at -80 °C until use. Freshly collected mouse faecal samples were immediately frozen using liquid nitrogen and stored at -80 °C until use. Bacterial DNA was isolated and purified from the faecal samples according to enzymatic lysis methods[48].

qPCR analysis for faecal samples

The DNA yield was assessed using Qubit dsDNA HS Assay Kits and a Qubit 3.0 or 4.0 fluorometer (Thermo Fisher Scientific Inc., USA). Bacterial DNA was quantified using the Applied Biosystems™ StepOnePlus™ Real-time quantitative PCR system (Thermo Fisher Scientific Inc., USA) and universal primers for the 16S rRNA gene V1-V2 region 27Fmod (5′-agrgttgatymtggctcag-3′) and 338R (5′-tgctgcctcgtaggagt-3′). The *Escherichia coli* 16S rRNA gene sequence was used as standard. Fast PCR proceeded using Fast SYBR® Green Master Mix (Thermo Fisher Scientific Inc., USA) as described by the manufacturer.

Metagenomic sequencing and data analysis

The metagenomic shotgun library from patients with MS was prepared using the ThruPLEX DNA-Seq kit (TaKaRa Bio. Inc., Japan) from 50 ng metagenomic DNA according to the manufacturer's protocol. The metagenomic shotgun library from healthy participants was prepared using the Accel-NGS 1S Plus DNA (Swift Bioscience) from 250ng metagenomic DNA according to the manufacturer's protocol. After quantifying the prepared DNA library with the Kapa Library Quantification Kit (Illumina, Inc., USA), sequencing was performed using the NovaSeq 6000 sequencing system with the 150bp paired-end mode (Illumina, Inc., USA). For sequences of the ACCEL library, the first 15bp of Read 2 was trimmed. Quality filtering of the metagenomic sequences was performed using fastp (v0.20.0)[49] and ParDRe (v2.1.5)[50] to remove low-quality reads

(length <50bp and mean QV <20) and duplicated reads, respectively. PhiX and human reads were excluded by mapping the quality-filtered reads to the phiX and human genomes (hg38) using minimap2[51] with short-read mapping option (v2.13-r850). Taxonomic profiles were obtained using mOTU profiler (v3.0.1)[52] using 10 million reads per sample. Numbers in parentheses next to the species names indicate the corresponding mOTU ID. Metagenomic assembly was performed using metaSPAdes with default parameters (v3.14.0)[53]. Metagenome-assembled genomes (MAGs) were constructed using MetaBAT2 with default parameters (v2.11.2)[54]. The quality of the MAGs was assessed using CheckM (v1.1.3)[55] and GUNC (v1.0.5)[56], selecting high-quality (HQ) MAGs with estimated completeness ≥90%, contamination ≤5%, and strain heterogeneity ≤0.5%, and passing GUNC filtering. Representative HQ MAGs were selected by dereplicating the MAGs using dRep (v3.2.2)[57] with 99% identity and 90% coverage. The ANI between HQ MAGs and complete genomes of *T.nexilis* were calculated by pyani with the ANIm option[58].

Long-read sequencing

For the metagenomic Pacific Biosciences (PacBio) Sequel II sequencing (PacBio, Inc., USA), SMRTbell libraries were constructed from 5 μg faecal metagenomic DNA according to the manufacturer's protocol (Part Number 101-853-100 Version 05). Isolated bacterial DNA was obtained from 20 ml of the cultured medium by the same DNA preparation method for faecal samples. For the PacBio Sequel II multiplex sequencing of isolated bacteria, SMRTbell libraries were constructed from 1 μg DNA according to the manufacturer's protocol (Part Number 101-696-100 Version 02). HiFi reads were generated using CCS with --min-passes 3 and --min-rq 0.95 parameters (https://github.com/PacificBiosciences/ccs).

Long-read metagenomic assembly

The HiFi reads were assembled using hiCanu (v2.1.1)[59] with minReadLength=1500, minOverlapLength=1500, genomeSize=100m, maxInputCoverage=100000, minInputCoverage=1, stopOnLowCoverage=1, and -pacbio-hifi options. The circular contigs were detected by a terminal direct repeat with >95% identity and >1,000bp alignment length using minimap2. The HQ genome satisfying completeness ≥90%, contamination ≤5%, and GUNC filtering were selected from circular

contigs. The ORFs of the HQ genomes were aligned to 20 bacterial marker genes of the *T.nexilis* mOTU (ref_mOTU_v3_03689), and HQ *T.nexilis* genomes were determined, satisfying >96.5% average nucleotide similarity of the marker genes. We successfully constructed a closed circular genome of *T.nexilis* strain A3, B2, and B3 using this strategy (**Extended data Fig.4a**). In the assembly of *T.nexilis* strain A2, we could not identify *T.nexilis* HQ genome from circular contigs. Then, short reads were mapped to long-read contigs, and binning was performed using MetaBAT2. The HQ bins were identified with the same criteria as those for the circular contigs and identified as the HQ bins of the *T.nexilis* satisfying >96.5% average nucleotide similarity of the 20 bacterial marker genes of the ref_mOTU_v3_03689. The phylogenetic tree of the 120 bacterial marker genes determined using GTDB-Tk indicated that *T.nexilis*_strain A2 HQ bins were classified into cluster A. Therefore, reference-guided assembly was performed to construct a closed circular genome of the strain A2. The metagenomic HiFi reads were mapped to the strain A2 HQ bin and the strain A1 complete genome with >99% identity and >80% coverage. The mapped reads were assembled using hifiasm_meta (v0.13)[60] with default parameters and the *T.nexilis* strain A2 closed circular genome was obtained, satisfying the criteria for curating HQ genomes (**Extended data Fig.4b**).

Long-read assembly of the isolated genomes

The HiFi reads were assembled by hiCanu with minReadLength=5000, minOverlapLength=5000, genomeSize=4m, and -pacbio-hifi options. The circular contigs were detected by a terminal direct repeat with >95% identity and >1,000bp alignment length using minimap2. The genomic quality of the circular contigs was assessed by CheckM and GUNC, and high-quality genome satisfying completeness ≥95%, contamination ≤5%, and passing GUNC filtering were selected. The lack of the MGEs in the short-read MAGs compared with the complete genome obtained by long-read was visualised by Gview[61].

Abundance and prevalence of the *Tyzzerella nexilis* strains

The bacterial marker genes were predicted by fetchMG from strain A and B MAGs[52]. The 11 genes with <97% nucleotide similarity between strain A and B MAGs were selected as marker genes specific to each MAG (**Extended data Table 3**). The RPM and prevalence were obtained by

mapping short reads to the concatenated sequence of 11 marker genes using Bowtie2 with 100% identity and >10% covered fraction[62].

Comparative genomics of the *Tyzzerella nexilis* strains

The ORFs of each *T.nexilis* genome were predicted using prokka[63] with default parameters (v1.13.3). The visualisation of the complete genomes was performed by Proksee[64]. The ORFs were aligned to the eggNOG database (v5.0.2) using eggNOG mapper (ultra-sensitive mode) (v2.1.5)[65] with e-value <1e-10, and assigned K number was mapped to the KEGG pathway using the KEGG mapper[66]. The ORFs were aligned to the comprehensive virulence factor database (VFDB)[67] (download date 2022/01/14) using DIAMOND blastp (v2.0.11) (sensitive mode) with e-value <1e-20 for the annotation of the virulent genes. The ORFs were aligned to the prokaryotic virus remote homologous groups (PHROGs) (version 4)[68] using DIAMOND blastp (v2.0.11) (sensitive mode) with >70% identity and >50% coverage for the annotation of the phage-related genes. The 16S rRNA gene was predicted using RNAmmer (v1.2) with default parameter[69], and the similarity of that among the *T.nexilis* genomes was obtained from alignments using BLASTN. ICE-related genes were annotated by the alignment between the ORFs and ICEberg database (last update version: May 2, 2018)[70] using DIAMOND blastp (sensitive mode) with >50% identity and coverage. The insertion sequences were predicted using the ISEscan (v1.7.2.3)[71] with default parameters. The ICE signature genes were determined using the ICEscreen (v1.2.0)[72] with default parameters. The number of ICE in each genome was counted for the complete ICE which had all ICE signature genes determined using the ICEscreen. The defence system was predicted using the defense finder (v1.2.2)[73] with default parameters. The publicly available *T.nexilis* genomes were obtained from the Genbank database and genomes satisfying completeness ≥90% and contamination ≤5% (CheckM) were used for the comparison. The multiple alignment of the 120 bacterial marker genes of the *T.nexilis* genomes was performed using GDTB-Tk with default parameters. The phylogenetic tree was constructed using the maximum likelihood method using MEGA X[74]. The dotplot for comparing the genomic structures was obtained using Gepard[75]. To compare the number of MGEs of the *T.nexilis* genomes with those of other Firmicutes genomes, we downloaded representative complete genomes from the GenBank database (Release 255).

Genomic analysis related to the horizontal gene transfer of the sulphur metabolic genes and flagellar genes

The genomic region including sulphur metabolic genes of the *T.nexilis*_strain B1 (780,456 – 822,674) was aligned to genome sequences from human gut isolates[29] with >99% nucleotide similarity and >2kb alignment length using BLASTN and found significant alignments with *R.gnavus*_strain AF33-12 (GCA_003475365.1). The *fliC* genes in the gut bacteria were obtained from the HumGut database[76]. The ORFs of the HumGut bacterial genomes were predicted by prodigal (v2.6.3)[77] and aligned to the eggNOG database (v5.0.2) using eggNOG mapper (ultra-sensitive mode) (v2.1.5) with e-value <1e-10 and assigned K number. The K02406 (*fliC*) annotated genes were assigned as flagellin, and short (<300aa) or long genes (>600aa) were excluded from further analysis. The multiple alignment of the *fliC* genes from the *T.nexilis*_strain B1 and HumGut was obtained by mafft (v7.310)[78] with default parameters. The multiple alignment based on the 120 bacterial marker genes, including the *T.nexilis*_strain B1 and Firmicutes genomes of the HumGut, was constructed by GDTB-tk with default parameters. The phylogenetic trees of the *fliC* gene and 120 bacterial marker genes were constructed by FastTree (v2.1.10)[79] with default parameters. The visualisation of the phylogenetic tree was performed by iTOL[80]. The conserved region between the *T.nexilis*_strain B1 and B2 genomes neighbouring the flagellar gene cluster was identified by the alignment using BLASTN with >95% similarity and >10kb alignment length. Direct repeats at the edge of the composite transposon and ICE were manually identified. The abundance of *fliC* genes of strain B1 was calculated as RPM using short-read mapping by BLASTX with >80% identity and >50% coverage.

Mice

Germ-free (GF) C57BL/6 mice were originally purchased from CLEA Japan and bred in the GF facility of the NCNP. We used female mice because MS is more common in females[44]. Mice were randomly divided into groups. No statistical estimations were performed to determine sample size. The experiments were not randomised, and investigators were not blinded to allocation during experiments and outcome assessment.

728 **Bacteria**

729 All faecal samples were homogenised and diluted with phosphate-buffered saline (PBS) containing
730 20% glycerol and then stored at -80 °C for the subsequent bacterial isolation. To isolate both
731 *T.nexilis*_strain A1 and B1, the cryopreserved faecal dilutions were plated on Eggerth Gagnon agar
732 plates and incubated under anaerobic conditions (80% N_2, 10% H_2, and 10% CO_2) in an anaerobic
733 chamber (Coy Laboratory Products) at 37 °C for two or four days. The 16S rRNA genes from
734 individual colonies were amplified using the 27F and 1492R primers. Then, Sanger sequencing was
735 conducted with the 27F and 519R primers to determine the colonies corresponding to *T.nexilis*_strain
736 A1 and B1. These bacteria were grown in YCFA liquid media and stored in 20% glycerol at −80 °C.
737

738 **Gnotobiotic experiments**

739 After 18 h of anaerobic culture for *T.nexilis*_strain A1 or B1 using YCFA liquid media at 37 ℃, 200
740 µl of the media containing the same number of bacteria by turbidity measurement were orally
741 administered to female GF C57BL/6 mice at 5-7 weeks of age. Colonisation of bacteria was
742 confirmed by qPCR of faecal samples one week after the last administration (above mentioned).
743 Control GF mice were gavaged with culture media. The same procedure of bacterial administration
744 was repeated in one week. All gnotobiotic mice were kept in isolators for 2 weeks after the initial
745 administration and used for experiments.
746

747 **EAE induction**

748 For EAE induction, mice were injected subcutaneously at two sites on the back with 200 mg MOG_{35-}
749 $_{55}$ peptide (synthesised by Toray Research Center, Tokyo, Japan) and 1 mg heat-killed
750 Mycobacterium tuberculosis H37RA emulsified in complete Freund's adjuvant (Difco, KS, USA).
751 Four hundred nanograms of Pertussis toxin (List Biological Laboratories, CA, USA) were injected
752 intraperitoneally on days 0 and 2 after immunisation. EAE was clinically scored daily (0, no clinical
753 signs; 0.5, tail weakness; 1, partial tail paralysis; 1.5, severe tail paralysis; 2, flaccid tail; 2.5, flaccid
754 tail and hind limb weakness; 3, partial hind limb paralysis; 4, total hind limb paralysis; 5, hind and
755 foreleg paralysis). The mice given a score of 5 were immediately euthanised.
756

Preparation of lymphocytes

The isolation of intestinal lamina propria lymphocytes was performed as previously described[81]. Briefly, the intestines were incubated in Hank's balanced salt solution (HBSS) supplemented with 2% foetal bovine serum (FBS), 1 mM dithiothreitol and 20 mM ethylenediamine tetraacetic acid (EDTA) for 30 min at 37 °C with agitation. Then, the epithelial layer was removed by vigorous shaking in HBSS. The remaining tissues were minced and incubated in Roswell Park Memorial Institute (RPMI) 1640 supplemented with 2% FBS, 400 U/ml collagenase D (Roche), 0.25 U/ml dispase (BD Biosciences) and 0.1mg/ml DNase I (Wako) for 30 min at 37 °C with agitation. The digested tissues were suspended in 37% Percoll (GE Healthcare) and overlaid onto 70% Percoll followed by centrifugation at 800 × g for 20 min. Lymphocytes at the interface were collected. To isolate lymphocytes from the CNS, murine tissues from the mice were minced and incubated in RPMI1640 supplemented with 2% FBS, 1.33 mg/ml collagenase H (Roche) and 20 μg/ml of DNase I (Wako) for 45 min at 37 °C with agitation. Lymphocytes were collected from the 37% and 70% Percoll interphase.

Flow cytometry

For intracellular cytokine staining, the cells were stimulated with 25 ng/ml phorbol myristate acetate (PMA; Sigma) and 1 μg ionomycin (Sigma) in the presence of GoldiStop (BD Biosciences) for 4 h at 37 °C. Dead cells were stained with Zombie Aqua dye (BioLegend). After Fc receptors were blocked with anti-CD16/32 (BD Biosciences), the cells were stained with anti-CD3 (BioLegend), anti-CD4 (eBioscience), anti-CD8 (BioLegend), anti-IL-17A (BioLegend), anti-IFN-γ (BioLegend), anti-RORγt (BD Biosciences), and anti-FOXP3 (BioLegend) antibodies. All data were collected on a FACS Canto II cytometer (BD Biosciences) and analysed using Flowjo (v.10.8, Tree Star).

Th17 cell differentiation from naïve $CD4^+$ T cells

Naïve $CD4^+$ T cells were prepared from splenocyte suspensions using a $CD4^+$ T cells isolation kit according to the manufacturer's instructions (Miltenyi Biotec) with an AutoMACS Pro Instrument (Miltenyi Biotec) followed by flow cytometric sorting of $TCR\beta^+CD4^+CD25^-CD62L^{high}CD44^{low}$ cells. The 96-well flat bottom plates were coated with 20 mg/ml anti-hamster IgG at 37°C in PBS for 4 h

before washing with PBS and blocking with complete media. Approximately 2.5 × 10^4 naïve $CD4^+$ T cells were added per well and stimulated with anti-CD3 (Biolegend) and anti-CD28 (Biolegend) antibodies in the presence of the combinations of recombinant cytokines: condition 1, 20 ng/ml IL-6 (Biolegend) and 2.5 ng/ml TGF-β (R&D Systems); condition 2, 20 ng/ml IL-6 and 5 μg/ml SAA1 (R&D Systems). Supernatants were collected at 96 h, and the concentrations of IL-17 (R&D Systems) and GM-CSF (BD Biosciences) were assessed using enzyme-linked immunosorbent assay (ELISA).

Coculture with lamina propria dendritic cells

FACS-sorted $CD11c^{high}CD11b^{high}$ lamina propria dendritic cells (5 × 10^3) were cultured in the presence of heat-inactivated (60 ℃ 30 min) bacterial culture media (*T.nexilis*_strain A1 or B1) with or without 1 μM TLR5 antagonist (TH1020; MedChemExpress). Supernatants were collected at 18 h and the concentration of IL-6 was assessed by ELISA (BD Biosciences).

Isolation of colonic ECs

The harvested colon was cut open longitudinally and washed well with PBS. Tissues were incubated with 5mM EDTA in HBSS at 37 °C for 30 min with shaking to facilitate dissociation of ECs from the lamina propria. ECs were collected by centrifugation, supplemented with RNAprotect Cell Reagent (QIAGEN) and stored at -80 °C.

RNA-seq analysis of colonic epithelial cells

Total RNA was extracted from the colonic ECs using the RNeasy mini kit (Qiagen). Total RNA concentration was calculated using Quant-IT RiboGreen (Invitrogen). To assess the integrity of the total RNA, samples were run on the TapeStation RNA screentape (Agilent). Only HQ RNA preparations were used for RNA library construction. A library was independently prepared with 1 μg of total RNA for each sample using the Illumina TruSeq Stranded mRNA Sample Prep Kit (Illumina, Inc., USA). The first step in the workflow involved purifying the poly-A-containing mRNA molecules using poly-T-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperatures. The cleaved RNA

fragments were copied into first-strand cDNAs using SuperScript II reverse transcriptase (Invitrogen) and random primers. This was followed by second-strand cDNA synthesis using DNA polymerase I, RNase H, and dUTP. These cDNA fragments were then subjected to an end repair process, the addition of a single "A" base, and then ligation of the adapters. The products were then purified and enriched with PCR to create the final cDNA library. The libraries were quantified using Kapa Library Quantification kits for Illumina Sequencing platforms according to the qPCR Quantification Protocol Guide (Illumina) and qualified using the TapeStation D1000 ScreenTape (Agilent). Indexed libraries were then submitted to an Illumina NovaSeq (Illumina, Inc.), and the paired-end (2×100 bp) sequencing was performed by Macrogen Incorporated. The raw reads from the sequencer were preprocessed to remove low-quality reads and the adapter sequence before analysis, and the processed reads were aligned to the *Mus musculus (mm10)* using HISAT (v2.1.0)[82]. HISAT utilises two types of indexes for alignment (a global, whole-genome index and tens of thousands of small local indexes). Both types of indexes are constructed using the same Burrows–Wheeler transform (BWT) and graph FM index (GFM) as those of Bowtie2. Because of its use of these efficient data structures and algorithms, HISAT generates spliced alignments several times faster than Bowtie and BWA, which are widely used. The reference genome sequence of *Mus musculus (mm10)* and annotation data were downloaded from the UCSC table browser (http://genome.uscs.edu). After alignment, StringTie was used to assemble aligned reads into transcripts and to estimate their abundance[83 84]. It provides the relative abundance estimates as read count values of the transcript and gene expressed in each sample. Additionally, transcript assembly of known transcripts, novel transcripts, and alternative splicing transcripts was processed using StringTie. Based on that result, transcript and gene expression abundances were calculated as read counts or fragments per kilobase of exon per million mapped reads (FPKM) value per sample. Gene-enrichment and functional annotation analyses for significant gene lists were performed based on gProfiler (https://biit.cs.ut.ee/gprofiler/orth).

qPCR analysis for colonic ECs

Total RNA was extracted from the colonic ECs using the RNeasy mini kit (Qiagen), and cDNA was obtained with PrimeScript™ RT Master Mix (TaKaRa). qPCR was performed using SYBR Premix

Ex Taq (TaKaRa) on a Thermal Cycler Dice Real Time System (TaKaRa). The following primer pairs were used: *GAPDH*, 5'-GTCGTGGAGTCTACTGGTGTCTTC-3' and 5'-GTCATATTTCTCGTGGTTCACACC-3'; *SAA1*, 5'- TGTTCACGAGGCTTTCCAAG-3' and 5'-CCCGAGCATGGAAGTATTTG-3'.

SEM

After the removal of intestinal contents, the large intestine tissues were washed well with PBS and fixed in 2.5% glutaraldehyde in 50 mM phosphate (pH 7.2). The plate-cultured bacterial colony was also fixed in 2.5% glutaraldehyde in 50 mM phosphate (pH 7.2). The samples were post-fixed in 1% osmium tetroxide in 50 mM phosphate buffer (pH 7.2) and dehydrated with increasing ethanol concentrations. After dehydration, the samples were point-dried using carbon dioxide (EMCPD300, Leica), coated with platinum (intestinal tissue) or osmium (cultured bacteria), and observed under a scanning electron microscope (SU8240; Hitachi High-Technologies).

TLR5 reporter HEK293 cell assay

After 18 h anaerobic culture for *T.nexilis*_strain A1 or strain B1 using YCFA liquid media at 37 ℃ with negative control just containing media, the samples were prepared. The TLR5-simulating capacity was assessed using HEK-Blue^{TM} mTLR5 cells (InvivoGen) and heat-inactivated (60 ℃ 30 min) bacterial culture media, following the manufacturer's protocol.

Validation of bacterial signatures with public data

We obtained publicly available metagenomic data (iMSMS: ERP115476; IBD: IBDMDB website [https://ibdmdb.org]; rheumatoid arthritis: PRJEB6997; colorectal cancer: PRJEB10878; Parkinson's disease: PRJNA834801) and clinical information from the cohorts of iMSMS[3], IBD[18], rheumatoid arthritis[19], colorectal cancer[20], and Parkinson's disease[21] as validation cohorts. We analysed 1,139 samples (435 patients with RRMS, 133 patients with progressive MS, and 571 HCs) whose sequence depth of metagenomic reads pooled based on sample ID exceeded one million reads in the iMSMS cohort. We also analysed 104 samples (26 non-IBD individuals, 29 patients with ulcerative colitis, and 49 patients with Crohn's disease), 141 samples (80 HCs and 61 patients with rheumatoid

arthritis), 126 samples (53 HCs and 73 patients with colorectal cancer), and 725 samples (234 HCs and 491 patients with Parkinson's disease) whose sequence depth of metagenomic reads exceeded 10 million reads in these cohorts. Taxonomic profiles were performed with mOTU profiler (v3.0.1) to compare with our original data.

Sulphur metabolomics

The freshly collected faecal samples were frozen and stored at −80 ℃ until use. Sulphur metabolomic analyses were outsourced to Sulphur Index service (Tokyo), Japan, with liquid chromatography coupled to a tandem mass spectrometry (LC-MS/MS) system as described previously[85]. Briefly, the sulphur-containing compounds in the samples were extracted by adding methanol and were converted to fluorescent derivatives with monobromobimane. The target metabolite levels were determined from the peak area by mass chromatography and were represented as relative amounts after normalisation with the peak area of the internal standard (D-camphor-10-sulphonic acid). In the process of data analysis, we added 0.0001 (<the lowest value) to all values of glutathione and glutathione disulphides to calculate the ratio of glutathione-to-glutathione disulphides because glutathione disulphides were not detected in several samples from GF mice.

Measurement of H_2S production

After 48 h of anaerobic culture for YCFA liquid media (1ml) supplemented with *T.nexilis*_strain A1 or B1 with the same number of bacteria by turbidity measurement in a sealed 50ml conical tube at 37 ℃, the concentration of H_2S in the gas phase above the liquid phase was semi-quantitatively assessed by lead acetate paper (Supelco) attached to the inner wall of the tube.

Statistical analysis

All statistical analyses were conducted with R version 4.2.0 or GraphPad Prism version 9.3.1. To compare two groups of data, statistical differences were evaluated using the two-tailed unpaired *t*-test (for normally distributed variables) or the two-tailed unpaired Wilcoxon rank-sum test (non-normally distributed variables). MaAsLin2[17] was used to compare the microbiome data between participant groups. Differences between more than three groups were evaluated using one-way ANOVA

followed by Tukey's multiple comparisons test (for normally distributed variables) or the Kruskal-Wallis test followed by Dunn's multiple comparison test (for non-normally distributed variables). Spearman's rank correlation coefficient and MaAsLin2[17] were used in the correlation analysis. The chi-square test was used to compare the patient demographic data. The significance level was set at p-value < 0.05.

Data availability

The short-read metagenomic sequences from 77 patients with MS and 29 HCs analysed in the present study will be deposited in DDBJ/GenBank/EMBL before publication. The long-read metagenomic sequences from 3 patients with MS will be also deposited in DDBJ/GenBank/EMBL before publication. The closed circular genomic sequences from seven *T.nexilis* strains (DSM1787, A1, A2, A3, B1, B2, and B3) will be deposited in DDBJ/GenBank/EMBL before publication.

ACKNOWLEDGMENTS

The authors thank M. Hattori for providing the research environment and supporting the projects. The authors thank C. Shindo, N. Kumar, Z. Iioka, Y. Hattori, H. Yamaguchi, A. Kimura, and H. Morita for technical support. The authors thank A. Tomita, D. Noto, M. Nakamura, N. Chihara, T. Matsuoka, and R. Ikeguchi for obtaining informed patient consent and collecting samples. The authors thank H. Sasaki, R. Sato, and M. Sato for supporting the electron microscopic imaging analysis. The authors thank I. Ohtsu for analysing sulphur metabolomics in the faecal samples. The authors thank RIKEN-IMS Genome Platform for high-throughput sequencing. This study was supported by the Practical Research Project for Rare/Intractable Diseases from Japan Agency for Medical Research and Development, AMED (19ek0109417h0001, 20ek0109417h0002, 21ek0109417h0003) and AMED-CREST (18gm1010011h0001, 19gm1010011h0002, 20gm1010011h0003, 21gm1010011h0004, 22gm1010011h0005, and 23gm1010011h0006).

AUTHOR CONTRIBUTIONS

D.Takewaki, Y.Kiguchi, W.Suda, and T.Yamamura planned the study. D.Takewaki, W.Sato, S.Miyake, and T.Yamamura contributed to the collection of samples and clinical data. Y.Kiguchi, H.Masuoka, Y.Ogata, R.Kurokawa, and W.Suda contributed to the collection and analysis of the primary sequence data of faecal samples. D.Takewaki, Y.Kiguchi., and H.Masuoka mainly conducted the metagenomic analysis. D.Takewaki, H.Masuoka, and S.Narushima isolated the bacteria. D.Takewaki, H.Masuoka, S.Narushima, Y.Ozawa., S.Yagishita, T.Araki, and T.Yamamura contributed to the gnotobiotic animal experiments. D.Takewaki, M.Manu, and B.Raveney conducted immunological experiments. Y.Kimura and N.Sato conducted brain MRI analyses. D.Takewaki and Y.Kiguchi wrote the first draft. W.Suda and T.Yamamura contributed to the completion of the manuscript. All authors read, critically revised for important intellectual contents, and approved the final manuscript.

COMPETING INTERESTS

D.Takewaki received support for research from Yakult Bio-Science Foundation. W. Sato received support for clinical trials from Chugai Pharmaceutical Co., Ltd. and Biogen Japan Ltd. T. Yamamura

served on the scientific advisory board of Biogen Japan Ltd. and received support for research and clinical trials from Sanofi K.K., Chugai Pharmaceutical Co., Ltd., UCB Japan Co., Ltd., Biogen Japan Ltd., Chiome Bioscience Inc., Novartis Pharma K.K., and Mebix, Inc.

EXTENDED DATA

Extended data Fig.1. Universality and specificity in the association between *Tyzzerella nexilis* and progressive MS

(a-e) The relative abundance of *T.nexilis* (id03689) among the HC, RRMS, and SPMS (or progressive MS) groups in our original (a: Japan) and iMSMS cohorts (b: United States, c: United Kingdom, d: Argentina, e: untreated). (f) mOTU analysis using the publicly available and original metagenomic data. The 40 mOTUs whose relative abundance was significantly enriched or depleted in the SPMS group relative to healthy control (HC) group are shown (p <0.05 and |coefficient| >1). We compared the abundance of these 40 mOTUs between the HC and other disease groups, including RRMS (571 HCs and 435 patients)[3], IBD (26 non-IBD individuals, 29 patients with ulcerative colitis, and 49 patients with Crohn's disease)[18], rheumatoid arthritis (80 HCs and 61 patients)[19], colorectal cancer (53 HCs and 73 patients)[20], and Parkinson's disease (234 HCs and 491 patients)[21]. A coefficient value based on MaAsLin2 in a comparison between the disease and healthy groups is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. Data are presented as mean ± S.D. *p <0.05, **p <0.01, ***p <0.001; MaAsLin2.

Extended data Fig.2. Associations between microbial abundance and a clinical severity score

(a-b) The relative abundance of four mOTUs at each EDSS score separately in the original (n = 77) and iMSMS (n = 568) cohorts of patients with MS. The four mOTUs were selected as the species whose relative abundance was the most positively correlated with an EDSS score other than *Tyzzerella nexilis* (id03689) based on a coefficient value of MaAsLin2 among 40 mOTUs that were highly enriched or depleted in the original SPMS group compared with the HC group. Curve fitting using non-linear regression for the average of the relative abundance in each group was conducted: the non-linear curves are shown in red.

Extended data Fig.3. Correlation between microbial abundance and quantitative brain MRI parameters

(a) Among the 77 patients with MS, 60 patients who underwent brain MRI within 1 year before or after the faecal sampling were included in this analysis, and the brain MRI parameters between

patients with RRMS (n=48) and SPMS (n=12) were compared. (b) The 40 mOTUs whose relative abundance was significantly enriched or depleted in the SPMS group compared with the HC group are shown (p <0.05 and |coefficient| >1). We analysed the correlation between the abundance of these 40 mOTUs and four brain MRI parameters including total brain volume, grey matter volume, white matter volume, and total T2 lesion volume among 60 patients with MS. The Spearman coefficient R in the correlation between the relative abundance of each mOTU and brain MRI parameters is depicted from the lowest (purple) to the highest (orange) according to the scale shown on the right. *p <0.05, **p <0.01; Wilcoxon rank-sum test (a), Spearman's rank correlation coefficient (b).

Extended data Fig.4. Workflow of the long-read metagenomic assembly

(a) Workflow for constructing the *Tyzzerella nexilis* genome from closed circular contigs. The strain A3, B2, and B3 genomes were constructed using this strategy. First, the metagenomic HiFi reads were assembled by hiCanu. Then, the high-quality (HQ) chromosomal contigs were selected by CheckM and GUNC from circular contigs. Finally, HQ genomes having marker genes of *T.nexilis* (ref_mOTU_v3_03689) with >96.5% identity were selected. (b) Workflow for combinational analysis of long-read metagenomic assembly, binning, and reference-guided assembly. The *T.nexilis*_strain A2 genome was constructed using this strategy. The metagenomic short reads were mapped to the linear contigs obtained by metagenomic assembly using HiFi reads. Then, the binning analysis and quality check of bins were performed by metabat2 and CheckM, respectively. The marker genes of *T.nexilis* (ref_mOTU_v3_03689) were aligned to HQ-bins and obtained HQ-bin of *T.nexilis* (cluster A) consisting of three contigs. For collecting *T.nexilis*-related HiFi reads, all metagenomic HiFi reads were mapped to the HQ-bin of *T.nexilis* and the complete genome of strain A1. Finally, *T.nexilis*-related HiFi reads were assembled by hifiasm_meta and obtained a closed circular genome of the *T.nexilis*_strain A2. (a,b) Details of these workflows are described in the Method section.

Extended data Fig.5 Comparisons of the genomic structures between MS patients-derived *Tyzzerella nexilis* strains

Dot plot diagram for genomic similarity among the six genomes of *T.nexilis*_strain A1, A2, A3, B1, B2, and B3.

Extended data Fig.6. The genomic comparison between short-read MAG and complete genome in *Tyzzerella nexilis*_strain B1

Comparison of genomic structures between the short-read MAG and the isolation-based complete genome of mobile genetic elements (MGEs)-enriched *T.nexilis*_strain B1. The predicted regions of integrative and conjugative elements (ICE), insertion sequences (IS), and prophages are shown. Black arrows indicate the specific regions predicted as MGEs among the missing regions in short-read MAGs.

Extended data Fig 7. Functional analysis in the comparison between *Tyzzerella nexilis* strains

(a) Quantification of 16S rRNA gene copies in the faeces of GF and *T.nexilis*_strain A1 or B1 mono-colonised mice (n=3 mice). (b) Representative FACS plots (gated on $CD3^+CD4^+CD8^-$). (c) Frequency of regulatory T cells in the small intestine (n=5 mice), large intestine (n=5 mice), and central nervous system (GF, n=13; *T.nexilis*_strain A1 and B1, n=11). (d-f) Gene expression profile of colonic ECs between *T.nexilis*_strain B1 mono-colonised mice (n=4), *T.nexilis*_strain A1 mono-colonised mice (n=4), and germ-free (GF) mice (n=4). Differential gene expression in the colonic ECs was analysed by RNA-seq. (d) Gene ontology (GO) terms significantly enriched in up-regulated gene sets in the colonic ECs derived from the strain B1 mono-colonised mice compared with those from the strain A1 mono-colonised mice are shown. (e) Comparison of the genes annotated to α-defensin between the GF, strain A1 mono-colonised, and strain B1 mono-colonised mice. The z-score based on the transcripts per kilobase million (TPM) is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. (f) TPM of *SAA1* between the GF, strain A1 mono-colonised, and strain B1 mono-colonised mice. (g) The presence or absence of KEGG orthologies (KOs) included in strain B1, but not in strain A1 in each module within the cationic antimicrobial peptide resistance (map01503). Data are presented as mean ± S.D. ns $p > 0.05$, *p

<0.05, **p <0.01, ***p <0.001, ****p <0.0001, two-sided unpaired t-test (c), one-way ANOVA with Tukey's multiple comparisons test (f).

Extended data Fig.8. Genome comparisons among seven *Tyzzerella nexilis* strains

The presence or absence of KEGG orthologies (KOs) included in flagellar assembly, sulphur metabolism, and CAMP resistance pathways within the genomes of *T.nexilis*_DSM1787, strain A1, A2, A3, B1, B2, and B3.

Extended data Fig.9. Extended data Fig.9. Phylogeny of the *fliC* genes

Phylogenetic tree of the *fliC* genes from *T.nexilis*_strain B1 and the HumGut database. Each branch colour shows the phylum of the genomes encoding *fliC* genes.

Extended data Fig.10. The nature and functions of bacteria associated with multiple sclerosis progression

Mobile genetic elements (MGEs)-enriched *T.nexilis* strains are abundant in the gut of patients with SPMS. Some of the strains acquired pathogenic genes associated with flagella formation and sulphate reduction via horizontal gene transfer from another microorganism. These strains-derived flagella potentially induce pathogenic Th17 cells via the combination of TLR5 stimulation and promotion of adherence to colonic ECs. These T cells might migrate into the CNS and accelerate neuroinflammation. The sulphate reduction accompanied by hydrogen sulphide (H_2S) production is potentially associated with neurodegeneration. The depletion of defence systems and subsequent enrichment of MGEs might have shaped the pathogenic bacteria underlying MS progression. Abbreviations: IS = insertion sequence; ICE = integrative conjugative element; TLR5 = toll-like receptor 5; SAA1 = serum amyloid A1; CAMP; cationic antimicrobial peptide resistance.

Extended data Table 1. Sequence statistics of the NovaSeq and Sequel sequencers

The number of filter-passed* NovaSeq reads and total bases of filter-passed* NovaSeq reads are presented. The number of Sequel HiFi reads, total bases of Sequel HiFi reads, and average reads length are also presented.

Extended data Table 2. Significantly enriched or depleted mOTUs in SPMS compared with RRMS

The 38 mOTUs whose relative abundance was significantly enriched or depleted in the SPMS group compared with the RRMS group are presented in the descending order of coefficient value of MaAsLin2. mOTUs whose coefficient value >0 are SPMS-enriched mOTUs and those whose coefficient value <0 are RRMS-enriched mOTUs.

Extended data Table 3. Strain-specific marker genes of *Tyzzerella nexilis*_strain A or B MAG

The list of 11 bacterial marker genes for calculating *T.nexilis*_strain A and B MAGs are shown. All genes have <97% nucleotide similarity between strain A and B MAGs.

Extended data Table 4. The number of mobile genetic elements in seven *Tyzzerella nexilis* strains and 726 representative Firmicutes strains

Seven *T.nexilis* genomes used in the present study and 726 representative complete genomes assigned to Firmicutes in the GenBank database were analysed. The number of MGEs including insertion sequences (IS), integrative and conjugative elements (ICE), and prophages are presented.

Extended data Table 5. 16S rRNA gene similarity between the genomes of seven *Tyzzerella nexilis* strains

Similarity of nucleotide sequences in full-length 16S rRNA genes between the seven *T.nexilis* genomes.

Extended data Table 6. Gene annotations of *Tyzzerella nexilis*_strain B1

Gene annotations of the *T.nexilis*_strain B1 genome. Gene annotations based on Pfam, KEGG orthologies, and ICE-related genes are shown.

Extended data Table 7. Demographics and characteristics of the patients and controls

Sixty-two patients with RRMS, 15 patients with SPMS, and 29 healthy controls were recruited. Data are represented as mean ± SEM. Abbreviations: BMI = body mass index; ARR = annual relapse rate; EDSS = expanded disability status scale; PSL = prednisolone; IS = immunosuppressive drugs; DMDs; disease-modifying drugs; IFN-β = interferon-β; GA = glatiramer acetate; FTY = fingolimod; NTZ = natalizumab; DMF = dimethyl fumarate; HC = healthy control; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis.

Extended data Table 8. Clinical information of patients and controls

Detailed clinical information including recruitment site, ethnicity, age, sex, body mass index (BMI), year of onset, disease duration, treatments, and EDSS score is presented.

METHOD REFERENCES

45 Thompson, A. J. *et al.* Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *Lancet Neurol* **17**, 162-173, doi:10.1016/s1474-4422(17)30470-2 (2018).

46 Rovaris, M. *et al.* Secondary progressive multiple sclerosis: current knowledge and future challenges. *Lancet Neurol* **5**, 343-354, doi:10.1016/s1474-4422(06)70410-0 (2006).

47 Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* **33**, 1444-1452 (1983).

48 Kim, S. W. *et al.* Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. *DNA research : an international journal for rapid publication of reports on genes and genomes* **20**, 241-253, doi:10.1093/dnares/dst006 (2013).

49 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890, doi:10.1093/bioinformatics/bty560 (2018).

50 González-Domínguez, J. & Schmidt, B. ParDRe: faster parallel duplicated reads removal tool for sequencing studies. *Bioinformatics* **32**, 1562-1564, doi:10.1093/bioinformatics/btw038 (2016).

51 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100, doi:10.1093/bioinformatics/bty191 (2018).

52 Milanese, A. *et al.* Microbial abundance, activity and population genomic profiling with mOTUs2. *Nat Commun* **10**, 1014, doi:10.1038/s41467-019-08844-4 (2019).

53 Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome research* **27**, 824-834, doi:10.1101/gr.213959.116 (2017).

54 Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359, doi:10.7717/peerj.7359 (2019).

55 Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25**, 1043-1055, doi:10.1101/gr.186072.114 (2015).

56 Orakov, A. *et al.* GUNC: detection of chimerism and contamination in prokaryotic genomes. *Genome biology* **22**, 178, doi:10.1186/s13059-021-02393-0 (2021).

57 Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *Isme j* **11**, 2864-2868, doi:10.1038/ismej.2017.126 (2017).

58 Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G. & Toth, I. K. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods* **8**, 12-24, doi:10.1039/c5ay02550h (2016).

59 Nurk, S. *et al.* HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome research* **30**, 1291-1305, doi:10.1101/gr.263566.120 (2020).

60 Feng, X., Cheng, H., Portik, D. & Li, H. Metagenome assembly of high-fidelity long reads with hifiasm-meta. *Nature methods* **19**, 671-674, doi:10.1038/s41592-022-01478-3 (2022).

61 Petkau, A., Stuart-Edwards, M., Stothard, P. & Van Domselaar, G. Interactive microbial genome visualization with GView. *Bioinformatics* **26**, 3125-3126, doi:10.1093/bioinformatics/btq588 (2010).

62 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

63 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).

64 Grant, J. R. *et al.* Proksee: in-depth characterization and visualization of bacterial genomes. *Nucleic Acids Res* **51**, W484-w492, doi:10.1093/nar/gkad326 (2023).

65 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol* **38**, 5825-5829, doi:10.1093/molbev/msab293 (2021).

66 Kanehisa, M. & Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* **29**, 28-35, doi:10.1002/pro.3711 (2020).

67 Liu, B., Zheng, D., Zhou, S., Chen, L. & Yang, J. VFDB 2022: a general classification scheme for bacterial virulence factors. *Nucleic Acids Res* **50**, D912-d917, doi:10.1093/nar/gkab1107 (2022).

68 Terzian, P. *et al.* PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform* **3**, lqab067, doi:10.1093/nargab/lqab067 (2021).

69 Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100-3108, doi:10.1093/nar/gkm160 (2007).

70 Liu, M. *et al.* ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Research* **47**, D660-D665, doi:10.1093/nar/gky1123 (2018).

71 Xie, Z. & Tang, H. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics* **33**, 3340-3347, doi:10.1093/bioinformatics/btx433 (2017).

72 Lao, J. *et al.* ICEscreen: a tool to detect Firmicute ICEs and IMEs, isolated or enclosed in composite structures. *NAR Genom Bioinform* **4**, lqac079, doi:10.1093/nargab/lqac079 (2022).

73 Tesson, F. *et al.* Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nat Commun* **13**, 2561, doi:10.1038/s41467-022-30269-9 (2022).

74 Stecher, G., Tamura, K. & Kumar, S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Mol Biol Evol* **37**, 1237-1239, doi:10.1093/molbev/msz312 (2020).

75 Krumsiek, J., Arnold, R. & Rattei, T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. *Bioinformatics* **23**, 1026-1028, doi:10.1093/bioinformatics/btm039 (2007).

76 Hiseni, P., Rudi, K., Wilson, R. C., Hegge, F. T. & Snipen, L. HumGut: a comprehensive human gut prokaryotic genomes collection filtered by metagenome data. *Microbiome* **9**, 165, doi:10.1186/s40168-021-01114-w (2021).

77 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (2010).

78 Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059-3066, doi:10.1093/nar/gkf436 (2002).

79 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641-1650, doi:10.1093/molbev/msp077 (2009).

80 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* **47**, W256-w259, doi:10.1093/nar/gkz239 (2019).

81 Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature* **504**, 446-450, doi:10.1038/nature12721 (2013).

82 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).

83 Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**, 1650-1667, doi:10.1038/nprot.2016.095 (2016).

84 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295, doi:10.1038/nbt.3122 (2015).

85 Kawano, Y. *et al.* Involvement of the yciW gene in l-cysteine and l-methionine metabolism in Escherichia coli. *J Biosci Bioeng* **119**, 310-313, doi:10.1016/j.jbiosc.2014.08.012 (2015).

Figure 1

Horizontal gene transfer shapes pathogenic bacteria in multiple sclerosis

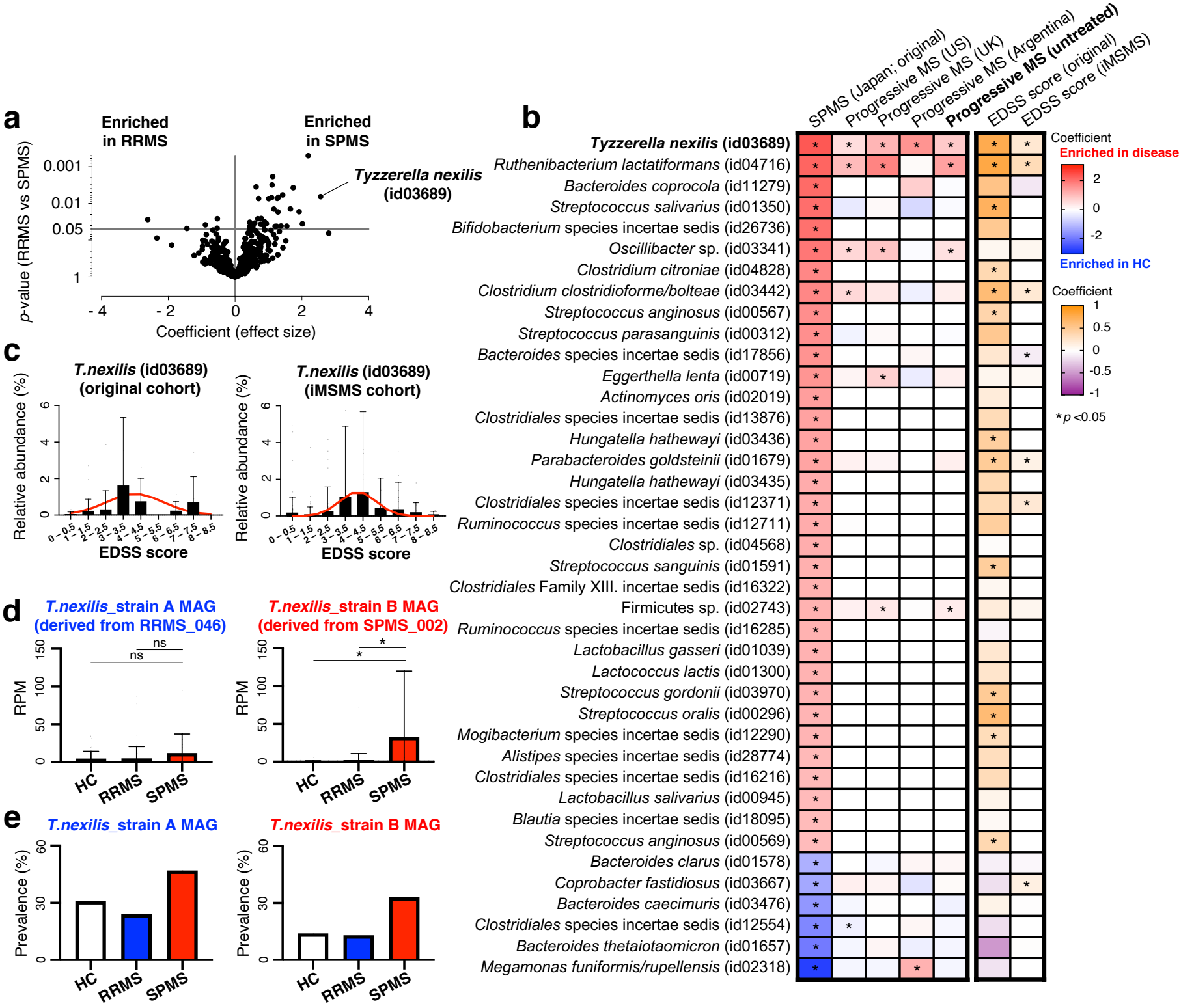
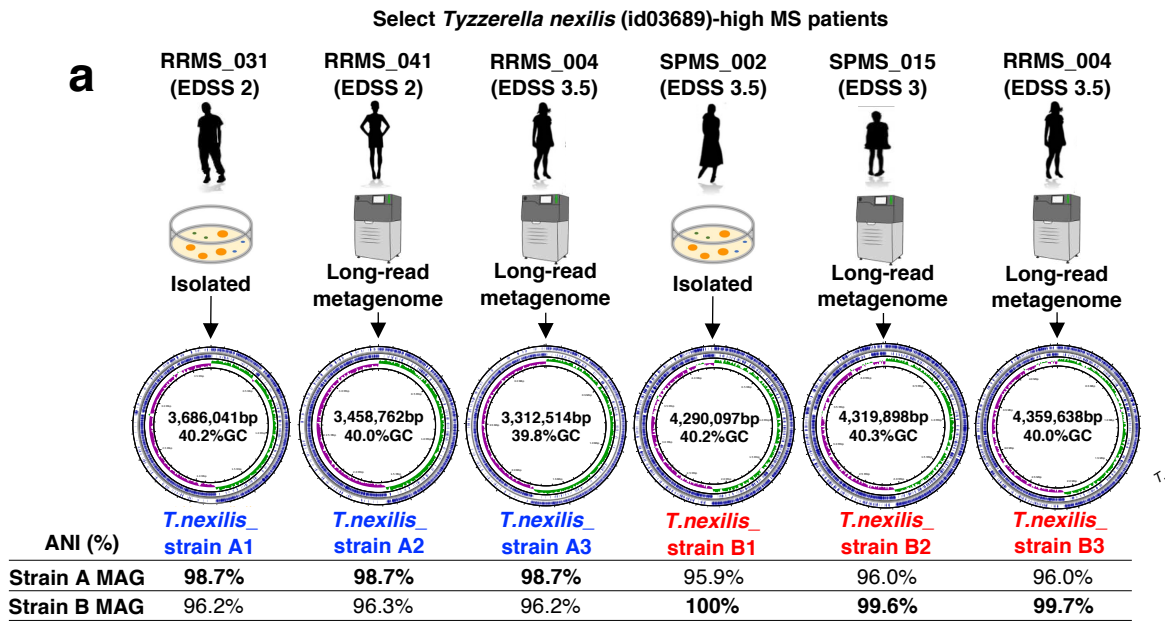


Fig. Identification of gut bacteria associated with MS progression

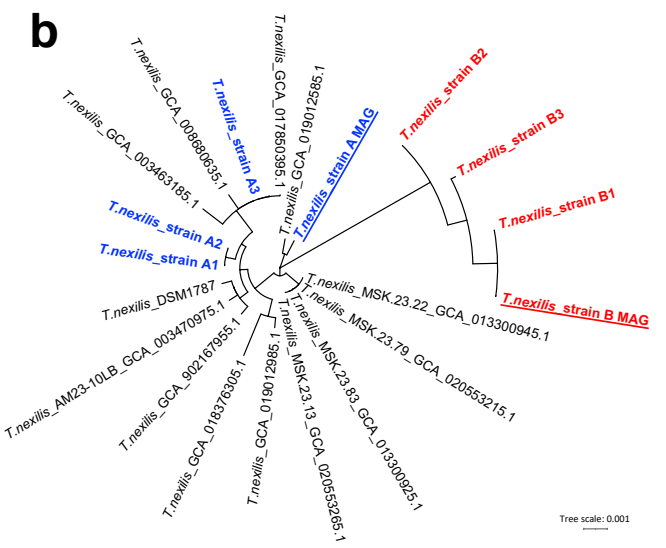
(**a**) The coefficient value based on MaAsLin2 and -$\log_{10}$ *p*-value obtained from the comparison between the 62 patients with RRMS and 15 patients with SPMS are shown in the volcano plot. (**b**) Global microbiome alterations in progressive MS compared with HC. SPMS-enriched or depleted ($p < 0.05$ and lcoefficientl >1) mOTUs are shown in the left column. The differences of these SPMS-enriched or depleted mOTUs in the comparison between the progressive MS and HC groups in each country using the validation cohort of international MS microbiome study (iMSMS) are shown. A coefficient value based on MaAsLin2 in the comparison between the disease and healthy groups is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. The coefficient values based on MaAsLin2 in the correlation between the relative abundance of each mOTU and an expanded disability scale (EDSS) score in the original and iMSMS cohorts are depicted from the lowest (purple) to the highest (orange) according to the scale shown on the right. (**c**) The relative abundance of *Tyzzerella nexilis* (id03689) at each EDSS score separately in the original (n = 77) and iMSMS (n = 568) cohorts of patients with MS. The non-linear regression curves for the average of the relative abundance in each EDSS group are shown in red. (**d-e**) Number of mapped reads per million (RPM) to the 11 strain-specific marker genes between *T.nexilis*_strain A and B MAGs (**d**) and their prevalence (**e**) among the 29 HCs, 62 patients with RRMS, and 15 patients with SPMS. Data are presented as the mean ± S.D. ns $p > 0.05$, *$p < 0.05$; MaAsLin2.

Figure 2

Horizontal gene transfer shapes pathogenic bacteria in multiple sclerosis



ANI (%)	T.nexilis_ strain A1	T.nexilis_ strain A2	T.nexilis_ strain A3	T.nexilis_ strain B1	T.nexilis_ strain B2	T.nexilis_ strain B3
Strain A MAG	**98.7%**	**98.7%**	**98.7%**	95.9%	96.0%	96.0%
Strain B MAG	96.2%	96.3%	96.2%	**100%**	**99.6%**	**99.7%**



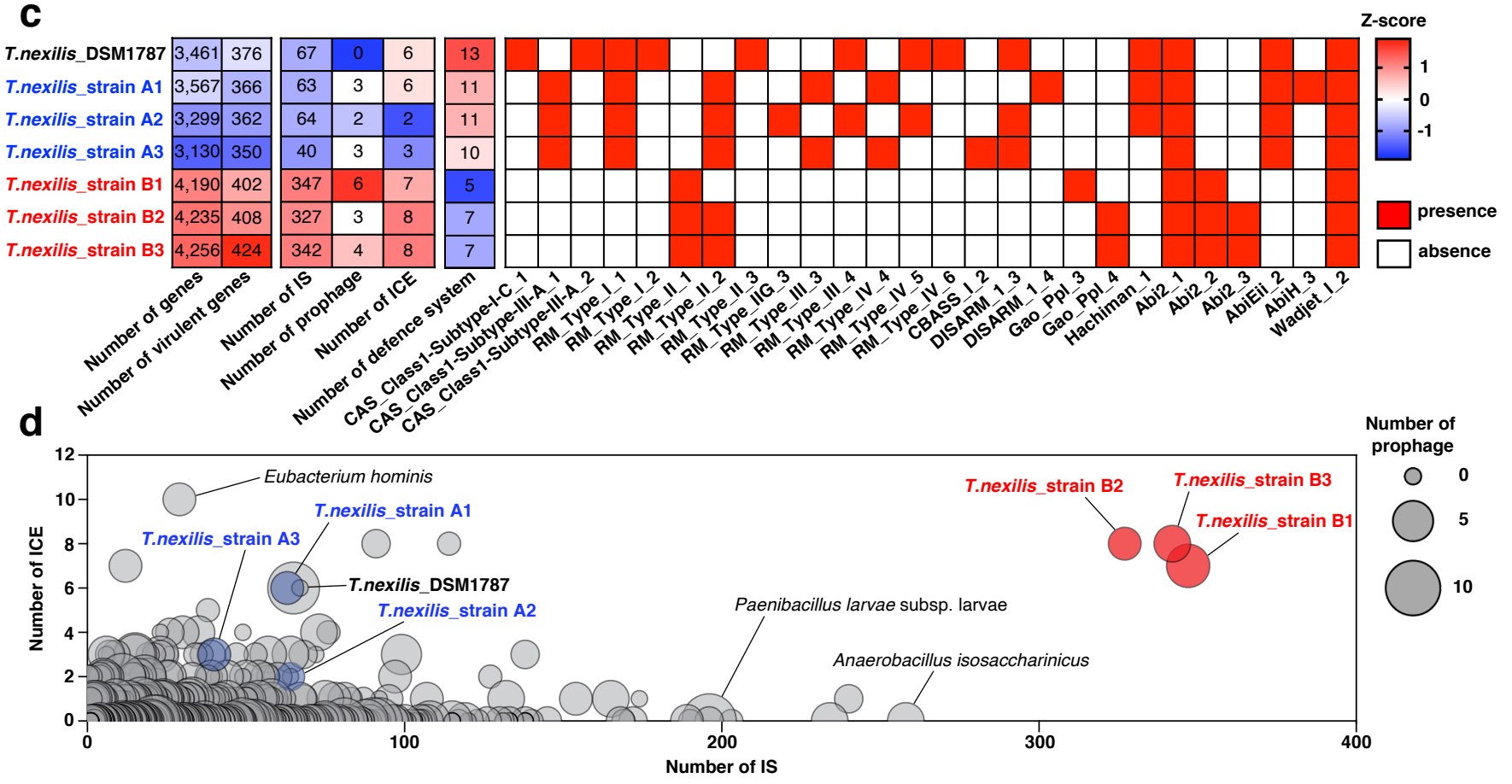


Fig.2. Genome comparison among various *Tyzzerella nexilis* strains

(**a**) Schematic representation of the strategy for construction of complete *T.nexilis* genomes. Five faecal samples derived from patients with RRMS and SPMS were selected as sources of *T.nexilis* strains. Closed circular genomes were obtained from two isolated strains (*T.nexilis*_strain A1 and B1) and from four non-isolated strains (*T.nexilis*_strain A2, A3, B2, and B3). The structure of the six closed circular *T.nexilis* genomes is shown. The average nucleotide identity (ANI) between *T.nexilis*_strain A MAG (or strain B MAG) and the obtained *T.nexilis* genomes is shown. (**b**) Phylogenetic tree of the short-read MAGs, seven closed circular genomes, and publicly available 12 genomes of *T.nexilis* based on 120 bacterial marker genes. (**c**) Comparisons of the various genomic characteristics among various *T.nexilis* genomes [number of total genes, number of virulent genes in VFDB, number of insertion sequence (IS)/ prophage/ integrative and conjugative elements (ICE), and number of defence system]. The presence or absence of anti-MGE defence systems including Abortive infection (Abi), CRISPR-Cas9, DISARM, Gao, Hachiman, Restriction modification (RM), and Wadjet is also shown. The z-score based on the number of genes is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. (**d**) Comparisons of the number of mobile genetic elements among the 727 representative Firmicutes genomes in the GenBank and seven *T.nexilis* strains.

Figure 3

Horizontal gene transfer shapes pathogenic bacteria in multiple sclerosis

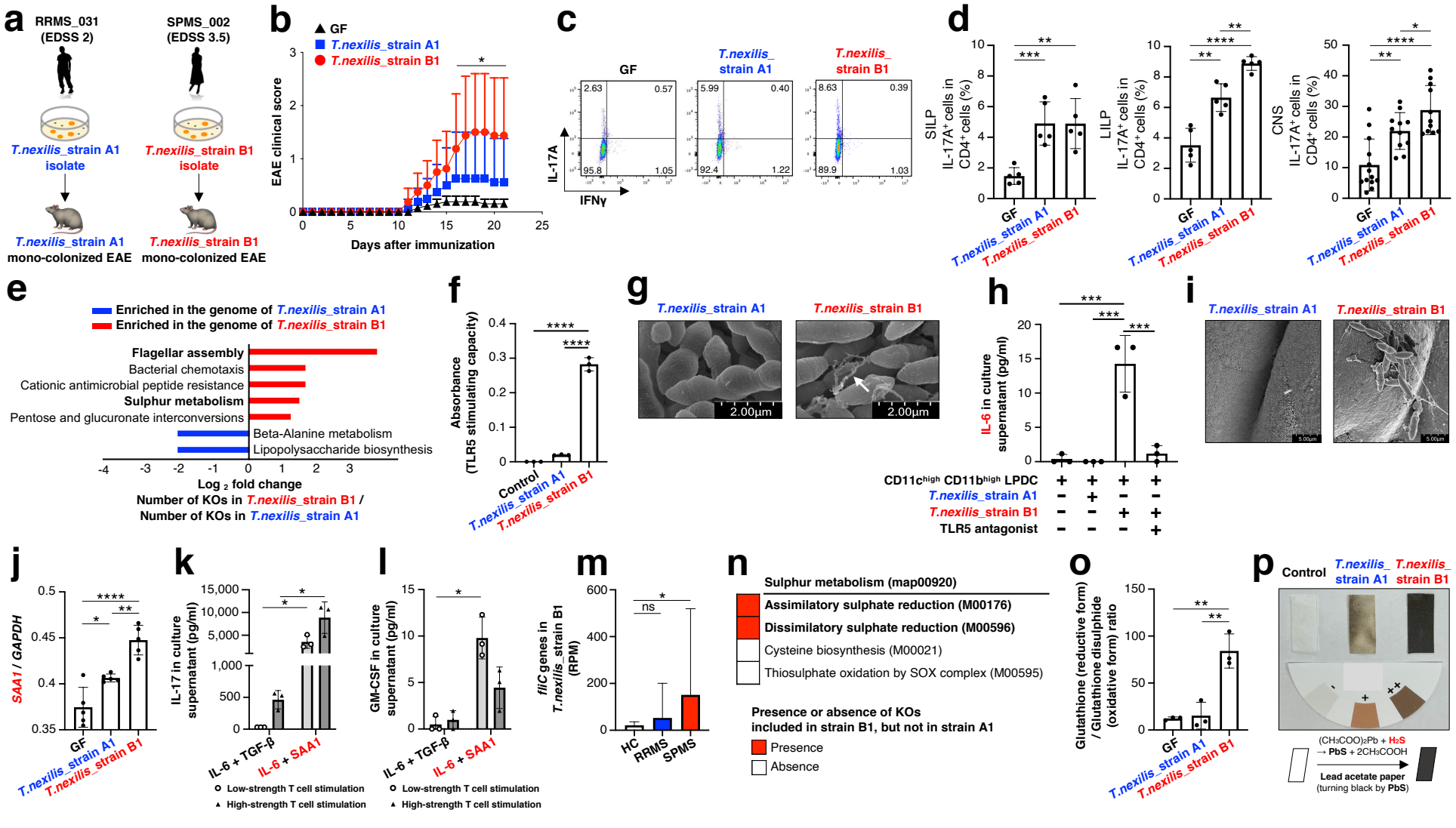


Fig.3. Functional characteristics of *Tyzzerella nexilis* strains

(**a**) Schematic of the experimental flow. (**b**) EAE score of germ-free (GF) mice inoculated with *T.nexilis*_strain A1 or B1. Combined results of two independent experiments are shown (GF, n=13; strain A1 and B1, n=11). (**c**) Representative FACS plots (gated on $CD3^+CD4^+CD8^-$). (**d**) Frequency of Th17 cells in the small intestine (n=5 mice), large intestine (n=5 mice), and central nervous system (GF, n=13; strain A1 and B1, n=11). (**e**) Differences of gene contents between strain A1 and B1 based on KEGG pathways. (**f**) Co-culture of heat-inactivated strains with TLR5 reporter HEK cells. (**g**) SEM images of cultured strains. A flagella-like structure is indicated by an arrow. (**h**) Co-culture of heat-inactivated strains with $CD11c^{high}CD11b^{high}$ intestinal dendritic cells. (**i**) SEM images on the surface of the colon in the mono-colonised mice. (**j**) Gene expression of *SAA1* on the colonic epithelial cells was assessed by qPCR. (**k, l**) The assay for Th17 cell differentiation. The concentrations of IL-17 (**k**) and GM-CSF (**l**) were assessed by ELISA. (**m**) Number of mapped reads per million (RPM) to *fliC* genes of *T.nexilis*_strain B1 among the 29 HCs, 62 RRMS, and 15 SPMS patients. (**n**) The presence or absence of KEGG orthologies (KOs) included in strain B1 but not in strain A1 in each module within sulphur metabolism (map00920). (**o**) The ratio of glutathione to glutathione disulphide in the faeces of GF and strain A1 or B1 mono-colonised mice (n=3 mice). (**p**) The production of hydrogen sulphide (H_2S) in YCFA liquid culture media was assessed by lead acetate paper. Data are presented as the mean ± S.D. ns $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$; Kruskal-Wallis test with Dunn`s multiple comparisons test (**b, m**), two-sided unpaired t-test (**d, k, l**), one-way ANOVA with Tukey's multiple comparisons test (**f, h, j, o**).

Figure 4

Horizontal gene transfer shapes pathogenic bacteria in multiple sclerosis

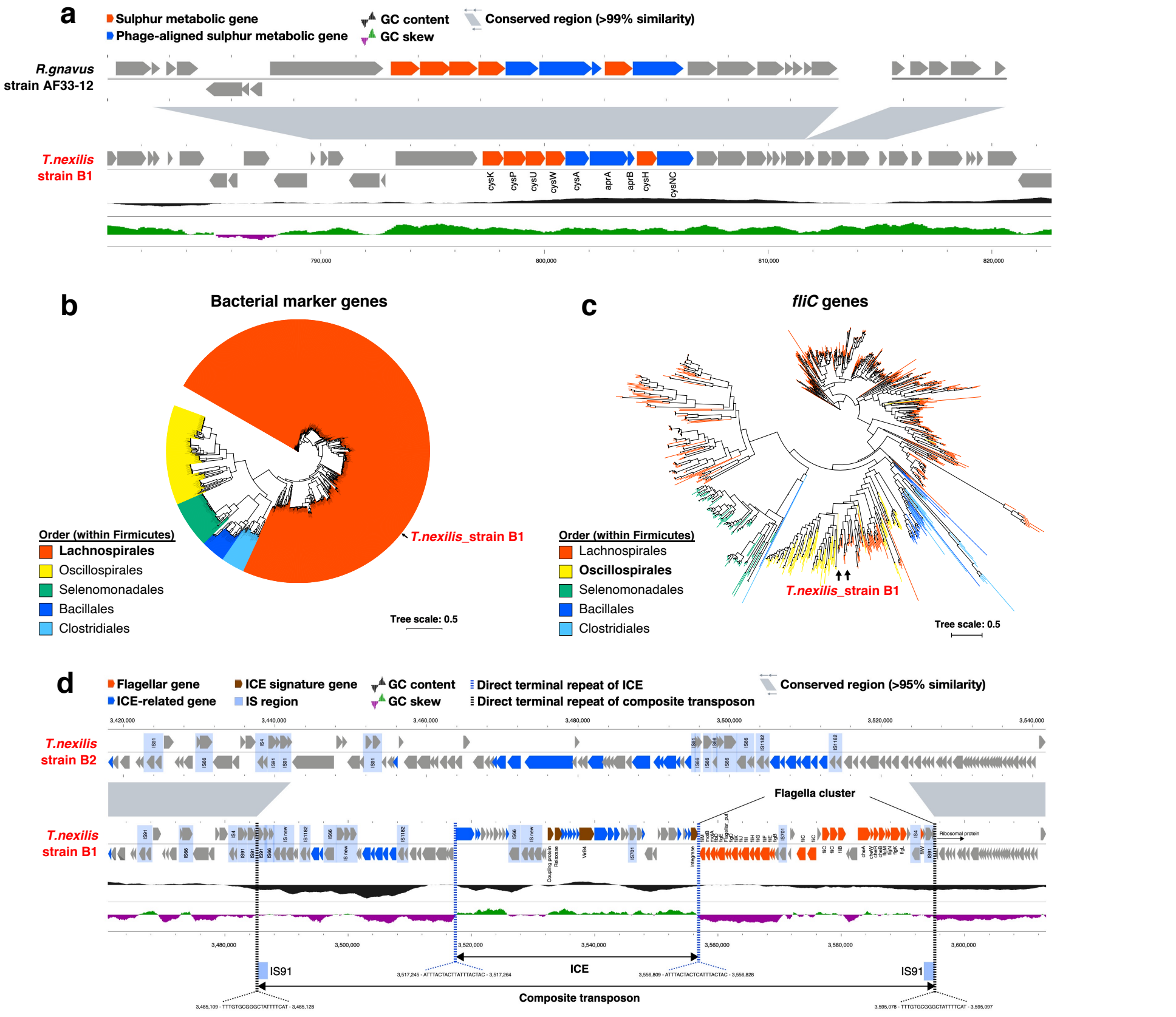
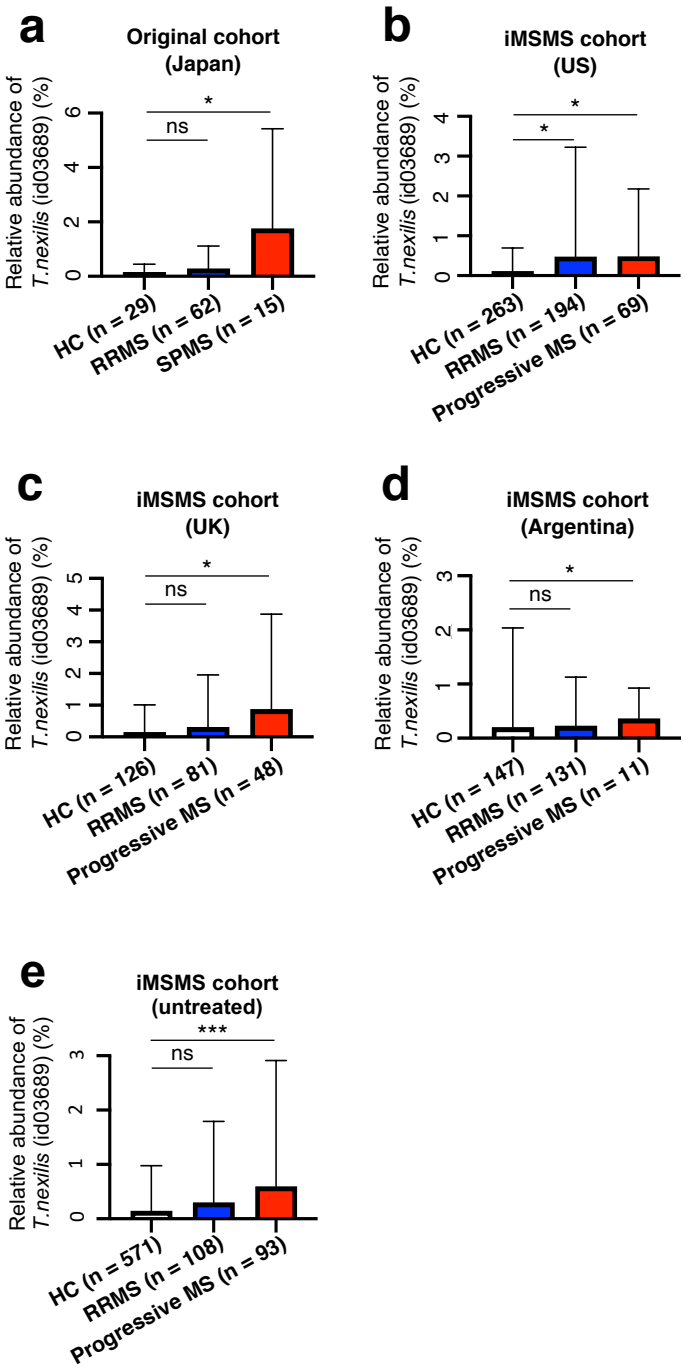


Fig.4. Acquisition mechanism of potentially pathogenic genes

(**a**) Genomic comparison between *Tyzzerella nexilis*_strain B1 and *Ruminococcus gnavus*_strain AF33-12. The ORF annotations of sulphur metabolic genes and phage-aligned sulphur metabolic genes are shown coloured in red and blue, respectively. Grey connections indicate conserved regions between two genomes. (**b**) Phylogenetic tree of the Firmicutes genomes encoding *fliC* genes from the *T.nexilis*_strain B1 and HumGut based on 120 bacterial marker genes. Each colour shows the order of each genome. (**c**) Phylogenetic tree of the *fliC* genes from *T.nexilis*_strain B1 and HumGut database in five orders in Firmicutes. Each branch colour shows the order of the genomes encoding *fliC* genes. (**d**) Genomic comparison between *T.nexilis*_strain B1 (with flagellar genes) and B2 (without flagellar genes). The ORF annotations of flagellar genes, ICE-related genes, and ICE signature genes are shown coloured in red, blue, and brown, respectively. Insertion sequence (IS) regions are shown as blue boxes. The name of each IS indicates a family of IS and IS new means a novel family of IS. Grey connections indicate conserved regions between two genomes. The direct terminal repeat sequences and the coordinate at the edge of ICE and composite transposon including flagellar genes are shown with dashed lines of blue and black, respectively.

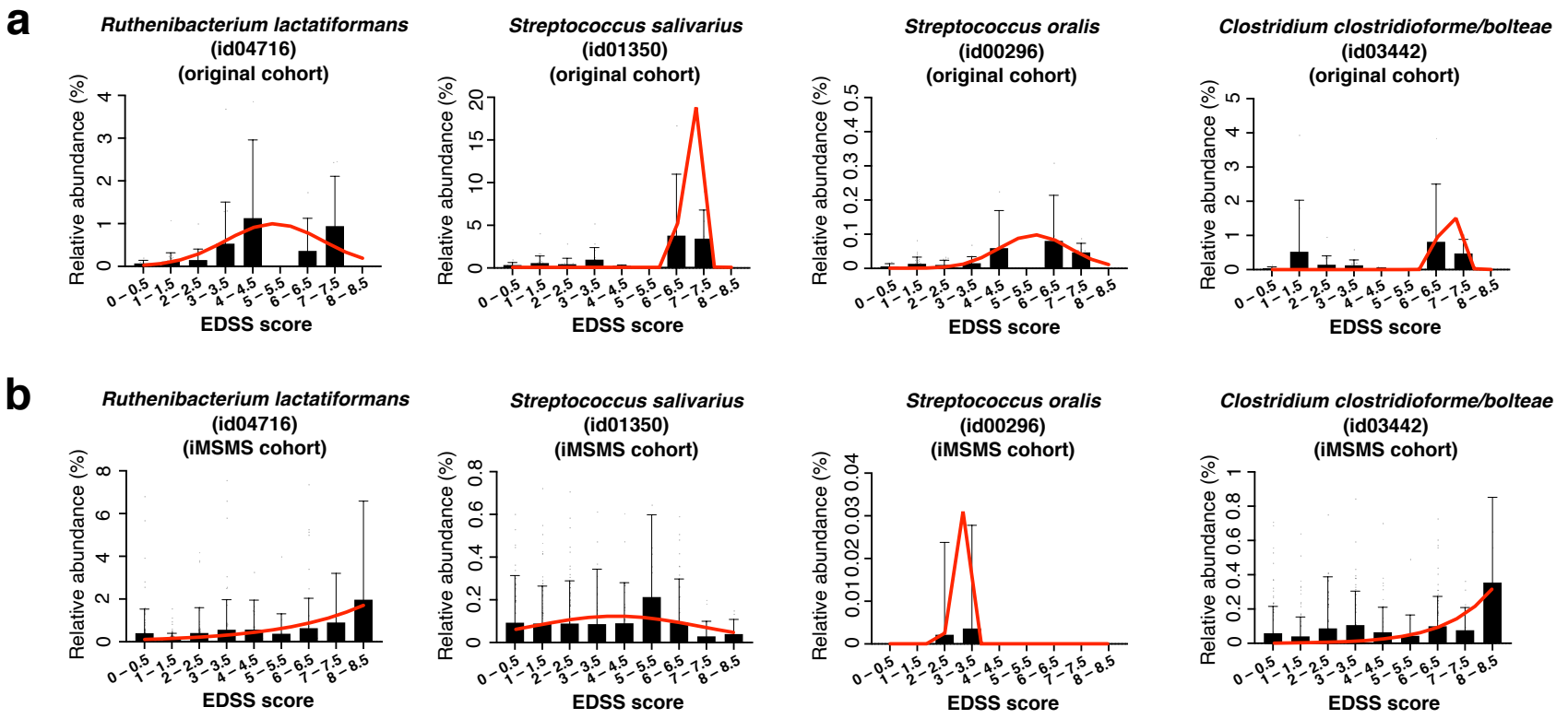
Extended data Figure 1



Extended data Fig.1. Universality and specificity in the association between *Tyzzerella nexilis* and progressive MS

(**a-e**) The relative abundance of *T.nexilis* (id03689) among the HC, RRMS, and SPMS (or progressive MS) groups in our original (**a**: Japan) and iMSMS cohorts (**b**: United States, **c**: United Kingdom, **d**: Argentina, **e**: untreated). (**f**) mOTU analysis using the publicly available and original metagenomic data. The 40 mOTUs whose relative abundance was significantly enriched or depleted in the SPMS group relative to healthy control (HC) group are shown (p <0.05 and lcoefficientl >1). We compared the abundance of these 40 mOTUs between the HC and other disease groups, including RRMS (571 HCs and 435 patients)[3], IBD (26 non-IBD individuals, 29 patients with ulcerative colitis, and 49 patients with Crohn's disease)[18], rheumatoid arthritis (80 HCs and 61 patients)[19], colorectal cancer (53 HCs and 73 patients)[20], and Parkinson's disease (234 HCs and 491 patients)[21]. A coefficient value based on MaAsLin2 in a comparison between the disease and healthy groups is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. Data are presented as mean ± S.D. *p <0.05, **p <0.01, ***p <0.001; MaAsLin2.

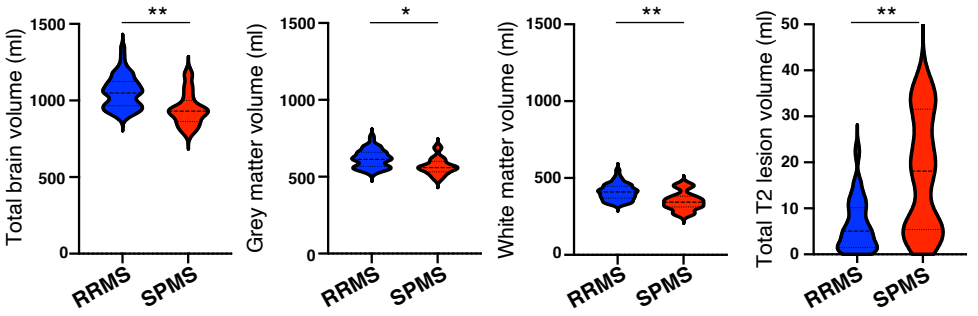
Extended data Figure 2

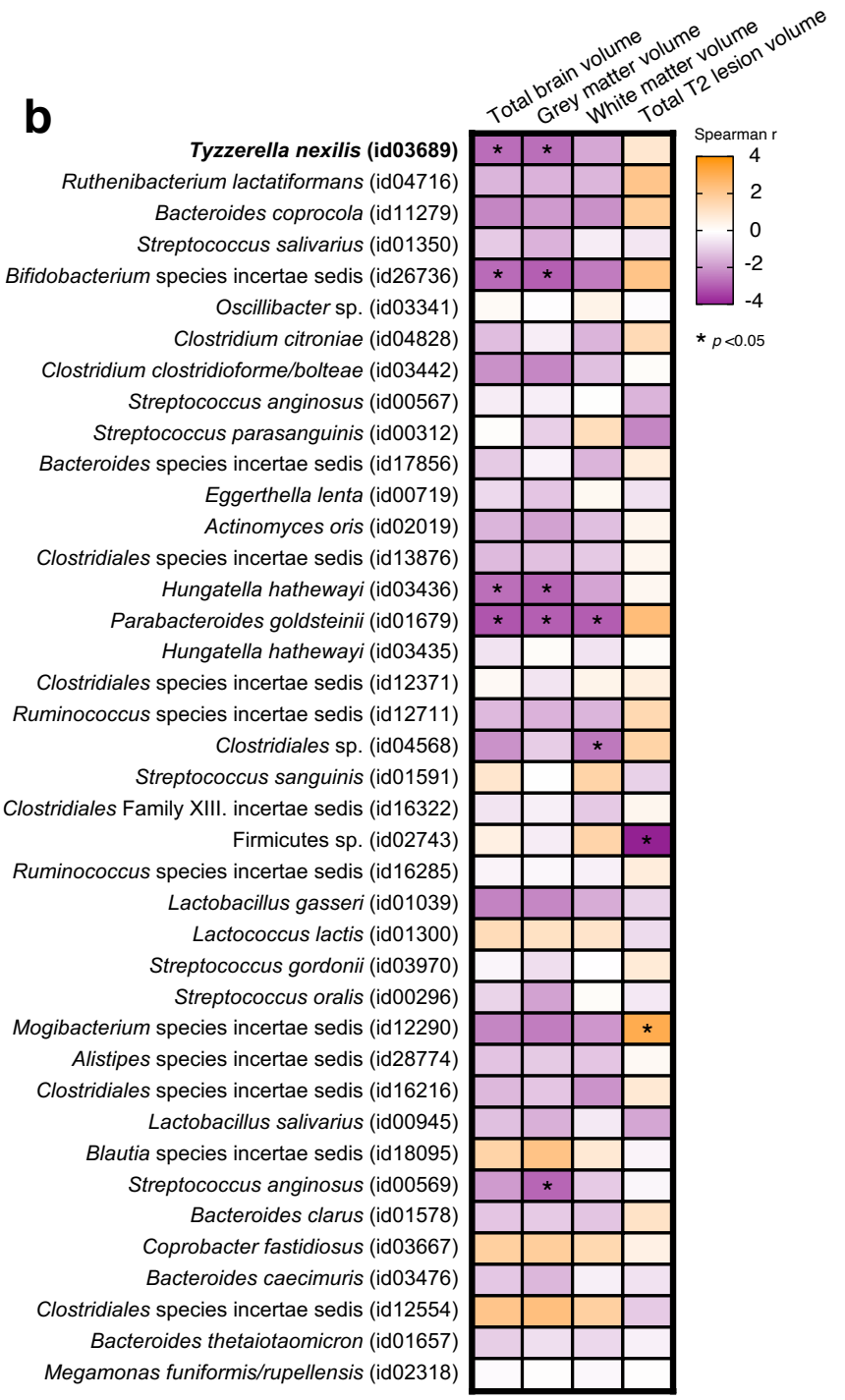


Extended data Fig.2. Associations between microbial abundance and a clinical severity score

(**a-b**) The relative abundance of four mOTUs at each EDSS score separately in the original (n = 77) and iMSMS (n = 568) cohorts of patients with MS. The four mOTUs were selected as the species whose relative abundance was the most positively correlated with an EDSS score other than *Tyzzerella nexilis* (id03689) based on a coefficient value of MaAsLin2 among 40 mOTUs that were highly enriched or depleted in the original SPMS group compared with the HC group. Curve fitting using non-linear regression for the average of the relative abundance in each group was conducted: the non-linear curves are shown in red.

Extended data Figure 3

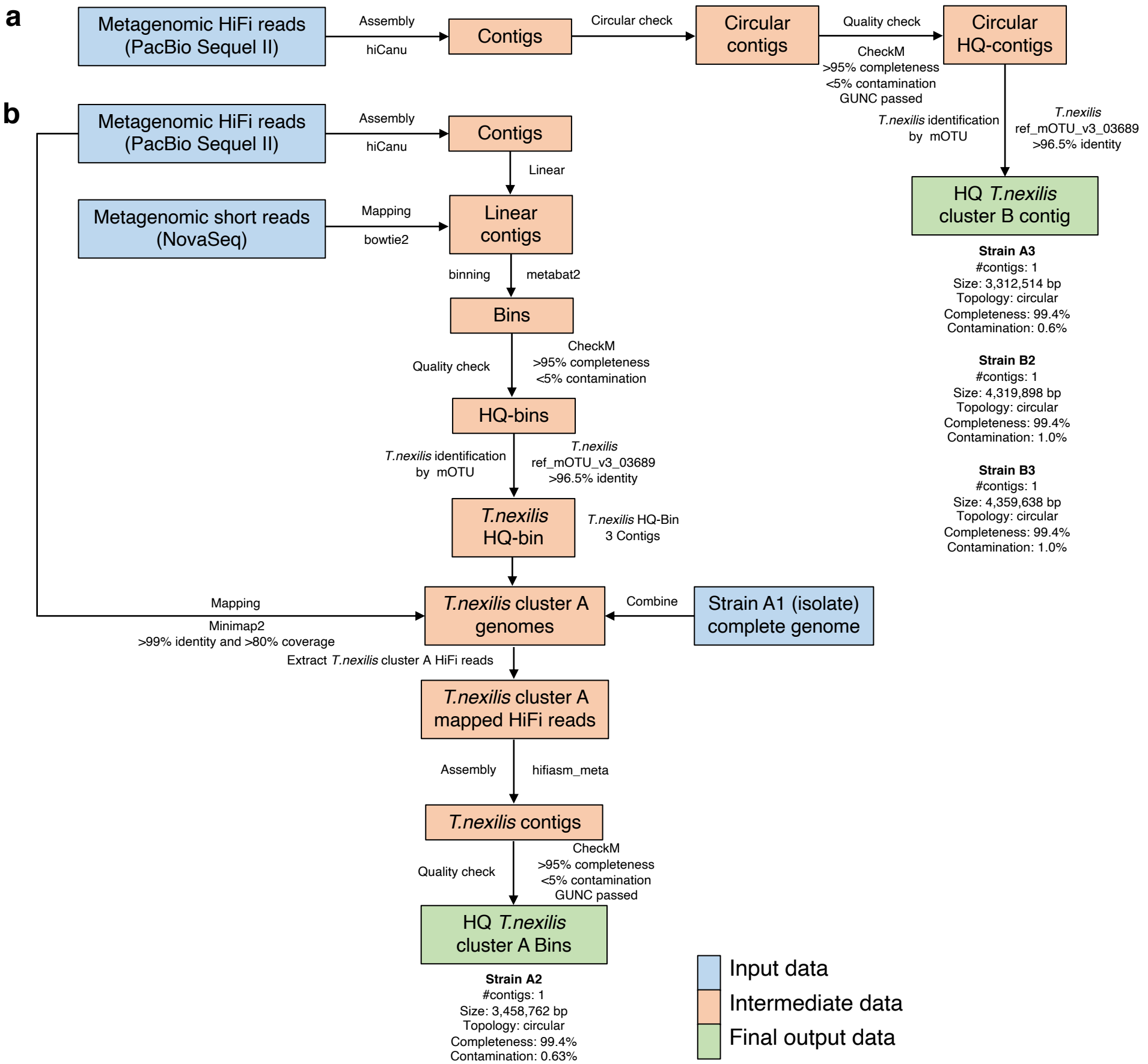




Extended data Fig.3. Correlation between microbial abundance and quantitative brain MRI parameters

(**a**) Among the 77 patients with MS, 60 patients who underwent brain MRI within 1 year before or after the faecal sampling were included in this analysis, and the brain MRI parameters between patients with RRMS (n=48) and SPMS (n=12) were compared. (**b**) The 40 mOTUs whose relative abundance was significantly enriched or depleted in the SPMS group compared with the HC group are shown (p <0.05 and lcoefficientl >1). We analysed the correlation between the abundance of these 40 mOTUs and four brain MRI parameters including total brain volume, grey matter volume, white matter volume, and total T2 lesion volume among 60 patients with MS. The Spearman coefficient R in the correlation between the relative abundance of each mOTU and brain MRI parameters is depicted from the lowest (purple) to the highest (orange) according to the scale shown on the right. *p <0.05, **p <0.01; Wilcoxon rank-sum test (**a**), Spearman's rank correlation coefficient (**b**).

Extended data Figure 4



Extended data Fig.4. Workflow of the long-read metagenomic assembly

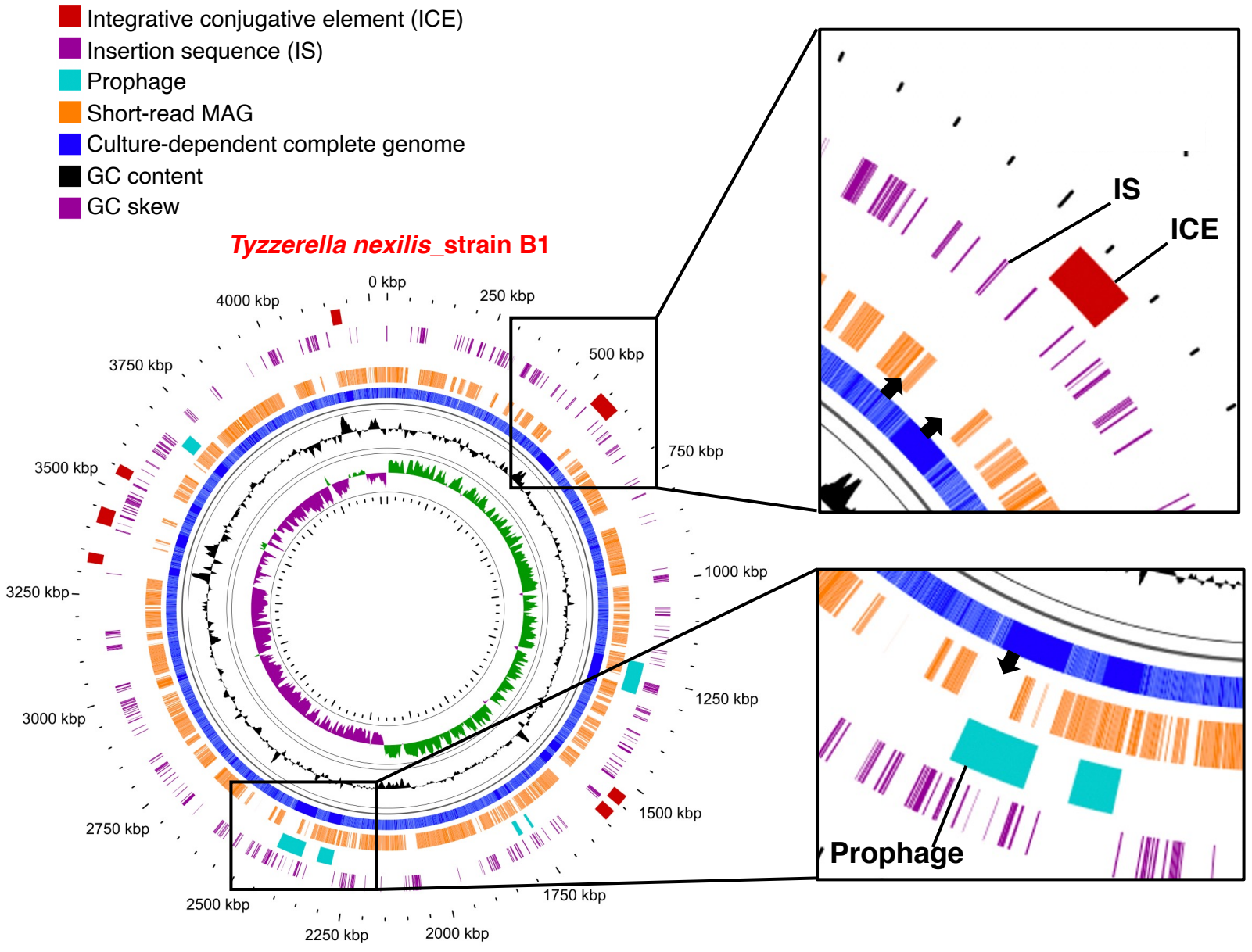
(**a**) Workflow for constructing the *Tyzzerella nexilis* genome from closed circular contigs. The strain A3, B2, and B3 genomes were constructed using this strategy. First, the metagenomic HiFi reads were assembled by hiCanu. Then, the high-quality (HQ) chromosomal contigs were selected by CheckM and GUNC from circular contigs. Finally, HQ genomes having marker genes of *T.nexilis* (ref_mOTU_v3_03689) with >96.5% identity were selected. (**b**) Workflow for combinational analysis of long-read metagenomic assembly, binning, and reference-guided assembly. The *T.nexilis*_strain A2 genome was constructed using this strategy. The metagenomic short reads were mapped to the linear contigs obtained by metagenomic assembly using HiFi reads. Then, the binning analysis and quality check of bins were performed by metabat2 and CheckM, respectively. The marker genes of *T.nexilis* (ref_mOTU_v3_03689) were aligned to HQ-bins and obtained HQ-bin of *T.nexilis* (cluster A) consisting of three contigs. For collecting *T.nexilis*-related HiFi reads, all metagenomic HiFi reads were mapped to the HQ-bin of *T.nexilis* and the complete genome of strain A1. Finally, *T.nexilis*-related HiFi reads were assembled by hifiasm_meta and obtained a closed circular genome of the *T.nexilis*_strain A2. (**a,b**) Details of these workflows are described in the Method section.

Extended data Figure 5

Extended data Fig.5 Comparisons of the genomic structures between MS patients-derived *Tyzzerella nexilis* strains

Dot plot diagram for genomic similarity among the six genomes of *T.nexilis*_strain A1, A2, A3, B1, B2, and B3.

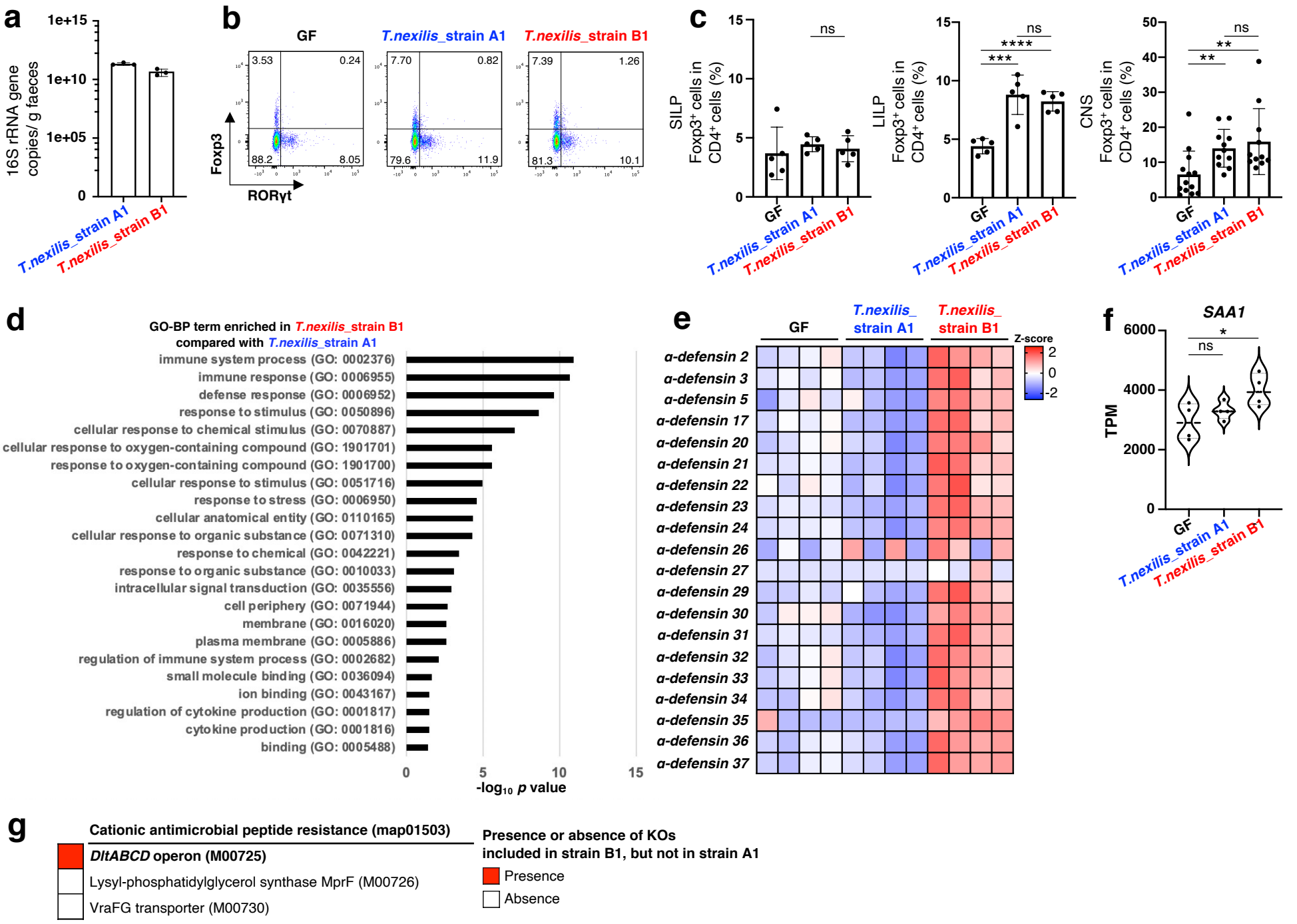
Extended data Figure 6



Extended data Fig.6. The genomic comparison between short-read MAG and complete genome in *Tyzzerella nexilis*_strain B1

Comparison of genomic structures between the short-read MAG and the isolation-based complete genome of mobile genetic elements (MGEs)-enriched *T.nexilis*_strain B1. The predicted regions of integrative and conjugative elements (ICE), insertion sequences (IS), and prophages are shown. Black arrows indicate the specific regions predicted as MGEs among the missing regions in short-read MAGs.

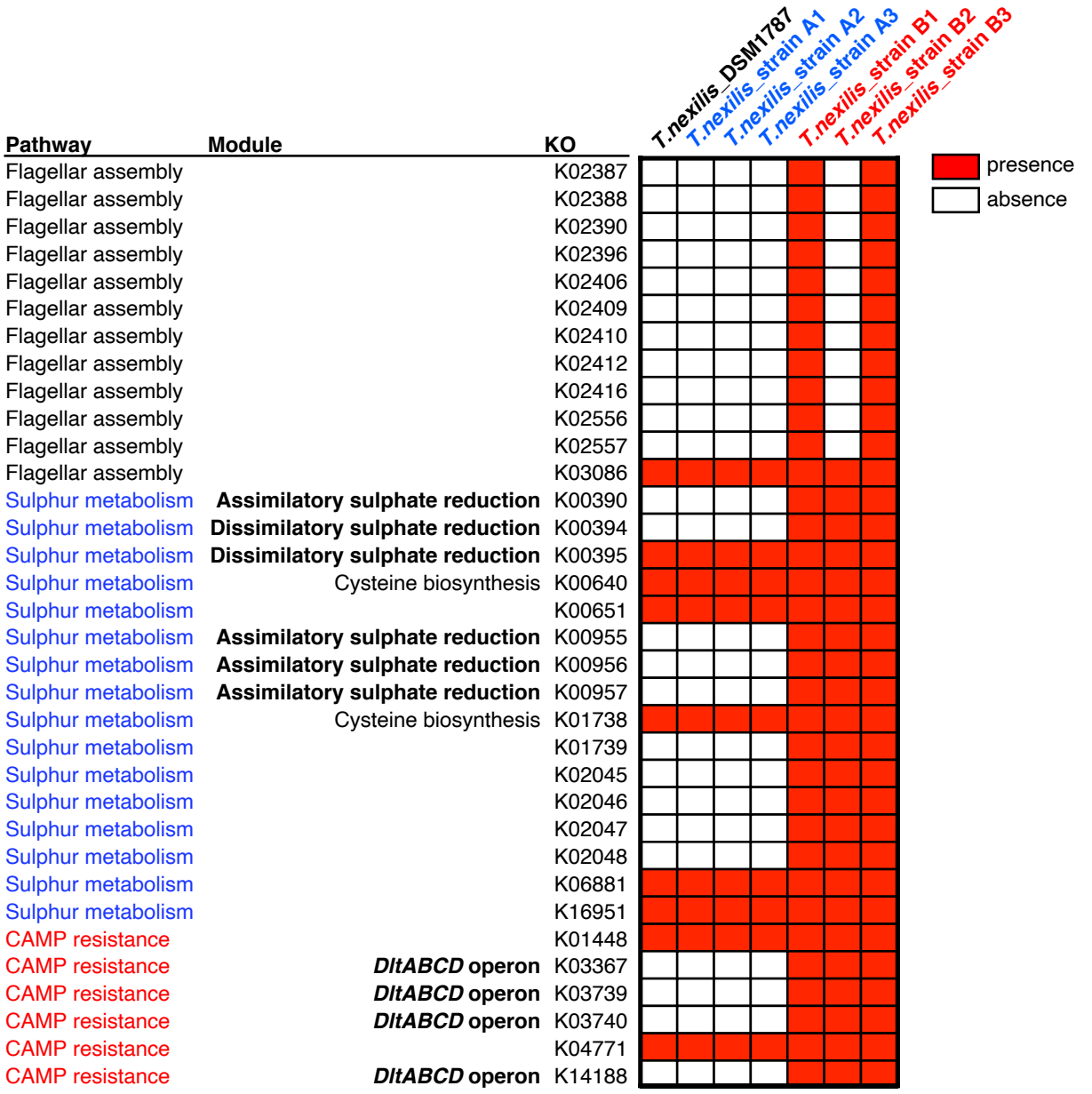
Extended data Figure 7



Extended data Fig 7. Functional analysis in the comparison between *Tyzzerella nexilis* strains

(**a**) Quantification of 16S rRNA gene copies in the faeces of GF and *T.nexilis*_strain A1 or B1 mono-colonised mice (n=3 mice). (**b**) Representative FACS plots (gated on $CD3^+CD4^+CD8^-$). (**c**) Frequency of regulatory T cells in the small intestine (n=5 mice), large intestine (n=5 mice), and central nervous system (GF, n=13; *T.nexilis*_strain A1 and B1, n=11). (**d-f**) Gene expression profile of colonic ECs between *T.nexilis*_strain B1 mono-colonised mice (n=4), *T.nexilis*_strain A1 mono-colonised mice (n=4), and germ-free (GF) mice (n=4). Differential gene expression in the colonic ECs was analysed by RNA-seq. (**d**) Gene ontology (GO) terms significantly enriched in up-regulated gene sets in the colonic ECs derived from the strain B1 mono-colonised mice compared with those from the strain A1 mono-colonised mice are shown. (**e**) Comparison of the genes annotated to α-defensin between the GF, strain A1 mono-colonised, and strain B1 mono-colonised mice. The z-score based on the transcripts per kilobase million (TPM) is depicted from the lowest (blue) to the highest (red) according to the scale shown on the right. (**f**) TPM of *SAA1* between the GF, strain A1 mono-colonised, and strain B1 mono-colonised mice. (**g**) The presence or absence of KEGG orthologies (KOs) included in strain B1, but not in strain A1 in each module within the cationic antimicrobial peptide resistance (map01503). Data are presented as mean ± S.D. ns $p > 0.05$, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$, two-sided unpaired t-test (**c**), one-way ANOVA with Tukey's multiple comparisons test (**f**).

Extended data Figure 8

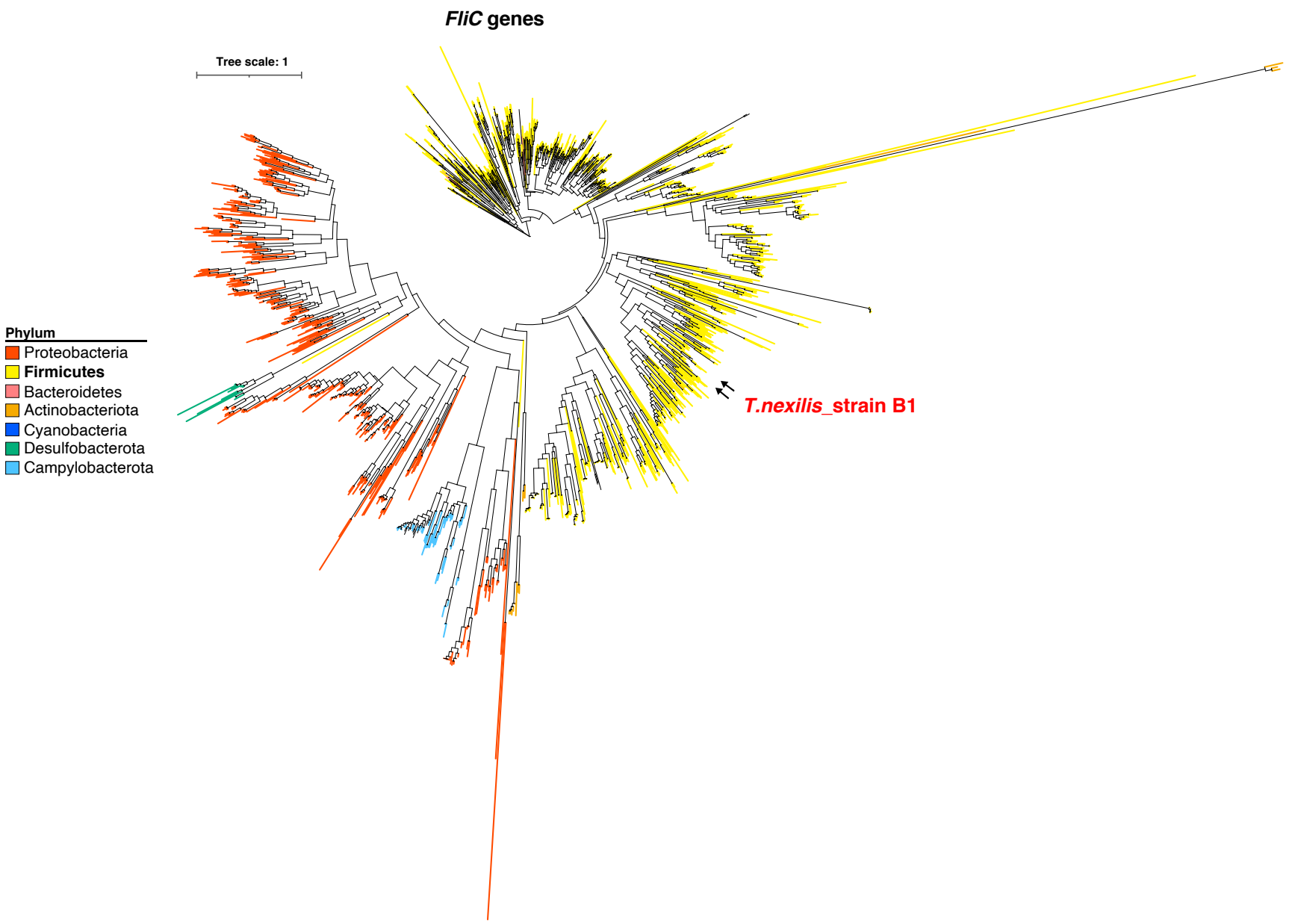


Pathway	Module	KO
Flagellar assembly		K02387
Flagellar assembly		K02388
Flagellar assembly		K02390
Flagellar assembly		K02396
Flagellar assembly		K02406
Flagellar assembly		K02409
Flagellar assembly		K02410
Flagellar assembly		K02412
Flagellar assembly		K02416
Flagellar assembly		K02556
Flagellar assembly		K02557
Flagellar assembly		K03086
Sulphur metabolism	**Assimilatory sulphate reduction**	K00390
Sulphur metabolism	**Dissimilatory sulphate reduction**	K00394
Sulphur metabolism	**Dissimilatory sulphate reduction**	K00395
Sulphur metabolism	Cysteine biosynthesis	K00640
Sulphur metabolism		K00651
Sulphur metabolism	**Assimilatory sulphate reduction**	K00955
Sulphur metabolism	**Assimilatory sulphate reduction**	K00956
Sulphur metabolism	**Assimilatory sulphate reduction**	K00957
Sulphur metabolism	Cysteine biosynthesis	K01738
Sulphur metabolism		K01739
Sulphur metabolism		K02045
Sulphur metabolism		K02046
Sulphur metabolism		K02047
Sulphur metabolism		K02048
Sulphur metabolism		K06881
Sulphur metabolism		K16951
CAMP resistance		K01448
CAMP resistance	***DltABCD* operon**	K03367
CAMP resistance	***DltABCD* operon**	K03739
CAMP resistance	***DltABCD* operon**	K03740
CAMP resistance		K04771
CAMP resistance	***DltABCD* operon**	K14188

Extended data Fig.8. Genome comparisons among seven *Tyzzerella nexilis* strains

The presence or absence of KEGG orthologies (KOs) included in flagellar assembly, sulphur metabolism, and CAMP resistance pathways within the genomes of *T.nexilis*_DSM1787, strain A1, A2, A3, B1, B2, and B3.

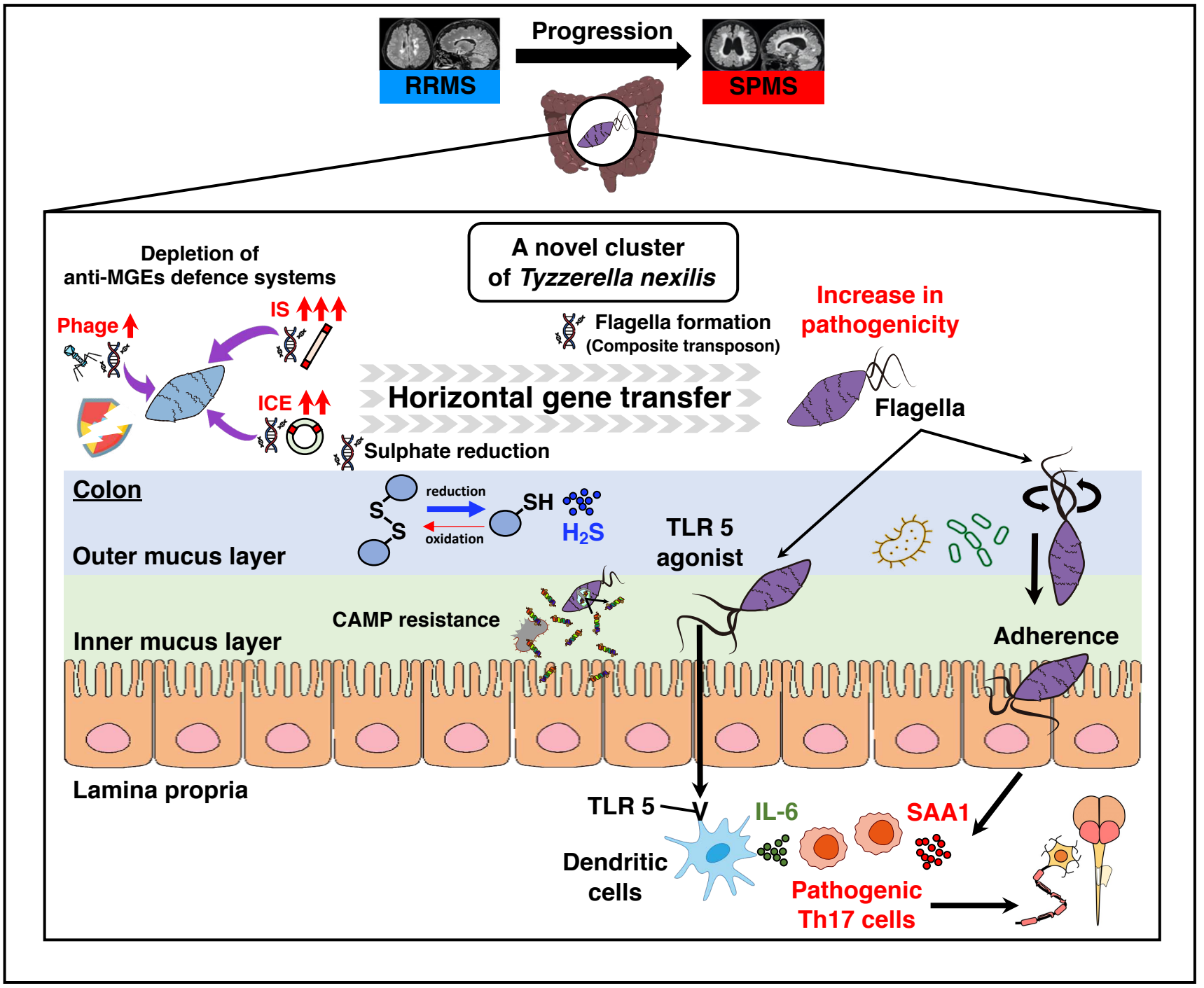
Extended data Figure 9



Extended data Fig.9. Extended data Fig.9. Phylogeny of the *fliC* genes

Phylogenetic tree of the *fliC* genes from *T.nexilis*_strain B1 and the HumGut database. Each branch colour shows the phylum of the genomes encoding *fliC* genes.

Extended data Figure 10



Extended data Fig.10. The nature and functions of bacteria associated with multiple sclerosis progression

Mobile genetic elements (MGEs)-enriched *T.nexilis* strains are abundant in the gut of patients with SPMS. Some of the strains acquired pathogenic genes associated with flagella formation and sulphate reduction via horizontal gene transfer from another microorganism. These strains-derived flagella potentially induce pathogenic Th17 cells via the combination of TLR5 stimulation and promotion of adherence to colonic ECs. These T cells might migrate into the CNS and accelerate neuroinflammation. The sulphate reduction accompanied by hydrogen sulphide (H_2S) production is potentially associated with neurodegeneration. The depletion of defence systems and subsequent enrichment of MGEs might have shaped the pathogenic bacteria underlying MS progression. Abbreviations: IS = insertion sequence; ICE = integrative conjugative element; TLR5 = toll-like receptor 5; SAA1 = serum amyloid A1; CAMP; cationic antimicrobial peptide resistance.

Extended data Table 1

Sample ID	NovaSeq reads		Sequel II HiFi reads		
	Number of filter-passed* NovaSeq reads	Total bases of filter-passed* NovaSeq reads	Number of HiFi reads	Total bases of HiFi reads	Average read length
RRMS_001	71,815,702	10,759,546,579	-	-	-
RRMS_002	77,212,788	11,276,617,433	-	-	-
RRMS_003	42,411,368	6,354,219,649	-	-	-
RRMS_004	90,684,058	13,291,269,220	2,252,056	21,906,770,313	9,728
RRMS_005	69,275,172	10,086,797,435	-	-	-
RRMS_006	93,470,332	13,606,048,815	-	-	-
RRMS_007	55,415,138	8,294,899,346	-	-	-
RRMS_008	92,992,880	13,569,374,061	-	-	-
RRMS_009	81,033,534	11,840,134,612	-	-	-
RRMS_010	102,303,182	15,017,651,121	-	-	-
RRMS_011	57,083,870	8,538,636,575	-	-	-
RRMS_012	93,579,296	13,705,678,362	-	-	-
RRMS_013	49,584,782	7,418,247,083	-	-	-
RRMS_014	103,614,988	15,235,318,747	-	-	-
RRMS_015	61,769,138	9,238,860,025	-	-	-
RRMS_016	90,892,678	13,251,336,823	-	-	-
RRMS_017	90,169,114	13,272,172,122	-	-	-
RRMS_018	46,891,090	6,910,676,861	-	-	-
RRMS_019	77,679,732	11,372,599,971	-	-	-
RRMS_020	67,240,518	10,067,186,534	-	-	-
RRMS_021	98,437,522	14,463,046,456	-	-	-
RRMS_022	75,115,786	10,935,224,611	-	-	-
RRMS_023	94,634,896	13,741,751,263	-	-	-
RRMS_024	132,708,216	19,164,322,098	-	-	-
RRMS_025	109,255,820	15,917,834,231	-	-	-
RRMS_026	118,125,228	17,300,266,035	-	-	-
RRMS_027	123,133,244	17,940,399,025	-	-	-
RRMS_028	109,324,824	16,038,893,762	-	-	-
RRMS_029	91,150,392	13,270,498,314	-	-	-
RRMS_030	112,992,168	16,446,970,975	-	-	-
RRMS_031	87,653,118	12,693,791,476	-	-	-
RRMS_032	122,371,928	17,737,330,230	-	-	-
RRMS_033	94,018,444	13,386,909,852	-	-	-
RRMS_034	117,753,716	17,276,602,266	-	-	-
RRMS_035	102,228,276	14,942,873,546	-	-	-
RRMS_036	105,936,266	15,369,346,740	-	-	-
RRMS_037	98,092,500	14,311,892,648	-	-	-
RRMS_038	101,113,704	14,706,895,320	-	-	-
RRMS_039	76,830,190	11,111,114,309	-	-	-
RRMS_040	131,033,768	18,959,733,016	-	-	-
RRMS_041	110,030,896	16,143,545,055	1,937,559	12,239,256,839	6,317
RRMS_042	105,098,422	15,450,738,757	-	-	-
RRMS_043	58,184,854	8,711,748,783	-	-	-
RRMS_044	60,586,618	9,074,941,882	-	-	-
RRMS_045	103,186,632	14,960,271,736	-	-	-
RRMS_046	69,209,384	10,048,106,871	-	-	-
RRMS_047	78,618,816	11,748,246,396	-	-	-
RRMS_048	52,444,362	7,851,478,352	-	-	-
RRMS_049	54,553,944	8,166,317,718	-	-	-
RRMS_050	78,726,224	11,525,290,192	-	-	-
RRMS_051	71,372,748	10,666,703,375	-	-	-
RRMS_052	59,867,710	8,972,625,582	-	-	-
RRMS_053	64,063,440	9,602,248,220	-	-	-
RRMS_054	76,736,138	11,501,419,834	-	-	-
RRMS_055	64,774,948	9,707,004,090	-	-	-
RRMS_056	66,236,162	9,926,776,944	-	-	-
RRMS_057	60,724,948	9,085,661,409	-	-	-
RRMS_058	63,462,548	9,508,821,816	-	-	-
RRMS_059	52,590,858	7,876,885,968	-	-	-
RRMS_060	96,292,072	14,422,753,663	-	-	-
RRMS_061	64,336,754	9,638,018,718	-	-	-
RRMS_062	48,487,890	7,170,509,056	-	-	-
SPMS_001	98,696,618	14,412,439,162	-	-	-
SPMS_002	510,636,458	72,943,520,736	-	-	-
SPMS_003	80,204,252	12,016,311,616	-	-	-
SPMS_004	160,763,816	23,534,667,107	-	-	-
SPMS_005	139,732,710	20,517,285,185	-	-	-
SPMS_006	130,558,414	19,124,788,380	-	-	-
SPMS_007	149,835,056	21,957,145,163	-	-	-
SPMS_008	95,678,020	13,981,977,760	-	-	-
SPMS_009	131,344,640	19,120,523,073	-	-	-
SPMS_010	107,307,636	15,777,110,378	-	-	-
SPMS_011	122,868,862	17,988,019,539	-	-	-
SPMS_012	56,839,782	8,518,499,041	-	-	-
SPMS_013	92,458,144	13,813,620,610	-	-	-
SPMS_014	84,293,450	12,629,972,863	-	-	-
SPMS_015	64,916,930	9,720,715,038	2,682,547	23,612,546,464	8,802
HC_001	67,181,110	9,568,348,321	-	-	-
HC_002	77,581,460	11,048,930,174	-	-	-
HC_003	94,454,542	13,453,045,642	-	-	-
HC_004	77,225,970	10,999,234,072	-	-	-
HC_005	35,756,698	5,056,701,513	-	-	-
HC_006	108,966,222	15,519,189,758	-	-	-
HC_007	80,646,890	11,485,685,988	-	-	-
HC_008	85,730,540	12,208,597,656	-	-	-
HC_009	67,179,748	9,566,623,711	-	-	-
HC_010	165,948,304	23,631,606,242	-	-	-
HC_011	52,041,532	7,411,948,141	-	-	-
HC_012	276,719,104	39,407,836,820	-	-	-
HC_013	57,390,380	8,173,117,823	-	-	-
HC_014	186,439,548	26,554,264,176	-	-	-
HC_015	181,631,782	25,868,425,465	-	-	-
HC_016	140,519,818	20,011,443,683	-	-	-
HC_017	61,192,772	8,714,773,151	-	-	-
HC_018	62,650,628	8,923,064,017	-	-	-
HC_019	72,644,758	10,346,655,497	-	-	-
HC_020	285,893,080	40,717,037,710	-	-	-
HC_021	50,380,976	7,175,419,517	-	-	-
HC_022	145,192,270	20,679,931,903	-	-	-
HC_023	175,358,410	24,975,573,815	-	-	-
HC_024	171,173,838	24,377,963,136	-	-	-
HC_025	162,077,852	23,082,260,866	-	-	-
HC_026	190,499,698	27,131,375,725	-	-	-
HC_027	199,526,250	28,416,204,905	-	-	-
HC_028	210,907,982	30,032,840,859	-	-	-
HC_029	213,885,590	30,459,990,817	-	-	-
Average	**103,391,965**	**15,015,450,648**	**2,290,721**	**19,252,857,872**	**8,282**
Max	**510,636,458**	**72,943,520,736**	**2,682,547**	**23,612,546,464**	**9,728**
Min	**35,756,698**	**5,056,701,513**	**1,937,559**	**12,239,256,839**	**6,317**
Total	**10,959,548,274**	**1,591,637,768,718**	**6,872,162**	**57,758,573,616**	**24,847**

*** Filter-passed; quality filtered and unmapped reads with human genome**

Extended data Table 1. Sequence statistics of the NovaSeq and Sequel sequencers

The number of filter-passed* NovaSeq reads and total bases of filter-passed* NovaSeq reads are presented. The number of Sequel HiFi reads, total bases of Sequel HiFi reads, and average reads length are also presented.

Extended data Table 2

Taxonomy ID	coefficient value	p-value
***Tyzzerella nexilis* (id03689)**	**2.559**	**0.007**
Bifidobacterium species incertae sedis (id26736)	2.193	0.001
Alistipes putredinis (id03683)	2.015	0.036
Bacteroides coprocola (id11279)	1.917	0.017
Streptococcus salivarius (id01350)	1.742	0.004
Coprococcus sp. (id01683)	1.656	0.014
Bilophila wadsworthia (id04300)	1.549	0.042
Clostridium glycyrrhizinilyticum (id03672)	1.427	0.006
Parabacteroides distasonis (id03640)	1.375	0.041
Bacteroides species incertae sedis (id17856)	1.299	0.007
Clostridium species incertae sedis (id17741)	1.283	0.017
uncultured *Eubacterium* sp. (id13063)	1.264	0.007
Parabacteroides species incertae sedis (id26659)	1.246	0.042
Bacteroides species incertae sedis (id17683)	1.200	0.042
Hungatella hathewayi (id03435)	1.161	0.014
Clostridium citroniae (id04828)	1.127	0.002
Streptococcus anginosus (id00567)	1.110	0.003
Alistipes species incertae sedis (id28774)	1.094	0.007
Streptococcus species incertae sedis (id28879)	1.007	0.007
Gemella sanguinis (id04303)	0.938	0.035
Streptococcus gordonii (id03970)	0.887	0.032
Actinomyces sp. ICM47 (id01914)	0.862	0.040
Clostridium sp. CAG:273 (id12673)	0.829	0.037
Streptococcus species incertae sedis (id26680)	0.787	0.026
Streptococcus species incertae sedis (id19491)	0.764	0.037
Streptococcus oralis (id00290)	0.704	0.007
Lachnospiraceae species incertae sedis (id18295)	0.684	0.026
Solobacterium species incertae sedis (id12387)	0.632	0.004
uncultured *Clostridium* sp. (id11611)	0.589	0.034
Erysipelotrichaceae species incertae sedis (id16196)	0.532	0.043
Actinomyces marseillensis/pacaensis (id03846)	0.487	0.042
Streptococcus oralis (id00292)	0.443	0.042
Streptococcus pneumoniae (id00282)	0.441	0.032
Lachnoclostridium species incertae sedis (id16281)	-0.528	0.026
uncultured *Eubacterium* sp. (id08868)	-0.536	0.047
Bacteroides intestinalis (id02809)	-0.888	0.038
Bacteroides thetaiotaomicron (id01657)	-1.441	0.048
Faecalibacterium prausnitzii (id06110)	-2.613	0.028

Extended data Table 2. Significantly enriched or depleted mOTUs in SPMS compared with RRMS

The 38 mOTUs whose relative abundance was significantly enriched or depleted in the SPMS group compared with the RRMS group are presented in the descending order of coefficient value of MaAsLin2. mOTUs whose coefficient value >0 are SPMS-enriched mOTUs and those whose coefficient value <0 are RRMS-enriched mOTUs.

Extended data Table 3

Annotation	Nucleotide sequence similarity between strain A and B MAGs (%)
Ribosome-binding ATPase YchF	96.3
50S ribosomal protein L11	96.9
DNA-directed RNA polymerase subunit beta	93.5
50S ribosomal protein L3	83.5
30S ribosomal protein S8	95.3
50S ribosomal protein L15	96.4
Protein translocase subunit SecY	91.3
DNA-directed RNA polymerase subunit alpha	95.4
Cysteine--tRNA ligase	96.0
Leucine--tRNA ligase	96.5
Signal recognition particle receptor FtsY	96.2

Extended data Table 3. Strain-specific marker genes of *Tyzzerella nexilis*_strain A or B MAG

The list of 11 bacterial marker genes for calculating *T.nexilis*_strain A and B MAGs are shown. All genes have <97% nucleotide similarity between strain A and B MAGs.

Extended data Table 4

Assembly_accession	Taxonomy ID	Kingdom	Phylum	Class	Order	Family	Genus	Species	Number of IS	Number of ICE	Number of prophage
N/A	N/A	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Tyzzerella	Tyzzerella nexilis DSM1787	67	6	0
N/A	N/A	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Tyzzerella	Tyzzerella nexilis_strain A1	63	6	3
N/A	N/A	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Tyzzerella	Tyzzerella nexilis_strain A2	64	2	2
N/A	N/A	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Tyzzerella	Tyzzerella nexilis_strain A3	40	3	3
N/A	N/A	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Tyzzerella	Tyzzerella nexilis_strain B1	347	7	6
N/A	N/A	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Tyzzerella	Tyzzerella nexilis_strain B2	327	8	3
N/A	N/A	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Tyzzerella	Tyzzerella nexilis_strain B3	342	8	4
GCA_000007085.1	273068	Bacteria	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacteraceae	*Caldanaerobacter*	*Caldanaerobacter subterraneus*	73	0	0
GCA_000007625.1	212717	Bacteria	Firmicutes	Clostridia	Eubacteriales	Clostridiaceae	*Clostridium*	*Clostridium tetani*	23	0	3
GCA_000008445.1	261594	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	*Bacillus*	*Bacillus cereus* group	16	0	4
GCA_000009785.1	235909	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	*Geobacillus*	*Geobacillus thermoleovorans* group	97	0	2
GCA_000009905.1	292459	Bacteria	Firmicutes	Clostridia	Eubacteriales	Symbiobacteriaceae	*Symbiobacterium*	*Symbiobacterium thermophilum*	47	0	0
GCA_000010265.1	583346	Bacteria	Firmicutes	Clostridia	Eubacteriales	Clostridiaceae	*Clostridium*	*Clostridium kluyveri*	50	2	3
GCA_000011245.1	221109	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	*Oceanobacillus*	*Oceanobacillus iheyensis*	17	0	1
GCA_000012865.1	246194	Bacteria	Firmicutes	Clostridia	Thermoanaerobacterales	Thermoanaerobacteraceae	*Carboxydothermus*	*Carboxydothermus hydrogenoformans*	17	0	1
GCA_000014505.1	278197	Bacteria	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	*Pediococcus*	*Pediococcus pentosaceus*	8	0	2
⋮											
GCA_900638585.1	54006	Bacteria	Firmicutes	Tissierellia	Tissierellales	Peptoniphilaceae	*Peptoniphilus*	*Peptoniphilus ivorii*	8	0	0
GCA_901482605.1	1498	Bacteria	Firmicutes	Clostridia	Eubacteriales	Clostridiaceae	*Hathewaya*	*Hathewaya histolytica*	23	0	4
GCA_901544385.1	1302	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	*Streptococcus gordonii*	7	0	0
GCA_901553735.1	1340	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	*Streptococcus porcinus*	4	2	2
GCA_902387955.1	301301	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	*Roseburia*	*Roseburia hominis*	23	4	0
GCA_902729405.1	1561005	Bacteria	Firmicutes	Clostridia	Eubacteriales	Peptococcaceae	*Acididesulfobacillus*	*Acididesulfobacillus acetoxydans*	103	0	0
GCA_903886475.1	1308	Bacteria	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	*Streptococcus*	*Streptococcus thermophilus*	62	0	0
GCA_940670685.1	2934315	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	*Anaeropeptidivorans*	*Anaeropeptidivorans aminofermentans*	33	1	2
GCA_940677205.1	36835	Bacteria	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	*Tyzzerella*	*[Clostridium] colinum*	11	0	5

※ Row 18 to 724 cannot be displayed due to lack of space. The complete version will be provided by Excel format.

Extended data Table 4. The number of mobile genetic elements in seven *Tyzzerella nexilis* strains and 726 representative Firmicutes strains

Seven *T.nexilis* genomes used in the present study and 726 representative complete genomes assigned to Firmicutes in the GenBank database were analysed. The number of MGEs including insertion sequences (IS), integrative and conjugative elements (ICE), and prophages are presented.

Extended data Table 5

16S rRNA gene similarity (%)	*T.nexilis*_DSM1787	*T.nexilis*_strain A1	*T.nexilis*_strain A2	*T.nexilis*_strain A3	*T.nexilis*_strain B1	*T.nexilis*_strain B2	*T.nexilis*_strain B3
***T.nexilis*_DSM1787**	-	99.934	99.868	98.947	99.474	99.539	99.539
***T.nexilis*_strain A1**	99.934	-	99.934	98.881	99.474	99.539	99.539
***T.nexilis*_strain A2**	99.868	99.934	-	98.815	99.408	99.474	99.474
***T.nexilis*_strain A3**	98.947	98.881	98.815	-	98.421	98.487	98.487
***T.nexilis*_strain B1**	99.474	99.474	99.408	98.421	-	99.934	99.737
***T.nexilis*_strain B2**	99.539	99.539	99.474	98.487	99.934	-	99.803
***T.nexilis*_strain B3**	99.539	99.539	99.474	98.487	99.737	99.803	-

Average similarity (%)	
cluster A - cluster A	99.40
cluster B - cluster B	99.83
cluster A - cluster B	99.24

Extended data Table 5. 16S rRNA gene similarity between the genomes of seven *Tyzzerella nexilis* strains

Similarity of nucleotide sequences in full-length 16S rRNA genes between the seven *T.nexilis* genomes.

Extended data Table 6

Category	start	end	strand	Pfam	KEGG Orthology	ICEBerg annotation	Amino acid similarity with ICEBerg gene (%)	ICE signature genes predicted by ICEscreen
CDS	101	1,486	+	Bac_DnaA,Bac_DnaA_C,DnaA_N	K02313	-	-	-
CDS	1,758	2,867	+	DNA_pol3_beta,DNA_pol3_beta_2,DNA_pol3_beta_3	K02338	-	-	-
CDS	2,871	3,080	+	S4_2	K14761	-	-	-
CDS	3,103	4,200	+	SMC_N	K03629	-	-	-
CDS	4,211	6,127	+	DNA_gyraseB,DNA_gyraseB_C,HATPase_c,Toprim	K02470	-	-	-
⋮								
CDS	797,240	798,175	+	PALP	K01738	-	-	-
CDS	798,178	799,182	+	SBP_bac_11	K02048	-	-	-
CDS	799,163	800,038	+	BPD_transp_1	K02046, K15496	-	-	-
CDS	800,061	800,933	+	BPD_transp_1	K02047	-	-	-
CDS	800,938	801,999	+	ABC_tran,TOBE_2	K02045, K02052	-	-	-
CDS	802,024	803,736	+	FAD_binding_2,Succ_DH_flav_C	K00394	-	-	-
CDS	803,720	804,034	+	Fer4,Fer4_9	K00395	-	-	-
CDS	804,135	805,034	+	PAPS_reduct	K00390, K00957	ICEValHN437|GenBank|KT072771|1..94290 Vibrio alginolyticus strain HN437 transposon integrating conjugative element ICEValHN437, complete sequence.	67	-
CDS	805,036	806,673	+	APS_kinase,GTP_EFTU,GTP_EFTU_D2,GTP_EFTU_D3	K00955, K00956	-	-	-
⋮								
CDS	3,517,457	3,520,501	+	Gram_pos_anchor	-	Tn6079|GenBank|GU951538|462..28872 Uncultured bacterium MID12 genomic sequence.	93	-
CDS	3,520,740	3,521,066	+	DUF961	-	Tn6079|GenBank|GU951538|462..28872 Uncultured bacterium MID12 genomic sequence.	94	-
CDS	3,521,082	3,521,465	+	DUF961	-	CTn7|GenBank|AM180355|3938102..3963961 Clostridium difficile 630 complete genome.	95	-
CDS	3,521,588	3,522,361	+	-	-	-	-	-
CDS	3,522,438	3,522,965	+	GAD-like	-	-	-	-
CDS	3,523,066	3,523,569	+	-	-	-	-	-
CDS	3,523,609	3,524,436	+	DUF5037	-	-	-	-
CDS	3,524,482	3,524,994	+	-	-	-	-	-
CDS	3,525,037	3,525,621	+	-	-	-	-	-
CDS	3,525,726	3,526,031	+	FtsK_SpoIIIE	-	Tn6194-like|GenBank|HG475346|1..28014 [Clostridium] difficile Tn6194-like conjugative transposon, strain CII7.	95	-
CDS	3,526,078	3,527,694	-	DDE_Tnp_IS66,DDE_Tnp_IS66_C,LZ_Tnp_IS66,zf-IS66	-	-	-	-
CDS	3,528,113	3,529,150	-	Phage_int_SAM_1,Phage_int_SAM_4,Phage_int_SAM_5,Phage_integrase	K04763	-	-	-
CDS	3,529,147	3,530,127	-	Phage_integrase	-	-	-	-
CDS	3,530,124	3,531,347	-	-	-	-	-	-
CDS	3,531,459	3,531,671	-	TnpB_IS66	K07484	-	-	-
CDS	3,531,665	3,532,078	-	HTH_23,HTH_Tnp_1	-	-	-	-
CDS	3,532,381	3,533,355	+	FtsK_SpoIIIE	-	Tn6194-like|GenBank|HG475346|1..28014 [Clostridium] difficile Tn6194-like conjugative transposon, strain CII7.	98	**Coupling protein**
CDS	3,533,539	3,534,735	+	HTH_3,HTH_31,Rep_trans	K07467	Tn6194-like|GenBank|HG475346|1..28014 [Clostridium] difficile Tn6194-like conjugative transposon, strain CII7.	100	**Relaxase**
CDS	3,534,748	3,534,882	+	DUF3789	-	CTn1|GenBank|AM180355|428851..453332 Clostridium difficile 630 complete genome.	75	-
CDS	3,534,883	3,535,104	+	-	-	Tn6194-like|GenBank|HG475346|1..28014 [Clostridium] difficile Tn6194-like conjugative transposon, strain CII7.	99	-
CDS	3,535,181	3,535,654	+	-	-	CTnBST|GenBank|AY345595|1..100903 Bacteroides uniformis strain WH207 transposon CTnBST, complete sequence.	54	-
CDS	3,535,756	3,536,052	+	-	-	CTn1|GenBank|AM180355|428851..453332 Clostridium difficile 630 complete genome.	92	-
CDS	3,535,992	3,536,474	+	ArdA	-	CTn7|GenBank|AM180355|3938102..3963961 Clostridium difficile 630 complete genome.	91	-
CDS	3,536,492	3,536,995	+	ArdA	-	Tn6194-like|GenBank|HG475346|1..28014 [Clostridium] difficile Tn6194-like conjugative transposon, strain CII7.	99	-
CDS	3,537,114	3,537,512	+	TcpE	-	CTn7|GenBank|AM180355|3938102..3963961 Clostridium difficile 630 complete genome.	99	-
CDS	3,537,490	3,539,940	+	AAA_10	-	Tn5386|GenBank|DQ321786|67..29238 Enterococcus faecium strain D344R transposon Tn5386, complete sequence.	97	**VirB4**
CDS	3,539,940	3,542,150	+	-	-	CTn1|GenBank|AM180355|428851..453332 Clostridium difficile 630 complete genome.	92	-
CDS	3,542,147	3,543,154	+	Lysozyme_like,NLPC_P60	-	Tn6194-like|GenBank|HG475346|1..28014 [Clostridium] difficile Tn6194-like conjugative transposon, strain CII7.	98	-
CDS	3,543,171	3,544,082	+	TpcC	-	CTn1|GenBank|AM180355|428851..453332 Clostridium difficile 630 complete genome.	97	-
CDS	3,544,261	3,545,118	+	Acetyltransf_1,Cass2,HTH_18	-	-	-	-
CDS	3,545,111	3,545,488	+	MatE	-	-	-	-
CDS	3,545,668	3,546,753	+	DDE_Tnp_1	-	-	-	-
CDS	3,546,887	3,547,816	+	MatE	-	-	-	-
CDS	3,547,813	3,548,037	+	-	-	CTn1|GenBank|AM180355|428851..453332 Clostridium difficile 630 complete genome.	93	-
CDS	3,548,083	3,549,375	-	HATPase_c,HisKA	-	-	-	-
CDS	3,549,369	3,550,055	-	Response_reg,Trans_reg_C	-	-	-	-
CDS	3,550,249	3,550,743	+	-	-	-	-	-
CDS	3,550,755	3,551,978	+	FtsX	-	-	-	-
CDS	3,551,982	3,553,427	+	FtsX,MacB_PCD	K02004	-	-	-
CDS	3,553,441	3,554,076	+	ABC_tran	-	CTn5|GenBank|AM180355|2137789..2181291 Clostridium difficile 630 complete genome.	60	-
CDS	3,554,471	3,554,791	+	-	-	-	-	-
CDS	3,554,776	3,555,009	+	HTH_16	-	Tn6194-like|GenBank|HG475346|1..28014 [Clostridium] difficile Tn6194-like conjugative transposon, strain CII7.	56	-
CDS	3,555,323	3,555,526	+	Tn916-Xis	-	CTn1|GenBank|AM180355|428851..453332 Clostridium difficile 630 complete genome.	100	-
CDS	3,555,605	3,556,795	+	Integrase_DNA,Phage_int_SAM_5,Phage_integrase	-	CTn1|GenBank|AM180355|428851..453332 Clostridium difficile 630 complete genome.	99	**Tyrosine integrase**
⋮								
CDS	4,286,634	4,286,846	-	Haemolytic	K08998	-	-	-
CDS	4,286,850	4,287,200	-	Ribonuclease_P	K03536	-	-	-
CDS	4,287,250	4,287,384	-	Ribosomal_L34	K02914	ICEAmeAS1|GenBank|NC_019393|1271627..1383296 Alteromonas mediterranea DE1, complete genome.	71	-
CDS	4,287,896	4,289,569	-	DDE_Tnp_IS66	-	-	-	-
CDS	4,289,595	4,289,759	-	-	-	-	-	-

※ Row 7 to 764, 774 to 3460, 3505 to 4260 cannot be displayed due to lack of space. The complete version will be provided by Excel format.

Extended data Table 6. Gene annotations of *Tyzzerella nexilis*_strain B1

Gene annotations of the *T.nexilis*_strain B1 genome. Gene annotations based on Pfam, KEGG orthologies, and ICE-related genes are shown.

Extended data Table 7

Demography	HC (n = 29)	RRMS (n = 62)	SPMS (n = 15)	p-value
Age, year	38.6 ± 2.28	39.0 ± 1.00	43.3 ± 2.60	0.26
Sex (female ; male)	17 ; 12	46 ; 16	9 ; 6	0.26
BMI, kg/m2	21.4 ± 0.40	22.0 ± 0.41	21.3 ± 0.69	0.15
Onset age, year		30.0 ± 1.08	27.3 ± 2.25	0.27
Disease duration, year		9.0 ± 0.84	15.9 ± 1.54	0.0004
ARR		0.68 ± 0.11	0.13 ± 0.13	0.02
EDSS score		1.88 ± 0.17	5.53 ± 0.45	<0.0001
Immunotherapy				
Oral PSL, %		40 (n = 25)	60 (n = 9)	
IS, %		5 (n = 3)	27 (n = 4)	
DMDs				
IFN-β, %		37 (n = 23)	13 (n = 2)	
GA, %		5 (n = 3)	0	
FTY, %		6 (n = 4)	13 (n = 2)	
NTZ, %		0	7 (n = 1)	
DMF, %		3 (n = 2)	0	

Extended data Table 7. Demographics and characteristics of the patients and controls

Sixty-two patients with RRMS, 15 patients with SPMS, and 29 healthy controls were recruited. Data are represented as mean ± SEM. Abbreviations: BMI = body mass index; ARR = annual relapse rate; EDSS = expanded disability status scale; PSL = prednisolone; IS = immunosuppressive drugs; DMDs; disease-modifying drugs; IFN-β = interferon-β; GA = glatiramer acetate; FTY = fingolimod; NTZ = natalizumab; DMF = dimethyl fumarate; HC = healthy control; RRMS = relapsing-remitting multiple sclerosis; SPMS = secondary progressive multiple sclerosis.

Extended data Table 8

ID	disease	site	ethnicity	age	sex	BMI	year of onset	disease duration	treatment status	treatments	EDSS score
RRMS_001	RRMS	Tokyo	Asian	41	F	24.2	15	26	Untreated	Untreated	1.5
RRMS_002	RRMS	Tokyo	Asian	32	M	23.1	16	16	Treated	Interferon	1
RRMS_003	RRMS	Tokyo	Asian	33	F	26.7	26	7	Treated	Interferon	1
RRMS_004	RRMS	Tokyo	Asian	43	F	23.6	34	9	Treated	Steroid	3.5
RRMS_005	RRMS	Tokyo	Asian	33	F	19.4	26	7	Untreated	Untreated	0
RRMS_006	RRMS	Tokyo	Asian	36	M	20.7	33	3	Untreated	Untreated	0
RRMS_007	RRMS	Tokyo	Asian	31	F	21.6	25	6	Untreated	Untreated	0
RRMS_008	RRMS	Tokyo	Asian	34	F	23.4	27	7	Treated	Interferon	1
RRMS_009	RRMS	Tokyo	Asian	44	F	19.1	38	6	Treated	Interferon	1
RRMS_010	RRMS	Tokyo	Asian	53	F	22.2	35	18	Treated	Steroid	3
RRMS_011	RRMS	Tokyo	Asian	40	F	19.7	21	19	Treated	Interferon	1
RRMS_012	RRMS	Tokyo	Asian	33	F	21.8	32	1	Untreated	Untreated	0
RRMS_013	RRMS	Tokyo	Asian	38	M	22.5	18	20	Treated	Interferon	2.5
RRMS_014	RRMS	Tokyo	Asian	41	F	23.1	35	6	Treated	Steroid	2
RRMS_015	RRMS	Tokyo	Asian	50	F	20.1	40	10	Treated	Steroid	7.5
RRMS_016	RRMS	Tokyo	Asian	30	F	32.4	26	4	Untreated	Untreated	2
RRMS_017	RRMS	Tokyo	Asian	35	M	20.5	22	13	Treated	Steroid	4
RRMS_018	RRMS	Tokyo	Asian	53	F	25.2	39	14	Treated	Steroid	4.5
RRMS_019	RRMS	Tokyo	Asian	26	F	18.7	23	3	Untreated	Untreated	2
RRMS_020	RRMS	Tokyo	Asian	43	F	20.7	39	4	Untreated	Untreated	1
RRMS_021	RRMS	Tokyo	Asian	33	F	20.1	19	14	Untreated	Untreated	1
RRMS_022	RRMS	Tokyo	Asian	44	F	19.6	22	22	Treated	Fingolimod + steroid	2
RRMS_023	RRMS	Tokyo	Asian	56	F	23.3	48	8	Treated	Steroid + immunosuppressant	1.5
RRMS_024	RRMS	Tokyo	Asian	30	M	25.6	18	12	Treated	Interferon	1.5
RRMS_025	RRMS	Tokyo	Asian	35	F	19.1	18	17	Treated	Interferon + steroid	3
RRMS_026	RRMS	Tokyo	Asian	35	M	26.7	25	10	Treated	Interferon	1
RRMS_027	RRMS	Tokyo	Asian	46	F	21.1	41	5	Treated	Interferon	1
RRMS_028	RRMS	Tokyo	Asian	36	F	21.8	23	13	Untreated	Untreated	1.5
RRMS_029	RRMS	Tokyo	Asian	37	F	22.8	25	12	Treated	Steroid	3
RRMS_030	RRMS	Tokyo	Asian	45	F	19.6	36	9	Treated	Steroid	1.5
RRMS_031	RRMS	Tokyo	Asian	46	M	23.7	40	6	Treated	Interferon + steroid	2
RRMS_032	RRMS	Tokyo	Asian	38	F	21.7	36	2	Untreated	Untreated	1
RRMS_033	RRMS	Tokyo	Asian	42	M	20.4	37	5	Treated	Interferon + steroid	3
RRMS_034	RRMS	Tokyo	Asian	37	F	17.2	31	6	Treated	Steroid	2
RRMS_035	RRMS	Tokyo	Asian	38	M	30.1	32	6	Treated	Dimethyl fumarate	1
RRMS_036	RRMS	Tokyo	Asian	35	F	17.3	23	12	Treated	Interferon	1
RRMS_037	RRMS	Tokyo	Asian	39	M	20.8	37	2	Treated	Interferon + steroid	0
RRMS_038	RRMS	Tokyo	Asian	25	F	24.2	22	3	Treated	Steroid	1
RRMS_039	RRMS	Tokyo	Asian	40	M	21.2	31	9	Treated	Steroid	2.5
RRMS_040	RRMS	Tokyo	Asian	38	F	26.0	30	8	Treated	Fingolimod	1
RRMS_041	RRMS	Tokyo	Asian	42	F	21.3	38	4	Treated	Steroid + immunosuppressant	2
RRMS_042	RRMS	Tokyo	Asian	31	F	19.8	30	1	Treated	Steroid	2
RRMS_043	RRMS	Tokyo	Asian	50	F	19.0	18	32	Treated	Fingolimod	3.5
RRMS_044	RRMS	Tokyo	Asian	33	F	23.4	30	3	Treated	Steroid	3
RRMS_045	RRMS	Tokyo	Asian	49	F	24.5	27	22	Treated	Steroid	4
RRMS_046	RRMS	Tokyo	Asian	49	F	30.4	44	5	Treated	Interferon	1.5
RRMS_047	RRMS	Tokyo	Asian	47	M	26.5	29	18	Treated	Interferon	1
RRMS_048	RRMS	Tokyo	Asian	44	F	23.4	29	15	Treated	Steroid	2.5
RRMS_049	RRMS	Tokyo	Asian	29	M	20.8	27	2	Treated	Glatiramer acetate	1
RRMS_050	RRMS	Tokyo	Asian	35	F	19.9	23	12	Treated	Interferon	1.5
RRMS_051	RRMS	Tokyo	Asian	34	F	20.7	34	0	Treated	Fingolimod	3.5
RRMS_052	RRMS	Tokyo	Asian	59	F	25.4	52	7	Treated	Glatiramer acetate	2
RRMS_053	RRMS	Tokyo	Asian	42	M	21.2	41	1	Treated	Interferon + steroid	3
RRMS_054	RRMS	Tokyo	Asian	23	M	16.7	22	1	Treated	Steroid	1
RRMS_055	RRMS	Tokyo	Asian	45	F	18.4	38	7	Treated	Interferon + steroid + immunosuppressant	3
RRMS_056	RRMS	Tokyo	Asian	37	M	23.4	27	10	Treated	Interferon	3.5
RRMS_057	RRMS	Tokyo	Asian	42	F	17.9	36	6	Treated	Interferon	1
RRMS_058	RRMS	Tokyo	Asian	32	F	19.2	24	8	Treated	Interferon	2.5
RRMS_059	RRMS	Tokyo	Asian	26	F	20.8	18	8	Treated	Interferon	1
RRMS_060	RRMS	Tokyo	Asian	36	F	16.4	33	3	Treated	Dimethyl fumarate	0
RRMS_061	RRMS	Tokyo	Asian	37	F	19.2	32	5	Untreated	Untreated	1
RRMS_062	RRMS	Tokyo	Asian	54	F	21.4	46	8	Treated	Glatiramer acetate	3
SPMS_001	SPMS	Tokyo	Asian	37	F	22.3	22	15	Treated	Steroid	6.5
SPMS_002	SPMS	Tokyo	Asian	39	F	19.7	27	12	Untreated	Untreated	3.5
SPMS_003	SPMS	Tokyo	Asian	42	M	19.9	32	10	Treated	Steroid	3.5
SPMS_004	SPMS	Tokyo	Asian	44	M	20.7	25	19	Treated	Interferon	7
SPMS_005	SPMS	Tokyo	Asian	33	M	19.9	24	9	Treated	Steroid	6
SPMS_006	SPMS	Tokyo	Asian	35	F	18.6	21	14	Treated	Steroid + immunosuppressant	4
SPMS_007	SPMS	Tokyo	Asian	39	F	21.3	18	21	Treated	Interferon + steroid	2.5
SPMS_008	SPMS	Tokyo	Asian	33	M	21.3	27	6	Treated	Fingolimod	6
SPMS_009	SPMS	Tokyo	Asian	42	M	29.1	32	10	Treated	Fingolimod	6
SPMS_010	SPMS	Tokyo	Asian	66	F	19.5	50	16	Treated	Steroid + immunosuppressant	7.5
SPMS_011	SPMS	Tokyo	Asian	46	F	18.4	20	26	Treated	Steroid + immunosuppressant	7
SPMS_012	SPMS	Tokyo	Asian	37	F	21.0	23	14	Treated	Steroid	6
SPMS_013	SPMS	Tokyo	Asian	53	M	21.2	31	22	Treated	Immunosuppressant	7
SPMS_014	SPMS	Tokyo	Asian	63	F	25.0	40	23	Treated	Natalizmab + steroid	7.5
SPMS_015	SPMS	Tokyo	Asian	40	F	21.2	18	22	Untreated	Untreated	3
HC_001	Control	Tokyo	Asian	32	M	26.2	N/A	N/A	N/A	N/A	N/A
HC_002	Control	Tokyo	Asian	46	F	20.7	N/A	N/A	N/A	N/A	N/A
HC_003	Control	Tokyo	Asian	31	M	25.9	N/A	N/A	N/A	N/A	N/A
HC_004	Control	Tokyo	Asian	47	M	17.6	N/A	N/A	N/A	N/A	N/A
HC_005	Control	Tokyo	Asian	57	F	21.5	N/A	N/A	N/A	N/A	N/A
HC_006	Control	Tokyo	Asian	65	F	19.1	N/A	N/A	N/A	N/A	N/A
HC_007	Control	Tokyo	Asian	61	F	21.7	N/A	N/A	N/A	N/A	N/A
HC_008	Control	Tokyo	Asian	30	F	21.2	N/A	N/A	N/A	N/A	N/A
HC_009	Control	Tokyo	Asian	60	F	22.2	N/A	N/A	N/A	N/A	N/A
HC_010	Control	Tokyo	Asian	25	F	21.9	N/A	N/A	N/A	N/A	N/A
HC_011	Control	Tokyo	Asian	62	M	24.1	N/A	N/A	N/A	N/A	N/A
HC_012	Control	Tokyo	Asian	28	F	18.1	N/A	N/A	N/A	N/A	N/A
HC_013	Control	Tokyo	Asian	23	F	19.7	N/A	N/A	N/A	N/A	N/A
HC_014	Control	Tokyo	Asian	31	M	23.9	N/A	N/A	N/A	N/A	N/A
HC_015	Control	Tokyo	Asian	33	M	22.5	N/A	N/A	N/A	N/A	N/A
HC_016	Control	Tokyo	Asian	35	M	21.2	N/A	N/A	N/A	N/A	N/A
HC_017	Control	Tokyo	Asian	29	F	20.9	N/A	N/A	N/A	N/A	N/A
HC_018	Control	Tokyo	Asian	30	M	22.7	N/A	N/A	N/A	N/A	N/A
HC_019	Control	Tokyo	Asian	26	M	21.8	N/A	N/A	N/A	N/A	N/A
HC_020	Control	Tokyo	Asian	24	F	21.5	N/A	N/A	N/A	N/A	N/A
HC_021	Control	Tokyo	Asian	47	F	18.8	N/A	N/A	N/A	N/A	N/A
HC_022	Control	Tokyo	Asian	43	F	23.5	N/A	N/A	N/A	N/A	N/A
HC_023	Control	Tokyo	Asian	36	M	20.9	N/A	N/A	N/A	N/A	N/A
HC_024	Control	Tokyo	Asian	37	F	19.4	N/A	N/A	N/A	N/A	N/A
HC_025	Control	Tokyo	Asian	35	M	21.9	N/A	N/A	N/A	N/A	N/A
HC_026	Control	Tokyo	Asian	31	F	23.2	N/A	N/A	N/A	N/A	N/A
HC_027	Control	Tokyo	Asian	35	F	18.0	N/A	N/A	N/A	N/A	N/A
HC_028	Control	Tokyo	Asian	42	M	22.4	N/A	N/A	N/A	N/A	N/A
HC_029	Control	Tokyo	Asian	37	F	19.2	N/A	N/A	N/A	N/A	N/A

Extended data Table 8. Clinical information of patients and controls

Detailed clinical information including recruitment site, ethnicity, age, sex, body mass index (BMI), year of onset, disease duration, treatments, and EDSS score is presented.