Noncontiguous finished genome sequence of *Megasphaera* sp. ASD88, isolated from faeces of a child with autism spectrum disorder

A. V. Chaplin¹, S. I. Polyakova², T. O. Kochetkova³, S. R. Sokolova¹, A. Y. Goltsov¹, D. Y. Trofimov³ and B. A. Efimov¹

¹) Department of Microbiology and Virology, Pirogov Russian National Research Medical University, Moscow, Russia, ²) Hospital Pediatrics Department Named After Academician V. A. Tabolin, Pirogov Russian National Research Medical University, Moscow, Russia and ³) Laboratory of Molecular Genetic Methods, Research Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia

Abstract

We report here a draft genome sequence of *Megasphaera* sp. ASD88, a strain from the intestinal microbiota of a child with autism spectrum disorder, representing a previously undescribed species of the genus *Megasphaera*. The assembled sequence consists of 88 scaffolds, and the total size is 2.59 Mb.

© 2018 The Author(s). Published by Elsevier Ltd.

Keywords: Autism spectrum disorder, Gut microbiota, *Megasphaera*, Negativicutes, Novel species, Taxonogenomics

Original Submission: 3 October 2017; Revised Submission: 7 December 2017; Accepted: 20 December 2017

Article published online: 8 January 2018

Corresponding author: A. V. Chaplin, Department of Microbiology and Virology, Pirogov Russian National Research Medical University, Ostrovitianov Str. 1, Moscow, 117997, Russia.
E-mail: okolomedik@gmail.com

Autism spectrum disorder (ASD) is a lifelong neurodevelopmental disorder with onset in early childhood. It is characterized by deficits in communication and social interaction as well as restricted, stereotyped patterns of interests or behavior. Clinical studies have demonstrated increased frequency of gastrointestinal symptoms in children with ASD [1]. Multiple metabolites, produced by intestinal microbiota, can affect gut–brain interaction and possibly ASD development [2].

Nowadays, the majority of studies of intestinal microbiota uses metagenomic or amplicon library sequencing approaches. Despite the fact that culturomic studies lack the capability to cover all taxonomic groups within this microbial community, researchers today successfully apply culturing techniques to isolate and describe the properties of a vast number of previously uncultured microbial species [3,4]. Development of new culture media and identification methods, including matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS), facilitates the progress of culturomic approach.

We performed culture-based study of the faecal microbiome in children with ASD in mid-2017. The child’s parents agreed to participate in the study and signed informed consent. The ethics committee of Pirogov Russian National Research Medical University validated the study (protocol 165, 22 May 2017).

During the study, we isolated two strains of strictly anaerobic, Gram-negative bacteria that presumably belonged to the genus *Megasphaera* but could not be identified to the species level by using systematic MALDI-TOF MS screening with VITEK MS Plus (bioMérieux, Marcy l’Étoile, France). We obtained strain *Megasphaera* sp. ASD88 from the faecal sample of a 8-year-old boy, and strain *Megasphaera* sp. ASD97 from a sample taken from a 9-year-old girl. The children were siblings, and both had a diagnosis of atypical autism (International Classification of Disease F84.1). Preliminary analysis, based on clustering of mass spectrometry data (SARAMIS Premium software; AnagnosTec, Potsdam-Golm, Germany), had shown that these strains are highly similar (Fig. 1).

Faecal specimens were collected and transferred anaerobically. Samples were weighed, serially diluted with saline and spread onto several media including anaerobe base agar (Oxoid, Basingstoke, UK) with 5% (v/v) defibrinated sheep’s blood. Plates were incubated anaerobically (in an atmosphere of 85% N₂, 10% H₂, 5% CO₂) at 37°C for 72 hours in anaerobic jars (Schuett-Bioteck, Göttingen, Germany). Well-isolated single colonies were picked and streaked several times to obtain pure...
cultures on the same medium. Strain ASD88 was present in stool at a concentration of $\sim 1 \times 10^8$ colony-forming units per gram. Strain ASD97 was isolated at a concentration of $\sim 9 \times 10^8$ colony-forming units per gram. Interestingly, both stool samples also contained *Megasphaera elsdenii* at concentrations of $\sim 2 \times 10^8$ and $\sim 3 \times 10^9$, respectively. Upon isolation, strains ASD88 and ASD97 were cultured anaerobically in Schaedler broth (Oxoid) and were preserved by freeze-drying of bacterial suspensions in 10% (w/v) sucrose and 1% (w/v) gelatin solution.

After 48 hours’ incubation on Anaerobe base agar (Oxoid) with 5% (v/v) defibrinated sheep’s blood, strains ASD88 and ASD97 formed colonies 0.6 to 1.0 mm and 1.0 to 1.4 mm, respectively, in diameter. They appeared slightly greenish, glossy, slightly convex, opaque and smooth with entire circular edges. ASD97 had weak haemolytic activity. Cells of both strains were Gram stain-negative ovoid cocci (0.7–1.1 × 0.5–0.8 μm and 0.5–1.1 × 0.4–0.7 μm, respectively) occurring singly or in pairs, or as short chains. The strains could grow well on Anaerobe base agar (Oxoid) under anaerobic conditions but not under aerobic conditions. Both strains had no catalase activity and no oxidase activity. These strains were also nonmotile and non-endospore forming.

Because of the high similarity of strains, only ASD88 was selected for genome sequencing. We performed sequencing of genomic DNA using the Ion Proton system and de novo assembly with SPAdes 3.9.1 [5]. Genome coverage was 135×. The resulting genome sequence contained 88 scaffolds; the total sequence length was 2 589 137 bp.

The GC content of the genome was 52.7%. The National Center for Biotechnology Information (NCBI) Prokaryotic Genome Annotation Pipeline predicted a total of 2530 genes, including 2440 protein coding sequences (CDS). The genome contained four CRISPR arrays (with 148 spacers in total) with nearly identical repeat sequences and one set of cas genes. We did not find any high-coverage contigs that may have represented the plasmids. Several putative prophages were located within this genome sequence: loci CJ260_00510-CJ260_00995, CJ260_10060, CJ260_10300, CJ260_11310-CJ260_11490 and CJ260_11495-CJ260_11630.

To infer the phylogenetic position relative to known species (Fig. 2), we aligned the 16S rRNA gene sequences of the strain ASD88 and related strains using MUSCLE [6] and performed a maximum likelihood approach implemented in raxML [7] (General time-reversible (GTR) model with gamma distributed rate heterogeneity, 1000 rapid bootstrap resamplings). Strain ASD88 exhibited a 94.7% sequence identity of 16S rRNA gene sequence with *Megasphaera elsdenii* DSM 20460T (accession no. NR_102980.1), the type strain of the most similar by 16S rRNA gene sequence species with standing in nomenclature. The obtained value was lower than the 98.7% threshold recommended to delineate a new species. Thus, the strain *Megasphaera* sp. ASD88 represents a previously undescribed species of the genus *Megasphaera*.
FIG. 2. Phylogenetic tree of 16S rRNA gene sequences of Megasphaera strains. 16S rRNA genes of strain An286 was reassembled from publicly available reads. Accession number of assembled sequence or sequence read archive in DDBJ/European Molecular Biology Laboratory/GenBank is provided next to strain name. Node labels represent bootstrap confidence levels.

FIG. 3. Phylogenetic tree of concatenated sequences of 582 conserved proteins. Accession number in DDBJ/European Molecular Biology Laboratory/GenBank is provided next to strain name. Node labels represent bootstrap confidence levels.
To reconstruct a phylogenetic tree based on protein sequences, we used whole genome sequences of *Megasphaera* strains available from NCBI RefSeq as well as *Veillonella parvula* DSM 2008 as an outgroup. We performed clustering of encoded proteins using OrthoMCL [8]. As a result, we obtained a set of 582 orthogroups that were present in every genome and did not contain paralogs. Amino acid sequences within each group were aligned using MUSCLE, and then the alignments were concatenated. Phylogenetic inference was performed using RapidNJ [9] with 1000 bootstrap resamplings. On the inferred tree (Fig. 3), strain ASD88 was located on a separate branch from all described species of *Megasphaera* genus.

The non-type isolates with sequenced genomes closely related to *Megasphaera* sp. ASD88 were *Megasphaera* sp. strain An286, isolated from chicken caecum, and uncultured *Megasphaera* sp. isolate 2789STDY5834854, isolated from healthy human faeces. These three strains shared 96.9% to 97.0% average nucleotide identity ([Species WS server, ANIb calculation method [10]), indicating that they are distant but belong to the same undescribed species.

To compare 16S RNA gene sequences among isolates closely related to ASD88, we performed Sanger sequencing of 16S rRNA gene of ASD97 using the primer pair Bact8F/ Bact1492R [11]. Deposited draft genome sequence of strain An286 did not contain assembled 16S rRNA genes, so we created a reassembly from publicly available reads with Bowtie 2 [12] using 16S rRNA gene of ASD88 as a reference. We also added the 16S rRNA gene sequences of uncultured clone CFT114G8 (DQ456131.1) and isolate 2789STDY5834854 (contig FMFF01000017.1, coordinates 2059–519). All the listed strains grouped together on the phylogenetic tree (Fig. 2) and shared 16S rRNA gene sequence similarity of 99.6% to 99.9%, showing that they belong to the same species. To elucidate metabolic pathways of ASD88, we performed the functional annotation of the encoded proteins using BlastKOALA [13,14]. We found genes encoding enzymes of complete tryptophan biosynthesis pathway, but no genes involved in the production of neuroactive compounds (tryptamine, 5-hydroxytryptamine, dopamine, L-DOPA and γ-aminobutyric acid).

The 16S rRNA gene and genome sequence of *Megasphaera* sp. ASD88 were deposited in GenBank under accession numbers MF765328 and NQXW000000000, respectively. Strain ASD88 was deposited in the All-Russian Collection of Microorganisms (VKM) under number VKM B-3207. The 16S rRNA gene sequence of *Megasphaera* sp. ASD97 was deposited in GenBank under accession number MG321613.

The current study provides no information on linkage between ASD and colonization of intestine by *Megasphaera* species. Further research is required to elucidate a possible association.

**Acknowledgement**

This study was supported by Russian Science Foundation grant number 17-15-01488.

**Conflict of interest**

None declared.

**References**

[1] McElhanon BO, McCracken C, Karpen S, Sharp WG. Gastrointestinal symptoms in autism spectrum disorder: a meta-analysis. Pediatrics 2014;133:872–83.

[2] Li Q, Han Y, Dy ABC, Hagerman RJ. The gut microbiota and autism spectrum disorders. Front Cell Neurosci 2017;11:120.

[3] Lagier JC, Khelaiifi S, Alou MT, Ndongo S, Dione N, Hugon P, et al. Culture of previously uncultured members of the human gut microbiota by culturogenesis. Nat Microbiol 2016;1:16203.

[4] Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, et al. Culture of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. Nature 2016;533:543–6.

[5] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455–77.

[6] Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004;32:1792–7.

[7] Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3.

[8] Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of orthologous groups for eukaryotic genomes. Genome Res 2003;13:2178–89.

[9] Simonsen M, Mailund T, Pedersen CNS. Rapid neighbour-joining. In: Crandall KA, Lagergren J, editors. Algorithms in bioinformatics. WABI 2008. Lecture notes in computer science, 5251. Berlin: Springer; 2008. p. 113–22.

[10] Richter M, Rosselló-Mora R, Oliver Glöckner F, Peplies J. Species WS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 2016;32:929–31.

[11] Turner S, Pryer KM, Mao VP, Palmer JD. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J Eukaryot Microbiol 1999;46:327–38.

[12] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357–9.

[13] Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 2016;428:726–31.

[14] Kaneshia M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 2016;44:D457–62.