Comparative genomics of *Neisseria weaveri* clarifies the taxonomy of this species and identifies genetic determinants that may be associated with virulence

Hana Yi¹, Yong-Joon Cho², Seok-Hwan Yoon³, Sang-Cheol Park³ & Jongsik Chun¹,²,³,⁴

¹Institute of Molecular Biology and Genetics, Seoul, Korea; ²Chunlab, Inc., Seoul, Korea; ³School of Biological Sciences, Seoul National University, Seoul, Korea; and ⁴AICT, Gyeonggi-do, Korea

Correspondence: Jongsik Chun, School of Biological Sciences, Seoul National University, 599 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea. Tel.: +82 2 880 8153; fax: +82 2 874 8153; e-mail: jchun@snu.ac.kr

Received 1 December 2011; revised 9 December 2011; accepted 9 December 2011. Final version published online 6 January 2012.

DOI: 10.1111/j.1574-6968.2011.02485.x

Editor: Aharon Oren

Keywords

*Neisseria weaveri*; comparative genomics; reclassification.

Abstract

A group of bacterial strains formerly known as CDC group M-5 are opportunistic pathogens to humans. In 1993, a name, *Neisseria weaveri*, was proposed by two independent studies to harbor CDC group M-5 strains, namely *N. weaveri* Holmes et al. 1993 and *N. weaveri* Andersen et al. 1993, with two different ‘type’ strains. However, no study has been conducted on the relatedness of the two ‘type’ strains, although the close relationship of the two taxa has long been accepted unofficially. Formally, the status of the name *N. weaveri* Andersen et al. 1993 is illegitimate because it is a later homonym of *N. weaveri* Holmes et al., 1993; but the name of the strain is still validly published. In this study, we attempt to resolve the confusion caused by the apparent duplication of the species *N. weaveri* (with different type strains) using whole genome shotgun sequencing. We also sought to gain insight into the genetic characteristics of *N. weaveri* by conducting comparative genomics. On the basis of genomic similarities revealed through a comparative genomic study, we propose that *N. weaveri* Andersen et al. 1993 should be re-classified as a later heterotypic synonym of *N. weaveri* Holmes et al., 1993.

Introduction

The genus *Neisseria* is composed of commensal bacteria that colonize the mucous membranes of mammals. *Neisseria* encompasses two important pathogens – *Neisseria meningitidis* and *Neisseria gonorrhoeae* – as well as many other opportunistic pathogens (Janda & Knapp, 2003; Han et al., 2006). Extensive genomic analyses have been successfully applied to reveal pathogenic mechanisms and vaccine candidates (Sun et al., 2000; Dunning Hotopp et al., 2006; Kawai et al., 2006; Maiden, 2008; Schoen et al., 2008; Aspholm et al., 2010; Marri et al., 2010; Joseph et al., 2011), and as a result, 46 *Neisseria* genome sequences of human origin are available from public databases. However, none of the genomes from strains that originated from animals have been studied.

A group of bacterial strains previously known as CDC group M-5 are part of the normal canine oral flora, but it has known opportunistic pathogenicity in human peritonitis (Kocyigit et al., 2010), lower respiratory tract infections (Panagea et al., 2002), cervical and vaginal infections (Flores-Paz et al., 2003), wound infections (Captopini et al., 2002), and septicemia (Carlson et al., 1997). In 1993, a name, *Neisseria weaveri*, has been proposed by two independent studies to harbor CDC group M-5 strains, namely *N. weaveri* Holmes et al. 1993 and *N. weaveri* Andersen et al. 1993. The type strain defined by Holmes et al. (1993) was isolated from human dog bite wound in Göteborg, Sweden (1974) and that defined by Andersen et al. (1993a) was isolated from same type of wound in New York, USA (1962). Even though both taxa were published in the same volume of *International Journal of Systematic Bacteriology* (IJSB), *N. weaveri* Holmes et al. 1993 has page precedence over *N. weaveri* Andersen et al. 1993 according to Bacteriological Code Rules 51b (4) and 24b (2).
(Lagage et al., 1992). Thus, *N. weaveri* Andersen et al. 1993 is illegitimate because it is a later homonym of *N. weaveri* Holmes et al. 1993 (Euzéby, 1998). Although different type strains were deposited as representing *N. weaveri*, the close relationship of the two taxa has long been accepted because of the similar pathogenic characteristics of the two ‘type’ strains. However, there has been no systematic investigation about the relatedness of these two ‘type’ strains and thus the illegitimate name has remained as a homonym and not transferred as a synonym of *N. weaveri* Holmes et al. 1993.

Thus, in this study, we attempt to resolve the confusion caused by two *N. weaveri* species with different ‘type’ strains using whole genome shotgun sequencing. We also sought to gain insight into the genetic characteristics of *N. weaveri* by conducting comparative genomics.

**Materials and methods**

**Bacterial strains and DNA extraction**

The ‘type’ strains of *N. weaveri* Holmes et al. 1993 (LMG 5135<sup>T</sup>) and *N. weaveri* Andersen et al. 1993 (ATCC 51223<sup>T</sup>) were obtained from LMG and ATCC, respectively, and the genomic DNAs were extracted using DNeasy Blood & Tissue kit (Qiagen).

**Genome sequencing**

The draft genome sequences of strains LMG 5135<sup>T</sup> and ATCC 51223<sup>T</sup> were determined by paired-end shotgun sequencing using the Genome Analyzer IIX (Illumina) with > 1000× fold coverage. The sequencing reads were assembled using the CLC genomics wb4 (CLCbio) and CodonCode Aligner (CodonCode Co.).

**Annotation and comparative genomics**

Gene finding and annotation were achieved using the RAST server (Aziz et al., 2008). Orthologous gene prediction and comparative genomic analyses were conducted as described previously (Chun et al., 2009). In brief, a segment on target contig, which is homologous to a query open reading frame (ORF), was identified using the BLASTN program. This potentially homologous region was expanded in both directions by 2000 bp. Nucleotide sequences of the query ORF and selection of target homologous region were then aligned using a pairwise global alignment algorithm (Myers & Miller, 1988), and the resultant matched region in the subject contig was extracted and saved as a homolog. Orthologs and paralogs were differentiated by reciprocal comparison.

**Phylogenetic tree construction**

A set of orthologous ORFs (327 total, 118 543 bp) showing > 70% similar to *N. meningitidis* MC58 (NC_003112) was selected as highly conserved proteins of the genus *Neisseria* and then aligned using the CLUSTALX (Thompson et al., 2002). The resultant multiple alignments were concatenated and then used to construct a genome tree using the neighbor-joining (Saitou & Nei, 1987) method implemented in MEGA program (Kumar et al., 2008). An evolutionary distance matrix for the neighbor-joining tree was generated according to the model of Jukes & Cantor (1969).

**Average nucleotide identity calculation**

The average nucleotide identity (ANI) was calculated using BLAST as previously described (Goris et al., 2007). In a given pair of genomes, the query genome is spliced into 1020-nt fragments and then blasted against the subject genome. The average of reciprocal results was represented as an ANI value.

**Results and discussion**

The genome sequences of strains LMG 5135<sup>T</sup> and ATCC 51223<sup>T</sup> were assembled into 46 and 40 contigs (> 1 kb long), respectively, and deposited into GenBank as accession numbers AFWQ0000000 and AFWR0000000, respectively. Each genome was 2.1 Mb in size (excluding the gaps) and had a G + C content of 49.0%. The genomic contents of the two *N. weaveri* strains were very similar, containing 2233 and 2099 predicted coding sequences (CDSs), respectively.

The genome tree based on the highly conserved orthologous ORFs showed that the two different *N. weaveri* species were closely related, forming a monophyletic clade within the radiation of *Neisseria* (Fig. 1). This phylogenetic closeness of the two species was also supported by the 16S rRNA gene tree (Supporting Information, Fig. S1), in which they have identical 16S rRNA gene sequences. The 16S rRNA gene sequence obtained from the genome sequence was albeit different (3/1488 nt) from the previously known PCR-derived sequence (L10738). The genomic relatedness of the two *N. weaveri* species was calculated by ANI (Konstantinidis & Tiedje, 2005). It is known that 94%–96% of the ANI between a pair of genome sequences may substitute for 70% of the DNA–DNA hybridization value (Konstantinidis & Tiedje, 2005; Goris et al., 2007; Richter & Rossello-Mora, 2009; Auch et al., 2010). The ANI between the two *N. weaveri* strains was 99.1%, clearly indicating that the two strains belong to the same species.
Despite the close relatedness of the two *N. weaveri* strains, they could be distinguished by several genetic elements. Compared with strain ATCC 51223, strain LMG 5135 contains one unique prophage region, one integrative element, and six nonhypothetical genes, but lacks five genes (Table S1). Compared with other *Neisseria* strains, both *N. weaveri* strains contain a unique prophage region, five unique integrative elements, and 21 unique nonhypothetical genes (Table S2).

Many putative virulence genes (Marri et al., 2010) and repeat elements (Parkhill et al., 2000; Snyder & Saunders, 2006; Snyder et al., 2007; Marri et al., 2010) were also detected in *N. weaveri* (Table 1), which are known to play key roles in *Neisseria* virulence and are exchanged via genetic transformation, gene expression, and genome rearrangements (Marri et al., 2010; Joseph et al., 2011). The number of DNA uptake sequences [DUS; function in DNA uptake/transformation (Goodman & Scocca, 1988;
Qvarnstrom & Swedberg, 2006]) and the number of virulence genes were also within the known range of the commensal Neisseria genome (Marri et al., 2010; Joseph et al., 2011). The absence of the Opa family [opacity outer membrane proteins for attachment, invasion, immune cell signaling, and inflammation (Dehio et al., 1998; Marri et al., 2010)] and certain iron scavenging genes (Marri et al., 2010) (Table S3) also reflect the genetic characteristics of N. weaveri as a member of the commensal Neisseria. However, the number of DUS1 was markedly lower in N. weaveri compared with other Neisseria strains from humans. In contrast to human commensal Neisseria, neither the drs3 element (Parkhill et al., 2000; Bentley et al., 2007) nor Correia elements [CR; (Correia et al., 1986; Snyder et al., 2009)], which function in gene regulation and sequence variation in pathogenic Neisseria, were detected in either of the N. weaveri genomes (Table 1). Instead, N. weaveri strains exclusively contain vapBC loci: a type II toxin-antitoxin system (Robson et al., 2009) in which vapC encodes a toxin (PilT N-terminus) and vapB encodes a matching antitoxin (Cooper et al., 2009). The absence of these loci in other Neisseria strains and the homology of these loci to genes in distantly related bacteria suggest that this toxin-related operon was acquired relatively recently via horizontal gene transfer. The overall pattern of virulence factors associated with Neisseria suggests that its pathogenicity may differ from other Neisseria.

On the basis of the high genomic relatedness (99.1% ANI value) and the identical 16S rRNA gene sequences discovered in this study, we propose that the two N. weaveri species should be united as a single species. On the basis of time of publication and established rules of nomenclatural priority (Lagage et al., 1992), we propose to reclassify N. weaveri Andersen et al. 1993 as a later heterotypic synonym of N. weaveri Holmes et al., 1993. The results of this study demonstrate the effectiveness of applying genome-based analyses to microbial taxonomy, including the reclassification of bacterial strains.

### Acknowledgements

We thank Dr J.P. Euzéby for his advice on nomenclature. This work was supported by Priority Research Centers Program (#2010-0094020) and a National Research Foundation grant (#2011-0016498) through the National Research Foundation of Korea, funded by the Ministry of Education, Science, and Technology, Republic of Korea.

### Statement

The GenBank accession numbers for the genome sequences of strains LMG 5135T and ATCC 51223T are AFWQ00000000 and AFWR00000000, respectively.

### References


Robson J, McKenzie JL, Cursons R, Cook GM & Arcus VL (2009) The vapBC operon from Mycobacterium smegmatis is...


**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Fig. S1.** 16S rRNA gene-based neighbor-joining tree showing the evolutionary relationships among 48 *Neisseria* strains.

**Table S1.** Genetic elements that distinguish strain LMG 5135 from strain ATCC 51223.

**Table S2.** Genetic elements that distinguish the *N. weaveri* genome from other *Neisseria* genomes.

**Table S3.** Composition of the iron utilization system in the *Neisseria* genomes.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.