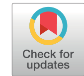
Veterinary Microbiology 218 (2018) 84–89

Contents lists available at ScienceDirect

Veterinary Microbiology

journal homepage: www.elsevier.com/locate/vetmic

Development of *Mycoplasma synoviae* (MS) core genome multilocus sequence typing (cgMLST) scheme



Mostafa Ghanem[a,b], Mohamed El-Gazzar[a,*]

[a] Department of Veterinary Preventive Medicine, College of Veterinary Medicine, The Ohio State University, Columbus, OH 43210, USA
[b] Faculty of Veterinary Medicine, Alexandria University, Rasheed El-Mahmoudeya, Markaz Rasheed, El Beheira Governorate, P.O. Box 22758, Egypt

ARTICLE INFO

Keywords:
Strain typing
cgMLST
MLST
Whole genome sequence
Mycoplasma Synoviae

ABSTRACT

Mycoplasma synoviae (MS) is a poultry pathogen with reported increased prevalence and virulence in recent years. MS strain identification is essential for prevention, control efforts and epidemiological outbreak investigations. Multiple multilocus based sequence typing schemes have been developed for MS, yet the resolution of these schemes could be limited for outbreak investigation. The cost of whole genome sequencing became close to that of sequencing the seven MLST targets; however, there is no standardized method for typing MS strains based on whole genome sequences. In this paper, we propose a core genome multilocus sequence typing (cgMLST) scheme as a standardized and reproducible method for typing MS based whole genome sequences. A diverse set of 25 MS whole genome sequences were used to identify 302 core genome genes as cgMLST targets (35.5% of MS genome) and 44 whole genome sequences of MS isolates from six countries in four continents were used for typing applying this scheme. cgMLST based phylogenetic trees displayed a high degree of agreement with core genome SNP based analysis and available epidemiological information. cgMLST allowed evaluation of two conventional MLST schemes of MS. The high discriminatory power of cgMLST allowed differentiation between samples of the same conventional MLST type. cgMLST represents a standardized, accurate, highly discriminatory, and reproducible method for differentiation between MS isolates. Like conventional MLST, it provides stable and expandable nomenclature, allowing for comparing and sharing the typing results between different laboratories worldwide.

1. Introduction

Mycoplasma synoviae (MS) is a significant poultry pathogen with worldwide distribution. Variability of infectivity, tissue tropism and pathogenicity between MS isolates have been reported previously (Kleven et al., 1975; Dijkman et al., 2016). Multiple sequence based typing methods have been developed; a single locus (*vlhA)* was the main target for MS sequence typing for many years with several modifications (Bencina et al., 2001; Hong et al., 2005; Wetzel et al., 2010; El-Gazzar et al., 2012). However, this method has multiple limitations (El-Gazzar et al., 2012, 2017). Three different MLST schemes were developed separately and coincidentally around the world. The first, (Cizelj et al., 2015), investigated the genotyping potential of 6 MS loci including housekeeping genes and highly polymorphic genes independently and collectively. The second, (Dijkman et al., 2016), identified 5 different core genes as a potential multilocus sequence typing targets. A third MLST scheme (El-Gazzar et al., 2017) was based on 7 housekeeping genes. In order to evaluate these multiple MLST schemes, a more robust genotyping scheme with high degree of reliability and discriminatory power needs to be developed. Recently, thanks to next generation sequencing, such typing schemes like cgMLST have been used in routine microbial diagnosis and epidemiological investigation of many human and animal pathogens (Mellmann et al., 2011; Bratcher et al., 2014; Kohl et al., 2014; Leopold et al., 2014; Schmid et al., 2014; Been et al., 2015; Ghanem et al., 2017). Despite increased accessibility to whole genome sequences, no standardized scheme for typing of MS based on whole genome sequences is currently available. Therefore, a core genome multilocus sequence-typing scheme (cgMLST) for *M. synoviae* was developed in this study.

Abbreviations: bps, base pairs; cgMLST, core genome multilocus sequence typing; DNA, deoxyribonucleic acid; MS, *Mycoplasma synoviae*; PBS, phosphate buffer saline; PCR, polymerase chain reaction; SNP, single nucleotide polymorphism; *vlhA*, variable lipoprotein and haemagglutinin A

* Corresponding author. Current address: Veterinary Diagnostic and Production Animal Medicine department Iowa State University, Ames, IA 50011, USA.
E-mail address: elgazzar@iastate.edu (M. El-Gazzar).

https://doi.org/10.1016/j.vetmic.2018.03.021
Received 26 October 2017; Received in revised form 16 March 2018; Accepted 17 March 2018
0378-1135/ © 2018 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

2. Materials and methods

2.1. MS strains, field isolates and DNA extraction

Forty MS isolates were genome sequenced, including laboratory-adapted strains and field isolates. Five of the isolates were collected in the UK between 1985 and 2012. One sample originated from each of the following countries Argentina, Canada, and Israel. Thirty-two isolates were from 11 different states within the US. These isolates were collected from Georgia [9 isolates], Arkansas [7 isolates], California [4 isolates], North Carolina [3 isolates], Alabama [2 isolates], Colorado [2 isolates], and one isolate from each of the following states Mississippi, Maine, Illinois, West Virginia, and Michigan. In addition to geographical diversity, this collection was temporally diverse as well, the oldest isolate was collected in 1955 and the most recent one collected in 2012. Most of the MS strains and field isolates were provided by the mycoplasma laboratory in the Poultry Diagnostic and Research Center, at the University of Georgia (Table S1) and were used previously for evaluation of the *vlhA* scheme in 2004 (Hong et al., 2005), and development of MS MLST (El-Gazzar et al., 2017). A total volume of 3 to 5 ml of modified Frey's color changed culture was centrifuged at 415 × *g* for 30 min. The pellets were suspended in 200 µl of PBS for genomic DNA extraction using the QIAamp DNA Mini Kit (QIAGEN, Valencia, CA) following the manufacturer's instructions. These isolates were confirmed to be MS using real time PCR (Raviv and Kleven, 2009)

2.2. Whole genome sequences from public databases

In this study, four Whole Genome Sequences (WGS) from public databases (GenBank and SRA databases) were used, three complete WGS and one WGS raw reads. All accession numbers are in Table S1.

2.3. Sequencing and assembly of Forty MS isolates

DNA extract from isolates was quantified using Qubit fluorometric analysis dsDNA HS (high sensitivity) Scheme Kit (Invitrogen, Carlsbad, CA). Paired-end libraries for next generation sequencing were prepared using the Illumina Nextera XT DNA library prep kit following manufacturer's protocols (Illumina, San Diego, CA). Sequencing was performed on Illumina MISeq at the molecular and cellular imaging center (MCIC) of the Ohio State University, Ohio Agricultural Research and Development Center (OARDC) in Wooster, Ohio, USA. SPAdes Genome Assembler (Bankevich et al., 2012) implemented within PATRIC; the pathosystems resource integration center (Wattam et al., 2013) was used for assembly of all MS raw reads.

The default preprocessing and read quality filtering were used. The quality of PATRIC generated assemblies were assessed and compared to each other using online quality assessment tool for genome assemblies (Quast) (http://quast.bioinf.spbau.ru/) (Gurevich et al., 2013).

2.4. MS-cgMLST development using SeqSphere+

Ridom SeqSphere+ version 4 (Jünemann et al., 2013) was used for MS-cgMLST development. Steps for cgMLST development using SeqSphere+ were described in previous studies (Mellmann et al., 2011; Kohl et al., 2014; Schmid et al., 2014; Antwerpen et al., 2015; Been et al., 2015; Ghanem et al., 2017). Briefly, twenty-five *M. synoviae* WGSs listed in Table S1 were used for scheme development. This collection was selected based on adhoc cgMLST analysis of 44 different *M. synoviae* WGSs to identify a group of diverse *M. synoviae* isolates that represent the entire diversity of MS population. One or more representative genomes from each cluster with 50 alleles or more different from the closest neighbor on a minimum spanning tree were selected. This collection included isolates from Canada, Israel, Argentina, Brazil, UK, and USA. Of these, the well characterized reference strain, MS2001-WVU1853-R (Accession number: GCA_000969765.1) was used as a reference and the remaining 24 WGS were used as query genomes. FASTA files of genome assemblies were loaded into SeqSphere+. Contigs/scaffolds of size ≥200 bps of the draft genomes were only included in the analysis. The cgMLST target genes identification involves two main steps; first step, MS2001-WVU1853-R reference genome filtration to exclude unfit gene targets for MLST typing. This step included the following filters: A) "homologous gene filter" to exclude all genes with high DNA similarity within a genome (with > 90% identity and > 100 bp overlap). B) "Start codon filter" to exclude all genes that devoid translation start codon at the beginning of the gene. C) "minimum length filter" to exclude all genes with length < 50 bp. D) "stop codon filter" to exclude all genes that are devoid of stop codon or have multiple stop codons, or a stop codon that is not located at the end of the gene. E) "gene overlap filter" the shortest of two overlapping genes by > 4 nucleotides is excluded according to this filter. Second step, query genomes pairwise comparison to select shared fit targets between the reference genome and 24 MS query genomes (core genome targets) using BLAST v2.2.12 (Altschul et al., 1990). All filtered genes of the reference genome that are shared with all query genomes with a sequence identity ≥90% and 100% overlap and passed the (default) SeqSphere+ settings formed the targets of final cgMLST scheme. The final selected genes were examined for their allele numbers and their nucleotide diversity percent.

2.5. MS-cgMLST target validation

Nineteen WGS labeled as E (Evaluation) under column name "Genome use" in Table S1 (in addition to the 25 WGS used for scheme development) were used for the purpose of target validation. This collection is used for validation of selected cgMLST targets.

2.6. cgMLST phylogenetic analysis of the 44 MS samples

To test the ability of the MS cgMLST scheme to differentiate the epidemiologically related samples from non-related samples, The 44 samples were loaded to SeqSphere+ and typed according to the 302-cgMLST gene targets.

2.7. Core genome SNP analysis for 44 MS samples

A core genome SNP analysis was performed by mapping the core genome genes in 44 MS isolates to identify single nucleotide variants from reference strain MS2001-WVU1853-R core genome genes and predict the phylogeny of these samples based on core genome SNPs. This analysis was performed using the tool "Find Nucleotide Variants" in SeqSphere+ with default filters (filter out Insertions/Deletions and filter out neighbor SNPs window within 10 bp). These filters were used to filter out any Insertion/Deletions (InDel) and neighboring SNPs within a window of 10 bases to eliminate any possible effect for genomic recombination induced SNPs. For each sample, a concatenated FASTA sequence containing the 302-cgMLST target gene sequences that were found in all 44 samples, in the order and orientation of the cgMLST target gene sequences of the reference strain MS2001-WVU1853-R, were mapped to the reference sequence. A table lists all variants, their position and the variant nucleotides in each sample was generated (Data set 1 S). Then, NJ tree (Fig. 2B) was generated based on the core genome SNP analysis after exclusion of targets with missing values. Two of the currently available MLST schemes (Dijkman et al., 2016; El-Gazzar et al., 2017) was used for evaluation of the study samples and was compared to the newly developed cgMLST see Supplemental file 1.

Download English Version:

https://daneshyari.com/en/article/8505442

Download Persian Version:

https://daneshyari.com/article/8505442

Daneshyari.com