EBioMedicine xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

Research Paper

Scarlet Fever Epidemic in China Caused by *Streptococcus pyogenes* Serotype M12: Epidemiologic and Molecular Analysis

Yuanhai You [a,1], Mark R. Davies [b,1], Melinda Protani [c], Liam McIntyre [b], Mark J. Walker [d,*], Jianzhong Zhang [a,*]

[a] State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China
[b] Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Melbourne, Victoria, Australia
[c] School of Public Health, The University of Queensland, Brisbane, QLD, Australia
[d] School of Chemistry and Molecular Biosciences, Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, QLD, Australia

ARTICLE INFO

Article history:
Received 14 December 2017
Received in revised form 10 January 2018
Accepted 10 January 2018
Available online xxxx

Keywords:
Scarlet fever
Group A *Streptococcus*
Mainland China
Epidemiological data
Genomic evolution

ABSTRACT

From 2011, Hong Kong and mainland China have witnessed a sharp increase in reported cases, with subsequent reports of epidemic scarlet fever in North Asia and the United Kingdom. Here we examine epidemiological data and investigate the genomic context of the predominantly serotype M12 *Streptococcus pyogenes* scarlet fever isolates from mainland China. Incident case data was obtained from the Chinese Nationwide Notifiable Infectious Diseases Reporting Information System. The relative risk of scarlet fever in recent outbreak years 2011–2016 was calculated using the median age-standardised incidence rate, compared to years 2003–2010 prior this outbreak. Whole genome sequencing was performed on 32 *emm12* scarlet fever isolates and 13 *emm12* non-scarlet fever isolates collected from different geographic regions of China, and compared with 203 published *emm12 S. pyogenes* genomes predominantly from scarlet fever outbreaks in Hong Kong (n = 134) and the United Kingdom (n = 63). We found during the outbreak period (2011–2016), the median age-standardised incidence in China was 4.14/100,000 (95% confidence interval (CI) 4.11-4.18), 2.62-fold higher (95% CI 2.57-2.66) than that of 1.58/100,000 (95% CI 1.56-1.61) during the baseline period prior to the outbreak (2003 − 2010). Highest incidence was reported for children 5 years of age (80.5/100,000). Streptococcal toxin encoding prophage φHKU.vir and φHKU.ssa in addition to the macrolide and tetracycline resistant ICE-*emm*12 and ICE-HKU397 elements were found amongst mainland China multi-clonal *emm12* isolates suggesting a role in selection and expansion of scarlet fever lineages in China. Global dissemination of toxin encoded prophage has played a role in the expansion of scarlet fever *emm12* clones. These findings emphasize the role of comprehensive surveillance approaches for monitoring of epidemic human disease.

1. Introduction

Scarlet fever ranked as one of the most severe infectious diseases prior to the widespread use of antibiotics in the 1940s. Scarlet fever is caused by the Gram positive bacterium *Streptococcus pyogenes* (group A *Streptococcus*, GAS) which is also responsible for several other diseases including suppurative pharyngitis and tonsillitis, impetigo and erysipelas, cellulitis, toxic shock and necrotizing fasciitis. Repeated infection also triggers the autoimmune sequelae rheumatic fever, rheumatic heart disease and acute post-streptococcal glomerulonephritis (Walker et al., 2014). Scarlet fever outbreaks began in 2011 in mainland China, Hong Kong, Vietnam and South Korea. *S. pyogenes emm12* is most frequently isolated from scarlet fever cases in China (Luk et al., 2012; Park et al., 2017; Tse et al., 2012; Yang et al., 2013; You et al., 2013; Wong and Yuen, 2012; Wu et al., 2013). In 2014, a scarlet fever epidemic has been reported in the United Kingdom caused by *S. pyogenes* isolates of *emm3*, *emm4*, *emm1* and *emm12* (Basetti et al., 2017; Chalker et al., 2017; Lamagni et al., 2017; Turner et al., 2016).

Several factors have been proposed to play a role in triggering these outbreaks, including changing bacterial population structure, enhanced capacity of GAS to cause scarlet fever through gene acquisition, changes in host herd immunity, meteorological factors and potential association between the bacteria and co-infection with an as yet unidentified factor that may predispose the host to scarlet fever caused by GAS. The underlying cause(s) of these outbreaks remain unresolved (Ben Zakour et al., 2015; Chalker et al., 2017; Coleman, 2016; Davies et al., 2015; Duan et al., 2017; Lee et al., 2017; Soderholm et al., 2017; Tse et al., 2012). Previous studies have shown that the predominantly *emm12* GAS isolates responsible for the Hong Kong outbreak have acquired mobile genetic

* Corresponding authors.
E-mail addresses: mark.walker@uq.edu.au (M.J. Walker), zhangjianzhong@icdc.cn (J. Zhang).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.ebiom.2018.01.010

elements including integrative and conjugative elements encoding tetracycline and macrolide antibiotic resistance and prophage encoding superantigens SSA, SpeC and the DNase Spd1 (Tse et al., 2012; Davies et al., 2015). Related mobile genetic elements were also identified in *emm1* GAS causing scarlet fever in Hong Kong (Ben Zakour et al., 2015).

Despite the epidemic lasting over 6 years, there has not been a comprehensive nationwide epidemiological description at the epicenter of the North Asia outbreak. In this study, we provide an up to date epidemiological characterization of the scarlet fever outbreak in China. We have also collected both historical and outbreak-related *emm12* GAS isolates from distinct geographic regions of mainland China, and employed whole genome sequencing to gain insight into genomic evolution, virulence profile and antimicrobial resistance gene carriage of scarlet fever causing GAS.

2. Materials and Methods

2.1. Epidemiological Data Collection and Analysis

The epidemiological scarlet fever case data used in this study was obtained from the Chinese Nationwide Notifiable Infectious Diseases Reporting Information System (established in 2003). Population data was obtained from the National Bureau of Statistics of China (http://data.stats.gov.cn – accessed 20.12.17). The age-specific incidence of scarlet fever for mainland China was extracted from this surveillance system by selecting years from 2003 to 2016. Direct age-standardisation was performed using the age-specific incidence rates (0–14, 15–64, 65 +) of scarlet fever for each year, standardised to the most recent (2015) population data. 95% confidence intervals were calculated using the exact method. Age-adjusted risk ratios (RR) and 95% confidence intervals (CI) were calculated by dividing the age-standardised median incidence rates to examine relative differences of incidence of scarlet fever in 2011–2016, compared to 2003–2010. For the analysis of scarlet fever spatial distribution, we focus on data reported after 2011. The geographic incidence data were extracted from 2011 to 2016. The age and occupation data for scarlet fever incidence were extracted and shown for the year 2016, after a comparison of this data was made with the data from 2011 to 2015; 2016 data was found representative for the six outbreak years. We collected all available data for scarlet fever incidence in China. After the new People's Republic of China was founded in 1949, this data was reported from 1950. We collected the incidence data from 1950 to 2002 from the Health Statistics Yearbook of the National Health and Family Planning Commission (http://www.nhfpc.gov.cn/zwgkzt/tjnj/list.shtml). The above spatial, temporal and population distribution data were analyzed using Excel and Stata 14.0. This manuscript was written in accordance with RECORD guidelines.

2.2. GAS Isolates From Mainland China

A national molecular epidemiologic investigation was performed from 2004 to 2011, to characterize *emm* types, virulence factors and antimicrobial resistance of strains circulating in both northern areas of China with high scarlet fever incidence and southern areas with low incidence. Throat swab samples were collected from patients clinically diagnosed as scarlet fever or pharyngitis. Scarlet fever cases were defined as patients who presented with fever (>38 °C), sore throat, 'sandpaper-like' rash on the trunk and limbs/extremities, and 'strawberry-like' tongue. In total, forty-five GAS *emm12* isolates from mainland China were investigated (Supplementary Table 1). These include 32 isolated by throat swab from diagnosed scarlet fever cases collected from areas with high incidence of scarlet fever in 2011 when the outbreak of scarlet fever first began. These areas included representative strains from Heilongjiang, Tianjin, Beijing and Shenyang. Eight historic isolates were also included for genomic analysis. We also included 139 publicly available *emm12* genomes reported from Hong Kong which is close to Guangdong province and here used to represent the south east of China, 1 from Lebanon (Ibrahim et al., 2016), 63 from the United Kingdom (Chalker et al., 2017), 2 from the USA and 3 from Australia (Davies et al., 2015). Collectively, a total of 248 *emm12* genomes were characterized in this study.

2.3. Genome Sequencing of emm12 Isolates

S. pyogenes isolates were cultivated on Columbia agar base supplemented with 5% sheep blood and incubated at 37 °C for 24 h. DNA was extracted using a Qiagen Mini kit. GAS *emm12* isolates were sequenced using Illumina Hiseq2000. Paired-end libraries with 500-bp insertions were generated, and the read lengths were 90 bp. For each isolate, 450 Mb of high-quality raw data was generated. Publicly available genome data were obtained from GenBank or the European Nucleotide Short Read Archive and shredded to an estimated 75× coverage of paired-end 100 bp reads using SAMtools wgsim for inclusion in phylogenetic comparisons.

2.4. Phylogenetic Analysis

Sequencing reads were mapped to the 1,908,100 bp Hong Kong *emm12* scarlet fever reference genome HKU16 (Tse et al., 2012) using the genome aligner SMALT v0.7.4 (http://www.sanger.ac.uk/resources/software/smalt/). The minimum base call quality to call a single nucleotide polymorphism (SNP) was set at 50, and the minimum mapping quality to call a SNP was set at 30 (Harris et al., 2010). Reads aligning to regions of the HKU16 genome pertaining to repeat sequences and prophage regions were excluded from phylogenetic analyses as previously defined (Davies et al., 2015). To remove further phylogenetic ambiguities, regions of high SNP density within the *emm12* population were identified with Gubbins (Croucher et al., 2015). Using the combination of these approaches, a total 321,164 bases (16.8%) of the HKU16 reference genome was excluded, resulting in an alignment of 1,586,936 bp. A total of 2637 SNPs was used to determine the phylogenetic structure of the global *emm12* population. Maximum-likelihood analysis of consensus alignments was conducted using the AVX version of RAxML v8.2.8 (Stamatakis, 2006). The general time-reversible model with gamma correction performed with 100 bootstrap random re-samplings to assess support for the maximum-likelihood phylogeny.

2.5. Assessment of Virulence, Antibiotic Resistance, and Mobile Genetic Elements Carriage

Carriage of antibiotic resistance genes and exotoxin superantigen genes was assessed using ABRicate (https://github.com/tseemann/abricate) at a cut-off of 75% of sequence identity from draft SPAdes v3.11.1 genome assemblies. Assessment of mobile genetic elements (MGE) repertoire was determined through a read mapping approach against a database of 17 *emm12* MGE (10 prophage sequences and 9 ICE elements) as identified from population analyses of Hong Kong scarlet fever *emm12* (Davies et al., 2015). Reads were mapped with BWA v0.7.16 and depth counted with Samtools depth v1.6 for bases with a Phred score > 20. MGEs were considered present if 90% of the sequence had 10× coverage.

2.6. Ethical Considerations

The National Institute for Communicable Disease Control and Prevention, China Center for Disease Control approved this study.

2.7. Data Access

Illumina sequence reads of 45 mainland China *emm12* GAS sequenced in this study were submitted to the Sequence Read Archive

Please cite this article as: You, Y., et al., Scarlet Fever Epidemic in China Caused by *Streptococcus pyogenes* Serotype M12: Epidemiologic and Molecular Analysis, EBioMedicine (2018), https://doi.org/10.1016/j.ebiom.2018.01.010