Distribution and clonal relationship of cell surface virulence genes among *Streptococcus pneumoniae* isolates in Japan

S. Imai[1], Y. Ito[1], T. Ishida[2], T. Hirai[1], I. Ito[1], K. Yoshimura[3], K. Maekawa[1], S. Takakura[4], A. Niimi[1], Y. Iinuma[5], S. Ichiyama[4], M. Mishima[1] and the Kansai Community Acquired Pneumococcal Pneumonia Study Group

1) *Department of Respiratory Medicine, Graduate School of Medicine, Kyoto University, Kyoto,* 2) *Kurashiki Central Hospital, Okayama,* 3) *Department of Clinical Trial Design and Management Translational Research Center, Kyoto University Hospital,* 4) *Department of Clinical Laboratory Medicine, Graduate School of Medicine, Kyoto University, Kyoto and* 5) *Department of Infectious Diseases, Kanazawa Medical University, Kanazawa, Japan*

Abstract

Streptococcus pneumoniae resides on mucosal surfaces in the nasopharynx, where selection for horizontal transfer of antimicrobial resistance genes and virulence factors may provide a survival advantage. We investigated the distribution of genes for pneumococcal cell surface proteins and their correlations with multilocus sequence typing (MLST), Pneumococcal Molecular Epidemiology Network (PMEN) clones and antimicrobial resistance, to identify pneumococcal virulence factors predicting prevalent clones from 156 pneumococcal isolates recovered from adult patients with community-acquired pneumonia in Japan. Pneumococcal *eno*, *pavA*, *piuA*, *cbpA* and *cbpG* were present in all isolates, and *hyl* and *piaA* were distributed among the clinical isolates. In contrast, pneumococcal *rlrA*, *pclA*, *psrP*, *nanC* and *pspA* family 1-type genes were variably distributed and significantly associated with MLST (Wallace coefficients (W) were over 84%). Serotype was a weaker predictor of sequence type (W, 0.75) than vice versa (W, 0.97). A multiple logistic regression analysis adjusted to the presence of virulence genes, *pspA* family 1 genes and carriage serotypes revealed that *pclA* and *rlrA* correlated with PMEN clones and antimicrobial resistance, and are likely to contribute to the selection of prevalent clones.

Keywords: Carriage serotypes, multilocus sequencing typing, *pclA*, *rlrA*, *Streptococcus pneumoniae*
Original Submission: 4 August 2010; **Revised Submission:** 20 November 2010; **Accepted:** 6 December 2010
Editor: J.-L. Mainardi
Article published online: 11 December 2010
Clin Microbiol Infect 2011; **17:** 1409–1414
10.1111/j.1469-0691.2010.03446.x

Corresponding author: Y. Ito, Department of Respiratory Medicine, Graduate School of Medicine, Kyoto University, 54, Kawahara, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
E-mail: yutaka@kuhp.kyoto-u.ac.jp

Introduction

Streptococcus pneumoniae resides on mucosal surfaces in the nasopharynx, where selection for horizontal transfer of antimicrobial resistance genes from other pneumococci and commensal organisms occurs. Additionally, various cell surface virulence factors provide a survival advantage [1].

Capsular polysaccharide prevents entrapment in the nasal mucous, allowing access to epithelial surfaces [1]. The available epidemiological data have shown that certain serotypes are nasopharyngeal carriage isolates in children using serotype 14 as a reference [2]. The *rlrA* islet-encoding pili is important for adhesion, and its presence correlates with certain prevalent clonal types, including several Pneumococcal Molecular Epidemiology Network (PMEN) clones with wide geographical distribution and antimicrobial resistance or importance in disease (http://www.sph.emory.edu/PMEN/pmen_criteria.html) and clones associated with antimicrobial resistance [3,4]. However, less prevalent clones also had carriage serotypes, and not all of the PMEN clones carried *rlrA* [3,5].

As candidates for cell surface proteins, choline-binding protein A (CbpA) [6,7], neuraminidase A (NanA) [8], pneumococcal collagen-like protein A (PclA) [9], pneumococcal adhesin and virulence A (PavA) [10], enolase (Eno) [11], pneumococcal serine-rich repeat protein (PsrP) [12] and hyaluronidase (Hyl) [13] contributed to adherence or invasion of host cells. The cell-surface lipoprotein pneumococcal

surface antigen A (PsaA) binds E-cadherin [14] and is a component of an ATP-binding cassette (ABC) transport system that has specificity for manganese [15]. Additional lipoproteins, pneumococcal iron acquisition (PiaA) and pneumococcal iron uptake A (PiuA) are also components of an ABC transport system for iron uptake [16]. Among *cbpA* and members of the same family sharing a choline-binding domain (*cbpD*, *cbpE* and *cbpG*), *cbpD* and *cbpE* were universally distributed in *S. pneumoniae* isolates [6]. Two neuraminidases genes, *nanA* and *nanB*, were present in all strains and in 96% of strains, respectively, while the *nanB* homologue, *nanC*, was found in 51% of strains [8].

In our previous report, potential carriage serotypes, the presence of *rlrA* and multidrug resistance were significantly associated with PMEN clonal complexes (CCs) [5]. Thus, we investigated the distribution of genes for other pneumococcal cell surface proteins and their correlations with multilocus sequence typing (MLST), PMEN clones and antimicrobial resistance to identify pneumococcal virulence factors predicting prevalent clones in this study.

Materials and Methods

Bacterial strains and MLST

A total of 156 pneumococcal isolates that had been prospectively collected from patients over 15 years of age who were diagnosed with community-acquired pneumonia between May 2003 and February 2005 in Japan were previously described [17–19]. We ran eBURST with default settings, associating each sequence type (ST) with a CC on the entire MLST database (4844 STs) and newly assigned STs within our dataset. In this work, we named CCs according to the ST number of the eBURST-predicted founder, defined as the ST with the greatest number of single locus variant, a smaller number of STs in a group consisting of two different STs or a singleton itself (http://spneumoniae.mlst.net/eburst/). In our previous study, 141 isolates of the cohort were analysed by MLST, showing that 49 STs, 27 CCs and three singletons were included [5]. Furthermore, MLST and CCs were determined for the remaining 15 isolates to increase strain collection. We previously reported that non-susceptibility to penicillin and erythromycin was recorded in 72 (46.2%) and 124 (81.4%) of the isolates, respectively [18], and we determined non-susceptibility for trimethoprim-sulfamethoxazole (113 isolates, 72.4%) in this study. According to meta-analysis data from children [2], we defined serotypes 3, 6A, 6B, 15, 19A, 19F, 23A and 23F as carriage serotypes, and found that 107 isolates (68.6%) were carriage serotypes [18].

Detection of the genes for pneumococcal cell surface proteins

The genomic location of the *rlrA* islet and the presence of *pspA* family 1 genes were determined by PCR as previously reported in 141 isolates [5,17]. We investigated the presence of *eno*, *hyl*, *pavA*, *piaA*, *piuA*, *cbpA*, *cbpG*, *nanC*, *pclA* and *psrP* using PCR. Because the primer sets for *cbpA*, *cbpG* and *nanC* in previous studies included a diversity region or mismatch compared with the genes in the complete genomes of the TIGER4, R6, D39, G54, CGSP14, Hungary 19A-6, Taiwan 19F-14, P1031, JJA, 70585 and ATCC700669 strains, we designed novel primers. The presence of *cbpA* and *cbpG* was determined using two primer sets for each. One primer set for *cbpA* and *cbpG* (*cbpA* outer and *cbpG* outer) matched conserved regions flanking the genes. When the genes were absent, a lower fragment size was detectable. The other sets for *cbpA* (*cbpA* inner) matched the 5′-flanking region and a region inside the gene, and those for *cbpG* (*cbpG* inner) matched regions inside the gene. The PCR fragments were detectable only when the genes were present. Primer sets for *nanC* matched conserved regions encoding the gene product (Table 1).

Statistical analysis

Statistical analysis was performed using JMP version 6.03 (SAS Campus Drive, Cary, NC, USA). We used Wallace coefficients (W) and 95% confidence intervals (CIs) to estimate the power of ST and CC in predicting gene presence or absence and carriage serotypes, as well as the power of ST in predicting serotype and vice versa. A W of 1.0 is a consequence of the universal or quasi-universal presence of the genes. Wi is defined as the expected value of W when both classifications are independent [20] (http://www.comparingpartitions.info). Association of pneumococcal virulence factors with the PMEN clones or non-susceptibility to antimicrobials was tested by Pearson chi-square or Fisher exact test when appropriate. Odds ratios (ORs) and the respective 95% CIs were computed as estimates of relative risk. Variables were selected for multiple logistic regression analysis if probability values were <0.05 by univariate analysis. Probability values of 0.05 or less were considered statistically significant.

Results

Prevalence of pneumococcal virulence factors

Among the additional 15 isolates, we identified one new allele sequence (*spi*225), two new STs (one contained new allelic profiles of the known alleles (ST4671) and one