Molecular characterization of clinical isolates of M non-typable group A streptococci from invasive disease cases

Asha Tanna,1 Michaela Emery,1 Chenchal Dhami,1 Eve Arnold2 and Androulla Efstratiou1

Respiratory and Systemic Infection Laboratory1, and Statistics, Modelling and Bioinformatics Department2, Health Protection Agency Centre for Infections, 61 Colindale Avenue, London NW9 5HT, UK

Currently there are 93 validated M serotypes of Streptococcus pyogenes, Lancefield group A streptococcus (GAS), and >130 emm genotypes. A marked increase in the number of non-typable GAS isolates (2 % in 2000, 4 % in 2001 and 9 % in 2002) from invasive disease cases referred to the authors’ reference laboratory was noted during 2000–2002. A total of 217 (92 %) were from blood cultures, 14 (6 %) from deep abscesses and five (2 %) from aspirates. The clinical manifestations included bacteraemia, septicaemia, cellulitis, meningitis, necrotizing fasciitis and toxic-shock syndrome. In order to establish whether this increase was due to the emergence of novel types or the unavailability of M-typing sera, these isolates were subjected to emm sequencing. A total of 144 isolates (61 %) belonged to M types for which sera were no longer available; 112 (48 %) belonged to higher M types, including emm83.1 (9 %), emm94 (8 %) emm87 (6 %) and emm89 (6 %); and 32 (13 %) belonged to lower M types that were not commonly isolated in the UK, and included M25, M43, M49, M64, M73 and M74. Sixty-six (28 %) of the isolates belonged to newly designated emm types. Other isolates belonged to the novel emm types st2147, STNS1033 and st854, recently registered in the Centers for Disease Control (CDC) database by other laboratories. One novel emm type, st2161, was isolated from an injecting drug user. There were differences in the type distribution of these isolates according to geographic location. However, 90 % of emm93, one of seven predominant emm types identified amongst the collection of M non-typable (MNT) isolates, were isolated from the London region.

INTRODUCTION

The emm gene of Streptococcus pyogenes, Lancefield group A streptococcus (GAS), encodes a major virulence factor for these important pathogens, the M protein. The hypervariability of surface-exposed amino termini of M proteins provides the basis for identifying the different types (Beachey et al., 1981; Fischetti, 1989). Identification of clinical isolates of GAS for surveillance and other epidemiological studies has relied primarily upon serological typing of the surface T and M proteins with polyclonal sera (Colman et al., 1993; Maxted & Widdowson, 1972). Even with the availability of sera for known M types and provisional types, many strains that are now encountered have other M proteins with new serological specificities (Beall et al., 1997). There are 93 internationally recognized M types, each of which is often associated with specific T-antigen patterns (Colman et al., 1993; Facklam & Edwards, 1979; Johnson & Kaplan, 1995; Moody et al., 1996). However, it is sometimes difficult to detect M proteins serologically, because typing sera are not widely available, and are difficult and expensive to prepare. Many GAS isolates are also non-typable, probably because of the lack of M protein expression, the lack of reactivity of expressed M protein with available antisera, or the expression of a previously uncharacterized M protein.

GAS isolates that are deemed non-typable by serological methods can be genotyped through emm sequence determination. emm typing relies upon the amplification of the hypervariable sequence encoding M serospecificity. The usefulness of emm gene sequence analysis combined with T-antigen typing and OF phenotyping of GAS has been demonstrated (Beall et al., 1996, 1997). The emm-typing system is a useful and reliable epidemiological tool for subdividing GAS because it is independent of M protein expression and can often discriminate between GAS isolates that may be only weakly antigenic or non-typable; thus, emm sequence typing has the potential to classify isolates
that have been difficult to type by serological methods (Facklam et al., 1999).

A statistically significant increase \((P<0.0001)\) amongst 236 M non-typable (MNT) isolates from invasive disease cases referred to our reference laboratory was observed during 2000–2002, from 2 % in year 2000 to 9 % in year 2002.

The purpose of this study was to establish whether the increase in M non-typability amongst isolate referrals to the Streptococcus and Diphtheria Reference Unit (SDRU) of the Health Protection Agency (HPA) was due to the presence of new types that would not be identified serologically with the current set of M antisera, or was due to the presence of types not previously observed within the UK.

**METHODS**

**Source of clinical isolates.** GAS sterile-site invasive isolates submitted from hospital laboratories from England and Wales during 2000–2002 to our reference laboratory were typed using conventional serological methods (Colman et al., 1993; Johnson, 1996). A total of 236 MNT isolates were subjected to emm sequencing. Of these, 217 (92 %) were from blood cultures, 14 (6 %) were from deep abscesses and five (2 %) were isolated from aspirates.

**Serological typing.** T, M and OF typing was undertaken and performed as described previously (Johnson, 1996).

emm sequencing. Cultures were grown in 20 ml Todd–Hewitt broth (Difco, BD) and incubated overnight at 37 °C. Cultures were diluted tenfold in Todd–Hewitt broth for DNA extraction. The MagNA Pure DNA Isolation Kit III (Roche Applied Science) was used for the extraction of DNA. The procedure described in the MagNA Pure manual was followed. PCR amplification of the emm gene was carried out using the following primers: forward primer 5'-TATT(CG)GCTTAGAAAATTAA-3' and reverse primer 5'-GAAATTCTCTCGTGGTGT-3'. The PCR cycling times used were 30 cycles of 94 °C for 1 min, 46 °C for 1 min, 72 °C for 2.5 min, 1 cycle of 72 °C for 7 min (J. Wotton, J. Moeller & C. Johnson, personal communication).

PCR products were detected by agarose gel electrophoresis, and the products were cleaned using MilliporePCR Clean Up plates (Millipore).

The sequencing reaction was carried out using the forward primer 5'-TATTCGTTAGAAATATATATGGGAC-3' with cycling times of 30 cycles at 96 °C for 20 s, 50 °C for 20 s, 60 °C for 4 min (J. Wotton, J. Moeller & C. Johnson, personal communication). The DNA was precipitated using the standard ethanol precipitation method (Beckman Coulter CEQ8000 Generic Analysis System, Dye Terminator Cycle Sequencing Chemistry protocol) and dried using a Nalge vacuum apparatus.

Sequencing was performed by the dyeoxynucleotide method using the Dye Terminator Cycle Sequencing with Quick Start kit (Beckman Coulter). Sequencing reactions were run on a CEQ 2000 capillary sequencer (Beckman Coulter).

Sequence analysis. Sequence analysis was performed by a BLAST search on the Centers for Disease Control (CDC) streptococcal emm sequence database (http://www.cdc.gov/ncidod/biotech/strep/assigning.htm) to designate emm sequence type. A sequence was considered to be a given emm gene allele (or sequence type) if it had ≥95 % identity over at least the first 350 bases with the corresponding emm gene in the CDC database (Beall et al., 1997; Facklam et al., 1999). A subtype according to the CDC scheme described at http://www.cdc.gov/ncidod/biotech/strep/assigning.htm is based solely upon the first 150 bases encoding the predicted first 50 amino acids of the expressed M protein.

**Data analysis.** The software programs SAS 8.2 and StatXact 6.03 were used for statistical analyses of the data, and chi square tests were used to test for association between type and other variables. If numbers of isolates received per type were small, an exact test was used; otherwise a Monte Carlo estimation of an exact \(P\) value was calculated. To examine a possible association between sequence type and age, ANOVA was applied and a Cochran–Armitage exact test was used to test for any trends.

**RESULTS AND DISCUSSION**

Amongst the 236 GAS MNT isolates sequenced, 213 (92 %) were from bacteremic patients with one or more of the following conditions: cellulitis, necrosis, pyrexia, sepsis, pneumonia, deep abscess and chest infection. Intravenous drug use was the major predisposing factor amongst 84 patients.

emm sequence typing

Sequencing of the 5' end of the emm gene is recommended as the molecular ‘gold standard’ for GAS typing, and has a 100 % typability rate. In order to achieve similar results with other molecular typing methods, such as capture enzyme immunoassay (EIA) (Saunders et al., 1997), in which specific probes are used, the isolates would have to be tested against a very large range of probes, including those for the subtypes. Although EIA is a very useful technique for screening isolates against the common types, it would not recognize any new emm types that may emerge.

A total of 144/236 isolates (61 %) belonged to emm types for which M sera were no longer available, of which 112/144 (78 %) belonged to higher M types and 32/144 (22 %) belonged to lower M types that are not commonly isolated in the UK. A proportion of isolates, (66/236, 28 %) belonged to recently designated emm types (Fig. 1, Table 1). In contrast to the study on emm typing conducted in Brazil (Teixeira et al., 2001), in which 13 new emm types were identified during 1995–1999, we identified only one novel type, emm st2861 (Fig. 2), which was isolated from the blood culture of an intravenous drug user (IDU). This highlights the diversity of emm types amongst different populations. Only 26/236 (11 %) of the isolates belonged to commonly seen lower M types, including emm1.9, emm2, emm2.1, emm6.2, emm6.10, emm11 and emm12. Therefore, the increase in M non-typability was attributed largely to the increase in higher M types for which M-typing sera are no longer available, and also to the isolation of recently designated emm types.

**Distribution of emm types**

Amongst the predominant emm types from the MNTs over the 3 years, a statistically significant increase \((P<0.0001)\) in
some of the newly designated \textit{emm} types, including \textit{emm}102.3 and \textit{emm}94, and previously uncommon M types, including \textit{emm}43.3, \textit{emm}68, \textit{emm}77, \textit{emm}82, \textit{emm}83.1, \textit{emm}87, \textit{emm}89, \textit{emm}92 and \textit{emm}93, was observed. Statistically significant differences ($P<0.001$) were also observed in the regional distribution of these \textit{emm} types. When we examined the regional distribution of the seven predominant \textit{emm} types amongst the MNT isolates (Fig. 3), 90\% of \textit{emm}93 isolates were received from the London region, and 39 and 28\% of type \textit{emm}94 were received from the Northern and Yorkshire regions and the North West region, respectively. \textit{emm}43.4 was predominant in the Northern and Yorkshire regions, and \textit{emm}102.3 in Scotland.

\textbf{Association of \textit{emm} types and T types}

The association of \textit{emm} types and T types amongst GAS isolates that were MNT was difficult to assess, as 17\% (41/236) were T non-typable due to autoagglutination. Amongst the remaining 195 isolates that were T typable, 85\% (165/195) correlated with the conventional M- and T-type association, whereas 15\% (30/195) of isolates showed atypical correlations (Table 2). The most significant atypical correlation was observed with \textit{emm}68 and T3/13/B3264 (10/11 isolates). Typically, M/\textit{emm}68 is associated with T1. Apart from one isolate, \textit{emm}87.2, all isolates that exhibited atypical T and \textit{emm}/M correlations belonged to higher M types.

\begin{table}[h]
\centering
\caption{Uncommon \textit{emm} types amongst MNT isolates in this study}
\begin{tabular}{|l|l|}
\hline
Established M types & \textit{emm} types for which M sera are not available \\
\textit{emm}64.1 & \textit{emm}105 \\
\textit{emm}73 & \textit{emm}108 \\
\textit{emm}73.1 & \textit{emm}109 \\
\textit{emm}73.2 & \textit{emm}112 \\
\textit{emm}74 & \textit{emm}115 \\
\textit{emm}97.1 & \textit{emm}118 \\
\textit{emm}97.2 & \textit{emm}120 \\
\textit{emm}98 & ST2147 \\
\textit{emm}99 & ST2854 \\
\textit{emm}100 & STNS1033 \\
\hline
\end{tabular}
\end{table}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{emm_sequence_type_distribution.png}
\caption{\textit{emm} sequence type distribution amongst GAS MNTs.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{emm_type_sequence.png}
\caption{DNA sequence of novel \textit{emm} type st2681.}
\end{figure}
types, which were not amongst the GAS commonly isolated in the UK. Similarly, other recent studies have shown atypical T and M/emm correlations (Beall et al., 1997, 1998; Tanaka et al., 2002). Moses et al. (2003) found 43% of such isolates amongst MNT in their study in Israel. These newly recognized atypical T and emm/M associations would be useful in current and future epidemiological studies.

**Correlation between emm type, disease and age**

We found no statistically significant correlation between disease and emm type. However, in the case of IDUs, emm94 accounted for 18%, emm83.1 for 15%, emm93 for 13% and emm43.3 for 8% of infections. These emm types are not widely distributed amongst the general community, which

---

**Table 2. Atypical correlations between emm type and T type amongst GAS referred in the current study**

<table>
<thead>
<tr>
<th>No. of strains</th>
<th>emm type</th>
<th>T type</th>
<th>Number of atypical correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>25.3</td>
<td>14</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>43.4</td>
<td>15/17/23/47/19</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>15/17/23/47/19</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>58</td>
<td>3/13/B3264</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>61</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>62</td>
<td>3/13/B3264</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>65</td>
<td>3/13/B3264</td>
<td>1</td>
</tr>
<tr>
<td>11*</td>
<td>68</td>
<td>3/13/B3264</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>68</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>73.2</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>82</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>82</td>
<td>3/13/B3264</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>83.1</td>
<td>8/25/IMP19</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>97</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>91</td>
<td>8/25/IMP19</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>14</td>
<td>1</td>
</tr>
</tbody>
</table>

*The most significant atypical correlation*
suggests close-contact transmission amongst IDUs as a high-risk group for GAS infection. Similarly, Swiss workers (Lechot et al., 2001) have found a predominance of the uncommon serotypes M11 and M25 in a study of epidemics and endemic disease in IDUs, and Bohlen et al., 2000 have reported an outbreak of severe soft-tissue infections caused by GAS serotype M25 amongst drug users. An increase in invasive disease caused by GAS amongst IDUs has been reported by several workers (Navarro et al., 1993; Bernaldo de Quiros et al., 1997; Efstratiou et al., 2003; Engler et al., 2004).

There is only weak evidence ($P=0.086$) for an association between sequence type and age, although nearly a third of $emm_{94}$ was isolated from the age group 20–30 years. This could be explained by the fact that 18% of the isolates from IDU patients were $emm_{94}$, and most of these patients belonged to the age group 20–30 years.

With the increase in international travel, and the importation and use of illegal drugs, it is important to monitor trends of predominant types in different countries. This will enable the recognition of emerging new patterns internationally, and also allow the recognition of novel types identified by laboratories in other countries.

**ACKNOWLEDGEMENTS**

We thank all the hospitals in the UK for referring GAS isolates to the HPA SDRU.

**REFERENCES**


