Genome Sequence of a Serotype M28 Strain of Group A *Streptococcus*: Potential New Insights into Puerperal Sepsis and Bacterial Disease Specificity

Nicole M. Green,1,2,a Shizhen Zhang,2,a Stephen F. Porcella,1,a Michal J. Nagiec,2 Kent D. Barbian,1 Stephen B. Beres,2 Rance B. LeFebvre,3 and James M. Musser 1,2

1Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Disease, National Institutes of Health, Hamilton, Montana; 2Center for Human Bacterial Pathogenesis Research, Department of Pathology, Baylor College of Medicine, Houston, Texas; 3Department of Pathology, Microbiology, and Immunology, University of California—Davis, Davis

Puerperal sepsis, a major cause of death of young women in Europe in the 1800s, was due predominantly to the gram-positive pathogen group A *Streptococcus*. Studies conducted during past decades have shown that serotype M28 strains are the major group A *Streptococcus* organisms responsible for many of these infections. To begin to increase our understanding of their enrichment in puerperal sepsis, we sequenced the genome of a genetically representative strain. This strain has genes encoding a novel array of prophage virulence factors, cell-surface proteins, and other molecules likely to contribute to host-pathogen interactions. Importantly, genes for 7 inferred extracellular proteins are encoded by a 37.4-kb foreign DNA element that is shared with group B *Streptococcus* and is present in all serotype M28 strains. Proteins encoded by the 37.4-kb element were expressed extracellularly and in human infections. Acquisition of foreign genes has helped create a disease-specialist clone of this pathogen.

Puerperal sepsis, also known as "childbed fever," killed as many as 1 in every 6 mothers who gave birth in some hospitals in Europe in the mid-1880s [1]. The study of puerperal sepsis has an unusually rich history. Clinical investigation conducted in Austria in the 1840s by Semmelweis determined that puerperal sepsis was transmitted to pregnant women in labor by attending doctors [2]. These studies led to widespread implementation of infection-control measures such as hand washing, thereby changing the course of medical history. Subsequent studies [3–7] resulted in the understanding that puerperal sepsis is frequently caused by the human pathogenic bacterium group A *Streptococcus* (GAS).

Strains of GAS are categorized into serotypes on the basis of variation in the M protein, a highly polymorphic cell-surface molecule that is antiphagocytic and, hence, is a virulence factor [8, 9]. More than 125 distinct M protein types and emm gene types are known [10]. Epidemiologic studies conducted over many decades have repeatedly identified significant associations between certain M protein types and human infections [11–14]. For example, serotype M28 GAS strains are overrepresented among cases of puerperal sepsis and neonatal GAS infections [11, 12, 15–18]. Type M28 strains also are common causes of other types of invasive infections and pharyngitis in many countries [11, 12, 15, 17–29].

The molecular mechanisms responsible for the enrichment of serotype M28 strains in these infections are not known because of a general lack of understanding of the pathogenic processes underlying intraspecies disease specificity. As a first step toward gaining new insight into the microbial factors that contribute to infection specificity in GAS, we sequenced the genome of a serotype M28 strain. Our findings suggest that the acquisition of genes encoding novel extracellular pro-
teins has helped create a disease specialist clone of GAS, thereby broadening the ecological niche of this pathogen and modifying characteristics of infection.

**MATERIALS AND METHODS**

**Genome sequencing.** The genome of *emm28* strain MGAS6180 was sequenced to a minimum consensus base quality of Q40 by methods used for several other bacteria, including GAS [30–33]. This strain was isolated from a case of invasive infection in Texas in the 1990s. This strain is genetically representative of common serotype M28 strains, as assessed by the *emm28* allele and the spectrum of prophage-associated virulence genes. The genome was tiled by polymerase chain reaction (PCR) after closure to validate the assembly. Open-reading frames (ORFs) were identified with proprietary software (Integrated Genomics), annotated, and analyzed with the ERGO bioinformatics suite [34]. Strain MGAS6180 has been deposited in the American Type Culture Collection (BAA-1064), and the genome sequence has been deposited in GenBank (accession number SP_MGAS6180 CP000056).

**Analysis of a novel foreign genetic element by PCR tiling.** PCR was used to screen diverse GAS strains for a 37.4-kb novel foreign genetic element, designated “region of difference 2” (RD2) (see Results), identified in the genome of strain MGAS6180 (table 1). The presence and genome context of RD2 was assessed by use of 12 pairs of PCR primers (table 2) that amplify a 40-kb region encompassing RD2 and flanking chromosomal genes in strain MGAS6180.

**Analysis of prophage elements in natural populations of serotype M28 GAS from different localities and diseases.** A 2-step PCR-based method was used to determine the presence and genome location of prophages and prophage-associated virulence factor genes, as described elsewhere [36–38]. A random sample of 114 strains cultured from patients with pharyngitis and invasive infections in Texas, Ontario, and Finland was analyzed. The combination of prophage-associated virulence factor genes was defined as the virulence factor gene profile.

**Overexpression and purification of recombinant proteins.** The gene segments encoding the inferred mature forms of novel extracellular proteins were cloned by standard techniques. Five segments of the gene encoding M28_Spy1325 also were cloned by standard techniques. All cloned gene segments were sequenced to rule out the presence of spurious mutations. The recombinant proteins were purified to apparent homogeneity.

**RESULTS**

**Overview of the Genome Sequence of Strain MGAS6180 and Comparison with Other Published GAS Genome Sequences**

The sequenced genome of strain MGAS6180 is a circular chromosome of 1,897,573 bp with a guanine and cytosine (GC) content of 38.4% (figure 1). The GC content of the genome is essentially identical to that of all other GAS strains whose genomes have been sequenced (range, 38.5%–38.7%) [33, 39–43]. There are 1903 predicted coding sequences, which make up 1,675,787 kb (88.3%) of the genome. The gene content and context of the “core” chromosome (i.e., the part of the genome that does not include prophage-like and obvious foreign elements) is very similar to those described for strains SF370 and MGAS5005 (serotype M1), MGAS10394 (serotype M6), MGAS8232 (serotype M18), and MGAS315 (serotype M3) and SSI-1 (serotype M3) [33, 39–42]. The genome contains genes that encode many proven or putative secreted virulence factors, including streptolysin O, streptolysin S, pyrogenic toxin superantigens (streptococcal pyrogenic exotoxin C [SpeC], SpeK, SpeJ, and SmeZ), a secreted phospholipase A₂ (SlaA), collagen-like proteins (ScIA and ScIB), and proteases (SpeB, Mac, and ScpA) [9, 44–46].

**Putative Prophages or Prophage-Like Elements**

Four regions of the genome contain prophage-like elements, or apparent remnants of prophage-like elements (figure 1), which vary in size from 11.8 to 46.3 kb (for ease of description and discussion, these elements will be referred to as “prophages,” with the understanding that none of them has been shown to be a bona fide prophage). The 4 prophages are integrated in the half of the chromosome distal to the origin of replication, similar to the great majority of GAS prophages (figure 2). Prophage element DNA makes up 114,704 bp (6.0%) of the chromosome, which is less than in strains SF370 (7.1%), MGAS8232 (10.8%), and MGAS315 and SSI-1 (12.4%) [45].

**Prophage Φ6180.1**

Prophage Φ6180.1 (46.2 kb) has contiguous genes encoding SpeC and an extracellular DNase (Spd1)

---

Table 1. Distribution of selected region of difference 2 (RD2) genes among various M serotypes of group A streptococcus (GAS).

<table>
<thead>
<tr>
<th>Strain</th>
<th>Region of Difference 2 (RD2) Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF370</td>
<td>M1</td>
</tr>
<tr>
<td>MGAS5005</td>
<td>M6</td>
</tr>
<tr>
<td>MGAS10394</td>
<td>M3</td>
</tr>
<tr>
<td>MGAS8232</td>
<td>M18</td>
</tr>
<tr>
<td>MGAS315</td>
<td>M3</td>
</tr>
<tr>
<td>SSI-1</td>
<td>M3</td>
</tr>
</tbody>
</table>

The table is available in its entirety in the online edition of the *Journal of Infectious Diseases.*

---

Table 2. Primers used for region of difference 2 (RD2) analysis.

<table>
<thead>
<tr>
<th>Primers</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td></td>
</tr>
<tr>
<td>JID 2005:192 (1 September) • 761</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Atlas of the chromosome of group A Streptococcus serotype M28 strain MGAS6180. Arrowheads in the outermost ring depict the position and orientation of all transposase genes (light blue, clockwise; orange, counterclockwise); Ori, origin of replication. The middle 2 rings show the position of the 6 RNA operons (black) and open-reading frames (orange, clockwise; blue, counterclockwise) identified in the genome. The next inner ring shows the skew in the percentage guanine and cytosine content. The innermost ring shows the location and name of the prophage elements (6180.1 and 6180.2), prophage remnants (6180.3 and 6180.4), and 2 regions of difference (RD1 and RD2) identified in the genome. RD2 encodes 7 extracellular proteins (see figure 5). SlaA, streptococcal phospholipase A2; Spd1, streptococcal phage DNase 1; SpeC, streptococcal pyrogenic exotoxin C; SpeK, streptococcal pyrogenic exotoxin K.

and is inserted at the predicted T12_rmt site, the same location reported for prophages F315.2 (serotype M3), F8232.3 (serotype M18), and F10394.3 (serotype M6). However, each of these 4 prophage elements (F6180.1, F315.2, F8232.3, and F10394.3) encodes different proven or putative extracellular virulence factors (figures 2 and 3).

**Prophage F6180.2.** Prophage F6180.2 (42.3 kb) is inserted at a site analogous to the location of F315.4 in serotype M3 strain MGAS315; like F315.4, it has contiguous genes encoding SpeK and the recently discovered SlaA. Prophages F6180.2 and F315.4 are closely related in overall gene content and nucleotide sequence (figure 3)—evidence that these prophages have shared a very recent common ancestor.

**Prophage remnant F6180.3.** Prophage remnant F6180.3 (14.3 kb) is closely similar in size and nucleotide sequence to F570.4 in strain SF370 and to F10394.8 in serotype M6 strain MGAS10394 (figure 3), and it is located at the analogous integration site. These 3 elements lack a gene encoding a proven or putative virulence factor [33, 39]. Together, these characteristics indicate that F10394.8, F370.4, and F6180.3 have undoubtedly shared a common ancestor.

**Prophage remnant F6180.4.** Prophage remnant F6180.4 (11.7 kb) has no close relationship, overall, with prophages found in previously sequenced GAS genomes. F6180.4 does not encode apparent virulence factors.

**Serum Opacity Factor (SOF) Gene Region**

SOF is an extracellular lipoproteinase that makes human serum opaque, binds to fibronectin and fibrinogen, and is a virulence factor in mice infected with serotype M2 GAS [47–49]. GAS strains historically have been divided into 2 groups (SOF-pos-
Figure 2. Schematic of the group A Streptococcus (GAS) core chromosome, prophage element insertion sites, and prophage element-encoded virulence factors. **Circle**, GAS chromosome; **Ori**, origin of replication. The prophage elements are indicated with triangles that are color-coded to match the GAS source strain. **Stacked triangles**, prophage elements inserted at the same site. Nos. in triangles represent the clockwise order of prophage elements in each GAS strain whose genome has been sequenced. The 6 RNA operons are shown as black bars on the chromosome.

- **msfA**, macrolide efflux A
- **R6**, novel extracellular protein
- **sda**, streptodornase-α
- **sdn**, streptodornase
- **slaA**, streptococcal phospholipase A
- **spd1**, streptococcal phage DNase 1
- **spd3**, streptococcal phage DNase 3
- **spd4**, streptococcal phage DNase 4 (mitogenic factor 4)
- **speA**, streptococcal pyrogenic exotoxin A (speA1, speA2, speA3, and speA4 variants)
- **speC**, streptococcal pyrogenic exotoxin C
- **speH**, streptococcal pyrogenic exotoxin H
- **speI**, streptococcal pyrogenic exotoxin I
- **speK**, streptococcal pyrogenic exotoxin K
- **speL**, streptococcal pyrogenic exotoxin L
- **speM**, streptococcal pyrogenic exotoxin M
- **ssa**, streptococcal superantigen

Positive and SOF-negative strains) on the basis of their ability to make human serum opaque. The sof gene (3018 bp) in strain MGAS6180 is located on an ~5.3-kb chromosomal segment between homologues of the spy2034 and spy2037 genes of serotype M1 strain SF370 [50]. This genome segment also includes a 1947-bp gene (sfbX) that encodes a fibronectin-binding protein. The location and orientation of the sof gene region in strain MGAS6180 is analogous to that described in emm12, emm49, emm75, emm87, emm92, and emm114 GAS strains [50].

Streptin Gene Cluster Region

Streptin is a type A1 lantibiotic produced by certain GAS strains, including type emm28 organisms [51, 52]. The genome of strain MGAS6180 has a 10-ORF locus (figure 4) that is closely similar in nucleotide sequence to an analogous region that has recently been characterized in a GAS strain of undescribed emm type [53]. In contrast to serotype M1 strain SF370, all 10 ORFs appear to be intact, which is consistent with the observation that emm28 strains express lantibiotic activity.

Two Nonprophage Regions with a Unique Sequence

The genome of strain MGAS6180 has 2 regions of >10 kb with unique, largely nonprophage DNA not found in other sequenced GAS genomes. We will refer to these chromosomal segments as “RD1” and “RD2.” Like the 4 prophages, RD1 and RD2 are integrated in the half of the chromosome distal to the origin of replication (figure 1).

**RD1.** RD1 (11.1 kb) is integrated into a uracil-methyltransferase tRNA and appears to be a chimeric element composed of genes that are related to those associated with plasmids and phages. This region does not encode apparent secreted or virulence-related proteins. RD1 is flanked by 8-bp direct repeat sequences (ACGTGATG), and the GC content of this element is 30.4%, well below the whole-genome value of 38.4%.
**RD2.** RD2 (37.4 kb) is integrated in tRNA-Thr, is flanked by 16-bp direct repeats, and has 7 ORFs predicted to encode proteins with gram-positive secretion signal sequences (figures 1 and 5). Importantly, RD2 is similar in overall gene composition and sequence to regions described in the genome of serotype III and V group B Streptococcus (GBS) (table 3) [54, 55], a pathogen that is the primary cause of sterile-site neonatal infections. The GC content of RD2 (35.1%) is considerably less than the whole-genome value of 38.35% and closely approximates that of GBS (35.6% and 35.7%) [54, 55].

Four of these proteins associated with RD2 (M28_Spy1306, M28_Spy1326, M28_Spy1325, and M28_Spy1336) have an LPXTG motif located at the carboxy terminus that covalently anchors many proteins in gram-positive pathogens to the cell wall. M28_Spy1336 is the R28 protein that has been studied previously in GAS and GBS [56–59], whereas the other 3 proteins with LPXTGE motifs have not been found previously in GAS. M28_Spy1306, an inferred 254-aa protein, is 98% identical to a protein of unknown function made by the 2 GBS strains whose genomes have been sequenced [54, 55]. The gene encoding this protein has a GC content of 40.1%, a value that exceeds that of the core genome of strain MGAS6180 (38.4%).

---

**Figure 3.** Relationships among group A Streptococcus (GAS) prophage elements. Prophage element sequences present in the sequenced GAS genomes were aligned with CLUSTALW (available at: http://inn-prot.weizmann.ac.il/software, and an unrooted tree was generated with the DRAWTREE application in PHYLIP (available at: http://evolution.genetics.washington.edu/phylip.html). Proven or putative virulence factors encoded by each prophage element are given. MetA, macrolide efflux A; R6, novel extracellular protein; Sda, streptodornase-α; Sdn, streptodornase; SlaA, streptococcal phospholipase A2; Spd1, streptococcal phage DNase 1; Spd3, streptococcal phage DNase 3; Spd4, streptococcal phage DNase 4 (mitogenic factor 4); SpeA, streptococcal pyrogenic exotoxin A (SpeA1, SpeA2, SpeA3, and SpeA4 variants); SpeC, streptococcal pyrogenic exotoxin C; SpeH, streptococcal pyrogenic exotoxin H; SpeI, streptococcal pyrogenic exotoxin I; SpeK, streptococcal pyrogenic exotoxin K; SpeL, streptococcal pyrogenic exotoxin L; SpeM, streptococcal pyrogenic exotoxin M; SSA, streptococcal superantigen.
Figure 4. Schematic of streptin locus open-reading frame (ORF) map. Shown are the 10 ORFs inferred to encode the 10 genes that constitute the streptin locus, plus flanking ORFs. The streptin locus ORFs were named on the basis of homology with previously described genes in other bacterial strains, including group A streptococcus. Nos. inside each ORF arrow indicate the number of amino acids present in the inferred protein encoded by each ORF. LSU, ribosomal large subunit.

Figure 5. Open-reading frame (ORF) map of the 37.4-kb region of difference element (RD2) identified in strain MGAS6180. MacVector was used to identify ORFs by use of ATG, TTG, and GTG as potential start codons. Blue, putative mobility/transfer ORFs; yellow, putative gene regulators; red, inferred extracellular proteins; green, inferred hypothetical proteins of unknown function.
Table 3. Inferred proteins encoded by region of difference 2 (RD2) orthologous to serotype III and V group B streptococcus.

The table is available in its entirety in the online edition of the Journal of Infectious Diseases.

the United States, Canada, and Finland for the presence of 6 of the genes encoding inferred extracellular proteins encoded by RD2 (table 1). All 95 strains tested were PCR positive for these 6 amplicons. To determine whether the entire 37.4-kb RD2 segment was present, 9 randomly chosen serotype M28 strains were studied by PCR tiling across the entire region. All 9 strains were PCR positive for these 12 amplicons, which indicates that all strains had RD2 and that it was located in the same chromosomal location (data not shown). Hence, RD2 is common in M28 strains of GAS.

Inasmuch as RD2 has characteristics that indicate it might be transferred horizontally, we next tested the hypothesis that this region was present in other GAS strains. A sample of 248 strains representing 76 non-emm28 types was studied (table 1). Sixty-four of 248 strains representing 14 emm types were PCR positive for ≥1 of the 6 genes associated with RD2 that encode inferred extracellular proteins. Furthermore, PCR tiling of a subset of these 64 strains revealed the entire RD2 region to be present in 14 strains representing types emm2, emm4, emm48, and emm77 (data not shown). In the case of emm53 and emm87, an RD2-like element was located at a different position in the genome than in the other emm types, although the exact location was not determined. Thus, although our survey was not exhaustive, we found that RD2 was associated with a distinct subset of GAS strains that includes emm types epidemiologically associated with maternal-fetal infections.

Diversity of Prophage-Associated Virulence Gene Profiles and Elements in Natural Populations of Serotype M28 GAS from Different Localities and Diseases

Strains of serotype M28 GAS have been reported to be genetically diverse [24, 25, 67]. Inasmuch as variation in prophage content and chromosomal site of integration are key contributors to genomic diversity among GAS of the same serotypes [45], we next tested the hypothesis that many distinct combinations of prophages would be present in M28 strains. A sample of 114 M28 strains cultured from patients with pharyngitis or invasive infections in the United States, Canada, and Finland was tested for 14 prophage-associated genes that encode extracellular proven or putative virulence factors. Consistent with the hypothesis, 21 distinct virulence gene profiles were identified among the 114 emm28 strains (table 4). The majority of strains (54%) had a profile that included only the speC and spd1 genes. The profile that included speC, spd1, and

Figure 6. Diagram of protein M28_Spy1325 and cloned fragments. The alanine-repeat (AR) region (amino acid residues 201–434) has 3 copies of a 78-aa residue motif that is alanine rich. The proline-repeat (PR) region has 3 copies of a 40-amino acid residue motif that is proline rich. Divergent and conserved terminology refers to comparison with analogous regions in the antigen I/II protein family that has been described in various oral streptococcal species. C, conserved region (amino acid residues 767–1319), CWA, cell-wall anchor region that contains the LPSTG motif (amino acid residues 1320–1324); D, divergent region (amino acid residues 435–646); NAD region, N-terminal region, alanine-repeat region, and divergent region (amino acid residues 60–646); PC, proline-rich region and conserved region (amino acid residues 647–1319); SP, secretion signal peptide (amino acid residues 1–39).
sdn was the second most abundant (11%), and the third most abundant profile included spec, spd1, slaA, and speK (8%). Strains with unique virulence factor gene profiles accounted for 11% of the 114 strains.

**Characterization of Novel Extracellular Proteins Encoded by RD2**

Given the importance of many extracellular proteins in GAS-host interactions [9, 46], we elected to clone the genes for inferred novel extracellular proteins encoded by RD2, purify the proteins, and assess evidence that some of them were produced in vivo. We were unable to clone M28_Spy1325 in its entirety, so 5 gene segments were cloned, and the recombinant proteins were purified to apparent homogeneity (figures 6 and 7). Sequence analysis of the amino terminus of each of the 8 recombinant proteins confirmed that the correct protein had been purified (data not shown).

Western immunoblotting was used to infer whether M28_Spy1332, M28_Spy1306, M28_Spy1326, and M28_Spy1325 were produced in vivo during human GAS infections. Convalescent (but not acute) serum samples obtained from patients with GAS infection were reactive to the proteins (figure 7), which indicates that these extracellular proteins were made in humans with GAS disease.

**DISCUSSION**

**Comparative genomics of human bacterial pathogens.** Comparative genome sequencing of medically important pathogens has become an area of considerable interest [68–70]. Six GAS genome sequences are available publicly, which makes this one of the more deeply sequenced pathogenic bacterial pathogens. One critical theme that has emerged is that each genome has genes encoding previously undescribed extracellular proteins that influence human-GAS interaction. For example, several novel exotoxin genes were identified in the genome sequences of serotype M1, M3, and M18 strains, and the genome sequence of the serotype M6 strain revealed a new extracellular cell-wall–anchored protein [33, 39–43]. Similarly, the serotype M28 strain that we sequenced had at least 6 genes for novel inferred extracellular proteins that may participate in host-pathogen interaction. Discovery of these many new accessory virulence genes in GAS provides an important motivation for ongoing, in-depth analysis of the population genomics of this pathogen. Inasmuch as there are >125 emm types of GAS, largely reflecting distinct clones, we believe that much remains to be learned about novel extracellular proteins used by GAS to cause human disease [10, 71].

**Strain variation in prophage-associated virulence gene content.** We identified 21 distinct combinations of prophage-associated virulence genes among the 114 emm28 strains studied (table 4). These results are consistent with data from serotype M1, M3, M6, and M18 strains that have shown that prophages are responsible for the majority of variation in gene

---

**Table 4.** Serotype M28 strains used for prophage-associated virulence gene profiling.

The table is available in its entirety in the online edition of the *Journal of Infectious Diseases.*

---

**Figure 7.** Analysis of extracellular proteins encoded by region of difference (RD) 2. A, SDS-PAGE gel of purified recombinant proteins. The purified recombinant proteins were analyzed by PAGE. B–G, Representative Western immunoblots obtained by use of serum obtained from humans infected with serotype M28 strains of group A streptococcus. A, acute-phase serum; C, convalescent-phase serum (matched serum pairs were obtained from each patient). Patients had either invasive infections or pharyngitis: r1306, recombinant M28_Spy1306; r1307, recombinant M28_Spy1307; r1308, recombinant M28_Spy1308; r1325-AR, recombinant M28_Spy1325-alanine repeat region; r1325-D, recombinant M28_Spy1325-divergent region; r1325-NAD, recombinant M28_Spy1325-NAD region; r1326, recombinant M28_Spy1326; r1332, recombinant M28_Spy1332.
content among strains of the same M protein serotype [33, 38, 40, 41, 45].

Potential new insights into puerperal sepsis: RD2, lateral gene transfer, and the molecular genetic basis of bacterial disease specificity. A key discovery of our research was the identification of a 37.4-kb DNA segment that encodes 7 inferred extracellular proteins, including 4 with an LPXTG carboxy terminal motif characteristic of cell-surface anchored proteins. This region was not present in previously sequenced GAS genomes. Importantly, RD2 is related to regions found in the genomes of GBS (table 3), a pathogen that is the primary cause of invasive infections in neonates [54, 55].

One of the extracellular proteins encoded by RD2 (the R28 protein) promotes the adhesion of GAS to human epithelial cells grown in vitro and confers protective immunity in a mouse model of invasive disease—these findings have led to speculation that the R28 protein participates in the pathogenesis of puerperal sepsis [56–59]. Our discovery of RD2 and the multiple extracellular proteins encoded by it potentially adds a substantial new perspective to our view of the molecular factors that may contribute to the enrichment of emm28 strains in neonatal-maternal, urogenital, and perineal infections. Although the R28 protein functions as an epithelial cell adhesin in vitro, there was no significant difference in mouse virulence between the wild-type M28 strain and an isogenic mutant in which the R28 gene was genetically inactivated [58, 59]. It will be important to dissect the contribution to GAS pathogenesis of each of the extracellular proteins encoded by RD2. Our findings also provide motivation for enhanced study of the homologous proteins produced by GBS. For example, it may be that these extracellular proteins contribute to enhanced colonization of the female urogenital tract.

Expression of prophage-encoded virulence genes can differ significantly as a function of variation in growth environment [72, 73]. This observation led to the proposal that GAS can respond to variable environments with condition-dependent expression of prophage-encoded virulence factors [72]. It is possible that the RD2 genes encoding extracellular proteins are contingency genes, analogous to prophage-encoded virulence genes. We hypothesize that production of the extracellular proteins encoded by genes in RD2 also is differentially regulated in response to changes in the host environment. For example, it may be that transcription of certain RD2 genes is stimulated by exposure to and growth in the human upper respiratory tract, whereas other genes are differentially expressed in response to growth in blood, the female urogenital tract, amniotic fluid, and other normally sterile sites. If the differential expression of RD2 genes occurs, it may contribute to the ability of emm28 strains to survive and thrive in an unusually broad array of anatomic niches for GAS, including the throat, female urogenital tract, and bowel [11, 12, 15–18, 26]. Studies are under way to test this hypothesis.

Genome sequencing and molecular population-genomic analysis of GAS strains have shown that prophages are the major source of strain-to-strain variation in gene content [45]. Our discovery that RD2 is distributed among strains of many distinct emm types that have not recently diverged from a single common ancestor suggests that the lateral transfer of RD2 is relatively frequent—that is, the element is fairly promiscuous. Aside from prophages, RD2 is the most widely distributed large, exogenous genetic element that has been described thus far in GAS. Importantly, with the exception of emm28 strains, the presence of RD2 was a variable trait among strains of other emm types that we studied. This observation raises the possibility that RD2 contributes to infection specificity within strains of the same emm type. For example, it is possible that RD2 is overrepresented among strains of the same emm type that cause maternal-neonatal invasive infections (vs. pharyngitis). Regardless, our findings suggest that the acquisition of foreign genes has led to broadening of the ecological niche of GAS, thereby creating a disease-specialist clone of this pathogen.

Acknowledgments

We thank A. Henion and A. Mora, for assistance with statistical analysis and graphics, respectively; A. McGeer, D. Low, S. Shulman, and J. Vuopio-Varkila, for providing group A Streptococcus strains; and N. P. Hoe and F. DeLeo, for critical reading of the manuscript.

References


33. Banks DJ, Porcella SF, Barbian KD, et al. Progress toward characteri-


