MAJOR ARTICLE

Polymorphisms in Regulator of Protease B (RopB) Alter Disease Phenotype and Strain Virulence of Serotype M3 Group A *Streptococcus*

Randall J. Olsen,[1] Daniel R. Laucirica,[1] M. Ebru Watkins,[1] Marsha L. Feske,[1,2] Jesus R. Garcia-Bustillos,[1,7] Chau Vu,[1,5] Concepcion Cantu,[1] Samuel A. Shelburne III,[3] Nahuel Fittipaldi,[1] Muthiah Kumaraswami,[1] Patrick R. Shea,[1] Anthony R. Flores,[1,4] Stephen B. Beres,[1] Maguerite Lovgren,[8] Gregory J. Tyrrell,[8] Androulla Efstratiou,[10] Donald E. Low,[9] Chris A. Van Beneden,[6] and James M. Musser[1]

[1]Center for Molecular and Translational Human Infectious Disease Research, The Methodist Hospital Research Institute, Department of Pathology and Laboratory Medicine, The Methodist Hospital, [2]School of Public Health, University of Texas Health Sciences Center, [3]Department of Infectious Diseases, MD Anderson Cancer Center, and [4]Department of Pediatrics, Texas Children's Hospital, Baylor College of Medicine, Houston; [5]Paul L Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso; [6]Respiratory Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia; [7]School of Biotechnology and Health, Instituto Tecnológico y de Estudios Superiores de Monterrey, Nuevo Leon, Mexico; [8]Provincial Laboratory for Public Health and University of Alberta, Edmonton, and [9]Ontario Agency for Health Protection and Promotion and University of Toronto, Canada; and [10]Health Protection Agency, Centre for Infections, London, United Kingdom

Whole-genome sequencing of serotype M3 group A streptococci (GAS) from oropharyngeal and invasive infections in Ontario recently showed that the gene encoding regulator of protease B (RopB) is highly polymorphic in this population. To test the hypothesis that *ropB* is under diversifying selective pressure among all serotype M3 GAS strains, we sequenced this gene in 1178 strains collected from different infection types, geographic regions, and time periods. The results confirmed our hypothesis and discovered a significant association between mutant *ropB* alleles, decreased activity of its major regulatory target SpeB, and pharyngitis. Additionally, isoallelic strains with *ropB* polymorphisms were significantly less virulent in a mouse model of necrotizing fasciitis. These studies provide a model strategy for applying whole-genome sequencing followed by deep single-gene sequencing to generate new insight to the rapid evolution and virulence regulation of human pathogens.

Group A *Streptococcus* (GAS) is a human-specific pathogen that causes infections ranging in severity from asymptomatic colonization and uncomplicated pharyngitis (''strep throat'') to life-threatening necrotizing fasciitis (''flesh-eating disease'') and pneumonia [1]. Despite decades of research, many aspects of the molecular basis for host-GAS interactions remain poorly understood. In particular, little information bearing on the ability of GAS to infect anatomically diverse sites or the evolution of GAS virulence during the course of human infection is available [1]. We have recently used an unbiased whole-genome sequencing strategy to investigate the relationship between GAS strain genotypes and human disease phenotypes among infections caused by serotype M3 strains in Ontario, Canada [2–5]. Serotype M3 strains are particularly interesting because they commonly cause both invasive and oropharyngeal infections, display epidemic behavior with rapid shifts in disease frequency, and are associated with a disproportionate risk of death compared with other GAS serotype strains [1, 2, 6].

Sequencing the genomes of ~180 serotype M3 GAS strains recovered from patients with well-described disease manifestations in Ontario has generated several

Received 8 March 2011; accepted 20 June 2011; electronically published 18 January 2012.

Correspondence: Randall J. Olsen, MD, PhD, Department of Pathology and Laboratory Medicine, The Methodist Hospital Research Institute, RIB6-414, Houston, TX 77030 (rjolsen@tmhs.org).

The Journal of Infectious Diseases 2012;204:1719–29

DOI: 10.1093/infdis/jir825

new leads for studying bacterial pathogenesis [2, 3, 7]. One noteworthy finding was an unexpectedly high frequency of polymorphisms in the regulator of protease B gene (*ropB*) among strains in this population. RopB positively regulates the expression of the gene encoding the secreted streptococcal cysteine protease B (SpeB), a potent extracellular protease virulence factor implicated in tissue destruction and bacterial dissemination in humans and animal models [7–11]. RopB also directly or indirectly regulates the expression of several other genes encoding proven or putative virulence factors [1, 12]. Furthermore, a truncating mutation in the *ropB* allele encoded by a single serotype M1 GAS strain was recently shown to significantly alter SpeB expression [12]. Herein, we tested the hypothesis that *ropB* is subject to diversifying selection in serotype M3 GAS and studied the effect of *ropB* polymorphisms on SpeB activity and strain virulence. The results confirmed our hypothesis and provide new understanding of the host-pathogen interactions that underlie human infections.

MATERIALS AND METHODS

Bacterial Strains

We studied 1178 serotype M3 GAS strains, including organisms from an ongoing 20-year study of invasive infections in Ontario (1990–2010) [2], recent outbreak of invasive infections in England (2008–2009) [13], ongoing 10-year United States Active Bacterial Core surveillance (ABCs) program for invasive infections (1998–2008, http://www.cdc.gov/abcs), pharyngitis strains from 6 public health and private diagnostic laboratories in Ontario (2002–2010) [3], invasive strains collected in the German Democratic Republic (1969–1990) [14, 15], and pharyngitis strains collected in Alberta, Canada (2010, strains courtesy of G. Tyrrell). The first 3 collections represent prospective, comprehensive, population-based studies. Genome-wide comparison of the Ontario pharyngitis and invasive strains results in superimposable phylogenetic trees with nearly identical topology [3], confirming that although they do not comprise a comprehensive collection, the Ontario pharyngitis strains are representative of the underlying GAS population. The latter 2 collections are convenience samples that are included because they expand the diversity of the disease types, geographic regions, and time points of the strains studied.

Gene Sequencing

GAS strains were grown overnight on Todd-Hewitt agar supplemented with 0.2% yeast extract, and genomic DNA was extracted by alkaline-boiling lysis [16]. The *ropB* gene allele was determined by Sanger sequencing using a 3730xL DNA analyzer (Applied Biosystems by Life Technologies), and the *emm*-type was confirmed (primers are listed in Table 1). Electropherograms were visually inspected using Sequencher4.7 (Gene Codes), *ropB* alleles were evaluated using MacVector11 (MacVector), and phylogeny was inferred using SplitsTree4 (University of Tubingen) software. Statistical analyses were performed using XLStat2010 software (Addinsoft) or Statistical Analysis Software (SAS Institute), with differences considered significant at $P < .05$ for all tests performed, and graphs were created using Prism4 (GraphPad Software).

Table 1. Primers Used in This Study

Primer	Sequence (5′–3′)
ropB-3′	TTGAAAAAATCGCCCTGGACT
ropB-5′	CATAACCGACTATCATCCGAAC
emm-1	TATTSGCTTAGAAAATTAA
emm-2	GCAAGTTCTTCAGCTTGTTT

Measurement of SpeB Expression and Protease Activity

SpeB expression and protease activity were evaluated by Western immunoblot and casein milk plate hydrolysis, respectively, as described elsewhere [7]. SpeB was semiquantitatively scored as 1 (present at near-wild-type levels), 0.5 (present at reduced levels), or 0 (absent expression/activity). SpeB was evaluated in every strain with a unique *ropB* allele; ≥2 strains for each allele that was identified in >1 strain, including 56 (∼50%) V7I strains and 9 (∼1%) wild-type strains and all 87 pharyngitis strains with a variant allele. Protease activity was confirmed in the isoallelic strains using a quantitative chromogenic azocasein hydrolysis assay as described elsewhere [17].

Modeling of *ropB* Polymorphisms

The structure of RopB was modeled as described elsewhere [18]. Functional domains were predicted by homology matching using the I-TASSER structure prediction server (http://zhanglab.ccmb.med.umich.edu/I-TASSER), and the RopB structure was visually inspected using the program PyMOL [19].

Mouse Virulence Studies Using *ropB* Isoallelic Strains

Strain MGAS10870 was recovered from a patient in Ontario with an invasive soft-tissue infection [2]. Its genome has been fully sequenced, is generally representative of serotype M3 strains that cause invasive infections, and has a wild-type allele for all major regulatory genes [2]. Isogenic mutant strains lacking the gene encoding either *ropB* (designated MGAS10870Δ*ropB*) or *speB* (designated MGAS10870Δ*speB*) were generated from wild-type strain MGAS10870 by insertional inactivation with a spectinomycin cassette, as described elsewhere [11, 18]. Isoallelic mutant strains with the *ropB* allele encoding the V7I (designated MGAS10870-V7I), R226Q (designated MGAS10870-R226Q) or V7I/R226Q (designated MGAS10870-V7I/R226Q) amino acid changes were also generated from wild-type strain MGAS10870, as described elsewhere [18]. Ten immunocompetent CD1 mice (Harlan Laboratories) were inoculated in the right hind limb muscle with 10^7 colony-forming units of each strain and

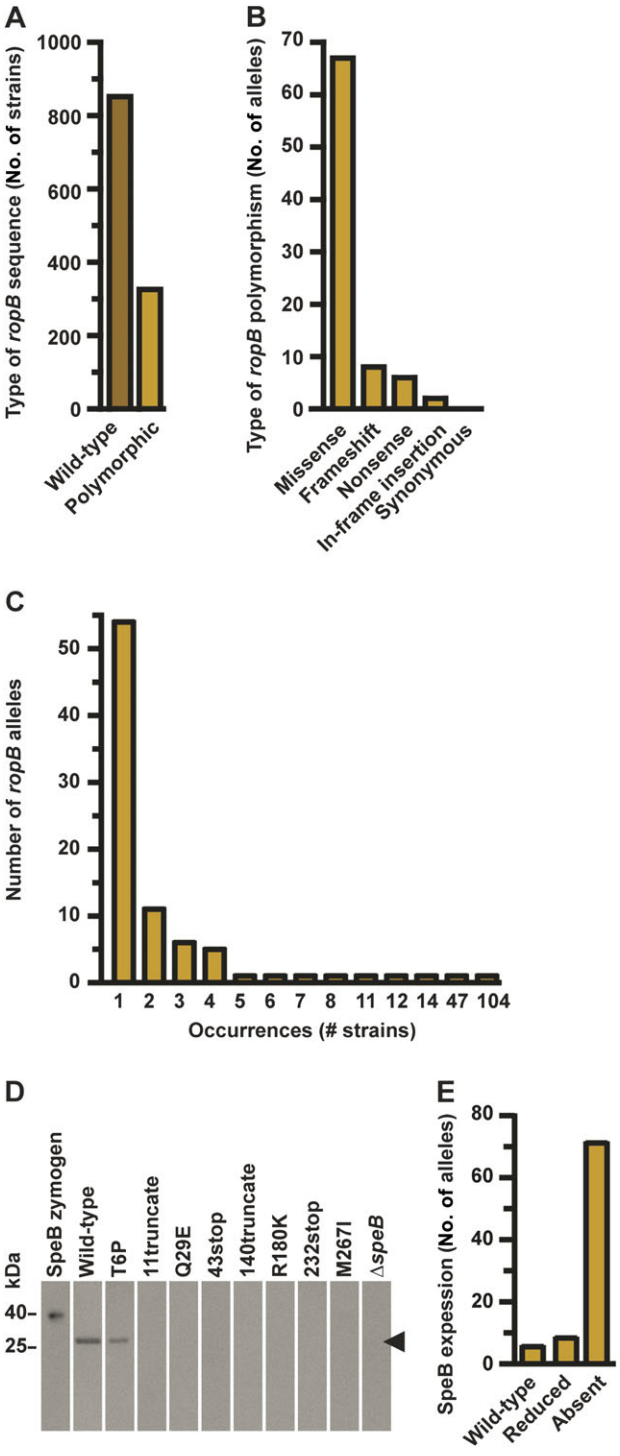

Figure 1. The *ropB* polymorphisms in serotype M3 group A streptococci (GAS) strains demonstrate a pattern of diversifying selection and significantly alter RopB regulatory activity. *A*, The *ropB* gene was sequenced in 1178 serotype M3 GAS strains collected from patients with different disease manifestations in diverse geographic regions and time periods. The numbers of strains with a wild-type (*dark brown bar*) or variant (*light brown bar*) *ropB* allele are shown. *B*, Gene sequencing identified 84 *ropB* alleles. Relative to the wild-type sequence, every polymorphism alters the RopB protein. The number of variant *ropB* alleles encoding each type of amino acid change is shown. *C*, Most of the 83 variant *ropB* alleles were present in only a few strains, with 54 alleles being specific to a single strain. *D*, RopB is the major regulator of secreted streptococcal cysteine protease B (SpeB), a proven virulence factor implicated in invasive disease. Most strains with a variant *ropB* allele demonstrated markedly reduced or absent SpeB expression as measured by Western immunoblot assay of culture supernatants. Representative strains are shown. Multiple immunoblots were performed simultaneously and processed identically, with lanes reordered such that *ropB* alleles were shown left to right with respect to the amino acid sequence. Recombinant SpeB zymogen and supernatant from a GAS strain with the gene encoding SpeB deleted (designated Δ*speB*) were used as a positive and negative control, respectively. *E*, Similarly, most strains with a variant *ropB* allele demonstrated markedly reduced or absent SpeB secreted protease activity, as measured by casein hydrolysis in milk plates.

monitored for near-mortality, as described elsewhere [7]. Visual and microscopic inspection of infected tissue was also performed. Lesions were excised and tissue was fixed in 10% phosphate-buffered formalin, decalcified, serially sectioned, and embedded in paraffin using automated standard instruments, as described elsewhere [7]. Hematoxylin-eosin– and Gram-stained sections were examined with a BX5 microscope and photographed with a DP70 camera (Olympus). Micrographs of tissue taken from the inoculation site that showed histopathology characteristic of each strain were selected for publication. The study was approved by the Institutional Animal Care and Use Committee of The Methodist Hospital Research Institute.

RESULTS

Polymorphism of *ropB* Among Serotype M3 GAS Strains

To test the hypothesis that *ropB* is highly polymorphic among all serotype M3 GAS, we sequenced the gene in 1178 strains from 6 collections that encompass a diverse array of infection types, geographic regions, and time periods. Among the 1178 strains studied (including strains from our whole-genome studies of serotype M3 GAS collected in Ontario [2, 3, 18]), 326 (28%) had a polymorphism in *ropB* (Figure 1*A*). In total, 84 distinct *ropB* alleles were identified, with the most common allele designated as the wild-type sequence (Table 2). Importantly, 64 of the 83 variant alleles have not been previously identified [2, 3, 18, 20]. Whereas only 1 of the previously identified *ropB* polymorphisms results in a protein alteration other than a single amino acid replacement, in this study we identified 17 polymorphisms that result in a major protein alteration by either premature termination or in-frame insertion (Figure 1*B*). We also identified 11 codons, which, compared with the wild-type sequence, were each affected by 2 distinct genetic changes resulting in different protein alterations in different strains (for example, independent R28K and R28G amino acid alterations occurred in 2 different strains) (Table 2). Additionally, we identified 5 alleles in which a single-nucleotide polymorphism

Table 2. *ropB* Alleles Identified in 1178 Serotype M3 GAS Strains

Allele No.	Codon	Nucleotide Change	Amino Acid Change	Occurrences, No.	Position in Codon	Strain Collections	SpeB Expression and Activity	Strains Tested for SpeB, No.[a]
1	Wild-type	None	None	852	…	All	Wild-type[b]	9
2	1	G3A	M1I	1	3	OI	Absent	1
3	1	T2A	M1K	1	2	US	Absent	1
4	6	A16C	T6P	1	1	OI	Reduced	1
5	7	G19A	V7I	104	1	OI, US, UK, OP	Reduced[c]	56
6	7, 61	G19A, T182C	V7I L61P	1	2	OP	Absent	1
7	7, 103	G19A, T307C	V7I S103P[d]	1	1	OP	Absent	1
8	7, 104	G19A, C311T	V7I T104I	1	2	OP	Absent	1
9	7, 226	G19A, G677A	V7I R226Q[d]	47	2	OI, OP	Absent[e]	28
10	7, 235	G19A, T704C	V7I I235T	1	2	US	Absent	1
11	11	G33-1	FS/truncation	1	1	US	Absent	1
12	22	G65A	C22Y	8	2	OI, US, OP	Absent	7
13	28	G83A	R28K	1	2	UK	Absent	1
14	28	A82G	R28G	1	1	GDR	Absent	1
15	29	C85G	Q29E	1	1	GDR	Absent	1
16	31	A92G	Y31C	1	2	US	Absent	1
17	33	C97 + 18	6 AA Insertion	4	In frame	US	Reduced	2
18	33	C97A	R33S	1	1	AP	Reduced	1
19	34	T100A	F34I	1	1	OI	Absent	1
20	43	C128A	Nonsense	1	2	OI	Absent	1
21	55	T164A	V55D	1	2	US	Absent	1
22	57	G169A	V57I	3	1	US	Wild-type	3
23	58	G172A	D58N	1	1	AP	Absent	1
24	62	T186 + 1	FS/truncation	7	1	AP	Absent[f]	7
25	63	T188C	I63T	12	2	US	Absent	2
26	68	A204C	K68N	2	3	US	Wild-type	1
27	72	G214T	Nonsense	1	1	US	Absent	1
28	73	T218G	F73C	1	2	US	Absent	1
29	75	G223A	D75N	1	1	US	Wild-type	1
30	76	G228A	M76I	1	3	US	Absent	1
31	80	A240C	K80N	14	3	US, OP	Absent	2
32	83	T248C	F83S	1	2	OI	Absent	1
33	85	G254A	C85Y	11	2	OI, US	Reduced	4
34	87	C261 + 1	FS/truncation	2	1	US	Absent	2
35	90	G268A	G90D	1	1	US	Absent	1
36	91	T272C	L91S	1	2	US	Absent	1
37	94	T281A	I94N	1	2	OI	Absent	1
38	100	G298 + 20	FS/truncation	1	2	US	Absent	1
39	103	T307C	S103P	5	1	OI, reference	Reduced	2
40	103	C308A	Nonsense	1	2	US	Absent	1
41	105	A314T	K105M	1	2	US	Absent	1
42	107	G321T	K107N	3	3	US	Wild-type	3
43	111	G331A	A111T	3	1	US	Absent	2
44	111	C332A	A111D	1	2	UK	Absent	1
45	134	C402-8	FS/truncation	1	1	US	Absent	1
46	136	C406A	L136I	1	1	US	Reduced	1
47	140	G418-1	FS/truncation	4	2	OP	Absent	4
48	140	G418 + 6	2 AA insertion	1	In frame	US	Absent	1
49	142	G425A	Nonsense	1	2	US	Absent	1
50	143	G428A	S143N	1	2	OI	Absent	1
51	150	T448A	F150I	1	3	US	Absent	1

Table 2 continued.

Allele No.	Codon	Nucleotide Change	Amino Acid Change	Occurrences, No.	Position in Codon	Strain Collections	SpeB Expression and Activity	Strains Tested for SpeB, No.[a]
52	151	A451T	N151Y	3	1	OI	Absent	2
53	151	T453A	N151K	1	3	US	Absent	1
54	152	A455T	N152I	1	3	US	Absent	1
55	154	G462A	M154I	1	3	OI	Absent	1
56	173	C517T	L173F	1	1	US	Absent	1
57	179	T536A	L179Q	1	2	US	Absent	1
58	180	G539A	R180K	1	2	GDR	Absent	1
59	184	T552A	N184K	2	3	OI	Absent	2
60	189	G567A	M189I	3	3	OI, US	Absent	2
61	196	T587G	L196W	2	2	UK	Absent	2
62	202	A605G	E202G	1	2	US	Absent	1
63	206	C617A	A206D	1	2	OI	Absent	1
64	206	C617-1	FS/truncation	1	3	UK	Absent	1
65	220	G658T	D220Y	3	1	UK	Absent	2
66	222	G665A	C222Y	4	2	OI, AP	Absent	4
67	224	T670C	Y224H	4	1	OI	Absent	2
68	226	G677A	R226Q	1	2	US	Absent	1
69	227	G680A	C227Y	1	2	OI	Absent	1
70	232	T696 + 1	Nonsense	2	1	OI, UK	Absent	1
71	232	T696-1	FS/truncation	6	1	US, GDR	Absent	2
72	237	G710T	G237V	1	2	OI	Absent	1
73	238	T713C	L238P	2	2	OP	Absent	2
74	245	G733A	A245T	2	1	OI, GDR	Absent	2
75	245	G733C	A245P	2	1	US	Absent	2
76	247	C739A	Nonsense	4	1	OI	Absent	2
77	249	G746A	C249Y	2	2	US	Absent	2
78	252	A754T	I252F	1	1	OI	Absent	1
79	252	C756-14	FS/truncation	2	1	US	Absent	2
80	256	T767C	F256S	1	2	US	Absent	1
81	260	A779C	N260T	1	2	UK	Wild-type	1
82	267	G801A	M267I	2	3	OI, US	Absent	2
83	268	T803G	F268C	1	2	OI	Absent	1
84	271	T811C	Y271H	1	1	AP	Reduced	1

Abbreviations: AA, amino acid; AP, Alberta pharyngitis strains; GDR, German Democratic Republic invasive strains; OI, Ontario invasive strains; OP, Ontario pharyngitis strains; SpeB, secreted streptococcal cysteine protease B; UK, United Kingdom invasive strains; US, United States Centers for Disease Control and Prevention invasive strains.

[a] Randomly selected invasive strains and all pharyngitis strains with a regulator of protease B gene (*ropB*) polymorphism were tested for SpeB.

[b] All strains tested secreted high levels of SpeB.

[c] Of 56 strains tested, 48 secreted SpeB (10 at near wild-type levels, 38 at reduced levels; 8 lacked detectable SpeB).

[d] The second single-nucleotide polymorphism occurring in the V7I genetic background of both these alleles also occurs alone.

[e] Of 28 strains tested, 4 secreted SpeB (1 at near wild-type levels, 3 at reduced levels; 24 lacked detectable SpeB).

[f] Of 7 strains tested, 1 secreted SpeB (at reduced levels; 6 lacked detectable SpeB).

(SNP) producing a V7I amino acid replacement occurred together with another SNP producing a second amino acid replacement (eg, the V7I and L61P amino acid alterations both occurred in the same strain) (Table 2). Of note, only 1 of these alleles with 2 changes relative to the wild-type sequence had been previously identified (V7I/R226Q) [2, 3, 18], and 2 of the involved amino acid alterations occurring together with the V7I replacement (R226Q and S103P) also occurred alone (Table 2).

Assuming that the background polymorphism rate across the genome of all serotype M3 GAS is similar to the rate calculated for Ontario strains used in our unbiased genome-wide

study [2], our data demonstrate that polymorphisms in *ropB* are significantly overrepresented among the 1178 strains studied herein (χ^2 test, $P < .001$, $\chi^2 = 939$, calculated using 83 polymorphisms observed and 6.27 polymorphisms expected if they were randomly distributed across the core serotype M3 genome [2]). The χ^2 statistic was also significant when calculated for *ropB* polymorphisms identified in each individual strain collection ($P < .001$ for each collection; $\chi^2 = 886$ for UK invasive strains, 615 for US invasive strains, 327 for Ontario invasive strains, 169 for Alberta pharyngitis strains, 115 for Ontario pharyngitis strains, and 47 for German invasive strains). Thus, the significant overabundance of *ropB* polymorphisms in serotype M3 GAS strains is not restricted to a single disease phenotype, geographic region, or time period.

Diversifying Selection and *ropB*

To test the hypothesis that *ropB* polymorphisms occur through diversifying selection, we evaluated the molecular consequence of each genetic change that defined the 83 variant alleles. Importantly, no synonymous (silent, not resulting in an amino acid replacement) polymorphisms were discovered (Figure 1*B*). That is, compared with the wild-type sequence, every allele encodes a protein with an altered RopB sequence. The nucleotide changes were significantly overrepresented in the first and second positions of the variant codons (Table 2), enabling us to reject the hypothesis of selective neutrality (χ^2 test, $P < .01$, $\chi^2 = 7.26$, calculated using 68 nucleotide changes observed in the first 2 positions and 54 nucleotide changes expected if the 81 polymorphisms were randomly distributed across the *ropB* coding sequence; note that codon positions could not be unambiguously assigned for the 2 in-frame insertions). Most *ropB* alleles occurred in a single strain or very few strains (Figure 1*C* and Table 2). Only 4 of the 83 variant alleles occurred in >10% of the strains tested (Table 2). Strains with these 4 *ropB* alleles usually had the same *emm3* allele and were present in a single location during a short time, suggesting that they share identity by descent rather than identity by independent evolutionary events (ie, convergence). Taken together, these genetic and epidemiologic data are consistent with a model of *ropB* evolution in which new alleles emerge through diversifying selection.

Effect of Amino Acid Alterations on Regulatory Activity of RopB

RopB is the major positive transcriptional regulator of secreted streptococcal cysteine protease B (SpeB) [1]. SpeB contributes to tissue destruction and mortality in animal models of necrotizing fasciitis and is believed to participate in the pathogenesis of some human invasive infections [7, 10, 11]. To test the hypothesis that amino acid replacements in RopB alter its regulatory activity, we assessed SpeB expression and protease activity in representative strains with each *ropB* allele. Strains with the wild-type allele secreted high levels of SpeB, whereas the majority of strains with a *ropB* polymorphism had either a marked reduction or complete absence of SpeB in vitro (Figure 1*D* and 1*E*). The results of the SpeB expression assays as measured by Western immunoblot, and protease activity assays, as measured by milk plate hydrolysis, were concordant for all strains tested (Table 2). Thus, the majority of *ropB* polymorphisms significantly decrease SpeB expression and secreted protease activity.

Domain Locations of Amino Acid Alterations in RopB

To investigate the potential molecular mechanism underlying *ropB* polymorphisms and decreased SpeB, we mapped the location of each genetic change to the *ropB* sequence (Figure 2*A*). Using a moving-average plot, we found that 6 regions within the *ropB* gene had a significant overabundance of polymorphisms (χ^2 test, $P < .05$ for each region, calculated using the observed number of changes centered on each codon in a moving-average window of 5 codons and 1.45 polymorphisms per window expected if the 81 genetic changes were randomly distributed across the whole gene sequence; note that SNPs resulting in the S103P and R226Q amino acid replacements occurring both in wild-type and V7I genetic backgrounds were counted only once) (Figure 2*B*). Importantly, each of these significantly over represented regions corresponded to a key predicted functional domain of RopB, which if altered, could be expected to alter SpeB regulation. One occurred in the DNA-binding motif, 1 occurred in the linker helix, 3 occurred in the tetratricopeptide ligand-binding motif, and 1 occurred in the oligomerization motif (Figure 2*A* and 2*B*). Moreover, 5 of the 11 codons altered by different genetic changes in different strains (eg, R28K and R28G) are located within these significantly overrepresented regions.

Next, we used molecular modeling to predict the consequence of these amino acid replacements to the structure of RopB (Figure 2*C*). Visual examination of the inferred RopB structure predicts that the amino acid replacements are highly represented in the scaffolding that supports the DNA-binding helix, but none are present in the DNA-binding helix itself (Figure 2*D*). Amino acid changes are also highly represented throughout the oligomerization domain (Figure 2*E*) and along the interior surface of the ligand-binding domain (Figure 2*F*). Thus, these data are consistent with the observation that the majority of *ropB* polymorphisms significantly alter the regulation of SpeB. Furthermore, the pattern of amino acid replacements affecting all key regulatory domains except the DNA-binding helix suggests an evolutionary model in which there is selection for mutations that change the global regulatory activity of RopB rather than interfere with its interaction with a single promoter sequence.

Effect of *ropB* Polymorphisms on Infection Type and GAS Strain Virulence

Next, to test the hypothesis that naturally occurring *ropB* polymorphisms contribute to disease-site specificity, the

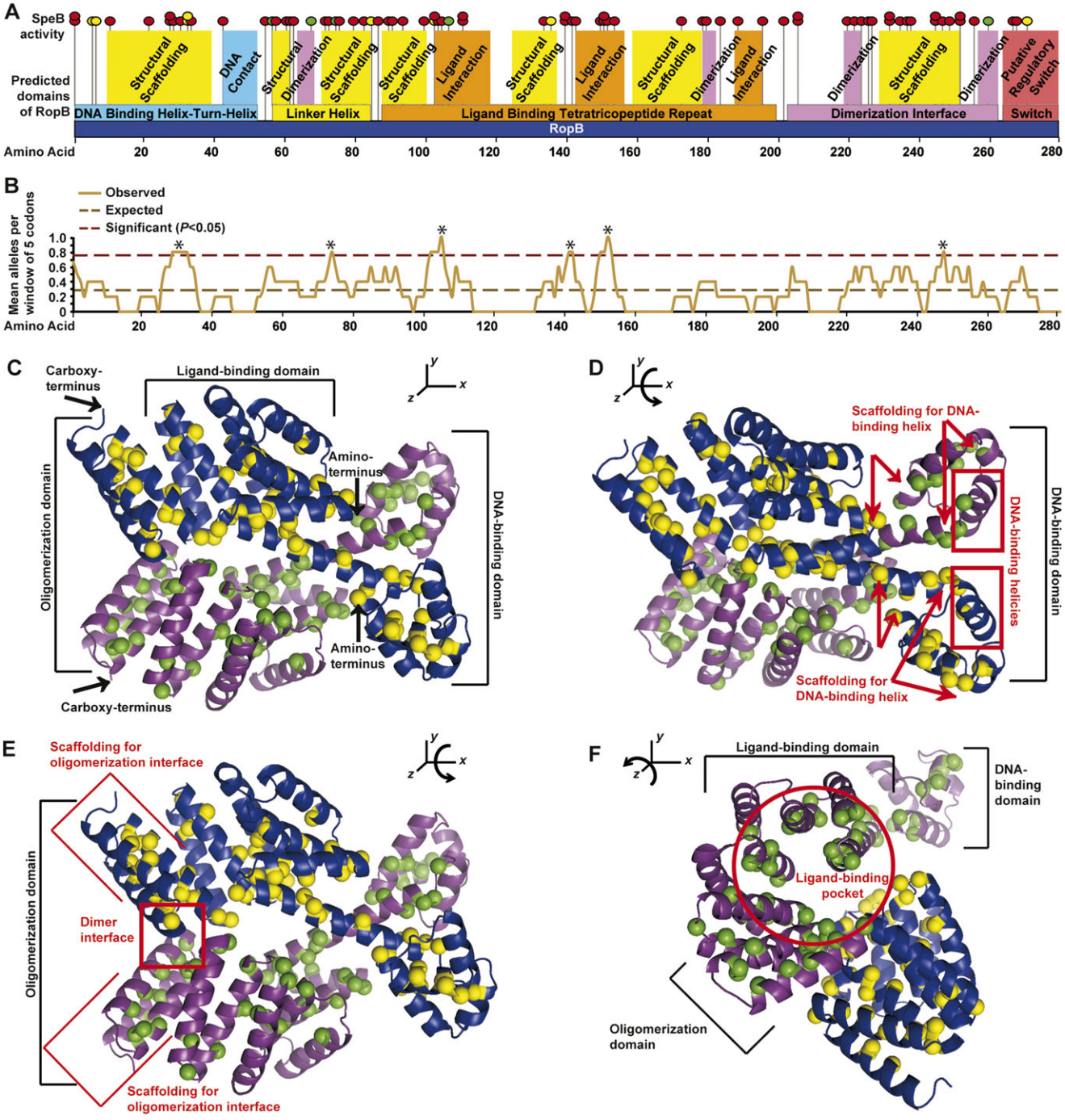

Figure 2. The *ropB* polymorphisms are predicted to alter RopB function. *A*, Location of each *ropB* polymorphism (*vertical black lines with attached spheres*) was mapped to the protein sequence (*horizontal dark blue box with amino acid positions labeled*). The DNA binding domain (*horizontal light blue box*), linker helix domain (*horizontal yellow box*), oligomerization domain (*horizontal pink box*), ligand interaction domain (*horizontal orange box*), and putative regulatory switch (*horizontal red box*) are shown. Protein regions predicted to form the DNA binding helix (*vertical light blue box*), ligand interaction interface (*vertical orange box*), oligomerization interface (*vertical pink box*) and structural scaffolding for these key functional domains (*vertical yellow boxes*) are also shown. The effect of each *ropB* polymoriphism on secreted streptococcal cysteine protease B (SpeB) expression and activity is also indicated (red, yellow, and green circles placed on top of each vertical line denoting the location of a particular polymorphism indicate absent, reduced or near wild-type levels of SpeB, respectively). The 11 codons altered by 2 different genetic changes in 2 different strains (eg, R28K and R28G) are indicated by partially overlapping stacked circles. *B*, Moving average plot of *ropB* polymorphisms per codon using a window of 5 amino acid residues (*solid light brown* line) was generated. Assuming a random distribution of altered codons, a rate of 0.29 polymorphisms per residue was expected (*dashed dark brown line*). Six regions with a significantly increased concentration of alterations were identified by the moving window analysis (asterisk above dashed maroon line indicates >0.77 polymorphisms per residue; χ^2 test, $P < .05$). *C*, Structural model of the RopB dimer is shown, with

frequencies of *ropB* polymorphisms in strains isolated from different infection types were compared. By univariate analysis, we found that strains with *ropB* polymorphisms were significantly associated with pharyngitis compared with invasive infections (Figure 3*A*). Similarly, strains with decreased SpeB activity were also significantly associated with pharyngitis (Figure 3*B*). The association between *ropB* polymorphisms, decreased SpeB, and pharyngitis remained significant if univariate analysis was restricted to strains isolated in Ontario and was independent of the frequently identified SpeB-positive V7I strains (Figure 3*A* and 3*B*). Because Alberta pharyngitis strains comprise a possibly biased convenience sample, we also performed the univariate analysis excluding this collection and again found a significant association (Figure 3*A* and 3*B*). Thus, these data unambiguously demonstrate that *ropB* polymorphisms and decreased SpeB are significantly associated with pharyngitis.

Finally, to test the hypothesis that amino acid replacements in RopB alter GAS strain virulence, isoallelic mutant strains were assessed in a mouse model of necrotizing fasciitis. Compared with the wild-type strain, the isoallelic strain with the frequently identified SNP that results in a V7I amino acid replacement secreted an intermediate level of SpeB protease activity in vitro (Figure 4*A*) and had intermediate virulence in mice (Figure 4*B* and 4*C*). In contrast, isoallelic strains with a *ropB* sequence that encodes a R226Q or V7I/R226Q amino acid replacement lacked SpeB production (Figure 4*A*), killed significantly fewer mice (Figure 4*B*), and caused markedly smaller lesions with less tissue destruction (Figure 4*C*). As expected, these features closely mimicked the characteristics observed for the isogenic strain in which the *speB* gene was deleted (Figure 4*B* and 4*C*).

DISCUSSION

The effect of SNPs and other minor gene mutations on human disease causation and susceptibility has been extensively investigated [21]. However, their consequence to bacterial virulence is poorly understood. We have recently used an unbiased whole-genome sequencing strategy to investigate strain genotype—disease phenotype relationships in infections caused by serotype M3 strains of GAS [22]. This strategy has been productive, leading to several new discoveries bearing on GAS virulence [22]. For example, we recently identified a naturally occurring SNP that disrupts the *mtsR-prsA*-SpeB virulence axis to significantly alter GAS necrotizing fasciitis capacity [4, 7]. Likewise, the gene encoding *ropB*, the major positive regulator of SpeB, was found to have the highest rate of nucleotide diversification among ~180 pharyngitis and invasive GAS strains recovered in Ontario [3, 4, 18]. Herein, *ropB* was demonstrated to be highly polymorphic in a collection of 1178 serotype M3 GAS strains recovered from patients with different disease manifestations in diverse geographic regions and time periods. The excess of *ropB* alterations was statistically significant compared with random expectation, and no silent mutations were identified. Consistent with the observation that nearly every polymorphism markedly decreased RopB function, molecular modeling predicted that the amino acid alterations are concentrated within key functional domains. Similarly, Ikebe et al [20] recently identified 4 variant *ropB* alleles, 2 of which were not identified herein (SNPs encoding F161Y and I162F amino acid alterations), among 26 serotype M3 GAS strains collected in Japan. Together, these data lead us to conclude that *ropB* evolves under diversifying selective pressure in serotype M3 GAS [23].

This finding is particularly unusual because RopB is a cytosolic transcriptional regulator that would not be expected to undergo direct selective pressure from the host immune system to change its antigenic presentation. An alternative explanation for the high diversity found in *ropB* is that changing the RopB-mediated transcriptome confers a selective advantage to serotype M3 GAS strains in some, but not all, disease conditions; otherwise, the *ropB* gene would be inactivated in all serotype M3 GAS strains. In particular, variant *ropB* alleles were significantly associated with pharyngitis, suggesting that serotype M3 strains with altered RopB function have enhanced fitness in the host oropharynx compared with sites of invasive infection. This idea is consistent with SpeB being a proven virulence factor for invasive infections but not essential for growth in human saliva [7, 24]. During invasive infections, SpeB directly causes severe tissue destruction by degrading host extracellular matrix molecules, such as fibronectin and vitronectin [25], and it indirectly causes further tissue damage by activating host matrix metalloproteases and disrupting coagulation [26–28]. SpeB also activates host proinflammatory mediators such as interleukin-1β [25], inactivates host innate immune molecules such as

Figure 2 continued. the backbone of each monomer represented as a ribbon (*violet and blue ribbons, respectively*) and each amino acid altered by a polymorphism represented as a space filling sphere (*green and yellow spheres, respectively*). The DNA-binding domain, ligand-binding domain, oligomerization domain, and amino- and carboxy-terminus of each chain are labeled. *D*, Compared with view in *C*, the RopB structure has been rotated ~45° around the x-axis to show the DNA-binding domain. Many amino acid replacements occur in the scaffolding helices (*boundaries denoted by red arrows*), but none occur in the DNA-binding helix (*red box*). *E*, Compared with view in *C*, the RopB structure has been rotated ~25° around the x-axis to show the oligomerization domain. Many amino acid replacements occur in both the dimer interface (*red box*) and the scaffolding helices (*boundaries denoted by red brackets*). *F*, Compared with view in *C*, the RopB structure has been rotated ~75° around the z-axis and ~25° around the y-axis to show the ligand-binding domain. Note that many amino acid replacements occur along the interior surface of the ligand-binding pocket (*red circle*).

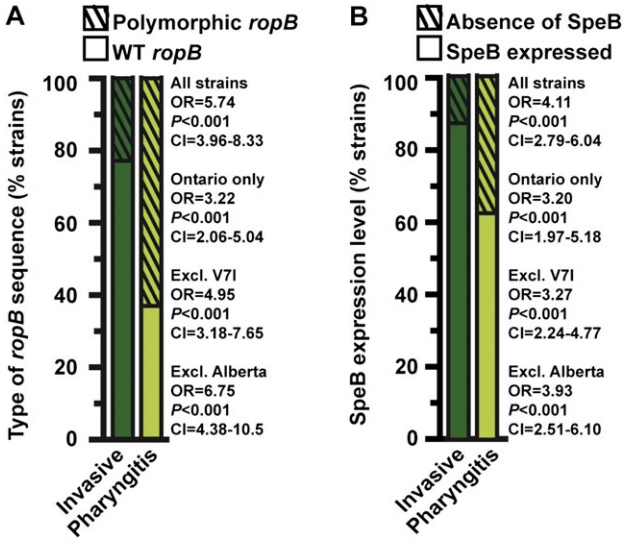

Figure 3. The *ropB* polymorphisms and decreased secreted streptococcal cysteine protease B (SpeB) are significantly associated with pharyngitis. *A*, Univariate analysis of the association between *ropB* alleles (wild type [*solid bar*] or variant [*hatched bar*]) and type of group A streptococci (GAS) infection (invasive [*dark green bar*] or oropharyngeal [*light green bar*]) is shown. *B*, Univariate analysis of the association between SpeB expression (SpeB expressed [*solid bar*] or absent [*hatched bar*]) and type of GAS infection (invasive [*dark green bar*] or oropharyngeal [*light green bar*]) is shown. Odds ratio (OR), probability (*P*), and confidence interval (CI) are shown for univariate analyses of all 1178 invasive and pharyngitis strains studied, only the Ontario invasive and pharyngitis strains, all strains except the frequently identified SpeB-positive V7I strains, and all strains except those from the Alberta collection.

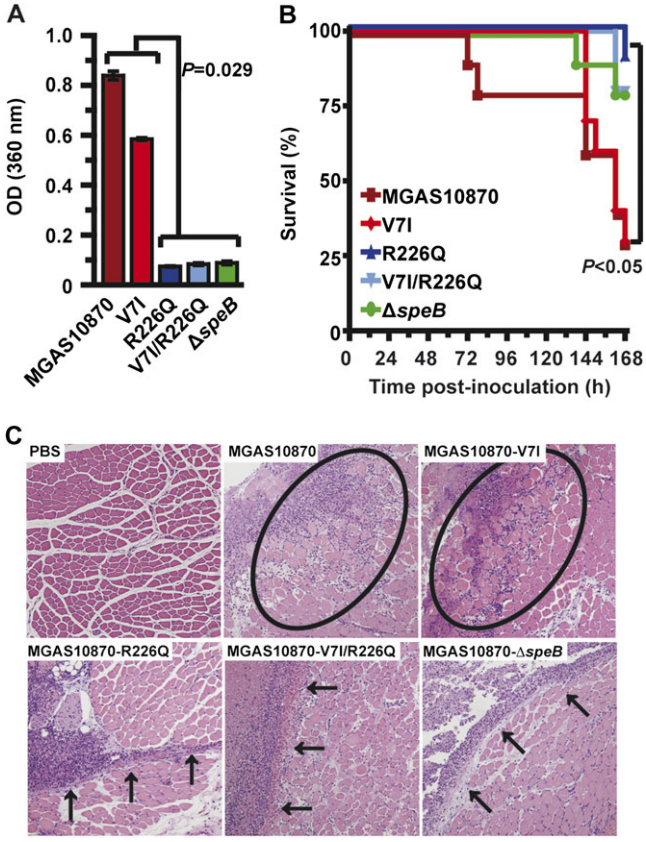

Figure 4. The *ropB* polymorphisms significantly alter group A streptococci (GAS) strain virulence. Isoallelic strains encoding a *ropB* sequence that results in the V7I, R226Q, and V7I/R226Q amino acid replacements or lacking the *speB* gene (designated Δ*speB*) were created from representative wild-type strain MGAS10870. *A*, Level of SpeB secreted protease activity of each isoallelic strain was confirmed using a quantitative chromogenic azocasein hydrolysis assay. The mean ± standard error of the mean (SEM) from 4 replicate measurements is shown (Mann-Whitney test, $P = .029$ for either strain MGAS10870 or V7I compared with either strain R226Q, V7I/R226Q or Δ*speB*). *B*, Virulence of these isoallelic strains was compared using a mouse model of necrotizing fasciitis. Results are graphically represented as a Kaplan-Meier survival curve (log-rank test, $P < .05$ for either strain MGAS10870 or V7I compared with either strain R226Q, V7I/R226Q or Δ*speB*; difference not significant for strain MGAS10870 compared with strain V7I). *C*, Histologic analysis of infected limb tissue (representative micrographs are shown, hematoxylin and eosin stain, 40× original magnification). Wild-type strain MGAS10870 caused severe muscle necrosis (*circled region*), whereas the isoallelic R226Q, V7I/R226Q and SpeB-deficient strains caused markedly smaller and less destructive lesions that were restricted to the major fascial planes (*black arrows*). The isoallelic strain with a V7I amino acid replacement had an intermediate virulence phenotype (note the 72 h difference in the time to the first near-mortality event and the slightly less severe tissue destruction caused by strain V7I compared with wild-type strain MGAS10870).

complement factor C3b and the antimicrobial peptide LL37 [29], and stimulates the release of proapoptotic molecules from host macrophages and pneumocytes [27]. Importantly, SpeB is abundantly present in infected human tissue [10, 30]. This model of GAS evolution is fully supported by our genome-wide analysis of ~180 serotype M3 strains recovered in Ontario, which also suggested that the oropharynx is the primary site of evolution for serotype M3 GAS strains, with invasive strains originating from lineages that cause pharyngitis [2, 3]. That is, most serotype M3 invasive strains are immediately descended from wild-type serotype M3 pharyngitis strains, not other invasive strains.

Inasmuch as RopB is best known for being the major positive regulator of SpeB [31], it also regulates the transcription of multiple other proven and putative virulence factors that may further contribute to decreased virulence of strains with variant *ropB* alleles. Additional targets of RopB regulation include the superantigens SpeK and SmeZ, the operon encoding the pilus structure that is critical to GAS adhesion to mucosal epithelium, the operon encoding the potent pore-forming cytotoxin streptolysin-S that inactivates host neutrophils, and the operon encoding the Opp oligopeptide transport system that is involved in nutrient acquisition [12, 18, 32]. Furthermore, RopB has also been implicated in the regulation of other GAS transcriptional regulators such as Mga (multiple gene activator), CcpA (catabolite control protein A), and the 2-component

control systems Ihk/Irr and CovR/CovS [18, 33]. Because the RopB-mediated transcriptome varies considerably between strains of different GAS serotypes and even between strains of the same serotype [12, 18, 20, 32, 34], additional studies using our isoallelic strains are needed to fully define the transcriptome differences mediated by variant *ropB* alleles. In addition, genome-wide analysis of serotype M3 GAS strains serially collected from individual pharyngitis patients or asymptomatic carriers would enable a near-real-time determination of the sequential evolutionary events that occur in vivo and help elucidate how such events contribute to human disease.

In summary, we used whole-genome sequencing followed by deep single-gene sequencing to test a hypothesis bearing on bacterial virulence and host-pathogen interactions. The comprehensive evaluation of naturally occurring diversity in a single gene among a large number of strains without geographic or temporal restrictions gives new insight to the evolution of a pathogen. It also provides a partial genetic explanation for the difference in virulence phenotypes observed between closely related strains. Given the recent application of whole-genome sequencing to many pathogenic bacteria, this approach will have widespread utility [35–37].

Notes

Acknowledgments. We thank Dr Bernard W. Beall for assistance with the Active Bacterial Core surveillance (ABCs) strain collection.

Financial support. This work was supported in part by the American Heart Association (grant AHA0775045).

Potential conflicts of interest. All authors: no reported conflicts

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Olsen RJ, Shelburne SA, Musser JM. Molecular mechanisms underlying group A streptococcal pathogenesis. Cell Microbiol **2009**; 11:1–12.
2. Beres SB, Carroll RK, Shea PR, et al. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. Proc Natl Acad Sci U S A **2010**; 107:4371–6.
3. Shea PR, Beres SB, Flores AR, et al. Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations. Proc Natl Acad Sci U S A **2011**; 108:5039–44.
4. Beres SB, Richter EW, Nagiec MJ, et al. Molecular genetic anatomy of inter- and intraserotype variation in the human bacterial pathogen group A *Streptococcus*. Proc Natl Acad Sci U S A **2006**; 103: 7059–64.
5. Beres SB, Sylva GL, Sturdevant DE, et al. Genome-wide molecular dissection of serotype M3 group A *Streptococcus* strains causing two epidemics of invasive infections. Proc Natl Acad Sci U S A **2004**; 101:11833–8.
6. O'Loughlin RE, Roberson A, Cieslak PR, et al. The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000–2004. Clin Infect Dis **2007**; 45:853–62.
7. Olsen RJ, Sitkiewicz I, Ayeras AA, et al. Decreased necrotizing fasciitis capacity caused by a single nucleotide mutation that alters a multiple gene virulence axis. Proc Natl Acad Sci U S A **2010**; 107:888–93.
8. Lukomski S, Montgomery CA, Rurangirwa J, et al. Extracellular cysteine protease produced by *Streptococcus pyogenes* participates in the pathogenesis of invasive skin infection and dissemination in mice. Infect Immun **1999**; 67:1779–88.
9. Neely MN, Lyon WR, Runft DL, Caparon M. Role of RopB in growth phase expression of the SpeB cysteine protease of *Streptococcus pyogenes*. J Bacteriol **2003**; 185:5166–74.
10. Johansson L, Thulin P, Sendi P, et al. Cathelicidin LL-37 in severe *Streptococcus pyogenes* soft tissue infections in humans. Infect Immun **2008**; 76:3399–404.
11. Olsen RJ, Watkins ME, Cantu CC, Beres SB, Musser JM. Virulence of serotype M3 group A *Streptococcus* strains in wax worms (*Galleria mellonella* larvae). Virulence **2011**; 2:111–9.
12. Hollands A, Aziz RK, Kansal R, Kotb M, Nizet V, Walker MJ. A naturally occurring mutation in *ropB* suppresses SpeB expression and reduces M1T1 group A streptococcal systemic virulence. PLoS One **2008**; 3: e4102.
13. Lamagni TL, Efstratiou A, Dennis J, Nair P, Kearney J, George R. Increase in invasive group A streptococcal infections in England, Wales and Northern Ireland, 2008–9. Euro Surveill **2009**; 14:pii=19110.
14. Kohler W, Gerlach D, Knoll H. Streptococcal outbreaks and erythrogenic toxin type A. Zentralbl Bakteriol Mikrobiol Hyg A **1987**; 266:104–15.
15. Musser JM, Nelson K, Selander RK, et al. Temporal variation in bacterial disease frequency: molecular population genetic analysis of scarlet fever epidemics in Ottawa and in eastern Germany. J Infect Dis **1993**; 167: 759–62.
16. Hartas J, Hibble M, Sriprakash KS. Simplification of a locus-specific DNA typing method (Vir typing) for *Streptococcus pyogenes*. J Clin Microbiol **1998**; 36:1428–9.
17. Lukomski S, Sreevatsan S, Amberg C, et al. Inactivation of *Streptococcus pyogenes* extracellular cysteine protease significantly decreases mouse lethality of serotype M3 and M49 strains. J Clin Invest **1997**; 99:2574–80.
18. Carroll RK, Shelburne SA 3rd, Olsen RJ, et al. Naturally occurring single amino acid replacements in a regulatory protein alter streptococcal gene expression and virulence in mice. J Clin Invest **2011**; 121:1956–68.
19. DeLano WL. Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol **2002**; 12:14–20.
20. Ikebe T, Ato M, Matsumura T, et al. Highly frequent mutations in negative regulators of multiple virulence genes in group A streptococcal toxic shock syndrome isolates. PLoS Pathog **2010**; 6:e1000832.
21. Jakobsson M, Scholz SW, Scheet P, et al. Genotype, haplotype and copy-number variation in worldwide human populations. Nature **2008**; 451:998–1003.
22. Musser JM, DeLeo FR. Toward a genome-wide systems biology analysis of host-pathogen interactions in group A *Streptococcus*. Am J Pathol **2005**; 167:1461–72.
23. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet **2002**; 18:486.
24. Shelburne SA 3rd, Granville C, Tokuyama M, Sitkiewicz I, Patel P, Musser JM. Growth characteristics of and virulence factor production by group A *Streptococcus* during cultivation in human saliva. Infect Immun **2005**; 73:4723–31.
25. Kapur V, Topouzis S, Majesky MW, et al. A conserved *Streptococcus pyogenes* extracellular cysteine protease cleaves human fibronectin and degrades vitronectin. Microb Pathog **1993**; 15:327–46.
26. Matsuka YV, Pillai S, Gubba S, Musser JM, Olmsted SB. Fibrinogen cleavage by the *Streptococcus pyogenes* extracellular cysteine protease and generation of antibodies that inhibit enzyme proteolytic activity. Infect Immun **1999**; 67:4326–3.
27. Tamura F, Nakagawa R, Akuta T, et al. Proapoptotic effect of proteolytic activation of matrix metalloproteinases by *Streptococcus pyogenes* thiol proteinase (*Streptococcus* pyrogenic exotoxin B). Infect Immun **2004**; 72: 4836–47.
28. Meinert Niclasen L, Olsen JG, Dagil R, Qing Z, Sorensen OE, Kragelund BB. Streptococcal pyogenic exotoxin B (SpeB) boosts the contact system via binding of α-1 antitrypsin. Biochem J **2011**; 434: 123–32.

29. Nyberg P, Rasmussen M, Bjorck L. alpha2-Macroglobulin-proteinase complexes protect *Streptococcus pyogenes* from killing by the antimicrobial peptide LL-37. J Biol Chem **2004**; 279:52820–3.
30. Thulin P, Johansson L, Low DE, et al. Viable group A streptococci in macrophages during acute soft tissue infection. PLoS Med **2006**; 3:e53.
31. Lyon WR, Gibson CM, Caparon MG. A role for trigger factor and an Rgg-like regulator in the transcription, secretion and processing of the cysteine proteinase of *Streptococcus pyogenes*. Embo J **1998**; 17:6263–75.
32. Dmitriev AV, McDowell EJ, Chaussee MS. Inter- and intraserotypic variation in the *Streptococcus pyogenes* Rgg regulon. FEMS Microbiol Lett **2008**; 284:43–51.
33. Chaussee MS, Watson RO, Smoot JC, Musser JM. Identification of Rgg-regulated exoproteins of *Streptococcus pyogenes*. Infect Immun **2001**; 69:822–31.
34. Dmitriev AV, McDowell EJ, Kappeler KV, Chaussee MA, Rieck LD, Chaussee MS. The Rgg regulator of *Streptococcus pyogenes* influences utilization of nonglucose carbohydrates, prophage induction, and expression of the NAD-glycohydrolase virulence operon. J Bacteriol **2006**; 188:7230–41.
35. Mwangi MM, Wu SW, Zhou Y, et al. Tracking the in vivo evolution of multidrug resistance in *Staphylococcus aureus* by whole-genome sequencing. Proc Natl Acad Sci U S A **2007**; 104:9451–6.
36. Smith EE, Buckley DG, Wu Z, et al. Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. Proc Natl Acad Sci U S A **2006**; 103:8487–92.
37. Croucher NJ, Harris SR, Fraser C, et al. Rapid pneumococcal evolution in response to clinical interventions. Science **2011**; 331: 430–4.