Article

Genome Sequence of an M3 Strain of *Streptococcus pyogenes* Reveals a Large-Scale Genomic Rearrangement in Invasive Strains and New Insights into Phage Evolution

Ichiro Nakagawa,[1,7] Ken Kurokawa,[2] Atsushi Yamashita,[3] Masanobu Nakata,[1] Yusuke Tomiyasu,[1] Nobuo Okahashi,[1] Shigetada Kawabata,[1,4] Kiyoshi Yamazaki,[2] Tadayoshi Shiba,[5] Teruo Yasunaga,[2] Hideo Hayashi,[6] Masahira Hattori,[3] and Shigeyuki Hamada[1]

[1]Department of Oral and Molecular Microbiology, Osaka University Graduate School of Dentistry, Suita-Osaka 565-0871, Japan; [2]Genome Information Research Center, Osaka University, Suita-Osaka 565-0871, Japan; [3]Kitasato Institute for Life Sciences, Kitasato University, Sagamihara-Kanagawa 228-8555, Japan; [4]PRESTO, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan; [5]School of Science, Kitasato University, Sagamihara-Kanagawa 228-8555, Japan; [6]Institute of Basic Medical Sciences, University of Tsukuba, Tsukuba-Ibaraki 305-8575, Japan

Group A streptococcus (GAS) is a gram-positive bacterial pathogen that causes various suppurative infections and nonsuppurative sequelae. Since the late 1980s, streptococcal toxic-shock like syndrome (STSS) and severe invasive GAS infections have been reported globally. Here we sequenced the genome of serotype M3 strain SSI-1, isolated from an STSS patient in Japan, and compared it with those of other GAS strains. The SSI-1 genome is composed of 1,884,275 bp, and 1.7 Mb of the sequence is highly conserved relative to strain SF370 (serotype M1) and MGAS8232 (serotype M18), and almost completely conserved relative to strain MGAS315 (serotype M3). However, a large genomic rearrangement has been shown to occur across the replication axis between the homologous *rrn*–*comX1* regions and between two prophage-coding regions across the replication axis. A total of 1 Mb of chromosomal DNA is inverted across the replication axis. Interestingly, the recombinations between the prophage regions are within the phage genes, and the genes encoding superantigens and mitogenic factors are interchanged between two prophages. This genomic rearrangement occurs in 65% of clinical isolates (64/94) collected after 1990, whereas it is found in only 25% of clinical isolates (7/28) collected before 1985. These observations indicate that streptococcal phages represent important plasticity regions in the GAS chromosome where recombination between homologous phage genes can occur and result not only in new phage derivatives, but also in large chromosomal rearrangements.

[Supplemental material is available online at www.genome.org. The sequence data from this study have been submitted to DDBJ under accession no. BA000034. The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: K. Kikuchi, S. Murai, M. Ohta, T. Ikebe, and H. Watanabe.]

Group A *Streptococcus pyogenes* (GAS) is responsible for a variety of suppurative infections, including pharyngitis, scarlet fever, impetigo, and cellulitis and for nonsuppurative sequelae, such as acute rheumatic fever, acute glomerulonephritis, and reactive arthritis (Cunningham 2000). Although GAS is still sensitive to classical antibiotics such as penicillin, an unexplained resurgence of GAS infections has been reported since the mid-1980s. Outbreaks of rheumatic fever have also been reported in several states of the USA and other countries (Cone et al. 1987; Kaplan 1991). Furthermore, in the late 1980s, streptococcal toxic shock-like syndrome (STSS), bacteremia, and severe invasive group A streptococcal skin and soft tissue infections were reported in the USA, Europe, and Japan (Musher et al. 1996; Holm et al. 1997; Murase et al. 1999). These severe and invasive diseases exhibit high morbidity and mortality. In severe cases, GAS invades the skin and soft tissues and destroys the infected tissues or limbs, and is therefore called the "flesh-eating" bacterium (Stevens 1999). Epidemiological data indicate a clonal expansion of serotype strains (i.e., serotypes M1 and M3) in severe invasive infections of GAS (Musser et al. 1991). These results suggest that unknown virulence factors that cause STSS have been acquired either by horizontal transfer or by genetic changes occurring in these strains (Inagaki et al. 2000). The genome of a serotype M1 organism isolated from a wound infection (Fer-

[7]Corresponding author.
E-MAIL ichiro@dent.osaka-u.ac.jp; FAX +81-6-6878-4755.

retti et al. 2001), a serotype M18 organism from a patient with acute rheumatic fever (Smoot et al. 2002), and a serotype M3 organism from an STSS patient (Beres et al. 2002) were recently sequenced. Comparative analysis of these three strains provided interesting genomic information about GAS pathogenicity. However, there is only limited genomic information available about GAS strains associated with severe, invasive infections. Comparative genomic analyses have been performed for other pathogenic bacteria such as *Helicobacter pylori* (Alm et al. 1999), *Mycobacterium* species (Cole 1998), and *Chlamydia* species (Read et al. 2000). These comparative studies indicate that the genes within closely related species are highly conserved, with the exception of inversions, translocations, phage integrations, and the mobile genetic elements. In particular, the genomic arrangement regions in *H. pylori* by inversion and translocation form a specific genetic segment called the "plasticity zone", and almost half of the strain-specific genes are included in this region (Alm et al. 1999). These observations suggest that the genes in plasticity zones have undergone genetic reorganization to a much higher degree than the rest of the chromosome, and this is thought to be related to the diversity of phenotypes seen in these organisms (Maeder et al. 1999; Hughes 2000). However, it is difficult to analyze genomic rearrangements using conventional methods such as pulsed-field gel electrophoresis (PFGE). Only comparative genomic analyses based on whole-genome sequences can provide useful information about genomic organization.

To develop effective prevention strategies or new therapeutic methods for these severe infections, it is necessary to understand the biology of this organism at the genomic level. As a first step in resolving the pathogenesis of severe infections, we sequenced the whole genome of a serotype M3 strain isolated from a patient with STSS in Japan, and compared the genomic sequence, genome structure, and gene variation with the genomes of M1, and M18 serotype organisms. In addition, we further compared two M3 organisms isolated in Japan and the United States. These genome sequences provide useful information about the evolutionary events associated with severe invasive GAS strains and the pandemic of recent GAS infections in advanced countries.

RESULTS AND DISCUSSION

General Features of the Strain SSI-I Genome

The genome of strain SSI-1 is a single, circular chromosome of 1,894,275 bp (Fig. 1). The size of the total genome is virtually identical to that of strains MGAS8232 and MGAS315, and is almost 38 kb larger than that of strain SF370 (Table 1). Strain SSI-1 contains 1861 open reading frames (ORFs) that cover 85.94% of the whole sequence, with an average size of 853 bp. Six regions in the SSI-1 genome are composed of prophage elements, and 341 ORFs are included in these regions. The sequence has been deposited in the DNA Data Bank of Japan (DDBJ; accession no. BA000034) and is available from our Web site (http://genome.gen-info.osaka-u.ac.jp/bacteria/spyo/).

### Extensive Genomic Rearrangements in Strain SSI-I

Alignment of the whole GAS genome (Fig. 2) shows the extent of genomic rearrangement in strain SSI-1 relative to the genomes of strains SF370, MGAS315, and MGAS8232. This is the most prominent difference and characteristic of the SSI-1 genome. The origins of DNA replication (*ori*; Suvorov and

Ferretti 2000) and *dif*-like termination sequence (Ferretti et al. 2001) are completely conserved among the four strains. Alignment analyses reveal clear X-shaped chromosomal rearrangements that are symmetrical across the *ori/ter* axis (replication axis; Fig. 2). We also determined the relative location of the homologous genes on the chromosomes of strains SSI-1 and other *S. pyogenes* strains (Fig. 3A). The locations of genes around the *ori* region (from 1650 kb to 230 kb) and around the *ter* region (from 920 kb to 1100 kb) were almost completely conserved between strain SSI-1 and the other strains. However, the other homologous genes were translocated to an inverted position on the chromosome, indicating that these two segments are translocated to other replichores (Fig. 2). On the other hand, the locations of genes between strains SF370 and MGAS8232, and between strains SF370 and MGAS315 are conserved except for the phage genes (Fig. 3B). Therefore, these chromosomal segments are syntenic, except for small gaps encoding the genes of prophage or phage-like elements. As a consequence, the inversion in strain SSI-1 relative to other GAS strains does not change gene orientation relative to the replication axis, and produces the characteristic X-shaped plot diagram (Fig. 3A; Eisen et al. 2000; Tillier and Collins 2000a). This characteristic X-shaped chromosomal inversion has also been found between *Pyrococcus horikoshii* and *P. abyssi* (Zivanovic et al. 1997; Makino and Suzuki 2001), and between *Chlamydia pneumoniae* and *C. trachomatis* (Read et al. 2000). However, comparison of the three GAS strains provides straightforward evidence of an intraspecies genomic rearrangement.

We further compared the order of orthologous genes between *S. pyogenes* and *S. pneumoniae* strains TIGR4 and R6, or *S. agalactiae* strain 2603V/R (Hoskins et al. 2001; Tettelin et al. 2001, 2002). As shown in Figure 3C, the X-shaped plot diagram is observed in comparison of GAS strain SSI-1 with *S. pneumoniae* TIGR4. Interestingly, an X-shaped plot diagram is also found in the comparison between the genomes of GAS strain SF370 and *S. pneumoniae* TIGR4. These observations indicate that the chromosomal inversion across the replication axis has frequently occurred in individual streptococcal species after branching from a common ancestor of the genus *Streptococcus* (Bentley et al. 1991). The gene order between strain SSI-1 and *S. agalactiae* is more highly conserved than between strain SF370 and *S. agalactiae* (Fig. 3D). The similarity of genome architecture of strain SSI-1 with *S. agalactiae* suggests two possibilities regarding the chromosomal inversion in the pyrogenic group of *Streptococcus*. One is that strain SSI-1, in which the ancestral chromosomal architecture is present, is the common ancestor of the pyrogenic group. The other is that this inversion has coincidentally occurred in both SSI-1 and *S. agalactiae*. However, the fact that the isolation frequency of clinical isolates with chromosomal inversion increased after 1985 (Suppl. Fig. A) clearly indicates that the chromosomal inversion found in SSI-1 and other recent isolates occurred very recently.

Genomic Analysis of the Rearrangement in the *rrn–comX* Regions

At the recombination sites near the *ori* region in strain SSI-1, we found a 5.6-kb-long homologous sequence, which included genes for 16S rRNA, tRNA-Ala, 23S rRNA, tRNA-Asn, tRNA-Arg, and the *comX1* homolog in both regions. Our sequencing data show that the rearrangement breakpoint is located 41 bp downstream from the *comX1* homolog genes
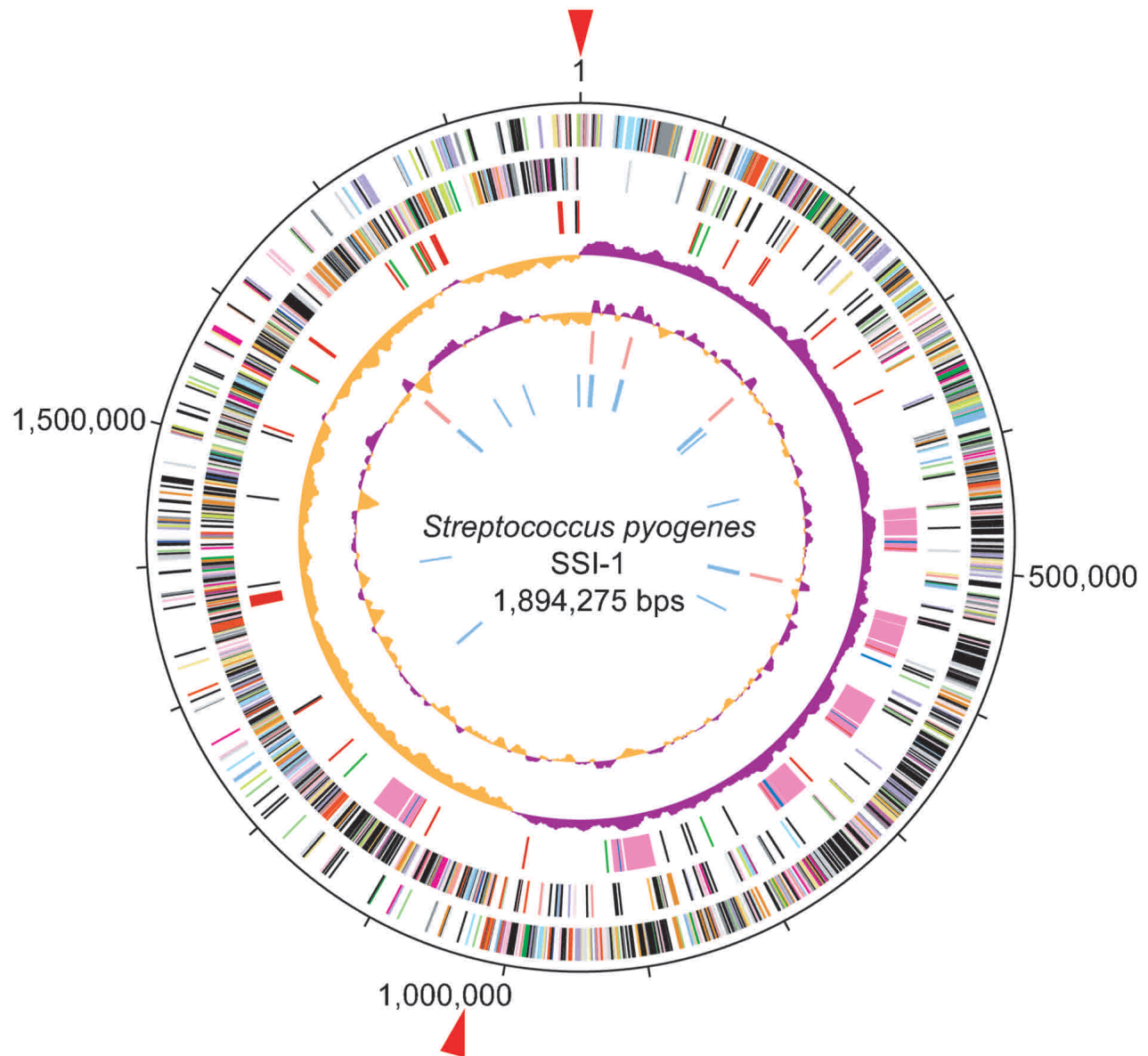
Figure 1 Circular map of GAS strain SSI-1. The outer circle shows the scale (bp). Rings 1 and 2 show the coding sequence by strands (ring 1, clockwise; ring 2, counterclockwise). The predicted ORFs are distinguished by different colors in the COG classification. Ring 3: The genes for bacteriophage (pink), transposase genes and insertion sequence (black), adhesion molecules (green), hyaluronidase genes (blue), and other putative virulence factors (red). Ring 4: The GC skew analysis. Ring 5: The G+C contents. Rings 6 and 7: The ribosomal RNA (red) genes and transfer RNA (blue) genes identified in the genome. Red arrows: The origin of DNA replication (*ori*; Suvorov and Ferretti 2000) and the putative region of replication terminus (*ter*).

(SPs0226 and SPs1640), and this rearrangement breakpoint is completely conserved within other strains (Fig. 4A). To confirm the break-point in this region, we designed long-PCR primers (1718F and 952R). An 18-kb PCR product was clearly amplified from strain SSI-1, but not from SF370 (Fig. 4B). These observations indicate that the chromosomal recombination in this site may be due to a *recA*-dependent recombination between two 5.6-kb inverted repeats containing ribosomal operon and *comX1* (Radding 1988; Tillier and Collins 2000b). In *Salmonella typhimurium*, inversion between large inverted repeats (over 5 kb) separated by large intervals (>60 kb) was shown to be *recA*- and *recB*-dependent (Segall and

Roth 1994). In addition, the chromosomal inversion between the ribosomal operons was reported between *S. typhimurium* and *S. paratyphi* A (Liu and Sanderson 1995). Therefore, our sequencing analysis represents clear evidence that long-repeated sequences across the replication axis can also induce large-scale chromosomal rearrangements within the same species of gram-positive bacteria.

Although the role of homologous recombination in genomic rearrangements across the replication axis has not been elucidated, Tillier and Collins (2000b) proposed an alternative model for the observed pattern of rearrangement. Gene translocation across the replication axis may result dur-

Table 1. General Features of *S. pyogenes* Strains SSI-1, SF370, MGAS315, and MGAS8232

Features	Strains			
	SSI-1 (M3)	SF370[a] (M1)	MGAS315[b] (M3)	MGAS8232[c] (M18)
Length of sequence (bp)	1,894,275	1,852,441	1,900,521	1,895,017
G+C content	37.93%	38.51%	38.6%	38.5%
Open reading frames				
Percentage coding	85.94%	83.77%	85.76%	84.94%
Protein coding region	1861	1696	1865	1845
Average gene length (bp)	853	915	874	872
SNPs (per gene)[d]				
Average number of SNPs	—	11.53	0.05	12.61
Silent SNPs (sSNPs)	—	6.40	0.02	7.26
Coding SNPs (cSNPs)	—	5.13	0.03	5.35
RNA				
Ribosomal RNA	5	6	6	6
Transfer RNA	57	60	60	60
tmRNA	1	1	1	2
Transposon or IS	21	17	22	25
Phage-related				
Pro-phage	6	4	6	5
Remnant	4	4	3	3

[a,b,c]The genomic sequences of other *S. pyogenes* strains (SF370:AE004092, MGAS8232: AR009949 and MGAS315: AE014074) were obtained through the Web site of the NCBI, http://www.ncbi.nlm.nih.gov.
[d]SNPs in orthologous genes of strain SSI-1 were compared with those of strains SF370, MGAS315, and MGAS8232.

ing the process of genome replication because homologous recombination equidistant from the *ori* region will occur in close physical proximity between two replication forks, and single- or double-stranded DNA breaks have been implicated in illegitimate recombination (Kuzminov and Stahl 1999). In fact, two *rrn-comX* regions are found equidistant from the *ori* region in the GAS genomes; thus, this model is in a good agreement with the genomic rearrangement mechanism of the GAS chromosome. In addition, sequence specificity may also affect the recombination of this site. It has been reported that a 240-bp chromosomal palindromic inverted repeat sequence is required for the repair of recombination from DNA cleavage of SbcCD nuclease (Leach et al. 1997). In the GAS genome, palindromic sequences (AAAAAAACAACAGGACAC TAATGTCCTGTTGTTTTTTT) are found just upstream of *comX* homologs (SPs0226 and SPs1640). This sequence specificity may also affect the site-specific recombination during homologous recombination.

## Unbalanced Genome Architecture by Phage Integration Affects the Chromosomal Inversion

As shown in Figure 1, the locations of *ori* and *ter* are not exactly opposite each other, leading to an "unbalanced" genome in strain SSI-1. This unbalanced genome was also found in strains SF370 (Ferretti et al. 2001) and MGAS315 (Beres et al. 2002). Only MGAS8232 (Smoot et al. 2002) had a balanced genome architecture. Strain SSI-1 has six prophages in the genome; however, five prophage regions are found in the one replichore. These lopsided phage integrations into chromosomal DNA may result in an unsymmetrical genome architecture across the replication axis. This unbalancing of the genome might induce the chromosomal rearrangement for stabilizing the genome architecture. This type of genomic rearrangement (the adopt-adapt model) was also found in *Salmonella enterica* serovar Pollorum (Liu et al. 2002). The genomic balance was disrupted by the 157-kb insertion near the

ter region in *S. enterica* serovar Pollorum, and the genomic inversions are considered to occur between two homologous *rrn* operons (*rrn*D/E and *rrn*H/G) and between two homologous insertions for rebalancing of the genome. In GAS genomes, we speculate that the "unbalanced" genomes found in strains SSI-1, SF370, and MGAS315 might be in flux by the phage integration. In fact, if the two phages found in strain SSI-1 (SPsP1 and SPsP2, totally about 80 kb) are excised, the genome will be in balance, because each replichore is almost equal in length. These observations suggest that the phage integrations affect the balance of the choromosomal architecture. Therefore, the imbalance and rebalancing by the phage integrations may cause the chromosomal inversion in GAS.

Chromosomal Inversion Triggers Streptococcal Phage Rearrangement

The genomic rearrangement sites of strain SSI-1 around the *ter* region are coincident with the phage integration sites. Two SSI-1 phages, SPsP5 and SPsP6, are integrated at positions equidistant from the *ter* region (Fig. 4C). In comparison with strain SF370, the rearrangement break-points were 43 bp upstream from the integrase (*int*) gene of SPsP5 (SPs0877) and the tmRNA region, which is predicted to be a streptococcal T12 phage integration site in the SF370 genome (McShan et al. 1997). On the other hand, the rearrangement sites of strain SSI-1 are also found within the phage regions of MGAS315 and MGAS8232. The rearrangements occur between the holin genes of phages SPsP6 (SPs1121) and 315.1 (*spyM3_0731*), and between the hydrase genes of phages SPsP5 (SPs0933) and 315.2 (*spyM3_0922*), respectively (Fig. 4C).

Rearrangements are also found between SPs0933 of SPsP5 and *spyM18_0777* of φSpeC in MGAS8232, and SPs1118 of SPsP6 and *spyM18_1237* of φSpeA in MGAS8232. To clarify the genomic recombination in these sites, we further compared the genomic sequences of streptococcal phages by dot-plot analysis (Fig. 5). Interestingly, about 8–10 kb of the SPsP5
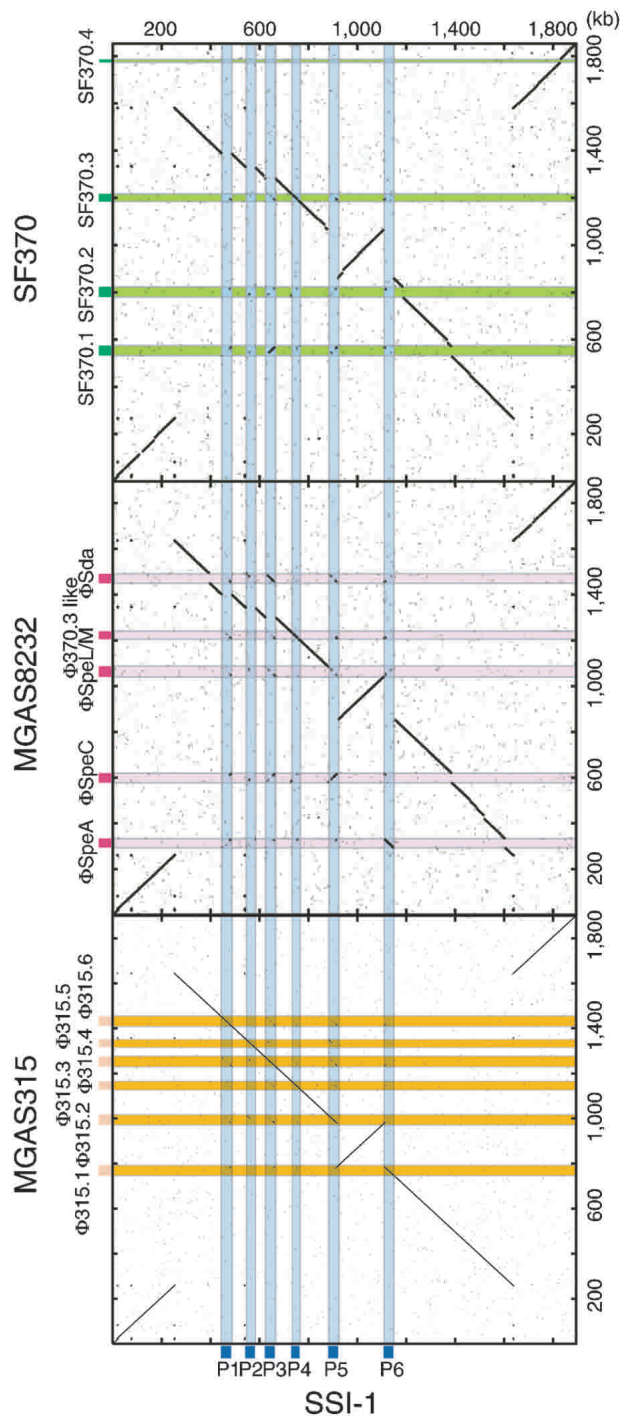
Figure 2 Comparison of the three sequenced GAS genomes based on the chromosomal organization of strain SSI-1. Dot-plot analyses were performed based on the genomic sequence of strain SSI-1 and the other three GAS strains (SF370, MGAS8232, and MGAS315). Deflection of segments along either axis indicates insertions of DNA segments. Segments not aligning along the diagonal line represent sequences that are similar but located in different parts of the genomes. Light blue, SSI-1-specific phages; pink, MGAS8232-specific phages; light green, SF370-specific phages; light orange, MGAS315-specific phages.

and SPsP6 phage regions near the attP-R sites are similar to those of other GAS phages. Therefore, genomic rearrangements in these regions might occur by homologous recombination of this region. However, in strain SF370, two phages in these regions should be excised. It is still unclear why homologous recombination of the *rrn-comX* region induces homologous recombination around the *ter* region. In comparative studies of *Salmonella* genomes, the orientation of the *ter* region did not affect chromosomal stability (Liu and Sanderson 1995; Hoskins et al. 2001; Makino and Suzuki 2001). On the other hand, the orientation of the *ter* region is thought to be important for chromosomal segregation in *Escherichia coli* (Perals et al. 2001). The orientation of the *ter* region in the GAS genome might be important to stabilize chromosomal segregation.

## Comparison of the Genes in Strain SSI-I and Strains SF370 and MGAS8232

Of the 1528 SSI-1 ORFs without phage-related genes, 1426 are common to all three strains. Most vegetative growth-related genes are included among the genes shared by the three GAS strains, and most of these genes are highly conserved, as revealed by SNP analysis (Table 1). As reported previously, strain SSI-1 also lacks tricarboxylic acid cycle and electron transport genes (Ferretti et al. 2001; Smoot et al. 2002). Of the other 102 ORFs, 38 are specific to SSI-1, 19 are shared by SSI-1 and SF370, and 45 are shared by SSI-1 and MGAS8232.

Comparison of the Genes in Strain SSI-I and Strain MGAS3I5

The gene contents of strain SSI-1 and strain MGAS315 are almost completely conserved as revealed by the SNP analysis (Table 1). These two strains belong to serotype M3, and were isolated from STSS patients in different countries. However, 44 genes were found to be different between the strains, with silent SNPs (sSNP) or coding SNPs (cSNP; Table 2). Of them, eight have two or more amino acid changes per gene, and consist of four hypothetical proteins, a putative domain protein (SPs0659), two transcriptional regulators (SPs0776 and SPs1742), and putative UDP-glucose 6-dehydrogenase (*hasC*; SPs1848). Recent epidemiological studies using PFGE or multilocus sequence typing for housekeeping genes indicated that clinical isolates of invasive M3 strains were composed of multiple clones. However, there is no clear evidence for an increased propensity of some M3 clones to be associated with invasive infections, compared with M3 clones prevalent in the general population (Enright et al. 2001; Johnson et al. 2002). Our genetic information will be useful for future epidemiological studies to analyze the temporal and geographic distributions of invasive M3 strains, because nonsynonymous changes in chromosomal proteins (SPs0659, SPs0776, SPs1742, and SPs1848) can be used for these studies.

Virulence Factors

Most of the genes for known virulence factors are highly conserved in these strains (E values $>1.0 \times 10^{-80}$) except for phage-related genes (Table 3). Genes encoding four superantigens (SPs0161, 0560, 0657, and 1119) and three streptodornase genes along with those for mitogenic factors SPs0455, 0700, and 1743 were found in the SSI-1 genome. We also found 22 genes predicted to encode extracellular proteins containing the LPXTG motif or other tripartite cell-wall anchor motif (Janulczyk and Rasmussen 2001). These genes are
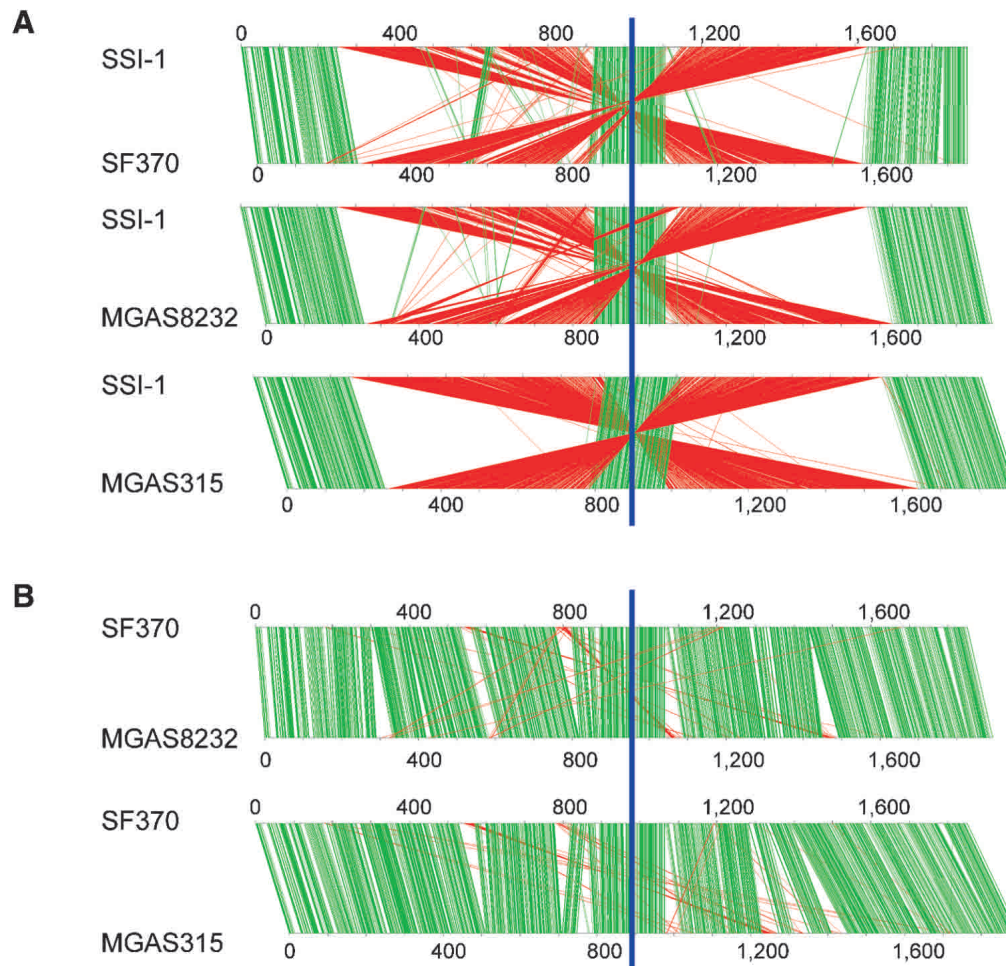
Figure 3 (Continued on next page)

also conserved among the four strains except for the genes for M protein, fibronectin-binding protein, and *scl* protein (Suppl. Table A). Genes for C5a peptidase, streptolysin O, and streptolysin S are well conserved (Table 3). These virulence factors are thought to influence the pathogenicity of GAS infections. However, no clear genetic variation in these genes was found. Therefore, the effects of these genes in severe GAS infection may arise from the regulation of gene expression (Cunningham 2000). In contrast, the genes for collagen-like proteins (SclA and SclB), Nra (Podbielski et al. 1999), Cpa, and fibronectin binding protein (F2-like protein) shared homology; however, these genes have different characteristics in each M-type strain. In particular, *nra* (SPs0099) and *lepA* (SPs0101) show significantly higher sSNPs and cSNPs (337 and 135 bases, respectively) compared with the average sSNPs (6.4 bases) and cSNPs (5.13 bases) in strains SSI-1 and SF370.

The gene for PrtF2-like protein (SPs0106) is conserved among strains SSI-1, MGAS315, and MGAS8232, although *nra* is disrupted in MGAS8232. The *nra-lepA* region had already been reported as the fibronectin- and collagen-binding proteins and T antigen (FCT) region, with an extensive intergenomic recombination site (Bessen and Kalia 2002). Such genetic variation also occurs in the *mga*-regulon, which includes the gene for M protein. All genes in both regions are thought

to play important roles in the adherence and invasion of this organism (Cunningham 2000). This genetic diversity of adhesive molecules may reflect the types of diseases caused by this organism and their relative incidence. On the other hand, the other ORFs encoding cell-anchor motifs found in gram-positive pathogenic bacteria are conserved among the four strains of GAS (Suppl. Table A; Navarre and Schneewind 1999; Janulczyk and Rasmussen 2001). Most of them encode proteins with enzymatic properties, indicating that these genes might be important in modulating other cell surface proteins, or in the acquisition of nutrients. Therefore, these genes are highly conserved in all GAS strains.

Streptococcal Phages Exchange Their Virulent Cassettes by Genomic Rearrangement

In contrast to the putative virulence factors, the gene contents of phages are variable among different serotypes of GAS organisms. Six regions of the genome of strain SSI-1 are composed of phage or phage-like elements, and they are almost completely conserved in strain MGAS315, serotype M3. Two phage regions in SSI-1 identified as SPsP1 and SPsP2, which are identical to phages ϕ315.6 and ϕ315.5, respectively, are specific for serotype M3 strains. SPsP5 is almost identical to
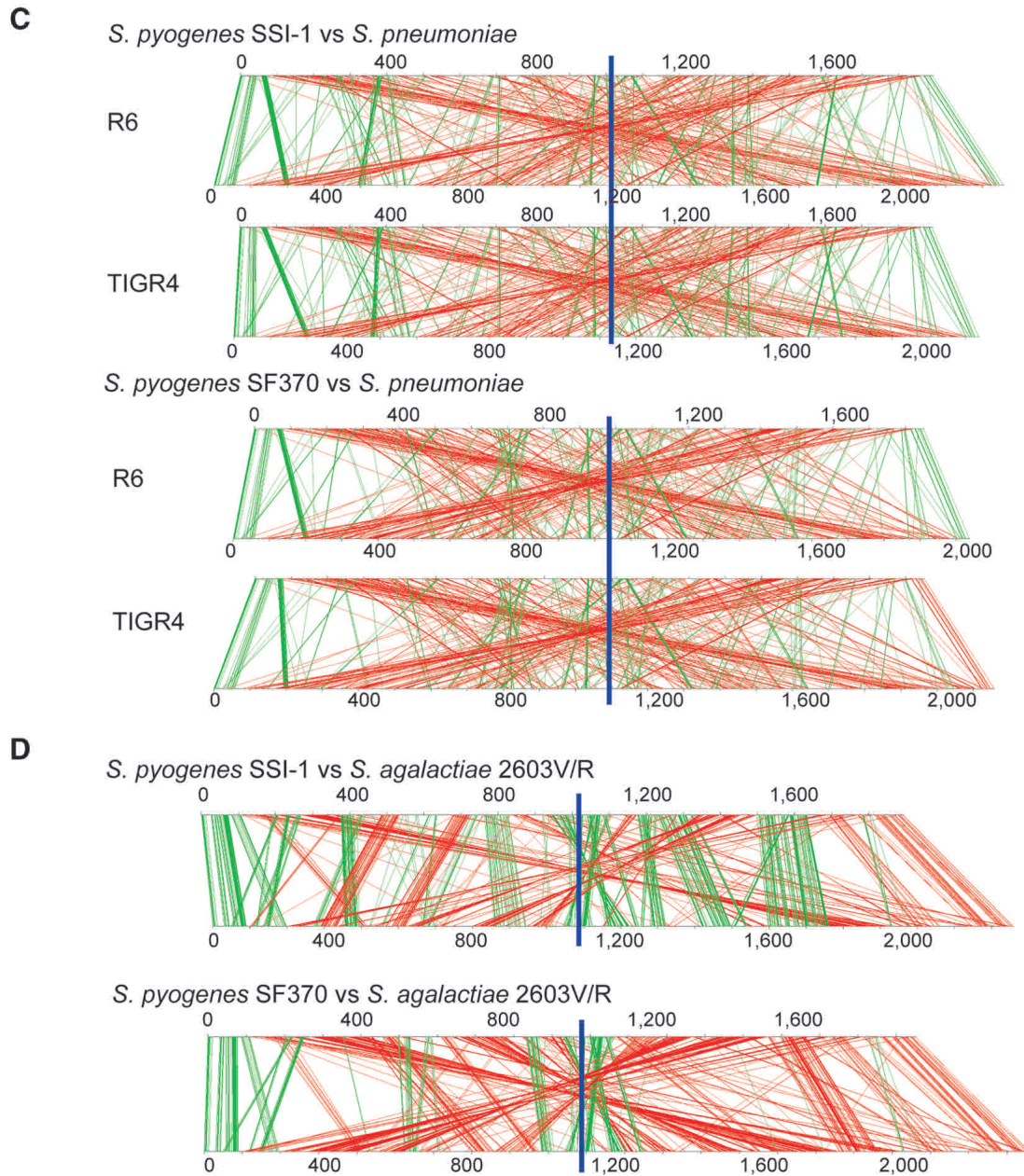
Figure 3 Comparison of genomic location of the homologous ORFs in GAS strains SSI-1 (M3) vs. SF370 (M1), SSI-1 vs. MGAS8232 (M18), and SSI-1 vs. MGAS315 (M3) (*A*), in GAS strains SF370 (M1) vs. MGAS8232 (M18) and SF370 vs. MGAS315 (M3) (*B*), in GAS strains and *Streptococcus pneumoniae* (*C*), or *S. agalactiae* 2603V/R (*D*). Pairs of homologous ORFs between two strains were generated by BLASTP analysis (E<1.0×10^{-5}). The homologous ORFs on each chromosome are shown by connected green lines (the relative location of each ORF is conserved) or red lines (the relative location of each ORF is in the opposite direction). The blue line indicates the putative region of the replication terminus (*ter*).

φ315.2 in MGAS315, and thus, this phage is also specific for M3 strains. Dot-plot analysis revealed that the last one-fourth of phage regions near the attP-R site are more similar to those of other phages (Fig. 5) than other regions of the phage sequences, even though these phages are only found in M3 strains. Interestingly, the streptococcal superantigen (SSA) gene in phage SPsP6 in strain SSI-1 is not found in the φ315.1 region, and is translocated to φ315.2 in strain MGAS315 (Fig. 3C). It is noted that the virulence genes encoding the superantigens, hyaluronidases and streptodornases, which are pre-

dicted as major virulence factors, are included in this region. These observations indicate that the virulence genes within the phage regions, which are a "streptococcal virulence cassette", are exchangeable if these phages are integrated in sites equidistant from the *ter* region (Fig. 6). Therefore, our genomic analyses clearly reveal an evolutionary mechanism for GAS phages.

We found that the holin genes, which are located near the attP-R site and found in the recombination site between SPsP5 and SPsP6, are classified by phylogenetic analysis into
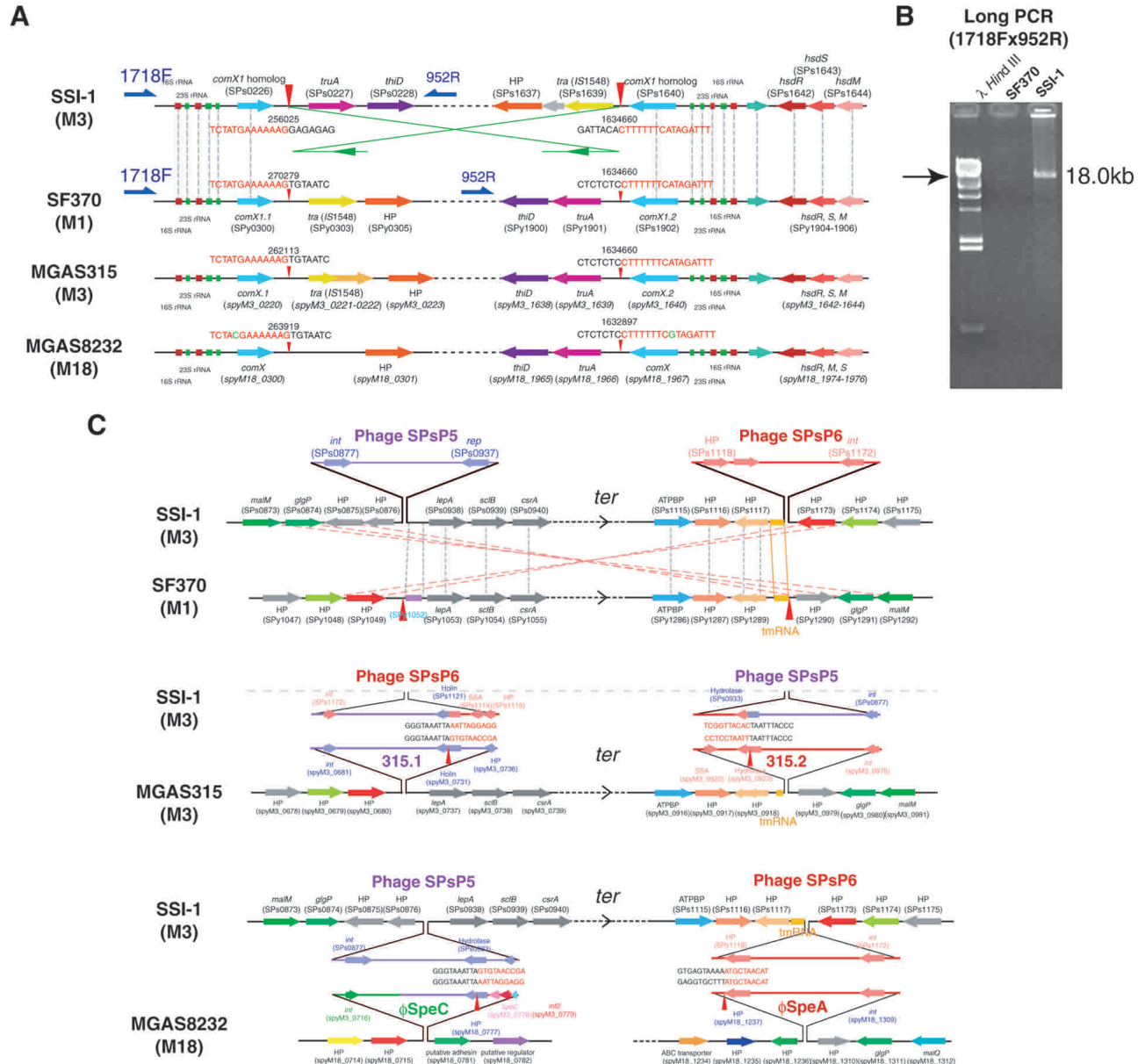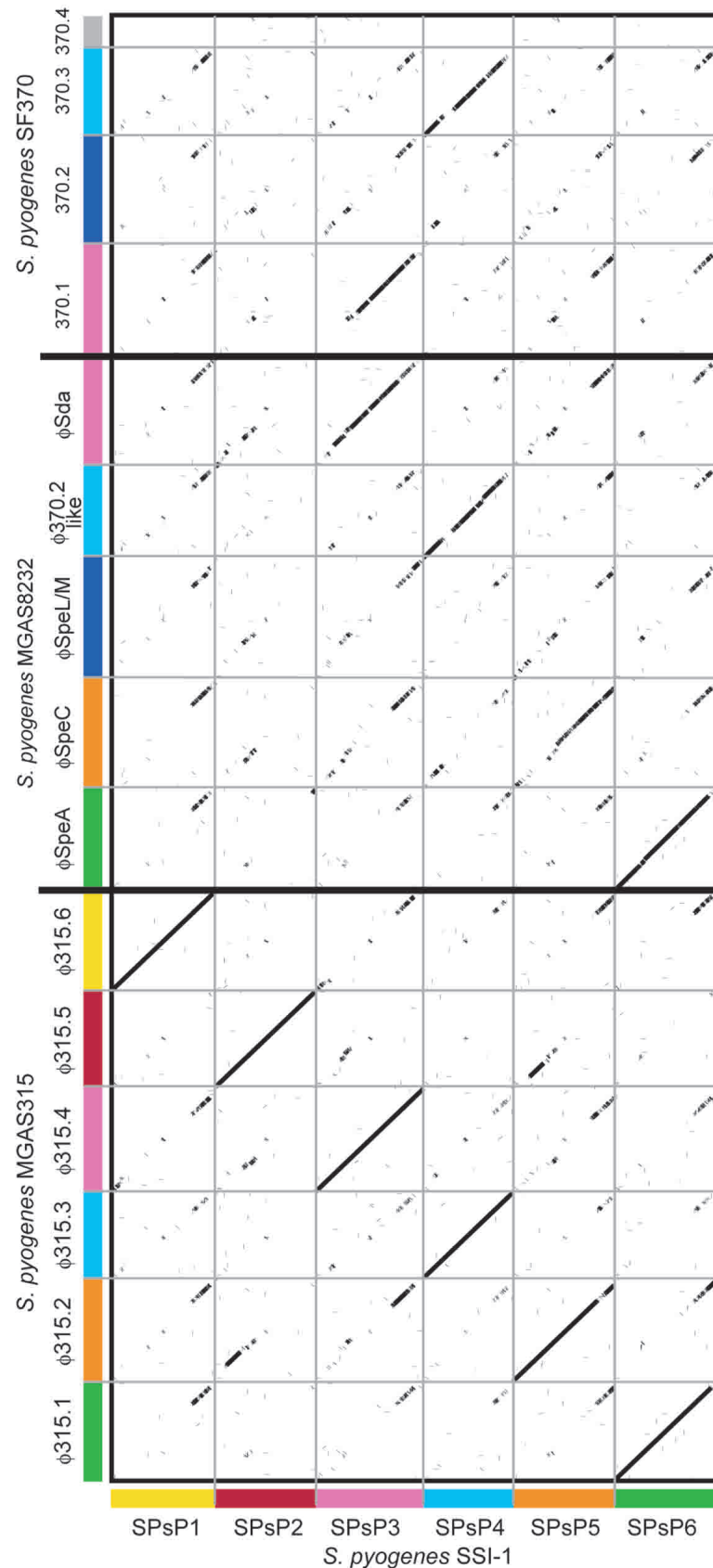
Figure 4 Expanded views of the *rrn-comX* genomic rearrangement site between strain SSI-1 and other GAS strains. (*A*) Comparison of *rrn-comX* region in strain SSI-1 and strains SF370, MGAS315, or MGAS8232. HP, hypothetical protein. Red arrowheads indicate the rearrangement break-point in strain SSI-1, and green lines and arrows indicate the orientation of the rearrangement in strain SSI-1 compared with strain SF370. Gray dotted lines indicate orthologous genes that are located in relatively identical positions. (*B*) Long-PCR analysis of *rrn-comX* rearrangement site. Long-PCR was performed with PCR primers 1718F and 952R, and analyzed by agarose gel electrophoresis. (*C*) Expanded views of the phage rearrangement site between strain SSI-1 and other GAS strains. The gene order and sequences of the rearrangement sites of strain SSI-1 are compared with those of strains SF370, MGAS315, and MGAS8232 with MUMmer and CLUSTALW programs. Arrows represent the relative orientation of the genes. HP, hypothetical protein. Red arrowheads: The rearrangement break-point in strain SSI-1. Gray and pink dotted lines indicate the orthologous genes that are located in relatively identical positions or in the inverted chromosomal regions, respectively.

one major group (including 12 of 15 holin homologs in GAS phages) and two minor groups. The integrase genes, however, are divergent and are classified into five major lineages based on their amino acid sequences (Suppl. Fig. B). In addition, the 12 homologous holin genes are highly or completely conserved, despite more than 100 members of the holin gene family having been identified, defining more than 30 orthologous groups (Wang et al. 2000). The sequence similarities of

the last one-fourth of the phage region in GAS phages may trigger not only the mechanism of homologous recombination in the GAS chromosome, but also contribute to the diversity of the distribution of superantigens or other virulence factors encoded in phage regions. Collectively, streptococcal phages are important contributors to genetic diversity, not only by inducing genomic rearrangements, but also by being a source of new genes.

Figure 5 Dot-plot analysis of the six phages in strain SSI-1 with those in strain SF370 (M1), MGAS8232 (M18), and MGAS315 (M3). All phage sequences were extracted from each genome sequence and aligned with the integrase genes as the starting position. Homologous phage sequences are indicated in the same axis color. The width of each column corresponds to relative phage length.

GAS Pathogenicity and Genomic Rearrangement

We also examined whether the genomic rearrangement is specific for strain SSI-1 or common to GAS strains. We analyzed the rearrangement in the *rrn-comX* region (Suppl. Fig. A) by long-PCR with clinical isolates from diverse localities. We divided these test strains into three groups: (A) strains isolated before 1985, (B) strains isolated from STSS patients after 1990 in Japan, and (C) strains isolated from non-STSS patients after 1990 in Japan. As shown in Supplemental Figure A, the typical genomic rearrangement in the *rrn-comX* region was observed in 81% (35/43) of strains isolated from STSS patients after 1990. In contrast, the genetic rearrangement was observed in only 25% (7/28) of strains isolated before 1985. Moreover, 57% of clinical strains (29/51) isolated from non-STSS patients after 1990 showed the rearrangement in this region. These results suggest that the genomic rearrangement in this region has occurred in many GAS strains with different M serotypes in the past. We note that almost all strains of serotypes M1, M3, and M28 from STSS patients have the rearrangement in the *rrn-comX* region. These observations may indicate that the X-shaped rearrangement found in strain SSI-1 was present in the recent past. This genomic rearrangement occurs in different M serotypes of GAS strains in recent isolates, and the phenomenon is becoming increasingly evident. Surprisingly, the increase in X-rearranged strains seems to coincide with the recent resurgence of rheumatic fever and severe invasive infections in Japan. It is possible that the resurgence of severe GAS infection is related to the clonal expansion of such invasive strains.

In conclusion, our results provide new insight into phage evolution as well as a mechanism for the origin of new clones and genetic variation in GAS. In a comparison of the genomes of three different serotype strains and the same serotype strains of GAS, we observed that the genetic variation of GAS was affected, not only by the integration of phages, but also by a large-scale genomic rearrangement of its chromosome. In addition, the genetic diversity of the GAS chromosome in phage regions might be induced by "virulence cassette" shuffling followed by genomic rearrangement. There remain more questions than answers regarding how these genomic rearrangements affect the gene expression and pathogenicity of GAS. We believe that this genetic information will assist in gaining

Table 2. Comparison of sSNP and cSNP in M3 Invasive Strains Isolated in Japan and the United States

ORF number						
SSI-1	MGAS315	Gene description	Amino acids Length[a]	sSNPs[b]	cSNPs[c]	Amino acids Changes (per gene)
*sps*0659	*spyM3*_1583	putative domain protein	96	17	128	69
*sps*0553	*spyM3*_1210	hypothetical protein (phage associated)	140	17	134	66
*sps*0135	*spyM3*_0131	hypothetical protein	109	21	103	52
*sps*0510	*spyM3*_1353	hypothetical protein	35	2	11	7
*sps*0525	*spyM3*_1246	hypothetical protein (phage associated)	84	8	21	15
*sps*1399	*spyM3*_0457	conserved hypothetical protein	50	2	7	3
*sps*1398	*spyM3*_0457	conserved hypothetical protein	190	0	7	4
*sps*0449	*spyM3*_1416	hypothetical protein (phage associated)	146	1	2	2
*sps*0776	*spyM3*_1089	putative transcriptional regulator	174	2	2	2
*sps*0450	*spyM3*_1415	hypothetical protein (phage associated)	98	0	1	1
*sps*0119	*spyM3*_0117	putative V-type Na+-ATPase subunit E	105	0	1	1
*sps*1742	*spyM3*_1744	putative transcription regulator	281	0	2	2
*sps*0857	*spyM3*_1001	hypothetical protein	150	0	1	1
*sps*0664	*spyM3*_1198	putative arginine repressor	158	0	1	1
*sps*1848	*spyM3*_1852	putative UDP-glucose 6-dehydrogenase	403	0	3	2
*sps*1358	*spyM3*_0496	putative proton-translocating ATPase, delta subunit	179	0	1	1
*sps*0219	*spyM3*_0213	putative D,D-carboxypeptidase, penicillin-binding protein	375	2	2	1
*sps*1618	*spyM3*_0241	putative cytoplasmic membrane protein	186	0	1	1
*sps*0043	*spyM3*_0041	50S ribosomal protein L4	208	0	1	1
*sps*1128	*spyM3*_0724	hypothetical protein	206	0	1	1
*sps*1645	*spyM3*_1645	putative response regulator (salivaricin regulon)	202	0	1	1
*sps*1124	*spyM3*_0728	hypothetical protein (phage associated)	211	0	1	1
*sps*1615	*spyM3*_0244	putative two-component response regulator (CsrR/CovR)	229	0	1	1
*sps*0879	*spyM3*_0977	putative repressor protein	237	0	1	1
*sps*0322	*spyM3*_1544	putative transcriptional pleiotropic repressor	261	0	1	1
*sps*0778	*spyM3*_1087	putative formate dehydrogenase	295	0	1	1
*sps*1442	*spyM3*_0413	putative prolipoprotein diacylglycerol transferase	260	0	1	1
*sps*0448	*spyM3*_1417	hypothetical protein (phage associated)	274	0	1	1
*sps*0385	*spyM3*_1482	putative tagatose 1,6-diphosphate aldolase	326	0	1	1
*sps*0455	*spyM3*_1409	putative sdalpha deoxyribonuclease	329	3	1	1
*sps*1191	*spyM3*_0662	putative acetoin dehydrogenase (TPP-dependent) beta chain	360	0	1	1
*sps*1858	*spyM3*_1862	putative transposase	329	0	1	1
*sps*1068	*spyM3*_0868	putative lipoprotein	351	0	1	1
*sps*0010	*spyM3*_0009	hypothetical protein	429	0	1	1
*sps*0535	*spyM3*_1326	conserved hypothetical protein	430	0	1	1
*sps*0743	*spyM3*_1121	putative structural protein (phage associated)	423	0	1	1
*sps*1420	*spyM3*_0435	putative peptidoglycan branched peptide synthesis protein	409	0	1	1
*sps*0358	*spyM3*_1509	conserved hypothetical protein	487	1	1	1
*sps*0669	*spyM3*_1193	conserved hypothetical protein	498	0	1	1
*sps*0835	*spyM3*_1025	putative PBP 5 synthesis repressor	484	0	1	1
*sps*0974	*spyM3*_0774	conserved hypothetical protein	511	0	1	1
*sps*1047	*spyM3*_0847	putative ABC transporter (ATP-binding protein)	515	0	1	1
*sps*1067	*spyM3*_0867	putative sugar ABC transporter (ATP-binding protein)	511	0	1	1
*sps*1454	*spyM3*_0401	putative signal recognition particle (docking protein)	517	0	1	1
*sps*1614	*spyM3*_0245	putative two-component sensor histidine kinase (CsrS/CovS)	501	0	1	1
*sps*0772	*spyM3*_1093	putative heavy metal-transporting ATPase	621	0	1	1
*sps*0790	*spyM3*_1073	putative competence protein	748	0	1	1
*sps*1228	*spyM3*_0625	putative DNA topoisomerase IV, subunit C	822	0	1	1
*sps*0077	*spyM3*_0076	putative DNA-dependent RNA polymerase beta prime subunit	1199	0	1	1
*sps*0576	*spyM3*_1285	putative beta-galactosidase	1169	0	1	1
*sps*1292	*spyM3*_0562	putative carbamoylphosphate synthetase	1059	0	1	1

[a]Amino acids length in strain SSI-1.
[b]SNPs in orthologous genes of strain SSI-1 were compared with those of strain MGAS315. sSNP and [c]cSNP indicate silent SNP and coding SNP (bp).

new insights into the molecular pathogenesis of GAS infection.

METHODS

Bacterial Strains

Streptococcus pyogenes strain SSI-1 was isolated in Japan in 1994 from a patient with STSS. This strain can induce necrotizing fasciitis in a mouse infection model (Okamoto et al. 2003). It produces pyrogenic exotoxins A (SpeA) and B (SpeB), and its serotype was determined to be M3 by sequence analysis of the *emm* gene of the organism (Murakami et al. 2002). Other clinical isolates and laboratory strains were isolated at Tokyo Women's Medical School, Nagoya University, Saga Prefectural Institute of Public Health, Ehime Prefectural Institute of Public Health, Osaka Prefectural Institute of Public Health, and the National Institute of Infectious Diseases, Japan. These strains were selected from our culture collection and checked for gene content, the production of putative exotoxins, *emm*

Table 3. Comparison of Putataive and Predicted Virulence Factors in Four Strains of GAS Genomes

	Strains				
Gene description	SSI-1 (*sps*)[a]	SF370 (*spy*)[a]	MGAS315 (*spyM3_*)[a]	MGAS8232 (*spyM18_*)[a]	Location[b]
Proteinase					
C3 degrading proteinase	0269	1851	1598	1914	Ch
Exfoliative toxin-like protein	1220	0918	0632	0975	Ch
IdeS/Sib38	1270	0861	0583	0921	Ch
Streptokinase A	1700	1979	1698	2042	Ch
C5a peptidase	1724	2010	1726	2074	Ch
SpeB/Cysteine proteinase	1739	2039	1742	2099	Ch
Serine proteinase	1860	2216	1864	2256	Ch
Adhesin					
GRAB	0828	1357	1032	1369	Ch
SclA	(1704)[c]	1983	1702/1703	—	Ch
SclB	0939	1054	0738	1029	Ch
M protein	1725	2018	1727	2076	Ch
Fibronectin-binding protein F2-like	0106	—	0104	0132	Ch
Fibronectin-binding protein (Fba)-like	—	2009	—	—	
Internalin-like protein	0825	1361	1035	1373	Ch
Laminin-binding protein	1723	2007	1725	2073	Ch
Hemolysin					
SLO	0132	0167	0130	0165	Ch
SLS	1366	0746	0488	0807	Ch
CAMP	1104	1273	0905	1221	Ch
Hemolysin (*hylA1*)	0709	1497	1153	1515	Ch
Hemolysin homolog (*hylX*)	1583	0378	0276	0432	Ch
Hemolysin homolog (*hyl III*)	1016	1159	0815	1119	Ch
Capsule synthesis and degradation					
Hyalunonate synthase (*hasA*)	1847	2200	1851	2236	Ch
Hyaluronidase	0567	1600	1294	1606	Ch
Hyaluronidase	0447	—	1418	—	Ph
Hyaluronidase	1127	—	1418	—	Ph
Hyaluronidase	0763	—	1108	—	Ph
Hyaluronidase	0927	—	1101	—	Ph
Hyaluronidase (*hylP1*)	0648	0701	1214	—	Ph
Hyaluronidase (*hylP2*)	—	0997	—	—	Ph
Hyaluronidase (*hylP3*)	—	1445	—	—	Ph
Hyaluronidase (*hyl*)	—	—	—	0385	Ph
Hyaluronidase (*hyl*)	—	—	—	0770	Ph
Hyaluronidase (*hyl*)	—	—	—	1254	Ph
Hyaluronidase (*hyl*)	—	—	—	1455	Ph
DNase					
Mitogenic factor (*mf*)	1743	2043	1745	2104	Ch
Mitogenic factor-like	0770	—	1095	—	Ph
MF2 (*mf2*)	—	0712	—	0779	Ch
MF3 (*mf3*)	—	1436	—	—	Ph
Sda	0455	—	1409	1746	Ph
Superantigen					
SpeA	0560	—	1301	0393	Ph
SpeC	—	0861	—	0778	Ph
SpeG	0161	0212	0155	0201	Ch
SpeH	—	1008	—	—	Ph
SpeI	—	1007	—	—	Ph
SpeJ	—	—	—	0919	Ch
SpeK	0657	—	1205	—	Ph
SpeL[d]	0657	—	1205	1238	Ph
SpeM	—	—	—	1239	Ph
SmeZ	—	1998	—	2064	Ch
SSA	1119	—	0920	—	Ph
Immunoreactive antigen					
Immunogenic secreted protein	0305	1801	1562	1870	Ch
Myosin crossreactive antigen	1525	0470	0332	0512	Ch
Immunogenic secreted protein	1728	2025	1731	2082	Ch
Regulators					
SpeB protease transcription regulator (*rgg*)	1742	2042	1744	2103	Ch
RopA (trigger factor)	0232	1896	1634	1961	Ch
M protein transacting positive regulator	1726	2019	1728	2077	Ch
CsrR/S (CovR/S)[e]	1614/1615	0337	0244/0245	328/0329	Ch
RofA	0163	0216	0157	0205	Ch

(continued)

Table 3. *Continued*

Gene description	SSI-1 (*sps*)[a]	SF370 (*spy*)[a]	MGAS315 (*spyM3_*)[a]	MGAS8232 (*spyM18_*)[a]	Location[b]
Bacteriocin					
Salivaricin precursor	1650	1915	1652	1983	Ch
SalK homolog	1646	(1910)[c]	1646	1979	Ch
Others					
Sic	—	2016	—	—	Ch

[a]Numbers in this table indicate ORF numbers annotated in GenBank Database (SF370:AE004092, MGAS8232: AR009949 and MGAS315: AE014074).
[b]"Location" indicates the genes located in backbone chromosome (Ch) or in the phage sequences (Ph).
[c]ORF number in parentheses indicates a disrupted gene by frameshift mutation.
[d]Nucleotide sequence of SpeL gene in GenBank (NP_438166) is different from that annotated in MGAS8232 (AAL97848.1)
[e]CsrR was not annotated in GenBank (AE004092), however, nucleotide sequence for CsrR is found in SF370 genome.

genotype, patient history, and the area of isolation recorded (Suppl. Table B).

Restriction Analyses by PFGE and Construction of the Physical Map of Strain SSI-I

We chose four restriction enzymes to construct the physical map of the strain SSI-1 genome. The restriction enzymes *Sfi*I and *Sgr*AI were selected according to the study of chromosomal DNA analysis of strain SF370 (Suvorov and Ferretti 1996). *Fse*I and *Asc*I were selected according to the nucleotide sequence of strain SF370 (GenBank accession no. AE004092), because they respectively result in only four or two fragments from the SF370 genome and three or three fragments from the SSI-1 genome. A total of 10 probes was selected from the predicted ORFs in the SF370 sequence (*leuS*, *csrS*, *dinG*, *fbp54*, *cfa*, *dnaX*, *pbp2X*, *gyrA*, *speB*, and *recA*), and these gene fragments were used for hybridization

with PFGE restriction fragments (Suppl. Fig. C). Southern blotting was performed with digoxigenin-labeled probes according to the manufacturer's instructions (Roche Diagnostics).

Genome Sequencing and Annotation

The initial stage of sequencing was performed using whole-genome random shotgun methods with sheared chromosomal DNA from strain SSI-1. We constructed a pUC18-based library containing 1–2 kb and 4–5 kb inserts, and sequenced 48,000 clones (12.6-fold coverage) with Big-Dye terminator chemistry and an ABI 3700 sequencer (Applied Biosystems) and with ET-Dye terminator chemistry and a MegaBACE 1000 sequencer (Amersham Biosciences). The sequence was assembled using Phred/Phrap/Consed (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998). Gaps in the sequence were filled by direct PCR sequencing, using primers con-
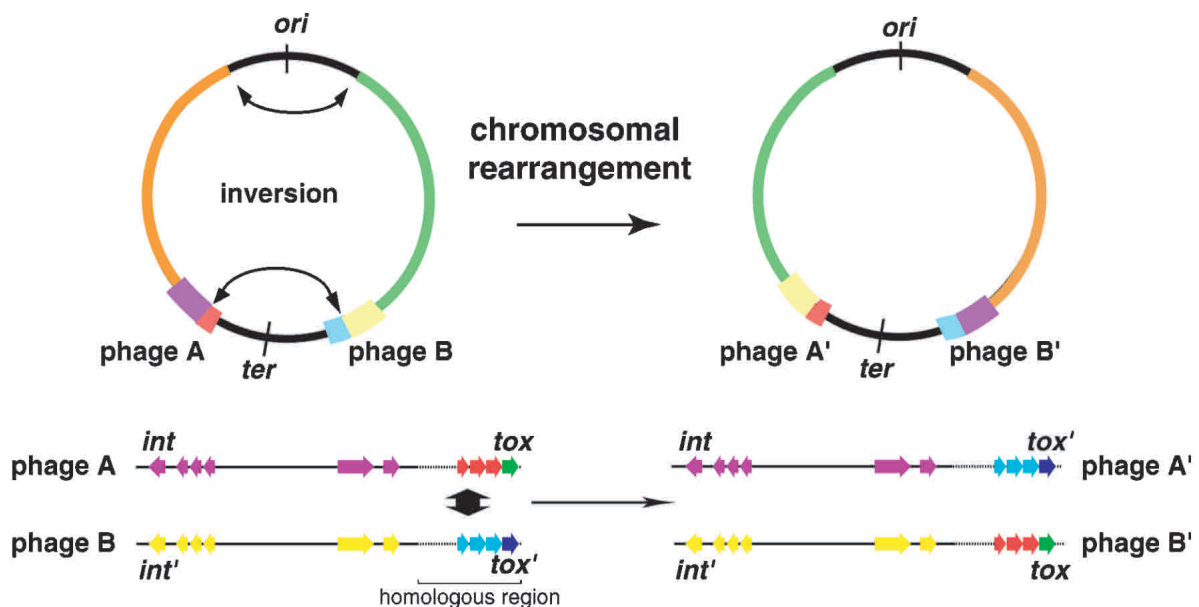
Figure 6 Schematic diagram of phage-related rearrangements by chromosomal inversion. Two phages integrated equidistant from the *ter* region exchange their virulent cassettes. *int*, integrase gene of phage region. *tox* indicates superantigen, mitogenic factor, or streptodornase genes.

structed to anneal to each end of neighboring contigs. Finally, the entire sequence was estimated to have an error rate of less than 1 per 10,000 bases (Phrap score ≥40). The final assembly was verified by the comparison of restriction-enzyme digest patterns using pulsed-field gel electrophoresis and Southern blot analysis with 40 specific genes as probes as described above. To verify and determine the assembled sequences, a total of 89 primer sets was constructed to cover whole chromosomal DNA of strain SSI-1 at the unique flanking sequence, and 18–25 kb of long-PCR was performed by the LA-PCR method (Takara). Large repeated elements in the genome (700–6000 bp) such as the 16S and 23s rRNA operons (*rrn*) were amplified from chromosomal DNA using Ex-Taq or LA-Taq (Takara), sequenced, and assembled independently, as described above.

ORFs >90 bp were identified and annotated separately using Genome Gambler version 1.47 (Sakiyama et al. 2000) and GLIMMER 2 (Delcher et al. 1999a; www.tigr.org). The predicted ORFs were reviewed individually by a manual search for start codons on the basis of ribosomal-binding motifs. ORFs were further compared across a nonredundant protein database using BLASTP software (version 2.2.3; Altschul et al. 1997). Functional motifs and the domains of proteins were identified by searches against Prosite, Blocks, and Pfam database (http://www.sanger.ac.uk/Software/Pfam/search.shtml) and phi-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). Protein localization and transmembrane domains were predicted by combining PSORT with the rule set for gram-positive bacteria (http://psort.nibb.ac.jp/), and the SOSUI/SOSUI signal program (http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html). Cell-wall attachment motifs (LPXTG) and secreted protein motifs (sortase recognition motif) were identified with the original Perl script (version 5.6). Functional categories based on the analysis of clusters of orthologous genes were assigned by using COGnitor (Tatusov et al. 2001) (http://www.ncbi.nlm.nih.gov/COG/xognitor.html). Transfer RNA genes were identified using tRNAscan-SE (Lowe and Eddy 1997).

Comparative Genomes of Four GAS Strains and Determination of Genomic Rearrangement

The genomic sequences of other *S. pyogenes* strains (SF370:AE004092, MAS8232: AR009949, and MGAS315: AE014074), *S. pneumoniae* (R6: AE007317 and TIGR4: AE005672), and *S. agalactiae* 2603V/R (AE009948) were obtained through the Web site of the National Center of Biological Information (NCBI, http://www.ncbi.nlm.nih.gov). Alignment of the complete genomic sequences of all four bacterial strains was accomplished with the MUMmer program (Delcher et al. 1999b), CONSERV (Goto et al. 2000), and FASTA3 (Pearson 1990). The predicted ORFs from these strains were aligned using BLASTP (maximum E-value = 10^{-5}). Single nucleotide polymorphism (SNP) analysis was also performed with CLUSTALW (Thompson et al. 1994) for each orthologous gene. To compare the lineage of phage genes in the four GAS strains, predicted amino acids sequences in the four GAS strains were compared with CLUSTALW (DNA Data Bank of Japan), and the phylogenetic tree was constructed using the neighbor-joining method (Kumar et al. 2001) on MEGA2.1 software (http://www. megasoftware.net/). Rearrangement sites were verified by the LA-PCR method (Takara) using site-specific primer pairs (*rrn-comX* region: 5′-TTGTCAAGAGCT TACTGACTGAGGCGACTGGGAC-3′ and 5′-AGCGATACTA GATGCAAAAGTACAGCCTGCGCC- 3′). Briefly, PCR was performed as follows: 95°C for 1 min for one cycle, 98°C for 10 sec and 68°C for 20 min for 30 cycles, and 72°C for 10 min for one cycle. The amplified fragments were separated on 0.8% agarose gel electrophoresis and visualized by ethidium bromide staining.

REFERENCES

Alm, R.A., Ling, L.S., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., et al. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397:** 176–180.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Bentley, R.W., Leigh, J.A., and Collins, M.D. 1991. Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences. *Int. J. Syst. Bacteriol.* **41:** 487–494.

Beres, S.B., Sylva, G.L., Barbian, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., et al. 2002. Genome sequence of a serotype M3 strain of group A *Streptococcus*: Phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc. Natl. Acad. Sci.* **99:** 10078–10083.

Bessen, D.E. and Kalia, A. 2002. Genomic localization of a T serotype locus to a recombinatoria zone encoding extracellular matrix-binding proteins in *Streptococcus pyogenes*. *Infect. Immun.* **70:** 1159–1167.

Cole, S.T. 1998. Comparative mycobacterial genomics. *Curr. Opin. Microbiol.* **1:** 567–571.

Cone, L.A., Woodard, D.R., Schlievert, P.M., and Tomory, G.S. 1987. Clinical and bacteriologic observations of a toxic shock-like syndrome due to *Streptococcus pyogenes*. *N. Engl. J. Med.* **317:** 146–149.

Cunningham, M.W. 2000. Pathogenesis of group A streptococcal infections. *Clin. Microbiol. Rev.* **13:** 470–511.

Delcher, A.L., Harmon, D., Kasif, S., White, O., and Salzberg, S.L. 1999a. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27:** 4636–4641.

Delcher, A.L., Kasif, S., Fleischmann, R.D., Peterson, J., White, O., and Salzberg, S.L. 1999b. Alignment of whole genomes. *Nucleic Acids Res.* **27:** 2369–2376.

Eisen, J.A., Heidelberg, J.F., White, O., and Salzberg, S.L. 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. *Genome Biol.* **1:** RESEARCH0011.

Enright, M.C., Spratt, B.G., Kalia, A., Cross, J.H., and Bessen, D.E. 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between emm type and clone. *Infect. Immun.* **69:** 2416–2427.

Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8:** 186–194.

Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8:** 175–185.

Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton, S., et al. 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci.* **98:** 4658–4663.

Gordon, D., Abajian, C., and Green, P. 1998. Consed: A graphical tool for sequence finishing. *Genome Res.* **8:** 195–202.

Goto, N., Kurokawa, K., and Yasunaga, T. 2000. CONSERV: A tool for finding exact matching conserved sequences in biological sequences. *Genome Informatics* **11:** 307–308.

Holm, S.E., Kohler, W., Kaplan, E.L., Schlievert, P.M., Alouf, J.E., Stevens, D.L., and Kotb, M. 1997. Streptococcal toxic shock syndrome (STSS). An update: A roundtable presentation. *Adv. Exp. Med. Biol.* **418:** 193–199.

Hoskins, J., Alborn Jr., W.E., Arnold, J., Blaszczak, L.C., Burgett, S.,

DeHoff, B.S., Estrem, S.T., Fritz, L., Fu, D.J., Fuller, W., et al. 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* **183:** 5709–5717.

Hughes, D. 2000. Evaluating genome dynamics: The constraints on rearrangements within bacterial genomes. *Genome Biol.* **1:** REVIEWS0006.

Inagaki, Y., Myouga, F., Kawabata, H., Yamai, S., and Watanabe, H. 2000. Genomic differences in *Streptococcus pyogenes* serotype M3 between recent isolates associated with toxic shock-like syndrome and past clinical isolates. *J. Infect. Dis.* **181:** 975–983.

Janulczyk, R. and Rasmussen, M. 2001. Improved pattern for genome-based screening identifies novel cell wall-attached proteins in gram-positive bacteria. *Infect. Immun.* **69:** 4019–4026.

Johnson, D.R., Wotton, J.T., Shet, A., and Kaplan, E.L. 2002. A comparison of group A streptococci from invasive and uncomplicated infections: Are virulent clones responsible for serious streptococcal infections? *J. Infect. Dis.* **185:** 1586–1595.

Kaplan, E.L. 1991. The resurgence of group A streptococcal infections and their sequelae. *Eur. J. Clin. Microbiol. Infect. Dis.* **10:** 55–57.

Kumar, S., Tanura, K., Jakobsen, I.B., and Nei, M. 2001. MEGA2: Molecular evolutonaty genetics analysis software. *Bioinformatics* **17:** 1244–1245.

Kuzminov, A. and Stahl, F.W. 1999. Double-strand end repair via the RecBC pathway in *Escherichia coli* primes DNA replication. *Genes & Dev.* **13:** 345–356.

Leach, D.R., Okely, E.A., and Pinder, D.J. 1997. Repair by recombination of DNA containing a palindromic sequence. *Mol. Microbiol.* **26:** 597–606.

Liu, G.R., Rahn, A. Liu, W.Q., Sanderson, K.E., Johnston, R.N., and Liu, S.L. 2002. The evolving genome of *Salmonella enterica* serovar Pullorum. *J. Bacteriol.* **184:** 2626–2633.

Liu, S.L. and Sanderson, K.E. 1995. Rearrangements in the genome of the bacterium *Salmonella typhi. Proc. Natl. Acad. Sci.* **92:** 1018–1022.

Lowe, T.M. and Eddy, S.R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25:** 955–964.

Maeder, D.L., Weiss, R.D., Dunn, D.M., Cherry, J.L., Gonzalez, J.M., DiRuggiero, J., and Robb, F.T. 1999. Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* **152:** 1299–1305.

Makino, S. and Suzuki, M. 2001. Bacterial genomic reorganization upon DNA replication. *Science* **292:** 803a.

McShan, W.M., Tang, Y.F., and Ferretti, J.J. 1997. Bacteriophage T12 of *Streptococcus pyogenes* integrates into the gene encoding a serine tRNA. *Mol. Microbiol.* **23:** 719–728.

Murakami, J., Kawabata, S., Terao, Y., Kikuchi, K., Totsuka, K., Tamaru, A., Katsukawa, C., Moriya, K., Nakagawa, I., Morisaki, I., et al. 2002. Distribution of emm genotypes and superantigen genes of *Streptococcus pyogenes* isolated in Japan from 1994 to 1999. *Epidemiol. Infect.* **128:** 397–404.

Murase, T., Suzuki, R., Osawa, R., and Yamai, S. 1999. Characteristics of *Streptococcus pyogenes* serotype M1 and M3 isolates from patients in Japan from 1981 to 1997. *J. Clin. Microbiol.* **37:** 4131–4134.

Musher, D.M., Hamill, R.J., Wright, C.E., Clarridge, J.E., and Ashton, C.M. 1996. Trends in bacteremic infection due to *Streptococcus pyogenes* (group A streptococcus), 1986–1995. *Emerg. Infect. Dis.* **2:** 54–56.

Musser, J.M., Hauser, A.R., Kim, M.H., Schlievert, P.M., Nelson, K., and Selander, R.K. 1991. *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: Clonal diversity and pyrogenic exotoxin expression. *Proc. Natl. Acad. Sci.* **88:** 2668–2672.

Navarre, W.W. and Schneewind, O. 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. *Microbiol. Mol. Biol. Rev.* **63:** 174–229.

Okamoto, S., Kawabata, S., Nakagawa, I., Okuno, Y., Goto, T., Sano, K., and Hamada, S. 2003. Influenza A virus-infected hosts boost an invasive type of *Streptococcus pyogenes* infection in mice. *J. Virol.* **77:** 4104–4112.

Pearson, W.R. 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods Enzymol.* **183:** 63–98.

Perals, K., Capiaux, H., Vincourt, J.B., Louarn, J.M., Sherratt, D.J., and Cornet, F. 2001. Interplay between recombination, cell division and chromosome structure during chromosome dimer resolution in *Escherichia coli. Mol. Microbiol.* **39:** 904–913.

Podbielski, A., Woischnik, M., Leonard, B.A., and Schmidt, K.H. 1999. Characterization of nra, a global negative regulator gene in group A streptococci. *Mol. Microbiol.* **31:** 1051–1064.

Radding, C.M. 1988. In *Genetic recombination* (eds. R. Kucherlapati and G.R. Smith) pp.193–229. American Society for Microbiology, Washington, DC.

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Utterback, T., Berry, K., et al. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28:** 1397–1406.

Sakiyama, T., Takami, H., Ogasawara, N., Kuhara, S., Kozuki, T., Doga, K., Ohyama, A., and Horikoshi, K. 2000. An automated system for genome analysis to support microbial whole-genome shotgun sequencing. *Biosci. Biotechnol. Biochem.* **64:** 670–673.

Segall, A.M. and Roth, J.R. 1994. Approaches to half-tetrad analysis in bacteria: Recombination between repeated, inverse-order chromosomal sequences. *Genetics* **136:** 27–39.

Smoot, J.C., Barbian, K.D., Van Gompel, J.J., Smoot, L.M., Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M., Porcella, S.F., Parkins, L.D., et al. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks. *Proc. Natl. Acad. Sci.* **99:** 4668–4673.

Stevens, D.L. 1999. The flesh-eating bacterium: What's next? *J. Infect. Dis.* **179:** 366–374.

Suvorov, A.N. and Ferretti, J.J. 1996. Physical and genetic map of an M type 1 strain of *Streptococcus pyogenes. J. Bacteriol.* **178:** 5546–5549.

Suvorov, A.N. and Ferretti, J.J. 2000. Replication origin of *Streptococcus pyogenes*, organization and cloning in heterogous systems. *FEMS Microbiol. Lett.* **185:** 293–297.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:** 22–28.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., et al. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae. Science* **293:** 498–506.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., et al. 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae. Proc. Natl. Acad. Sci.* **99:** 12391–12396.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Tillier, E.R. and Collins, R.A. 2000a. The contributions of replication orientation, gene direction, and signal sequences to base-composition asymmetries in bacterial genomes. *J. Mol. Evol.* **50:** 249–257.

———. 2000b. Genome rearrangement by replication-directed translocation. *Nat. Genet.* **26:** 195–197.

Wang, I.N., Smith, D.L., and Young, R. 2000. Holins: The protein clocks of bacteriophage infections. *Annu. Rev. Microbiol.* **54:** 799–825.

Zivanovic, Y., Lopez, P., Philippe, H., and Forterre, P. 1997. Pyrococcus genome comparison evidences chromosomal shuffling-driven evolution. *Nucleic Acids Res.* **30:** 1902–1910.

WEB SITE REFERENCES

http://www.ncbi.nlm.nih.gov/BLAST/; National Center for Biotechnology Information Web site for BLAST search.

http://www.ncbi.nlm.nih.gov/COG/xognitor.html; Phylogenetic classification of proteins.

http://www.megasoftware.net/; Phylogenetic tree analysis.

http://sosui.proteome.bio.tuat.ac.jp/sosuiframe0.html; Prediction program of membrane proteins.

http://psort.nibb.ac.jp; Prediction program of protein localizations.

http://genome.gen-info.osaka-u.ac.jp/bacteria/spyo/; *S. pyogenes* genome project Web site in Genome Information Research Center, Osaka University.

http://www.sanger.ac.uk/Software/Pfam/search.shtml; Sanger Center Web site for the Pfam database.

www.tigr.org; The Institution of Genomic Research Web site.