# Genetic variation in group A streptococci

# Genetic Variation in Group A Streptococcus

David J McMillan[1,2], Kadaba S Sriprakash[2] and Gursharan S Chhatwal[1]

[1]Department of Microbial Pathogenesis, Helmholtz Centre for Infection Research, Braunschweig, Germany

[2]Bacterial Pathogenesis Laboratory, Queensland Institute of Medical Research, Brisbane, Australia

**Corresponding author**

Prof. G.S. Chhatwal
Division of Microbiology
Helmholtz Centre for Infection Research
Inhoffenstraße 7, 38124 Braunschweig, Germany
Phone +49 531 6181 4400
Fax +49 531 6181 4499
eMail gsc@helmholtz-hzi.de

**Summary**

Group A streptococcus is a bacterial pathogen responsible a range of human disease that vary in their clinical manifestations and severity. While numerous virulence factors have been described in this species, the way these factors interact to promote different streptococcal diseases is less clear. In order to identify multifactorial relationship between GAS and human host, novel high throughput technology such as microarray is necessary. We have performed comparative studies using custom-designed virulence arrays to enhance our understanding of high degree of genotypic variation that occurs in streptococci. This study has pointed to mobile genetic elements as the major agents that promote variation. Our results show that multiple combinations of genes might bring about the similar clinical picture. This then adds a further complexity to unravel the intricate relationship between the pathogen and the host.

Keywords: group A streptococcus, *Streptococcus pyogenes*, comparative genomics

## Introduction

*Streptococcus pyogenes* (Group A streptococci, GAS) is a human-specific bacterial pathogen capable of causing a wide range of diseases that vary in their clinical symptoms and severity (Cunningham, 2000). These include pharyngitis, a self-limiting condition estimated to affect up to 600 million people each year (Carapetis, et al., 2005), skin diseases such as scarlet fever, impetigo and pyoderma, and post-infectious Rheumatic Fever (RF), Rheumatic Heart Disease (RHD) and post-streptococcal glomerulonephritis (PSGN). Once major causes of disease in western countries, the incidence of RF/RHF and scarlet fever have declined from their peaks in the 19$^{th}$ Century (Denny, 2000). Normally ascribed to improvements in living conditions, changes in the bacterial population structure of GAS may have also contributed the reduced incidence of these diseases (Shulman, et al., 2006).

GAS are subcategorised on the basis of variation in the M-protein, a major virulence determinant. The importance of M-typing lies in studies demonstrating associations between certain M-types and disease. M1, M3, M6 M18 and M24 are considered rheumatogenic, while and M1, M12, M55 and M57 are nephritogenic (Cunningham, 2000). After decades of decline in the US, an outbreak of RF was reported in Salt Lake City in 1985 (Veasy, et al., 1987). In the same decade, a resurgence of severe streptococcal invasive disease, defined by the ability to isolates GAS from normally sterile body sites, was also reported (Cone, et al., 1987, Kiska, et al., 1997). While predominantly associated with certain M1 and M3 strains, M89, M28 and other M-types are also commonly isolated from patients presenting with invasive disease (Vlaminckx, et al., 2005). The outbreak of RF, and the resurgence of serious

streptococcal invasive diseases caught the medical and scientific fraternity off guard and raise the question about the epidemiological and genetic basis for these new outbreaks.

One view is that the new strains with an altered genetic endowment that result in altered virulence characteristics, have emerged. Many virulence factors that contribute to GAS pathogenesis have been described. These include adhesins that bind to components of the host extracellullar matrix, and proteins that contribute to immune system avoidance modulation and evasion, bacterial transmission, toxins, and DNAses (Cunningham, 2000). Less obvious is how these factors contribute to influence the outcome of an infection i.e. what factors are important for simple pharyngitis, and which factors are critical for RHD, PSGN or invasive disease. Many epidemiological studies have attempted to answer these question by identifying common genetic factors that can be used to classify GAS based on disease causing propensity (Buchanan, et al., 2006, Chatellier, et al., 2000, Delvecchio, et al., 2002, Musser, et al., 1993, Schmitz, et al., 2003, Smoot, et al., 2002). Such information would enable a clearer picture of the mechanisms underlying each disease, and assist in development of novel treatments.

These studies have focussed on gene content differences, allelic distinctions, disparities in the expression of certain proteins and more recently, contrasting expression profiles (Graham, et al., 2006, Shelburne, et al., 2005). The associations that were identified in early studies have in general not held up under increased scrutiny. To illustrate this point, early studies implicated *speA* and *speC* in invasive disease (Hauser, et al., 1991, Musser, et al., 1993). This was subsequently

demonstrated to be due to the association of particular M1 and M3 strains in invasive disease cases in North America and Western Europe (Chatellier, et al., 2000, Cleary, et al., 1992). As some M-types that do not express these particular superantigen are capable of causing invasive disease (Chaussee, et al., 1996, Gorton, et al., 2005, Vlaminckx, et al., 2003) the presence of *speA* and *speC* is not the sole predictor of a strains capacity to cause invasive disease. Equally, the absence of these genes can not be used to infer a strain is non-invasive.

**Comparative genomics**

The advent of genomic sequencing has allowed a more rigorous genetic comparison of GAS. Currently eleven GAS genomes are available in the NCBI genomic database. Direct comparisons reveal up to 10% of the genes can differ between any two strains (Ferretti, et al., 2001, Smoot, et al., 2002). The large differences in gene content between genomes, the complexity of regulatory networks (Kreikemeyer, et al., 2003) and contradictory epidemiological studies have all contributed to the difficulties in identifying factors that consistently correlate with specific disease. The observation that only some strains within an M-type have the propensity to cause invasive disease suggests the clonal expansion of selected strains that have undergone some kind of genetic mutation that increased their invasive propensity (eg. M1T1 lineage (Aziz, et al., 2005, Beres, et al., 2004). Conversely, the observation that multiple M-types are capable of causing invasive disease demonstrates invasive traits to be found in strains that are not clonally related. To account for these contradictory observations one must assume that either 1) strains of different M-type independently developed internal mutations (eg frameshift, gene duplications) increasing their invasive potential or ii) the factor(s) that cause invasive disease are being transferred through the

streptococcal population by lateral genetic transfer (LGT). As invasive disease have only emerged in the last thirty years, the former seems unlikely. Thus the second scenario – the transfer or 'invasive' genes between strains seems more likely. However the inability to identify single genes that are diagnostic for invasive disease suggests that the acquisition of an 'invasive' gene through LGT is insufficient to convert a strain from non-invasive to invasive potential. Rather a combination of virulence factors (or allelic variants) some of which are acquired through LGT, and some of which are already harboured on the chromosome are probably required for an 'invasive' phenotype. The identification of such combinations can only be achieved through comparative genomic studies.

To investigate variability in the virulence gene repertoires of GAS we developed a targeted GAS virulence array consisting of 226 genes (McMillan, et al., 2006, McMillan, et al., 2007). Seventy four of these genes encode classical or suspected GAS virulence factors, or orthologues of virulence factors in other pathogens. The remainder encode putative extracellular proteins identified through HMMSPScan, NN-SPScan and LipoP bioinformatic prediction of signal sequences (Menne, et al., 2000, Nielsen, et al., 1997, Nielsen and Krogh, 1998) contained in open reading frames present in the M1, M3 and M18 genomes, and positive control sequences. Each of the genes was represented by conserved 70mers on the array. Our rationale for including these sequences was that extracellular proteins are more likely to interact with the host than cytoplasmic proteins and therefore have a greater chance of being involved in virulence. Using the array, we reported that 128 genes were present in 100% of 68 GAS isolates representing multiple M-types, and therefore potentially represent constituents of a core genome (McMillan, et al., 2007). In an expanded data

set of 115 isolates, 125 genes are ubiquitously present (Table 1). With the exception of a put. C3-degrading proteinase (C3degP) found in 114 of the 115 isolates, all other proteases examined were present in all isolates studied. This group incudes well studied proteases such as C5a peptidase, streptokinase and SpeB, and lesser studied proteins such as a putative exofoliative toxin (ET). Both C3degP and ET are important virulence factors in other bacterial species (Amagai, et al., 2002), but have received little attention in GAS. These comparisons also revealed four of eight chromosomally encoded toxin genes to be present in all isolates; the remaining four are present in greater than ninety five percent of strains. This data suggest these factors to be critical for the conserved virulence characteristics of these organisms.

Our data has demonstrated that forty seven unique gene repertoires, differing in at least one gene exist amongst the 115 isolates, independent of bacteriophage content. Such a large number of profiles is suggestive of ongoing evolutionary events in the streptococcal genome. Clustering of these strains based on their genetic repertoire demonstrates that overwhelmingly, synteny of strains of the same M-type is greater than strains of distinct M-type (Figure 1). These relationships hold true even in the absence of bacteriophage related genes, already known to have M-type related distributions. In this data there is little evidence linking genetic repertoires and disease (McMillan, et al., 2006). The genes that discriminate between M-type include major virulence factors, as well as genes that have no known biological function. Genetic mapping revealed these genes to occupy a limited number discrete loci (Table 2) (McMillan, et al., 2007). A defining feature in ten of these loci is their proximity to genes involved in lateral genetic transfer. These are both complete and remnant bacteriophage sequences, as well as incomplete transposon elements. The inference is

that bacteriophages and transposons equip strains with new genes and genetic repertoires. However incorrect insertion or excision of these elements, or genomic instability caused by the presence of new large DNA sequences can also result in deletion of genes that flank agents of LGT.

Amongst the M-specific loci, a large number included genes involved in adhesion. The FCT (Towers, et al., 2004) and SbX/sfbII (Bessen and Kalia, 2002, Jeng, et al., 2003) loci are the major location of adhesins in group A streptococcus. Adhesins are important in initial colonisation, enabling bacteria to attach to host extracellular matrix binding proteins. The multiplicity of Fibronectin Binding Protein (FBP) underscores their importance to GAS. Less is known of the importance of collagen binding proteins, although they too are abundant (Kreikemeyer, et al., 2005). Variation in Fibronectin binding capacity not only comes through variation in FBP repertoires; a significant level of amino acid diversity also occur within individual FBPs, some of which has been demonstrated to occur through modular recombination. The necessity for multiple FBP proteins, is unclear, although it has been demonstrated that at least one FBP also has a secondary function (Molinari, et al., 1997). Alternatively changing the FBP repertoire may change tissue specificity, or may be a mechanism to avoid host immune responses directed to these proteins. A third M-type specific loci, Spa, also contains a number of adhesin genes. Although the function of Spa itself remains unknown Spa-deficient mutant were shown to have an impaired virulence in mouse models (McLellan, et al., 2001).

**Conclusion**

Through lateral genetic transfer and internal mutational processes, many unique virulence repertoires that vary both in gene content and allelic type exist in the GAS population. Differences in genetic content between strains of the one M-type highlight the limitations in relying on a small numbers of strains to determine species genomes. By comparing multiple strains from different M-types and associated with different disease, a core genome that includes house keeping genes and key virulence genes whose function is critical to the organisms survival, is emerging. In addition to the core genome each organism also contains supplementary genes that provide additional genotypic and phenotypic attributes. While the distribution of the supplementary genes varies between strains, several have M-type specific distributions, and populate discrete loci in the genome. These loci typically contain evidence of MGEs and often contain genes encoding adhesins. Variation in the adhesion protein repertoire may be an important factor changing the host pathogen relationship, enabling GAS to stay one step ahead of the host immune system (Brochet, et al., 2006).

Although the overall size of the supplementary gene pool has not been defined Tettelin recently provided evidence suggesting GAS pan-genome to be open i.e. as more GAS genomes are sequenced, new GAS genes will continue to be identified (Tettelin, et al., 2005). The extension of this observation is that the number supplementary genes far outnumber the number of core genes. This is consistent with the group B streptococcal pan-genome, and would appear to be a feature of beta-hemolytic streptococci. Sharing of genes between different streptococcal species has been documented (Kalia, et al., 2001, Sriprakash and Hartas, 1996), and often occurs in those loci that differentiate between M-type (eg mga/scp locus). Interspecies acquisition of new genes is therefore another strategy utilised by GAS to increase

their pool of available virulence genes. Recombination and lateral genetic transfers will result in development of new genetic repertoires that will have new virulence properties. In this regard sharing of virulence genes between species may also be responsible for recent reports of group G streptococcal strains associated with invasive disease (Igwe, et al., 2003, Pinho, et al., 2006). As more genomes become available, genomic comparisons of relevant clinical GGS and GAS isolates will yield valuable new data pertinent to the identification of genes associated with streptococcal disease.

**Acknowledgements**

The work presented in this paper was supported by a Pathogenomik Grant from Federal Ministry of Education and Research, Germany to GSC. DJM and KSS are funded through the Australian National Health and Research Council Program Grant.

**References**

Amagai, M., Yamaguchi, T., Hanakawa, Y., Nishifuji, K., Sugai, M., and Stanley, J.R. 2002. Staphylococcal exfoliative toxin B specifically cleaves desmoglein 1. J Invest Dermatol 118**,** 845-850.

Aziz, R.K., Edwards, R.A., Taylor, W.W., Low, D.E., McGeer, A., and Kotb, M. 2005. Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1T1 clone of Streptococcus pyogenes. J Bacteriol 187**,** 3311-3318.

Beres, S.B., Sylva, G.L., Sturdevant, D.E., Granville, C.N., Liu, M., Ricklefs, S.M., Whitney, A.R., Parkins, L.D., Hoe, N.P., Adams, G.J., Low, D.E., DeLeo, F.R., McGeer, A., and Musser, J.M. 2004. Genome-wide molecular dissection of serotype

M3 group A Streptococcus strains causing two epidemics of invasive infections. Proc Natl Acad Sci U S A 101**,** 11833-11838.

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

Brochet, M., Couve, E., Zouine, M., Vallaeys, T., Rusniok, C., Lamy, M.C., Buchrieser, C., Trieu-Cuot, P., Kunst, F., Poyart, C., and Glaser, P. 2006. Genomic diversity and evolution within the species Streptococcus agalactiae. Microbes Infect 8**,** 1227-1243.

Buchanan, J.T., Simpson, A.J., Aziz, R.K., Liu, G.Y., Kristian, S.A., Kotb, M., Feramisco, J., and Nizet, V. 2006. DNase expression allows the pathogen group A Streptococcus to escape killing in neutrophil extracellular traps. Curr Biol 16**,** 396-400.

Carapetis, J.R., Steer, A.C., Mulholland, E.K., and Weber, M. 2005. The global burden of group A streptococcal diseases. Lancet Infect Dis 5**,** 685-694.

Chatellier, S., Ihendyane, N., Kansal, R.G., Khambaty, F., Basma, H., Norrby-Teglund, A., Low, D.E., McGeer, A., and Kotb, M. 2000. Genetic relatedness and superantigen expression in group A streptococcus serotype M1 isolates from patients with severe and nonsevere invasive diseases. Infect Immun 68**,** 3523-3534.

Chaussee, M.S., Liu, J., Stevens, D.L., and Ferretti, J.J. 1996. Genetic and phenotypic diversity among isolates of Streptococcus pyogenes from invasive infections. J Infect Dis 173**,** 901-908.

Cleary, P.P., Kaplan, E.L., Handley, J.P., Wlazlo, A., Kim, M.H., Hauser, A.R., and Schlievert, P.M. 1992. Clonal basis for resurgence of serious Streptococcus pyogenes disease in the 1980s. Lancet 339**,** 518-521.

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.

Delvecchio, A., Currie, B.J., McArthur, J.D., Walker, M.J., and Sriprakash, K.S. 2002. Streptococcus pyogenes prtFII, but not sfbI, sfbII or fbp54, is represented more frequently among invasive-disease isolates of tropical Australia. Epidemiol Infect 128**,** 391-396.

Denny, F.W. 2000. History of the haemolytic streptococcus and associated diseases, p. 1-18. *In* D. L. Stevens and E. L. Kaplan (ed.), Streptococcal Infections. OUP, New York.

Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton, S., Lai, H.S., Lin, S.P., Qian, Y., Jia, H.G., Najar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A., and McLaughlin, R. 2001. Complete genome sequence of an M1 strain of Streptococcus pyogenes. Proc Natl Acad Sci U S A 98**,** 4658-4663.

Gorton, D., Norton, R., Layton, R., Smith, H., and Ketheesan, N. 2005. Presence of fibronectin-binding protein gene prtF2 in invasive group A streptococci in tropical Australia is associated with increased internalisation efficiency. Microbes Infect 7**,** 421-426.

Graham, M.R., Virtaneva, K., Porcella, S.F., Gardner, D.J., Long, R.D., Welty, D.M., Barry, W.T., Johnson, C.A., Parkins, L.D., Wright, F.A., and Musser, J.M. 2006. Analysis of the transcriptome of group a streptococcus in mouse soft tissue infection. Am J Pathol 169**,** 927-942.

Hauser, A.R., Stevens, D.L., Kaplan, E.L., and Schlievert, P.M. 1991. Molecular analysis of pyrogenic exotoxins from Streptococcus pyogenes isolates associated with toxic shock-like syndrome. J Clin Microbiol 29**,** 1562-1567.

Igwe, E.I., Shewmaker, P.L., Facklam, R.R., Farley, M.M., van Beneden, C., and Beall, B. 2003. Identification of superantigen genes speM, ssa, and smeZ in invasive strains of beta-hemolytic group C and G streptococci recovered from humans. FEMS Microbiol Lett 229**,** 259-264.

Jeng, A., Sakota, V., Li, Z., Datta, V., Beall, B., and Nizet, V. 2003. Molecular genetic analysis of a group A Streptococcus operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. J Bacteriol 185**,** 1208-1217.

Kalia, A., Enright, M.C., Spratt, B.G., and Bessen, D.E. 2001. Directional gene movement from human-pathogenic to commensal-like streptococci. Infect Immun 69**,** 4858-4869.

Kiska, D.L., Thiede, B., Caracciolo, J., Jordan, M., Johnson, D., Kaplan, E.L., Gruninger, R.P., Lohr, J.A., Gilligan, P.H., and Denny, F.W., Jr. 1997. Invasive group A streptococcal infections in North Carolina: epidemiology, clinical features, and genetic and serotype analysis of causative organisms. J Infect Dis 176**,** 992-1000.

Kreikemeyer, B., McIver, K.S., and Podbielski, A. 2003. Virulence factor regulation and regulatory networks in Streptococcus pyogenes and their impact on pathogen-host interactions. Trends Microbiol 11**,** 224-232.

Kreikemeyer, B., Nakata, M., Oehmcke, S., Gschwendtner, C., Normann, J., and Podbielski, A. 2005. Streptococcus pyogenes collagen type I-binding Cpa surface protein. Expression profile, binding characteristics, biological functions, and potential clinical impact. J Biol Chem 280**,** 33228-33239.

McLellan, D.G., Chiang, E.Y., Courtney, H.S., Hasty, D.L., Wei, S.C., Hu, M.C., Walls, M.A., Bloom, J.J., and Dale, J.B. 2001. Spa contributes to the virulence of type 18 group A streptococci. Infect Immun 69**,** 2943-2949.

McMillan, D.J., Beiko, R.G., Geffers, R., Buer, J., Schouls, L.M., Vlaminckx, B.J., Wannet, W.J., Sriprakash, K.S., and Chhatwal, G.S. 2006. Genes for the majority of group a streptococcal virulence factors and extracellular surface proteins do not confer an increased propensity to cause invasive disease. Clin Infect Dis 43**,** 884-891.

McMillan, D.J., Geffers, R., Buer, J., Vlaminckx, B., and Sriprakash, K.S. 2007. Variations in the distribution of genes encoding virulence and extracellular proteins in group A streptococcus are largely restricted to eleven genomic loci. Microb Infect accepted (MICINF_2955).

Menne, K.M., Hermjakob, H., and Apweiler, R. 2000. A comparison of signal sequence prediction methods using a test set of signal peptides. Bioinformatics 16**,** 741-742.

Molinari, G., Talay, S.R., Valentin-Weigand, P., Rohde, M., and Chhatwal, G.S. 1997. The fibronectin-binding protein of Streptococcus pyogenes, SfbI, is involved in the internalization of group A streptococci by epithelial cells. Infect Immun 65**,** 1357-1363.

Musser, J.M., Kapur, V., Kanjilal, S., Shah, U., Musher, D.M., Barg, N.L., Johnston, K.H., Schlievert, P.M., Henrichsen, J., Gerlach, D., and et al. 1993. Geographic and temporal distribution and molecular characterization of two highly pathogenic clones of Streptococcus pyogenes expressing allelic variants of pyrogenic exotoxin A (Scarlet fever toxin). J Infect Dis 167**,** 337-346.

Nielsen, H., Engelbrecht, J., Bunak, S., and von Heijne, G. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein engineering 10**,** 1-6.

Nielsen, H., and Krogh, A. 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol 6**,** 122-130.

Pinho, M.D., Melo-Cristino, J., and Ramirez, M. 2006. Clonal relationships between invasive and noninvasive Lancefield group C and G streptococci and emm-specific differences in invasiveness. J Clin Microbiol 44**,** 841-846.

Schmitz, F.J., Beyer, A., Charpentier, E., Normark, B.H., Schade, M., Fluit, A.C., Hafner, D., and Novak, R. 2003. Toxin-gene profile heterogeneity among endemic invasive European group A streptococcal isolates. J Infect Dis 188**,** 1578-1586.

Shelburne, S.A., 3rd, Sumby, P., Sitkiewicz, I., Granville, C., DeLeo, F.R., and Musser, J.M. 2005. Central role of a bacterial two-component gene regulatory system of previously unknown function in pathogen persistence in human saliva. Proc Natl Acad Sci U S A 102**,** 16037-16042.

Shulman, S.T., Stollerman, G., Beall, B., Dale, J.B., and Tanz, R.R. 2006. Temporal changes in streptococcal M protein types and the near-disappearance of acute rheumatic fever in the United States. Clin Infect Dis 42**,** 441-447.

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., Beres, S.B., Campbell, D.S., Smith, T.M., Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.G., and Musser, J.M. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks. Proc Natl Acad Sci U S A 99**,** 4668-4673.

Sriprakash, K.S., and Hartas, J. 1996. Lateral genetic transfers between group A and G streptococci for M-like genes are ongoing. Microb Pathog 20**,** 275-285.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., and Fraser, C.M. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 102**,** 13950-13955.

Towers, R.J., Gal, D., McMillan, D., Sriprakash, K.S., Currie, B.J., Walker, M.J., Chhatwal, G.S., and Fagan, P.K. 2004. Fibronectin-binding protein gene recombination and horizontal transfer between group A and G streptococci. J Clin Microbiol 42**,** 5357-5361.

Veasy, L.G., Wiedmeier, S.E., Orsmond, G.S., Ruttenberg, H.D., Boucek, M.M., Roth, S.J., Tait, V.F., Thompson, J.A., Daly, J.A., Kaplan, E.L., and et al. 1987. Resurgence of acute rheumatic fever in the intermountain area of the United States. N Engl J Med 316**,** 421-427.

Vlaminckx, B.J., Mascini, E.M., Schellekens, J., Schouls, L.M., Paauw, A., Fluit, A.C., Novak, R., Verhoef, J., and Schmitz, F.J. 2003. Site-specific manifestations of invasive group a streptococcal disease: type distribution and corresponding patterns of virulence determinants. J Clin Microbiol 41**,** 4941-4949.

Vlaminckx, B.J., van Pelt, W., Schouls, L.M., van Silfhout, A., Mascini, E.M., Elzenaar, C.P., Fernandes, T., Bosman, A., and Schellekens, J.F. 2005. Long-term surveillance of invasive group A streptococcal disease in The Netherlands, 1994-2003. Clin Microbiol Infect 11**,** 226-231.

**Legend to Fig. 1**

Hierarchical clustering of 115 group A streptococcal strains based on gene content. Genes are represented on the X-axis and strains on the Y-axis. The M-type of major clusters is shown on the right. Strains represent isolates collected from patients presenting with invasive disease, rheumatic fever or superficial infection. Genes present in specific strains are shown as white bars. Absent genes are depicted as black bars.

**Table 1.** Core genes encoding virulence factors and putative extracellular proteins of group A streptococcus.[a]

**Adhesins**
put. collagen-like protein (spy1054)
put. choline binding protein (spyM3_0025)
glyceraldehyde-3-phosphate dehydrogenase, plasmin receptor (spyM3_0201)
put. adhesion protein (spyM3_0466)
put. enolase (spyM3_0479)
put. fibronectin-binding protein-like protein A (spyM3_0652)

**Proteases**
put. protease (spyM3_0418)
streptokinase A precursor (spyM3_1698)
pyrogenic exotoxin B (spyM3_1742)
N-terminal fragment of Sib38 homolgue (spyM18_0921)
put. exfoliative toxin (spyM3_0632)
C5A peptidase precursor (spyM3_1726)
put. serine protease (spyM3_1864)

**Toxins**
mitogenic factor 25K precursor (spyM3_1745)
put. hemolysin (spyM3_0276)
pyrogenic exotoxin G precursor (spyM3_0155)
streptolysin O (spyM3_0130)
mitogenic exotoxin Z (spy1998)
put. hemolysin (spyM3_1153)
put. hemolysin III (spyM3_0815)
streptolysin O (spyM3_0130)

**Miscellaneous factors**
67 kDa Myosin-crossreactive streptococcal antigen (spyM3_0332)
arginine deiminase (spy1547)
collagen-like protein SclB (spyM3_0738)
extracellular hyaluronate lyase (spyM3_0665)
hyaluronan synthase (spyM3_1851)
immunogenic secreted protein precursor homolog (spyM3_1562)
lantibiotic salivaricin A precursor (spyM3_1652)
nicotine adenine dinucleotide glycohydrolase precursor (spyM3_0128)
oligopeptide permease (spyM3_0215)
put. 6-phospho-beta-galactosidase (spyM3_1653)
put. ABC transporter (lipoprotein) (spyM3_0868)
put. ABC transporter substrate-binding protein (spyM3_0653)
put. acid phosphatase (class B) (spyM3_0773)
put. acid phosphatase (spyM3_1625)
put. carbamate kinase (spyM3_1191)
put. cell division regulator (spyM3_0477)
put. competence associated hyp. protein (spyM3_0503)

**Miscellaneous factors cont.**
put. competence protein (spyM3_1074)
put. competence protein (spyM3_0080)
put. competence protein (spyM3_0081)
put. competence protein ComX1 (spyM3_1640)
put. cytoplasmic membrane protein (spyM3_0241)
put. D,D-carboxypeptidase, penicillin-binding protein (spyM3_0755)
put. D,D-carboxypeptidase (spyM3_1535)
put. D-alanyl-D-alanine carboxypeptidase (spyM3_0214)
put. dehydrogenease / oxidoreductase (spyM3_0309)
put. dipeptidase (spyM3_0465)
put. dipeptidase (spyM3_1763)
put. DNA/pantothenate metabolism flavoprotein (spyM3_0862)
put. DNA-entry nuclease (spyM3_0504)
put. endolysin, (spyM3_0425)
put. ferric uptake regulator (spyM3_0147)
put. glutathione peroxidase (spyM3_0428)
put. GTP-binding protein LepA (spyM3_0737)
put. hyaluronidase (spyM3_1294)
put. internalin A precursor (spyM3_1035)
put. manganese-dependent inorganic pyrophosphatase (spyM3_0278)
put. metal binding protein of ABC transporter (spyM3_0318)
put. methionine sulfoxide reductase (spyM3_1267)
put. peptidoglycan hydrolase (spyM3_0582)
put. peptidyl-prolyl cis-trans isomerase (spyM3_0321)
put. periplasmic-iron-binding protein (spyM3_0747)
put. protease maturation protein (spyM3_1059)
put. protease maturation protein (spyM3_1740)
put. PTS system IIB component (spyM3_1679)
put. pullulanase (spyM3_1694)
put. put. proton-translocating ATPase subunit b (spyM3_0495)
put. secreted 5'-nucleotidase (spyM3_0591)
put. secreted endoglycosidase (spyM3_1568)
put. secreted protein (spyM3_0014)
put. sugar ABC transporter substrate binding protein (spyM3_1289)
put. sugar transporter sugar binding lipoprotein (spy0252)
put. surface exclusion protein (spyM3_0197)
put. two-component sensor histidine kinase (spyM3_1201)
put. UDP-glucose pyrophosphorylase (spyM3_0160)
put. UDP-glucose pyrophosphorylase (spyM3_1853)
put. uridine diphosphate N-acetylmuramoyl-L-alanine:D-glutamate ligase (spyM3_1176)
put. uridine kinase (spyM3_1042)
surface lipoprotein DppA (spyM3_1718)

**Hypothetical proteins**
hyp. protein (spy0116)
hyp. protein (spy0317)
hyp. protein (spyM18_1050)
hyp. protein (spyM18_1343)
hyp. protein (spyM3_0009)
hyp. protein (spyM3_0034)
hyp. protein (spyM3_0082)
hyp. protein (spyM3_0084)
hyp. protein (spyM3_0090)
hyp. protein (spyM3_0105)
hyp. protein (spyM3_0154)
hyp. protein (spyM3_0194)
hyp. protein (spyM3_0261)
hyp. protein (spyM3_0299)
hyp. protein (spyM3_0316)
hyp. protein (spyM3_0331)
hyp. protein (spyM3_0343)
hyp. protein (spyM3_0345)
hyp. protein (spyM3_0414)
hyp. protein (spyM3_0415)
hyp. protein (spyM3_0427)
hyp. protein (spyM3_0437)
hyp. protein (spyM3_0536)
hyp. protein (spyM3_0548)
hyp. protein (spyM3_0563)
hyp. protein (spyM3_0633)
hyp. protein (spyM3_0670)
hyp. protein (spyM3_0740)
hyp. protein (spyM3_0756)
hyp. protein (spyM3_0813)
hyp. protein (spyM3_0844)
hyp. protein (spyM3_0898)
hyp. protein (spyM3_0979)
hyp. protein (spyM3_0999)
hyp. protein (spyM3_1070)
hyp. protein (spyM3_1148)
hyp. protein (spyM3_1150)
hyp. protein (spyM3_1268)
hyp. protein (spyM3_1376)
hyp. protein (spyM3_1477)
hyp. protein (spyM3_1626)
hyp. protein (spyM3_1631)
hyp. protein (spyM3_1699)
hyp. protein (spyM3_1737)
hyp. protein (spyM3_1762)
hyp. protein (spyM3_1843)

[a]Adapted from McMillan et al, 2007.

**Table 2**. Loci that differentiate between GAS M-types

| Chromosomal Loci | Genes[a] |
|---|---|
| mga/scp | drs, sic, transposase (spy2009) |
| spa | streptococcal protective antigen, amidase, collagen-like surface protein |
| sfbX/sfbII | sfbII, sfbX, collagen like protein (M3_1702), collagen like protein (M3_1703) |
| maltose transport operon | put. cyclomaltodextrin glucanotransferase, put. maltose/maltodextrin-binding protein, |
| FCT | prtF15, prtF2, sfbI, spy0129, spy0130, spyM3_0101, put. collagen binding protein (spyM18_0126), put. collagen binding protein (spyM3_0098) |
| cit | citE, M3_0823 |
| sagA | put. extracellular matrix binding protein |
| R28 region | R28 |
| streptolysin O | spy0171, spyM3_0131 |
| speJ | spy0430, spy0433, spy0435, spy0437, speJ, spyM18_0477, pyM18_480, spyM3_0307 |
| spyM3_1830 | spyM3_1830 |
| penicillin | put. penicillin binding protein 1A |

[a]Gene with significant M-type distributions

**Figure 1**