Published in final edited form as:

Infect Genet Evol. 2009 July ; 9(4): 581–593. doi:10.1016/j.meegid.2009.03.00.

Population Biology of the Human Restricted Pathogen, *Streptococcus pyogenes*

Debra E. Bessen

Department of Microbiology & Immunology, New York Medical College, Valhalla, NY, USA 10595,
debra_bessen@nymc.edu, +1-914-594-4193

Abstract

S. pyogenes, also referred to as β -hemolytic group A streptococci, are strictly human pathogens with a global distribution and high prevalence of infection. The organisms are characterized by high levels of genetic recombination, extensive strain diversity, and a narrow habitat. This review highlights many key features of the population genetics and molecular epidemiology of this biologically diverse bacterial species, with special emphasis on ecological subdivisions and tissue-specific infections, strain diversity and population dynamics in communities, selection pressures arising from the specific host immune response and antibiotic exposure, and within-host selection during the course of invasive disease.

The *Streptococcus* genus contains > 50 named species that inhabit a broad range of hosts, including humans and domesticated animals, where they often colonize as part of the normal flora and/or cause infection. The most clinically significant of the human pathogens are *Streptococcus pyogenes*, the causative agent of “strep” throat, and *S. pneumoniae*, a leading cause of bacterial pneumonia. *S. pyogenes* and *S. pneumoniae* are exclusive to humans, existing in a carrier state or as pathogens, whereby both natural immunity and therapeutic or preventive interventions (egs., antibiotics, vaccines) have the potential to shape their population biology. The evolution of *S. pneumoniae* clones was recently described in an excellent review (Henriques-Normark *et al.*, 2008). This article focuses on *S. pyogenes*, which lie within the pyogenic branch of the 16S rRNA-based phylogenetic tree, along with other streptococcal species displaying a β -hemolytic pattern of growth on blood agar, such as the human and animal pathogen *S. agalactiae* (Facklam, 2002). *S. pyogenes* also contain the Lancefield serogroup A carbohydrate on their cell surface, and are often referred to as the large colony-forming β -hemolytic group A streptococci.

1. Overview of key genome features

S. pyogenes contain a single circular chromosome of ~ 1.9 Mb; the complete genome sequence has been determined for > 10 independent isolates (Ferretti *et al.*, 2001, Beres *et al.*, 2002, McShan *et al.*, 2008, Smoot *et al.*, 2002, Banks *et al.*, 2003, Green *et al.*, 2005, Beres & Musser, 2007, Holden *et al.*, 2007, Nakagawa *et al.*, 2003). A striking feature of the *S. pyogenes* genome is the abundance of exogenous genetic elements, including prophage and integrated conjugative elements (ICE), occupying ~ 10% of the chromosome [reviewed in (Beres & Musser, 2007, Banks *et al.*, 2002)]. That these elements undergo horizontal exchange is

strongly supported by their mosaic structure and unequal distribution across strains. Recombination between prophage genes can also lead to large-scale chromosomal inversions (Nakagawa *et al.*, 2003). Genes encoding virulence factors and antibiotic resistance are often found in association with the exogenous genetic elements, however, unlike many bacterial pathogens, *S. pyogenes* appear to be completely devoid of pathogenicity islands (Schmidt & Hensel, 2004).

Several genome-wide analyses on the evolution of *S. pyogenes* have been performed. About half of *S. agalactiae* genes have homologs in both *S. pyogenes* and *S. pneumoniae*, with conservation of gene synteny being more pronounced between *S. agalactiae* and *S. pyogenes* (Tettelin *et al.*, 2002). A comparison of numerous genomes of *S. agalactiae* and *S. pyogenes* reveals that *S. pyogenes* has a smaller pan-genome and more recombination in its core-genome, perhaps reflecting the narrower habitat diversity of *S. pyogenes* (Lefebure & Stanhope, 2007). Many individual strains of *S. pyogenes* appear to have acquired a variety of genes from related streptococcal species, such as *S. agalactiae* and *S. dysgalactiae* subspecies *equisimilis*, via lateral gene transfer (Stalhammar-Carlemalm *et al.*, 1999, Franken *et al.*, 2001, Kalia & Bessen, 2004, Bessen *et al.*, 2005, Beres & Musser, 2007). In fact, up to 50% of the ~ 200 *S. pyogenes* virulence genes examined were present in *S. dysgalactiae* ssp. *equisimilis* isolates evaluated by comparative genomic hybridization (Davies *et al.*, 2007), indicative of a close genetic relationship between *S. pyogenes* and this largely commensal species.

2. Disease, transmission and ecology

S. pyogenes are responsible for a minimally estimated 616 million cases of throat infection (pharyngitis, tonsillitis) worldwide per year, and 111 million cases of skin infection (primarily non-bullous impetigo) in children of less developed countries (Carapetis *et al.*, 2005). Streptococcal pharyngitis and impetigo are superficial, self-limiting infections that usually cause only a mild illness. A clinically inapparent oropharyngeal infection can also occur, in which the host mounts a significant immune response, yet lacks obvious clinical symptoms of illness. Infections at the throat and skin are self-limiting, and typically resolve within 2 weeks, coincident with the rise of specific host immune defenses.

Most morbidity and mortality due to *S. pyogenes* arises from invasive and autoimmune disease, although each is somewhat rare in occurrence when compared to the mild infections at the throat and skin. Rheumatic fever is an autoimmune disorder that can have serious cardiac manifestations. Invasive disease is the result of streptococcal infection of deep tissue and can be moderate to quite severe, such as in cases of toxic shock syndrome and necrotizing fasciitis.

S. pyogenes is a free-living organism, however, its ecological niche appears to be quite narrow: Its' only known biological host is the human, and like many other streptococci, lacks an environmental reservoir of known importance. Thus, direct human-to-human transmission, which mostly occurs via respiratory droplets or skin contact, is critical for its maintenance. Streptococcal organisms tend to reside extracellularly, but can also be found within the mammalian host cells of the respiratory tract (Osterlund *et al.*, 1997, Cleary *et al.*, 1998). In addition to causing clinical infection at the throat and skin, *S. pyogenes* can persist in a quiescent carrier state; it is not uncommon for throat carriage rates to exceed 15–20% within communities of school-aged children during seasonal peaks (Bisno & Stevens, 2005). The carrier state elicits little or no immune response to streptococcal antigens, and throat carriage can persist for weeks or months (Kaplan, 1980). Secondary transmission of streptococci from an impetigo lesion to the oropharynx is well-documented, and will typically result in a carrier state (Bisno & Stevens, 2005, Ferrieri *et al.*, 1972); carriage at the skin is less well understood, and may be only transient. Importantly, the mucosal epithelium of the oropharynx and the epidermal layer of

the skin constitute the primary ecological niches of *S. pyogenes*. It is at these two tissue sites that the organism most often enters and exits its biological host.

Reproductive success of *S. pyogenes* at the host epithelium probably follows at least 2 different dynamics. During infection, the generation of new progeny is likely to be high; in an experimental model for impetigo, a > 5,000-fold increase in colony forming units (cfu) is observed (Scaramuzzino *et al.*, 2000, Lizano *et al.*, 2007). An exponential rise in the number of new progeny leads to larger infective doses, which in turn, increases the chance for successful transmission. Yet the window of opportunity for transmission may be narrow due to rapid clearance of the infection by the mounting host immune response. During carriage, the organisms may be in a persistent dormant state (Wood *et al.*, 2005, Leonard *et al.*, 1998), with a low yield of new progeny and a feeble host immune response, but with a much longer time frame for potential transmission events. The physiological state of the organism during its exposure to a new host may also impact transmission success, providing yet another possible factor shaping the distinct dynamic processes for infection versus carriage.

The relative incidence of disease caused by *S. pyogenes* varies throughout the world, in accordance with both season and locale (Carapetis *et al.*, 1999). In temperate regions, pharyngitis is highly prevalent during the winter months, perhaps as a consequence of indoor crowding; impetigo, although less common, is most often encountered during warm and humid weather conditions. Impetigo caused by *S. pyogenes* is far more prevalent than pharyngeal infection in many tropical regions. Thus, climate is an important risk factor for the different forms of streptococcal disease, which in turn, creates spatial and temporal distances between the organisms causing throat versus skin infections.

3. Strain typing based on surface structures

Epidemiological markers are useful for investigating outbreaks of disease and providing a reference point for deciphering the genetic organization of a bacterial population. During the 1920's, Dr. Rebecca Lancefield began work aimed at understanding the basis for protective immunity to *S. pyogenes* infection. Antiserum raised to extractable surface antigens known as M proteins led to opsonophagocytosis and killing of the strain from which the M protein was derived (Lancefield, 1962); however, antiserum directed to the M protein of one organism often failed to protect against many other isolates. A serological typing scheme arose through the development of antiserum directed to M proteins of different isolates. More than 80 distinct M types were identified, and strong protective immunity is M type-specific.

M proteins form hair-like fibrils extending about 60 nm from the surface of the bacterial cell (Fischetti, 1989). The determinants of serological type lie at the amino-termini, which correspond to the distal fibril tips. The serologically-based M protein typing scheme was replaced about 12 years with a highly correlated nucleotide (nt) sequence-based *emm* typing scheme (Beall, 2008, Beall *et al.*, 1996). *emm* type is defined by the 5' end region of the *emm* gene, and genes assigned the same *emm* type have > 92% nt sequence identity. More than 200 *emm* types are currently listed in the online C.D.C. database (Beall, 2008).

S. pyogenes display heterogeneity in the number and arrangement of 4 basic forms of *emm* and *emm*-like genes that are defined by differences in the region encoding the cell wall-spanning domain (Hollingshead *et al.*, 1994, Hollingshead *et al.*, 1993). A strain can have up to 3 distinct *emm* or *emm*-like genes, lying in a contiguous stretch. There are 5 recognized arrangements of *emm* and *emm*-like genes that are denoted *emm* patterns. Together, the 5 *emm* patterns comprise 3 main groups known as pattern A–C, D and E, with *emm* type residing in either the first (pattern A–C) or middle (patterns D and E) gene. Nearly all *emm* types examined are restricted to a single *emm* pattern grouping (McGregor *et al.*, 2004b). So while there may be horizontal

movement of the *emm* gene between strains, any one *emm* type seems to be largely restricted to the same *emm* pattern group.

A second serological typing scheme for *S. pyogenes* is based on the T antigen, originally defined by its resistance to trypsin. There appears to be a far greater number of distinct M serotypes than T serotypes (Johnson *et al.*, 2006). The *tee* gene of a T6 type strain was initially mapped to the highly recombinatorial FCT region of the *S. pyogenes* genome, positioned ~ 300 kb from the *emm* region (Schneewind *et al.*, 1990, Bessen & Kalia, 2002). More recently, T antigens have been characterized as forming elongated pili that function as adhesins (Mora *et al.*, 2005, Kreikemeyer *et al.*, 2005, Manetti *et al.*, 2007, Abbot *et al.*, 2007). A new approach for nt sequence-based typing of the genes encoding pilus structural proteins was recently introduced (Falugi *et al.*, 2008).

Serum opacity factor (SOF) is a protein produced by ~ 50 % of *S. pyogenes* isolates, both in a surface-bound and secreted form. The amino-terminal region binds apolipoprotein present in serum, leading to its opacification, whereas the carboxy-terminal region functions as an adhesin via binding to fibronectin (Gillen *et al.*, 2008). A serological typing scheme was developed in which SOF type-specific antiserum inhibits serum opacity activity. *sof*-positive strains and *emm* pattern E generalists comprise nearly perfect overlapping groups.

S. pyogenes also possess a surface polysaccharide capsule that functions as an important virulence factor. The quantity of capsule produced by *S. pyogenes* strains can be elevated in association with increased levels of transmission via a respiratory route (Stollerman & Dale, 2008). However, the capsular material is uniformly composed of hyaluronic acid, a poor immunogen. The other streptococcal species causing significant disease in humans - *S. pneumoniae* and *S. agalactiae* – differ from *S. pyogenes* in having a variety of capsular polysaccharide antigenic forms that provide the basis for serological typing as well as targets for protective immunity.

4. Tissue tropisms among *S. pyogenes*

With serological typing schemes in place, decades of epidemiological studies revealed that there are serotypes of *S. pyogenes* having a strong tendency to cause throat infection, but not skin infection, and similarly, there are other serotypes often seen with impetigo, but much less so with pharyngitis (Bisno & Stevens, 2005, Wannamaker, 1970). This observation gave rise to the concept of distinct throat and skin types, suggesting that there exists a degree of specialization among strains belonging to this species. Indeed, the *emm* pattern genotypes have proven to be fairly strong markers for preferred tissue site of infection among *S. pyogenes* strains.

Several recent reports on population-based surveillance for *S. pyogenes* associated with pharyngitis or impetigo infections, when taken together, yield *emm* typing data for > 3,700 isolates collected from all 6 major continents (Bessen *et al.*, 2000, Dicuonzo *et al.*, 2001, Sakota *et al.*, 2006, Shulman, 2004, Smeesters *et al.*, 2006, Tewodros & Kronvall, 2005, Bessen *et al.*, 2008, Alberti *et al.*, 2003, Espinosa *et al.*, 2003, Brandt *et al.*, 2001, McDonald *et al.*, 2007b). Since isolates sharing the same *emm* type almost always belong to the same *emm* pattern group (McGregor *et al.*, 2004b), *emm* pattern can be inferred with a high degree of accuracy, based on knowledge of the *emm* type (Table 1).

The combined population analysis reveals a strong trend whereby the *emm* pattern AC group has a strong predilection to cause infection at the throat, and *emm* pattern D strains have a strong tendency to cause impetigo. The relationship between *emm* pattern and infection site is statistically significant when the data for each of the surveys is combined (Table 1); the differences in the distribution of *emm* pattern A–C versus D strains are highly significant for

collections of both pharyngitis ($t = 0.005$; paired t -test, 2-tailed) and impetigo ($t = 0.008$) isolates. The *emm* pattern A–C strains are considered throat specialists, whereas *emm* pattern D strains are skin specialists. As a group, the pattern E strains are designated generalists, readily causing infections at both tissue sites. Thus, *S. pyogenes* strains can be divided into 3 clinically and ecologically relevant groups based on *emm* pattern genotype.

Conceivably, tissue site preferences for infection at the throat versus skin might be explained by tissue-specific colonization factors, such as those encoded by the FCT region genes that give rise to surface pili. There is some correlative data to support this idea (summarized in Table 2). The *cpa* gene, encoding a pilus accessory protein, is present in nearly all *emm* pattern D strains, but absent from the majority of *emm* pattern A–C strains (Kratovac *et al.*, 2007); the well-characterized Cpa protein of an M49 strain displays high affinity-binding to human collagen type I (Kreikemeyer *et al.*, 2005), a form that is abundant in the dermis of human skin. Cpa is essential for streptococcal virulence in an experimental model for impetigo (Lizano *et al.*, 2007).

The FCT region gene *prtF1* encoding a fibronectin-binding protein shows an inverse relationship to that of *cpa*, being present in most *emm* pattern A–C strains and absent from most *emm* pattern D strains (Kratovac *et al.*, 2007). The *emm* pattern E generalists mostly harbor both *cpa* and *prtF1*, perhaps reflecting their tropism for both the skin and throat. However, *emm* pattern D organisms can readily undergo secondary transmission to the throat suggesting that they can colonize both tissues, even though disease is largely restricted to the skin. Furthermore, in addition to *cpa* and *prtF1*, *S. pyogenes* possess other adhesins, including ≥ 5 other fibronectin-binding proteins (encoded by *prtF2*, *sof*, *sfbX*, *fbaA* and *fbp54*), most of which have a differential presence among strains (Kratovac *et al.*, 2007, Jeng *et al.*, 2003, Terao *et al.*, 2001, Courtney *et al.*, 1996, Jaffe *et al.*, 1996), and at least one of which (*prtF2*) contributes to virulence in experimental skin infection (Lizano *et al.*, 2007). Thus, FCT region-derived colonization factors may facilitate tissue-specific infection, yet additional factors appear to be at play as well. Of the > 10 *S. pyogenes* isolates whose complete genome sequence has been published, 10 *emm* types are represented, but none of the strains belong to *emm* pattern group D; therefore, it is plausible that new tissue-specific colonization factors remain to be discovered.

The M proteins are mosaics of functional domains that bind a variety of host factors present in plasma and tissue. The high-affinity plasminogen-binding domain of M protein (Sanderson-Smith *et al.*, 2007, Berge & Sjöbring, 1993) is restricted to a subset of *emm* pattern D strains (Svensson *et al.*, 1999) (Table 2). Plasminogen-binding M proteins, also known as PAM, capture host plasminogen on the bacterial cell surface. Bacterial surface-bound plasmin activity, due to a broad spectrum protease, arises via interaction of the PAM-bound form of plasminogen with the streptococcal exoprotein streptokinase (Ringdahl *et al.*, 1998). Streptokinase is encoded by the *ska* locus, positioned ~ 35 kb from the *emm* chromosomal region. The *ska* alleles of *S. pyogenes* form 2 distinct phylogenetic lineages, and all PAM-positive *emm* pattern D strains examined share sub-cluster 2b alleles (Kalia & Bessen, 2004) (Table 2); *ska* is also essential for full virulence in an experimental model for streptococcal impetigo (Svensson *et al.*, 2002). Striking functional differences have been described for streptokinase encoded by alleles of different lineages (McArthur *et al.*, 2008, Nordstrand *et al.*, 2000). Additional phenotypic correlates of the *emm* pattern D genotype include elevated levels of secreted cysteine protease activity due to the streptococcal exoprotein SpeB, which is also a key virulence factor in experimental impetigo (Svensson *et al.*, 2000).

The *mga* gene encodes a transcriptional activator of *emm* and *emm*-like genes, and lies immediately upstream of the *emm* region [reviewed in (Hondorp & McIver, 2007)]. All *S. pyogenes* strains examined have a *mga* locus, but the *mga* alleles comprise 2 discrete

phylogenetic lineages of ~ 24% nt sequence divergence. Nearly all *emm* pattern A–C strains harbor a *mga-1* lineage allele whereas pattern D and E strains have *mga-2* lineage alleles (Bessen *et al.*, 2005) (Table 2). *Mga* acts as a global regulator of transcription, and it is plausible that *mga-1* and *mga-2* products differentially regulate genes that act as key determinants of tissue-specific infection.

Two mutually exclusive lineages of alleles, having ~ 33% nt sequence divergence, are present in a second locus (*rofA/nra*) that encodes transcription regulatory proteins, for which FCT region genes are a target (Bessen *et al.*, 2005, Kreikemeyer *et al.*, 2003, Podbielski *et al.*, 1999, Fogg *et al.*, 1994). Nearly all *emm* pattern A–C and E strains have *rofA*, whereas most *emm* pattern D strains harbor *nra* (Bessen *et al.*, 2005) (Table 2). The association of *rofA* and *nra* with *emm* pattern group differs from the distribution of *mga* alleles, in that the pattern E generalists tend to share *rofA* with pattern A–C strains, but share *mga-2* with pattern D strains. Tissue-specific infection by *S. pyogenes* may require a particular set of virulence factor genes, combined with the differential expression of multiple virulence factors, acting in concert.

To gain further insight on the evolution of *S. pyogenes*, the gain or loss of genes which correlate with *emm* pattern genotypes was assessed by maximum parsimony. The genes were assigned a character state (presence or absence, or lineage) at 6 loci of the *emm* or FCT regions, for > 100 *S. pyogenes* isolates representing different *emm* types (Figure 1). The phylogenetic tree shows that the majority of isolates cluster in accordance with their *emm* pattern group. Most strains belonging to the *emm* pattern E group of generalists are positioned in between the throat and skin specialists, consistent with the concept that specialists tend to emerge from generalists (Elena & Sanjuan, 2003).

Not only are the throat and skin specialists physically separated by their ecological niches, but there is geographic partitioning in terms of the prevalence of throat versus skin infection in different parts of the world (Carapetis *et al.*, 1999). The non-overlapping seasonal peak incidence for pharyngitis and impetigo introduces even more distance. The spatial and temporal separation might impose barriers to lateral gene exchange which, in turn, raises the question of whether *emm* pattern A–C and D strains are at the beginning stages of forming new species.

5. Overview of mechanisms underlying genetic change

The mechanisms underlying genetic change in *S. pyogenes* presumably include both specialized and generalized transduction via bacteriophage and conjugation via ICE. Replicative plasmids have also been described for some isolates (Woodbury *et al.*, 2008). *S. pyogenes* lack the full complement of competence genes mediating transformation that are found in *S. pneumoniae* (Martin *et al.*, 2006), however, some *S. pyogenes* strains have homologs (*sil* locus) involved in quorum sensing and DNA transfer (Eran *et al.*, 2007, Hidalgo-Grass *et al.*, 2002).

There is evidence for differences among strains in their potential for genetic change, rooted in their genetic machinery. The presence of clustered regularly interspaced short palindromic repeat (CRISPR) regions within the genomes of some *S. pyogenes* strains may result in selective blocking of bacterial cell infection by new phage via an RNA interference mechanism (McShan *et al.*, 2008, Barrangou *et al.*, 2007). ICE harboring genes that encode type II restriction-modification cassettes may prevent the uptake of foreign DNA originating from certain sources (Euler *et al.*, 2007). The mutator phenotype, leading to high rates of mutation in some strains, can arise from a defect in DNA mismatch repair that is controlled by prophage excision and re-integration at a chromosomal site upstream from the *mutL* locus (Scott *et al.*, 2008). Thus, individual strains may differ in the quality and frequency of genetic change.

Based on whole-genome alignments derived from 12 *S. pyogenes* isolates, genetic flux was found to be dominated by gain and loss of prophage genes. Estimation of the rates of genetic flux indicate that, at least for some strains, prophage integration may have accelerated in recent time (Didelot *et al.*, 2008). Another study yielded similar findings, showing that the within-species branches of a phylogenetic tree constructed from isolates of several *Streptococcus* species, including *S. pyogenes*, have a higher rate of gene gain or loss than the branches leading to a species (Marri *et al.*, 2006). Furthermore, a higher proportion of genes that underwent recent lateral transfer show evidence for positive selection, when compared to ancient genes, indicating that some of the newly acquired genes probably serve an adaptive function. Yet in another phylogenomic study taking a slightly different approach, the proportion of *S. pyogenes* genes exhibiting signs of positive selection were about equal for accessory and core genes (Anisimova *et al.*, 2007). Species-specific genes identified for *S. pyogenes*, such as those encoding bacteriocins, may have conferred a survival advantage (Marri *et al.*, 2006).

The impact of homologous recombination and point mutation on the population genetic structure of *S. pyogenes* was estimated using a multilocus sequencing typing (MLST) approach, as described below.

6. Population genetic structure defined by MLST

MLST based on neutral housekeeping genes is routinely used to define clones of *S. pyogenes* with greater precision (Enright *et al.*, 2001). The MLST data posted at www.mlst.net currently lists 463 sequence types (ST) of *S. pyogenes*, based on allelic profiles at 7 loci (Aanensen, 2008). Numerous investigators have generously contributed to this rich data set.

Using the MLST data, a population snapshot generated by the eBURST program (Feil *et al.*, 2004) reveals 463 STs (dots) and 71 clonal complexes (CCs), in which the connected STs (lines) are single locus variants (SLVs) sharing 6 of the 7 housekeeping alleles (Figure 2A). A high proportion of STs (43.8%) differ from all others by ≥ 2 alleles (singletons), and the largest CC contains only 3.02% of the total STs. This places the population structure of *S. pyogenes* at the far end of the spectrum for bacteria, approaching that of the gut pathogen *Helicobacter pylori*. Organisms falling in this part of the scale are characterized by high levels of genetic diversification, due to both point mutation and homologous recombination (Turner *et al.*, 2007). The population snapshot of *S. pyogenes* is strikingly different from that of its close genetic relative *S. agalactiae* (Figure 2B) (Jolley, 2008), which has a high % of STs in large straggly (and possibly unreliable) eBURST groups (Turner *et al.*, 2007).

Recombination and mutation can be envisioned as two opposing forces in shaping the population genetic structure of a bacterial species (Hanage *et al.*, 2006b). There are several methods for assessing the relative amount of recombination and mutation that contributes to genetic change. Statistical tests were used to measure the congruence between pair wise combinations of the phylogenetic trees corresponding to each of the 7 housekeeping genes used for MLST of *S. pyogenes*, based on representative strains from each deep branch (Feil *et al.*, 2001). Of the 42 possible pair wise tree comparisons, no significant congruence between trees was observed. The lack of congruence is indicative of relatively high levels of recombination within *S. pyogenes*.

Using MLST data for an expanded sample set of *S. pyogenes* isolates and the multilocus infinite alleles model, high rates of recombination are predicted once again (Hanage *et al.*, 2006a). The value for the estimated rate of mutation (θ) is 7.1 for *S. pyogenes*, which is similar to that calculated for *S. pneumoniae* ($\theta = 7.4$). The estimated rate of recombination (ρ) is even higher for *S. pyogenes* (51.2) than for *S. pneumoniae* ($\rho = 29.6$). *S. pneumoniae* is similar to *S. pyogenes* in its lack of congruency among the topologies of phylogenetic trees based on

housekeeping gene sequences (Feil et al., 2001). Both sets of measures are also consistent with population snapshot findings on *S. pneumoniae* showing a slightly higher % of STs associated with the largest CC (Turner et al., 2007). Together, these data support the notion that *S. pyogenes* exhibits a slightly higher level of recombination, as compared to *S. pneumoniae*.

An empirical estimate of the number of alleles changed by recombination compared to mutation, based on SLVs generated by eBURST analysis, yields a ratio of 8.9 to 1 for *S. pneumoniae* (Feil et al., 2001). Empirical estimates for the recombination to mutation ratio in *S. pyogenes* have been more difficult to attain because of the few SLVs observed in this species. In a recent analysis of a highly diverse set of nearly 600 *S. pyogenes* isolates collected from > 25 countries worldwide, representing 156 *emm* types and 259 STs, only 56 SLVs were detected by eBURST (Bessen et al., 2008). The genetic mechanism of descent was estimated as recombination for ≥ 33 of the 56 SLVs, yielding a recombination to mutation ratio of ≥ 1.4 , which is perhaps a bit lower than expected in light of the other types of measurements. However, when stratified according to *emm* pattern group, the ratio of recombination to mutational events is 0.14, 4.0 and 3.0 for *emm* patterns A–C, D and E, respectively. Thus, diversification by mutation seems to predominate among the group of throat specialists, whereas skin specialists and generalists tend to diversify by recombination.

Analysis of the population genetic structure of *emm* pattern-defined groups of *S. pyogenes* isolates are based on small sample sets and data needs to be interpreted with caution. Nonetheless, *emm* pattern A–C strains show the highest level of congruence for housekeeping gene tree topologies, indicative of relatively lower levels of recombination, whereas *emm* pattern D strains displayed the lowest congruency, indicating higher recombination (Kalia et al., 2002). Therefore, the different methodologies point to the same general trend, whereby the throat specialists have the highest tendency to diversify by mutation. Differences among the *emm* pattern-defined groups in terms of their predominant mechanism for genetic change point to different dynamics in shaping the population structures of the ecologically distinct organisms. Conceivably, this may be a result of higher incidences of co-colonization or co-infection with multiple strains for skin specialists and generalists (Carapetis et al., 1995), increasing the chances for lateral gene transfer, and/or differences in the machinery underlying genetic change.

Ecological barriers can give rise to allopatric speciation. It was of interest to ascertain whether there are signatures of early stages of speciation within the *S. pyogenes* population. Numerous analyses on housekeeping alleles - including phylogenetic trees of concatenated alleles, splits graphs, fixed nt differences, distribution of shared alleles - all provide strong support for extensive recombination involving housekeeping genes of the 3 *emm* pattern groupings (Kalia et al., 2002). These findings extend to isolates known to be recovered from the upper respiratory tract versus skin lesions, regardless of *emm* pattern group (Kalia et al., 2002). Figure 3 shows a phylogenetic tree based on concatenated nt sequences of the 7 housekeeping genes used in MLST, for 114 strains representing 114 STs and 113 distinct *emm* types. There are only a few branches showing strong bootstrap support, and the nt diversity (π) is very low. Strains assigned to *emm* pattern groups A–C, D and E are scattered along the length of the tree, and fail to exhibit strong clustering. When depicted as a circular tree, the concatenated sequences display a “bush-like” structure (data not shown), consistent with rapid genetic exchange. A subset of these 114 strains are represented in Figure 1, showing ecological congruence based on several adaptive loci.

In summary, there is a lack of clonal congruence among strains corresponding to the 3 *emm* pattern-defined groups, as demonstrated with housekeeping genes. Thus, despite some niche separation created by distinct epidemiologic trends (temporal and spatial) and innate tissue tropisms, there is no evidence for housekeeping gene sequence divergence between strains of

differing *emm* patterns, or between isolates known to be recovered from the upper respiratory tract versus skin lesions. This finding provides further support for ecological congruence among isolates of each *emm* pattern group, whereby loci exhibiting a high degree of linkage with *emm* pattern genotypes (Figure 1, Table 2), and which also map to distal positions on the genome, likely play a direct role in the adaptations leading to throat or skin infection.

7. Strain diversity and dynamics within defined host populations

Surveillance studies conducted on small spatial scales provide evidence for ample migration of *S. pyogenes*. In a remote tropical island community of aboriginal Australians, periodic surveillance was conducted over 2 years by collecting specimens from the throat and skin lesions (Bessen et al., 2000, McGregor et al., 2004a). Of ~ 500 patient visits involving 224 individual subjects, 16 *S. pyogenes* were recovered in association with throat carriage and 121 isolates were obtained from skin lesions; there were no cases of pharyngitis. Together, the 137 *S. pyogenes* isolates represent 32 *emm* types, 35 STs and 35 unique *emm* type-ST combinations, with only 2 CCs. Using the *emm* type-ST combination to define clone, the Simpson's diversity index (Grundmann et al., 2001), D , is measured as 0.9589 (95% CI, 0.9505, 0.9673), indicative of a highly diverse bacterial population.

The *S. pyogenes* strains afflicting this remote island community most likely migrated in from elsewhere. At least 90% of the *emm* types were recovered from places outside of Australia (McGregor et al., 2004a). There is no known reservoir for *S. pyogenes* aside from the human host. Also, the genomes of the many *S. pyogenes* strains examined carry genetic material for only a single *emm* type (and perhaps a few minor variants), as opposed to a broad genetic repertoire for assembling the complete *emm* type set.

Similarly high levels of *S. pyogenes* strain diversity have been observed in other remote human populations sampled over relatively short periods. The combined data for 3 remote Australian aboriginal communities on the mainland yielded 350 *S. pyogenes* isolates from 49 households and 1173 individuals, periodically screened over 1 or 2 years (McDonald et al., 2007b, McDonald et al., 2007a); 43 *emm* types were found among these isolates, yet long-term throat carriage of the same *emm* type was uncommon. In far-western Nepal, 120 *S. pyogenes* isolates were collected from a single screening of 60 children in each of 8 villages, yielding 45 *emm* types and 51 STs (Sakota et al., 2006). Thus, despite rather limited outside contacts, many distinct strains of *S. pyogenes* circulate within remote human populations over a short time frame.

The 3 studies cited above were dominated by cases of superficial skin infection, with few or no instances of pharyngitis, although asymptomatic throat carriage was evident. A large multiple site surveillance study of streptococcal pharyngitis in children was recently conducted for several urban and suburban areas in the United States and Canada, yielding nearly 2,000 isolates over 2 years (Shulman, 2004). In year 1 and year 2, the number of distinct *emm* types recovered was 29 and 31, respectively, with 23 *emm* types shared between both years. Thus, a highly diverse set of *S. pyogenes* strains are also responsible for oropharyngeal disease, yielding a Simpson's diversity index D of 0.8907 (95% CI, 0.8848, 0.8966) when *emm* type is used to define clone. Interestingly, the diversity value for the multisite pharyngitis study is somewhat less than that for the remote aboriginal Australian island population (adjusted D = 0.9379, when clone is defined by *emm* type only; 95% CI, 0.9296, 0.9462) (Bessen et al., 2000). This might be explained by differences in the sampling approaches and/or the diversity of throat versus skin specialist strains.

The pharyngitis survey also reveals striking differences between the multiple study sites in terms of the distribution of predominant *emm* types, although the predominant *emm* types fluctuate only slightly from year to year (Shulman, 2004). Yet, a comparison of the M type

distribution among pharyngitis isolates obtained from pediatric patients in Chicago during the 1960s versus 40 years later shows marked changes in the predominating M types (Shulman *et al.*, 2006). The factors underlying these trends are probably complex, and may include longer term cyclical changes in herd immunity. Age-related changes in *emm* type distribution are observed in childhood cases of pharyngitis, with older children showing increased infection by uncommon *emm* types, and a corresponding decrease in common *emm* types (Jaggi *et al.*, 2005). Conceivably, these shifts are due to the acquisition of protective immunity following exposure to highly prevalent strains earlier in life.

8. *emm* type as a target of host immune selection

For the many *S. pyogenes* strains examined, protective immunity is *emm* type-specific, whereby M type-specific antibodies mediate opsonization and overcome the antiphagocytic property of M protein (Lancefield, 1962, Beachey *et al.*, 1981). Because they elicit a highly protective host immune response, the M type determinants are the targets of a promising *S. pyogenes* vaccine currently under development (Bisno *et al.*, 2005, Dale *et al.*, 2005, McNeil *et al.*, 2005). However, given the large number of *emm* types present throughout the world, the practicality of an M type-specific vaccine for the prevention of *S. pyogenes* disease has been the subject of debate.

The C.D.C. website currently lists ~ 217 *emm*-types identified among *S. pyogenes*, represented by ~ 883 partial *emm* alleles (i.e., subtypes) as defined by the 150 nt encoding the 50 amino-terminal residues of the mature M protein (Beall, 2008). The *emm* type-determining regions of alleles assigned to the same *emm* type give high quality alignments, and small indels and nt substitutions are readily apparent (Bessen *et al.*, 2008). However, alignments of the *emm* type-determining regions derived from multiple *emm* types are often characterized by numerous gaps and high levels of sequence heterogeneity, and therefore it becomes more difficult to establish the phylogenetic relationships between different *emm* types (Wertz *et al.*, 2007). Trees based on amino acid sequence alignments that include a large portion of the signal peptide show a clear segregation of *emm* types derived from SOF-positive versus SOF-negative strains (Sakota *et al.*, 2006), which closely correspond to *emm* patterns E versus A–C and D groups, respectively.

Immune selection pressures imposed by the host may drive diversification of the *emm* type-specific region, leading to immune escape. Experimental findings provide evidence for amino acid changes in the M type-specific region that allow for immune escape (Jones *et al.*, 1988, Beres *et al.*, 2006, Eriksson *et al.*, 2001), and such changes may explain the recent emergence of an important M3 type clone. A signature of selection lies in the ratio of nonsynonymous substitutions per nonsynonymous site (K_a) and synonymous substitutions per synonymous site (K_s), whereby nonsynonymous substitutions lead to a change in amino acid residue. The K_a and K_s values were calculated for 105 alignments of partial *emm* alleles (i.e., subtypes) corresponding to 105 *emm* types, derived from a global collection of > 500 *S. pyogenes* isolates (Bessen *et al.*, 2008). The ratio of the mean K_a to mean K_s value for the 105 alignments was 1.96, indicative of positive diversifying selection acting on the type-specific region. For the *emm* pattern A–C group of throat specialists, this ratio value was ~ 3- to 4-fold higher than for skin specialists and generalists, suggestive of stronger immune selection pressures acting on the M type-specific region in throat strains.

Despite the large number of distinct *emm* types and *emm* allele subtypes that exist, there also appear to be functional constraints which put the brakes on genetic diversification. The M type-specific regions of most *S. pyogenes* strains bind the complement regulator C4b-binding protein (C4BP). C4BP binding is achieved in the absence of a shared amino acid sequence motif and even though substitutions can introduce antigenic change without altering C4BP binding

activity (Persson *et al.*, 2006), it seems likely that there are functional constraints on sequence variation. Of probable relevance is the finding that most isolates lacking C4BP binding activity have *emm* types characteristic of pattern A–C strains (egs., M types 1, 3, 5, 6, 12, 19, 24, 26, 30, 39) (Persson *et al.*, 2006, McGregor *et al.*, 2004b). When combined with the K_a to K_s ratio data (Bessen *et al.*, 2008), the findings are suggestive of higher levels of purifying (negative) selection on pattern D and E *emm* types in order to preserve C4BP binding activity.

Immune escape might also arise following a change in *emm* type mediated by recombinational replacement of all or part of the *emm* gene. Among a genetically diverse set of nearly 600 *S. pyogenes* isolates, a small proportion of the STs examined (5.4%) were found in association with a substantially larger fraction (> 20%) of the total *emm* types evaluated; nearly all of the *emm*-variable STs were restricted to *emm* pattern groups A–C and D (Bessen *et al.*, 2008). There are also numerous instances of the same *emm* type (~ 50%) associated with distant STs differing at ≥ 5 of the 7 housekeeping alleles. Horizontal transfer of an *emm* type to a distant ST was observed primarily among the *emm* pattern D and E strains. Thus, relationships between *emm* type and genetic background differ among the 3 host tissue-related groups; any differences in selection pressures and mechanisms for genetic change may have important implications for vaccine design. This analysis also shows that *emm* type is a reasonably good marker for ST or CC among the throat specialist strains, but a rather poor marker for clone among skin specialists and generalists.

A given M or *emm* type can be found in association with numerous SOF types (or *sof* alleles) and T types, as summarized in an analysis of > 40,000 isolates collected worldwide over a 50 year period (Johnson *et al.*, 2006). These findings provide further evidence that *emm* type undergoes horizontal exchange with other strains. There are numerous examples of *emm* types associated with > 1 SOF or *sof* type, suggestive of independent lateral transfer events involving the *sof* locus, which is positioned ~ 16 kb upstream of *emm*. A 450 nt region at the 5' end of the *sof* gene is used to define the partial *sof* allele and it strongly correlates with SOF serological types (Beall *et al.*, 2000). Attempts to generate multiple sequence alignments of the 5' end region of *sof* alleles has been thwarted by extensive sequence heterogeneity and evidence for past recombination that likely involved short tandem duplications and inverted repeats (Wertz *et al.*, 2007).

The number of *S. pyogenes* “strains” or “clones” – as defined by unique combinations of *emm* type, *sof* type, T or pilus gene type and ST – remains to be established. In addition to *emm*, *sof* and *tee* (i.e., FCT region) genes, there may be additional immunodominant surface antigens that shape the population structure of this species.

9. Antibiotic resistance

β -lactams such as penicillin are the drug of choice for treating upper respiratory tract infections caused by *S. pyogenes*. Fortunately, resistance to β -lactams has not emerged in *S. pyogenes*, perhaps due to the organisms' failure to acquire plasmids harboring a β -lactamase gene, or a high biological cost is incurred when penicillin-binding proteins involved in cell wall synthesis undergo mutations that lower their affinity for penicillin (Horn *et al.*, 1998). However, *S. pyogenes* have evolved resistance to other antibiotics, which can adversely affect the clinical course of disease and shape the population biology of the species by causing shifts in fitness when placed in an antibiotic infused environment.

Macrolides and lincosamides are the primary treatment for GAS infections in patients with β -lactam hypersensitivity or chronic, recurrent pharyngitis due to prior treatment failure; clindamycin (a lincosamide) is the first choice drug for patients with life-threatening soft tissue infections, such as necrotizing fasciitis, because it halts exotoxin production. Studies in Japan, Finland and elsewhere show a strong correlation between macrolide consumption and

resistance in *S. pyogenes* (Freeman & Shulman, 2002, Fujita *et al.*, 1994, Garcia-Rey *et al.*, 2002, Perez-Trallero *et al.*, 2001, Seppala *et al.*, 1997). In a recent 3-year longitudinal surveillance study in Pittsburgh, 100% of *S. pyogenes* isolates recovered from children were macrolide-sensitive until the third year of study, when macrolide resistance shot up to 48% of all isolates; all resistant isolates were M6 (Martin *et al.*, 2002). Thus, a macrolide resistant clone can have a significant survival advantage under certain conditions.

At least 3 genes confer resistance to macrolides in *S. pyogenes*: *erm(A)* and *erm(B)*, leading to ribosomal modification, and *mef(A)*, which promotes drug efflux. In a survey of 212 macrolide-resistant isolates obtained from throughout the world, at least 49 independent acquisitions of macrolide resistance were estimated, based on unique combinations of *emm* type, ST and the resistance gene type (Robinson *et al.*, 2006). Numerous genetic elements harboring resistance genes have been defined, and include plasmids, ICEs and non-conjugative elements containing features of prophage, plasmids and/or transposons, some of which are shared with related species such as *S. pneumoniae* (Varaldo *et al.*, 2008, Banks *et al.*, 2003). Some elements carry determinants for resistance to both macrolides and tetracyclines, the latter often being conferred by *tet(M)* or *tet(O)*. Resistance to tetracyclines is estimated to have been acquired via ≥ 80 independent horizontal gene transfer events (Ayer *et al.*, 2007). Considering the higher global prevalence of tetracycline resistant *S. pyogenes*, it is plausible that tetracycline usage drives the acquisition of macrolide resistance to some extent, via genetic elements harboring both resistance genes.

Some strains of *S. pyogenes* (eg., M6 isolates) are resistant to fluoroquinolones; this phenotype can be found in isolates recovered prior to the development of these synthetic antibiotics (Orscheln *et al.*, 2005). Non-therapeutic antibiotics to which at least some *S. pyogenes* strains are susceptible include bacteriocins (lantibiotics), produced by *S. pyogenes* or by other bacterial species that are part of the normal flora of the oral cavity (Upton *et al.*, 2001, Wescombe *et al.*, 2006, Walls *et al.*, 2003). Operons encoding lantibiotic synthesis genes are listed among the prominent species-specific genes of *S. pyogenes* (Marri *et al.*, 2006). These natural compounds have the potential to profoundly shape the ecology of a microbial community.

10. Population biology considerations in invasive disease

One of the more serious consequences of *S. pyogenes* infection is invasive disease, wherein the organism gains access to normally sterile tissue. A transient bacteremia probably occurs in the ~ 50% of patients who develop invasive infections without an obvious portal of entry (Bisno & Stevens, 2005). Host factors that modulate the immune response are key determinants of the severity of invasive streptococcal disease (Abdeltawab *et al.*, 2008). Several clones which are highly correlated with invasive disease cases have been identified. However, there is strong evidence for strain prevalence in the community, rather than innate virulence potential, as a major factor in the tight association between certain strains, as defined by *emm* type, and invasive disease (Rogers *et al.*, 2007, Shulman, 2004, Cockerill *et al.*, 1997).

The M1T1 clone is among the most significant of *S. pyogenes* strains, displaying both a high degree of lethality due to invasive disease and a widespread prevalence over the past 25 years (Aziz & Kotb, 2008). Injection of a pharyngeal-derived M1 isolate into mice led to the recovery of organisms from deeper tissue having a mutation in genes encoding the 2-component signal transduction system CovRS, resulting in an altered transcriptome (Sumby *et al.*, 2006). Thus, genetic change among one or a few of the infecting organisms may be a prerequisite for gaining access to and/or surviving in deep tissue. The findings would appear to imply that bacteria must reproduce at the upper respiratory tract in sufficient numbers in order to increase the probability of generating mutants having increased fitness for invasive disease; however, the pharyngeal infection that precedes invasive disease is typically asymptomatic. Perhaps an

unusual ability of M1 strains to induce clinically inapparent infection at the throat is an evolved trait that confers a survival advantage to the organism because the patient fails to seek antibiotic treatment.

Unique genetic features of the hypervirulent MIT1 clone also include the presence of prophage carrying genes for the pyrogenic exotoxin SpeA and SdaI nuclease; these genes are absent from the M1 strains which have been in circulation decades before emergence of the hypervirulent MIT1 clone (Aziz *et al.*, 2005, Aziz *et al.*, 2004). SpeA is a superantigen which recognizes subsets of human T lymphocytes in a manner that eventually leads to an outpouring of cytokines; the relative lack of *speA* in earlier *S. pyogenes* isolates may have resulted in low levels of neutralizing antibody to SpeA, thereby contributing to the increased incidence of severe invasive disease in recent years (Aziz & Kotb, 2008). SdaI is degraded by the SpeB cysteine protease whose expression, in turn, is controlled by CovRS. In the absence of SpeB, the DNA degrading activity of SdaI is able to facilitate the escape of bacteria from neutrophil extracellular traps and in doing so, SdaI may act as a selective force that promotes the emergence and invasive spread of *covRS* mutants (Walker *et al.*, 2007). This important new concept - that selection pressure exerted by the innate immune response of the host generates hypervirulent genetic variants - may also be broadly applicable to invasive disease caused by other *S. pyogenes* strains, as well as by other bacterial pathogens.

One aspect of the epithelium-to-invasive genetic transition that remains to be more firmly established is the epidemiological association between the pharyngeal genotype (or pharyngeal transcriptome profile) and recovery from the oropharynx, and the invasive genotype (or invasive transcriptome profile) and isolation from deep tissue. To what extent is the pharyngeal genotype recovered from deep tissue, or the invasive genotype recovered from the epithelium? Is the invasive genotype essential for deep tissue infection having a clear portal of entry? Is the invasive genotype capable of causing primary infection at the epithelial surface and transmitting to new hosts? This information should provide added insights into the epidemiology and population biology of severe *S. pyogenes* disease.

11. Future directions

Despite our extensive knowledge, there remain many unanswered questions on the population biology of *S. pyogenes*. Although identification of new *emm* types has slowed somewhat in recent years, it is difficult to predict the total number of extant *emm* types; some parts of the world, notably much of Africa and portions of Asia and South America, have yet to be intensively sampled. Because sequence alignment of many *emm* types is generally poor, one needs to wonder whether there are intermediate genetic forms in circulation (but not yet sampled), or if they no longer exist, why did they become extinct, whether by random genetic drift or selection. How best to define strains of *S. pyogenes* also remains to be established; “strain” can be defined in epidemiologic terms by the loci that affect its transmission, rather than its entire genetic constitution (Gupta *et al.*, 1996). The type-specific determinants of M protein satisfy this criteria because of the role played by M type as a target of protective immunity, yet it remains unclear as to which additional factors affect transmission (egs., T antigens, SOF). The number of STs as defined by MLST is bound to grow with more intensive sampling and as a consequence, perhaps measures of the relative rates of recombination to mutation will become more accurate.

Many of the dynamic qualities of *S. pyogenes* population biology can only be roughly estimated at this time. Little is known about strain migration, beyond that it appears to occur at a high rate. Herd immunity undoubtedly influences the transmission success of individual strains by establishing the number of susceptible hosts. The reason for why there exist > 200 *emm* types in nature is probably explained in part by strain migration, herd immunity, and the size and

density of the human host population. Human genetic loci that confer susceptibility to invasive disease and autoimmune sequelae are recognized (Abdeltawab et al., 2008, Guilherme & Kalil, 2007), but less is known about the potential role of host genetics in the ability to cause mild infections at the throat and skin.

Defining a collection of bacteria as a species can be challenging because the boundaries are often fuzzy for highly recombinogenic organisms (Hanage *et al.*, 2005). The complex microbial communities in which *S. pyogenes* often resides - at the human throat and skin - typically include other streptococcal species and gram positive organisms whose gene sequences and genetic machinery are occasionally compatible. Gaining a more complete understanding of the co-inhabitants may, in turn, sharpen our understanding of *S. pyogenes* by placing it in a broader context.

Acknowledgments

Work on the population biology of *S. pyogenes* has received generous support from The National Institutes of Health (GM060793, AI053826, AI061454, AI065572). Special thanks to Daniel Godoy (Imperial College London) for recent updates to the *S. pyogenes* database available at www.mlst.net.

References

- Aanensen, DM. Imperial College; London: 2008. <http://spyogenes.mlst.net>
- Abbot EL, Smith WD, Siou GP, Chiriboga C, Smith RJ, Wilson JA, Hirst BH, Kehoe MA. Pili mediate specific adhesion of *Streptococcus pyogenes* to human tonsil and skin. *Cell Microbiol* 2007;9:1822–1833. [PubMed: 17359232]
- Abdeltawab NF, Aziz RK, Kansal R, Rowe SL, Su Y, Gardner L, Brannen C, Nooh MM, Attia RR, Abdelsamed HA, Taylor WL, Lu L, Williams RW, Kotb M. An unbiased systems genetics approach to mapping genetic loci modulating susceptibility to severe streptococcal sepsis. *PLoS pathogens* 2008;4:e1000042. [PubMed: 18421376]
- Alberti S, Garcia-Rey C, Dominguez MA, Aguilar L, Cercenado E, Gobernado M, Garcia-Perea A. Survey of emm gene sequences from pharyngeal *Streptococcus pyogenes* isolates collected in Spain and their relationship with erythromycin susceptibility. *J Clin Microbiol* 2003;41:2385–2390. [PubMed: 12791853]
- Anisimova M, Bielawski J, Dunn K, Yang Z. Phylogenomic analysis of natural selection pressure in *Streptococcus* genomes. *BMC evolutionary biology* 2007;7:154. [PubMed: 17760998]
- Ayer V, Tewodros W, Manoharan A, Skariah S, Luo F, Bessen DE. Tetracycline resistance in group A streptococci: emergence on a global scale and influence on multiple-drug resistance. *Antimicrobial agents and chemotherapy* 2007;51:1865–1868. [PubMed: 17307980]
- Aziz RK, Edwards RA, Taylor WW, Low DE, McGeer A, Kotb M. Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1T1 clone of *Streptococcus pyogenes*. *Journal of bacteriology* 2005;187:3311–3318. [PubMed: 15866915]
- Aziz RK, Kotb M. Rise and persistence of global M1T1 clone of *Streptococcus pyogenes*. *Emerging infectious diseases* 2008;14:1511–1517. [PubMed: 18826812]
- Aziz RK, Pabst MJ, Jeng A, Kansal R, Low DE, Nizet V, Kotb M. Invasive M1T1 group A *Streptococcus* undergoes a phase-shift in vivo to prevent proteolytic degradation of multiple virulence factors by SpeB. *Molecular microbiology* 2004;51:123–134. [PubMed: 14651616]
- Banks DJ, Beres SB, Musser JM. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. *Trends in microbiology* 2002;10:515–521. [PubMed: 12419616]
- Banks DJ, Porcella SF, Barbican KD, Martin JM, Musser JM. Structure and distribution of an unusual chimeric genetic element encoding macrolide resistance in phylogenetically diverse clones of group A *Streptococcus*. *Journal of Infectious Diseases* 2003;188:1898–1908. [PubMed: 14673771]

- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* (New York, NY) 2007;315:1709–1712.
- Beachey EH, Seyer JM, Dale JB, Simpson WA, Kang AH. Type-specific protective immunity evoked by synthetic peptide of *Streptococcus pyogenes* M protein. *Nature* 1981;292:457–459. [PubMed: 6166868]
- Beall, B. Atlanta: 2008. <http://www.cdc.gov/ncidod/biotech/strep/strepindex.htm>
- Beall B, Facklam R, Thompson T. Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. *J Clin Microbiol* 1996;34:953–958. [PubMed: 8815115]
- Beall B, Gherardi G, Lovgren M, Forwick B, Facklam R, Tyrrell G. *Emm* and *sof* gene sequence variation in relation to serological typing of opacity factor positive group A streptococci. *Microbiology* 2000;146:1195–1209. [PubMed: 10832648]
- Beres SB, Musser JM. Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. *PLoS ONE* 2007;2:e800. [PubMed: 17726530]
- Beres SB, Richter EW, Nagiec MJ, Sumby P, Porcella SF, Deleo FR, Musser JM. Molecular genetic anatomy of inter- and intraserotype variation in the human bacterial pathogen group A *Streptococcus*. *Proceedings of the National Academy of Sciences of the United States of America* 2006;103:7059–7064. [PubMed: 16636287]
- Beres SB, Sylva GL, Barbian KD, Lei B, Hoff JS, Mammarella ND, Liu MY, Smoot JC, Porcella SF, Parkins LD, Campbell DS, Smith TM, McCormick JK, Leung DY, Schlievert PM, Musser JM. Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage- encoded toxins, the high-virulence phenotype, and clone emergence. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99:10078–10083. [PubMed: 12122206]
- Berge A, Sjöbring U. PAM, a novel plasminogen-binding protein from *Streptococcus pyogenes*. *J Biol Chem* 1993;268:25417–25424. [PubMed: 8244975]
- Bessen DE, Carapetis JR, Beall B, Katz R, Hibble M, Currie BJ, Collingridge T, Izzo MW, Scaramuzzino DA, Sriprakash KS. Contrasting molecular epidemiology of group A streptococci causing tropical and non-tropical infections of the skin and throat. *J Infect Dis* 2000;182:1109–1116. [PubMed: 10979907]
- Bessen DE, Kalia A. Genomic localization of a T-serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in *Streptococcus pyogenes*. *Infect Immun* 2002;70:1159–1167. [PubMed: 11854196]
- Bessen DE, Manoharan A, Luo F, Wertz JE, Robinson DA. Evolution of transcription regulatory genes is linked to niche specialization in the bacterial pathogen *Streptococcus pyogenes*. *Journal of bacteriology* 2005;187:4163–4172. [PubMed: 15937178]
- Bessen DE, McGregor KF, Whatmore AM. Relationships between *emm* and multilocus sequence types within a global collection of *Streptococcus pyogenes*. *BMC microbiology* 2008;8:59. [PubMed: 18405369]
- Bisno AL, Rubin FA, Cleary PP, Dale JB. Prospects for a group A streptococcal vaccine: Rationale, feasibility, and obstacles - Report of a National Institute of Allergy and Infectious Diseases workshop. *Clinical Infectious Diseases* 2005;41:1150–1156. [PubMed: 16163634]
- Bisno, AL.; Stevens, DL. *Streptococcus pyogenes*. In: Mandell, GL.; Douglas, RG.; Dolin, R., editors. *Principles and Practice of Infectious Diseases*. Philadelphia: Churchill Livingstone; 2005.
- Brandt CM, Spellerberg B, Honscha M, Truong ND, Hoevenner B, Lutticken R. Typing of *Streptococcus pyogenes* strains isolated from throat infections in the region of Aachen, Germany. *Infection* 2001;29:163–165. [PubMed: 11440388]
- Carapetis J, Currie B, Kaplan E. Epidemiology and prevention of group A streptococcal infections: Acute respiratory tract infections, skin infections, and their sequelae at the close of the twentieth century. *Clin Infect Dis* 1999;28:205–210. [PubMed: 10064227]
- Carapetis J, Gardiner D, Currie B, Mathews JD. Multiple strains of *Streptococcus pyogenes* in skin sores of Aboriginal Australians. *J Clin Microbiol* 1995;33:1471–1472. [PubMed: 7650169]
- Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases [Review]. *The Lancet Infectious Diseases* 2005;5:685–694. [PubMed: 16253886]

- Cleary PP, McLandsborough L, Ikeda L, Cue D, Krawczak J, Lam H. High-frequency intracellular infection and erythrogenic toxin A expression undergo phase variation in M1 group A streptococci. *Mol Microbiol* 1998;28:157–167. [PubMed: 9593304]
- Cockerill FR, MacDonald KL, Thompson RL, Roberson F, Kohner Pc, Besser-Wiek J, Manahan JM, Musser JM, Schlievert PM, Talbot J, Frankfort B, Steckelberg JM, Wilson WR, Osterholm MT. An outbreak of invasive group A streptococcal disease associated with high carriage rates of the invasive clone among school-aged children. *J Amer Med Assoc* 1997;277:38–43.
- Courtney HS, Dale JB, Hasty DI. Differential effects of the streptococcal fibronectin-binding protein, FBP54, on adhesion of group A streptococci to human buccal cells and HEp-2 tissue culture cells. *Infect Immun* 1996;64:2415–2419. [PubMed: 8698460]
- Dale JB, Penfound T, Chiang EY, Long V, Shulman ST, Beall B. Multivalent group A streptococcal vaccine elicits bactericidal antibodies against variant M subtypes. *Clinical & Diagnostic Laboratory Immunology* 2005;12:833–836. [PubMed: 16002631]
- Davies MR, McMillan DJ, Beiko RG, Barroso V, Geffers R, Sriprakash KS, Chhatwal GS. Virulence profiling of *Streptococcus dysgalactiae* subspecies *equisimilis* isolated from infected humans reveals two distinct genetic lineages which do not segregate with their phenotypes or propensity to cause diseases. *Clin Infect Dis*. 2007in press
- Dicuonzo G, Gherardi G, Lorino G, Angeletti S, DeCesaris M, Fiscarelli E, Bessen DE, Beall B. Group A streptococcal genotypes from pediatric throat isolates in Rome, Italy. *J Clin Microbiol* 2001;39:1687–1690. [PubMed: 11325974]
- Didelot X, Darling A, Falush D. Inferring genomic flux in bacteria. *Genome research*. 2008
- Elena SF, Sanjuan R. Climb every mountain? *Science (New York, NY)* 2003;302:2074–2075.
- Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of *Streptococcus pyogenes* and the relationship between *emm*-type and clone. *Infect Immun* 2001;69:2416–2427. [PubMed: 11254602]
- Eran Y, Getter Y, Baruch M, Belotserkovsky I, Padalon G, Mishalian I, Podbielski A, Kreikemeyer B, Hanski E. Transcriptional regulation of the *sil* locus by the *SilCR* signalling peptide and its implications on group A streptococcus virulence. *Molecular microbiology* 2007;63:1209–1222. [PubMed: 17238919]
- Eriksson BKG, Villaseñor-Sierra A, Norgren M, Stevens DL. Opsonization of T1M1 group A *Streptococcus*: Dynamics of antibody production and strain specificity. *Clinical Infectious Diseases* 2001;32:E24–E30. [PubMed: 11170937]
- Espinosa LE, Li ZY, Barreto DG, Jaimes EC, Rodriguez RS, Sakota V, Facklam RR, Beall B. M protein gene type distribution among group A streptococcal clinical isolates recovered in Mexico City, Mexico, from 1991 to 2000, and Durango, Mexico, from 1998 to 1999: Overlap with type distribution within the United States. *Journal of clinical microbiology* 2003;41:373–378. [PubMed: 12517875]
- Euler CW, Ryan PA, Martin JM, Fischetti VA. M. *SpyI*, a DNA methyltransferase encoded on a *mefA* chimeric element, modifies the genome of *Streptococcus pyogenes*. *Journal of bacteriology* 2007;189:1044–1054. [PubMed: 17085578]
- Facklam R. What happened to the streptococci: Overview of taxonomic and nomenclature changes [Review]. *Clinical Microbiology Reviews* 2002;15:613–630. [PubMed: 12364372]
- Falugi F, Zingaretti C, Pinto V, Mariani M, Amodeo L, Manetti AG, Capo S, Musser JM, Orefici G, Margarit I, Telford JL, Grandi G, Mora M. Sequence Variation in Group A *Streptococcus* Pili and Association of Pilus Backbone Types with Lancefield T Serotypes. *The Journal of infectious diseases*. 2008
- Feil EJ, Holmes EC, Bessen DE, Chan MS, Day NPJ, Enright MC, Goldstein R, Hood D, Kalia A, Moore CE, Zhou J, Spratt BG. Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci, USA* 2001;98:182–187. [PubMed: 11136255]
- Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol* 2004;186:1518–1530. [PubMed: 14973027]
- Ferretti JJ, McShan WM, Ajdic D, Savic DJ, Savic G, Lyon K, Primeaux C, Sezate S, Suvorov AN, Kenton S, Lai HS, Lin SP, Qian Y, Jia HG, Najjar FZ, Ren Q, Zhu H, Song L, White J, Yuan X,

- Clifton SW, Roe BA, McLaughlin R. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci USA* 2001;98:4658–4663. [PubMed: 11296296]
- Ferrieri P, Dajani AS, Wannamaker LW, Chapman SS. Natural history of impetigo. I. Site sequence of acquisition and familial patterns of spread of cutaneous streptococci. *J Clin Invest* 1972;51:2851–2862. [PubMed: 5080412]
- Fischetti VA. Streptococcal M protein: Molecular design and biological behavior. *Clin Microbiol Rev* 1989;2:285–314. [PubMed: 2670192]
- Fogg GC, Gibson CM, Caparon MG. The identification of *rofa*, a positive-acting regulatory component of *prtF* expression: use of an mu-gamma-delta-based shuttle mutagenesis strategy in *Streptococcus pyogenes*. *Mol Microbiol* 1994;11:671–684. [PubMed: 8196542]
- Franken C, Haase G, Brandt C, Weber-Heynemann J, Martin S, Lammler C, Podbielski A, Luttkicken R, Spellerberg B. Horizontal gene transfer and host specificity of beta-haemolytic streptococci: the role of a putative composite transposon containing *scpB* and *lmb*. *Molecular microbiology* 2001;41:925–935. [PubMed: 11532154]
- Freeman AF, Shulman ST. Macrolide resistance in group A *Streptococcus*. *Pediatric Infectious Disease Journal* 2002;21:1158–1160. [PubMed: 12488669]
- Fujita K, Muroto K, Yoshikawa M, Murai T. Decline of erythromycin resistance of group A streptococci in Japan. *The Pediatric infectious disease journal* 1994;13:1075–1078. [PubMed: 7892073]
- Garcia-Rey C, Aguilar L, Baquero F, Casal J, Martin JE. Pharmacoepidemiological Analysis of Provincial Differences between Consumption of Macrolides and Rates of Erythromycin Resistance among *Streptococcus pyogenes* Isolates in Spain. *Journal of clinical microbiology* 2002;40:2959–2963. [PubMed: 12149359]
- Gillen CM, Courtney HS, Schulze K, Rohde M, Wilson MR, Timmer AM, Guzman CA, Nizet V, Chhatwal GS, Walker MJ. Opacity factor activity and epithelial cell binding by the serum opacity factor protein of *Streptococcus pyogenes* are functionally discrete. *The Journal of biological chemistry* 2008;283:6359–6366. [PubMed: 18180300]
- Green NM, Zhang S, Porcella SF, Nagiec MJ, Barbian KD, Beres SB, LeFebvre RB, Musser JM. Genome sequence of a serotype M28 strain of group A streptococcus: potential new insights into puerperal sepsis and bacterial disease specificity. *Journal of Infectious Diseases* 2005;192:760–770. [PubMed: 16088825]
- Grundmann H, Hori S, Tanner G. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *Journal of clinical microbiology* 2001;39:4190–4192. [PubMed: 11682558]
- Guilherme L, Kalil J. Rheumatic fever: from innate to acquired immune response. *Annals of the New York Academy of Sciences* 2007;1107:426–433. [PubMed: 17804571]
- Gupta S, Maiden MCJ, Feavers IM, Nee S, May RM, Anderson RM. The maintenance of strain structure in populations of recombining infectious agents. *Nature medicine* 1996;2:437–442.
- Hanage WP, Fraser C, Spratt BG. Fuzzy species among recombinogenic bacteria. *BMC biology* 2005;3:6. [PubMed: 15752428]
- Hanage WP, Fraser C, Spratt BG. The impact of homologous recombination on the generation of diversity in bacteria. *Journal of theoretical biology* 2006a;239:210–219. [PubMed: 16236325]
- Hanage WP, Spratt BG, Turner KM, Fraser C. Modelling bacterial speciation. *Philosophical transactions of the Royal Society of London* 2006b;361:2039–2044. [PubMed: 17062418]
- Henriques-Normark B, Blomberg C, Dagerhamn J, Battig P, Normark S. The rise and fall of bacterial clones: *Streptococcus pneumoniae*. *Nat Rev Microbiol* 2008;6:827–837. [PubMed: 18923410]
- Hidalgo-Grass C, Ravins M, Dan-Goor M, Jaffe J, Moses AE, Hanski E. A locus of group A *Streptococcus* involved in invasive disease and DNA transfer. *Molecular microbiology* 2002;46:87–99. [PubMed: 12366833]
- Holden MT, Scott A, Cherevach I, Chillingworth T, Churcher C, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Mungall K, Quail MA, Price C, Rabinowitsch E, Sharp S, Skelton J, Whitehead S, Barrell BG, Kehoe M, Parkhill J. Complete genome of acute rheumatic fever-associated serotype M5 *Streptococcus pyogenes* strain manfredo. *Journal of bacteriology* 2007;189:1473–1477. [PubMed: 17012393]

- Hollingshead SK, Readdy T, Arnold J, Bessen DE. Molecular evolution of a multi-gene family in group A streptococci. *Mol Biol Evol* 1994;11:208–219. [PubMed: 8170362]
- Hollingshead SK, Readdy TL, Yung DL, Bessen DE. Structural heterogeneity of the *emm* gene cluster in group A streptococci. *Mol Microbiol* 1993;8:707–717. [PubMed: 8332063]
- Hondorp ER, McIver KS. The Mga virulence regulon: infection where the grass is greener. *Molecular microbiology* 2007;66:1056–1065. [PubMed: 18001346]
- Horn D, Zabriskie J, Austrian R, Cleary P, Ferretti J, Fischetti V, Gotschlich E, Kaplan E, McCarty M, Opal S, Roberts R, Tomasz A, Wachtfogel Y. Why have group A streptococci remained susceptible to penicillin? Report on a symposium. *Clin Infect Dis* 1998;26:1341–1345. [PubMed: 9636860]
- Jaffe J, Natanson-Yaron S, Caparon MG, Hanski E. Protein F2, a novel fibronectin-binding protein from *Streptococcus pyogenes*, possesses two binding domains. *Mol Microbiol* 1996;21:373–384. [PubMed: 8858591]
- Jaggi P, Tanz RR, Beall B, Shulman ST. Age influences the emm type distribution of pediatric group A streptococcal pharyngeal isolates. *The Pediatric infectious disease journal* 2005;24:1089–1092. [PubMed: 16371871]
- Jeng A, Sakota V, Li ZY, Datta V, Beall B, Nizet V. Molecular genetic analysis of a group A *Streptococcus* operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. *Journal of bacteriology* 2003;185:1208–1217. [PubMed: 12562790]
- Johnson DR, Kaplan EL, VanGheem A, Facklam RR, Beall B. Characterization of group A streptococci (*Streptococcus pyogenes*): correlation of M-protein and emm-gene type with T-protein agglutination pattern and serum opacity factor. *Journal of medical microbiology* 2006;55:157–164. [PubMed: 16434707]
- Jolley, K. University of Oxford; 2008. <http://pubmlst.org/>
- Jones KF, Hollingshead SK, Scott JR, Fischetti VA. Spontaneous M6 protein size mutants of group A streptococci display variation in antigenic and opsonogenic epitopes. *Proc Natl Acad Sci USA* 1988;85:8271–8275. [PubMed: 2460864]
- Kalia A, Bessen DE. Natural Selection and Evolution of Streptococcal Virulence Genes Involved in Tissue-Specific Adaptations. *J Bacteriol* 2004;186:110–121. [PubMed: 14679231]
- Kalia A, Spratt BG, Enright MC, Bessen DE. Influence of recombination and niche separation on the population genetic structure of the pathogen *Streptococcus pyogenes*. *Infect Immun* 2002;70:1971–1983. [PubMed: 11895961]
- Kaplan EL. The group A streptococcal upper respiratory tract carrier state: An enigma. *J Pediatr* 1980;97:337–345. [PubMed: 6997450]
- Kratovac Z, Manoharan A, Luo F, Lizano S, Bessen DE. Population genetics and linkage analysis of loci within the FCT region of *Streptococcus pyogenes* *J Bacteriol* 2007;189:1299–1310.
- Kreikemeyer B, McIver KS, Podbielski A. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends in microbiology* 2003;11:224–232. [PubMed: 12781526]
- Kreikemeyer B, Nakata M, Oehmcke S, Gschwendtner C, Normann J, Podbielski A. *Streptococcus pyogenes* collagen type I-binding Cpa surface protein - Expression profile, binding characteristics, biological functions, and potential clinical impact. *Journal of Biological Chemistry* 2005;280:33228–33239. [PubMed: 16040603]
- Lancefield RC. Current knowledge of the type specific M antigens of group A streptococci. *J Immunol* 1962;89:307–313. [PubMed: 14461914]
- Lefebure T, Stanhope MJ. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. *Genome biology* 2007;8:R71. [PubMed: 17475002]
- Leonard BAB, Woischnik M, Podbielski A. Production of stabilized virulence factor-negative variants by group A streptococci during stationary phase. *Infect Immun* 1998;66:3841–3847. [PubMed: 9673270]
- Lizano S, Luo F, Bessen DE. Role of streptococcal T-antigens in superficial skin infection. *J Bacteriol* 2007;189:1426–1434. [PubMed: 17012387]
- Manetti AG, Zingaretti C, Falugi F, Capo S, Bombaci M, Bagnoli F, Gambellini G, Bensi G, Mora M, Edwards AM, Musser JM, Graviss EA, Telford JL, Grandi G, Margarit I. *Streptococcus pyogenes*

- pili promote pharyngeal cell adhesion and biofilm formation. *Molecular microbiology* 2007;64:968–983. [PubMed: 17501921]
- Marri PR, Hao W, Golding GB. Gene gain and gene loss in streptococcus: is it driven by habitat? *Molecular biology and evolution* 2006;23:2379–2391. [PubMed: 16966682]
- Martin B, Quentin Y, Fichant G, Claverys JP. Independent evolution of competence regulatory cascades in streptococci? *Trends in microbiology* 2006;14:339–345. [PubMed: 16820295]
- Martin JM, Green M, Barbadora KA, Wald ER. Erythromycin-resistant group A streptococci in schoolchildren in Pittsburgh. *The New England journal of medicine* 2002;346:1200–1206. [PubMed: 11961148]
- McArthur JD, McKay FC, Ramachandran V, Shyam P, Cork AJ, Sanderson-Smith ML, Cole JN, Ringdahl U, Sjöbring U, Ranson M, Walker MJ. Allelic variants of streptokinase from *Streptococcus pyogenes* display functional differences in plasminogen activation. *Faseb J* 2008;22:3146–3153. [PubMed: 18511548]
- McDonald MI, Towers RJ, Andrews R, Bengler N, Fagan P, Currie BJ, Carapetis JR. The dynamic nature of group A streptococcal epidemiology in tropical communities with high rates of rheumatic heart disease. *Epidemiol Infect* 2007a;1–11.
- McDonald MI, Towers RJ, Fagan P, Carapetis JR, Currie BJ. Molecular typing of *Streptococcus pyogenes* from remote Aboriginal communities where rheumatic fever is common and pyoderma is the predominant streptococcal infection. *Epidemiol Infect* 2007b;135:1398–1405. [PubMed: 17306049]
- McGregor K, Bilek N, Bennett A, Kalia A, Beall B, Carapetis J, Currie B, Sriprakash K, Spratt B, Bessen D. Group A streptococci from a remote community have novel multilocus genotypes but share *emm*-types and housekeeping alleles. *J Infect Dis* 2004a;189:717–723. [PubMed: 14767827]
- McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, Bessen DE. Multi-locus sequence typing of *Streptococcus pyogenes* representing most known *emm*-types and distinctions among sub-population genetic structures. *J Bacteriol* 2004b;186:4285–4294. [PubMed: 15205431]
- McNeil SA, Halperin SA, Langley JM, Smith B, Warren A, Sharratt GP, Baxendale DM, Reddish MA, Hu MC, Stroop SD, Linden J, Fries LF, Vink PE, Dale JB. Safety and immunogenicity of 26-valent group A *Streptococcus* vaccine in healthy adult volunteers. *Clinical Infectious Diseases* 2005;41:1114–1122. [PubMed: 16163629]
- McShan WM, Ferretti JJ, Karasawa T, Suvorov AN, Lin S, Qin B, Jia H, Kenton S, Najjar F, Wu H, Scott J, Roe BA, Savic DJ. Genome sequence of a nephritogenic and highly transformable M49 strain of *Streptococcus pyogenes*. *Journal of bacteriology* 2008;190:7773–7785. [PubMed: 18820018]
- Mora M, Bensi G, Capo S, Falugi F, Zingaretti C, Manetti AGO, Maggi T, Taddei AR, Grandi G, Telford JL. Group A *Streptococcus* produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:15641–15646. [PubMed: 16223875]
- Nakagawa I, Kurokawa K, Yamashita A, Nakata M, Tomiyasu Y, Okahashi N, Kawabata S, Yamazaki K, Shiba T, Yasunaga T, Hayashi H, Hattori M, Hamada S. Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome research* 2003;13:1042–1055. [PubMed: 12799345]
- Nordstrand A, McShan WM, Ferretti JJ, Holm SE, Norgren M. Allele substitution of the streptokinase gene reduces the nephritogenic capacity of group A streptococcal strain NZ131. *Infection and immunity* 2000;68:1019–1025. [PubMed: 10678902]
- Orscheln RC, Johnson DR, Olson SM, Presti RM, Martin JM, Kaplan EL, Storch GA. Intrinsic reduced susceptibility of serotype 6 *Streptococcus pyogenes* to fluoroquinolone antibiotics. *The Journal of infectious diseases* 2005;191:1272–1279. [PubMed: 15776373]
- Osterlund A, Popa R, Nikkila T, Scheynius A, Engstrand L. Intracellular reservoir of *Streptococcus pyogenes* in vivo: a possible explanation for recurrent pharyngotonsillitis. *Laryngoscope* 1997;107:640–647. [PubMed: 9149167]
- Perez-Trallero E, Fernandez-Mazarrasa C, Garcia-Rey C, Bouza E, Aguilar L, Garcia-de-Lomas J, Baquero F. Antimicrobial susceptibilities of 1,684 *Streptococcus pneumoniae* and 2,039 *Streptococcus pyogenes* isolates and their ecological relationships: results of a 1-year (1998–1999) multicenter surveillance study in Spain. *Antimicrobial agents and chemotherapy* 2001;45:3334–3340. [PubMed: 11709305]

- Persson J, Beall B, Linse S, Lindahl G. Extreme sequence divergence but conserved ligand-binding specificity in *Streptococcus pyogenes* M protein. *PLoS pathogens* 2006;2:e47. [PubMed: 16733543]
- Podbielski A, Woischnik M, Leonard BAB, Schmidt KH. Characterization of nra, a global negative regulator gene in group A streptococci. *Molecular microbiology* 1999;31:1051–1064. [PubMed: 10096074]
- Ringdahl U, Svensson M, Wistedt A, Renné T, Kellner R, Müller-Esterl W, Sjöbring U. Molecular co-operation between protein PAM and streptokinase for plasmin acquisition by *Streptococcus pyogenes*. *The Journal of biological chemistry* 1998;273:6424–6430. [PubMed: 9497374]
- Robinson D, Sutcliffe J, Tewodros W, Manoharan A, Bessen D. Evolution and global dissemination of macrolide resistant group A streptococci. *Antimicrob Agents & Chemo* 2006;50:2903–2911.
- Rogers S, Commons R, Danchin MH, Selvaraj G, Kelpie L, Curtis N, Robins-Browne R, Carapetis JR. Strain prevalence, rather than innate virulence potential, is the major factor responsible for an increase in serious group A streptococcus infections. *The Journal of infectious diseases* 2007;195:1625–1633. [PubMed: 17471432]
- Sakota V, Fry AM, Lietman TM, Facklam RR, Li ZY, Beall B. Genetically diverse group A streptococci from children in Far-Western Nepal share high genetic relatedness with isolates from other countries. *Journal of clinical microbiology* 2006;44:2160–2166. [PubMed: 16757615]
- Sanderson-Smith ML, Dowton M, Ranson M, Walker MJ. The plasminogen-binding group A streptococcal M protein-related protein Prp binds plasminogen via arginine and histidine residues. *Journal of bacteriology* 2007;189:1435–1440. [PubMed: 17012384]
- Scaramuzzino DA, McNiff JM, Bessen DE. Humanized *in vivo* model for streptococcal impetigo. *Infect Immun* 2000;68:2880–2887. [PubMed: 10768985]
- Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. *Clin Microb Rev* 2004;17:14–56.
- Schneewind O, Jones KF, Fischetti VA. Sequence and structural characterization of the trypsin-resistant T6 surface protein of group A streptococci. *J Bacteriol* 1990;172:3310–3317. [PubMed: 2188957]
- Scott J, Thompson-Mayberry P, Lahmamsi S, King CJ, McShan WM. Phage-associated mutator phenotype in group A streptococcus. *Journal of bacteriology* 2008;190:6290–6301. [PubMed: 18676670]
- Seppala H, Klaukka T, Vuopio-Varkila J, Muotiala A, Helenius H, Lager K, Huovinen P. The effect of changes in the consumption of macrolide antibiotics on erythromycin resistance in group A streptococci in Finland. Finnish Study Group for Antimicrobial Resistance. *The New England journal of medicine* 1997;337:441–446. [PubMed: 9250845]
- Shulman S. Group A streptococcal pharyngitis serotype surveillance in North America, 2000–2002. *Clin Infect Dis* 2004;39:325–332. [PubMed: 15306998]
- Shulman ST, Stollerman G, Beall B, Dale JB, Tanz RR. Temporal changes in streptococcal M protein types and the near-disappearance of acute rheumatic fever in the United States. *Clin Infect Dis* 2006;42:441–447. [PubMed: 16421785]
- Smeesters PR, Vergison A, Campos D, de Aguiar E, Deyi VY, Van Melderen L. Differences between Belgian and Brazilian Group A *Streptococcus* Epidemiologic Landscape. *PLoS ONE* 2006;1:e10. [PubMed: 17183632]
- Smoot JC, Barbian KD, Van Gompel JJ, Smoot LM, Chaussee MS, Sylva GL, Sturdevant DE, Ricklefs SM, Porcella SF, Parkins LD, Beres SB, Campbell DS, Smith TM, Zhang Q, Kapur V, Daly JA, Veasy LG, Musser JM. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proceedings of the National Academy of Sciences of the United States of America* 2002;99:4668–4673. [PubMed: 11917108]
- Stalhammar-Carlalm M, Areschoug T, Larsson C, Lindahl G. The R28 protein of *Streptococcus pyogenes* is related to several group B streptococcal surface proteins, confers protective immunity and promotes binding to human epithelial cells. *Molecular microbiology* 1999;33:208–219. [PubMed: 10411737]
- Stollerman GH, Dale JB. The importance of the group A streptococcus capsule in the pathogenesis of human infections: a historical perspective. *Clin Infect Dis* 2008;46:1038–1045. [PubMed: 18444821]

- Sumby P, Whitney AR, Graviss EA, DeLeo FR, Musser JM. Genome-wide analysis of group A streptococci reveals a mutation that modulates global phenotype and disease specificity. *PLoS pathogens* 2006;2:e5. [PubMed: 16446783]
- Svensson MD, Scaramuzzino DA, Sjöbring U, Olsen A, Frank C, Bessen DE. Role for a secreted cysteine proteinase in the establishment of host tissue tropism by group A streptococci. *Mol Microbiol* 2000;38:242–253. [PubMed: 11069651]
- Svensson MD, Sjöbring U, Bessen DE. Selective distribution of a high-affinity plasminogen binding site among group A streptococci associated with impetigo. *Infect Immun* 1999;67:3915–3920. [PubMed: 10417156]
- Svensson MD, Sjöbring U, Luo F, Bessen DE. Roles of the plasminogen activator streptokinase and plasminogen-associated M protein in an experimental model for streptococcal impetigo. *Microbiology* 2002;148:3933–3945. [PubMed: 12480897]
- Terao Y, Kawabata S, Kunitomo E, Murakami J, Nakagawa I, Hamada S. Fba, a novel fibronectin-binding protein from *Streptococcus pyogenes*, promotes bacterial entry into epithelial cells, and the fba gene is positively transcribed under the Mga regulator. *Mol Microbiol* 2001;42:75–86. [PubMed: 11679068]
- Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, Wessels MR, Paulsen IT, Nelson KE, Margarit I, Read TD, Madoff LC, Wolf AM, Beanan MJ, Brinkac LM, Daugherty SC, DeBoy RT, Durkin AS, Kolonay JF, Madupu R, Lewis MR, Radune D, Fedorova NB, Scanlan D, Khouri H, Mulligan S, Carty HA, Cline RT, Van Aken SE, Gill J, Scarselli M, Mora M, Iacobini ET, Brettoni C, Galli G, Mariani M, Vegni F, Maione D, Rinaudo D, Rappuoli R, Telford JL, Kasper DL, Grandi G, Fraser CM. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *PNAS* 2002;99:12391–12396. [PubMed: 12200547]
- Tewodros W, Kronvall G. M protein gene emm type analysis of group A beta-hemolytic streptococci from Ethiopia reveals unique patterns. *Journal of clinical microbiology* 2005;43:4369–4376. [PubMed: 16145079]
- Turner KM, Hanage WP, Fraser C, Connor TR, Spratt BG. Assessing the reliability of eBURST using simulated populations with known ancestry. *BMC microbiology* 2007;7:30. [PubMed: 17430587]
- Upton M, Tagg JR, Wescombe P, Jenkinson HF. Intra- and interspecies signaling between *Streptococcus salivarius* and *Streptococcus pyogenes* mediated by SalA and SalA1 lantibiotic peptides. *J Bacteriol* 2001;183:3931–3938. [PubMed: 11395456]
- Valardo PE, Montanari MP, Giovanetti E. Genetic Elements Responsible for Erythromycin Resistance in *Streptococci*. *Antimicrobial agents and chemotherapy*. 2008
- Walker MJ, Hollands A, Sanderson-Smith ML, Cole JN, Kirk JK, Henningham A, McArthur JD, Dinkla K, Aziz RK, Kansal RG, Simpson AJ, Buchanan JT, Chhatwal GS, Kotb M, Nizet V. DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. *Nature medicine* 2007;13:981–985.
- Walls T, Power D, Tagg J. Bacteriocin-like inhibitory substance BLIS production by the normal flora of the nasopharynx: potential to protect against otitis media? *Journal of medical microbiology* 2003;52:829–833. [PubMed: 12909662]
- Wannamaker LW. Differences between streptococcal infections of the throat and of the skin. *N Engl J Med* 1970;282:23–31. [PubMed: 4905440]
- Wertz JE, McGregor KF, Bessen DE. Detecting key structural features within highly recombined genes. *PLoS computational biology* 2007;3:e14. [PubMed: 17257051]
- Wescombe PA, Upton M, Dierksen KP, Ragland NL, Sivabalan S, Wirawan RE, Inglis MA, Moore CJ, Walker GV, Chilcott CN, Jenkinson HF, Tagg JR. Production of the lantibiotic salivaricin A and its variants by oral streptococci and use of a specific induction assay to detect their presence in human saliva. *Applied and environmental microbiology* 2006;72:1459–1466. [PubMed: 16461700]
- Wood DN, Chaussee MA, Chaussee MS, Buttar BA. Persistence of *Streptococcus pyogenes* in stationary-phase cultures. *Journal of bacteriology* 2005;187:3319–3328. [PubMed: 15866916]
- Woodbury RL, Klammer KA, Xiong Y, Bailiff T, Glennen A, Bartkus JM, Lynfield R, Van Beneden C, Beall BW. Plasmid-Borne erm(T) from invasive, macrolide-resistant *Streptococcus pyogenes* strains. *Antimicrobial agents and chemotherapy* 2008;52:1140–1143. [PubMed: 18180360]

Figure 1.

Maximum parsimony tree based on *emm* and FCT region genotypes. FCT region genes (*rofA/nra*, *prtF1*, *cpa*, *prtF2*) and *emm* region genes (*mga*, *sof*) provide the 6 characters for the unrooted tree generated by maximum parsimony. Character states correspond to gene presence/absence or lineage. Taxa (N = 11), representing $\geq 1\%$ of the > 100 *S. pyogenes* strains evaluated, were used to generate the tree by the branch-and-bound search method. Each circle at the nodes represents an individual strain, and is further defined as belonging to *emm* pattern group A–C (open), D (filled) or E (hatched). The starting set of 114 strains, from which the 11 taxa are derived, is also depicted in Figure 3.

A, *S. pyogenes* **B, *S. agalactiae*** **Figure 2.**

Population snapshots generated by eBURST. MLST data from online sources (Aanensen, 2008, Jolley, 2008) is displayed as single eBURST diagrams, by setting the group definition to 0 of 7 shared housekeeping alleles. STs that are single locus variants (SLVs) are connected by a line. (A), *S. pyogenes*, and (B), *S. agalactiae*.

Figure 3.

Phylogenetic tree based on concatenated housekeeping gene sequences. The nucleotide sequences of the 7 housekeeping genes used in MLST were concatenated ($N = 3,134$ sites) and a tree constructed by the minimum evolution method using the Kimura 2-parameter substitution model. Bootstrap analysis (500 replicates) was performed. Included are data from 114 *S. pyogenes* isolates representing 114 distinct STs and 113 distinct *emm* types, as described in (Bessen et al., 2005); a subset of these strains are depicted in Figure 1. The *emm* pattern group of each strain is indicated by bars positioned to the right of each taxon: A–C (open), D (filled) and E (hatched). Nucleotide diversity (π) equals 0.00756.

Table 1 The association of *emm* pattern with tissue site of infection

Location of population based surveillance	Pharyngitis (or tonsillitis)				# of isolates	Impetigo				Reference for study
	# of isolates	% of isolates that are emm pattern:				# of isolates	% of isolates that are emm pattern			
		A-C	D	E			A-C	D	E	
Australia, tropical	NR				125	13	46	41	Bessen et al., 2000	
Rome (Italy)	114	50	1	48	n.d.				Dicunzo et al., 2001	
Germany	216	51	0	49	n.d.				Brandt et al., 2001	
Spain	520	32	1	68	n.d.				Alberti et al., 2003	
Mexico	282	54	1	44	n.d.				Espinosa et al., 2003	
USA	> 1,900	53	1	47	n.d.				Shulman et al., 2004	
Ethiopia	57	26	28	35	47	0	32	47	Tewodros et al., 2005	
Brazil	52	17	21	62	58	3	57	40	Smeesters et al. 2006	
Brussels (Belgium)	163	55	0	45	NR				Smeesters et al. 2006	
Nepal	NR				53	19	30	51	Sakota et al., 2006	
Australia, tropical	NR				129	13	53	35	McDonald et al., 2007b	
average		42	7	50		10	44	43		
s.d.		15	11	10		8	12	6		
paired t-test, A-C versus D	0.005					0.008				

NR, < 10 isolates recovered; n.d., not determined

Table 2
Summary of genotype and phenotype correlates of *emm* pattern

Genotype or phenotype	Equivalent genome position (kb)	<i>emm</i> pattern group		
		A-C	D	E
Preferred tissue site of infection	n/a	throat (specialist)	skin (specialist)	throat and skin (generalist)
RofA/Nra, transcriptional regulator	115	nearly all have <i>rofA</i> lineage alleles	nearly all have <i>nra</i> lineage alleles	nearly all have <i>rofA</i> lineage alleles
<i>prfI</i> , encoding fibronectin-binding protein	117	most positive	most negative	most positive
<i>cpa</i> , encoding collagen-binding protein of pilus	120	most negative	most positive	most positive
Streptokinase, plasminogen activator	1674	all have (sub)cluster 1 or 2a alleles	many have subcluster 2b alleles	nearly all have cluster 1 alleles
PAM, plasminogen-binding M protein	1711	all negative	many positive	all negative
Mga, transcriptional regulator	1714	nearly all have lineage 1 alleles	all have lineage 2 alleles	all have lineage 2 alleles
<i>sof</i> , encoding fibronectin-binding protein	1728	nearly all negative	nearly all negative	all positive

* Based on the 1.90 Mb genome of M28 strain MGAS6180