JOURNAL OF CLINICAL MICROBIOLOGY, July 2009, p. 2046–2054 Vol. 47, No. 7
0095-1137/09/$08.00+0 doi:10.1128/JCM.00246-09

Genetic Relationships Deduced from *emm* and Multilocus Sequence Typing of Invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* Recovered from Isolates Collected in the United States▽

Yusra Ahmad,[1] Robert E. Gertz, Jr.,[1] Zhongya Li,[1] Varja Sakota,[1] Laura N. Broyles,[2] Chris Van Beneden,[1] Richard Facklam,[1] P. Lynn Shewmaker,[1] Arthur Reingold,[4,5] Monica M. Farley,[2,3] and Bernard W. Beall[1]*

Respiratory Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia[1]; Division of Infectious Diseases, Emory University School of Medicine, Atlanta, Georgia[2]; Atlanta Veterans Affairs Medical Center, Atlanta, Georgia[3]; California Emerging Infections Program, Oakland, California[4]; and University of California at Berkeley School of Public Health, Berkeley, California[5]

Received 4 February 2009/Accepted 16 April 2009

Beta-hemolytic group C and G streptococci cause a considerable invasive disease burden and sometimes cause disease outbreaks. Little is known about the critical epidemiologic parameter of genetic relatedness between isolates. We determined the *emm* types of 334 *Streptococcus dysgalactiae* subsp. *equisimilis* isolates, and attempted *emm* typing of 5 *Streptococcus canis* isolates from a recent population-based surveillance for invasive isolates. Thirty-four *emm* types were observed, including one from *S. canis*. We formulated multilocus sequence typing (MLST) primers with six of the seven loci corresponding to the *Streptococcus pyogenes* MLST scheme. We performed MLST with 65 of the 334 surveillance isolates (61 *S. dysgalactiae* subsp. *equisimilis* isolates, 4 *S. canis* isolates) to represent each *emm* type identified, including 2 to 3 isolates for each of the 25 redundantly represented *emm* types. Forty-one MLST sequence types (STs) were observed. Isolates within 16 redundantly represented *S. dysgalactiae* subsp. *equisimilis emm* types shared identical or nearly identical STs, demonstrating concordance between the *emm* type and genetic relatedness. However, seven STs were each represented by two to four different *emm* types, and 7 of the 10 *S. dysgalactiae* subsp. *equisimilis* eBURST groups represented up to six different *emm* types. Thus, *S. dysgalactiae* subsp. *equisimilis* isolates were similar to *S. pyogenes* isolates, in that strains of the same *emm* type were often highly related, but they differed from *S. pyogenes*, in that *S. dysgalactiae* subsp. *equisimilis* strains with identical or closely similar STs often exhibited multiple unrelated *emm* types. The phylogenetic relationships between *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes* alleles revealed a history of interspecies recombination, with either species often serving as genetic donors. The four *S. canis* isolates shared highly homologous alleles but were unrelated clones without evidence of past recombination with *S. dysgalactiae* subsp. *equisimilis* or *S. pyogenes*.

Streptococcus pyogenes, *Streptococcus dysgalactiae* subsp. *equisimilis*, and *Streptococcus canis* are beta-hemolytic pyogenic species that are highly genetically related on the basis of 16S rRNA sequence comparisons, with *S. canis* being the closest relative to *S. pyogenes* (39, 16). In recent years, *S. dysgalactiae* subsp. *equisimilis* has increasingly been reported as a cause of invasive disease (11, 24, 36), yet the epidemiology and population genetics of this species is poorly understood. *S. canis* is a commensal and opportunistic pathogen of dogs and other animals (13, 16). *S. canis* rarely causes disease in humans (7, 39); however, its incidence in human disease is unknown.

As with *S. pyogenes*, *S. dysgalactiae* subsp. *equisimilis* isolates are almost always *emm* typeable (8, 11, 24, 39), with over 50 known *emm* types (*emm* types are downloadable from ftp://ftp.cdc.gov/pub/infectious_diseases/biotech/tsemm/). During 1998, we documented the occurrence of *emm* type *stG1389* from an invasive *S. canis* infection in a dog (unpublished data), and we also recently noticed the same sequence in the GenBank database (GenBank accession number EU195120). To our knowledge, this is the only *emm* type documented from *S. canis*. Several other virulence determinants are known to be shared between *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes* (9, 10, 12, 19, 25). *S. pyogenes* exotoxin genes have been detected within *S. dysgalactiae* subsp. *equisimilis* and *S. canis*, and lysogenic transfers of prophages carrying superantigen genes between *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis* have been documented (20, 22, 31, 33, 38).

M-protein gene (*emm*) typing has been useful as a simple genetic tool for identifying and resolving *Streptococcus pyogenes* strains for epidemiologic studies (18, 27, 29, 35). This is consistent with the observation that within given regions, group A streptococcal (GAS) *emm* types are often restricted to only one or two defined clonal complexes, and often, these common clones are predominant in diverse locations (15) (available at www.mlst.net). In direct contrast, a previous report described a poor concordance between the *emm* type and genetic relatedness in *S. dysgalactiae* subsp. *equisimilis* (23). That report additionally showed evidence of the lateral movement of housekeeping genes between the two species, with the majority of gene transfer events involving *S. pyogenes* donors and *S. dysgalactiae* subsp. *equisimilis* recipients.

* Corresponding author. Mailing address: CDC *Streptococcus* Lab, 1600 Clifton Rd., NE, MS-C02, Atlanta, GA 30329. Phone: (404) 639-1237. Fax: (404) 639-4215. E-mail: beb0@cdc.gov.

▽ Published ahead of print on 22 April 2009.

We recently carried out population-based surveillance for invasive disease caused by beta-hemolytic streptococci of groups other than A and B, and the results from 2 years of surveillance were reported (7). The majority of case isolates (about 80%) were *S. dysgalactiae* subsp. *equisimilis*. The only other *emm*-typeable species identified by active surveillance was *S. canis*, of which only one of five *S. canis* isolates was *emm* typeable. We found that the burden of invasive disease caused by *S. dysgalactiae* subsp. *equisimilis* is comparable to that caused by invasive GAS (29) and affects similar adult populations (7). In the work presented here, we describe newly discovered relationships between multilocus sequence types and *emm* types among an expanded collection of invasive *S. dysgalactiae* subsp. *equisimilis* and *S. canis* isolates recovered in the United States from 2002 to 2005.

MATERIALS AND METHODS

Isolates. Surveillance for invasive streptococcal disease caused by beta-hemolytic streptococci of groups other than A and B was conducted as part of CDC's Active Bacterial Core surveillance from 2002 to 2005 in the eight-county metropolitan Atlanta, GA, area (population, 4,464,200 in 2003) and in 2003 and 2004 in the three-county San Francisco, CA, Bay Area (population, 3,213,848 in 2003). Isolates were recovered from a normally sterile site (e.g., blood, cerebrospinal fluid, or bone) from cases that included only residents of the two surveillance areas. The isolates were initially identified as non-group A and non-group B streptococci at local hospital laboratories by using standard, commercially available latex agglutination kits. All available isolates were sent to CDC for verification and further characterization. The results of *emm* typing for 212 *S. dysgalactiae* subsp. *equisimilis* isolates and 5 *S. canis* isolates were reported previously (7). An additional 122 *S. dysgalactiae* subsp. *equisimilis* isolates were collected through continued surveillance through 2005. The combined group of 344 invasive *S. dysgalactiae* subsp. *equisimilis* and *S. canis* isolates were further evaluated by multilocus sequence typing (MLST).

Species identification. Conventional biochemical tests were used for the identification of the beta-hemolytic streptococcal species, as described previously (7) and at http://www.cdc.gov/ncidod/biotech/strep/strep-doc/index.htm.

***emm* typing.** Crude template preparation and *emm* typing were performed as described at http://www.cdc.gov/ncidod/biotech/strep/protocols.htm. For *S. pyogenes*, the M-protein gene (*emm*) types were determined exactly as described at http://www.cdc.gov/ncidod/biotech/strep/protocol_emm-type.htm. All *emm* type-specific sequences described here are downloadable from the CDC database (ftp://ftp.cdc.gov/pub/infectious_diseases/biotech/tsemm/).

MLST. The 65 isolates were chosen from among 334 invasive *S. dysgalactiae* subsp. *equisimilis* and *S. canis* isolates to represent all different *emm* types and, when possible, to represent multiple independent isolates that shared the same *emm* type. When possible, we selected isolates that shared the same *emm* type and that were collected in different years and from different surveillance areas. Included among these 65 isolates were 4 of the 5 *S. canis* isolates recovered, 1 of which was *emm* typeable.

MLST was performed with seven primer sets based on the deduced closest matches of *S. dysgalactiae* subsp. *equisimilis* to the *S. pyogenes* MLST targets (*gki*, *gtr*, *murI*, *mutS*, *recP*, *xpt*, and *yqiL*), described at http://spyogenes.mlst.net/. These primer sequences are posted at http://www.cdc.gov/ncidod/biotech/strep/protocols.htm. To formulate these primers, we employed a nearly completed *S. dysgalactiae* subsp. *equisimilis* genome sequence provided by Awdhesh Kalia (381 contigs, approximately 2,000,000 bp, 300 to 20,000 reads each). Within this genome sequence, no target was highly similar (>93% identity) to *yqiL* sequences previously described from *S. dysgalactiae* subsp. *equisimilis* (23); however, we did identify a less similar putative acetyl coenzyme A acetyltransferase gene that we designated *yqiZ* and that shared about 65% identity to the closest matches from *S. pyogenes*. We were unable to amplify a *yqiL*-like gene; however, portions of *yqiZ* were variably amplified with primers described previously (23) or with the *S. pyogenes yqiL*-specific primers described at http://spyogenes.mlst.net/. Thus, we formulated optimal *yqiZ*-specific primers and substituted *yqiZ* into our scheme in place of *yqiL*. Our MLST scheme additionally differed from that described previously for *S. dysgalactiae* subsp. *equisimilis* (23) in that instead of generating 430-bp *gtr* sequences, we generated 450-base *gtr* sequences that correspond to the *S. pyogenes* MLST scheme. These seven primer sets also appear to provide a potentially reliable MLST scheme for *S. canis*. In summary, our MLST scheme directly corresponds to the existing system for *S. pyogenes*, except for the inclusion of *yqiZ* instead of *yqiL*.

MLST allele nomenclature. Four-digit designations (Table 1) were assigned to alleles newly discovered during this study, with the exception that the sequences of *S. canis mutS* alleles 5426, 6318, and 8881 and *S. dysgalactiae* subsp. *equisimilis mutS* allele 8343 perfectly matched the sequences with GenBank accession numbers AJ413209, AJ413210, EF619610, and AJ413208, respectively. Three-digit designations correspond to alleles discovered during a previous study (23). Two-digit designations (*xpt31* and *recp70*) correspond to alleles previously documented at the *S. pyogenes* MLST database (available at www.mlst.net). Five-digit designations (three digits separated by a hyphen from two digits) correspond to previously documented designations for *S. dysgalactiae* subsp. *equisimilis* alleles (23) that are also documented in the *S. pyogenes* MLST database (the first three digits and last two digits correspond to previous designations in *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes*, respectively; see, for example, *mutS106-46* in Table 1).

eBURST analysis. eBURST (version 3) analysis was performed with the program provided at http://spyogenes.mlst.net/eburst/with the MLST data generated from the study isolates. An eBURST group consists of a set of seven-locus allelic profiles in which each profile is related to at least one other profile in the set by sharing six of seven alleles.

Phylogenetic analysis. Wisconsin Package programs (Wisconsin Package, version 10.3; Accelrys Inc., San Diego, CA) were used for all DNA sequence manipulations. Individual alleles (405 to 498 bases in length) and six-allele concatenated sequences totaling 2,700 bases (*yqiL* was not used since we could not amplify this gene from *S. dysgalactiae* subsp. *equisimilis*) from each different allelic combination found during this study were aligned precisely by use of the PileUp program. We used the Distances program to generate a matrix of the pairwise evolutionary distance between the aligned sequences. For analysis of individual loci, all known *S. pyogenes* MLST alleles and the alleles from *S. dysgalactiae* subsp. *equisimilis* and *S. canis* found in this study were included. For analysis of concatenated six-allele profiles, all study STs and 21 STs selected from the *S. pyogenes* MLST database were employed. These *S. pyogenes* STs were inclusive of certain major *S. pyogenes* clones found during U.S. surveillance and also of highly diverse STs from *S. pyogenes* that contained *S. dysgalactiae* subsp. *equisimilis*-like alleles. Tree construction was done by the neighbor-joining approach by use of the Jukes-Cantor model. One thousand bootstrap replicates were performed with the same data by using the Paupsearch program to give the frequencies of occurrence of bipartitions.

Nucleotide sequence accession numbers. GenBank accession numbers EU768803 to EU768808 were assigned to the seven newly discovered *gki* alleles; GenBank accession numbers EU768892, EU938013 to EU938015, EU981290, EU999155, EU999156, FJ238468, and FJ414797-FJ515799 were assigned to the *gtr* alleles; GenBank accession numbers EU999157, FJ151276 to FJ151278, FJ156731, FJ128469, and FJ238470 were assigned to the *murI* alleles; GenBank accession numbers FJ238471 to FJ238474 were assigned to the new *mutS* alleles; GenBank accession numbers FJ238475 to FJ238481 and EU283342 were assigned to the *recP* alleles; GenBank accession numbers FJ238482 to FJ238488 were assigned to the *xpt* alleles; and GenBank accession numbers FJ348682, FJ263615, FJ263616, and FJ238646 to FJ268656 were assigned to the *yqiZ* alleles. The sequences of *S. canis mutS* alleles 5426, 6318, and 8881 and *S. dysgalactiae* subsp. *equisimilis mutS* allele 8343 perfectly matched the sequences with GenBank accession numbers AJ413209, AJ413210, EF619610, and AJ413208, respectively.

RESULTS

***emm* type diversity.** Thirty-three *emm* types were observed among the 334 invasive *S. dysgalactiae* subsp. *equisimilis* isolates, with five types each accounting for 20 to 40 isolates (Fig. 1). These *S. dysgalactiae* subsp. *equisimilis* isolates consisted of 275 group G streptococcal (GGS) strains, 58 group C streptococcal (GCS) strains, and 1 group L streptococcal (GLS) strain. Nine *emm* types represented both GGS and GCS strains. The type found from the single GLS strain, *stL1929*, has previously been shown to be associated only with GLS strains isolated as long as 28 years ago (unpublished observations; see http://www.cdc.gov/ncidod/biotech/strep/types_emm103-124.htm and ftp://ftp.cdc.gov/pub/infectious_diseases/biotech/tsemm/ for limited information regarding individual *emm* types). All five *S. canis* isolates were

TABLE 1. *emm* type, allelic compositions, and demographics associated with 10 *S. dysgalactiae* subsp. *equisimilis* eBURST groups and 41 MLST types of *S. dysgalactiae* subsp. *equisimilis* and *S. canis*

Genetic group	ST, group(s)	*emm* type(s) (no. of isolates, state, yr recovered)[a]	Allele						
			gki[b,c]	*gtr*[b,c]	*murI*[b,c]	*mutS*[b]	*recP*[b,c]	*xpt*[b,c]	*yqiZ*[d]
eBURST 1	Se6 (G)	***stG11*** (1, GA, 2005)	**106**	**105**	**105**	**106-46**[e]	**113-50**[e]	104	**8337**
	Se7 (G)	***stG480*** (1, CA, 2005)	**106**	**108**	**105**	**106-46**[e]	**113-50**[e]	104	**8343**
	Se8 (G)	*stC839* (1, GA, 2003) ***stG480*** (1, GA, 2002) *stG7860* (1, GA, 2004) ***stG11*** (1, CA, 2003)	**106**	**108**	**105**	**106-46**[e]	**113-50**[e]	104	**8337**
	Se9 (G)	*stC6979* (1, GA, 2004)	**106**	**108**	**105**	**106-46**[e]	**113-50**[e]	104	**6458**
eBURST 2	Se12	*stG643* (1, GA, 2002)	**107**	**6458**	**4127**	**105**	**6458**	**109**	**6458**
	Se16	*stG3442* (1, CA, 2003)	**107**	**6458**	**105**	**105**	**6458**	104	**6458**
	Se17	***stG2078*** (3, CA, GA, 04, 05), *stG485* (1, GA, 2004)	**107**	**6458**	**105**	**105**	**6458**	**109**	**6458**
	Se36	*stG245* (1, CA, 2003)	**107**	**6458**	**105**	**105**	**6458**	**109**	**8343**
eBURST 3	Se18	***stG7882***(1, GA, 2003)	**107**	**106**	**110**	**106-46**[e]	104	**108**	**6458**
	Se19	***stC6979*** (1, CA, 2003)	**108-74**[e]	**6742**	**110**	**106-46**[e]	104	**108**	**6458**
	Se31	***stG6792*** (1, CA, 2004)	**108-74**[e]	**106**	**110**	**106-46**[e]	104	**108**	**6458**
	Se4	***stG97*** (2, GA, 2002, 2005), ***stC5344*** (2, GA, 2005), ***stG6792*** (1, GA, 2004), ***stG7882*** (1, CA, 2005)	**105**	**106**	**110**	**106-46**[e]	104	**108**	**6458**
eBURST 4	Se27	***stC6746*** (1, GA, 2004)	**105**	**106**	**110**	**106-46**[e]	**7917**	**6746**	**6746**
	Se5	***stC6746*** (1, CA, 2003)	**105**	**106**	**110**	**106-46**[e]	**7505**	**6746**	**6746**
eBURST 5	Se20	***stG62647*** (2, CA, GA, 2005)	**108-74**[e]	**105**	**107**	**6532**	105-2[e]	**109**	**6532**
	Se33	***stG62647*** (1, CA, 2003)	**108-74**[e]	**6742**	**107**	**6532**	105-2[e]	**109**	**6532**
eBURST 6	Se34	***stC1400*** (2, GA, 2003, 2004), *stG5063* (1, CA, 2005)	**108-74**[e]	**7934**	**110**	**106-46**[e]	7934	**31sp**[e]	**3499**
	Se35	*stC7934* (1, GA, 2003)	**108-74**[e]	**7934**	**110**	**106-46**[e]	7934	**31sp**[e]	**6458**
eBURST 7	Se26	***stG840*** (1, CA, 2003)	**108-74**[e]	**106**	**105**	**106-46**[e]	102	**112**	**6458**
	Se30	***stG840*** (1, GA, 2004)	**108-74**[e]	**106**	**110**	**106-46**[e]	102	**112**	**6458**
eBURST 8	Se15	***stG10*** (2, CA, GA, 02, 2004), *stG6* (1, CA, 2003), *stG166B* (1, GA, 2002), *stG652* (1, GA, 2002)	**108-74**[e]	**105**	**107**	**105**	105-2[e]	**110**	**6458**
	Se21	*stG245* (1, GA, 2002)	**108-74**[e]	**6742**	**107**	**105**	105-2[e]	**110**	**6458**
eBURST 9	Se10	*stC7505* (1, GA, 2005)	**9182**	**6458**	**7505**	**7505**	**7505**	**9182**	**9182**
	Se14	*stC9431* (1, GA, 2004)	**9182**	**6458**	**7505**	**6461**	**7505**	**9182**	**9182**
eBURST 10	Se24	*stG6* (1, GA, 2005)	**108-74**[e]	**106**	**105**	**8343**	4012	103-2[e]	**8343**
	Se25	***stG5420*** (2, CA, GA, 2004, 2005), *stG166B* (1, CA, 2004)	**108-74**[e]	**106**	**105**	**8343**	102	103-2[e]	**8343**
Single	Se3	***emm57*** (2, CA, GA, 2005), ***stG653*** (2, GA, 2002, 2005)	**103**	**105**	**110**	**106-46**[e]	101-83[e]	107-71[e]	**6317**
	Se1	*stL1929* (1, GA, 2003)	**9182**	**9182**	**9182**	**6461**	**7505**	**9182**	**9182**
	Se11	*stC7901* (1, CA, 2004)	**7901**	**105**	**110**	**106-46**[e]	106-4[e]	**108**	**9544**
	Se13	*stC9431* (1, CA, 2005)	**9182**	**6461**	**9182**	**6461**	**7505**	**9182**	**6461**
	Se2	*stG5063* (1, GA, 2004)	**103**	**6458**	**110**	**106-46**[e]	106-4[e]	**31sp**[e]	**6458**
	Se22	*stG643* (1, GA, 2004)	**108-74**[e]	**105**	**107**	**6532**	**113-50**[e]	**6532**	**6532**
	Se23	*stG2574* (1, CA, 2005)	**108-74**[e]	**105**	**110**	**106-46**[e]	**112**	104	**6317**
	Se28	*stC1400* (1, CA, 2004)	**108-74**[e]	**105**	**110**	**105**	70sp[e]	8983	**6458**
	Se29	***stC74a*** (2, CA, 2003)	**108-74**[e]	**106**	**110**	**105**	102	104	**8343**
	Se32	*stG652* (1, CA, 2004)	**108-74**[e]	**106**	**110**	**106-46**[e]	**107**	**112**	**9544**
	Se37	*stG485* (1, CA, 2005)	**108-74**[e]	**106**	**105**	**106-46**[e]	107	104	**8343**
	Sc1	*stG1389* (1, GA, 2002)	*4074*	*4074*	*4074*	*4074*	*8881*	Not done	*4074*
	Sc2	Nontypeable (1, GA, 2002)	*6318*	*5426*	*6318*	*6318*	*6318*	*6318*	*6318*
	Sc3	Nontypeable (1, CA, 2003)	*8881*	*8881*	*8881*	*8881*	*8881*	*8881*	*8881*
	Sc4	Nontypeable (1, GA, 2005)	*5426*	*5426*	*5426*	*5426*	*5426*	*5426*	*5426*

[a] Isolates of the 16 boldface *emm* types exhibit identical or highly similar multilocus STs (no more than one allelic difference), with the exception that one of the three *stC1400* isolates was genetically divergent from the other two. *emm* types that are underlined are represented by only a single isolate.

[b] Three-digit designations correspond exactly to alleles discovered during the previous study (23), with the exception that our *gtr* sequences were 450 bases in length (as in http://spyogenes.mlst.net/) but shared identity in their overlap with the 430-base entries described previously (23). Four-digit designations correspond to alleles discovered during this study.

[c] Boldface entries correspond to cluster II GGS/GCS-like alleles, according to previously described phylogenic analysis and the analysis performed during this study. Nonboldface entries correspond to cluster I GAS-like alleles. *S. canis* alleles belong to neither category and are italicized.

[d] *yqiZ* is specific for the MLST scheme presented in this study, since we were unable to amplify the *yqiL* homolog of *S. pyogenes* that was described in the previous study (23).

[e] Designations indicate alleles additionally found within *S. pyogenes* according to our previous data and according to data recorded at http://spyogenes.mlst.net/. In designations with hyphens (e.g., gki108-74), the first three-digit number refers to the designation described for *S. dysgalactiae* subsp. *equisimilis* in the previous study (23), while the number after the hyphen refers to the designation given for the same allele in *S. pyogenes* at http://spyogenes.mlst.net/. *recP70sp* and *xpt31sp* were not previously identified from *S. dysgalactiae* subsp. *equisimilis* and correspond to alleles found in the *S. pyogenes* MLST database.

GGS, with only one of the isolates yielding an *emm* amplicon. This *S. canis emm* type was included in the CDC *emm* database in 1999 and corresponded to *st1389* from an *S. canis* isolate recovered from a dog. The sequence of type *stG1389* shares 99% identity in its 402-base overlap with the sequence with GenBank accession number EU195120 (from an *S. canis* isolate from the throat of a dog), deposited in the GenBank database in December 2007. None of the 34 *emm* types (including *stG1389* from *S.*

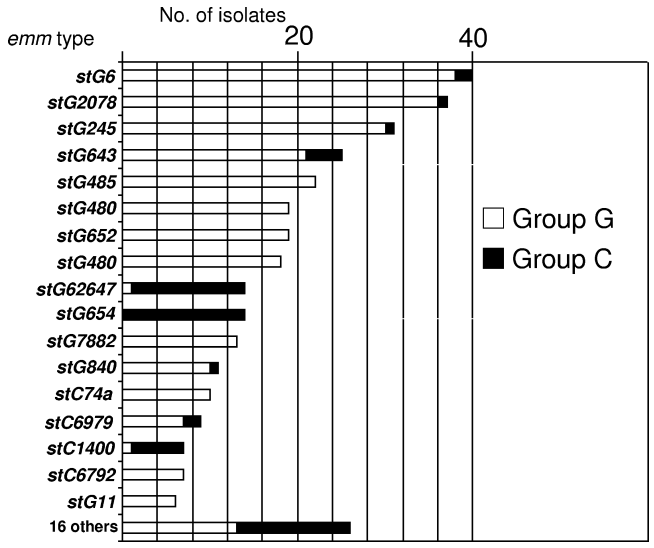

FIG. 1. *emm* type distribution observed among 334 *S. dysgalactiae* subsp. *equisimilis* isolates recovered from population-based surveillance of invasive streptococcal disease in areas of California and Georgia from 2002 to 2005.

canis) were obvious derivatives of each other, and all shared less than 87% identity within the 50 codons encoding the predicted N terminus of the M protein, with most types sharing less than 50% identity. One type, *emm57*, is documented from reference strains of M serotype 57 *S. pyogenes* (4, 40) and was found in two GCS *S. dysgalactiae* subsp. *equisimilis* isolates recovered in California and Georgia in 2005 (Table 1; *emm57* corresponds to MLST type Se3, which was not included within an eBURST group).

MLST results. For the seven targets, the number of different alleles encountered among the 61 *S. dysgalactiae* subsp. *equisimilis* isolates ranged from as few as 6 (*murI* and *mutS*) to as many as 14 (*recP*) (Table 1). Sixty-two distinct alleles were found, 27 of which were identical to previously discovered alleles within a set of 34 *S. dysgalactiae* subsp. *equisimilis* isolates (23).

Our results differed markedly from those of a previous study in which each *S. dysgalactiae* subsp. *equisimilis* isolate corresponded to 1 of 34 distinct allelic profiles (23). In our study of 61 *S. dysgalactiae* subsp. *equisimilis* isolates, we found 37 allelic profiles, with 9 profiles each accounting for multiple (two to five) isolates. Also in contrast to the findings of the previous study (23), in which only a single pair of isolates shared the same eBURST group (by virtue of sharing six of seven identical alleles), our *S. dysgalactiae* subsp. *equisimilis* isolate set revealed 10 eBURST groups that represented 46 of the 61 isolates, with up to four different allelic profiles and 6 different isolates being represented in individual eBURST groups (Table 1). In addition, three individual STs that were not within eBURST groups accounted for three pairs of isolates.

We successfully applied the MLST protocol used for *S. dysgalactiae* subsp. *equisimilis* to our small set of four *S. canis* isolates, except for the failure to amplify *xpt* from one *S. canis* isolate. There were no related allelic profiles among the four *S. canis* isolates (Table 1), with only one instance of a shared allele between two isolates (*recP8881*) occurring.

Concordance between *emm* type and genetic relatedness in *S. dysgalactiae* subsp. *equisimilis*. Twenty-five *emm* types found among *S. dysgalactiae* subsp. *equisimilis* isolates were represented in this MLST study by two to three independent isolates (Table 1). For 16 of these 25 *emm* types, concordance was observed between the *emm* type and genetic relatedness, in that the isolates shared the same eBURST group; and for 10 of these 16 concordant sets, isolates with the same *emm* type shared the same ST. Fifteen of these 16 concordant sets displayed close genetic relatedness (allelic identity or a single divergent allele) among all members of an *emm* type included in the study (two isolates each of types *emm57*, *stC5344*, *stC6746*, *stC74a*, *stG10*, *stG11*, *stG480*, *stG5420*, *stG653*, *stG6792*, *stG7882*, *stG840*, and *stG97* and three isolates each of *stG2078* and *stG62647*), with *stC1400* representing two isolates with an identical allelic profile and a third isolate with a divergent allelic profile. In addition to the 16 observed examples of concordance between *emm* type and genetic relatedness, we found that the ST-Se37 (*stG485*) detected in this study shared allelic identity (excluding *yqiZ*) with a previously determined genotype (designated ST21) from a GCS isolate of the same *emm* type that was recovered prior to 2001 (23).

High degree of genetic relatedness between *S. dysgalactiae* subsp. *equisimilis* isolates with different *emm* types. The 10 small eBURST groups, consisting of two to four STs each, consisted solely of highly related isolates (Table 1). In only 3 of the 10 eBURST groups were double-locus variants observed (data not shown), and in all other eBURST groups, all isolates were single-locus variants of one another. As mentioned above, in 15 of the 16 examples of observed concordance between the *emm* type and genetic relatedness, all isolates within a given *emm* type were encompassed within a single eBURST group or ST.

It was striking that only 2 of the 10 eBURST groups were represented by a single *emm* type. Seven individual STs were represented by two to four distinct *emm* types, and 8 of 10 eBURST groups represented two to five different *emm* types. In contrast to *S. pyogenes*, in which single- and double-locus variants usually share the same *emm* types, four of these *S. dysgalactiae* subsp. *equisimilis* eBURST groups displayed five different *emm* types.

Genotypes shared between GGS and GCS isolates. For five different *emm* types, both GGS and GCS isolates were chosen for genetic analysis. Unrelated genotypes (only zero to three identical alleles) were observed between GGS and GCS isolates within three of these five *emm* types (*stC1400*, *stG5063*, and *stG643*), while GGS and GCS isolates within types *st5344* and *st62647* shared six to seven identical alleles.

Evidence of housekeeping gene transfer between *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes*. Since *S. pyogenes*, *S. canis*, and *S. dysgalactiae* subsp. *equisimilis* are close relatives, we wished to use the MLST scheme described previously (15, 23) and a more recent collection of invasive *S. dysgalactiae* subsp. *equisimilis* and *S. canis* isolates (7) to assess potential housekeeping gene flow between the three species.

Within *gki*, *mutS*, *recP*, and *xpt* there were a total of 10 alleles found in our *S. dysgalactiae* subsp. *equisimilis* data set that have also been documented within *S. pyogenes* (Table 1). These included, *gki108-74*, *mutS106-46*, *recP101-83*, *recP105-2*, *recP106-4*, *recP113-50*, *recP70sp*, *xpt103-2*, *xpt107-*

71, and *xpt31sp*. Fifty *S. dysgalactiae* subsp. *equisimilis* isolates contained one to three of these alleles that had previously been documented in *S. pyogenes*.

***gtr* and *murI* data.** Unlike the previous study (23), our study set *S. dysgalactiae* subsp. *equisimilis gtr* and *murI* alleles showed no examples of allelic identity or similarity between the two species. As reported previously (23), the *S. pyogenes gtr* and *murI* alleles display very little variation. The 8 *S. dysgalactiae* subsp. *equisimilis gtr* alleles shared ≥97.8% sequence identity but only ≤82% identity with 78 of 79 known *S. pyogenes gtr* alleles. The *gtr60* allele in the *S. pyogenes* MLST database differed by only a single base from the *S. dysgalactiae* subsp. *equisimilis* allele *gtr106* (99.8% sequence identity); however, there is no existing documentation of an *S. pyogenes* strain with this allele. The 6 *S. dysgalactiae* subsp. *equisimilis murI* alleles shared ≥98.4% sequence identity but ≤77% identity with the 70 known *S. pyogenes murI* alleles. These alleles were closely similar to 5 *S. dysgalactiae* subsp. *equisimilis murI* alleles from a previous study (23), and these 11 alleles clustered together in a species-specific cluster in the phylogram shown in Fig. 2A.

***gki* data.** The *gki108-74* allele shared a much higher degree of similarity with other *S. dysgalactiae* subsp. *equisimilis* alleles (≥96%) than nearly all other *S. pyogenes* alleles (≤92% identity to 98/101 *S. pyogenes* alleles), indicating that the potential interspecies transfer of this allele was from an *S. dysgalactiae* subsp. *equisimilis* donor to *S. pyogenes*; however, there is no past documentation of the association of the *gki74* allele with any specific GAS allelic profile or strain. Two additional *S. dysgalactiae* subsp. *equisimilis*-like *gki* alleles are evident in the *S. pyogenes* database, and these share 95 to 98.2% identity with the *S. dysgalactiae* subsp. *equisimilis gki* alleles in Table 1. These include *gki79* and *gki75*, found in GAS strains recovered in Australia and Nepal, respectively (28, 34).

***mutS* data.** The *S. dysgalactiae* subsp. *equisimilis mutS* alleles shared ≥98.3% identity and 86 to 92% identity with the 56 *S. pyogenes*-specific *mutS* alleles (available at www.mlst.net). The *mutS106-46* allele, found in both *S. pyogenes* and in this study, shared ≥98.8% identity with the other five *S. dysgalactiae* subsp. *equisimilis mutS* alleles encountered in this study, suggesting the possible transfer of this allele from *S. dysgalactiae* subsp. *equisimilis* into *S. pyogenes*. With the exception of the *mutS62* allele, all *S. dysgalactiae* subsp. *equisimilis*, *S. pyogenes*, and *S. canis mutS* alleles were each resolved by phylogenetic analysis into separate, well-supported, species-specific clusters (data not shown), as shown for the data for *murI* (Fig. 2A). The *mutS62* allele, found in two separate instances from *S. pyogenes* isolates causing invasive disease in rhesus monkeys (17), shared only 95.1% and 92% identity with the closest-matching alleles of *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis*, respectively.

***recP* and *xpt* data.** Unlike *gki*, *gtr*, *murI*, and *mutS*, for which phylogenetic analysis primarily resolved the alleles from *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis* into two simple clusters, the *recP* and *xpt* allele clusters were heterogeneous (the *recP* phylogram is shown in Fig. 2B).

Of the 14 *S. dysgalactiae* subsp. *equisimilis recP* alleles found in this study, 5 (alleles 70, 101-83, 105-2, 106-4, 113-50) have previously been observed in reported *S. pyogenes* isolates, although only *recP* alleles 70, 2, 4, and 50 are associated at mlst.net with complete STs that correspond to verified *S. pyogenes* strains with typical *S. pyogenes emm* types. By using a tree based upon the Distances program (Fig. 2B), *recP*113-50 and five other *S. dysgalactiae* subsp. *equisimilis recP* alleles from this study (alleles 112se, 7505, 7917, 107se, and 6458) clustered together within a distinct branch with six *S. dysgalactiae* subsp. *equisimilis* alleles from a previous study (alleles 114seb, 110seb, 111seb, 112seb, 115seb, and 108seb) that were described as cluster II alleles characteristic of *S. dysgalactiae* subsp. *equisimilis* (21). Additionally, the *S. pyogenes recP* alleles (*recP* alleles 15, 47, 29, 94, and 51) clustered within this main group of *S. dysgalactiae* subsp. *equisimilis*-like *recP* alleles (Fig. 2B). The remaining eight *recP S. dysgalactiae* subsp. *equisimilis* alleles found during this study (alleles 70, 4012, 7934, 101-83, 102se, 104se,105-2, and 106-4) clustered together with the majority of known *S. pyogenes* alleles and are categorized as cluster I GAS-like alleles by use of the criteria from the previous study (23).

The *xpt* and *recP* data sets from *S. dysgalactiae* subsp. *equisimilis* shared similarities, in that several alleles were previously documented in *S. pyogenes*, revealing that both *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes* could serve as recombinational donors or recipients of these housekeeping alleles. Three *xpt* alleles found in this study (alleles 103-2, 107-71, and 31sp) have previously been documented in *S. pyogenes*. *xpt* allele 31sp clustered with five additional closely similar *S. dysgalactiae* subsp. *equisimilis* alleles from this study (alleles 6746, 110se, 6532, 112se, and 108se) and four more *S. dysgalactiae* subsp. *equisimilis* cluster II *S. dysgalactiae* subsp. *equisimilis*-like alleles from the previous study (23) (alleles 114seb, 115seb, 113seb, and 111seb). Included within cluster II were *S. pyogenes xpt* alleles 73, 51, 29, and 62. *xpt* alleles 103-2 and 107-71 were typical of cluster I and were closely similar to various *S. pyogenes* alleles. Additional *S. dysgalactiae* subsp. *equisimilis* cluster I alleles found in this study were *xpt104se* and *xpt898304*, which were in the same major branch as four previously documented *S. dysgalactiae* subsp. *equisimilis* alleles (data not shown).

***yqiZ* data.** As described in Materials and Methods, we were unable to amplify sequences from our *S. dysgalactiae* subsp. *equisimilis* isolates that corresponded to the *yqiL* MLST alleles described earlier (23), and a closely similar target was also not apparent in a nearly completed *S. dysgalactiae* subsp. *equisimilis* genome sequence (Awdhesh Kalia, personal communication). Among our *S. dysgalactiae* subsp. *equisimilis* isolate set, 10 different alleles of *yqiZ* were observed, and these shared ≥97.2% identity and only 65 to 85% identity to various *S. dysgalactiae* subsp. *equisimilis yqiL* alleles described previously (23).

***S. canis* MLST data.** There were no examples of the detection of alleles in the four *S. canis* isolates that were also found in the two other species. Within the seven MLST targets, all the *S. canis* alleles shared at least 97.5% identity (for six alleles, ≥98.5% identity) and shared no more than 79 to 94.2% identity within individual MLST targets to the closest *S. pyogenes* or *S. dysgalactiae* subsp. *equisimilis* alleles.

Phylogenetic analysis. We used the Distances program to generate a matrix of the pairwise evolutionary distance between aligned sequences. We found that the dendrograms generated from these data clustered sequences in a manner concordant with the previous, more detailed phylogenetic analysis

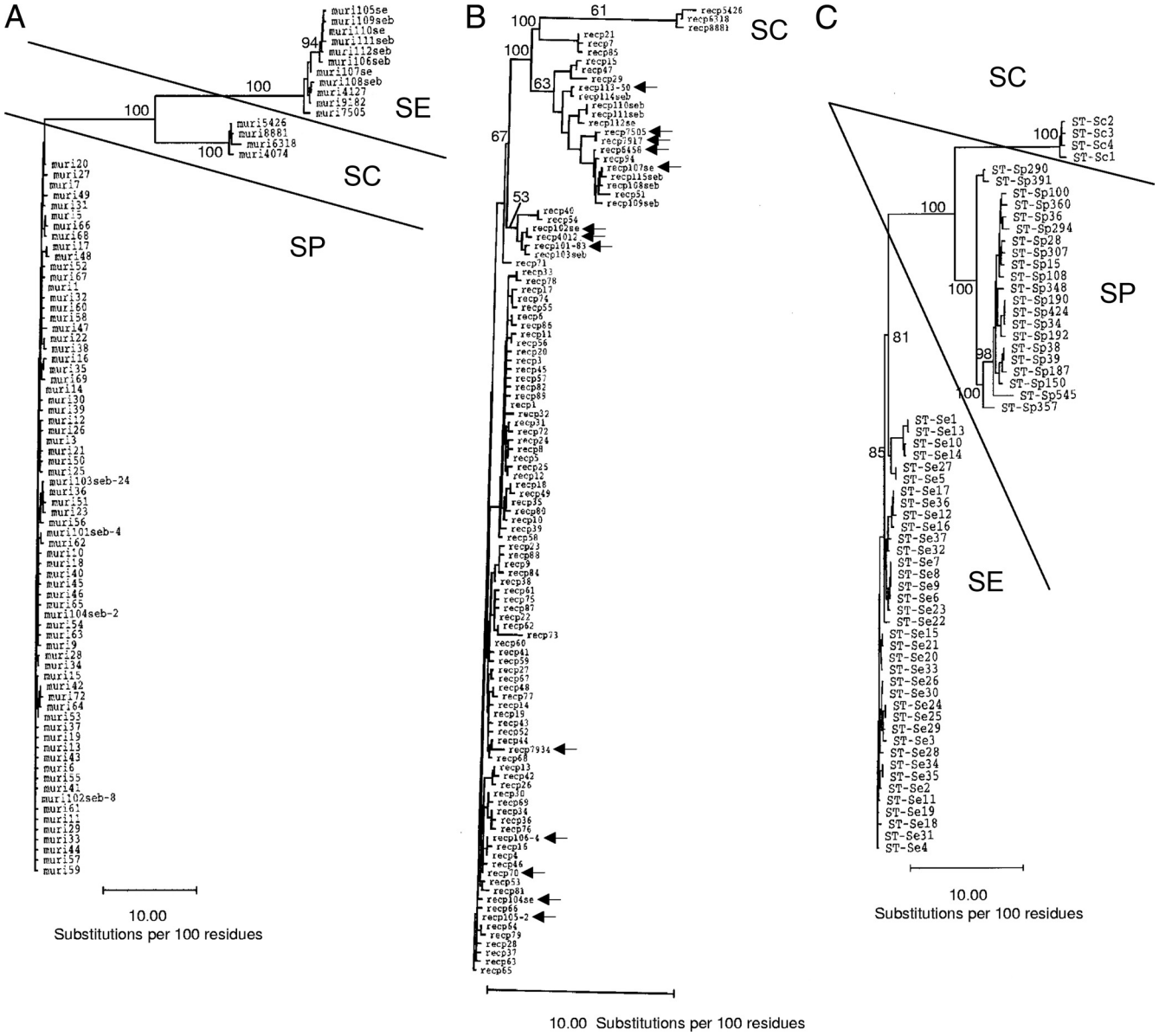

FIG. 2. Phylograms of *murI* (A); *recP* (B); and concatenated *gki*, *gtr*, *murI*, *mutS*, *recP*, and *xpt* (C) MLST data. We employed the neighbor-joining approach, using the Jukes-Cantor model. One thousand bootstrap replicates were performed on the same data by using the Paupsearch program to give the indicated frequencies of occurrence of bipartitions. seb, alleles exclusively found in a previous study of *S. dysgalactiae* subsp. *equisimilis* isolates. All known *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis murI* and *recP* alleles (as of June 2008) were also included in the analysis. Alleles with hyphens have been documented in both *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis* (B and C). Arrows show the *S. dysgalactiae* subsp. *equisimilis recP* alleles found during this study (B). In panel C, *murI* alleles are distinctly clustered in the indicated species-specific clusters (*S. dysgalactiae* subsp. *equisimilis*, *S. canis*, and *S. pyogenes*), while *recP* alleles from *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes* are not as distinctly separated. For the phylogram in panel C, the analysis included all STs from *S. dysgalactiae* subsp. *equisimilis* and *S. canis* found in this study and the *S. pyogenes* STs included in Table 2. For ST-Sc1 in panel C, the *xpt6318* allele from another *S. canis* strain was employed since we were unable to amplify *xpt* from this strain. SP, *S. pyogenes*; SE, *S. dysgalactiae* subsp. *equisimilis*; SC, *S. canis*.

(23). Using our dendrograms to construct trees for allelic data generated for combined *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes* data, we assigned different alleles as GAS-like or GGS/GCS-like according to how the alleles clustered with typical *S. pyogenes* alleles represented within the entire *S. pyogenes* MLST database and with the *S. dysgalactiae* subsp. *equisimilis* alleles generated in the previous study (23). The trees for *murI* and *recP* are shown in Fig. 2A and B, respectively. Equivalent trees for the other five MLST alleles are available upon request.

Resolution of the three species by use of MLST data. For the construction of phyograms from concatenated MLST data, we selected the most diverse existing *S. pyogenes* allelic profiles as well as allelic profiles from some of the most common global clones, including the prevalent M1, M3, M75, M12, and M4 strains. Four typical *S. pyogenes* allelic profiles (from predom-

inant types *emm1*, *emm3*, and *emm12* and the rarer type *emm100* clone) were characterized by the use of highly conserved MLST alleles (data not shown). We also included diverse allelic profiles containing known *S. dysgalactiae* subsp. *equisimilis*-like alleles and other obviously divergent alleles.

We employed concatenated MLST allele sequences from six of the seven targets (excluding *yqiZ*) of all 65 GGS/GCS isolates from this study and the corresponding sequences from documented *S. pyogenes* strains. Despite clear instances in which *S. pyogenes*-like and *S. dysgalactiae* subsp. *equisimilis*-like MLST alleles were shared between *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes*, phylograms of concatenated allelic data from all 37 *S. dysgalactiae* subsp. *equisimilis* STs and 21 *S. pyogenes* STs clearly resolved the two species (Fig. 2C).

DISCUSSION

A major purpose of this study was to examine the genetic diversity of a recent sample of population-based invasive *S. dysgalactiae* subsp. *equisimilis* isolates that were recovered in the United States and that represented all *emm* types recovered. We also wished to make limited observations from an analogous analysis of a very small sample of four *S. canis* isolates. The main contribution of this work regarding *S. canis* is simply a method for MLST that should prove useful for epidemiologic and other studies. While we were unable to amplify *xpt* from one of the four distinct *S. canis* strains, we doubt that this will be a serious impediment for large MLST projects. Even with *S. pyogenes* and pneumococci, we find rare individual loci that will not amplify with standard MLST primers.

Among the *S. pyogenes* isolates recovered from economically developed geographic regions and associated with either invasive disease or pharyngitis, *emm* types are often highly associated with specific clonal types (1, 2, 3, 6, 14, 21). In order to assess clonal relationships among *S. dysgalactiae* subsp. *equisimilis* isolates, our study included multiple members of 25 different *emm* types. It has been known for at least the past decade that certain *S. pyogenes emm* types are associated with multiple clonal types that can easily be resolved by phenotypic or genotypic testing (1, 4, 5, 34). For example, based upon a combination of phenotypic and genotypic data (T typing, antiopacity factor testing, macrorestriction profile analysis, *sof* gene sequence typing), classic M44-M61 and M27-M77 *S. pyogenes* reference strain pairs are obviously divergent clones that share the same *emm* sequence types (1, 2, 4). Ten *S. dysgalactiae* subsp. *equisimilis emm* types presented here (*stC1400*, *stC6979*, *stC9431*, *stG166B*, *stG245*, *stG485*, *stG5063*, *stG6*, *stG643*, *stG652*) each show associations with two different clonal types (Table 1; note that there was a concordant relationship between *emm* type and genetic relatedness in two of the three *stC1400* isolates).

In the previous study, 34 distinct allelic profiles were observed among 34 *S. dysgalactiae* subsp. *equisimilis* isolates, with only a single pair of conserved but nonidentical allelic profiles being detected (23). In striking contrast, we found that more than 75% (46/61) of our invasive *S. dysgalactiae* subsp. *equisimilis* isolates were resolved into 10 different eBURST groups comprised of highly similar allelic profiles, with 9 of 37 allelic profiles being completely shared between multiple isolates. In addition, in the previous study (23) no genetic relatedness between isolates within all nine *emm* types that represented multiple isolates was observed, while we found a high degree of genetic relatedness between multiple isolates of the majority of *emm* types examined (16/25 [64%]). These marked differences may be due in part to the greater geographic and temporal diversities of the isolates described in the previous study (23) compared to the diversities of the isolates evaluated in the present study, all of which were recovered within a 3-year period in the United States. In addition, at least 6 of the 34 isolates from the previous study were not recovered from sterile sites. We do note, however, the identical allelic profile between an invasive *stG485* isolate from our study (ST-Se37) that was recovered at least 8 years later than an invasive *stG485* isolate described in the previous study (designated ST21). We received this *stG485* strain, whose isolation date is unknown, in 1997 and supplied it for the study by Kalia et al. (23).

Similar to observations made with *S. pyogenes* (1, 5), in at least some cases the *S. dysgalactiae* subsp. *equisimilis emm* type is not predictive of genetic relatedness. For example, the two isolates of the most common *emm* type chosen, *stG6* (Fig. 1), had divergent allelic profiles (Table 2), and in total, 9 of the 25 *emm* types represented by multiple isolates displayed no concordance between genetic relatedness and *emm* type. It is likely that additional tests could easily resolve strains that share an identical *emm* type distributed among two or more genetic backgrounds in the same manner that T-antigen profiles or *sof* sequence types currently resolve the analogous situation in GAS (1, 2, 3, 14).

A considerable fraction (7/37 [19%]) of the allelic profiles represented multiple *emm* types, with up to four *emm* types associated with individual allelic profiles. Furthermore, for the majority (8/10 [80%]) of eBURST groups, within which isolates generally varied from one another by only a single locus, multiple *emm* types were observed (up to six different types). It is unusual for *S. pyogenes* isolates that differ by one or two MLST loci to represent multiple *emm* types (28). It is even more rare for a single allelic profile to be associated with *S. pyogenes* isolates of divergent *emm* types, although within at least some of these instances, evidence of intragenic changes within the *emm* gene that result in a new *emm* type is apparent (28).

Our study shares basic similarities to a previous study of clonal relationships between *S. dysgalactiae* subsp. *equisimilis* invasive and noninvasive disease isolates collected in Portugal over a 6-year period (30), in which a majority of the clones deduced by pulsed-field gel electrophoresis included more than one *emm* type and the same *emm* type was sometimes found among isolates with diverse genetic backgrounds. This study also indicated a significant association in *S. dysgalactiae* subsp. *equisimilis* between certain clone-*emm* type combinations and invasive disease.

The results of our study also indicate that the *emm* type is likely to be a useful genetic marker for epidemiologic studies and other investigations involving *S. dysgalactiae* subsp. *equisimilis*, as it is useful for studies of *S. pyogenes*. It is entirely conceivable that the 16 instances in which all or nearly all *S. dysgalactiae* subsp. *equisimilis* isolates within a given *emm* type shared identical or closely similar allelic profiles are generally reflective of the clonality of these *emm* types at the national

level. Obviously, a more thorough analysis of invasive *S. dysgalactiae* subsp. *equisimilis* isolates sharing the same *emm* type is necessary to verify such a projection. It also appears that additional parameters will be necessary to resolve major strains of *S. dysgalactiae* subsp. *equisimilis*, since multiple *emm* types are often superimposed upon the same allelic profile or set of highly related allelic profiles (Table 1). While it is likely that in *S. dysgalactiae* subsp. *equisimilis* certain *emm* types are largely predictive of clonal type in the same manner that certain *emm* types are predictive of clonal type in *S. pyogenes*, the situation appears to be more complex in *S. dysgalactiae* subsp. *equisimilis*. For example, while the common *emm* type *stG2078* was associated with ST-Se17 among all three isolates of this *emm* type recovered from Georgia and California, an isolate carrying an unrelated *emm* type (*stG485*) also shared ST-Se17 (Table 2). In order to use *S. dysgalactiae* subsp. *equisimilis emm* types as a meaningful parameter in epidemiologic studies, it is important to elucidate and understand *emm* type correlates with clonal relatedness. From the findings of the work presented here, it appears that *emm* type deduction alone, while informative, is often insufficient for the resolution of *S. dysgalactiae* subsp. *equisimilis* invasive clones.

Besides the obvious assessment of correlations of *S. dysgalactiae* subsp. *equisimilis emm* types with specific clones, we wished to assess if we could deduce a history of housekeeping gene transfer between *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes*, and if so, we wondered if we could deduce an obvious net directionality in transfer events between the two species. Previous work suggested a net directionality of housekeeping gene transfer from *S. pyogenes* donors to *S. dysgalactiae* subsp. *equisimilis* recipients, resulting in highly mosaic *S. dysgalactiae* subsp. *equisimilis* genomes. This conclusion was based primarily upon the observations that at the time known *S. pyogenes* housekeeping gene alleles displayed little diversity and a heterogeneous collection of *S. dysgalactiae* subsp. *equisimilis* isolates exhibited a great deal of genetic diversity because they contained one to four *S. pyogenes*-like housekeeping gene sequences. The great contrast between our observations and those from the previous study (23) in the degree of mosaicism observed within *S. dysgalactiae* subsp. *equisimilis* due to the presence of *S. pyogenes*-like MLST alleles is quite apparent when the MLST data are compared. Whereas 10 of 34 isolates in the previous study revealed *S. pyogenes*-like alleles for three to four MLST targets, no isolates in our data set contained more than two *S. pyogenes*-like alleles. *S. pyogenes*-like alleles were completely absent within *gki*, *gtr*, *murI*, and *mutS* loci from our 61-isolate data set, while 9 of 34 isolates from the previous study contained one to two *S. pyogenes*-like alleles within these four targets. Typical alleles of these four targets are much less related between the two species, while the *S. dysgalactiae* subsp. *equisimilis* and *S. pyogenes recP* and *xpt* alleles are much more closely related, as evidenced by the relatively smaller number of substitutions per 100 residues seen between *recP* allelic clusters most characteristic of each species compared to the markedly greater separation in *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis murI* alleles (Fig. 2A and B). In fact, when the five isolates from the previous study with an *S. pyogenes*-like allele at either *gki*, *gtr*, *murI*, or *mutS* were included within the same phylogenetic analysis with concatenated alleles shown in Fig. 2C, the five isolates were poorly resolved from other *S. dysgalactiae* subsp. *equisimilis* strains and formed a separate branch (data not shown). The four isolates from the previous study with two *S. pyogenes*-like alleles within these four loci actually showed greater shared relatedness to *S. pyogenes* than to *S. dysgalactiae* subsp. *equisimilis* and formed a subgroup connected with the main *S. pyogenes* branch shown (data not shown). At present, it appears that assessments of housekeeping gene transfer directionality between *S. pyogenes* and *S. dysgalactiae* subsp. *equisimilis* are heavily dependent upon the sources of the isolate sets. The majority of MLST results that have been documented for *S. pyogenes* have been based on collections from developed countries. More recent limited studies from more diverse *S. pyogenes* isolate sets from less developed countries have revealed a disproportionate number of *S. dysgalactiae* subsp. *equisimilis*-like alleles (26, 34). We have also found an *S. pyogenes* strain (represented by ST545 in Fig. 2C) that was recovered on two different occasions from monkeys with invasive infections (17) (data not shown) and that contained two different *S. dysgalactiae* subsp. *equisimilis*-like housekeeping alleles (Table 2). It is possible that the cumulative MLST databases are not sufficiently reflective of the global diversity of either of these species to make generalizations regarding the net directionality of gene flow between them.

TABLE 2. Concatenated MLST types from documented *S. pyogenes* strains used for comparison with *S. dysgalactiae* subsp. *equisimilis* and *S. canis* MLST types from this study

ST	Associated *emm* type	*S. dysgalactiae* subsp. *equisimilis*-like or other divergent allele	Reference or contributor of *S. pyogenes* allele
ST34	*emm49*	*recP7*	15
ST38	*emm4*	*recP15*	15
ST39	*emm4*	*recP15*	15
ST100	*emm55*	*xpt29*	26, 34
ST108	*st3765*	*xpt31*	26
ST150	*emm75*	*recp113-50*[a]	37
ST187	*st213*	*recP47*	26
ST190	*emm49*	*recP21*, *mutS101-7*[a]	26
ST192	*emm25*	*xpt51*	26
ST290	*emm81*	*gki79*	26
ST294	*st6030*	*xpt51*	34
ST348	*emm57*	*mutS106-46*[a]	34
ST357	*emm75*	*gki75*	34
ST360	*st1759*	*xpt62*	34
ST391	*emm4*	*recp15*	32
ST424	*emm49*	*recP85*, *mutS101-7*	D. R. Martin
ST545	*st804*	*recP94*, *xpt73*	17
ST15	*emm3*	None	15
ST28	*emm1*	None	15
ST36	*emm12*	None	15
ST307	*emm100*	None	27

[a] *recP113-50*, *mutS101-7*, and *mutS106-46* correspond to *S. pyogenes* alleles *recP50*, *mutS7*, and *mutS46*, respectively.

ACKNOWLEDGMENTS

We sincerely thank Awdhesh Kalia for sharing unpublished sequence data from his *S. dysgalactiae* subsp. *equisimilis* genome project. We are grateful to the California and Georgia Active Bacterial Core surveillance participants in this study.

REFERENCES

1. **Beall, B., G. Gherardi, M. Lovgren, R. R. Facklam, B. A. Forwick, and G. J. Tyrrell.** 2000. *emm* and *sof* gene sequence variation in relation to serological typing of opacity-factor-positive group A streptococci. Microbiology **146:** 1195–1209.
2. **Beall, B., R. R. Facklam, J. A. Elliott, A. R. Franklin, T. Hoenes, D. Jackson, L. Laclaire, T. Thompson, and R. Viswanathan.** 1998. Streptococcal *emm* types associated with T-agglutination types and the use of conserved *emm* gene restriction fragment patterns for subtyping group A streptococci. J. Med. Microbiol. **47:**893–898.
3. **Beall, B., R. Facklam, T. Hoenes, and B. Schwartz.** 1997. Survey of *emm* gene sequences and T-antigen types from systemic *Streptococcus pyogenes* infection isolates collected in San Francisco, California; Atlanta, Georgia; and Connecticut in 1994 and 1995. J. Clin. Microbiol. **35:**1231–1235.
4. **Beall, B., R. Facklam, and T. Thompson.** 1996. Sequencing *emm*-specific PCR products for routine and accurate typing of group A streptococci. J. Clin. Microbiol. **34:**953–958.
5. **Bessen, D. E., K. F. McGregor, and A. M. Whatmore.** 2008. Relationships between *emm* and multilocus sequence types within a global collection of *Streptococcus pyogenes*. BMC Microbiol. **8:**59.
6. **Bessen, D. E., J. R. Carapetis, B. Beall, R. Katz, M. Hibble, B. J. Currie, T. Collingridge, M. W. Izzo, D. A. Scaramuzzino, and K. S. Sriprakash.** 2000. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. J. Infect. Dis. **182:**1109–1116.
7. **Broyles, L. N., C. Van Beneden, B. Beall, R. Facklam, P. L. Shewmaker, P. Malpiedi, P. Daily, A. Reingold, and M. M. Farley.** 2009. Population-based study of invasive disease caused by beta-hemolytic streptococci of groups other than A and B. Clin. Infect. Dis. **48:**706–712.
8. **Campo, R. E., D. R. Schultz, and A. L. Bisno.** 1995. M proteins of group G streptococci: mechanisms of resistance to phagocytosis. J. Infect. Dis. **171:** 601–606.
9. **Chhatwal, G. S., and S. R. Talay.** 2000. Pathogenicity factors in C and G streptococci, p. 177–183. *In* V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, and J. I. Rood (ed.), Gram-positive pathogens. American Society for Microbiology, Washington, DC.
10. **Cleary, P. P., J. Peterson, C. Chen, and C. Nelson.** 1991. Virulent human strains of group G streptococci express a C5a peptidase enzyme similar to that produced by group A streptococci. Infect. Immun. **59:**2305–2310.
11. **Cohen-Poradosu, R., J. Jaffe, D. Lavi, S. Grisariu-Greenzaid, R. Nir-Paz, L. Valinsky, M. Dan-Goor, C. Block, B. Beall, and A. E. Moses.** 2004. Group G streptococcal bacteremia in Jerusalem. Emerg. Infect. Dis. **10:**1455–1460.
12. **Davies, M. R., D. J. McMillan, R. G. Beiko, V. Barroso, R. Geffers, K. S. Sriprakas, and G. S. Chhatwal.** 2007. Virulence profiling of *Streptococcus dysgalactiae* subspecies *equisimilis* isolated from infected humans reveals 2 distinct genetic lineages that do not segregate with their phenotypes or propensity to cause diseases. Clin. Infect. Dis. **44:**1442–1454.
13. **Devriese, L. A., J. Hommez, R. Kilpper-Bälz, and K.-H. Schleifer.** 1986. *Streptococcus canis* sp. nov.: a species of group G streptococci from animals. Int. J. Syst. Bacteriol. **36:**422–425.
14. **Dicuonzo, G., G. Gherardi, G. Lorino, S. Angeletti, M. De Cesaris, E. Fiscarelli, D. E. Bessen, and B. Beall.** 2001. Group A streptococcal genotypes from pediatric throat isolates in Rome, Italy. J. Clin. Microbiol. **39:** 1687–1690.
15. **Enright, M. C., B. G. Spratt, A. Kalia, J. H. Cross, and D. E. Bessen.** 2001. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. Infect. Immun. **69:**2416–2427.
16. **Facklam, R.** 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15:**613–630.
17. **Garcia, A., K. Paul, B. Beall, and H. McClure.** 2006. Toxic shock due to Streptococcus pyogenes in a rhesus monkey (Macaca mulatta). J. Am. Assoc. Lab. Anim. Sci. **45:**79–82.
18. **Gooskens, J., A. J. Neeling, R. J. Willems, J. W. Wout, and E. J. Kuijper.** 2005. Streptococcal toxic shock syndrome by an iMLS resistant M type 77 *Streptococcus pyogenes* in The Netherlands. Scand. J. Infect. Dis. **37:**85–89.
19. **Humar, D., V. Datta, D. J. Bast, B. Beall, J. C. De Azavedo, and V. Nizet.** 2002. Streptolysin S and necrotising infections produced by group G streptococcus. Lancet **359:**124–129.
20. **Igwe, E. I., P. L. Shewmaker, R. R. Facklam, M. M. Farley, C. van Beneden, and B. Beall.** 2003. Identification of superantigen genes *speM*, *ssa*, and *smeZ* in invasive strains of beta-hemolytic group C and G streptococci recovered from humans. FEMS Microbiol. Lett. **229:**259–264.
21. **Johnson, D. R., E. L. Kaplan, A. VanGheem, R. R. Facklam, and B. Beall.** 2006. Characterization of group A streptococci (*Streptococcus pyogenes*): correlation of M-protein and *emm*-gene type with T-protein agglutination pattern and serum opacity factor. J. Med. Microbiol. **55:**157–164.
22. **Kalia, A., and D. E. Bessen.** 2003. Presence of streptococcal pyrogenic exotoxin A and C genes in human isolates of group G streptococci. FEMS Microbiol. Lett. **219:**291–295.
23. **Kalia, A., M. C. Enright, B. G. Spratt, and D. E. Bessen.** 2001. Directional gene movement from human-pathogenic to commensal-like streptococci. Infect. Immun. **69:**4858–4869.
24. **Lopardo, H. A., P. Vidal, M. Sparo, P. Jeric, D. Centron, R. R. Facklam, H. Paganini, N. G. Pagniez, M. Lovgren, and B. Beall.** 2005. Six-month multicenter study on invasive infections due to *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis* in Argentina. J. Clin. Microbiol. **43:** 802–807.
25. **Malke, H.** 2000. Genetics and pathogenicity factors of group C and G streptococci, p. 163–176. *In* V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, and J. I. Rood (ed.), Gram-positive pathogens. American Society for Microbiology, Washington, DC.
26. **McGregor, K. F., N. Bilek, A. Bennett, A. Kalia, B. Beall, J. R. Carapetis, B. J. Currie, K. S. Sriprakash, B. G. Spratt, and D. E. Bessen.** 2004. Group A streptococci from a remote community have novel multilocus genotypes but share *emm* types and housekeeping alleles with isolates from worldwide sources. J. Infect. Dis. **189:**717–723.
27. **McGregor, K. F., and B. G. Spratt.** 2005. Identity and prevalence of multilocus sequence typing-defined clones of group A streptococci within a hospital setting. J. Clin. Microbiol. **43:**1963–1967.
28. **McGregor, K. F., B. G. Spratt, A. Kalia, A. Bennett, N. Bilek, B. Beall, and D. E. Bessen.** 2004. Multilocus sequence typing of *Streptococcus pyogenes* representing most known *emm* types and distinctions among subpopulation genetic structures. J. Bacteriol. **186:**4285–4294.
29. **O'Loughlin, R. E., A. Roberson, P. R. Cieslak, R. Lynfield, K. Gershman, A. Craig, B. A. Albanese, M. M. Farley, N. L. Barrett, N. L. Spina, B. Beall, L. H. Harrison, A. Reingold, C. Van Beneden, and Active Bacterial Core Surveillance Team.** 2007. The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000-2004. Clin. Infect. Dis. **45:**853–862.
30. **Pinho, M. D., J. Melo-Cristino, and M. Ramirez.** 2006. Clonal relationships between invasive and noninvasive Lancefield group C and G streptococci and *emm*-specific differences in invasiveness. J. Clin. Microbiol. **44:**841–846.
31. **Proft, T., P. D. Webb, V. Handley, and J. D. Fraser.** 2003. Two novel superantigens found in both group A and group C *Streptococcus*. Infect. Immun. **71:**1361–1369.
32. **Robinson, D. A., J. A. Sutcliffe, W. Tewodros, A. Manoharan, and D. E. Bessen.** 2006. Evolution and global dissemination of macrolide-resistant group A streptococci. Antimicrob. Agents Chemother. **50:**2903–2911.
33. **Sachse, S., P. Seidel, D. Gerlach, E. Günther, J. Rödel, E. Straube, and K. H. Schmidt.** 2002. Superantigen-like gene(s) in human pathogenic *Streptococcus dysgalactiae*, subsp *equisimilis*: genomic localisation of the gene encoding streptococcal pyrogenic exotoxin G (*speG*(dys)). FEMS Immunol. Med. Microbiol. **34:**159–167.
34. **Sakota, V., A. M. Fry, T. M. Lietman, R. R. Facklam, Z. Li, and B. Beall.** 2006. Genetically diverse group A streptococci from children in far-western Nepal share high genetic relatedness with isolates from other countries. J. Clin. Microbiol. **44:**2160–2166.
35. **Shulman, S. T., R. R. Tanz, W. Kabat, K. Kabat, E. Cederlund, D. Patel, Z. Li, V. Sakota, J. B. Dale, B. Beall, and U.S. Streptococcal Pharyngitis Surveillance Group.** 2004. Group A streptococcal pharyngitis serotype surveillance in North America, 2000-2002. Clin. Infect. Dis. **39:**325–332.
36. **Sylvetsky, N., D. Raveh, Y. Schlesinger, B. Rudensky, and A. M. Yinnon.** 2002. Bacteremia due to beta-hemolytic Streptococcus group G: increasing incidence and clinical characteristics of patients. Am. J. Med. **112:**622–626.
37. **Szczypa, K., E. Sadowy, R. Izdebsk, and W. Hryniewicz.** 2004. A rapid increase in macrolide resistance in *Streptococcus pyogenes* isolated in Poland during 1996-2002. J. Antimicrob. Chemother. **54:**828–831.
38. **Vojtek, I., Z. A. Pirzada, B. Henriques-Normark, M. Mastny, R. P. Janapatla, and E. Charpentier.** 2008. Lysogenic transfer of group A Streptococcus superantigen gene among streptococci. J. Infect. Dis. **197:**225–234.
39. **Whatmore, A. M., K. H. Engler, G. Gudmundsdottir, and A. Efstratiou.** 2001. Identification of isolates of *Streptococcus canis* infecting humans. J. Clin. Microbiol. **39:**4196–4199.
40. **Whatmore, A. M., V. Kapur, D. J. Sullivan, J. M. Musser, and M. A. Kehoe.** 1994. Non-congruent relationships between variation in *emm* gene sequences and the population genetic structure of group A streptococci. Mol. Microbiol. **14:**619–631.